

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

---

DOTTORATO DI RICERCA IN STATISTICA E APPLICAZIONI

Ciclo XXV



MARGINAL PARAMETRIZATIONS  
FOR CONDITIONAL INDEPENDENCE  
MODELS  
AND GRAPHICAL MODELS FOR  
CATEGORICAL DATA

PhD Thesis by: Federica Nicolussi

Supervisor:

Chiar.mo Prof.

Roberto Colombi

Anno Accademico 2012/2013





## Contents

Chapter 1. Introduction	7
Log-linear and marginal models	7
Graphical models	9
Contribution of this work	11
The structure of this work	12
1.1. Basic Notions for Contingency Tables	12
1.2. Conditional independence	13
1.3. Models for conditional independencies	15
Chapter 2. Graphical Models	17
2.1. Basic concepts for graphs	17
2.2. Markov Properties	21
Chapter 3. Alternative Markov properties	27
3.1. The new condition (C2*a)	31
3.2. The new condition (C3*b)	35
Appendix to Chapter 3	39
Proof of Lemma 1	39
Proof of Theorem 1	40
Proof of Theorem 3.	42

Chapter 4. Models for categorical data	45
4.1. Log-linear Models	46
4.2. Marginal Models (MM)	49
4.3. Marginal Models and CI models	54
 Chapter 5. Graphical models associated with marginal models	 61
5.1. Four-step Method	61
5.2. Graphical Models with MP I	62
5.3. Graphical Models with MP II	67
5.3.1. Parametrization based on (C3b)	68
5.3.2. Parametrization based on (C2*a) and (C3*b)	73
5.3.3. Mixed parametrization	82
5.3.4. Comparative examples of the three parametrizations proposed for GM II	86
5.4. Graphical Models with MP III	90
5.5. Graphical Models with MP IV	95
 Appendix to Chapter 5	 99
Proof of Theorem 10 (Section 5.3.1)	99
Proof of Lemma 2 (section 5.3.2)	100
Proof of Theorem 11 (section 5.3.2)	101
Proof of Theorem 12 (section 5.3.3)	103
Proof of Theorem 13 (section 5.4)	104
 Chapter 6. Application based on data from the European Values Study (EVS), 2008	 107

CONTENTS

	5
Conclusion	115
Appendix A	117
Glossary of graph terminology	117
Appendix B	121
Bibliography	123

ABSTRACT. The graphical models (**GM**) for categorical data are models useful to representing conditional independencies through graphs. The parametric marginal models for categorical data have useful properties for the asymptotic theory. This work is focused on finding which **GMs** can be represented by marginal parametrizations. Following theorem 1 of Bergsma, Rudas and Németh [9], we have proposed a method to identify when a **GM** is parametrizable according to a marginal model. We have applied this method to the four types of **GMs** for chain graphs, summarized by Drton [22]. In particular, with regard to the so-called **GM** of type II and **GM** of type III, we have found the subclasses of these models which are parametrizable with marginal models, and therefore they are smooth. About the so-called **GM** of type I and **GM** of type IV, in the literature it is known that these models are smooth and we have provided new proof of this result. Finally we have applied the mean results concerning the **GM** of type II on the EVS data-set.

*Key words:* Categorical data, chain graph, hierarchical and complete parametrizations, log-linear parameters, Markov properties, marginal parameters, smoothness.

## CHAPTER 1

### Introduction

**Log-linear and marginal models.** The analysis of discrete data, grouped into contingency tables, was the subject of many research studies in past decades. The probabilistic relationships of independence (dependence) among discrete data are discussed in this work using different tools. The most well-known and commonly used tools are certainly log-linear parameters, which are contrasts of logarithms of probabilities. Models based on log-linear parameters are able to describe the probabilistic relationships of conditional independence between the variables considered in the model. This is possible through linear constraints on the log-linear parameters. There is extensive literature regarding these models, see Goodman (1970, 1971), Cox (1972), Haberman (1974), Andersen (1974), Darroch (1980) and Agresti et al. (1993) among others for details [**32, 33, 16, 3, 19, 2**].

However, the log-linear models have a limit: they cannot easily represent conditional independencies in marginal distributions. McCullagh and Nelder (1989) [**48**] and later Liang, Zeger and Qaqish (1992), Lang and Agresti (1994), Molenberghs and Lesaffre (1994), Glonek and McCullagh (1995) and Colombi (1995) [**45, 40, 49, 31, 12**], discussed this topic extensively, opening the way to the development of models characterized by constraints on marginal distributions. In this contest, Glonek and McCullagh (1995) proposed the multivariate logistic model where the parameters (called multivariate logistic parameters) are the log-linear



interactions evaluated in the marginal contingency table of the set of variables to which the interactions refer. Kauermann (1997) [39] showed the benefits of this last model, pointing out that multivariate logistic models work very well when they must represent marginal independencies of subsets of variables but, on the other hand, they do not fit as well with conditional independencies.

In order to handle this last topic, Bergsma and Rudas (2002) [7] proposed marginal models that are models where the log-linear parameters are defined in marginal distributions and where, once a collection of marginal sets has been selected, these interactions are uniquely defined. In fact, the marginal models of Bergsma and Rudas are the generalization of both classical log-linear models and multivariate logistic models. Their suitable properties led to their large diffusion. First, the marginal models are able to represent many kinds of independence, even the conditional independence of a subset of variables. Secondly, Bergsma and Rudas provided some suitable results about estimating parameters. In particular, they proved that marginal models are always smooth, that is the parametrization has continuous derivatives up to some desired order, it is invertible and its inverse function has continuous derivatives up to some desired order. For more details see [29, 30, 37, 23]. However, a list of independencies cannot always be represented by a marginal model. The simplest example to highlight this problem is represented by the independencies  $A \perp\!\!\!\perp B$  and  $A \perp\!\!\!\perp B|C$ , where  $A$ ,  $B$  and  $C$  are disjoint sets of variables. In fact, there is no marginal model capable of representing these independencies. Bergsma, Rudas and Németh (2010) and Forcina, Lupparelli and Marchetti (2010) [9], [27], showed which lists of independencies are compatible with marginal models.

During the past, the development of research on marginal models has led to generalizations of the marginal models of Bergsma and Rudas using a large range of parameters which are generalizations of log-linear parameters (see Douglas et al. (1990), [21] for the different types of parameters). These new models were introduced by Colombi and Forcina (2001), Bartolucci, Colombi and Forcina (2007) and Cazzaro and Colombi (2008) [14, 6, 11]. Due to these models, two more important aspects are considered. First, by adding inequality constraints over the linear constraints, different situations of dependencies and stochastic orderings can be represented. Secondly, by adding new types of parameters, the relationships between ordinal variables can be described more appropriately.

**Graphical models.** During these same years, many authors started to study models that could provide an immediate representation of the probabilistic structures of the variables. These models use the graphs  $G = \{V, E\}$  which are mathematical objects defined by a set of vertices  $V$  and a set of edges  $E$ . In graphical models every variable is associated with one vertex of the graph, and the presence (absence) of an edge between two variables denotes dependence (independence) between those variables. The rules that are needed to read the list of independencies from a graph are called Markov properties (**MP**) and they change according to the kind of graph. Different kinds of graphs are useful for representing different situations. In particular, graphs with only undirected edges (called undirected graph **UG**), are used to explain situations of conditional independencies, most of the time applied to models of spatial dependence and image analysis. See Hammersley and Clifford (1971) and Darroch et al. (1980)[36, 19]. Graphs with only directed edges, where there are no cycles, are called DAG (Directed Acyclic

Graphs) and they are usually used to describe and to verify dependence relationships among variables. See Pearl (1988), Lauritzen and Spiegelhalter (1988) and Shachter and Kenley (1989) [50, 44, 54]. Chain graphs **CG** are a generalization of the two types of graphs cited above. In these kinds of graphs, in fact, both undirected and directed edges may appear and there are not either directed or semi directed cycles (see Lauritzen (1996)). Obviously, these tools are able to represent both situations of associative and structural dependencies. For an exhaustive discussion of graphical models, see Whittaker (1990), Lauritzen (1996) and Studený (2004) [58, 41, 55].

Over recent years, chain graphs have been studied extensively by Lauritzen and Wermuth (1989), Frydenberg (1990), Andersson, Madigan and Perlman (2001), Wermuth and Cox (2004), Drton (2009) and Marchetti and Lupporelli (2011) [42, 28, 4, 57, 22, 47], who pointed out different Markov properties to read independencies from chain graphs. These authors distinguish four types of graphical models associated with the different Markov properties, each of which is characterized by three conditions. The first describes the macro relationship between the components, and it is common to any type of Markov properties for chain graphs. The second condition evaluates the relationship within a component, and there are two kinds of these conditions. Finally, the third condition refers to the relationship between a group of vertices in a component and a group of vertices in the parent components; there are two kind of conditions in this case as well. The combination of these produces four types, I II III and IV, of Markov properties, summarized in Drton (2009) and in Chapter 2 of this work.

**Contribution of this work.** The main issue, discussed in this work, refers to the connection between graphical models of chain graphs and marginal models. In fact, we investigate which lists of independencies, compatible with a graph, can also be represented by a marginal model. The aim is to profit from both the quality of representation and interpretation of the graphical models and the broad theory that supports marginal models. We deal with this topics applying the theorem 1 of Bergsma, Rudas and Németh (2010), [9], where they provide guidelines to build a marginal parametrization capable of representing the probabilistic relationships described by a list of independencies. In particular, we propose an original method, based on this theorem, that can determine which graphical models can be parametrized according to marginal parameters.

Lauritzen (1996) [41] conducted in-depth studies of graphical models based on the Markov properties proposed by Lauritzen, Wermuth and Frydenberg (called type I), and he showed that these models can always be represented by a marginal model. Furthermore, Drton (2009) and Marchetti and Lupporelli (2011) [22, 47] proved that the same result also holds for graphical models that follow the type IV Markov property. In this work these known results will be presented with a new approach.

Furthermore, in order to analyze the graphical models of type II, we have proposed three different marginal parametrizations. The first is directly derived from the Markov property discussed by Andersson, Madigan and Perlman (2001). The second parametrization is obtained using conditions that are equivalent to the type II Markov property. Finally, the third parametrization, called “mixed”, is a mixture of the previous two. Applying the theorem proposed by Bergsma,

Rudas and Németh (2010) to these three marginal parametrizations we find the subclasses of the graphical model of type II which can be represented by a marginal model, and which are therefore smooth.

Finally, we apply the method to the graphical model of type III (introduced by Drton (2009)). In this case as well, we find a subclass of graphical models of type III which are also marginal models.

**The structure of this work.** In chapter 1, we will propose some basic concepts regarding conditional independence models for categorical variables. In chapter 2, we will take an in-depth look at the graphical models for chain graphs. Instead, in chapter 3 will propose two new conditions to read the independencies between the components of a chain graph. Furthermore, we will prove that these new conditions are equivalent to known Markov properties. Chapter 4 will be focused on the marginal models for categorical variables and on the main results of these models. In particular, section 4.3 will be presented the theorem 1 of Bergsma, Rudas and Németh (2010).

In chapter 5, our method to find which graphical model is also a marginal model will be presented and we will apply this method to the four types of graphical models obtaining the original results mentioned above. Last chapter is reserved to an application of the GM II to the EVS datasets, showing the main results highlighted in Chapter 5.

### 1.1. Basic Notions for Contingency Tables

This work deals with categorical variables that are variables that assume a finite number of categories on both nominal and ordinal scale.

Let us consider  $V_1, \dots, V_q$ , categorical variables taking value in the set  $I_j = \{i_{j1}, \dots, i_{jd_j}\}$ , where  $d_j$  denotes the number's categories of the  $j$ -th variable,  $j = 1, \dots, q$ . We refer to the whole set of variables with the symbol  $V$ . The set  $I = \times_{j=1}^q I_j$  defines a contingency table of  $q$  variables, while the cells of the table are denoted with the vector  $\mathbf{i}$  which contains one category for each variable according to the order of the variables. Let's consider a sample of  $m$  units for which we observe the  $q$  categorical variables. The observed data are displayed in a contingency table  $I$ . We suppose the trials are independent, then the random variable that counts the number of units that have the modalities  $\mathbf{i} \in I$  of the  $q$  variables follows the multinomial distribution with parameters  $m$  and  $\pi_V(\mathbf{i})$ ,  $\mathbf{i} \in I$ . Let us assume that the joint probability distribution function of  $q$  variables  $\pi : I \rightarrow (0, 1]$  is a strictly positive function, which assigns to each cell  $\mathbf{i}$  a probability  $\pi_V(\mathbf{i}) = P(V_j = i_j, j = 1, \dots, q)$ , where  $i_j \in I_j$ . The set of probability distribution function of  $V$  is  $\Pi = \{\pi : \pi_V(\mathbf{i}) > 0, \mathbf{i} \in I, \sum_{\mathbf{i} \in I} \pi_V(\mathbf{i}) = 1\}$ . Assuming that  $A$  and  $B$  are two disjoint subsets of  $V$ , the marginal probability that refers to  $A \subset V$  is  $\pi_A(\mathbf{i}_A) = \sum_{\mathbf{i} \in I_{V \setminus A}} \pi_V(\mathbf{i})$ ,  $\forall \mathbf{i}_A \in I_A$  where  $I_A = \times_{j:(V_j \in A)} I_j$  and the conditional probability of  $A$  given a set  $B$  is  $\pi_{A|B}(\mathbf{i}_A|\mathbf{i}_B) = \frac{\pi_{AB}(\mathbf{i}_{AB})}{\pi_B(\mathbf{i}_B)}$ , where  $\pi_{AB}(\mathbf{i}_{AB})$  denotes the joint probability distribution of  $A \cup B$ .

## 1.2. Conditional independence

The aim of this work is to investigate the relationships of conditional independence within a group of variables.

**Definition.** Let  $V$  be the set of variables  $(V_1, \dots, V_q)$  with probability function  $\pi$  and let  $A, B$  and  $C$ , with  $A, B \neq \emptyset$ , be three disjoint subsets in  $V$ ; then the set  $A$

is *conditionally independent* of  $B$  given  $C$ , and it is denoted with  $A \perp\!\!\!\perp B|C$ , if and only if  $\pi_{AB|C}(\mathbf{i}_{AB}|\mathbf{i}_C) = \pi_{A|C}(\mathbf{i}_A|\mathbf{i}_C)\pi_{B|C}(\mathbf{i}_B|\mathbf{i}_C) \forall \mathbf{i} \in I$ .

When the conditional set  $C$  is empty the independence is called *marginal* and is denoted  $A \perp\!\!\!\perp B$ . Note that marginal independence is not implied by conditional independence, thus if  $A \perp\!\!\!\perp B|C$  holds, it is not necessarily  $A \perp\!\!\!\perp B$ , and if  $A \perp\!\!\!\perp B$ , the independence  $A \perp\!\!\!\perp B|C$  is not necessarily true.

The following statements derive from the definition of conditional independence.

$$(1.2.1) \quad A \perp\!\!\!\perp B|C \iff \pi_{ABC}(\mathbf{i}_{ABC}) = \pi_{AC}(\mathbf{i}_{AC})\pi_{BC}(\mathbf{i}_{BC})/\pi_C(\mathbf{i}_C)$$

$$(1.2.2) \quad A \perp\!\!\!\perp B|C \iff \pi_{A|BC}(\mathbf{i}_A|\mathbf{i}_{BC}) = \pi_{A|C}(\mathbf{i}_A|\mathbf{i}_C)$$

$$(1.2.3) \quad A \perp\!\!\!\perp B|C \iff \pi_{ABC}(\mathbf{i}_{ABC}) = h(\mathbf{i}_{AC})k(\mathbf{i}_{BC})$$

$$(1.2.4) \quad A \perp\!\!\!\perp B|C \iff \pi_{ABC}(\mathbf{i}_{ABC}) = \pi_{A|C}(\mathbf{i}_A|\mathbf{i}_C)\pi_{BC}(\mathbf{i}_{BC})$$

for some  $h(\cdot)$  and  $k(\cdot)$  in (1.2.3).

With the help of these statements it is possible to demonstrate the following properties:

$$\text{(P1): } A \perp\!\!\!\perp B|C \Rightarrow B \perp\!\!\!\perp A|C$$

$$\text{(P2): } A \perp\!\!\!\perp (B, C)|D \Rightarrow A \perp\!\!\!\perp B|D$$

$$\text{(P3): } A \perp\!\!\!\perp (B, C)|D \Rightarrow A \perp\!\!\!\perp B|(C, D)$$

$$\text{(P4): } (A \perp\!\!\!\perp B|D \ \& \ A \perp\!\!\!\perp C|(B, D)) \Rightarrow A \perp\!\!\!\perp (B, C)|D$$

where  $A$ ,  $B$ ,  $C$  and  $D$  are disjoint sets. For more details see [58, 41]. If the probability distribution function is strictly positive, then the following property is also true:

$$\mathbf{(P5)}: (A \perp\!\!\!\perp B|(C, D) \ \& \ A \perp\!\!\!\perp C|(B, D)) \Rightarrow A \perp\!\!\!\perp (B, C)|D$$

For more details see [58, 41, 55].

By means of the properties **P1**, **P2**, **P3**, **P4** and **P5**, it is easy to see that the relationship among variables can be expressed with different lists of independencies. For example, given four variables  $V_1, V_2, V_3, V_4$ , the relationship of conditional independence  $\{V_1 \perp\!\!\!\perp V_2, V_3|V_4\}$  can be replaced with the two following statements:  $V_1 \perp\!\!\!\perp V_2|V_3, V_4$  and  $V_1 \perp\!\!\!\perp V_3|V_2, V_4$ . Applying the property **P5** to the last two statements it is easy to see that we get the first independence.

### 1.3. Models for conditional independencies

The relationships of independence of  $q$  variables can be described by a model of conditional independence, that is a joint probability function  $\pi \in \Pi$  that satisfies a list of  $k$  conditional independencies.

**Definition 1.** Let  $A_i \perp\!\!\!\perp B_i|C_i$ ,  $i = 1, \dots, k$ , be a list of independencies for the variables in  $V$ ; models of conditional independence are given by:

$$(1.3.1) \quad CI_k = \cap_{i=1}^k \{\pi \in \Pi : A_i \perp\!\!\!\perp B_i|C_i(\pi)\},$$

Thus, a list of conditional independencies is the set of constraints that the probability function must satisfy. In chapter 2 we will introduce the graphical models that are models of conditional independence where the list of independencies is obtained by a graph applying rules called Markov properties. Chapter 4 will present the marginal models, which are parametric models which can describe lists of conditional independencies through constraints on interaction parameters. Below are some features of lists of conditional independencies.



**Definition 2.** A list of conditional independencies is called *non redundant* if there is no independence that implies another element of the list.

For example  $V_1 \perp\!\!\!\perp (V_2, V_3) | V_4$  and  $V_1 \perp\!\!\!\perp V_2 | (V_3, V_4)$  is a redundant list because applying the property **P3** it is possible to obtain the second independence from the first.

**Definition.** Two lists of conditional independencies are *equivalent* if it is possible to obtain all elements of one list from the elements of the other list and vice-versa.

## CHAPTER 2

### Graphical Models

Graphical models are a relatively recent tool in the field of statistics, even though there are some applications from as long as the early 1900s. In fact, in recent years, the number of scientific techniques that use them has greatly increased. The main reason for this rapid development is the ability to represent problems in a simplified way. In this section we will use graphical models to represent the relationships of independencies among categorical variables. We will see that there are many types of graphical models, based on the type of graph associated. In particular, there are four types of graphical models associated with chain graphs.

#### 2.1. Basic concepts for graphs

Given three variables, the structure of conditional independence can be checked easily, but as the number of variables increases, the analysis of independence becomes a more complex problem. In these cases, it is necessary to find suitable tools capable of representing situations of independence (and also dependence) in a simple and immediate fashion. These tools are graphs, mathematical structures identified by two sets:  $G = \{V, E\}$ , where  $V$  is the set of finite vertices, or nodes,  $V_1, \dots, V_q$  which represents the  $q$  variables, and  $E$  is the set of edges, or arcs, which represents the relationships of dependence or independence between variables,  $E \subseteq V \times V$ . This section will report the main aspects of the graphs, leaving the detailed list of the definitions for the appendix A. We use the notation presented by Drton

(2009). Section **2.2** will describe the rules for obtaining a list of independencies from a graph.

In order to describe different situations, in the literature, many kind of graphs are used. Here, the more relevant types of graphs will be mentioned with special emphasis on chain graphs.

An *Undirected Graph* (**UG**) is a graph  $G_U = \{V, E\}$  where the edge set  $E$  is composed of undirected arcs, symbolized by  $(-)$ , and it is such that if the couple  $(V_i, V_j)$  belongs to the set  $E$ , then  $(V_j, V_i) \in E$ .

A *Directed Graph* (**DG**) is a graph  $G_D = \{V, E\}$  where the edge set  $E$  is composed of directed arcs, or arrows, symbolized by  $(\rightarrow)$ .

A *directed cycle* is a directed path (a sequence of arrows) that starts and ends in the same vertex.

Let us define a *Directed Acyclic Graph* (**DAG**) as a particular directed graph  $G_{DA} = \{V, E\}$ , with no directed cycles. This kind of graph was studied by Lauritzen et al. [38].

A *Mixed Graph* (**MG**) is a graph  $G_M = \{V, E\}$  such that, the edge set  $E$  is composed by both directed and undirected arcs.

Let a *semi-directed cycle* be a ordered path, composed of directed and undirected arcs, that starts and ends in the same vertex, where the directed arcs preserve the direction.

A *Directed Acyclic Mixed Graph* (**DAMG**) is a mixed graph  $G_{DAM} = \{V, E\}$  with no directed cycles, but with possible semi-directed cycles between the components. This type of graph was studied by Richardson and Spirtes (2003)[53].

A **Chain Graph (CG)** is a mixed graph  $G_C = \{V, E\}$  with no cycles either directed or semi-directed. In a chain graph the nodes in  $V$  can be partitioned in components  $V = T_1 \cup \dots \cup T_s$ , such that the vertices within the same components are joined only by undirected arcs, while the vertices of two different components are linked only by arrows. These components are called **chain components**.

We consider the partial order of the components of the **CG**, where the component  $T_j$  precedes  $T_h$  if there is a direct path from  $T_j$  to  $T_h$  in the graph. The collection of all chain components  $\mathcal{T} = \{T_1; \dots; T_s\}$  is well ordered if  $j < h$  implies that  $T_j$  precedes  $T_h$ .

It's easy to see that a chain graph with only one component is an undirected graph and a chain graph where all components are singleton (set of only one element) is a directed acyclic graph. Indeed, every **CG** is associated with a **DAG**:  $G_D = \{(T_1, \dots, T_s), E_D\}$  where any element of the set of vertices refers to a chain component in  $\mathcal{T}$  and where there is an arrow between two components if there is at least one element of the first component that points to one element in the second component. The following example shows this structure.

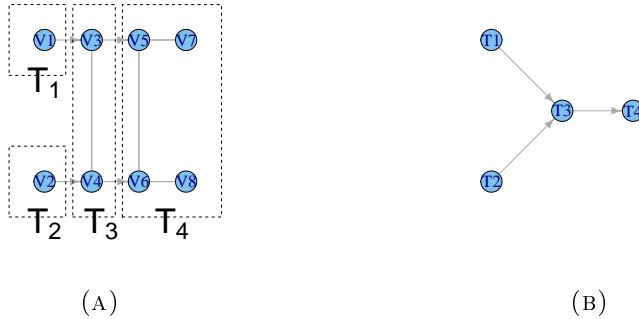


FIGURE 2.1.1. Example of Chain Graph

**Example 1.** Figure (2.1.1.A) shows a chain graph with vertices  $V = \{V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8\}$  and components  $\mathcal{T} = \{T_1; T_2; T_3; T_4\}$  where  $T_1 = V_1$ ,  $T_2 = V_2$ ,  $T_3 = (V_3, V_4)$ ,  $T_4 = (V_5, V_6, V_7, V_8)$ . It is easy to see that the order of class  $\mathcal{T}$  is not unique; in fact, if the first two component  $T_1$  and  $T_2$  are reversed, class  $\mathcal{T}$  is still well ordered. Instead, figure (2.1.1.B) shows the associated DAG in which the vertices are the components in  $\mathcal{T}$ .

A *graphical model* is a representation of a probabilistic model of conditional independence where the independencies are given by a graph. These models make use of graphs to represent probabilistic relationships among variables. The nodes of the graph act as random variables and the edges explain the relationships among variables. In this work we will study graphical models associated with chain graphs because they can represent the dependence structures of both undirected graph and DAG. The absence of the arc between two vertices  $V_i$  and  $V_j$  is a symptom of conditional independence among the variables that these vertices represent. If, on the other hand, two variables are linked by an arc, the situation of conditional dependence among the associated variables depends on whether the arc is directed or undirected. In an intuitive way, undirected arcs represent a symmetrical relationship among variables, on the other hand, the directed arcs emphasize a non-symmetrical relationship (which can often be interpreted as the relationship between response variables and the potential explanatory variables).

The following section describes the rules to get a list of independencies from a graph. These rules are called Markov properties and they change according to whether the graph is undirected or directed. As we will see in the next section, chain graphs combine these two cases.

## 2.2. Markov Properties

Let us consider a chain graph  $G = \{V, E\}$  where any vertex in  $V$  represents a categorical variable. The Markov Properties (**MP**) are rules that make it possible to read a list of independencies from the graph. Generally, given a graph, two variables are in some way independent if there are no edges or arrows between the vertices that represent the variables. In the case of chain graphs, there are four types of Markov properties (see Drton (2009), [22]), each one based on different interpretations of the relationship among variables within the same component and among variables of different components of the graph. The four Markov properties are characterized by the following rules. Let  $G = (V, E)$  be a chain graph with components  $\mathcal{T} = \{T_1, \dots, T_s\}$  and let  $G_D(\mathcal{T}, E_D)$  be the associated directed acyclic graph.

The first rule, shared by any type of **MP**, describes the relationships among the elements of  $G_D$ :

$$(\mathbf{C1}^*): T_h \perp\!\!\!\perp [nd(T_h) \setminus pa_D(T_h)] \mid pa_D(T_h) \quad \forall T_h \in \mathcal{T};$$

This rule specifies that any component is conditionally independent of its previous component, given the parent components.

Marchetti and Lupporelli have shown in Theorem 1 of [47] that this first condition can be rewritten considering the elements of the well ordered class  $\mathcal{T}$ . The alternative condition is:

$$(\mathbf{C1}): T_h \perp\!\!\!\perp \left( \cup_{j=1}^{h-1} T_j \setminus pa_D(T_h) \right) \mid pa_D(T_h), \quad \forall h = 1, \dots, s.$$

In this work we use condition **(C1)** instead of **(C1\*)**.

The following example shows an application of condition **(C1)** to a **CG**. In order to simplify the notation, in the examples the vertices  $V_j$  of the graphs will be replaced with their subscripts  $j$ .

**Example 2.** Referring to the chain graph in figure (2.1.1.A) and to its associated DAG in figure (2.1.1.B), the independencies that hold from the condition **(C1)** are the following:

$$1 \perp\!\!\!\perp 2 \quad (5, 6, 7, 8) \perp\!\!\!\perp (1, 2) \mid (3, 4)$$

The second rule regards variables within the same component  $T_h$ ,  $\forall T_h \in \mathcal{T}$ . There are two kinds of statements involving these relationships. The **(C2a)** is basically the global Markov property for undirected graph applied to any component  $T_h$ , for more details see Lauritzen (1996) [41]. This rule gives the independencies among variables in the same component. A different approach is used in **(C2b)** where, in the conditional set,  $nb(A)$  does not appear, that is a subset of  $T_h$ . This means that the condition **(C2b)** explains the relationships of variables within the same component as marginal independencies (dependencies).

$$\text{(C2a): } A \perp\!\!\!\perp [T_h \setminus Nb(A)] \mid [pa_D(T_h) \cup nb(A)] \quad \forall A \subseteq T_h, \forall h = 1, \dots, s;$$

$$\text{(C2b): } A \perp\!\!\!\perp [T_h \setminus Nb(A)] \mid [pa_D(T_h)] \quad \forall A \subseteq T_h, \forall h = 1, \dots, s.$$

Below is shown how to apply the previous conditions.

**Example 3.** (*Example 2 continued*) The condition **(C2a)** produces the following independencies:

$$\begin{aligned} 5 \perp\!\!\!\perp 8 \mid (3, 4, 6, 7); & \quad 6 \perp\!\!\!\perp 7 \mid (3, 4, 5, 8); \\ 7 \perp\!\!\!\perp (6, 8) \mid (3, 4, 5); & \quad 8 \perp\!\!\!\perp (5, 7) \mid (3, 4, 6). \end{aligned}$$

And from **(C2b)** the following independencies are obtained:

$$\begin{aligned} 5 \perp\!\!\!\perp 8 \mid (3, 4); & \quad 6 \perp\!\!\!\perp 7 \mid (3, 4); \\ 7 \perp\!\!\!\perp (6, 8) \mid (3, 4); & \quad 8 \perp\!\!\!\perp (5, 7) \mid (3, 4). \end{aligned}$$

The lists of independencies generated by both the rules contain redundant elements. Next is an instance of non-redundant lists which express the same relationships of independencies. For the two respective rules the lists are:

$$7 \perp\!\!\!\perp (6, 8) \mid (3, 4, 5); \quad 8 \perp\!\!\!\perp (5, 7) \mid (3, 4, 6).$$

and

$$7 \perp\!\!\!\perp (6, 8) \mid (3, 4); \quad 8 \perp\!\!\!\perp (5, 7) \mid (3, 4).$$

The third condition explains the connection among variables in parents and children components. In this case as well, there are two kinds of Markov properties that differ by the conditional set.

$$\mathbf{(C3a):} \quad A \perp\!\!\!\perp [pa_D(T_h) \setminus pa_G(A)] \mid [pa_G(A) \cup nb(A)] \quad \forall A \subseteq T_h, \forall h = 1, \dots, s;$$

$$\mathbf{(C3b):} \quad A \perp\!\!\!\perp [pa_D(T_h) \setminus pa_G(A)] \mid [pa_G(A)] \quad \forall A \subseteq T_h, \forall h = 1, \dots, s.$$

The lack of  $nb(A)$  in the conditional sets of the **(C3b)** does not consider the influence of other variables belonging to  $T_h$  in the independence statements.



**Example 4.** (*Example 2 continued*) Applying the condition **(C3a)** to the graph in figure (2.1.1.A), we get the following independencies:

$$\begin{array}{lll}
3 \perp\!\!\!\perp 1|(2, 4); & 4 \perp\!\!\!\perp 2|(1, 3); & 5 \perp\!\!\!\perp 4|(3, 6, 7); \\
6 \perp\!\!\!\perp 3|(4, 5, 8); & 7 \perp\!\!\!\perp (3, 4)|5; & 8 \perp\!\!\!\perp (3, 4)|6; \\
(7, 5) \perp\!\!\!\perp 4|(3, 6); & (6, 8) \perp\!\!\!\perp 3|(4, 5); & (7, 8) \perp\!\!\!\perp (3, 4)|(5, 6); \\
(5, 8, 7) \perp\!\!\!\perp 4|(3, 6) & (6, 8, 7) \perp\!\!\!\perp 3|(4, 5). &
\end{array}$$

On the contrary, applying the condition **(C3b)**, we get:

$$\begin{array}{llll}
3 \perp\!\!\!\perp 1|4; & 4 \perp\!\!\!\perp 2|3; & 5 \perp\!\!\!\perp 4|3; & 6 \perp\!\!\!\perp 3|4; \\
7 \perp\!\!\!\perp (3, 4); & 8 \perp\!\!\!\perp (3, 4); & (7, 5) \perp\!\!\!\perp 4|3; & (6, 8) \perp\!\!\!\perp 3|4; \\
(5, 8, 7) \perp\!\!\!\perp 4|3; & (6, 8, 7) \perp\!\!\!\perp 3|4; & (7, 8) \perp\!\!\!\perp (3, 4). &
\end{array}$$

Note that, even in this case, in both the lists there are redundant elements.

The previous relationship of independencies can be expressed through non redundant lists. For instance, from **(C3a)**:

$$\begin{array}{ll}
3 \perp\!\!\!\perp 1|(2, 4); & 4 \perp\!\!\!\perp 2|(1, 3); \\
(5, 8, 7) \perp\!\!\!\perp 4|(3, 6); & (6, 8, 7) \perp\!\!\!\perp 3|(4, 5).
\end{array}$$

and from **(C3b)**:

$$\begin{array}{ll}
3 \perp\!\!\!\perp 1|4; & 4 \perp\!\!\!\perp 2|3; \\
(5, 8, 7) \perp\!\!\!\perp 4|3; & (6, 8, 7) \perp\!\!\!\perp 3|4.
\end{array}$$

The four types of **MP** for chain graphs are built by combining the different kinds of conditions **(C2)** and **(C3)**:

**MP I:** this Markov property is described by **(C1)**, **(C2a)** and **(C3a)** rules, and was introduced by Lauritzen and Wermuth (1989) [43] and Frydenberg (1990) [28]: (LWF).

**MP II:** this Markov property is described by **(C1)**, **(C2a)** and **(C3b)**, and was proposed by Andersson Madigan and Perlman (2001) [4].

**MP III:** this Markov property is described by **(C1)**, **(C2b)** and **(C3a)**, and was studied by Drton (2009)[22].

**MP IV:** this Markov property is described by **(C1)**, **(C2b)** and **(C3b)**, and was proposed by Wermuth and Cox (2004)[57].

The differences between rules of types **a** and **b** are always given by the conditioning event.

It is possible to identify two particular cases according to special structures of **CG**. If all components in  $\mathcal{T}$  are singletons, the **CG** is a **DAG** and, any variable  $V_j$  has no neighbor. In this case for both **(C2)** and **(C3)** conditions **a** and **b** are the same, thus there is only one **MP**. On the other hand, if the graph is composed of only one component, conditions **(C1)** and **(C3)** do not produce any independence, thus the **CG** is a **UG** and the only two possible **MP** are identified respectively by **(C2a)** and **(C2b)**.

Since there are four types of Markov properties, there are also four lists of independencies that can be read by a graph. We discern four types of graphical models according to the Markov properties **MP I**, **MP II**, **MP III** or **MP IV**. Thus, the graphical models of type I, denoted with **GM I**, are conditional independence models where the independencies, read by the chain graph, obey the type I Markov property. The same holds for the other models. Respectively,

graphical models for type II (**GM II**), graphical models for type III (**GM III**) and graphical models for type IV (**GM IV**) are characterized by independencies obeying Markov properties **MP II**, **MP III** and **MP IV**.

It is worth noting that the condition (**C2a**), that refers to the relationship within chain components, is the generalization of the Markov property for undirected graphs. In particular, since the graphical model for undirected graphs are always log-linear models (See Lauritzen [41]), **GM I** and **GM II** use the log-linear approach for the variables within the same component. On the other hand, the condition (**C2b**), is able to represent marginal independencies among the variable within the same component. Thus, in order to represent a list of conditional independencies, **GM I** and **GM II** are preferred to **GM III** and **GM IV**. Instead, when it is useful to describe marginal independencies for variables in the same component, we tend to prefer **GM III** and **GM IV**.

To investigate the relationships among the variables in different chain components, we can interpret the components with no parents as groups of “purely explicative” variables, the components with no children as groups of “purely response” variables and the remaining components as groups of “intervening” variables. The condition (**C3b**) considers any subgroup of the response variables as a function of only its explicative variables. This makes it easier to interpret the results. Instead, with the condition (**C3a**), the relationship among a group of variables in a component  $T_h$  and its parent component, is also dependent on some element of  $T_h$ . This makes **GM II** and **GM IV** more useful.

## CHAPTER 3

### Alternative Markov properties

In the previous chapter we cited four different types of graphical models. For several reasons, discussed in greater depth in chapter 5, it is useful to see which kind of graphical model defines a marginal model (discussed in chapter 4). As we will see, the graphical models that obey **MP I** and **MP IV** are described by different lists of independencies that are always parametrizable by a marginal model (**MM**). This result is not always true for **MP II** and **MP III** graphical models. Chapter 5 is dedicated to investigating when a list of independencies corresponding to a graphical model can be represented by a marginal model. In this chapter we will instead focus on the condition (**C2a**), that appears in **GM I** and in **GM II** and the condition (**C3b**), that appears in **GM II** and in **GM IV**. Specifically, we will propose two equivalent conditions, denoted with (**C2\*a**) and (**C3\*b**) that will aid us in section 5.2. Furthermore, the new (**C3\*b**) produces a no-redundant list of independencies. Notice that, given a graph, a unique not redundant list of independencies does not exist. We then propose some definitions regarding particular structures identifiable in graphs, with the goal of obtaining a new rule. For the sake of brevity, we will report three graphs where it is possible to identify these structures.

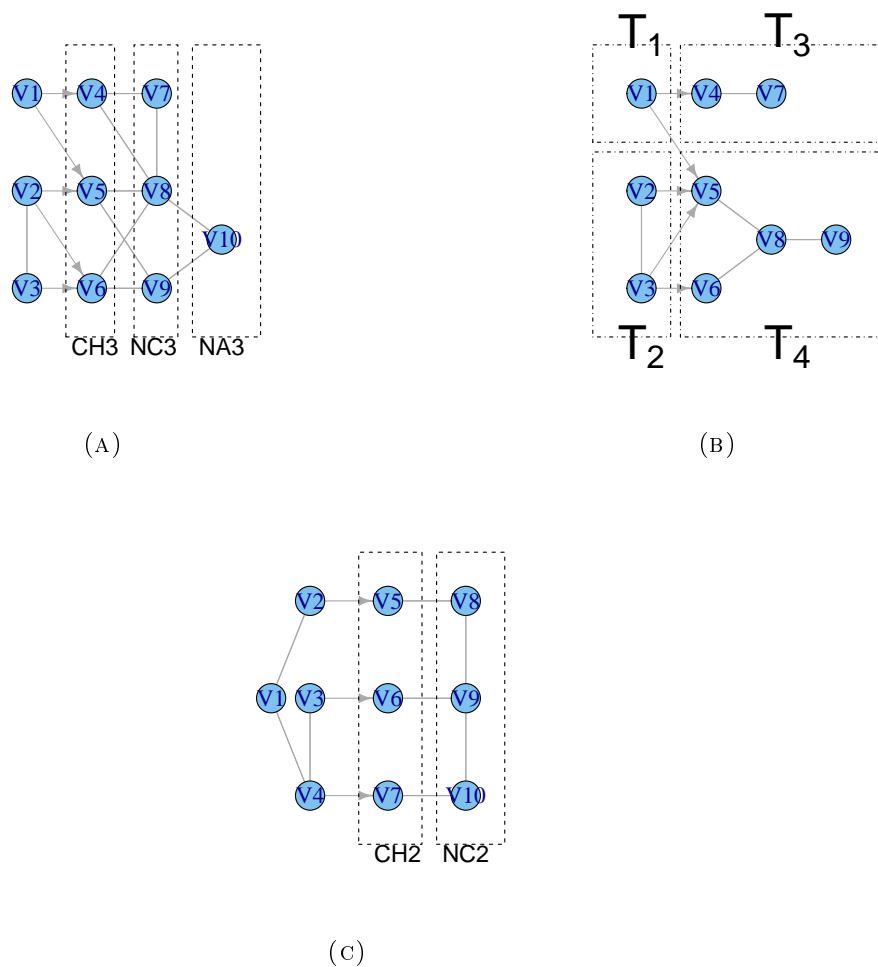


FIGURE 3.0.1. Three examples of graphs with particular structures

**Definition 3.** Given a chain graph  $G$ , any component  $T_h$  can be partitioned in three subsets called  $CH_h$ ,  $NC_h$  and  $NA_h$ , defined as follows:

- The *set of children*  $CH$ , which contains all the elements of  $T$  that have at least one parent. Any element in  $T$  which is the endpoint of at least one arrow.

- The *set of neighbors*  $NC$ , which contains all the elements of  $T$  that do not have parents and are adjacent to at least one element in  $CH$ .
- The *set of non-adjacent vertices*  $NA$ , which contains all the elements of  $T$  that do not have parents and they are not adjacent to any element in  $CH$ .

In order to simplify the notation, from here, we indicate the different elements of a class of sets by using the notation “ $\{A, B; B, C\}$ ” instead of the more laborious “ $\{\{A, B\}, \{B, C\}\}$ ”.

**Example 5.** In figure (3.0.1a), the components  $T_1 = \{1\}$  and  $T_2 = \{2, 3\}$  have no parents, so the partition of these components is formed by  $CH_1 = CH_2 = \{\emptyset\}$  and  $NC_1 = NC_2 = \{\emptyset\}$ ; thus all elements belong to the set of non-adjacent vertices  $NA$ :  $NA_1 = \{1\}$  and  $NA_2 = \{2, 3\}$ . The remaining component  $T_3 = \{4, 5, 6, 7, 8, 9, 10\}$  is the only one with parents:  $pa_D(T_3) = \{1, 2, 3\}$  and so we may distinguish three subsets:  $CH_3 = \{4, 5, 6\}$ ,  $NC_3 = \{7, 8, 9\}$  and  $NA_3 = \{10\}$ .

In figure (3.0.1b), the components  $T_1$  and  $T_2$  are formed only by the set  $NC$ :  $NC_1 = \{1\}$  and  $NC_2 = \{2, 3\}$  for the same reasons seen above. In component  $T_3$  we are able to identify the set  $CH_3 = \{4\}$  and the set  $NC_3 = \{5\}$ . Finally, the component  $T_4$  contains all the sets:  $CH_4 = \{6, 7\}$ ,  $NC_4 = \{8\}$  and  $NA_4 = \{9\}$ .

In figure (3.0.1c), there are two chain components. The first one has only the set  $NA$ :  $NA_1 = \{1, 2, 3, 4\}$  and the component  $T_2$ , as highlighted in the graph, contains the sets  $CH_2 = \{5, 6, 7\}$  and  $NC_2 = \{8, 9, 10\}$ .

**Definition 4.** Given the component  $T_h \in \mathcal{T}$ , the class  $PA_h$  of *sets of parents with the same children* in  $T_h$ , is a partition of  $pa_D(T_h)$  such as, for any element  $\mathcal{A}$  of  $PA_h$ , the vertices  $V_j, V_i \in \mathcal{A}$  iff  $ch(V_j) \cap T_h = ch(V_i) \cap T_h$ , with  $V_i, V_j \in pa_D(T_h)$ .

We consider the elements of this class partially ordered according to the following rule:  $\forall \mathcal{A}, \mathcal{B} \in PA_h$  if  $|ch(\mathcal{B})| < |ch(\mathcal{A})|$  then  $\mathcal{A} < \mathcal{B}$ .

**Example 6.** Let us consider the graph in figure (3.0.1a), the components  $T_1$  and  $T_2$  do not have parents, so  $PA_1 = PA_2 = \emptyset$ . With regard to  $T_3$ , the set of parents with the same children is  $PA_3 = \{1; 2; 3\}$ , in fact,  $ch(1) \cap T_3 = (4, 5)$ ,  $ch(2) \cap T_3 = (5, 6)$  and  $ch(3) \cap T_3 = 6$  which are all different. Note that, in this case,  $PA_3 = \{2; 1; 3\}$  is still well ordered because  $|ch(1) \cap T_3| = |ch(2) \cap T_3| = 2$ .

In figure(3.0.1b), the components  $T_1$  and  $T_2$  do not have parents, thus the class of parents with the same children of  $T_1$  and  $T_2$ , respectively  $PA_1$  and  $PA_2$ , do not exist. With regard to the component  $T_3$ , the class of  $PA_3$  is  $\{1\}$  because the only parent of  $T_3$  is 1. On the other hand,  $PA_4$  is  $\{3; 1, 2\}$  because the vertices 1 and 2 have exactly the same child in  $T_4$ : the vertex 6; the vertex 3 has  $(6, 7)$  as children. Since  $|ch(1, 2)| < |ch(3)|$ , the element  $(1, 2)$  must be preceded by 3 in  $PA_4$ .

Finally, in figure (3.0.1c),  $PA_2 = \{2; 3; 4; 1\}$ . Note that the vertex 1 does not have children, thus must be the last element of  $PA_2$ .

Note that all subsets  $\mathcal{A}$  are composed of variables in  $pa_D(T_h)$  and they have exactly the same children, thus  $\mathcal{A} \subseteq pa_G(ch_G(\mathcal{A}))$ ,  $\forall \mathcal{A} \in PA_h$ .

**Definition 5.** Given the component  $T_h$ , the *in-degree* of a set  $A \subseteq CH$ , denoted by  $d_I(A)$ , is the number of sets of  $PA$  that have  $A$  as child.

**Example 7.** In figure (3.0.1a), the only possible set of children is  $CH_3 = \{4, 5, 6\}$  and the class of parents with the same children is  $PA_3 = \{1; 2; 3\}$ . The in-degree of  $V_4$  is 1 ( $d_I(4) = 1$ ) because  $pa_G(4) = 1$  and it is composed only of one element of  $PA_3$ . The in-degree of  $V_5$ :  $d_I(5) = 2$  because  $pa_G(5) = \{1, 2\}$ . Finally the in-degree of  $V_6$ :  $d_I(6) = 2$  because its parents are  $pa_G(6) = \{2, 3\}$ .

In figure (3.0.1b), the component  $T_3$  has set of children  $CH_3 = \{4\}$  and class of parents with the same children  $PA_3 = \{1\}$ . Obviously  $d_I(4) = 1$ . In component  $T_4$  there are  $CH_4 = \{6, 7\}$  and  $PA_4 = \{3; 1, 2\}$ . The parent set of vertex  $V_6$  is  $pa_G(6) = \{1, 2, 3\}$ , thus its in-degree is  $d_I(6) = 2$ . Similarly  $d_I(7) = 1$ .

In figure (3.0.1c), the set of children of  $T_2$  is  $CH_2 = \{5, 6, 7\}$  and the class of parents is  $PA_2 = \{2; 3; 4; 1\}$ . The in-degrees of the children of component  $T_2$  are respectively  $d_I(5) = 1$ ,  $d_I(6) = 1$  and  $d_I(7) = 1$ .

### 3.1. The new condition (C2\*a)

As condition (C2a) plays an important role in Chapter 5, it is worthwhile to examine some topics regarding this statement in greater depth until to define an alternative condition (C2\*a).

First, this Markov property is the generalization of the global Markov property for an UG, see Lauritzen [41]. This implies that the statement  $A \perp\!\!\!\perp T_h \setminus Nb(A) \mid pa_D(T_h) \cup nb(A)$ ,  $A \subseteq T_h$ , is equivalent to the local Markov property  $V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j)$ , for any  $V_j \in T_h$ .

Furthermore, the following lemma and theorem show another equivalence that will be useful to demonstrate theorems 11 and 12 in Chapter 5:



**Lemma 1.** *Let  $B = B_1 \cup B_2$  be a complete subset of  $T_h$ , where  $V_j \in B_1$  iff  $nb(V_j) = B \setminus V_j$  and  $B_2 = B \setminus B_1$ . Then the following list of independencies:*

$$(3.1.1) \quad B_1 \perp\!\!\!\perp T_h \setminus B \mid pa_D(T_h) \cup B_2$$

$$(3.1.2) \quad (B_1 \cup V_j) \perp\!\!\!\perp T_h \setminus Nb(B_1 \cup V_j) \mid pa_D(T_h) \cup nb(B_1 \cup V_j) \quad \forall V_j \in B_2$$

*is equivalent to the list:*

$$(3.1.3) \quad V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j) \quad \forall V_j \in B.$$

Note that, if there is a set  $B_1$  such that set  $B_2$  is empty, then  $B_1 = B = T_h$ , the component  $T_h$  is complete and there is not any independence. On the other hand, if a set  $B_1 = \emptyset$ , then  $B_2 = B$  and the list in 3.1.2 matches the list in 3.1.3. Let  $\mathcal{Cl}_h$  be the family of clique of the component  $T_h$ , that are the maximal complete sets of the component  $T_h$ .

**Theorem 1.** *Let  $B_i = B_{1i} \cup B_{2i}$  be the cliques of the component,  $B_i \in \mathcal{Cl}_h$ , where  $V_j \in B_{1i}$  iff  $nb(V_j) = B_i \setminus V_j$  and  $B_{2i} = B_i \setminus B_{1i}$ . Then the following list of independencies:*

$$(3.1.4) \quad V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j) \quad \forall V_j \in T_h.$$

*is equivalent to the list:*

$$(3.1.5) \quad B_{1i} \perp\!\!\!\perp T_h \setminus B_i \mid pa_D(T_h) \cup B_{2i} \quad \forall i = 1, \dots, n$$

(3.1.6)

$$(\mathbf{B}_{1,V_j} \cup V_j) \perp\!\!\!\perp T_h \setminus Nb(\mathbf{B}_{1,V_j} \cup V_j) \mid pa_D(T_h) \cup nb(\mathbf{B}_{1,V_j} \cup V_j) \quad \forall V_j \in B_{2i},$$

$\forall i = 1, \dots, n$ . where  $\mathbf{B}_{1,V_j} = (\cup_{B_{1k} \subset nb(V_j)} B_{1k})$  is the union of all sets  $B_{1k}$  that belong to the set  $nb(V_j)$ ,  $\forall V_j \in \cup_{i=1}^n B_{2i}$ .

The proof of this theorem is shown in the Appendix to this chapter.

Now we can define the following condition (**C2\*a**):

**Definition 6.** We consider the class  $B_h^* = \{B_i : B_i \in Cl_h, B_i = B_{1i} \cup B_{2i}, Nb(B_{1i}) = B_i, CH_h \cap B_i \neq \emptyset\}$ . The condition (**C2\*a**) is described by the following list of independencies:

$$(3.1.7) \quad B_{1i} \perp\!\!\!\perp T_h \setminus B_i \mid pa_D(T_h) \cup B_{2i} \quad \forall B_{1i} \subseteq B_i \in B^*$$

(3.1.8)

$$(\mathbf{B}_{1,V_j} \cup V_j) \perp\!\!\!\perp T_h \setminus Nb(\mathbf{B}_{1,V_j} \cup V_j) \mid pa_D(T_h) \cup nb(\mathbf{B}_{1,V_j} \cup V_j) \quad \forall V_j \in T_h \setminus (\cup_{i:B_i \in B^*} B_{1i})$$

$$\mathbf{B}_{1,V_j} = (\cup_{B_{1k} \subset nb(V_j)} B_{1k}).$$

From lemma 1 and theorem 1 the following theorem derives.

**Theorem 2.** *The list of conditional independencies generated by the rule (**C2\*a**) is equivalent to the list of independencies given by (**C2a**).*

Below are reported some exaples to showing this new condition.

**Example 8.** Let us consider the graph in figure 3.0.1a. The family of the cliques of the component  $T_3$  is  $\mathcal{Cl}_3 = \{4, 7, 8; 5, 8; 5, 9; 6, 8; 6, 9; 8, 10; 9, 19\}$ . Secondly we define the sets  $B_i \in \mathcal{Cl}_3$  such that  $B_i \cap CH_3 = \emptyset$ , that are  $B_3^* = \{4, 7, 8; 5, 8; 5, 9; 6, 8; 6, 9\}$ . For all these sets we consider only the sets  $B_{1i}$  such that  $Nb(B_{1i}) = B_i$ , thus we get  $\{4, 7\}$ . For the remaining vertices  $V_j$ , that are  $\{5; 6; 8; 9; 10\}$ , we define the sets  $\mathbf{B}_{1,V_j}$  that are  $\mathbf{B}_{1,5} = \mathbf{B}_{1,6} = \mathbf{B}_{1,9} = \mathbf{B}_{1,10} \emptyset$  and  $\mathbf{B}_{1,8} = 4, 7$ . According the **(C2\*a)** are the following independencies:

$$\begin{array}{lll}
 4, 7 \perp\!\!\!\perp 5, 6, 9, 10 | 1, 2, 3, 8 & 5 \perp\!\!\!\perp 4, 7, 6, 10 | 1, 2, 3, 8, 9 & 6 \perp\!\!\!\perp 4, 7, 5, 10 | 1, 2, 3, 8, 9 \\
 4, 7, 8 \perp\!\!\!\perp 9 | 1, 2, 3, 5, 6, 10 & 9 \perp\!\!\!\perp 4, 7, 8 | 1, 2, 3, 5, 6, 10 & 10 \perp\!\!\!\perp 4, 5, 6, 7 | 1, 2, 3, 8, 9
 \end{array}$$

**Example 9.** Let us consider the graph in figure 3.0.1b. The family of the cliques of the component  $T_4$  is  $\mathcal{Cl}_4 = \{5, 8; 6, 8; ; 8, 9\}$ . Secondly we define the sets  $B_i \in \mathcal{Cl}_4$  such that  $B_i \cap CH_4 = \emptyset$ , that are  $B_4^* = \{5, 8; 6, 8\}$ . For all these sets we consider only the sets  $B_{1i}$  such that  $Nb(B_{1i}) = B_i$ , thus we get  $\{5; 6\}$ . For the remaining vertices  $V_j$ , that are  $\{8; 9\}$ , we define the sets  $\mathbf{B}_{1,V_j}$  that are  $\mathbf{B}_{1,8} = 5, 6$  and  $\mathbf{B}_{1,9} = \emptyset$ . According the **(C2\*a)** are the following independencies:

$$5 \perp\!\!\!\perp 6, 9 | 1, 2, 3, 8 \quad 6 \perp\!\!\!\perp 5, 9 | 1, 2, 3, 8 \quad 9 \perp\!\!\!\perp 5, 6 | 1, 2, 3, 8$$

**Example 10.** Let us consider the graph in figure 3.0.1c. The family of the cliques of the component  $T_2$  is  $\mathcal{Cl}_2 = \{5, 6; 6, 9; ; 7, 10 8, 9; 9 10\}$ . Secondly we define the sets  $B_i \in \mathcal{Cl}_2$  such that  $B_i \cap CH_2 = \emptyset$ , that are  $B_2^* = \{5, 8; 6, 9; 7, 10\}$ . For all these sets we consider only the sets  $B_{1i}$  such that  $Nb(B_{1i}) = B_i$ , thus we get  $\{5; 6; 7\}$ . For the remaining vertices  $V_j$ , that are  $\{8; 9; 10\}$ , we define the sets  $\mathbf{B}_{1,V_j}$  that are  $\mathbf{B}_{1,8} = 5$ ,  $\mathbf{B}_{1,9} = 6$  and  $\mathbf{B}_{1,10} = 7$ . According the (C2\*a) are the following independencies:

$$\begin{array}{lll}
 5 \perp\!\!\!\perp 6, 7, 9, 10 | 1, 2, 3, 4, 8 & 6 \perp\!\!\!\perp 5, 7, 8, 10 | 1, 2, 3, 4, 9 & 7 \perp\!\!\!\perp 5, 6, 8, 9 | 1, 2, 3, 4, 10 \\
 5, 8 \perp\!\!\!\perp 6, 7, 10 | 1, 2, 3, 4, 9 & 6, 9 \perp\!\!\!\perp 5, 7 | 1, 2, 3, 4, 8, 10 & 7, 10 \perp\!\!\!\perp 5, 6, 8 | 1, 2, 3, 4, 9
 \end{array}$$

Note that any element of this list is essential to describe the relationships between the variables, thus this list is non redundant.

### 3.2. The new condition (C3\*b)

In this section we will introduce a new (C3\*b) condition in place of (C3b). First we will propose the new condition (C3\*b) that describes the relationship between a set and its children. After that, it will be shown that this new condition is equivalent to the condition (C3b). One of the advantages of this new condition is that it ensures a non-redundant list of independencies that are *maximal* on the component  $T$ .

**Definition 7.** Given three sets  $A$ ,  $B$  and  $C$  such that  $A \subseteq T_h$  and  $B \not\subseteq T_h$ , the statement of independence  $A \perp\!\!\!\perp B | C$ , is called *maximal “on the component  $T_h$ ”* if it is not possible to add any element of  $T_h$  to the set  $A \subseteq T_h$  while keeping the independence true.

Indeed, the new condition **(C3\*b)** follows this definition.

**Definition 8.** The new condition, concerning the relationships between vertices in different components, is:

$$\mathbf{(C3*b):} \mathcal{A} \perp\!\!\!\perp [T_h \setminus ch(\mathcal{A})] | (pa_D(T_h) \setminus \mathcal{A}), \quad \forall \mathcal{A} \in PA_h \quad \forall h = 1, \dots, s.$$

Note that  $\mathcal{A}$  does not belong to  $T_h$  but to the class of set  $PA_h$ .

**Example 11.** Let us take the figure (3.0.1a). The class of parents with the same children is  $PA_3 = \{1; 2; 3\}$ . From the **(C3\*b)** condition we obtain:

$$1 \perp\!\!\!\perp (6, 7, 8, 9, 10) | (2, 3) \quad 2 \perp\!\!\!\perp (4, 7, 8, 9, 10) | (1, 3) \quad 3 \perp\!\!\!\perp (4, 5, 7, 8, 9, 10) | (1, 2)$$

In the figure (3.0.1b) the class of parents with the same children of  $T_3$  is  $PA = \{1\}$  and the class of parents with the same children of  $T_4$  is  $PA_4 = \{3; 1, 2\}$ . The list of independencies given by **(C3\*b)** is:

$$1 \perp\!\!\!\perp 5 \quad 3 \perp\!\!\!\perp 6, 8, 9 | 1, 2 \quad 1, 2 \perp\!\!\!\perp 7, 8, 9 | 3$$

In figure (3.0.1c) we have the class of parents of  $T_2$  that is  $PA_2 = \{2; 3; 4; 1\}$  and the independencies according to **(C3\*b)** rule are:

$$\begin{aligned} 1 \perp\!\!\!\perp (5, 6, 7, 8, 9, 10) | (2, 3, 4); & \quad 2 \perp\!\!\!\perp (6, 7, 8, 9, 10) | (1, 3, 4); \\ 3 \perp\!\!\!\perp (5, 7, 8, 9, 10) | (1, 2, 4); & \quad 4 \perp\!\!\!\perp (5, 6, 8, 9, 10) | (1, 2, 3). \end{aligned}$$

**Theorem 3.** *The list of conditional independencies generated by the rule **(C3\*b)** is equivalent to the list of independencies given by **(C3b)**.*

The proof of this theorem appears in the appendix to this chapter.

**Example 12.** We saw in Chapter 2 that the graph in figure (2.1.1a) has the following list of independencies obtained by applying the rule **(C3b)**:

$$\begin{array}{cccc} 4 \perp\!\!\!\perp 1|2; & 3 \perp\!\!\!\perp 2|1; & 5 \perp\!\!\!\perp 4|3; & 6 \perp\!\!\!\perp 3|4; \\ 7 \perp\!\!\!\perp (3,4); & 8 \perp\!\!\!\perp (3,4); & (7,5) \perp\!\!\!\perp 4|3; & (6,8) \perp\!\!\!\perp 3|4; \\ (7,8) \perp\!\!\!\perp (3,4); & (6,7,8) \perp\!\!\!\perp 3|4 & & \end{array}$$

Applying the rule **(C3\*b)**, instead, we get the class of parents with the same children for any component, respectively  $PA_1 = \emptyset$ ,  $PA_2 = \emptyset$ ,  $PA_3 = \{V_1; V_2\}$  and  $PA_4 = \{V_3; V_4\}$ . The list of independencies results:

$$1 \perp\!\!\!\perp 4|??2 \quad 2 \perp\!\!\!\perp 3|1 \quad 4 \perp\!\!\!\perp (5,7,8)|3 \quad 3 \perp\!\!\!\perp (6,7,8)|4$$

The first two independencies of both lists are equal. It is the same for the last independence. Furthermore, independence  $3 \perp\!\!\!\perp (6,7,8)|4$ , obtained from **(C3\*b)**, implies  $3 \perp\!\!\!\perp 6|4$  (the fourth independence in the first list),  $3 \perp\!\!\!\perp (6,8)|4$  (the eighth independence) and  $3 \perp\!\!\!\perp (7,8)|4$ . Hence,  $4 \perp\!\!\!\perp (5,7,8)|3$  implies  $4 \perp\!\!\!\perp 5|3$  (the third independence of the first list),  $4 \perp\!\!\!\perp (5,7)$  (the seventh in dependence) and  $4 \perp\!\!\!\perp (7,8)|3$ . Now, applying the property **(P5)** to  $3 \perp\!\!\!\perp (7,8)|4$  and  $4 \perp\!\!\!\perp (7,8)|3$  we obtain  $(3,4) \perp\!\!\!\perp (7,8)$ , (the ninth independence). Finally from  $(3,4) \perp\!\!\!\perp (7,8)$  we get the last two remaining independencies:  $(3,4) \perp\!\!\!\perp 7$  and  $(3,4) \perp\!\!\!\perp 8$ .

**Remark 1.** Under **(C3b)**, all the independencies induced by **(C3\*b)** are maximal on  $T_h$ .

In fact, by adding even just one vertex to the set  $T_h \setminus ch_G(\mathcal{A})$ , this new element necessarily has parents in  $\mathcal{A}$ , and the independence is no longer true.

**Remark 2.** The rule **(C3\*b)** leads to a smaller number of independencies than **(C3b)**.

It is sufficient to note that the cardinality of  $PA$  is lower than the cardinality of  $\mathcal{P}(T)$ , where  $\mathcal{P}(T)$  is called the power set and is the class of all subsets of  $T$ .

**Remark 3.** The list of independencies obtained is not redundant.

Remember that the independence  $A \perp\!\!\!\perp B|C$  implies all independencies  $\alpha \perp\!\!\!\perp \beta|C \cup \gamma$ , where  $\alpha \subseteq A$ ,  $\beta \subseteq B$  and  $\gamma \in \{\mathcal{P}(A \setminus \alpha) \cup \mathcal{P}(B \setminus \beta)\}$ . Since the condition **(C3\*b)** defines one independence for any element  $\mathcal{A} \in PA_h$ , where  $PA_h$  is a particular partition of  $pa_D(T_h)$ , in the list of independencies obtained from **(C3\*b)** the statement  $\mathcal{B} \perp\!\!\!\perp T_h \setminus ch(\mathcal{B}) | pa_D(T_h) \setminus \mathcal{B}$ ,  $\forall \mathcal{B} \subseteq \mathcal{A}$  and  $\forall \mathcal{A} \in PA_h$ , never appears. Hence, even if it is possible that  $T_h \setminus ch(\mathcal{A}) \cap T_h \setminus ch(\mathcal{B}) \neq \emptyset$ , for some  $\mathcal{A}, \mathcal{B} \in PA_h$ , the statement of independencies concerning these two sets refer to a different relationship because  $\mathcal{A} \cap \mathcal{B} = \emptyset$ , and it is impossible to obtain one statement of independence from the other. This shows that any statement in the list of independencies obtained from **(C3\*b)** is essential.

The importance of these properties for a list of independencies will be discussed in Section 5.2.2 where we will propose a parametrization based on condition **(C3\*b)**.

## Appendix to Chapter 3

### Proof of Lemma 1.

**Lemma. 1** *Let  $A = A_1 \cup A_2$  be a complete subset of  $T_h$ , where  $V_j \in A_1$  iff  $nb(V_j) = A \setminus V_j$  and  $A_2 = A \setminus A_1$ . Then the following list of independencies:*

$$A_1 \perp\!\!\!\perp T_h \setminus A \mid pa_D(T_h) \cup A_2$$

$$(A_1 \cup V_j) \perp\!\!\!\perp T_h \setminus Nb(A_1 \cup V_j) \mid pa_D(T_h) \cup nb(A_1 \cup V_j) \quad \forall V_j \in A_2$$

*is equivalent to the list:*

$$V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j) \quad \forall V_j \in A.$$

PROOF. Note that, according to the assumptions of the theorem,  $nb(A_1) = A_2$  and  $nb(A_2) = A_1 \cup nb(A)$ ; hence  $Nb(A_1) = A$  and  $Nb(A_2) = Nb(A) = A \cup nb(A)$ .

Applying the property **P3** to any element of  $A_1$ , the independence in (3.1.1) implies  $V_i \perp\!\!\!\perp T_h \setminus A \mid pa_D(T_h) \cup A_2 \cup A_1 \setminus V_i$ . Since,  $\forall V_j \in A_1$ ,  $nb(V_j) = A \setminus V_j$  and  $Nb(V_j) = A$ , the previous formula can be rewritten as  $V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j)$ ,  $\forall V_j \in A_1$ .

In the same way, from any element of the list in (3.1.2) we get  $V_j \perp\!\!\!\perp T_h \setminus Nb(A_1 \cup V_j) \mid pa_D(T_h) \cup nb(A_1 \cup V_j) \cup A_1$ ,  $\forall V_j \in A_2$ . When  $V_j \in A_2$ ,  $nb(V_j) = A_1 \cup A_2 \setminus V_j \cup (nb(A) \cap$



$nb(V_j)$ ), since  $nb(A_1) = A_2$  and  $nb(V_j \cup A_1) = (nb(V_j) \cup nb(A_1)) \setminus (V_j \cup A_1)$ , we get  $nb(V_j \cup A_1) = (A_2 \cup A_1 \cup A_2 \setminus V_j \cup (nb(A) \cap nb(V_j))) \setminus (V_j \cup A_1) = A_2 \setminus V_j \cup (nb(A) \cap nb(V_j))$ , so  $A_1 \cup nb(A_1 \cup V_j) = A_1 \cup A_2 \setminus V_j \cup (nb(A) \cap nb(V_j)) = nb(V_j)$  and  $Nb(A_1 \cup V_j) = Nb(V_j)$ . Due to the latter considerations, we get  $V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j)$ ,  $V_j \in A_2$ .

On the other hand, we apply the property **P5** to the element of 3.1.3  $\forall V_i \in A_1$ , obtaining 3.1.1. Similarly, applying **P5** to the element of 3.1.3  $\forall V_i \in A_2$  and to 3.1.1, we get the list in 3.1.2.  $\square$

### Proof of Theorem 1.

**Theorem. 1** *Let  $A_i = A_{1i} \cup A_{2i}$  be the cliques of the component,  $A_i \in Cl_h$ , where  $V_j \in A_{1i}$  iff  $nb(V_j) = A_i \setminus V_j$  and  $A_{2i} = A_i \setminus A_{1i}$ . Then the following list of independencies:*

$$V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j) \quad \forall V_j \in T_h.$$

is equivalent to the list:

$$A_{1i} \perp\!\!\!\perp T_h \setminus A_i \mid pa_D(T_h) \cup A_{2i} \quad \forall i = 1, \dots, n$$

$$(\mathbf{A}_{1,V_j} \cup V_j) \perp\!\!\!\perp T_h \setminus Nb(\mathbf{A}_{1,V_j} \cup V_j) \mid pa_D(T_h) \cup nb(\mathbf{A}_{1,V_j} \cup V_j) \quad \forall V_j \in A_{2i},$$

$\forall i = 1, \dots, n$ . where  $\mathbf{A}_{1,V_j} = (\cup_{A_{1k} \subset nb(V_j)} A_{1k})$  is the union of all sets  $A_{1k}$  that belong to the set  $nb(V_j)$ ,  $\forall V_j \in \cup_{i=1}^n A_{2i}$ .

**PROOF.** To begin with, we should highlight some results. First,  $A_{1i} \cap A_{1k} = \emptyset$ ,  $\forall i, k = 1, \dots, n$ ,  $i \neq k$ . In fact, by contradiction, if there was a vertex  $V_j$  such that  $A_{1i} \cap A_{ki} = V_j$ , by the assumptions of the theorem,  $nb(V_j) = A_k \setminus V_j$  and  $nb(V_j) = A_i \setminus V_j$ . Since  $A_i$  is a maximal complete set  $\forall i = 1, \dots, n$ ,  $A_i = A_k$ . Note

that, given a set  $A_i$ , the corresponding set  $A_{1i}$  is unique, hence  $A_{1i} = A_{1k}$  and these infringe the condition  $i \neq k$ . Secondly, given a set  $A_{1i}$ , the corresponding set  $A_{2i}$  such that  $nb(A_{1i}) = A_{2i}$  is unique. This result arises easily from the previous considerations. Finally, given a set  $A_{2i}$ , the set  $nb(A_{2i})$  is equal to  $\mathbf{A}_{1,A_{2i}} \cup nb(A_i) \setminus \mathbf{A}_{1,A_{2i}}$ , where  $\mathbf{A}_{1,A_{2i}} = (\cup_{A_{1k} \subset nb(A_{2i})} A_{1k})$ . In fact, it is easy to derive from the theorem that  $nb(A_{2i}) = A_{1i} \cup nb(A_i)$ , where  $nb(A_i) \neq \emptyset$ . Note that, it can occur that there are some sets  $A_{1k} \subset nb(A_i)$ , for  $k \neq i$ . The set  $\mathbf{A}_{1,A_{2i}}$  contains all these sets in addition to  $A_{1i}$ . Thus,  $nb(A_{2i}) = \mathbf{A}_{1,A_{2i}} \cup nb(A_i) \setminus \mathbf{A}_{1,A_{2i}}$ .

Now we prove that, from the list of independencies obtained by applying (3.1.5) and (3.1.6) we can deduce the list in 3.1.4 and vice versa. Applying the property **P3** to any element of  $A_{1i}$ , the independence in (3.1.5) implies  $V_i \perp\!\!\!\perp T_h \setminus A_i \mid pa_D(T_h) \cup A_{2i} \cup A_{1i} \setminus V_i$ . Since,  $\forall V_j \in A_{1i}$ ,  $nb(V_j) = A_i \setminus V_j$  and  $Nb(V_j) = A_i$ , the previous formula can be rewritten as  $V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j)$ ,  $\forall V_j \in A_{1i}$ .

In the same way, from any element of the list in (3.1.6) we get  $V_j \perp\!\!\!\perp T_h \setminus Nb(\mathbf{A}_{1,V_j} \cup V_j) \mid pa_D(T_h) \cup nb(\mathbf{A}_{1,V_j} \cup V_j) \cup A_{1i}$ ,  $\forall V_j \in A_{2i}$ ,  $\forall i = 1, \dots, n$ . When  $V_j \in A_{2i}$ ,  $nb(V_j) = \mathbf{A}_{1,V_j} \cup A_{2i} \setminus V_j \cup (nb(A_i) \cap nb(V_j)) \setminus \mathbf{A}_{1,V_j}$ , since  $nb(\mathbf{A}_{1,V_j}) = A_{2i}$  and  $nb(V_j \cup \mathbf{A}_{1,V_j}) = (nb(V_j) \cup nb(\mathbf{A}_{1,V_j})) \setminus (V_j \cup \mathbf{A}_{1,V_j})$ , we get  $nb(V_j \cup \mathbf{A}_{1,V_j}) = (A_{2i} \cup \mathbf{A}_{1,V_j} \cup A_{2i} \setminus V_j \setminus (nb(A_i) \cap nb(V_j)) \setminus \mathbf{A}_{1,V_j}) \setminus (V_j \cup \mathbf{A}_{1,V_j}) = A_{2i} \setminus V_j \cup (nb(A_i) \cap nb(V_j))$ , so  $\mathbf{A}_{1,V_j} \cup nb(\mathbf{A}_{1,V_j} \cup V_j) = \mathbf{A}_{1,V_j} \cup A_{2i} \setminus V_j \cup (nb(A_i) \cap nb(V_j)) = nb(V_j)$  and  $Nb(\mathbf{A}_{1,V_j} \cup V_j) = Nb(V_j)$ . Due to the latter considerations, we get  $V_j \perp\!\!\!\perp T_h \setminus Nb(V_j) \mid pa_D(T_h) \cup nb(V_j)$ ,  $V_j \in A_{2i}$ .

On the other hand, we apply the property **P5** to the element of 3.1.4  $\forall V_i \in A_{1i}$ , obtaining 3.1.5. Similarly, applying the **P5** to the element of 3.1.4  $\forall V_i \in A_{2i}$ ,  $\forall i = 1, \dots, n$  and to 3.1.5  $\forall i = 1, \dots, n$ , we get the list in 3.1.6.  $\square$

**Proof of Theorem 3.**

**Theorem. 3.** *The list of conditional independencies generated by the rules (C3\*b) is equivalent to the list of independencies given by (C3b).*

PROOF. We start by showing that the list of independencies generated by (C3\*b) is implied by the independencies of (C3b). Since the (C3b) condition holds for all subsets  $A$  of  $T_h$ , we take  $A = T_h \setminus ch(\mathcal{A})$ . Then, the set  $pa_D(T_h) \setminus pa_G(A)$ , that is the set of  $pa_D(T_h)$  with no children in  $A$ , is  $\mathcal{A} \cup P$ , where  $P = pa_D(T_h) \setminus pa_G(T_h)$  is the set of  $pa_D(T_h)$  with no children. The set of parents of  $A$  becomes  $pa_G(A) = pa_G(T_h \setminus ch(\mathcal{A})) = pa_D(T_h) \setminus (\mathcal{A} \cup P)$ . Thus, for any  $\mathcal{A} \in PA_h$ , a set  $A = T_h \setminus ch(\mathcal{A})$  exists in  $T_h$  so that the independence  $[T_h \setminus ch(\mathcal{A})] \perp\!\!\!\perp \mathcal{A} \cup P \mid (pa_D(T_h) \setminus (\mathcal{A} \cup P))$  holds. Subsequently, we apply the property (P3) and we get  $[T_h \setminus ch(\mathcal{A})] \perp\!\!\!\perp \mathcal{A} \mid (pa_D(T_h) \setminus \mathcal{A})$ .

The second step consists of showing that the list of independencies generated by (C3\*b) implies all independencies generated by (C3b). That is the independencies  $\mathcal{A} \perp\!\!\!\perp [T_h \setminus ch(\mathcal{A})] \mid (pa_D(T_h) \setminus \mathcal{A})$ ,  $\mathcal{A} \in PA_h$  imply the independencies  $A \perp\!\!\!\perp pa_D(T_h) \setminus pa_G(A) \mid pa_G(A)$ ,  $A \subseteq T_h$ .

Let us consider set  $A \subseteq T_h$ , with in-degree  $d_I(A) = (n - r)$ , with  $r = 0, \dots, n$ , where  $n = |PA|$ . This means that in  $PA$  there are exactly  $r$  sets  $\mathcal{A}_i \in PA_h$ ,  $i = 1, \dots, r$ , so that  $A \subseteq T_h \setminus ch(\mathcal{A}_i)$ . Thus  $A \subseteq \cap_{i=1}^r T_h \setminus ch(\mathcal{A}_i) = T_h \setminus (\cup_{i=1}^r ch(\mathcal{A}_i))$ . It follows that  $pa_G(A) = pa_D(T_h) \setminus (\cup_{i=1}^r \mathcal{A}_i)$  and  $pa_D(T_h) \setminus pa_G(A) = (\cup_{i=1}^r \mathcal{A}_i)$ . Now we consider

the statements of independencies related to this sets  $\mathcal{A}_i$ ,  $i = 1, \dots, r$ :

$$\left\{ \begin{array}{l} \mathcal{A}_1 \perp\!\!\!\perp T_h \setminus ch(\mathcal{A}_1) \mid (pa_D(T_h) \setminus \mathcal{A}_1) \\ \dots \\ \mathcal{A}_r \perp\!\!\!\perp T_h \setminus ch(\mathcal{A}_r) \mid (pa_D(T_h) \setminus \mathcal{A}_r) \end{array} \right.$$

which can be rewritten as:

$$\left\{ \begin{array}{l} \mathcal{A}_1 \perp\!\!\!\perp T_h \setminus ch(\mathcal{A}_1) \mid ((pa_D(T_h) \setminus (\cup_{i=1}^r \mathcal{A}_i)) \cup (\cup_{i=1}^r \mathcal{A}_i \setminus \mathcal{A}_1)) \\ \dots \\ \mathcal{A}_r \perp\!\!\!\perp T_h \setminus ch(\mathcal{A}_r) \mid ((pa_D(T_h) \setminus (\cup_{i=1}^r \mathcal{A}_i)) \cup (\cup_{i=1}^r \mathcal{A}_i \setminus \mathcal{A}_r)) \end{array} \right.$$

Since  $A \subseteq T_h \setminus ch(\mathcal{A}_i)$ ,  $\forall i = 1, \dots, r$ , and applying the property **(P2)** we get:

$$\left\{ \begin{array}{l} \mathcal{A}_1 \perp\!\!\!\perp A \mid ((pa_D(T_h) \setminus (\cup_{i=1}^r \mathcal{A}_i)) \cup (\cup_{i=1}^r \mathcal{A}_i \setminus \mathcal{A}_1)) \\ \dots \\ \mathcal{A}_r \perp\!\!\!\perp A \mid ((pa_D(T_h) \setminus (\cup_{i=1}^r \mathcal{A}_i)) \cup (\cup_{i=1}^r \mathcal{A}_i \setminus \mathcal{A}_r)) \end{array} \right.$$

Finally, applying the property **(P5)** we obtain:  $(\cup_{i=1}^r \mathcal{A}_i) \perp\!\!\!\perp A \mid ((pa_D(T_h) \setminus (\cup_{i=1}^r \mathcal{A}_i)))$ .

Thus, from the list of independencies read with the **(C3\*b)** condition we may obtain the list read with the **(C3b)** condition and vice versa.  $\square$



## CHAPTER 4

### Models for categorical data

In the previous chapters we focused on models which could to represent independencies among groups of variables using graphs. Indeed, given a graph, with one of the Markov properties explained in the previous chapter, we are able to obtain a list of independencies. In this section some parametric models for categorical data will be presented which make it possible to study the probabilistic relationships among  $q$  categorical variables, respectively with  $d_1, \dots, d_q$  categories. In section 4.1 we will propose the log-linear models, widely discussed in the literature. For details see Agresti (2002) [1]. These models describe the probabilistic structure of the variables by parameterizing the probability function with a set of log-linear parameters. A log-linear parameter is a contrast of logarithms of probabilities. The main two types of log-linear parameters are the *baseline logits* and contrasts of these or the *local logits* and contrasts of these. The baseline parameters are used for nominal variables, while the local parameters are used for ordinal variables.

In section 4.2 we will describe the marginal models, proposed by Bergsma and Rudas (2002) [7]. These models generalize the log-linear models by defining the log-linear parameters on different marginal distributions. The interactions obtained in this way are contrasts of logarithms of sums of probabilities.

A further generalization of these models is proposed by Bartolucci, Colombi and Forcina (2007), where the parameters, called generalized marginal interactions, are not necessarily log-linear interactions.

### 4.1. Log-linear Models

In this section we deal with basic concepts of log-linear models in order to facilitate understanding of the models defined in chapter 3.2. For a more extensive analysis, see Agresti [1].

Let's  $\mathcal{L}$  be a non empty subset of  $V$ ,  $\mathcal{L} \subseteq V$ ,  $\mathcal{L} \neq \emptyset$ . The baseline log-linear interactions are:

$$(4.1.1) \quad \lambda_{\mathcal{L}}(\mathbf{i}_{\mathcal{L}}) = \sum_{K \subseteq \mathcal{L}} (-1)^{|\mathcal{L} \setminus K|} \log(\pi(\mathbf{i}_K \mathbf{1}_{V \setminus K})) \quad \forall \mathbf{i}_{\mathcal{L}} \in I_{\mathcal{L}}$$

where  $\mathbf{1}_K$  is a vector of ones with dimension  $K$  and  $I_{\mathcal{L}}$  is the table referring to the variable in  $\mathcal{L}$  (see section 1.1). The set  $\mathcal{L}$  is called *interaction set* and it denotes the group of variables to which the parameters  $\lambda_{\mathcal{L}}(\mathbf{i}_{\mathcal{L}})$  refer.

For any interaction set  $\mathcal{L} \subseteq V$  there are  $K_{\mathcal{L}} = \prod_{V_j \in \mathcal{L}} (d_j - 1)$  log-linear parameters defined by (4.1.1). Thus, given  $q$  variables, there are  $(|I| - 1)$  possible parameters. Let  $\boldsymbol{\pi}$  be the vector of probabilities arranged under the lexicographic order, and it is possible to rewrite the formula (4.1.1) in matricial form:

$$(4.1.2) \quad \boldsymbol{\lambda}_{\mathcal{L}} = \mathbf{C}_{\mathcal{L}} \log \boldsymbol{\pi}$$

where  $\mathbf{C}_{\mathcal{L}}$  is a  $(k_{\mathcal{L}} \times |I|)$  matrix of contrasts. All possible parameters are arranged in the  $(|I|-1)$ -dimensional vector  $\boldsymbol{\lambda}$  obtained by stacking the vectors  $\boldsymbol{\lambda}_{\mathcal{L}}, \forall \mathcal{L} \in \mathcal{P}(V)$ , where  $\mathcal{P}(V)$  is called a power set and is the class of all subsets of  $V$ .

In order to build a model it is necessary for the previous parameters to represent a parametrization of the probability function.

We reiterate the definition:

**Definition.** A function  $f(\boldsymbol{\pi}) : \Pi \rightarrow B \subset \mathbb{R}^{|I|-1}$  is a parametrization of  $\boldsymbol{\pi} \in \Pi$  if it is a one-to-one correspondence between  $\Pi$  and  $B$ .

Thus, the inverse function  $f^{-1}(b) : B \rightarrow \Pi$  must exist.

Let  $\lambda$  be the function of  $\boldsymbol{\pi}$  which, applied to any probabilities  $\pi_V(\mathbf{i}) \in \Pi$ , gives the whole set of log-linear parameters  $\boldsymbol{\lambda}$ . Thus  $\lambda(\boldsymbol{\pi}) : \Pi \rightarrow \Lambda \subseteq \mathbb{R}^{|I|-1}$ . This function is a parametrization of the probability function because it is invertible and its inverse function is:

$$(4.1.3) \quad \pi_V(\mathbf{i}) = \frac{\exp(\sum_{\mathcal{L} \subseteq \mathcal{P}(V) \setminus \emptyset} \lambda_{\mathcal{L}}(\mathbf{i}_{\mathcal{L}}))}{\sum_{\mathbf{i} \in I} \exp(\sum_{\mathcal{L} \subseteq \mathcal{P}(V) \setminus \emptyset} \lambda_{\mathcal{L}}(\mathbf{i}_{\mathcal{L}}))} \quad \forall \mathbf{i} \in I, \forall \mathbf{i}_{\mathcal{L}} \in I_{\mathcal{L}}$$

We can rewrite it in matricial form:

$$(4.1.4) \quad \boldsymbol{\pi} = \frac{\exp\{\mathbf{Z}\boldsymbol{\lambda}\}}{\mathbf{1}' \exp\{\mathbf{Z}\boldsymbol{\lambda}\}}$$

where  $\mathbf{Z}$  is the design matrix of elements  $\{0, 1\}$ .

A log-linear model is a parametrization of a probability distribution capable of representing situations of conditional independence. The log-linear model characterized by the whole set of the previous parameters is called a saturated model and it is representative of situations where independencies are lacking. The models



used to describe relationships of independencies are sub-models of these, obtained constraining some parameters to zero. In particular, if the independence  $A \perp\!\!\!\perp B|C$  holds, all parameters  $\lambda_{\mathcal{L}}(\mathbf{i}_{\mathcal{L}})$  are equal to zero, where  $\mathcal{L}$  contains at least one element of  $A$  and at least one element of  $B$ . For the proofs see [1, 58]. Thus, given a conditional independence model  $CI$ , introduced in formula (1.3.1), such that  $\forall(A_i \perp\!\!\!\perp B_i|C_i) \in CI, A_i \cup B_i \cup C_i = V$ , with  $i = 1, \dots, k$ , the log-linear model that satisfies these independencies is described by the set of linear constraints:

$$(4.1.5) \quad \{\lambda_{\mathcal{L}_i} = 0, \forall \mathcal{L}_i \in D_i, i = 1, \dots, k\}$$

where  $D_i$  is the following class:

$$(4.1.6) \quad D_i = \{\mathcal{P}(V) \setminus (\mathcal{P}(A_i \cup C_i) \cup \mathcal{P}(B_i \cup C_i))\}$$

or in the alternative:

$$(4.1.7) \quad D_i = \{\mathcal{L} : \mathcal{L} = a \cup b \cup c, \emptyset \neq a \subseteq A, \emptyset \neq b \subseteq B, c \subseteq C\}.$$

Note that, according to the formula 4.1.5, if  $\lambda_{\mathcal{L}_k} = 0$  then  $\lambda_{\mathcal{L}_j} = 0, \forall \mathcal{L}_j \supset \mathcal{L}_k$ .

A statement of independence may involve a smaller number of variables, so the union of the three sets  $A$ ,  $B$  and  $C$  is strictly a subset of  $V$ . These hypotheses are not easily represented by log-linear models. Bergsma and Rudas (2002) have proposed a more general model which may represent these kinds of independencies. This model will be explained in the next section.

## 4.2. Marginal Models (MM)

As already mentioned, the model proposed by Bergsma and Rudas in (2002) [7] is able to represent a more exhaustive set of independencies. The parameters used are called marginal log-linear parameters, because they are calculated on marginal distributions. Given  $\mathcal{L} \subseteq \mathcal{M} \subseteq V$ , the marginal baseline log-linear parameters, referring to the set of variables in  $\mathcal{L}$  and evaluated on the marginal distribution  $\boldsymbol{\pi}_{\mathcal{M}}$  are:

$$(4.2.1) \quad \eta_{\mathcal{L}}^{\mathcal{M}}(\mathbf{i}_{\mathcal{L}}) = \sum_{\mathcal{K} \subseteq \mathcal{L}} (-1)^{|\mathcal{L} \setminus \mathcal{K}|} \log(\boldsymbol{\pi}_{\mathcal{M}}(\mathbf{i}_{\mathcal{K}} \mathbf{1}_{\mathcal{M} \setminus \mathcal{K}})) \quad \forall \mathbf{i}_{\mathcal{L}} \in I_{\mathcal{L}}$$

where the set  $\mathcal{L}$  denotes the group of variables to which the parameter  $\eta_{\mathcal{L}}^{\mathcal{M}}(\mathbf{i}_{\mathcal{L}})$  refers and it is called an *interaction set*. The set  $\mathcal{M} \subseteq V$ , instead denotes the marginal distribution where the parameter is evaluated and it is called *marginal set*. The vector of all log-linear parameters,  $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$ , referring to the interaction set  $\mathcal{L}$ , is obtained by stacking the parameters  $\eta_{\mathcal{L}}^{\mathcal{M}}(\mathbf{i}_{\mathcal{L}})$  in formula 4.2.1 in lexicographic order.

In order to build the whole list of marginal log-linear parameters it is necessary for the marginal sets and interaction sets to follow some properties proposed by Bergsma and Rudas (2002).

**Definition.** Let  $\mathcal{H} = \{\mathcal{M}_1, \dots, \mathcal{M}_s = V\}$  be a *class of marginal sets*; this class is *hierarchical* if the sets  $\mathcal{M}_1, \dots, \mathcal{M}_s$  are such that  $\mathcal{M}_h \not\subseteq \mathcal{M}_k$  for all  $k = 1, \dots, h$ ,  $h = 2, \dots, s$ .

Given  $s$  marginal set, the hierarchical class always exists but it might not be unique.

**Definition.** Let  $\mathcal{F}_h = \{\mathcal{L}_1, \dots, \mathcal{L}_c\}$  be the *class of interaction sets* assigned to the marginal set  $\mathcal{M}_h$ , with  $\mathcal{L}_j \subseteq \mathcal{M}_h, \forall \mathcal{L}_j \in \mathcal{F}_h$ . The vector of interactions  $\boldsymbol{\eta}$  obtained by stacking the vectors  $\{\boldsymbol{\eta}_{\mathcal{L}_j}^{\mathcal{M}_h}\}_{\mathcal{M}_h \in \mathcal{H}, \mathcal{L}_j \in \mathcal{F}_h}$ , is called **complete** if the union  $\cup_{h=1}^s \mathcal{F}_h$  of all classes of interactions is equal to the power set of  $V$ ,  $\mathcal{P}(V)$  and  $\mathcal{F}_h \cap \mathcal{F}_j = \emptyset, \forall h = 2, \dots, s$  and  $j = 1, \dots, h$ ; thus any interaction set is defined one time. The vector  $\boldsymbol{\eta}$  is called **hierarchical** if any interaction set  $\mathcal{L}_j$  is defined exactly in the first marginal that contains it. It follows that:

- (1)  $\mathcal{F}_1 = \mathcal{P}(\mathcal{M}_1) \setminus \emptyset$ ;
- (2)  $\mathcal{F}_h = \mathcal{P}(\mathcal{M}_h) \setminus \cup_{k < h} \mathcal{F}_k$ .

The parameters of the marginal log-linear model are built in accordance with these two rules.

The next example reports some instances of possible classes of interactions and marginal sets, some of which in accordance with the properties seen above.

**Example 13.** Let  $V = \{V_1, V_2, V_3\}$  the vector of 3 variables, the power set of  $V$  is  $\mathcal{P}(V) = \{\emptyset; V_1; V_2; V_1, V_2; V_3; V_1, V_3; V_2, V_3; V_1, V_2, V_3\}$ . Reported below, are some possible lists of vectors of parameters regarding the three variables:

$$(4.2.2) \quad \boldsymbol{\eta}_{\emptyset}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_1}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_2}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_3}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_3}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_2 V_3}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2 V_3}^{V_1 V_2 V_3}.$$

$$(4.2.3) \quad \boldsymbol{\eta}_{\emptyset}^{\emptyset}, \quad \boldsymbol{\eta}_{V_1}^{V_1}, \quad \boldsymbol{\eta}_{V_2}^{V_2}, \quad \boldsymbol{\eta}_{V_3}^{V_3}, \quad \boldsymbol{\eta}_{V_1 V_2}^{V_1 V_2}, \quad \boldsymbol{\eta}_{V_1 V_3}^{V_1 V_3}, \quad \boldsymbol{\eta}_{V_2 V_3}^{V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2 V_3}^{V_1 V_2 V_3}.$$

$$(4.2.4) \quad \boldsymbol{\eta}_{\emptyset}^{V_1}, \quad \boldsymbol{\eta}_{V_1}^{V_1}, \quad \boldsymbol{\eta}_{V_2}^{V_2 V_3}, \quad \boldsymbol{\eta}_{V_3}^{V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_3}^{V_1 V_2 V_3}, \quad \boldsymbol{\eta}_{V_2 V_3}^{V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2 V_3}^{V_1 V_2 V_3}.$$

$$(4.2.5) \quad \boldsymbol{\eta}_{\emptyset}^{\emptyset}, \quad \boldsymbol{\eta}_{V_1}^{V_1}, \quad \boldsymbol{\eta}_{V_1}^{V_1 V_2}, \quad \boldsymbol{\eta}_{V_2}^{V_1 V_2}, \quad \boldsymbol{\eta}_{V_3}^{V_3}, \quad \boldsymbol{\eta}_{V_1 V_3}^{V_1 V_3}, \quad \boldsymbol{\eta}_{V_2 V_3}^{V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2 V_3}^{V_1 V_2 V_3}.$$

$$(4.2.6) \quad \boldsymbol{\eta}_{\emptyset}^{V_1}, \quad \boldsymbol{\eta}_{V_1}^{V_1 V_2}, \quad \boldsymbol{\eta}_{V_2}^{V_1 V_2}, \quad \boldsymbol{\eta}_{V_3}^{V_3}, \quad \boldsymbol{\eta}_{V_1 V_2}^{V_1 V_2}, \quad \boldsymbol{\eta}_{V_1 V_3}^{V_1 V_3}, \quad \boldsymbol{\eta}_{V_2 V_3}^{V_2 V_3}, \quad \boldsymbol{\eta}_{V_1 V_2 V_3}^{V_1 V_2 V_3}.$$

**Example.** (*Continued*) The parameters in (4.2.2) are the classic log-linear parameters, usually indicated with  $\lambda_A, \forall A \in \mathcal{P}(V)$ ; in (4.2.3) there are parameters called multivariate logistic parameters proposed by Glonek and McCullagh (1995). The first three lists satisfy the properties of both hierarchy and completeness. On the contrary, the parameters in (4.2.5) violate completeness in two ways there is no parameter referred to  $(V_1, V_2)$  jointly, and there are two parameters regarding  $V_1$ . The last set of parameters violates hierarchy, since  $\mathcal{L} = V_1$  is defined in  $\mathcal{M}_2 = (V_1, V_2)$  instead of  $\mathcal{M}_1 = V_1$ . Thus, only the first three lists can be considered marginal log-linear parameters.

It is easy to see that the log-linear model is a special type of marginal log-linear model, with the only marginal set  $V$ . The Glonek and McCullagh model is also a special type of marginal log-linear model. In fact, when  $\mathcal{M} = \mathcal{L}, \forall \mathcal{L} \in \mathcal{P}(V)$  the two models are equal. It is important to note that, given an ordered class of marginal sets, the choice of complete and hierarchical interactions is unique. The vector  $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$  of all marginal log-linear parameters, the elements of which are described by the formula (4.2.1), can be rewritten in matricial form as

$$(4.2.7) \quad \boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} = \mathbf{C}_{\mathcal{L}}^{\mathcal{M}} \log \mathbf{M}_{\mathcal{L}} \boldsymbol{\pi}$$

where  $\mathbf{M}_{\mathcal{L}}$  is a matrix of 0s and 1s whose rows are used to form the marginal distributions and  $\mathbf{C}_{\mathcal{L}}^{\mathcal{M}}$  is a matrix of contrasts, which changes depending on whether the parameters are baseline or local interactions. Colombi and Forcina (2001), [14], in Appendix A, provided an algorithm to build this matrices.

The next example shows a list of possible marginal log-linear parameters.

**Example 14.** Let  $V = \{V_1, V_2, V_3\}$  be a vector of variables of level respectively  $d_1 = d_2 = 2$  and  $d_3 = 3$ . We assume the following hierarchical class of marginal sets  $\mathcal{H} = \{\mathcal{M}_1; \mathcal{M}_2; \mathcal{M}_3\} = \{V_1; V_2, V_3; V_1, V_2, V_3\}$ . According to the properties of completeness and hierarchy, the classes of interaction sets corresponding to the marginal sets are:  $\mathcal{F}_1 = \{\mathcal{L}_1\} = \{V_1\}$ ,  $\mathcal{F}_2 = \{\mathcal{L}_1; \mathcal{L}_2; \mathcal{L}_3\} = \{V_2; V_3; V_2, V_3\}$ ,  $\mathcal{F}_3 = \{\mathcal{L}_1; \mathcal{L}_2; \mathcal{L}_3\} = \{V_1, V_2; V_1, V_3; V_1, V_2, V_3\}$ . The marginal log-linear parameters aggregated under the baseline criterion are:

$$\boldsymbol{\eta} = (\eta_{V_1}^{V_1}, \eta_{V_2}^{V_2 V_3}, \eta_{V_3}^{V_2 V_3}, \eta_{V_2}^{V_2 V_3}, \eta_{V_1}^{V_1 V_2 V_3}, \eta_{V_1}^{V_1 V_2 V_3}, \eta_{V_1}^{V_1 V_2 V_3})'$$

where

$$\eta_{V_1}^{V_1} = \left[ \log\left(\frac{\pi_1(2)}{\pi_1(1)}\right) \right]$$

$$\eta_{V_2}^{V_2 V_3} = \left[ \log\left(\frac{\pi_{23}(2,1)}{\pi_{23}(1,1)}\right) \right]$$

$$\eta_{V_3}^{V_2 V_3} = \left[ \log\left(\frac{\pi_{23}(1,2)}{\pi_{23}(1,1)}\right), \log\left(\frac{\pi_{23}(1,3)}{\pi_{23}(1,1)}\right) \right]'$$

$$\eta_{V_2 V_3}^{V_2 V_3} = \left[ \log\left(\frac{\pi_{23}(1,1)\pi_{23}(2,2)}{\pi_{23}(1,2)\pi_{23}(2,1)}\right), \log\left(\frac{\pi_{23}(1,1)\pi_{23}(2,3)}{\pi_{23}(1,3)\pi_{23}(2,1)}\right) \right]'$$

$$\eta_{V_1 V_2}^{V_1 V_2 V_3} = \left[ \log\left(\frac{\pi_{123}(1,1,1)\pi_{123}(1,2,2)}{\pi_{123}(1,1,2)\pi_{123}(1,2,1)}\right), \log\left(\frac{\pi_{123}(1,1,1)\pi_{123}(1,2,3)}{\pi_{123}(1,1,3)\pi_{123}(1,2,1)}\right) \right]'$$

$$\eta_{V_1 V_3}^{V_1 V_2 V_3} = \left[ \log\left(\frac{\pi_{123}(1,1,1)\pi_{123}(2,1,2)}{\pi_{123}(1,1,2)\pi_{123}(2,1,1)}\right), \log\left(\frac{\pi_{123}(1,1,1)\pi_{123}(2,1,3)}{\pi_{123}(1,1,3)\pi_{123}(2,1,1)}\right) \right]'$$

$$\eta_{V_1 V_2 V_3}^{V_1 V_2 V_3} = \left[ \log\left(\frac{\pi_{123}(2,2,2)\pi_{123}(1,1,2)}{\pi_{123}(2,1,2)\pi_{123}(1,2,2)}\right) - \log\left(\frac{\pi_{123}(2,1,1)\pi_{123}(1,2,1)}{\pi_{123}(2,2,1)\pi_{123}(1,1,1)}\right), \right.$$

$$\left. \log\left(\frac{\pi_{123}(2,2,3)\pi_{123}(1,1,3)}{\pi_{123}(2,1,3)\pi_{123}(1,2,3)}\right) - \log\left(\frac{\pi_{123}(2,1,1)\pi_{123}(1,2,1)}{\pi_{123}(2,2,1)\pi_{123}(1,1,1)}\right) \right]'$$

Once the parameters are defined, it is necessary to show that these parameters are a parametrization of the probability distribution. Thus, the function that gives all marginal parameters  $\eta(\boldsymbol{\pi}) : \Pi \rightarrow B \subseteq \mathbb{R}^{|I|-1}$  must be invertible. Bergsma and Rudas in [7] proved that this function is invertible. This means that the vector  $\boldsymbol{\eta}$  of marginal log-linear parameters, is a parametrization of the joint probability function  $\boldsymbol{\pi}_V$ . Furthermore, they proved that this parametrization has the smoothness property. This result is reported after a brief discussion about the importance of this property.

The smoothness property is one of most important features to require because it guarantees the standard asymptotic theory of the maximum likelihood estimators. Next we will review topics regarding the smoothness of a function. For more details see Geiger and Meek (1998) and Geiger, Heckerman, King and Meek (2001)[29, 30].

**Def.:** The function  $\eta(\boldsymbol{\pi}) : \Pi \rightarrow B \subseteq \mathbb{R}^{|I|-1}$  is **smooth** if it is a diffeomorphism onto  $B$ , that is the function  $\eta(\boldsymbol{\pi}) \in \mathcal{C}^\infty$ , is invertible and its inverse function  $\eta^{-1}(\mathbf{b}) \in \mathcal{C}^\infty$ .

Smoothness is a sufficient condition for the existence of second derivatives of the log-likelihood function, which is necessary to calculate Fisher's information matrix and then for the asymptotic distribution of the estimators. In this regard the following theorem gives a less strong sufficient condition for the asymptotic theory.

**Theorem 4.** (*Inverse function theorem*) *Let  $f$  be a function in  $\mathcal{C}^q$ ,  $q \geq 1$ , from an open set  $B \subset E^n$  into  $E^n$ . If the Jacobian matrix of  $f$  has full rank in  $t_0$ :*

$J(f(t_0)) \neq 0$ , then an open set  $B_0$  exists, containing  $t_0$  such that the inverse  $g$  of  $f$  restricted to  $B_0$  is of class  $C^q$ .

Bergsma and Rudas in [7], Theorem 2, gave the following condition of smoothness for complete and hierarchical models:

**Theorem 5.** *If the vector of marginal log-linear parameter  $\boldsymbol{\eta}$  follows the properties of completeness and hierarchy, then  $\boldsymbol{\eta}$  is a smooth parametrization of  $\Pi$ .*

So any model that can be represented by the above parametrization is smooth. Again, Bergsma and Rudas in [7], theorem 3, provided a necessary condition for smoothness:

**Theorem 6.** *If an interaction set  $\mathcal{L}$  is defined in both  $\mathcal{M}_i$  and  $\mathcal{M}_j$ , then the parametrization  $\boldsymbol{\eta}$  is not smooth.*

As before, conditional independence models are obtained by constraining certain parameters to zero. Unfortunately, as we will see in the next section, not all conditional independence models are representable by marginal models.

### 4.3. Marginal Models and CI models

Similarly to log-linear models, we define a marginal model as the set of probability distributions parametrized by a vector of hierarchical and complete interactions  $\boldsymbol{\eta}$ . If all parameters are not null, then the model is saturated. Marginal models are a generalization of log-linear models since they can parametrize a larger set of conditional independencies because the constraint  $A \cup B \cup C = V$  for  $A \perp\!\!\!\perp B|C$  is

removed. By constraining certain parameters  $\eta_{\mathcal{L}_i}^{\mathcal{M}_i}$  to zero, we are able to represent a list of  $k$  independencies  $A_i \perp\!\!\!\perp B_i | C_i$ , with  $i = 1, \dots, k$ . As before:

$$(4.3.1) \quad A_i \perp\!\!\!\perp B_i | C_i \iff \eta_{\mathcal{L}_i}^{\mathcal{M}_i} = 0 \quad \forall \mathcal{L}_i \in D_i$$

where  $\mathcal{M}_i = A_i \cup B_i \cup C_i$  and

$$(4.3.2) \quad D_i = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(A_i \cup B_i \cup C_i) \setminus (\mathcal{P}(A_i \cup C_i) \cup \mathcal{P}(B_i \cup C_i))\} \quad \forall i = 1, \dots, k.$$

is the class of all null interactions. Note that this class can be rewritten as follows:

$$(4.3.3) \quad D_i = \{\mathcal{L} : \mathcal{L} = a \cup b \cup c, \emptyset \neq a \subseteq A_i, \emptyset \neq b \subseteq B_i, c \subseteq C_i\}$$

Not every list of conditional independencies can be represented by a marginal model, because, given a list of independencies, completeness and hierarchy must be respected. The next example shows the more simple conditional independence model that is not parametrizable with a marginal model.

**Example 15.** Let us consider the independencies  $1 \perp\!\!\!\perp 2$  and  $1 \perp\!\!\!\perp 2 | 3$ . In order to represent the first independence it is necessary to constrain the parameter  $\eta_{12}^{12}$  to zero. According to the second independence, parameters  $\eta_{12}^{123}$  and  $\eta_{123}^{12}$  must be set to zero. Note that the parameter referring to the interaction 12 is twice defined, thus completeness is violated. According to theorem 6, the model of conditional independence is not smooth.

There are cases where completeness and hierarchy are only apparently violated. The next example highlights this situation.



**Example 16.** Let us consider the following list of independencies:  $\{1 \perp\!\!\!\perp 2|3; 1 \perp\!\!\!\perp 3, 4|2\}$ . According to 4.3.1 we get  $\mathcal{M}_1 = \{1, 2, 3\}$  and  $\mathcal{M}_2 = \{1, 2, 3, 4\}$  and according to 4.3.2,  $D_1 = \{1, 2; 1, 2, 3\}$  and  $D_2 = \{1, 3; 1, 4; 1, 3, 4; 1, 2, 3; 1, 2, 4; 1, 2, 3, 4\}$ . Note that, if the interaction sets  $(1, 3)$  and  $(1, 2, 3)$  in  $D_2$  are defined in  $\mathcal{M}_2$  according to 4.3.1 the properties of completeness and hierarchy are violated. But, if we rewrite the list of conditional independencies in the following equivalent way  $\{1 \perp\!\!\!\perp 2|3; 1 \perp\!\!\!\perp 3|2; 1 \perp\!\!\!\perp 4|2, 3\}$ , sets  $D_i$  become  $D_1 = \{1, 2; 1, 2, 3\}$ ,  $D_2 = \{1, 3; 1, 2, 3\}$  which must be correctly defined as  $\mathcal{M}_1 = 1, 2, 3$  and  $D_2 = \{1, 4; 1, 3, 4; 1, 2, 4; 1, 2, 3, 4\}$  that must be correctly defined as  $\mathcal{M}_2 = \{1, 2, 3, 4\}$ .

Bergsma, Rudas and Németh (2010) (hereinafter referred to as BRN) and Forcina, Lupparelli and Marchetti (2010) (hereinafter referred to as FLM) have discussed the problems described in the previous example.

In particular, BRN consider a list of  $k$  independencies  $A_i \perp\!\!\!\perp B_i|C_i$ ,  $i = 1, \dots, k$  concerning the variables in  $V$ , and they define  $\forall \mathcal{L} \in \mathcal{P}(V)$ , marginal set  $\mathcal{M}(\mathcal{L})$ , that is the first marginal set in the hierarchical class  $\mathcal{H} = \{\mathcal{M}_1, \dots, \mathcal{M}_m\}$  which contains the interaction  $\mathcal{L}$ . Then the BRN's theorem 1, describes when the list of  $k$  independencies is a marginal model.

**Theorem 7.** *If  $\forall \mathcal{L} \in D_i$  the next condition is satisfied*

$$(4.3.4) \quad C_i \subseteq \mathcal{M}(\mathcal{L}) \subseteq (A_i \cup B_i \cup C_i), \quad i = 1, \dots, k$$

*then the following statements hold:*

- The conditional independence model  $CI_k = \cap_{i=1}^k \{\pi \in \Pi : A_i \perp\!\!\!\perp B_i|C_i(\pi)\}$ ,  
if and only if

$$\eta_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})} = 0, \quad \forall \mathcal{L} \in \cup_{i=1}^k D_i.$$

- The model  $CI_k$  can be parametrized by

$$(4.3.5) \quad \{\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{i=1}^k D_i\}$$

which is a complete and hierarchical marginal parametrization;

- The number of degrees of freedom of  $CI_k$  is

$$\sum_{\mathcal{L} \in \cup_{i=1}^k \mathbb{D}_i} \prod_{j \in \mathcal{L}} (d_j - 1).$$

Note that, according to this theorem, given the conditional independence model, the marginal class is not unique. But, since the interaction  $\mathcal{L} = A_i \cup B_i \cup C_i \in D_i$ ,  $\forall i = 1, \dots, k$ , then the marginal sets  $\mathcal{M}_i = A_i \cup B_i \cup C_i$  will be always defined.

The next two examples show how to apply this theorem.

**Example 17.** Let us consider the following list of independencies:  $\{(1 \perp\!\!\!\perp 2); (1 \perp\!\!\!\perp 2|3)\}$  proposed in example 15. BRN define the following classes about any independence:  $D_1 = \{1, 2\}$  and  $D_2 = \{1, 2; 1, 2, 3\}$ . We define the marginal sets considering, for all independencies, the set  $\mathcal{M}_i = A_i \cup B_i \cup C_i$ . Thus, we obtain the hierarchical class of marginal sets  $\mathcal{H} = \{1, 2; 1, 2, 3\}$ . Now, according to the first statement of independence  $1 \perp\!\!\!\perp 2$  we must constrain all parameters  $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$  to zero, where  $\mathcal{L} \in D_1$ , so  $\boldsymbol{\eta}_{12}^{12} = \mathbf{0}$ . Since  $\mathcal{M}(\mathcal{L}) = A \cup B \cup C$ , the requirements of theorem 7 are satisfied. Similarly, according to the second independence  $1 \perp\!\!\!\perp 2|3$  we have  $\boldsymbol{\eta}_{12}^{12} = \mathbf{0}$  and  $\boldsymbol{\eta}_{123}^{123} = \mathbf{0}$ . The first constrain does not satisfy the condition of theorem 7  $C \notin \mathcal{M}(\mathcal{L}) \subseteq A \cup B \cup C$ , where, in this case  $C = 3$ . We may conclude that there are no marginal parametrizations that describe this list of independencies, as we saw in example 15.

**Example 18.** Let us consider the following independencies  $\{1 \perp\!\!\!\perp 2|3; 1 \perp\!\!\!\perp 3, 4|2\}$ . According to BRN, for all independencies we consider the marginal sets  $\mathcal{M}_i = A_i \cup B_i \cup C_i$ . Thus, we obtain the hierarchical class of marginal sets  $\mathcal{H} = \{1, 2, 3; 1, 2, 3, 4\}$ . Otherwise, the classes of null parameters are  $D_1 = \{1, 2; 1, 2, 3\}$  and  $D_2 = \{1, 3; 1, 4; 1, 3, 4; 1, 2, 3; 1, 2, 4; 1, 2, 3, 4\}$ . According to  $1 \perp\!\!\!\perp 2|3$  we must constrain the parameters  $\eta_{12}^{123}$  and  $\eta_{123}^{123}$  to zero and, according to  $1 \perp\!\!\!\perp 3, 4|2$ , we must constrain the parameters  $\eta_{13}^{1234}$ ,  $\eta_{123}^{1234}$ ,  $\eta_{14}^{1234}$ ,  $\eta_{134}^{1234}$ ,  $\eta_{124}^{1234}$  and  $\eta_{1234}^{1234}$  to zero. Note that the first two sets of parameters are defined in a smaller marginal set than  $\mathcal{M} = A \cup B \cup C = 1, 2, 3, 4$ . But, since  $C \subseteq \mathcal{M}(\mathcal{L}) \subseteq A \cup B \cup C$ , where, in this case  $C = 2$ , then the condition of theorem 7 is satisfied, thus the marginal model described by  $\{\eta_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}\}_{\mathcal{L} \in \cup D_i; \mathcal{M} \in \mathcal{H}}$ , is able to represent the previous set of independencies.

On the other hand, FLM (2010) consider one independence at time. For any statement  $A \perp\!\!\!\perp B|C$  belonging to the conditional independence model  $CI_k$ , they select the following marginal sets:

$$(4.3.6) \quad \mathcal{M}_j = A_j \cup B_j \cup C, \quad j = 1, \dots, m$$

where  $A_j \subseteq A$  and  $B_j \subseteq B$ , and where  $m = 2^{(|A|-1)}2^{(|B|-1)}$ . We also collect these marginal sets in the hierarchical class  $\mathcal{H} = \{\mathcal{M}_j\}_{j=1, \dots, m}$ .

They considered the classes  $\mathcal{R}_j$ ,  $j = 1, \dots, m$  defined as

$$(4.3.7) \quad \mathcal{R}_j = \{\mathcal{L} : \mathcal{L} = a \cup b \cup c, \emptyset \neq a \subseteq A_j, \emptyset \neq b \subseteq B_j, c \subseteq C, \mathcal{L} \notin R_l, l < j\}$$

Note that the  $\cup_{j=1}^m \mathcal{R}_j$  for any independencies is equal to class in formula 4.3.2.

FML have proved the following theorem.

**Theorem 8.** *Within the family of hierarchical and complete marginal log-linear parametrization, the condition that  $\boldsymbol{\eta}_{\mathcal{R}_j}^{\mathcal{M}_j} = \mathbf{0}$  for all  $j = 1, \dots, m$  is necessary and sufficient for  $A \perp\!\!\!\perp B|C$  to hold.*

According to this theorem, FLM proposed an algorithm to find, if possible, a marginal parametrization for a list of independencies.

In the next chapter we will analyze the four types of graphical models presented in Chapter 2, in order to determine which conditions these models can be parametrized to using marginal models. In particular we will propose a four-step method based on BRN's theorem 7 and we will apply this method to the different graphical models.



## CHAPTER 5

### Graphical models associated with marginal models

The aim of this chapter is to find which models of conditional independence are both graphical models and marginal models. In particular, given a list of independencies compatible with a graph, we determine whether this list can be parametrized by a marginal model. With this aim, we apply theorem 7 to the list of independencies obtained from a graphical model. The result is a four-step method as explained in section 5.1.

In Sections 5.2 and 5.5 we will provide a review of known results for **GM I** and **GM IV**. Section 5.3 is dedicated to investigating **GM II**. We will propose three marginal parametrizations, each of which is able to represent a sub-class of **GM II**. In particular, subsection 5.3.3 shows a better parametrization for **GM II**. Section 5.4 is dedicated to **GM III**.

#### 5.1. Four-step Method

Given a list of independencies  $\{A_i \perp\!\!\!\perp B_i | C_i\}_{i=1,\dots,k}$  from a graphical model, our four-step method consists of:

- Step 1* We define the hierarchical class of marginal sets  $\mathcal{H} = \{\mathcal{M}_1, \dots, \mathcal{M}_k\}$  which must contains at least the sets  $\mathcal{M}_i = A_i \cup B_i \cup C_i$ ,  $\forall A_i \perp\!\!\!\perp B_i | C_i$ .
- Step 2* We define the set of hierarchical and complete parameters associated with the previous marginal class  $\{\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}}$ .

*Step 3* We define the classes of null interactions  $D_i$  according to the formula 4.3.2.

*Step 4* We check whether condition  $C_i \subseteq \mathcal{M}(\mathcal{L}) \subseteq (A_i \cup B_i \cup C_i)$ , of theorem 7, holds  $\forall \mathcal{L} \in D_i, \forall i = 1, \dots, h$ .

Note that class  $\mathcal{H}$  depends on the list of the independencies, and that different  $\mathcal{H}$  classes can correspond to equivalent lists of independencies.

If the requirement at the fourth step is satisfied, then the marginal parametrization proposed is capable of representing the graphical model.

Lauritzen, in [41], showed that any type I graphical model has a log-linear structure. It was also known that **GM IV** can be expressed through marginal models (see [22] and [47]). With this method we will provide new proofs to known results regarding type I and IV graphical models. Furthermore, we will find subclasses of type II and III graphical models that can always be represented by marginal parametrizations.

## 5.2. Graphical Models with MP I

Given a CG, **GM I** is a set of probability functions that satisfies the list of independencies generated by properties **(C1)**, **(C2a)** and **(C3a)**. We want to investigate if there is a marginal parametrization that is capable of representing the list of independencies of a GM I. As mentioned in the introduction to this chapter we will apply the method described in the previous chapter.

In the *first step*, we define the class of marginal sets  $\mathcal{H} = \{A_i \cup B_i \cup C_i, i = 1, \dots, k\}$ , for any independence  $A_i \perp\!\!\!\perp B_i | C_i$ . In order to consider the three conditions **(C1)**, **(C2a)** and **(C3a)** we use the following marginal sets:

- In condition **(C1)**, the sets of variables involved are  $T_h \cup (\cup_{j=1}^{h-1} T_j \setminus pa_D(T_h)) \cup pa_D(T_h) = \cup_{j=1}^h T_j$ ,  $\forall h = 1, \dots, s$ . Thus the first kind of marginal sets is given by:  $\mathcal{M}_h^1 = \cup_{j \leq h} T_j$ ,  $h = 1, \dots, s$ .
- In condition **(C2a)**, the sets involved are:  $A \cup [T_h \setminus Nb(A)] \cup nb(A) \cup pa_D(T_h) = T_h \cup pa_D(T_h)$ ,  $\forall h = 1, \dots, s$ , and  $\forall A \subseteq T_h$ . Thus, the marginal sets associated to the second condition are  $\mathcal{M}_h^{2a} = T_h \cup pa_D(T_h)$ ,  $\forall h = 1, \dots, s$ .
- The third condition **(C3a)** is  $A \perp\!\!\!\perp pa_D(T_h) \setminus pa_G(A) \mid (nb(A) \cup pa_G(A))$ . Note that,  $\forall A \subseteq T_h$ , according to **(C2a)**, there is also the independence  $A \perp\!\!\!\perp T_h \setminus Nb(A) \mid (pa_D(T_h) \cup nb(A))$ , which we can rewrite as  $A \perp\!\!\!\perp T_h \setminus Nb(A) \mid (pa_G(A) \cup pa_D(T_h) \setminus pa_G(A) \cup nb(A))$ . Now, applying the property **(P4)**, we get that,  $\forall A \in \mathcal{P}(T_h)$ ,  $A \perp\!\!\!\perp (T_h \setminus Nb(A) \cup pa_D(T_h) \setminus pa_G(A)) \mid (pa_G(A) \cup nb(A))$ . This independence, according to the property **(P5)** is equivalent to the following two:

$$\begin{aligned} A \perp\!\!\!\perp pa_D(T_h) \setminus pa_G(A) \mid pa_G(A) \cup T_h \setminus A \\ A \perp\!\!\!\perp T_h \setminus Nb(A) \mid (pa_G(A) \cup nb(A)) \end{aligned}$$

$\forall A \subseteq T_h$ . Note that the second independence is exactly **(C2a)**, while the first independence involves all variables in  $T_h \cup pa_D(T_h)$ . Thus, no new marginal set is needed in addition to  $\mathcal{M}_h^{2a}$  and  $\mathcal{M}_h^1$ ,  $h = 1, \dots, s$ .

The marginal sets above defined, are in accordance with the relationship  $\mathcal{M}_h^{2a} \subseteq \mathcal{M}_h^1$ ,  $\forall h = 1, \dots, s$ . The hierarchical class  $\mathcal{H}_I$  for the **MP I** models is provided by

$$(5.2.1) \quad \mathcal{H}_I = \{(\mathcal{M}_h^{2a}; \mathcal{M}_h^1)\}_{h=1, \dots, s}.$$

If  $\mathcal{M}_h^{2a} = \mathcal{M}_h^1$ , we use to retain only the set  $\mathcal{M}_h^{2a}$ .



Once the class of marginal sets has been defined, in the *second step*, we introduce the associated hierarchical and complete parametrization.

**Example 19.** Let us consider the graph in figure (2.1.1.A). According to **MP I** the marginal sets  $\mathcal{M}_h^1$  are  $\{1; 1, 2; 1, 2, 3, 4; 1, 2, 3, 4, 5, 6, 7, 8\}$  and the marginal sets  $\mathcal{M}_h^{2a}$  are  $\{1; 2; 1, 2, 3, 4; 3, 4, 5, 6, 7, 8\}$ . So the class  $\mathcal{H}_I$  is  $\mathcal{H}_I = \{\mathcal{M}_1^{2a} = 1; \mathcal{M}_2^{2a} = 2; \mathcal{M}_2^1 = 1, 2; \mathcal{M}_3^{2a} = 1, 2, 3, 4; \mathcal{M}_4^{2a} = 3, 4, 5, 6, 7, 8; \mathcal{M}_4^1 = 1, 2, 3, 4, 5, 6, 7, 8\}$ . In the marginal sets we allocate the interaction sets respecting the hierarchy and completeness properties.

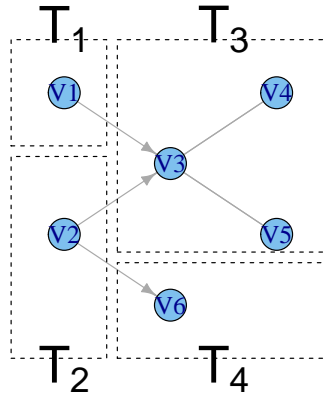


FIGURE 5.2.1

**Example 20.** Let's consider the graph in figure 5.2.1. In this case the class of components is not univocal. We choose the order  $\mathcal{T} = \{T_1, T_2, T_3, T_4\}$ . The two kinds of marginal sets  $\mathcal{M}_h^1$  and  $\mathcal{M}_h^{2a}$  are respectively  $\mathcal{M}_1^1 = (1)$ ,  $\mathcal{M}_2^1 = (1, 2)$ ,  $\mathcal{M}_3^1 = (1, 2, 3, 4, 5)$ ,  $\mathcal{M}_4^1 = (1, 2, 3, 4, 5, 6)$ ,  $\mathcal{M}_1^{2a} = (1)$ ,  $\mathcal{M}_2^{2a} = (2)$ ,  $\mathcal{M}_3^{2a} = (1, 2, 3, 4, 5)$  and  $\mathcal{M}_4^{2a} = (2, 6)$ . So, the class of marginal sets is:  $\mathcal{H}_I = \{1; 2; 1, 2; 1, 2, 3, 4, 5; 2, 6; 1, 2, 3, 4, 5, 6\}$ .

In the *third step*, we define the sub classes of interaction sets  $D_i = \{\mathcal{P}(A_i \cup B_i \cup C_i) \setminus \mathcal{P}(A_i \cup C_i) \cup \mathcal{P}(B_i \cup C_i)\}$ , according to the formula (4.3.2). Thus:

- Applying the formula (4.3.2) to the condition **(C1)** we get the class:

$$(5.2.2) \quad D_h^1 = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(\cup_{j=1}^h T_j) \setminus (\mathcal{P}(T_h \cup pa_D(T_h)) \cup \mathcal{P}(\cup_{j=1}^{h-1} T_j))\}$$

$\forall h = 1, \dots, s.$

- For the condition **(C2a)** we get  $D_h^{2a} = \cup_{A \subseteq T_h} D_{h,A}^{2a}$ , where for any  $A \subseteq T_h$

$$(5.2.3) \quad D_{h,A}^{2a} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \cup pa_D(T_h)) \setminus (\mathcal{P}(Nb(A) \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus A \cup pa_D(T_h)))\}$$

and  $\forall h = 1, \dots, s.$

- According to the independence  $A \perp\!\!\!\perp pa_D(T_h) \setminus pa_G(A) \mid pa_G(A) \cup T_h \setminus A$ , obtained from **(C3a)**, we have  $D_h^{3a} = \cup_{A \subseteq T_h} D_{h,A}^{3a}$  for any  $A \subseteq T_h$

$$(5.2.4) \quad D_{h,A}^{3a} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \cup pa_D(T_h)) \setminus \mathcal{P}(T_h \cup pa_G(A)) \cup \mathcal{P}(T_h \setminus A \cup pa_D(T_h))\}$$

and  $\forall h = 1, \dots, s.$

**Example 21.** Let us consider the graph in figure 5.2.1. According to condition **(C1)** we must set to zero all parameters which refer to the elements of the classes  $D_2^1 = \{1, 2\}$ ,  $D_4^1 = \{3, 6; 4, 6; 5, 6; 3, 4, 6; 3, 5, 6; 4, 5, 6; 3, 4, 5, 6; 1, 3, 6; 1, 4, 6; 1, 5, 6; 1, 3, 4, 6; 1, 3, 5, 6; 1, 4, 5, 6; 1, 3, 4, 5, 6; 2, 3, 6; 2, 4, 6; 2, 5, 6; 2, 3, 4, 6; 2, 3, 5, 6; 2, 4, 5, 6; 2, 3, 4, 5, 6; 1, 2, 3, 6; 1, 2, 4, 6; 1, 2, 5, 6; 1, 2, 3, 4, 6;$

1, 2, 3, 5, 6; 1, 2, 4, 5, 6; 1, 2, 3, 4, 5, 6}. According to condition **(C2a)** the parameter to set to zero refer to the interaction set in  $D_3^{2a} = \{4, 5; 1, 4, 5; 2, 4, 5; 1, 2, 4, 5\}$ . Finally, according to the **(C3a)** we get class  $D_3^{3a} = \{1, 4; 2, 4; 1, 2, 4; 1, 5; 2, 5; 1, 2, 5; 1, 4, 5; 2, 4, 5; 1, 2, 4, 5; 1, 3, 4; 2, 3, 4; 1, 2, 3, 4; 1, 3, 5; 2, 3, 5; 1, 2, 3, 5; 1, 3, 4, 5; 2, 3, 4, 5; 1, 2, 3, 4, 5\}$ .

As a *final step*, we must determine whether the condition 4.3.4 of theorem 7 is satisfied. In this case, we have three requirements to satisfy:

$$(5.2.5) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (\cup_{j=1}^h T_j), \quad \forall \mathcal{L} \in D_h^1.$$

It is easy to see that the first marginal set  $\mathcal{M}(\mathcal{L})$  in  $\mathcal{H}_I$  that contains any element  $\mathcal{L}$  in  $D_h^1$  is  $\mathcal{M}_h^1 = (\cup_{j=1}^h T_j)$ .

$$(5.2.6) \quad pa_D(T_h) \cup nb(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h) \quad \forall \mathcal{L} \in D_{h,A}^{2a}, \quad \forall A \subseteq T_h.$$

In this case,  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_h^{2a} = T_h \cup pa_D(T_h)$ , for all  $\mathcal{L} \in D_h^{2a}$ .

$$(5.2.7) \quad pa_G(A) \cup nb(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h) \quad \forall \mathcal{L} \in D_{h,A}^{3a}, \quad \forall A \subseteq T_h$$

Even in this case,  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_h^{2a} = T_h \cup pa_D(T_h)$ , for all  $\mathcal{L} \in D_h^{3a}$ .

$\forall h = 1, \dots, s$ . From these considerations it is easy to see that the following theorem holds.

**Theorem 9.** *A graphical model of type I is always a marginal model of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3a}), \mathcal{M} \in \mathcal{H}_I\}$ .*

From this result we can conclude that a graphical model I is smooth.

### 5.3. Graphical Models with MP II

The issue of which **GM II** are smooth is still open. Drton (2009) [22] showed that **GM II** are not always smooth proving that the graph in figure 5.3.1 did not have this property. In fact, the graph 5.3.1 describes the list of independencies  $\{1 \perp\!\!\!\perp 2, 4; 2 \perp\!\!\!\perp 4 | 1, 3\}$ . Drton (2009), in chapter 5, proved that, in order to satisfy this **MP II**, it is necessary to constrain the interactions referring to the variables  $(1, 2, 4)$  in both distributions  $\pi_{24|13}(\mathbf{i}_{24|13})$  and  $\pi_{124}(\mathbf{i}_{124})$ . Therefore, completeness is violated and according Bergsma and Rudas theorem 6 the model is not smooth.

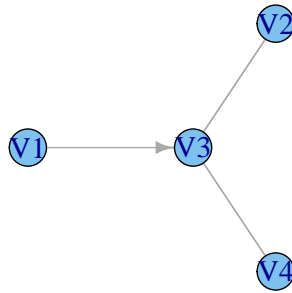


FIGURE 5.3.1

In this chapter we propose some conditions that, if satisfied, assure the smoothness of the model. We deal with this problem by finding what graphs, according to **MP II**, give a list of independencies that permits a marginal parametrization. In the following chapters we will propose three different parametrizations with their strong points and their disadvantages. In particular, the parametrization in the subsection 5.3.3 includes the advantages of the parametrizations in the previous subsections. Any parametrization is found by applying the method proposed in section 5.1.

**5.3.1. Parametrization based on (C3b).** We consider the independencies obtained from **MP II**. Let us remember that these conditional independencies follow conditions **(C1)**, **(C2a)** and **(C3b)**.

It is useful to introduce the class of subsets of  $T_h$ , having all possible parents  $J_h = \{A : A \in \mathcal{P}(T_h), pa_G(A) = pa_D(T_h)\}$ .

In the *first step*, we define the class of marginal sets  $\mathcal{H} = \{A_i \cup B_i \cup C_i, i = 1, \dots, k\}$  for any independence  $A_i \perp\!\!\!\perp B_i | C_i$ , thus we have three kind of marginal sets,  $\mathcal{M}^1$ ,  $\mathcal{M}^{2a}$  and  $\mathcal{M}^{3b}$ , according to the conditions considered:

- According to condition **(C1)**, we have the sets  $\mathcal{M}_h^1 = \cup_{j \leq h} T_j$ ,  $h = 1, \dots, s$ .
- According to **(C2a)**, the marginal sets are  $\mathcal{M}_h^{2a} = T_h \cup pa_D(T_h)$ ,  $\forall h = 1, \dots, s$ .
- According to **(C3b)**, the marginal sets are  $\mathcal{M}_{h,A}^{3b} = A \cup pa_D(T_h)$ ,  $\forall A \in \mathcal{P}(T_h) \setminus J_h$  and  $\forall h = 1, \dots, s$ . If  $pa_G(A) = pa_D(T_h)$ , then no interaction must be set equal to zero in the marginal  $A \cup pa_D(T_h)$ , so it is unnecessary to define this marginal set.

Note that sets  $\mathcal{M}_h^1$  and  $\mathcal{M}_h^{2a}$ ,  $\forall h = 1, \dots, s$  are the same for the **MP I** (see section 5.2).

The new kind of marginal sets satisfy the following relationships:  $\mathcal{M}_{h,A}^{3b} \subseteq \mathcal{M}_h^{2a}$ ,  $\forall A \in T_h$  where equality holds for  $A = T_h$ . Furthermore,  $\mathcal{M}_h^1 \not\subseteq \mathcal{M}_{h,A}^{3b}$ ,  $\forall A \in \mathcal{P}(T_h) \setminus J_h$  and  $\forall h = 1, \dots, s$ . The hierarchical class of marginal sets is given by

$$(5.3.1) \quad \mathcal{H}'_{II} = \{(\mathcal{M}_{h,A}^{3b}, \mathcal{M}_h^{2a}, \mathcal{M}_h^1), A \in \mathcal{P}(T_h)\}_{h=1, \dots, s}$$

If  $\mathcal{M}_{h,A}^{3b} = \mathcal{M}_h^{2a}$  or  $\mathcal{M}_h^{2a} = \mathcal{M}_h$ , we customarily retain only  $\mathcal{M}_h^{2a}$ .

The *second step* consists of specifying the classes of interaction following the hierarchical and complete properties. Thus, we allocate the interaction sets in the first marginal that contains them. Hence, the whole set of parameters  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}'_{II}}$  characterizes a saturated marginal model.

Below are some examples of marginal class  $\mathcal{H}'_{II}$ .

**Example 22.** With regard to figure (2.1.1A), the marginal class  $\mathcal{H}'_{II}$  is described by the following elements  $\mathcal{M}_1^{2a} = 1$ ,  $\mathcal{M}_2^{2a} = 2$ ,  $\mathcal{M}_2^1 = (1, 2)$ ,  $\mathcal{M}_{3,3}^{3b} = (1, 2, 3)$ ,  $\mathcal{M}_{3,4}^{3b} = (1, 2, 4)$ ,  $\mathcal{M}_3^{2a} = (1, 2, 3, 4)$ ,  $\mathcal{M}_{4,5}^{3b} = (3, 4, 5)$ ,  $\mathcal{M}_{4,6}^{3b} = (3, 4, 6)$ ,  $\mathcal{M}_{4,7}^{3b} = (3, 4, 7)$ ,  $\mathcal{M}_{4,8}^{3b} = (3, 4, 8)$ ,  $\mathcal{M}_{4,57}^{3b} = (3, 4, 5, 7)$ ,  $\mathcal{M}_{4,58}^{3b} = (3, 4, 5, 8)$ ,  $\mathcal{M}_{4,67}^{3b} = (3, 4, 6, 7)$ ,  $\mathcal{M}_{4,68}^{3b} = (3, 4, 6, 8)$ ,  $\mathcal{M}_{4,78}^{3b} = (3, 4, 7, 8)$ ,  $\mathcal{M}_{4,578}^{3b} = (3, 4, 5, 7, 8)$ ,  $\mathcal{M}_{4,678}^{3b} = (3, 4, 6, 7, 8)$ ,  $\mathcal{M}_4^{2a} = (3, 4, 5, 6, 7, 8)$ ,  $\mathcal{M}_4^1 = (1, 2, 3, 4, 5, 6, 7, 8)$ .

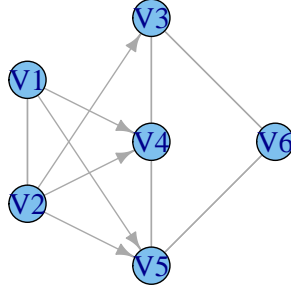


FIGURE 5.3.2

**Example 23.** Let us consider the graph in the figure (5.3.2). Some elements of component  $T_2$  have parents equal to the parents of the component:  $pa_G(A) = pa_D(T_2) = (1, 2)$ , where  $A$  belongs to  $J_2 = \{4; 5; 3, 4; 3, 5; 4, 5; 4, 6; 5, 6; 3, 4, 5; 3, 4, 6; 3, 5, 6; 4, 5, 6; 3, 4, 5, 6\}$ . Since the marginal sets  $\mathcal{M}_{h,A}^{3b}$  are evaluated for all  $A \in \mathcal{P}(T_h) \setminus J_h$ , the class of marginal sets  $\mathcal{H}'_{II}$  is composed by the followings sets  $\mathcal{M}_1^{2a} = 1, 2$ ;  $\mathcal{M}_{2,3}^{3b} = 1, 2, 3$ ;  $\mathcal{M}_{2,6}^{3b} = 1, 2, 6$ ;  $\mathcal{M}_{2,36}^{3b} = 1, 2, 3, 6$ ;  $\mathcal{M}_2^{2a} = 1, 2, 3, 4, 5, 6$ .

The *third step* consists of finding classes  $D_i = \{\mathcal{P}(A \cup B \cup C) \setminus (\mathcal{P}(A \cup C) \cup \mathcal{P}(B \cup C))\}$  according to the formula (4.3.2). As before, there are three kinds of these classes,  $D^1$ ,  $D^{2a}$  and  $D^{3b}$ :

- $D_h^1 = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(\cup_{j=1}^h T_j) \setminus (\mathcal{P}(T_h \cup pa_D(T_h)) \cup \mathcal{P}(\cup_{j=1}^{h-1} T_j))\}, \forall h = 1, \dots, s.$
- $D_h^{2a} = \cup_{A \in T_h} D_{h,A}^{2a}, \forall h = 1, \dots, s$  where  $D_{h,A}^{2a} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \cup pa_D(T_h)) \setminus (\mathcal{P}(Nb(A) \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus A \cup pa_D(T_h)))\}.$

- $D_h^{3b} = \cup_{A \in \mathcal{P}(T_h) \setminus J_h} D_{h,A}^{3b}$ ,  $\forall h = 1, \dots, s$ , where  $D_{h,A}^{3b} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(A \cup pa_D(T_h)) \setminus \mathcal{P}(A \cup pa_G(A)) \cup \mathcal{P}(pa_D(T_h))\}$ .

Note that the first two classes  $D_h^1$  and  $D_h^{2a}$  are the same as presented in the section 5.1.

Finally, in the *last step*, we must check if the conditions of the theorem 7 are satisfied. As before, there are three conditions to verify:

$$(5.3.2) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (\cup_{j=1}^h T_j), \quad \forall \mathcal{L} \in D_h^1.$$

$$(5.3.3) \quad pa_D(T_h) \cup nb(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h) \quad \forall \mathcal{L} \in D_{h,A}^{2a}, \quad \forall A \subseteq T_h.$$

$$(5.3.4) \quad pa_G(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq A \cup pa_D(T_h) \quad \forall \mathcal{L} \in D_{h,A}^{3b}, \quad \forall A \in \mathcal{P}(T_h) \setminus J_h.$$

$\forall h = 1, \dots, s$ . Note that,  $\mathcal{M}_h^1$  is the only marginal set that contains all interaction sets of  $D_h^1$ , thus the first condition is always satisfied. It is easy to see that,  $\forall \mathcal{L} \in D_{h,A}^{3b}$ , the sets  $\mathcal{M}_{h,A}^{3b} = A \cup pa_D(T_h)$  is the first marginal set that contains the interaction set  $\mathcal{L}$ . Thus, even condition (5.3.4) still holds. This is not always true for condition (5.3.3). In fact, it may happen that there is a set  $\mathcal{L} \in D_{h,A}^{2a}$  such that  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3b}$  that might not contain the set  $nb(A)$ . Theorem 10 gives the conditions according to which condition (5.3.3) holds.

Let  $\mathcal{K}_h$  be the class of all non connected subsets of  $T_h$ .



**Theorem 10.** *A graphical model of type II is a marginal model of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b})$ ,  $\mathcal{M} \in \mathcal{H}'_{II}\}$  if,  $\mathcal{K}_h \subseteq J_h$ ,  $\forall h = 1, \dots, s$ .*

That is, for any subset  $A$  of  $T_h$ , if  $A$  is a non connected set, then it must have the parent set equal to the parent set of the component:  $pa_G(A) = pa_D(T_h)$ .

The proof of this theorem is postponed in the appendix to this chapter.

Following are some examples of this parametrization.

**Example 24.** Let us consider the graph in figure (2.1.1A). According to theorem 10, we may not conclude that graphical model II associated with the graph in figure 2.1.1 is compatible with the parametrizations  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b})$ ,  $\mathcal{M} \in \mathcal{H}'_{II}\}$ . In fact, the class  $\mathcal{K}_h \setminus J_h = \{5, 8; 6, 7; 7, 6, 8; 5, 7, 8\} \neq \emptyset$ .

**Example 25.** According to theorem 10, graphical model II associated with the graph in figure (5.3.2) permits representation with a marginal model, because all non connected sets have parents equal to the parents of the component. In fact, the non complete sets in component  $T_2$  are  $\{3, 5; 4, 6\}$  and  $pa_G(3, 5) = pa_G(4, 6) = pa_D(T_2) = 1, 2$ . Hence the model is smooth.

**Example 26.** Let us consider the graph in figure 5.3.3. The class of marginal sets is given by  $\mathcal{H}'_{II} = \{1, 2; 1, 2, 3; 1, 2, 4; 1, 2, 5; 1, 2, 6; 1, 2, 3, 5; 1, 2, 3, 6; 1, 2, 4, 5; 1, 2, 4, 6; 1, 2, 5, 6; 1, 2, 3, 5, 6; 1, 2, 4, 5, 6; 1, 2, 3, 4, 5, 6\}$ . In this case as well, the conditions of theorem 10 are not satisfied, because the non complete subsets of  $T_4$ ,  $K = 3, 6$  and  $K = 4, 5$  do not have all parents, that is  $pa_G(K) \neq pa_D(T_2) = 1, 2$ .

It should be noted that, if a graph has all complete components, then the class  $\mathcal{K}_h = \{\emptyset\}$ ,  $\forall h = 1, \dots, s$ . This implies that for any  $J_h$ , the relation  $\mathcal{K}_h \subseteq J_h$  holds and

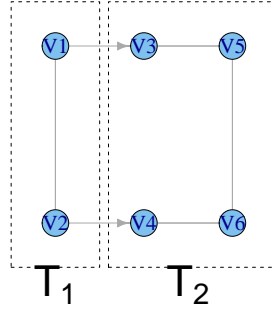


FIGURE 5.3.3

theorem 10 is satisfied. On the other hand, if a graph has only one component, then class  $J_h = \mathcal{P}(T_h)$  and relationship  $\mathcal{K}_h \subseteq J_h$  holds. Note that, in the last case, no one independence follows the condition **(C3b)** thus the GM II is equal to **GM I**. It's easy to note that in this case the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a}), \mathcal{M} \in \mathcal{H}'_{II}\}$  is equal to the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a}), \mathcal{M} \in \mathcal{H}_I\}$  proposed in chapter 5.2.

**5.3.2. Parametrization based on (C2\*a) and (C3\*b).** In chapter 3 we proposed two new conditions, called **(C2\*a)** and **(C3\*b)**, in order to explain the relationships among a group of variables belonging to a component  $T_h$  with the parents. We also saw that these conditions are equivalent to **(C2a)** and **(C3b)**. In order to explain the benefits of these, we apply the method proposed in section 5.1 to the graphical model of type II described by **(C1)**, **(C2\*a)** and **(C3\*b)**. The reasons that lead us to replace the **(C2a)** with **(C2\*a)** lies on the importance to have a non redundant list of independence (see section 3.1). Instead, as it will be shown in section 3.2, there are three main reasons to replace **(C3b)** with **(C3\*b)**. First, using **(C3\*b)**, the list of independencies is shorter than the list

obtained with **(C3b)** and, consequently, the list of marginals to define is shorter too. Secondly, given a graph, the list of independencies that obeys to **(C3\*b)** is non redundant. Finally, as we will see at the end of this section, the new class of marginal sets permits different conditions in order to find graphical models II which are also marginal model. Before proceeding, we define  $PA_h^+$  as the class which elements belong to  $\{\{pa_D(T_h)\} \cup PA_h\}$ . The elements of class  $PA_h^+$  are partially ordered according the following rule:  $\forall \mathcal{A}, \mathcal{B} \in PA_h$  if  $|ch(\mathcal{B})| < |ch(\mathcal{A})|$  then  $\mathcal{A} < \mathcal{B}$ .

In the *first step* of our method, introduced in section 5.1, we define the class of marginal sets  $\mathcal{H} = \{A_i \cup B_i \cup C_i, i = 1, \dots, k\}$  for any independence  $A_i \perp\!\!\!\perp B_i | C_i$ . In order to consider the three conditions **(C1)**, **(C2\*a)** and **(C3\*b)** we use the following marginal sets:

- From condition **(C1)**, we have the sets  $\mathcal{M}_h^1 = \cup_{j \leq h} T_j, h = 1, \dots, s$ .
- From **(C2\*a)**, the marginal sets unchange and are  $\mathcal{M}_h^{2a} = T_h \cup pa_D(T_h), \forall h = 1, \dots, s$ .
- From **(C3\*b)**, the marginal sets are  $\mathcal{M}_{h,\mathcal{A}}^{3*b} = pa_D(T_h) \cup (T \setminus ch(\mathcal{A})), \forall \mathcal{A} \in PA_h^+$  and  $\forall h = 1, \dots, s$ . Any marginal set  $\mathcal{M}_{h,\mathcal{A}}^{3*b}$  is composed of the  $pa_D(T_h)$  and of a subset of  $T_h$ . Notice that, when  $\mathcal{A} = pa_D(T)$ , the marginal set becomes  $\mathcal{M}_{h,pa_D(T_h)}^{3*b} = pa_D(T_h) \cup (NC \cup NA)$ , thus  $\mathcal{M}_{h,pa_D(T_h)}^{3*b} = \cap_{\mathcal{A} \in PA_h} \mathcal{M}_{h,\mathcal{A}}^{3*b}$ . In addition, if  $PA_h$  has only one element, then only set  $\mathcal{A} \in PA_h$  has children corresponding to the whole set  $CH_h$ , thus  $\mathcal{M}_{h,pa_D(T_h)}^{3*b} = \mathcal{M}_{h,\mathcal{A}}^{3*b}$ .

For all sets  $\mathcal{A} \in PA_h^+$  and for all  $h = 1, \dots, s$ , the three kinds of marginal sets are in accordance with the following relationship:

$$(5.3.5) \quad \mathcal{M}_{h,\mathcal{A}}^{3^*b} \subseteq \mathcal{M}_h^{2a} \subseteq \mathcal{M}_h^1 \quad \forall h = 1, \dots, s.$$

Thus, hierarchical class  $\mathcal{H}_{II}^*$  is given by

$$(5.3.6) \quad \mathcal{H}_{II}^* = \{(\mathcal{M}_{h,\mathcal{A}}^{3^*b}, \mathcal{M}_h^{2a}, \mathcal{M}_h^1), \mathcal{A} \in PA_h^+\}_{h=1,\dots,s}.$$

As before, if  $\mathcal{M}_h^{2a} = \mathcal{M}_h^1$  for some  $h$ , we customarily to retain only the set  $\mathcal{M}_h^{2a}$ ; if, instead, it occurs that  $\mathcal{M}_{h,\mathcal{A}}^{3^*b} = \mathcal{M}_h^{2a}$  or  $\mathcal{M}_{h,\mathcal{A}}^{3^*b} = \mathcal{M}_h^1$ , for some  $h$  and  $\mathcal{A}$ , then we retain respectively only the set  $\mathcal{M}_h^{2a}$  or  $\mathcal{M}_h^1$ , as shown in the following examples.

**Example 27.** With regards to figure (2.1.1), the components  $T_1$  and  $T_2$  do not have parents, thus  $PA_1 = \{\emptyset\}$  and  $PA_2 = \{\emptyset\}$ , so there are not marginal sets like  $\mathcal{M}_{h,\mathcal{A}}^{3^*b}$ . The components  $T_3$  and  $T_4$  have classes of parents with the same children, respectively  $PA_3 = \{1; 2\}$  and  $PA_4 = \{3; 4\}$ . In addition we have the classes  $PA_3^+ = \{1, 2; 1; 2\}$  and  $PA_4^+ = \{3, 4; 3; 4\}$ . The marginal class  $\mathcal{H}_{II}^*$  is composed by the following sets  $\mathcal{M}_1^{2a} = 1$ ,  $\mathcal{M}_2^{2a} = 2$ ,  $\mathcal{M}_2^1 = (1, 2)$ ,  $\mathcal{M}_{3,1}^{3^*b} = (1, 2, 4)$ ,  $\mathcal{M}_{3,2}^{3^*b} = (1, 2, 3)$ ,  $\mathcal{M}_3^{2a} = (1, 2, 3, 4)$ ,  $\mathcal{M}_{4,34}^{3^*b} = (3, 4, 7, 8)$ ,  $\mathcal{M}_{4,3}^{3^*b} = (3, 4, 6, 7, 8)$ ,  $\mathcal{M}_{4,4}^{3^*b} = (3, 4, 5, 7, 8)$ ,  $\mathcal{M}_4^{2a} = (3, 4, 5, 6, 7, 8)$ ,  $\mathcal{M}_4^1 = (1, 2, 3, 4, 5, 6, 7, 8)$ .

**Example 28.** The graph in figure (5.2.1) is composed of two components. The class of marginal sets of the graph  $\mathcal{H}_{II}^*$  is given by the following sets  $\mathcal{M}_1^{2a} = (1, 2)$ ,  $\mathcal{M}_{2,12}^{3^*b} = (1, 2, 6)$ ,  $\mathcal{M}_{2,1}^{3^*b} = (1, 2, 3, 6)$ ,  $\mathcal{M}_2^{2a} = (1, 2, 3, 4, 5, 6)$ .

According to the *second step* we specify the interaction sets in accordance with the hierarchical and complete properties.

In the *third step* we report the classes defined by the formula 4.3.2 according to three conditions **(C1)**, **(C2\*a)** and **(C3\*b)**. Note that, since we proved that **(C2a)** and **(C2\*a)** are equivalent, then the class  $D_h^{2a}$  and  $D_h^{2^*a}$  have the same elements.

- $D_h^1 = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(\cup_{j=1}^h T_j) \setminus (\mathcal{P}(T_h \cup pa_D(T_h)) \cup \mathcal{P}(\cup_{j=1}^{h-1} T_j))\}, \forall h = 1, \dots, s.$
- $D_h^{2a} = \cup_{A \subseteq T_h} D_{h,A}^{2a}, \forall h = 1, \dots, s,$  where  $D_{h,A}^{2a} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \cup pa_D(T_h)) \setminus (\mathcal{P}(Nb(A) \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus A \cup pa_D(T_h)))\},$  and .
- $D_h^{3^*b} = \cup_{\mathcal{A} \in PA_h^+} D_{h,\mathcal{A}}^{3^*b}, \forall h = 1, \dots, s$  where  $D_{h,\mathcal{A}}^{3^*b} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h)) \setminus (\mathcal{P}(pa_D(T_h)) \cup \mathcal{P}(T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h) \setminus \mathcal{A}))\},$  and

Where class  $D_h^1$  and  $D_h^{2a}$  are the same presented in both section 5.1 and section 5.2.1.

The *final step* consists of verifying when parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}^*}$  satisfies condition 4.3.4 of theorem 7. In this case as well, we have three conditions to check:

$$(5.3.7) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (\cup_{j=1}^h T_j), \quad \forall \mathcal{L} \in D_h^1$$

$$(5.3.8) \quad pa_D(T_h) \cup nb(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h), \quad \forall \mathcal{L} \in D_{h,A}^{2a}, \forall A \subseteq T_h.$$

$$(5.3.9) \quad pa_D(T_h) \setminus \mathcal{A} \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h), \quad \forall \mathcal{L} \in D_{h,\mathcal{A}}^{3^*b}, \forall \mathcal{A} \in PA_h^+$$

$\forall h = 1, \dots, s$ . The first condition is always checked, as it was shown in section 5.1. Unfortunately, conditions (5.3.8) and (5.3.9) are not always met. In particular, two different situations may occur.

First, when there is at least one interaction set  $\mathcal{L} \in D_{h,\mathcal{A}}^{3^*b}$  such that  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,\mathcal{B}}^{3^*b}$ , for  $\mathcal{A}, \mathcal{B} \in PA_h^+$ , with  $\mathcal{B} < \mathcal{A}$  in  $PA_h^+$ , condition (5.3.9) is satisfied, only if the relationship  $\mathcal{P}(T_h) \cup \mathcal{A} \subseteq \mathcal{M}_{h,\mathcal{B}}^{3^*b} \subseteq T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h)$  is met.

Secondly, when there is at least a set  $\mathcal{L} \in D_{h,\mathcal{A}}^{2^a}$  such that  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,\mathcal{A}}^{3^*b}$ , for some  $\mathcal{A}$ , condition (5.3.8) is met only if  $pa_D(T_h) \cup nb(\mathcal{A}) \subseteq \mathcal{M}_{h,\mathcal{A}}^{3^*b} \subseteq T_h \cup pa_D(T_h)$  holds.

The following example shows these situations.

**Example 29.** With regard to the graph in figure (5.3.3), the marginal sets in  $\mathcal{H}_{II}^*$  are  $\mathcal{M}_1^1 = \mathcal{M}_1^{2^a} = 1, 2$ ,  $\mathcal{M}_1^1 = \mathcal{M}_2^{2^a} = (1, 2, 3, 4, 5, 6)$ ,  $\mathcal{M}_{3,12}^{3^*b} = (1, 2, 5, 6)$ ,  $\mathcal{M}_{3,1}^{3^*b} = (1, 2, 3, 5, 6)$ ,  $\mathcal{M}_{3,2}^{3^*b} = (1, 2, 4, 5, 6)$ . The class of interaction sets referring to the null parameters are  $D_2^{3^*b} = \{1, 4; 1, 5; 1, 6; 1, 4, 5; 1, 4, 6; 1, 5, 6; 1, 4, 5, 6; 1, 2, 4; 1, 2, 5; 1, 2, 6; 1, 2, 4, 5; 1, 2, 4, 6; 1, 2, 5, 6; 1, 2, 4, 5, 6; 2, 3; 2, 5; 2, 6; 2, 3, 5; 2, 3, 6; 2, 5, 6; 2, 3, 5, 6; 1, 2, 3; 1, 2, 3, 5; 1, 2, 3, 6; 1, 2, 3, 5, 6\}$  and  $D_2^{2^a} = \{3, 4; 3, 6; 3, 4, 6; 1, 3, 4; 1, 3, 6; 1, 3, 4, 6; 2, 3, 4; 2, 3, 6; 2, 3, 4, 6; 3, 4, 5; 3, 5, 6; 3, 4, 5, 6; 1, 2, 3, 4; 1, 2, 3, 6; 1, 2, 3, 4, 6; 1, 3, 4, 5; 1, 3, 5, 6; 1, 3, 4, 5, 6; 2, 3, 4, 5; 2, 3, 5, 6; 2, 3, 4, 5, 6; 1, 2, 3, 4, 5; 1, 2, 3, 5, 6; 1, 2, 3, 4, 5, 6\}$ .

The first problem is addressed in the next lemma referring to a chain graph with only complete components.

**Lemma 2.** *A graphical model of type II is a marginal model of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b}), \mathcal{M} \in \mathcal{H}_{II}^*\}$ , if the following conditions are met:*

- (a) *All components of the graph are complete;*
- (b)  $\forall \mathcal{A}, \mathcal{B} \in PA_h: \mathcal{B} < \mathcal{A}$  in  $PA_h$ , if  $(CH_h \setminus ch(\mathcal{A})) \cap (CH_h \setminus ch(\mathcal{B})) \neq \emptyset$ ,  
 then  $(CH_h \setminus ch(\mathcal{B})) \subseteq (CH_h \setminus ch(\mathcal{A}))$ ,  $h = 1, \dots, s$ .

Note that according to theorem 10, each graphical model associated with a chain graph with complete components is a marginal model, thus Lemma 2 identifies a smaller class of GM II that are MM. However, the graphical model described by the parameters  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}_{II}^*}$  brings many benefits shown in theorem 11.

The proof of this lemma is reported in the appendix to this chapter. An example will show the applicability of Lemma 2.

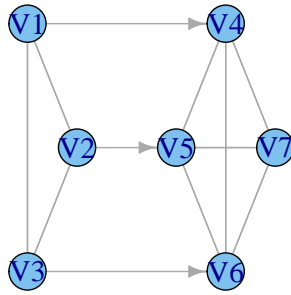


FIGURE 5.3.4

**Example 30.** Let's consider the graph in figure 5.3.4 where the component  $T_2$  is complete. The class of parents with the same children is  $PA_2 = \{1; 2; 3\}$  and  $PA_2^+ = \{1, 2, 3; 1; 2; 3\}$ . The marginal sets are  $\mathcal{M}_1^{2a} = (1, 2, 3)$ ,  $\mathcal{M}_{2,123}^{3*b} = (1, 2, 3, 7)$ ,  $\mathcal{M}_{2,1}^{3*b} = (1, 2, 3, 5, 6, 7)$ ,  $\mathcal{M}_{2,2}^{3*b} = (1, 2, 3, 4, 6, 7)$ ,  $\mathcal{M}_{2,3}^{3*b} = (1, 2, 3, 4, 5, 7)$ ,  $\mathcal{M}_2^{2a} = (1, 2, 3, 4, 5, 6, 7)$ .

We report the set of interest  $\forall \mathcal{A} \in PA_2$  mentioned in lemma 2:  $CH_2 \setminus ch(1) = (5, 6)$ ,  $CH_2 \setminus ch(2) = (4, 6)$  and  $CH_2 \setminus ch(3) = (4, 5)$ . It is easy to see that the requirement of the theorem has not been met, in fact  $(CH_2 \setminus ch(1)) \cap (CH_2 \setminus ch(2)) \neq \emptyset$  and  $(CH_2 \setminus ch(1)) \not\subseteq (CH_2 \setminus ch(2))$ . The same holds for  $\mathcal{A} = 3$ .

In fact, according to **(C3\*b)** we get the following independencies:

$$1 \perp\!\!\!\perp 5, 6, 7 \mid 2, 3 \quad 2 \perp\!\!\!\perp 4, 6, 7 \mid 1, 3 \quad 3 \perp\!\!\!\perp 4, 5, 7 \mid 1, 2$$

For example, the interaction set  $(1, 2, 6)$  is defined in the marginal  $\mathcal{M}(1, 2, 6) = \mathcal{M}_{2,567}^{3*b} = (1, 2, 3, 5, 6, 7)$  but, according to the second independence this marginal set should satisfy the relationship  $1, 3, 5 \subseteq \mathcal{M}_{2,567}^{3*b} \subseteq 1, 2, 3, 4, 5, 6, 7$ . Thus, the parametrization based on  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}_{II}^*}$  is not representative.

The following theorem shows that there are GM II that are marginal models of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b}), \mathcal{M} \in \mathcal{H}_{II}^*\}$  but not of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}'_{II}\}$ . This result justifies the use of this parametrization.

**Theorem 11.** *A graphical model of type II is a marginal model with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b}), \mathcal{M} \in \mathcal{H}_{II}^*\}$ , if the assumption (b) of Lemma 2 holds and if, for all  $V_j \in CH_h$  such that  $Nb(V_j) \notin \mathcal{C}_h$ ,  $\{K : K \in \mathcal{K}_h; K \cap nb(V_j) \neq \emptyset\} \subseteq J_h$ .*



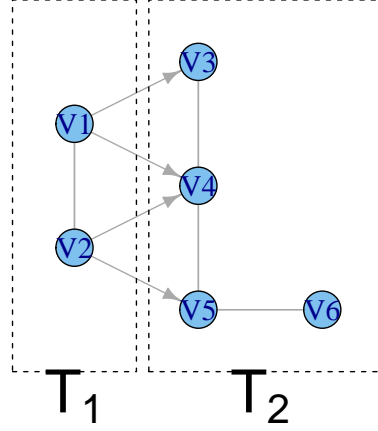


FIGURE 5.3.5

**Example 31.** In figure (5.3.5), regarding component  $T_2$ , we may recognize the set of children  $CH_2 = \{3, 4, 5\}$ , and the class of parents with the same children  $PA_2 = \{1; 2\}$  and  $PA_2^+ = \{1, 2; 1; 2\}$ . Note that,  $CH_2 \setminus ch(1) = 5$  and  $CH_2 \setminus ch(2) = 3$ . Since the previous two sets have null intersection, the assumption of lemma 2 is fulfilled. Now we must check whether condition (5.3.8) of theorem 11 has been met. For every vertices in  $CH_2$ , we have that  $Nb(3) = 3, 4$  is complete set but  $Nb(4) = 3, 4, 5$  and  $Nb(5) = 4, 5, 6$  are non complete sets. Note that  $\{K : K \in \mathcal{K}_2; K \cap nb(4) \neq \emptyset\} = \{3, 5; 3, 6; 3, 5, 6\}$ . Since  $(3, 6) \notin J_2$  the second condition of theorem 11 is not met.

The following examples compare the two parametrizations  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s$   
 $(D_h^1 \cup D_h^{2a} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}'_{II}\}$  and  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b}), \mathcal{M} \in$   
 $\mathcal{H}^*_{II}\}$ .

**Example 32.** According to the graph in figure (5.3.3),  $\forall V_j \in CH_2$ , the sets  $CH_2 \setminus ch(V_j)$  are incompatible, hence, condition *b*) of lemma 2 is fulfilled. Further,  $\forall V_j \in CH_2$ , the sets  $Nb(V_j)$  are complete, then theorem 11 is also satisfied. Note that the graphical model associated with this graph does not permit parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}'_{II}}$ . In fact, the class of all non connected sets is  $\mathcal{K}_2 = \{3, 2; 3, 5; 3, 6; 4, 5; 3, 4, 5; 3, 4, 6\}$ . Note that  $pa_G(3, 6) = 1 \neq pa_D(T_2) = 1, 2$ . Thus  $\mathcal{K}_2 \not\subseteq J_2$ , and the theorem (10) does not hold.

**Example 33.** As we showed in example 30, the graph in figure 5.3.4 does not satisfy lemma 2. Note that, in  $T_2$  the only two sets not connected are (3, 5) and (4, 6) which have  $pa_G(3, 5) = pa_G(4, 6) = pa_D(T_2)$ . Thus, theorem 10 holds, and this graphical model II is a marginal model based on  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}'_{II}}$ . The class of marginal set  $\mathcal{H}'_{II}$  is composed by  $\mathcal{M}_1^{2a} = (1, 2, 3)$ ,  $\mathcal{M}_{2,4}^{3b} = (1, 2, 3, 4)$ ,  $\mathcal{M}_{2,5}^{3b} = (1, 2, 3, 5)$ ,  $\mathcal{M}_{2,6}^{3b} = (1, 2, 3, 6)$ ,  $\mathcal{M}_{2,7}^{3b} = (1, 2, 3, 7)$ ,  $\mathcal{M}_{2,45}^{3b} = (1, 2, 3, 4, 5)$ ,  $\mathcal{M}_{2,46}^{3b} = (1, 2, 3, 4, 6)$ ,  $\mathcal{M}_{2,47}^{3b} = (1, 2, 3, 4, 7)$ ,  $\mathcal{M}_{2,56}^{3b} = (1, 2, 3, 5, 6)$ ,  $\mathcal{M}_{2,57}^{3b} = (1, 2, 3, 5, 7)$ ,  $\mathcal{M}_{2,67}^{3b} = (1, 2, 3, 6, 7)$ ,  $\mathcal{M}_{2,457}^{3b} = (1, 2, 3, 4, 5, 7)$ ,  $\mathcal{M}_{2,467}^{3b} = (1, 2, 3, 4, 6, 7)$ ,  $\mathcal{M}_{2,567}^{3b} = (1, 2, 3, 5, 6, 7)$ ,  $\mathcal{M}_2^{2a} = (1, 2, 3, 4, 5, 6, 7)$ .

In this case, any interaction set belonging to class  $D_2^{3b}$  is defined in the smallest marginal. This means that the graphical model is a marginal model based on  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}'_{II}\}$  and therefore it is smooth.

We should make some comments about the two parametrizations introduced. First, the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}'_{II}\}$  is characterized by the marginal sets  $\mathcal{M}_h^1$ ,  $\mathcal{M}_h^{2a}$  and  $\mathcal{M}_{h,A}^{3b}$ , which are obtained by considering the independencies of (C2a) and (C3b). In this way, we get a marginal set referring to any subset  $A$  of  $T_h$  that does not belong to the class  $J_h$ .

With this parametrization, as we saw in Section 5.2.1, condition (5.3.4) is always met, while (5.3.3) holds only if all non connected subsets belong to class  $J_h$ .

On the other hand, the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b}), \mathcal{M} \in \mathcal{H}_{II}^*\}$  is characterized by fewer marginal sets that are able to describe the independencies following **(C3\*b)**. It is possible to note that class of marginal sets  $\mathcal{H}_{II}^*$  is included in class  $\mathcal{H}'_{II}$ . Moreover, using the second parametrization, there are fewer graphs which satisfy condition (5.3.9) but more graphs which satisfy condition (5.3.8).

Note that the new condition **(C2\*a)** is important in the last step, when we must verify the conditions 5.3.8, 5.3.7 and 5.3.9.

In order to profit from both advantages, in the next section we will propose a parametrization that is a mix between the two seen above.

**5.3.3. Mixed parametrization.** The two previous parametrizations have different advantages. The following diagram represents the **GM II** that are marginal models according to theorem (10) (left area), according to theorem (11) (right area) and according to lemma (2) ( a subset of the intersection area). In particular, in the intersection area, we have the graphs which satisfy condition *b*) of lemma 2 and with the set  $\mathcal{K}_h \subseteq J_h$ .

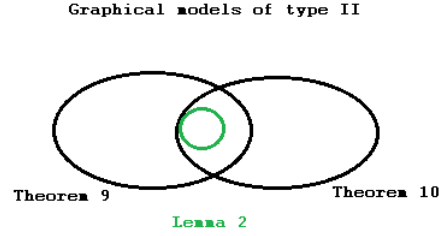


FIGURE 5.3.6

The difference between the two parametrizations derives from the choice of the hierarchical class of marginal sets. In order to define the mixed class of marginal sets we consider the whole list of marginal sets  $\mathcal{M}_{h,A}^{3*b}$ ,  $\forall A \in PA_h^+$  and we add certain elements of the list of marginal sets  $\mathcal{M}_{h,A}^{3b}$ ,  $A \in \mathcal{P}(T_h) \setminus J_h$ . The resulting class is

$$\mathcal{M}_{h,A}^{3bMIX} = pa_D(T_h) \cup (NC_h \cup NA_h) \cup A \quad \forall A \in \{\mathcal{P}(CH_h) \setminus J_h\}$$

$\forall h = 1, \dots, s$ , where the class  $J_h$  is the same class seen in Chapter 5.3.1:  $J_h = \{A \in \mathcal{P}(T_h) : pa_G(A) = pa_D(T_h)\}$ . It is easy to see that when  $A = \emptyset$ ,  $\mathcal{M}_{h,A}^{3bMIX} = pa_D(T_h) \cup (NC_h \cup NA_h)$  and it is the smallest marginal sets with elements in  $T_h$ . Furthermore, when  $A = CH_h \setminus ch(A)$ , then  $\mathcal{M}_{h,A}^{3bMIX} = \mathcal{M}_{h,A}^{3*b}$ . Since the relationship  $\mathcal{M}_{h,A}^{3bMIX} \subseteq \mathcal{M}_h^{2a}$  holds  $\forall A \in \{\mathcal{P}(CH_h) \setminus J_h; \emptyset\}$ , the class of marginal sets for this mixed parametrization is:

$$\mathcal{H}_{II}^{MIX} = \{(\mathcal{M}_{h,A}^{3bMIX}, \mathcal{M}_h^{2a}, \mathcal{M}_h^1), \forall A \in \{\mathcal{P}(CH_h) \setminus J_h\}_{h=1, \dots, s}\}$$

where the sets  $\mathcal{M}_{h,A}^{3bMIX}$  are ordered to obtain a hierarchical class of marginal sets. If  $\mathcal{M}_{h,A}^{3bMIX} = \mathcal{M}_h^{2a}$  we retain only set  $\mathcal{M}_h^{2a}$ . From the construction, it is easy to see that  $\mathcal{H}_{II}^* \subseteq \mathcal{H}_{II}^{MIX} \subseteq \mathcal{H}'_{II}$ .

In the *second step* of the method proposed in section 4.2, we define all interaction sets according to the property of completeness and hierarchy. Here is an example of this new parametrization applied to the following graph.

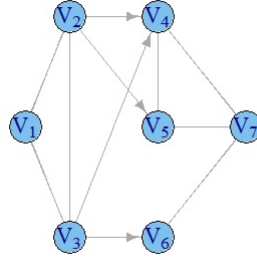


FIGURE 5.3.7

**Example 34.** The previous graph is composed of two components  $T_1$  and  $T_2$ . First, it is useful to note that class  $J_2$  is empty because vertex 1 has no children. Now, we are able to recognize the following marginal sets of the new type:  $\mathcal{M}_{2,\emptyset}^{3bMIX} = (1, 2, 3, 7)$ ,  $\mathcal{M}_{2,4}^{3bMIX} = (1, 2, 3, 4, 7)$ ,  $\mathcal{M}_{2,5}^{3bMIX} = (1, 2, 3, 5, 7)$ ,  $\mathcal{M}_{2,6}^{3bMIX} = (1, 2, 3, 6, 7)$ ,  $\mathcal{M}_{2,45}^{3bMIX} = (1, 2, 3, 4, 5, 7)$ ,  $\mathcal{M}_{2,46}^{3bMIX} = (1, 2, 3, 4, 6, 7)$ ,  $\mathcal{M}_{2,56}^{3bMIX} = (1, 2, 3, 5, 6, 7)$  and  $\mathcal{M}_{2,456}^{3bMIX} = (1, 2, 3, 4, 5, 6, 7)$ . The relations of independence represented by this graph are:

$$1 \perp\!\!\!\perp 4, 5, 6, 7 | 2, 3 \quad 2 \perp\!\!\!\perp 6, 7 | 1, 3 \quad 3 \perp\!\!\!\perp 5, 7 | 1, 2 \quad 4, 5 \perp\!\!\!\perp 6 | 1, 2, 3, 7$$

In the *third step* of the method, we take into account the classes of parameters to constrain to zero. Since we must satisfy conditions (C1), (C2\*a) and (C3\*b),

we take the same classes  $D_i$  proposed in section 5.2.2. Thus we get  $D_h^1$ ,  $D_h^{2^*a}$  and  $D_h^{3^*b}$  for all  $h = 1, \dots, s$ .

The *final step* consists of verifying when the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b}), \mathcal{M} \in \mathcal{H}_{II}^{MIX}\}$  satisfies condition 4.3.4 of theorem 7. In this case as well, we have three conditions to check.

$$(5.3.10) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (\cup_{j=1}^h T_j), \quad \forall \mathcal{L} \in D_h^1$$

$$(5.3.11) \quad pa_D(T_h) \cup nb(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h) \quad \forall \mathcal{L} \in D_{h,A}^{2^*a}, \forall A$$

$$(5.3.12) \quad pa_D(T_h) \setminus \mathcal{A} \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h) \quad \forall \mathcal{L} \in D_{h,\mathcal{A}}^{3^*b}, \forall \mathcal{A} \in PA_h$$

$\forall h = 1, \dots, s$ . As before, the set  $\mathcal{M}(\mathcal{L})$  for any  $\mathcal{L} \in D_h^1$  is exactly  $\mathcal{M}_h^1$ , thus condition (5.3.10) is always checked. Condition (5.3.12) is also checked. In fact, for any interaction set  $\mathcal{L} \in D_{h,\mathcal{A}}^{3^*b}$ ,  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,\mathcal{A}}^{3bMIX}$ , where  $A = CH_h \setminus ch(\mathcal{A})$ . One can easily note that  $pa_D(T_h) \setminus \mathcal{A} \subseteq \mathcal{M}_{h,\mathcal{A}}^{3bMIX} \subseteq T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h)$  holds for any  $\mathcal{A} \in PA_h^{3^*b}$ .

Thus, we obtain the same results as the standard parametrization where the only problematic statement is (5.3.11).

The following theorem describes when a graphical model of type II is a marginal model.

**Theorem 12.** *A graphical model of type II is a marginal model with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b})\}$ ,  $\mathcal{M} \in \mathcal{H}_{II}^{MIX}\}$ , if, for all  $V_j \in CH_h$  such that  $Nb(V_j) \notin \mathcal{C}_h$ ,  $\{K : K \in \mathcal{K}_h; K \cap nb(V_j) \neq \emptyset\} \subseteq J_h$ .*

The proof of this theorem is reported in the appendix to Chapter 5.

An example will show this result.

**Example 35.** *(continue from Example 26)* It is easy to see that the graph in figure (5.3.7) meets the condition of theorem 12. In fact, for any vertex in  $CH_2 = \{4, 5, 6\}$ , we have  $Nb(4) = Nb(5) = (4, 5, 7)$  which is complete and  $Nb(6) = (6, 7)$  which is also complete.

**Example.** Following it is shown that it is possible to arrive at the same result applying theorem 7.

In fact, according to independence  $4, 5 \perp\!\!\!\perp 6 | 1, 2, 3, 7$ , there are some interaction sets  $\mathcal{L} \in D_2^{2a}$  such that  $(4, 6) \subseteq \mathcal{L}$ . Note that, for these interaction sets, we have  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{2,46}^{3bMIX} = 1, 2, 3, 4, 6, 7$ . But, since  $C = 1, 2, 3, 7 \subseteq \mathcal{M}_{2,46}^{3bMIX}$ , theorem 7 is checked. The same arguments hold for interaction sets  $\mathcal{L} \in D_2^{2a}$  such that  $(5, 6) \subseteq \mathcal{L}$ . In this case the first marginal set  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{2,56}^{3bMIX} = 1, 2, 3, 5, 6, 7$ . Even here it is easy to check that theorem 7 is satisfied. Since all the remaining interaction sets  $\mathcal{L} \in D_2^{2a}$  are defined in  $\mathcal{M}_2^{2a} = 1, 2, 3, 4, 5, 6, 7$ , the parametrization proposed is marginal and the model is smooth.

**5.3.4. Comparative examples of the three parametrizations proposed for GM II.** In this section some examples will be presented to compared the three parametrizations proposed above. The first example shows a graphical model II that permits all the parametrizations proposed:  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s$

$(D_h^1 \cup D_h^{2a} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}'_{II}\}, \{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b}), \mathcal{M} \in \mathcal{H}^*_{II}\}$   
and  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b}), \mathcal{M} \in \mathcal{H}^{MIX}_{II}\}$ .

**Example 36.** Let us consider the graph in figure 5.3.2, table 1 reports the marginal class for the three parametrizations:

$\mathcal{H}'_{II}$				
$\mathcal{M}_1^{2a} = (1, 2)$	$\mathcal{M}_{2,3}^{3b} = (1, 2, 3)$	$\mathcal{M}_{2,6}^{3b} = (1, 2, 6)$	$\mathcal{M}_{2,36}^{3b} = (1, 2, 3, 6)$	$\mathcal{M}_2^{2a} = (1, 2, 3, 4, 5, 6)$
$\mathcal{H}^*_{II}$				
$\mathcal{M}_1^{2a} = (1, 2)$	$\mathcal{M}_{2,12}^{3*b} = (1, 2, 6)$	$\mathcal{M}_{2,6}^{3*b} = (1, 2, 3, 6)$	$\mathcal{M}_2^{2a} = (1, 2, 3, 4, 5, 6)$	
$\mathcal{H}^{MIX}_{II}$				
$\mathcal{M}_1^{2a} = (1, 2)$	$\mathcal{M}_{2,\emptyset}^{3bMIX} = (1, 2, 6)$	$\mathcal{M}_{2,3}^{3bMIX} = (1, 2, 3, 6)$	$\mathcal{M}_2^{2a} = (1, 2, 3, 4, 5, 6)$	

TABLE 1. Three classes of marginal sets  $\mathcal{H}'_{II}$ ,  $\mathcal{H}^*_{II}$  and  $\mathcal{H}^{MIX}_{II}$  which refer to 5.3.2.

According to theorem 10, theorem 11 and theorem 12, the graphical model II associated with this graph permits the parametrizations  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}'_{II}}$ ,  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}^*_{II}}$  and  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}^{MIX}_{II}}$  and therefore the model is smooth. In fact, the class of all non connects sets of  $T_2$  of  $\mathcal{K}_2 = \{3, 5; 4, 6\}$ . It is easy to see that  $\mathcal{K}_2 \in J_2$ , thus theorems 10 and 12 hold. Furthermore,  $CH_2 \setminus ch(V_1) = 3, 6$  contains  $CH_2 \setminus ch(V_2) = 6$ . In the example 25 we detailed the case of the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}'_{II}}$ . About the  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}^*_{II}}$ , we may consider the marginal class  $\mathcal{H}^*$ . The **(C3\*b)** leads only two independencies  $2 \perp\!\!\!\perp 6|1$  and  $1 \perp\!\!\!\perp 3, 6|2$ . Note that the interaction that refers to  $(1, 2, 6) \in D_2^{3*b}$  is defined in the marginal  $\mathcal{M}(1, 2, 6) = \mathcal{M}_{2,12}^{3*b}$ .



According to the first independence, the relationship  $1 \subset \mathcal{M}_{2,12}^{3^*b} \subset (1, 2, 3, 6)$  must hold and, according the second independence the  $2 \subset \mathcal{M}_{2,12}^{3^*b} \subset (1, 2, 3, 6)$  must hold. The **(C2a)** also leads two independencies  $4 \perp\!\!\!\perp 6 | 1, 2, 3, 5$  and  $3 \perp\!\!\!\perp 5 | 1, 2, 4, 6$  and the smaller marginal set where  $(4, 6)$  and  $(3, 5)$  occur, is exactly  $\mathcal{M}_2^*$ . Thus, theorem 7 guarantees that these parametrizations correspond to a marginal model. Since  $\mathcal{H}_{II}^{MIX} = \mathcal{H}_{II}^*$ , it is easy to see that even for theorem 12 the associated graphical model is smooth.

Example 37 shows a graph representable only by the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3^*b}), \mathcal{M} \in \mathcal{H}_{II}^{MIX}\}$ .

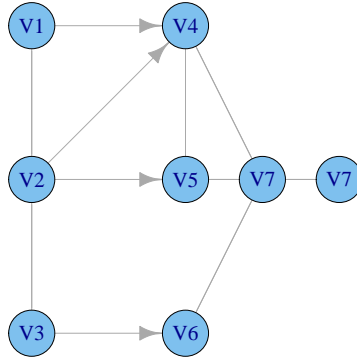


FIGURE 5.3.8

**Example 37.** Theorem 12 is the only one that assures the smoothness of **GM II** associated with the previous graph. In fact, since there are non connected sets  $(4, 8)$ ,  $(5, 6)$ ,  $(5, 8)$ ,  $(6, 8)$  and  $(5, 6, 8)$  which do not belong to the class  $J_h$ , the theorem 10 doesn't hold. Besides, with regard to the second parametrization proposed, we have class  $PA_2 = \{2; 1; 3\}$ . Note that, according the first two sets  $CH_2 \setminus ch(2) = 6$  and  $CH_2 \setminus ch(1) = 5, 6$ , Lemma 2 holds. Instead, according to the third set  $CH_2 \setminus ch(3) = 4, 5$ , since  $(5, 6) \notin (4, 5)$ , Lemma 2 does not hold. Finally, note that, for all vertices  $V_j$  in  $CH_2$ , the sets  $Nb(V_j)$  are complete. Next we will show that we get the same results by applying theorem 7 to these parametrizations.

We should make some comments about these three parametrizations proposed.

If a graphical model permits the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b})\}$ ,  $\mathcal{M} \in \mathcal{H}'_{II}\}$  or the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b})\}$ ,  $\mathcal{M} \in \mathcal{H}^*_{II}\}$ , then it is parametrizable with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b})\}$ ,  $\mathcal{M} \in \mathcal{H}^{MIX}_{II}\}$ .

If a graph has the sets  $NC_h \cup NA_h = \emptyset$  for all  $h = 1, \dots, s$ , then the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}'_{II}}$  and  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}^{MIX}_{II}}$  are equivalent.

#### 5.4. Graphical Models with MP III

Graphical models III were introduced by Drton in (2009) [22] to complete the treatment on graphical models for chain graphs and are characterized by conditions **(C1)**, **(C2b)** and **(C3a)**. The *first step* of our method consists of defining the class of marginal sets:

- According to **(C1)**, we must consider the marginal sets  $\mathcal{M}_h^1 = \cup_{j=1}^h T_j$ .
- According to **(C2b)**, the variables involved are  $A \cup T_h \setminus Nb(A) \cup pa_D(T_h)$ ,  $\forall A : A \in \mathcal{P}(T_h) \setminus \emptyset$ . Note that,  $A \cup T_h \setminus Nb(A)$ ,  $\forall A \in \mathcal{P}(T_h) \setminus \emptyset$ , is contained in the list of all not connected subsets of  $T_h$ . Hence, the following marginal sets contain all sets of interest:

$$(5.4.1) \quad \mathcal{M}_{h,K}^{2b} = K \cup pa_D(T_h), \quad \forall K \in \mathcal{K}_h,$$

where  $\mathcal{K}_h$  is the family of all non connected subsets of  $T_h$ .

- According to **(C3a)**,  $\mathcal{M}_{h,A}^{3a'} = Nb(A) \cup pa_D(T_h) \quad \forall A \in \mathcal{P}(T_h) \setminus J_h$  where  $J_h = \{A : pa_G(A) = pa_D(T_h)\}$ . Note that,  $\mathcal{M}_{h,A}^{3a'} \neq \mathcal{M}_{h,A}^{3a}$ , where  $\mathcal{M}_{h,A}^{3a}$  were introduced for the graphical model of type I.

We list the marginal sets  $\mathcal{M}_{h,A}^{3a'}$ ,  $\mathcal{M}_{h,K}^{2b}$  and  $\mathcal{M}_h^1$  according to the hierarchical property of the marginal sets  $\mathcal{H}_{III}$ .

The *second step* consists of specifying the classes of interaction following the hierarchical and complete properties.

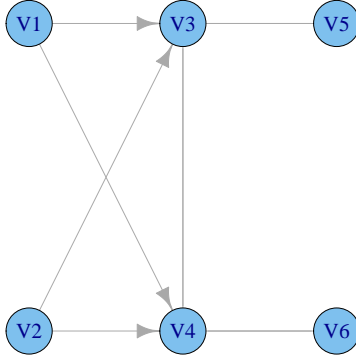


FIGURE 5.4.1

**Example 38.** Let us consider the previous graph. It is possible to note that conditions **(C2b)** and **(C3a)** produce statements of independence only on the component  $T_3$ . Thus we may recognize the following classes of interest:  $\mathcal{K}_3 = \{3, 6; 4, 5; 5, 6; 3, 5, 6; 4, 5, 6\}$  and  $\mathcal{P}(T_3) \setminus \mathcal{J}_3 = \{5; 6; 5, 6\}$ . The hierarchical class of marginal sets is:  $\mathcal{M}_1^1 = 1$ ,  $\mathcal{M}_2^1 = (1, 2)$ ,  $\mathcal{M}_{3,5}^{3a'} = (1, 2, 3, 5)$ ,  $\mathcal{M}_{3,36}^{2b} = (1, 2, 3, 6)$ ,  $\mathcal{M}_{3,45}^{2b} = (1, 2, 4, 5)$ ,  $\mathcal{M}_{3,6}^{3a'} = (1, 2, 4, 6)$ ,  $\mathcal{M}_{3,56}^{2b} = (1, 2, 5, 6)$ ,  $\mathcal{M}_{3,356}^{2b} = (1, 2, 3, 5, 6)$ ,  $\mathcal{M}_{3,456}^{2b} = (1, 2, 4, 5, 6)$ ,  $\mathcal{M}_{3,34}^{3a'} = (1, 2, 3, 4, 5, 6)$ .

Next, in the *third step*, the classes of interaction sets associated with the null parameters are reported, according to formula 4.3.2.

- $D_h^1 = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(\cup_{j=1}^h T_j) \setminus (\mathcal{P}(T_h \cup pa_D(T_h)) \cup \mathcal{P}(\cup_{j=1}^{h-1} T_j))\}, \forall h = 1, \dots, s.$
- $D_h^{2b} = \cup_{A \subset T_h} D_{h,A}^{2b}, \forall h = 1, \dots, s,$  where  $D_{h,A}^{2b} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \setminus nb(A) \cup pa_D(T_h)) \setminus (\mathcal{P}(A \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus Nb(A) \cup pa_D(T_h)))\}.$
- $D_h^{3a'} = \cup_{A \in \mathcal{P}(T_h) \setminus J_h} D_{h,A}^{3a'}, \forall h = 1, \dots, s,$  where  $D_{h,A}^{3a'} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(Nb(A) \cup pa_D(T_h)) \setminus (\mathcal{P}(Nb(A) \cup pa_G(A)) \cup \mathcal{P}(nb(A) \cup pa_D(T_h)))\}.$

Obviously classes  $D_h^1$  are the same presented in the previous sections.

In the *last step*, we must check the 4.3.4 requirement of theorem 7. In this case as well, we have three conditions to verify:

$$(5.4.2) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (\cup_{j=1}^h T_j), \quad \forall \mathcal{L} \in D_h^1.$$

$$(5.4.3) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq K \cup pa_D(T_h), \quad \forall \mathcal{L} \in D_{h,K}^{2b}, \forall K \in \mathcal{K}_h.$$

$$(5.4.4) \quad pa_D(T_h) \setminus pa_G(A) \cup nb(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq Nb(A) \cup pa_D(T_h), \quad \forall \mathcal{L} \in D_{h,A}^{3a'}, \forall A \in \mathcal{P}(T_h) \setminus J_h.$$

$\forall h = 1, \dots, s.$  Obviously the first condition (5.4.2) always holds. Furthermore, condition (5.4.3) is always verified. In fact, for all  $\mathcal{L}_K \in D_{h,K}^{2b}$ , where  $\mathcal{L}_K \cap T_h = K$ ,

the marginal set  $\mathcal{M}(\mathcal{L}_K)$  is equal to  $\mathcal{M}_{h,K}^{2b} = K \cup pa_D(T_h)$ . Thus, condition (5.4.3) is always satisfied. On the other hand, condition (5.4.4) is not always satisfied. In fact, there may be a set  $\mathcal{L}_A \in D_h^{3a'}$  such that  $\mathcal{M}(\mathcal{L}_A) = \mathcal{M}_{h,K}^{2b}$ . In this case, if  $A \in \mathcal{K}_h$ , then condition (5.4.4) does not hold because  $nb(A) \not\subseteq \mathcal{M}(\mathcal{L}_A)$ . Indeed the following theorem explains when a graphical model of type III is a marginal model.

**Theorem 13.** *A graphical model of type III is a marginal model with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2b} \cup D_h^{3'a'})$ ,  $\mathcal{M} \in \mathcal{H}_{III}\}$ , if any component of the graph meets both conditions  $\forall A \in \mathcal{P}(T_h) \setminus J_h$ ,  $Nb(A) = T_h$  and  $\mathcal{K}_h \subseteq J_h$ .*

The following example shows the special case of a graph where there are not independencies of type (C3b). In this case **GM IV** and **GM III** are equivalent, and, as we will see in section 5.4, the two parametrizations coincide.

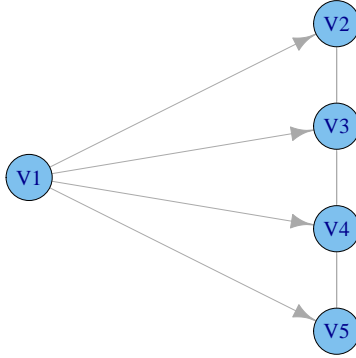


FIGURE 5.4.2

**Example 39.** Let's consider the previous graph, where the list of independencies implied by MP III is  $\{2 \perp\!\!\!\perp 5|1; 2 \perp\!\!\!\perp 4|1; 3 \perp\!\!\!\perp 1; 2 \perp\!\!\!\perp 4, 5|1; 2, 3 \perp\!\!\!\perp 5|1\}$ .

The class of all non connected sets of  $T_2$  is  $\mathcal{K}_2 = \{2, 4; 2, 5; 3, 5; 2, 3, 5; 2, 4, 5\}$ .

The next table reports the interaction sets defined in the respective marginal sets according to the properties of hierarchy and completeness. The elements of  $D_2^{2b}$  are denoted in bold.

<i>Marginal sets</i>	<b>1</b>	<b>1,2,4</b>	<b>1,2,5</b>	<b>1,3,5</b>	<b>1,2,3,5</b>	<b>1,2,4,5</b>	<b>1,2,3,4,5</b>
<i>Interaction sets</i>	1	2	5	3	2,3	4,5	3,4
		4	1,5	1,3	1,2,3	1,4,5	1,3,4
		<b>2,4</b>	<b>2,5</b>	<b>3,5</b>	<b>2,3,5</b>	<b>2,4,5</b>	2,3,4
		1,2	<b>1,2,5</b>	<b>1,3,5</b>	<b>1,2,3,5</b>	<b>1,2,4,5</b>	3,4,5
		1,4					1,2,3,4
		<b>1,2,4</b>					1,3,4,5
							2,3,4,5
						1,2,3,4,5	

TABLE 2

Note that all non connected sets belong to class  $J_h$ . Furthermore, class  $J_h = \mathcal{P}(T_h)$  thus, both assumptions of theorem 13 are verified.

The next example shows the relationships between the **GM I** and **GM III**; in fact, when any component of the graph is complete the two models are equal.

**Example 40.** We take the list of independencies from graph 5.3.4 with **MP III**, that is  $\{4 \perp\!\!\!\perp 2, 3|1, 5, 6, 7; 5 \perp\!\!\!\perp 1, 3|2, 4, 6, 7; 6 \perp\!\!\!\perp 1, 2|3, 4, 5, 7; 7 \perp\!\!\!\perp 1, 2, 3|4, 5, 6\}$ . It's easy to see that the union of all sets involved in the independencies is always equal to  $T_h \cup pa_D(T_h) = (1, 2, 3, 4, 5, 6, 7)$ . Thus, there is only one marginal set  $\mathcal{M}_h^{3a'} = T_h \cup pa_D(T_h)$ . Furthermore, since components  $T_1$  and  $T_2$  are complete, the classes  $\mathcal{K}_1$  and  $\mathcal{K}_2$  are empty. Thus this graphical model is smooth.

Finally, the next example introduces a graphical model III that is smooth.

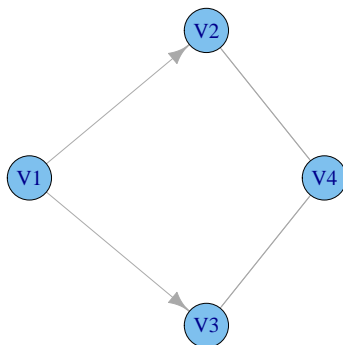


FIGURE 5.4.3

**Example 41.** In the underlying graph we may recognize the following independencies according to the Markov properties III:  $\{2 \perp\!\!\!\perp 3|1; 4 \perp\!\!\!\perp 1|2, 3\}$ . From this graph we have the classes  $J_2 = \{2; 3; 2, 3; 2, 4; 3, 4; 2, 3, 4\}$  and  $\mathcal{P}(T_2) \setminus J_2 = \{4\}$ . The condition of theorem 13 holds is met since  $Nb(4) = 2, 3, 4 = T_2$  and the only non connected set of  $T_2$ , that is  $K = 2, 3$  belongs to  $J_2$ .

### 5.5. Graphical Models with MP IV

The Markov Property of type IV is characterized by conditions **(C1)**, **(C2b)** and **(C3b)**. Bergsma Rudas and Németh in [9] and Marchetti and Lupporelli in [47] showed that the graphical model, where the independencies obey **MP IV**, is smooth. In this chapter we will come to the same results by applying our method explained in section 4.3.1.

In the *first step* of our method, we define the class of marginal sets  $\mathcal{H} = \{A_i \cup B_i \cup C_i, i = 1, \dots, k\}$ , for any independence  $A_i \perp\!\!\!\perp B_i|C_i$ .



- $\mathcal{M}_h^1 = \cup_{j=1}^h T_j$ , from condition **(C1)**;
- $\mathcal{M}_{h,K}^{2b} = pa_D(T_h) \cup K$ ,  $\forall K \in \mathcal{K}_h$ , from condition **(C2b)**, for more details about this marginal set see section 5.3;
- $\mathcal{M}_{h,A}^{3b} = pa_D(T_h) \cup A$ ,  $\forall A \in \mathcal{P}(T_h) \setminus J_h$ , from condition **(C3b)**;

Note that, the first class of the marginal sets is the same for all parametrization proposed. The marginal sets  $\mathcal{M}_{h,K}^{2b}$  are the same proposed for GM III and the marginal sets  $\mathcal{M}_{h,A}^{3a}$  are described in section 5.2.1 for the graphical model II.

The last two types of marginal sets can be compacted in the following way:

$$\mathcal{M}_{h,A}^{IV} = pa_D(T_h) \cup A, \quad A \in \mathcal{P}(T_h) \setminus (J_h \cap \mathcal{R}_h).$$

Where  $\mathcal{R}_h = \mathcal{P}(T_h) \setminus \mathcal{K}_h$  is the class of all connected subsets of  $T_h$ . Obviously  $\forall h = 1, \dots, s$ ,  $\mathcal{M}_{h,A}^{IV} \subseteq \mathcal{M}_h^1$ , and when the two sets are equivalent we only define the set  $\mathcal{M}_{h,A}^1$ .

The hierarchical class of marginal sets is

$$\mathcal{H}_{IV} = \{(\mathcal{M}_{h,A}; \mathcal{M}_h), A \in \mathcal{P}(T_h) \setminus (J_h \cap \mathcal{R}_h)\}_{h=1, \dots, s}$$

where the sets  $\mathcal{M}_{h,A}$  satisfy the hierarchical order.

Once the marginal class is specified, in the *second step*, we define the interaction sets respecting the properties of hierarchy and completeness.

**Example 42.** In the graph in figure 5.3.2, in the component  $T_1$  we may recognize the classes  $J_1 = \{1; 2; 1,2\}$  and  $\mathcal{R}_1 = \{1; 2; 1,2\}$ , thus  $\mathcal{P}(T_1) \setminus (J_1 \cap \mathcal{R}_1) = \emptyset$ . Instead, looking at the component  $T_2$ , we can recognize the class  $J_2 = \{4; 5; 3,4; 3,5; 4,5; 4,6; 5,6; 3,4,5; 3,4,6; 3,5,6; 4,5,6; 3,4,5,6\}$  and the class  $\mathcal{R}_2 = \{3; 4; 5; 6; 3,4; 3,6; 4,5; 5,6; 3,4,5; 3,4,6; 3,5,6; 4,5,6; 3,4,5,6\}$ . Hence, we choose any subset  $A$  of  $T_h$  that belongs to  $\mathcal{P}(T_2) \setminus (J_2 \cap \mathcal{R}_2) = \{3; 6; 3,5; 3,6; 4,6\}$ . The marginal sets of this graph with the interaction sets, according to MP IV are  $\mathcal{M}_1^1 = (1,2)$ ,  $\mathcal{M}_{2,3}^{IV} = (1,2,3)$ ,  $\mathcal{M}_{2,6}^{IV} = (1,2,6)$ ,  $\mathcal{M}_{2,35}^{IV} = (1,2,3,5)$ ,  $\mathcal{M}_{2,36}^{IV} = (1,2,3,6)$ ,  $\mathcal{M}_{2,46}^{IV} = (1,2,4,6)$ ,  $\mathcal{M}_2^1 = (1,2,3,4,5,6)$ .

In the *third step*, we apply the formula 4.3.2 to the three conditions identifying the following three classes:

- $D_h^1 = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(\cup_{j=1}^h T_j) \setminus (\mathcal{P}(T_h \cup pa_D(T_h)) \cup \mathcal{P}(\cup_{j=1}^{h-1} T_j))\}, \forall h = 1, \dots, s.$
- $D_h^{2b} = \cup_{A \subset T_h} D_{h,A}^{2b}, \forall h = 1, \dots, s$  where  $D_{h,A}^{2b} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \setminus nb(A) \cup pa_D(T_h)) \setminus (\mathcal{P}(A \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus Nb(A) \cup pa_D(T_h)))\}.$
- $D_h^{3b} = \cup_{A \in \mathcal{P}(T_h) \setminus J_h} D_{h,A}^{3b}, \forall h = 1, \dots, s$  where  $D_{h,A}^{3b} = \{\mathcal{L} : \mathcal{L} \in (\mathcal{P}(A \cup pa_D(T_h)) \setminus \mathcal{P}(A \cup pa_G(T_h))) \cup \mathcal{P}(pa_D(T_h))\}.$

According to theorem 7, the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}}\}_{\mathcal{M} \in \mathcal{H}_{IV}}$  is able to describe a graphical model IV if the following conditions are satisfied.

$$(5.5.1) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (\cup_{j=1}^h T_j), \quad \forall \mathcal{L} \in D_h^1.$$

$$(5.5.2) \quad pa_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T \setminus nb(K) \cup pa_D(T_h), \quad \forall \mathcal{L} \in D_{h,K}^{2b} \quad \forall K \in \mathcal{K}_h.$$

$$(5.5.3) \quad pa_G(A) \subseteq \mathcal{M}(\mathcal{L}) \subseteq A \cup pa_D(T_h), \quad \forall \mathcal{L} \in D_{h,A}^{3b}, \quad \forall A \in \mathcal{P}(T_h) \setminus J_h.$$

$\forall h = 1, \dots, s$ . It is easy to see that,  $\forall \mathcal{L} \in D_{h,K}^{2b}$ ,  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,K}^{IV} = K \cup pa_D(T_h)$ . Thus, condition (5.5.2) are always satisfied. Furthermore  $\forall \mathcal{L} \in D_{h,A}^{3b}$ ,  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{IV} = A \cup pa_D(T_h)$ , hence, even condition (5.5.3) is checked. Since condition (5.5.1) is always satisfied, the validity of the following theorem is evident.

**Theorem 14.** *A graphical model of type IV is always a marginal model with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2b} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}_{IV}\}$  and therefore it is always smooth.*

Note that, if we take a chain graph with all complete components, the condition **(C2b)** does not produce any independence, and the independencies obtained using the **MP II** and **MP IV** are the same and the marginal class  $\mathcal{H}_{IV}$  is equal to  $\mathcal{H}'_{II}$ .

## Appendix to Chapter 5

### Proof of Theorem 10 (Section 5.3.1)

**Theorem.** *A graphical model of type II is a marginal model of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3b}), \mathcal{M} \in \mathcal{H}'_{II}\}$  if,  $\forall h = 1, \dots, s, \mathcal{K}_h \subseteq J_h$ .*

PROOF. In order to prove this theorem we must see when conditions (5.3.2), (5.3.3) and (5.3.4), obtained by applying the formula 4.3.4 to **(C1)**, **(C2a)** and **(C3b)** hold. As shown in chapter 5.3.1, the only problematic situation occurs in condition (5.3.4). In fact, it may happen that, given an interaction set  $\mathcal{L} \in D_{h,A}^{2a}$ , there is a set  $B \subseteq A \cup nb(A)$  such that  $\mathcal{L} \in D_{h,B}^{3b}$ . In this case the first marginal set which contains  $\mathcal{L}$  is  $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,B}^{3b} = B \cup pa_D(T_h)$ , and the condition (5.3.4) becomes  $pa_D(T_h) \cup nb(A) \subseteq \mathcal{M}_{h,B}^{3b} \subseteq T_h \cup pa_D(T_h)$ , that holds only for strong condition  $B = nb(A) \cup A$ .

On the other hand, if  $D_h^{3b} \cap D_h^{2a} = \emptyset$  theorem 7 is satisfied. Since the interaction sets  $\mathcal{L} \in D_h^{2a}$  are such that  $\mathcal{L} \cap T_h \in \mathcal{P}(T_h) \setminus \mathcal{C}_h$  and  $\mathcal{L} \in D_h^{3b}$  are such that  $\mathcal{L} \cap T_h \in \mathcal{P}(T_h) \setminus J_h$ , then if  $\mathcal{P}(T_h) \setminus \mathcal{C}_h \subseteq J_h$  then  $D_h^{3b} \cap D_h^{2a} = \emptyset$ .

Note that, any element  $A$  of the class of non-complete sets  $\mathcal{P}(T_h) \setminus \mathcal{C}_h$  contains a non-connected set  $K$ . Thus, if the class of all non-connected sets  $\mathcal{K}_h \in J_h$  then, even  $\mathcal{P}(T_h) \setminus \mathcal{C}_h \in J_h$ . Thus, we may conclude that the condition of theorem 7 is never violated if  $\mathcal{K}_h \subseteq J_h$ . □

**Proof of Lemma 2 (section 5.3.2)**

**Lemma.** *A graphical model of type II is a marginal model of parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b}), \mathcal{M} \in \mathcal{H}_{II}^*\}$ . if both the conditions are satisfied:*

- (a): *All components of the graph are complete;*
- (b):  $\forall \mathcal{A}, \mathcal{B} \in PA_h$ :  $\mathcal{B} < \mathcal{A}$  in  $PA_h$ , if  $(CH_h \setminus ch(\mathcal{A})) \cap (CH_h \setminus ch(\mathcal{B})) \neq \emptyset$ , then  $(CH_h \setminus ch(\mathcal{B})) \subseteq (CH_h \setminus ch(\mathcal{A}))$ ,  $h = 1, \dots, s$ .

We consider a chain graph with all complete components. Thus, there is not any independence associated with **(C2a)** and the marginal class is  $\mathcal{H}_{II} = \{(\mathcal{M}_{h,\mathcal{A}}^{3^*b}, \mathcal{M}_h^1), \forall \mathcal{A} \in PA_h^+\}$ , where  $\mathcal{M}_{h,\mathcal{A}}^{3^*b} = pa_D(T_h) \cup T_h \setminus ch(\mathcal{A})$ . Section 5.2.2 introduces three conditions (5.3.7), (5.3.8) and (5.3.9) related to **(C1)**, **(C2a)** and **(C3b)**, which, if satisfied, allow us to conclude that the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b}), \mathcal{M} \in \mathcal{H}_{II}^*\}$  is a graphical model of type II. Since (5.3.7) always holds and since **(C2a)** does not produce independencies when components are complete, we must see if condition (5.3.9) is met.

For every  $A \subseteq T_h \setminus ch(\mathcal{A})$ ,  $\forall \mathcal{A} \in PA_h^+$ , let  $\mathcal{L}_A$  be an interaction set such that  $\mathcal{L}_A \cap T_h = A$ . Let  $\mathcal{B}$  be a set in  $PA_h$  such that  $\mathcal{B} < \mathcal{A}$ . Following are the possible situations which may arise:

When  $\mathcal{L}_A : \mathcal{M}(\mathcal{L}_A) = \mathcal{M}_{h,\mathcal{A}}^{3^*b} = pa_D(T_h) \cup T_h \setminus ch(\mathcal{A})$ , then condition (5.3.9) is always met.

When  $\mathcal{L}_A : \mathcal{M}(\mathcal{L}_A) = \mathcal{M}_{h,\mathcal{B}}^{3^*b} = pa_D(T_h) \cup T_h \setminus ch(\mathcal{B})$ , and  $\mathcal{B} = pa_D(T_h)$ , the marginal set becomes  $\mathcal{M}_{h,\mathcal{B}}^{3^*b} = pa_D(T_h) \cup NC_h \cup NA_h$  and the condition (5.3.9) becomes  $pa_D(T_h) \setminus \mathcal{A} \subseteq (pa_D(T_h) \cup NC_h \cup NA_h) \subseteq pa_D(T_h) \cup T_h \setminus ch(\mathcal{A})$  which always holds.

When  $\mathcal{L}_A : \mathcal{M}(\mathcal{L}_A) = \mathcal{M}_{h,\mathcal{B}}^{3^*b} = pa_D(T_h) \cup T_h \setminus ch(\mathcal{B})$ , and  $\mathcal{B} \neq pa_D(T_h)$ , condition (5.3.9) becomes  $pa_D(T_h \setminus \mathcal{A}) \subseteq (pa_D(T_h) \cup T_h \setminus ch(\mathcal{B})) \subseteq pa_D(T_h) \cup T_h \setminus ch(\mathcal{A})$  which is met only if  $T_h \setminus ch(\mathcal{B}) \subseteq T_h \setminus ch(\mathcal{A})$  or, equivalently  $CH_h \setminus ch(\mathcal{B}) \subseteq CH_h \setminus ch(\mathcal{A})$ .

Note that if  $(CH_h \setminus ch(\mathcal{A})) \cap (CH_h \setminus ch(\mathcal{B})) = \emptyset$ ,  $\forall \mathcal{A}, \mathcal{B} \in PA_h$ , then the problematic situation seen in the last point never occurs.

### Proof of Theorem 11 (section 5.3.2)

**Theorem.** *A graphical model of type II is a marginal model with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b})$ ,  $\mathcal{M} \in \mathcal{H}_{II}^*\}$ , if the assumption (b) of Lemma 2 holds and if, for all  $V_j \in CH_h$  such that  $Nb(V_j) \notin \mathcal{C}_h$ ,  $\{K : K \in \mathcal{K}_h; K \cap nb(V_j) \neq \emptyset\} \subseteq J_h$ .*

Given a graph, the parametrization  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b})$ ,  $\mathcal{M} \in \mathcal{H}_{II}^*\}$  is a graphical model II if conditions (5.3.7), (5.3.8) and (5.3.9) proposed in section 5.2.2 are met. We have already seen that (5.3.7) is always true, irrespective of the type of graphical model, and condition (5.3.9) holds when the assumption (b) of Lemma 2 is satisfied.

The next step consists of checking when condition (5.3.8) holds. Let remember that the condition **(C2\*a)** is described by the following list of independencies:

$$B_{1i} \perp\!\!\!\perp T_h \setminus B_i \mid pa_D(T_h) \cup B_i \setminus B_{1i} \quad \forall B_{1i} \subseteq B_i \in B^*$$

$$(\mathbf{B}_{1,V_j} \cup V_j) \perp\!\!\!\perp T_h \setminus Nb(\mathbf{B}_{1,V_j} \cup V_j) \mid pa_D(T_h) \cup nb(\mathbf{B}_{1,V_j} \cup V_j) \quad \forall V_j \in T_h \setminus (\cup_{i: B_i \in B^*} B_{1i})$$

where  $\mathbf{B}_{1,V_j} = (\cup_{B_{1k} \subset nb(V_j)} B_{1k})$ ,  $B_h^* = \{B_i : B_i \in Cl_h, B_i = B_{1i} \cup B_{2i}, Nb(B_{1i}) = B_i, CH_h \cap B_i \neq \emptyset\}$  and  $Cl_h$  is the class of cliques of the component  $T_h$ .

Further,  $nb(B_{1i}) = B_{2i}$  and  $nb(B_{2i}) = B_{1i} \cup nb(B_i)$ , where in  $nb(B_i)$  can be  $B_{1h}$  and  $B_{2h}$ , for some  $h$ .

If  $CH_h \subseteq (\cup_{i:B_i \in B^*} B_{1i})$ , condition (5.3.8) holds because, in this case,  $(pa_D(T_h) \cup B_i \setminus B_{1i}) \cap CH_h = \emptyset$  and also

$(pa_D(T_h) \cup nb(\mathbf{B}_{1,V_j} \cup V_j)) \cap CH_h = \emptyset$  as it follows from  $nb(\mathbf{B}_{1,V_j} \cup V_j) = B_i \setminus (B_{1i} \cup V_j) \cup (nb(V_j) \cap NC_h)$ . Condition  $CH_h \subseteq (\cup_{i:B_i \in B^*} B_{1i})$ , occurs if and only if,  $\forall V_j \in CH_h$ ,  $Nb(V_j)$  is a complete set.

In fact, all  $B_{1i}$  and  $Nb(B_{1i})$  are complete sets, so  $\forall V_j \in B_{1i}$ ,  $Nb(V_j)$  are also complete. If  $CH_h \subseteq (\cup_{i:B_i \in B^*} B_{1i})$ , then  $Nb(V_j)$  must be a complete set,  $\forall V_j \in CH_h$ .

On the contrary, if  $V_j \in CH_h$  is such that  $Nb(V_j)$  is complete set, then it is even a maximal complete set, thus we can choose a set  $B_i = Nb(V_j)$ , and, by definition, the vertex  $V_j$  belongs to the set  $B_{1i}$ . This holds for every  $V_j \in CH_h$ . Hence, if  $\forall V_j \in CH_h$ ,  $Nb(V_j)$  is complete,  $CH_h \subseteq (\cup_{i:B_i \in B^*} B_{1i})$ .

Instead, if it is  $CH_h \not\subseteq (\cup_{i:B_i \in B^*} B_{1i})$ , there is at least one vertex  $V_k \in CH_h \cap B_i \setminus B_{1i}$  having  $Nb(V_k)$  non complete. In this case, the conditional set of ??, that is  $pa_D(T_h) \cup B_i \setminus B_{1i}$ , contains the set  $V_k$  when  $V_k \in nb(B_{1i})$ , that occurs for every  $B_{1i} \subseteq nb(V_k)$ . Furthermore, conditional set of ??, that is  $pa_D(T_h) \cup nb(\mathbf{B}_{1,V_j} \cup V_j)$ , contains  $V_k$  whether  $V_k$  is neighbor of  $V_j \in B_i \setminus B_{1i}$ , that is for every  $V_j \in nb(V_k)$ .

For the effectiveness of condition (5.3.8), it is necessary that, according to independencies in ??  $\forall B_{1i} \subseteq \mathbf{B}_{1,V_k}$ , and in ??  $\forall V_j \in B_i \setminus B_{1i}$  such that  $V_k \in nb(V_j)$ , the corresponding interaction sets in  $D_{h,B_{1i}}^{2a}$  e  $D_{h,(\mathbf{B}_{1,V_k} \cup V_k)}^{2a}$  will be allocated in the marginal  $\mathcal{M}_h^{2a}$ .

In particular, from ??,  $\forall B_{1i} \subseteq \mathbf{B}_{1,V_k}$ , the interaction sets in  $D_{h,B_{1i}}^{2a} = \{\mathcal{P}(T_h \cup pa_D(T_h)) \setminus (\mathcal{P}(B_i \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus B_{1i} \cup pa_D(T_h)))\}$  must belong to  $J_h$ . Note that, in  $D_{h,B_{1i}}^{2a}$  there are all non complete sets with at least one element in  $B_{1i}$ . But, if all non-connected sets containing at least one element of  $B_{1i}$ , for all  $B_{1i} \subset nb(V_k)$ , that is  $\{K : K \in \mathcal{K}_h; K \cap \mathbf{B}_{1,V_k} \neq \emptyset\}$ , belong to  $J_h$ , this is even more so for all sets in  $D_{h,B_{1i}}^{2a} \subseteq J_h$ .

Similarly, from ??,  $\forall V_j \in B_i \setminus B_{1i}$  such that  $V_j \in nb(V_k)$ , the class  $\{\mathcal{P}(T_h \cup pa_D(T_h)) \setminus (\mathcal{P}(Nb(\mathbf{B}_{1,V_j} \cup V_j) \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus (\mathbf{B}_{1,V_j} \cup V_j) \cup pa_D(T_h)))\}$  must belong to  $J_h$ . Even in this case, if all sets  $\{K : K \in \mathcal{K}_h; K \cap (V_k \cup \mathbf{B}_{1,V_k}) \neq \emptyset\}$  belong to  $J_h$ , then all sets in  $D_{h,(\mathbf{B}_{1,V_k} \cup V_k)}^{2a} \subseteq J_h$ .

Summarizing,  $\forall V_k$  such that  $Nb(V_k) \notin \mathcal{C}_h$ ,  $\{K : K \in \mathcal{K}_h; K \cap nb(V_k) \neq \emptyset\} \subseteq J_h$ .

### Proof of Theorem 12 (section 5.3.3)

**Theorem.** *A graphical model of type II is a marginal model with  $\{\eta_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2a} \cup D_h^{3*b})\}$ ,  $\mathcal{M} \in \mathcal{H}_{II}^{MIX}\}$ , if, for all  $V_j \in CH_h$  such that  $Nb(V_j) \notin \mathcal{C}_h$ ,  $\{K : K \in \mathcal{K}_h; K \cap nb(V_j) \neq \emptyset\} \subseteq J_h$ .*

We have already seen in Chapter 5.3.3 that condition 5.3.10 is always met. Since  $(NC_h \cup NA_h) \subseteq \mathcal{M}_{h,A}^{3bMIX}$ ,  $\forall A \in \mathcal{P}(T_h) \setminus J_h$ , then, according to theorem 11, if  $\forall V_j \in CH_h$ , such that  $Nb(V_j) \notin \mathcal{C}_h$ ,  $\{K : K \in \mathcal{K}_h; K \cap nb(V_j) \neq \emptyset\} \subseteq J_h$ , condition 5.3.11 always holds.

Now we must to check condition (5.3.12) that is  $pa_D(T_h) \setminus \mathcal{A} \subseteq \mathcal{M}(\mathcal{L}) \subseteq pa_D(T_h) \cup T_h \setminus ch(\mathcal{A})$ ,  $\forall \mathcal{L} \in D_h^{3*b}$ .

Let  $\mathcal{L}_A$  be an interaction set such that  $\mathcal{L}_A \in D_{h,A}^{3*b}$ ,  $\mathcal{L}_A \cap T_h = A$ . Since  $D_{h,A}^{3*b} = \{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h)) \setminus \mathcal{P}(pa_D(T_h)) \cup \mathcal{P}(T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h) \setminus \mathcal{A})\}$ , if



$\mathcal{L}_A \in D_{h,A}^{3^*b}$ , then  $A \subseteq T_h \setminus ch(\mathcal{A})$ . Then  $\forall A \in \mathcal{P}(T_h) \setminus J_h$ , it is  $\mathcal{M}(\mathcal{L}_A) = \mathcal{M}_{h,B}^{3bMIX} = pa_D(T_h) \cup (NC_h \cup NA_h) \cup B$ , where  $B = A \cap CH_h$ . Thus, condition (5.3.12) that becomes  $pa_D(T_h) \setminus \mathcal{A} \subseteq pa_D(T_h) \cup (NC_h \cup NA_h) \cup B \subseteq pa_D(T_h) \cup T_h \setminus ch(\mathcal{A})$ , holds  $\forall \mathcal{A} : A \in T_h \setminus ch(\mathcal{A})$ .

### Proof of Theorem 13 (section 5.4)

**Theorem.** *A graphical model of type III is a marginal model with  $\{\eta_{\mathcal{L}}^M : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^s (D_h^1 \cup D_h^{2b} \cup D_h^{3'a})\}$ ,  $\mathcal{M} \in \mathcal{H}_{III}\}$ , if any component of the graph meets the both conditions:  $\forall A \in \mathcal{P}(T_h) \setminus J_h$ ,  $Nb(A) = T_h$  and  $\mathcal{K}_h \subseteq J_h$ .*

As before, we may verify conditions (5.4.3) and (5.4.4) proposed in Chapter 5.3, since condition (5.4.2) always holds. With regard to condition (5.4.3),  $\forall \mathcal{L}_K \in D_{h,K}^{2b}$ , where  $\mathcal{L}_K$  is such that  $\mathcal{L}_K \cap T_h = K$ , the first marginal set where it appears is  $\mathcal{M}(\mathcal{L}_K) = K \cup pa_D(T_h) = \mathcal{M}_{h,K}^{2b}$ . Thus, even condition (5.4.3) always holds. The same is not true for the third condition. In fact two problematic situations may occur:

$\exists \mathcal{L} \in D_{h,A}^{3a'} : \mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,B}^{3a'} = Nb(B) \cup pa_D(T_h)$ , where  $A \subseteq Nb(B)$ ;

$\exists \mathcal{L} \in D_{h,A}^{3a'} : \mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,K(A)}^{2b}$ , where  $K(A)$  is the first set in  $\mathcal{K}_h$  so that  $A \subset K(A)$ .

Let  $\mathcal{L}_A$  be a set such that  $\mathcal{L}_A \cap T_h = A$ . With regard the first situation, note that if  $A \subset Nb(B)$  and  $Nb(A) \subseteq Nb(B)$ , then  $\mathcal{M}(\mathcal{L}_A) = \mathcal{M}_{h,A}^{3a'}$ . If, instead,  $Nb(A) \not\subseteq Nb(B)$  then condition (5.4.4) is not satisfied, because  $pa_G(A) \cup nb(A) \not\subseteq \mathcal{M}_{h,B}^{3a'}$ . Note that if a set  $A \in \mathcal{P}(T_h) \setminus J_h$  has  $nb(A)$  so that  $Nb(A) = T_h$ , we have exactly one marginal set  $\mathcal{M}_{h,A}^{3a'} = T_h \cup pa_D(T_h)$ , and the first case never occurs.

With regard to the second condition,  $K(A) = A$  if  $A \in \mathcal{K}_h$ , or  $K(A) = A \cup V_j$ , where  $V_j \not\subseteq nb(A)$ , if  $A$  is connected. In both cases it is easy to see that condition

(5.4.4) is not met because  $pa_G(A) \cup nb(A) \notin \mathcal{M}_{h,K(A)}^{2b}$ . Note that, if  $\mathcal{K}_h \subseteq J_h$ , this second situation cannot occur. Thus, if  $A \in \mathcal{P}(T_h) \setminus J_h$ ,  $Nb(A) = T_h$  and  $\mathcal{K}_h \subseteq J_h$ , the condition (5.4.4) is satisfied.



## CHAPTER 6

### **Application based on data from the European Values Study (EVS), 2008**

In this chapter we illustrate the results obtained on **GM II** through an application based on data from the European Values Study (EVS) [24], 2008. This study is a research project on how Europeans think about *family, work, religion, politics* and *society*. For our purposes, we considered only the Italian data-set concerning 1520 individuals where we select the six variables described above:

*A*: Range of age (20 → 40, 40 → 60, > 60)

*C*: Children (Yes, No)

*E*: Employed (Yes, No)

*T*: Trust in the people (Yes, No)

*LS*: Life satisfaction (High; Low)

*OS*: Opinion on Society (High, Mean, Low)

We divided the variables into three groups. In the first group we placed the variables concerning the personal data of the respondents: *A*. In the second group there were variables regarding the social life of the respondents *C* and *E*. Finally, the last group regarded the variables perception of life for the respondents *T*, *LS* and *OS*. Each group of variables was represented with a component in the graphs. Thus, we used the components  $T_1 = \{A\}$ ,  $T_2 = \{C, E\}$  and  $T_3 = \{T, LS, OS\}$  to investigate how children and work can influence the opinion variables in  $T_3$

108 APPLICATION BASED ON DATA FROM THE EUROPEAN VALUES STUDY (EVS), 2008  
and, furthermore how they can influence the other variables. To this aim, we considered variable  $A$  as “purely explicative”, variables  $C$  and  $F$  as “intervening” and variables in the component  $T_3$  as “purely response”.

In order to find the most representative graphical model, we started by considering a graph with all complete components and where any vertex is child of any element of the parent component. In this case the only condition that produces independencies is **(C1)**. Thus we tested the independence

$$LS, T, OS \perp\!\!\!\perp A | C, E$$

The result is reported in table 2. Thus, if it is satisfied the previous independence, the DAG associated with these components will be compatible with the graph in figure 6.0.1.

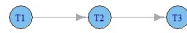


FIGURE 6.0.1

We proceeded by removing one by one the edges in component  $T_3$ . Therefore, we obtained three possible chain graphs depending on the edge removed. In particular, applying condition **(C2a)** we had the following independencies for the three graphs.

$$LS \perp\!\!\!\perp T | OS, C, E \quad LS \perp\!\!\!\perp OS | T, C, E \quad T \perp\!\!\!\perp OS | LS, C, E$$

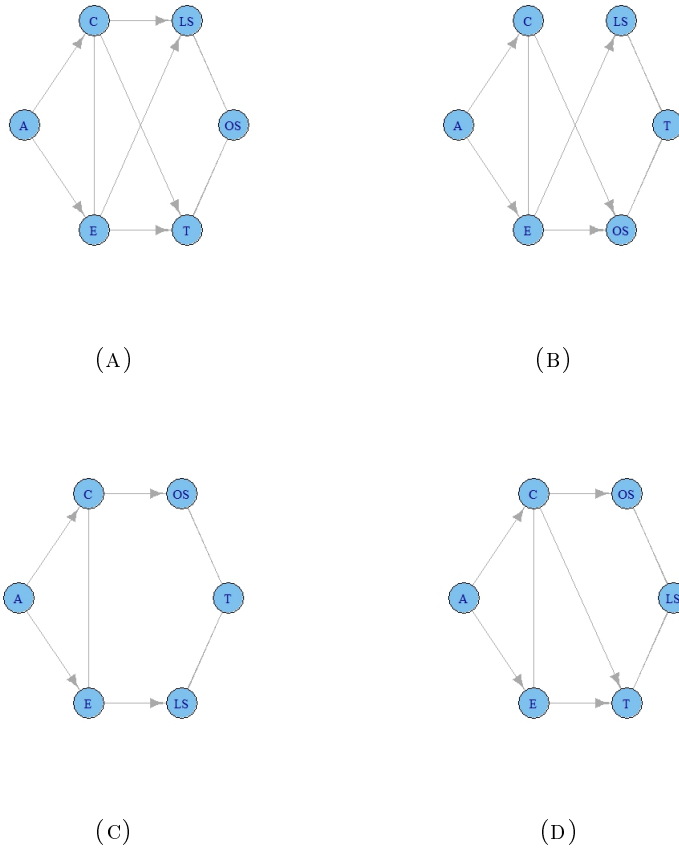


FIGURE 6.0.2. Graphs respectively concerning the models 3.1.1, 3.2.3, 3.2.8 and 3.3.4

Finally, one by one we removed the arcs between the components  $T_2$  and  $T_3$ , respecting the condition of theorem 12.

Table 1 shows the most interesting results for the **GM II**. The table is shared in 5 blocks. The first describes the **GM II** with only condition **(C1)**. The second block refers to the three **GM II** associated with different conditions **(C2a)** described above. The last three blocks refer to the just cited independencies **(C2a)** with different **(C3\*b)**s.

Ref.	Fig.	(C1)	(C2a)	(C3*b)	df	Gsq	p
1	6.0.1	$LS, T, OS \perp A C, E$	-	-	88	70.85958	0.90919
2.1		$LS, T, OS \perp A C, E$	$LS \perp T OS, C, E$	-	100	91.44389	0.71758
2.2		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	-	104	91.33411	0.80783
2.3		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	-	104	98.87589	0.62357
3.1.1	6.0.2a	$LS, T, OS \perp A C, E$	$LS \perp T OS, C, E$	$OS \perp C, E$	106	97.84138	<u>0.7017</u>
3.1.2		$LS, T, OS \perp A C, E$	$LS \perp T OS, C, E$	$C \perp OS E$ $E \perp LS, OS C$	112	135.7669	0.062808
3.1.3		$LS, T, OS \perp A C, E$	$LS \perp T OS, C, E$	$C \perp OS, LS E$ $E \perp OS C$	112	126.852	0.1597
3.1.4		$LS, T, OS \perp A C, E$	$LS \perp T OS, C, E$	$C, E \perp OS, LS$	115	140.4228	0.053715
3.1.5		$LS, T, OS \perp A C, E$	$LS \perp T OS, C, E$	$C \perp OS, LS E$ $E \perp T, OS C$	118	141.769	0.06722
3.2.1		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$T \perp C, E$	107	100.7471	0.65182
3.2.2		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C \perp T E$ $E \perp LS, T C$	111	113.1351	0.4258
3.2.3	6.0.2b	$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C \perp T, SL E$ $E \perp T C$	111	106.5838	<u>0.60083</u>
3.2.4		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C, E \perp LS, T$	113	117.9889	0.35521
3.2.5		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C \perp T, OS E$ $E \perp T C$	115	132.6055	0.1251
3.2.6		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C \perp T E$ $E \perp OS, T C$	115	113.4657	0.52297
3.2.7		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C \perp T, SO E$ $E \perp T, LS C$	119	144.9936	0.052797
3.2.8	6.0.2c	$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C \perp T, LS E$ $E \perp OS, T C$	119	137.1901	<u>0.47496</u>
3.2.9		$LS, T, OS \perp A C, E$	$LS \perp OS T, C, E$	$C, E \perp OS, T$	119	137.1901	0.12169
3.3.1		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$LS \perp C, E$	107	112.1012	0.34874
3.3.2		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C \perp T, LS E$ $E \perp LS C$	111	121.0358	0.24237
3.3.3		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C, E \perp LS, T$	113	125.5307	0.19807
3.3.4	6.0.2d	$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C \perp LS E$ $E \perp OS, LS C$	115	120.7303	<u>0.33887</u>
3.3.5		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C \perp OS, LS E$ $E \perp LS C$	115	144.6259	0.032138
3.3.6		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C \perp OS, LS E$ $E \perp T, LS C$	116	153.0183	0.019388
3.3.7		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C \perp T, LS E$ $E \perp LS, OS C$	119	129.665	0.23743
3.3.8		$LS, T, OS \perp A C, E$	$T \perp OS LS, C, E$	$C, E \perp LS, OS$	119	147.8547	0.037542

TABLE 1. Summary of main **GM II** proposed

The goal of this analysis consist in finding the simplest model which represents the data. From the previous table it is clear that the suitable models are 3.1.1, 3.2.3, 3.2.8 and 3.3.4. In figure (6.0.2) are the graphs associated to these models. The best model for the data is represented in figure 6.0.2c. Below a summary table on this model is reported. In the table, any column is indexed by the marginal sets of  $\mathcal{H}_{II}^{MI^X}$  and the cells contain the interaction sets defined according to the properties of hierarchy and completeness. The bold interactions refer to the null parameters according the model 3.2.4.



<i>Marginal set</i>	$\mathcal{M}_1^1$ =	$\mathcal{M}_2^{2a}$ =	$\mathcal{M}_{3,\emptyset}^{3bMIX}$ =	$\mathcal{M}_{3,LS}^{3bMIX}$ =	$\mathcal{M}_{3,OS}^{3bMIX}$ =	$\mathcal{M}_3^{2a}$ =	$\mathcal{M}_3^1$ =
<i>Interaction sets</i>	A	(A, E, C)	(E, C, T)	(E, C, LS, T)	(E, C, OS, T)	(E, C, LS, T, OS)	(A, E, C, LS, T, OS)
	(A) E	T		(E; C; E; C) LS	(E; C; E; C) OS	(E; C; E; C) LS, OS	(E; C; E; C) A, LS
	(A) C	E, T		(E; C; E; C) LS, T	(E; C; E; C) OS, T	(E; C; E; C) LS, OS, T	(E; C; E; C) A, T
	(A) E, C	C, T					(E; C; E; C) A, OS
		E, C, T					(E; C; E; C) A, LS, T
							(E; C; E; C) A, LS, OS
							(C; E; C; E) A, T, OS
							(E; C; E; C) A, LS, T, OS

TABLE 2. Summary of model 3.2.8

With the symbol (E; C; E, C) LS we denote the interaction sets E, LS; C, LS; C, E, LS.

By the previous table it is easy to see that, applying **GM II**, we have considerably simplified the model to describe the relationships among the variables. In fact, constraining to zero 111 parameters, only 35 parameters to estimate remain.

Since we have made more hypothesis of independencies on the marginal table concerning the variables  $E, C, LS, T, OS$ , it follows the respective distribution.

		<b>LS</b>	Low						High					
		<b>OS</b>	High		Low		Mean		High		Low		Mean	
C	E	<b>T</b>	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
NO	NO		2	0	5	2	20	12	2	1	11	5	87	34
	SI		6	0	7	1	19	8	10	1	23	8	116	61
SI	NO		8	1	7	1	60	13	15	4	9		155	78
	SI		3	1	7	2	44	15	20	3	11	9	183	130

TABLE 3. Main models for the data

We have used the statistical software R (R Core Team (2012)), with the help of the package "hmmm", (that is available from the comprehensive R Archive Network at <http://cran.r-project.org/web/packages/hmmm>) to test marginal models and estimate the parameters and the packages "gRbase" (<http://cran.r-project.org/web/packages/gRbase>) and "RBGL" (<http://www.bioconductor.org/packages/release/bioc/html/RBGL.html>) to plot the graphs.



## Conclusion

In this work we studied the graphical models focusing on the **GM II** and **GM III**. In particular we introduced the subclasses of these models which are also marginal models.

Future research will attempt to prove that the only graphical models corresponding to marginal models are these subclasses.

In addition will be interested to study how to model the non null parameters to representing different relationships of dependence.



## Appendix A

### Glossary of graph terminology

- A graph  $G = \{V, E\}$  is a mathematical object composed by a set of vertices  $V$  and a set  $E$  of unordered pairs of distinct element of  $V$ , called edges or arcs, such that  $E \subseteq V \times V$ .
- $G_A = \{A, E_A\}$  is a subgraph of  $G = \{V, E\}$ , if  $A \subset V$  and  $E_A = E \cap (A \times A)$ .
- The arcs of a graph can be **undirected** (denoted with segments:  $-$ ) or **directed** (denoted with arrows:  $\rightarrow$ ).
- Given a graph  $G = \{V, E\}$ ,  $\forall V_i, V_j \in V$ , the two vertices are called **adjacent** if there is an undirected arc between  $V_i$  and  $V_j$ .
- Given a graph  $G = \{V, E\}$ , a **path** is a sequence of adjacent vertices.
- Given a graph  $G = \{V, E\}$ , the set of all vertices which are adjacent to  $V_j$ ,  $\forall V_j \in V$ , is called **set of neighbors** of  $V_j$  and it is denoted with the symbol  $nb(V_j)$ .
- Given a graph  $G = \{V, E\}$  and a set  $A \subset V$ , the set of all vertices which are adjacent to  $A$ , but not in  $A$ , is called **set of neighbors** of  $A$  is  $nb(A) = \cup_{V_j \in A} nb(V_j) \setminus A$ .
- Given a graph  $G = \{V, E\}$  and a set  $A \subset V$ , the **neighborhood** of the set  $A$  is  $Nb(A) = A \cup nb(A)$ .

- A subset  $R$  of  $V$  is called **connected** if all pairs of vertices in  $C$  are linked by a path in  $R$ .
- A subset  $K$  of  $V$  is called **non connected** if there are at least two vertices non connected in  $K$ .
- A subset  $C$  of  $V$  is called **complete** if all pairs of vertices in  $C$  are adjacent.
- The class of all connected subsets of  $V$  is denoted with  $\mathcal{R}$ .
- The class of all non connected subsets of  $V$  is denoted with  $\mathcal{K}$ .
- The class of all complete subsets of  $V$  is denoted with  $\mathcal{C}$ .
- A graph  $G = \{V, E\}$  with only undirected arcs is called **Undirected Graph** (UG).
- Given a graph  $G = \{V, E\}$  and  $V_j, V_i \in V$ , if  $V_j \rightarrow V_i$  then  $V_j$  is called **parent** of  $V_i$  and  $V_i$  is called **child** of  $V_j$ .
- Given a graph  $G = \{V, E\}$ , an **ordered path** is a ordered sequence of vertices  $V_1, \dots, V_q$  where  $V_j$  is either adjacent or children of  $V_{(j-1)}$ .
- Given a graph  $G = \{V, E\}$ , a **directed cycle** is a sequence of vertices linked by arrows in an ordered path.
- Given a graph  $G = \{V, E\}$ , for all vertices  $V_j$  in  $V$ , the set of all children of  $V_j$  is denoted with  $ch(V_j)$  and the set of all parents is denoted with  $pa(V_j)$ .
- Given a graph  $G = \{V, E\}$ , the **set of the parents of a subset  $A$  of  $V$**  is denoted with  $pa(A)$  and is the set of the parents of all vertices in  $A$ :  
 $pa(A) = \left( \cup_{V_j \in A} pa_G(V_j) \right)$ .

- Given a graph  $G = \{V, E\}$ , the **set of children of  $A$** ,  $ch(A)$ , is the set of all vertices  $V_j \in T$  such that there is at least one vertex  $V_h \in A$  such that  $V_h \in pa_G(V_j)$ .
- A graph  $G = \{V, E\}$  with only directed arcs where there are non directed cycles is called **Directed Acyclic Graph** (DAG).
- Given a graph  $G = \{V, E\}$ , a **semi directed cycle** is a sequence of vertices linked by arrows and segments in an ordered path.
- A graph  $G = \{V, E\}$  with both directed and undirected arcs where there are non directed or semi directed cycles is called **Chain Graph** (CG).
- Given a CG, the components of the chain  $T_h$ ,  $h = 1, \dots, s$  induce undirected subgraphs of the CG. Between the components there are only directed arcs.
- The class of all component of a CG is denoted with  $\mathcal{T} = \{T_1, \dots, T_s\}$ .
- The DAG associated to a CG is the graph  $G_D = \{\mathcal{T}, E\}$  where the set of vertices is composed by the chain components, and where two components are linked by an arrow if there is at least a directed arc between the vertices in the components.
- Given a CG, the **set of the parents of the chain component  $T_h$** ,  $pa_D(T_h)$ , is the set of parents of the component  $T_h$  in the DAG associated to  $\mathcal{T}$ .
- Given a CG, the set of **non descendants**  $nd(T_h)$  of the chain component  $T_h$  is the set of all components  $T_j$  such that there is not any ordered path from  $T_j$  to  $T_h$ .
- Given a component  $T_h$ , we define the family of the cliques of the component  $\mathcal{Cl}_h$  the class of maximal complete sets such that  $\cup_{A \in \mathcal{Cl}_h} A = T_h$ .





## Appendix B

$\{A, B; B, C\}$ : Class with elements  $\{\{A; B\}, \{B; C\}\}$

$\lambda_{\mathcal{L}}$ : Vector of log-linear parameters referring to the variables in  $\mathcal{L}$

$\eta_{\mathcal{L}}^{\mathcal{M}}$ : Vector of marginal log-linear parameters referring to the variables in  $\mathcal{L}$  allocated in the marginal  $\mathcal{M}$

$\Pi$ : Set of probability distribution functions

**(C1)**, **(C2a)**, **(C2b)**, **(C3a)**, **(C3b)**, **(C3 \* b)**: Markov properties for chain graphs

**CG**: Chain Graphs

$CH_h$ : Set of children of the component  $T_h$

$ch(A)$ : Children of the set  $A$

$CI_k$ : Conditional independence model referring to  $k$  independencies

$Cl_h$ : Family of the cliques of the component  $T_h$

$D_i$ : Class of interaction sets referring to the null parameters according to the  $i$ -th independence

$d_I(A)$ : In-degree of the set  $A$

**DAG**: Directed Acyclic Graph

**DG**: Directed Graph

$\mathcal{F}$ : Class of hierarchical and complete interaction sets

**GM**: Graphical model

$\mathcal{H}$ : Class of hierarchical marginal sets

$\mathcal{H}_I$ : Hierarchical class of marginal sets for **GM I**

$\mathcal{H}'_{II}$ : Hierarchical class of marginal sets for the first parametrization for **GM II**

$\mathcal{H}^*_{II}$ : Hierarchical class of marginal sets for the second parametrization for **GM II**

$\mathcal{H}^{MIX}_{II}$ : Hierarchical class of marginal sets for the mixed parametrization for **GM II**

$\mathcal{H}_{III}$ : Hierarchical class of marginal sets for **GM III**

$\mathcal{H}_{IV}$ : Hierarchical class of marginal sets for **GM IV**

$I$ : Contingency table

$\mathcal{L}$ : Interaction set

$\mathcal{M}$ : Marginal set

**MG**: Mixed Graph

**MP**: Markov properties

$NA_h$ : Set of vertices that are non adjacent to the children in component  $T_h$

$Nb(A)$ : Set of neighborhood of  $A$

$nb(A)$ : Set of neighbors of  $A$

$NC_h$ : Set of vertices that are adjacent to  $CH_h$  in  $T_h$

$nd(T_h)$ : Set of non descendants of  $T_h$

$pa_D(T_h)$ : Set of parents of component  $T_h$

$pa_G(T_h)$ : Set of parents of set  $A$

$PA_h$ : Class of parents with the same children in  $T_h$

**UG**: Undirected graph

$\mathcal{P}(A)$ : Power set of  $A$

$\mathcal{T} = \{T_1, \dots, T_s\}$ : Class of components of a chain graph

## Bibliography

- [1] Agresti, A. (2002) “*Categorical Data Analysis*”. Wiley-Interscience.
- [2] Agresti, A. and Lang, J.B. (1993). “A proportional odds model with subject-specific effects for repeated ordered categorical responses”. *Biometrika*, **80** 527-534.
- [3] Andersen, A.H. (1974). “Multidimensional contingency tables”. *Scand. J. Statist.*, 115-127.
- [4] Andersson, S.A., Madigan, D. and Perlman, M.D. (2001). “Alternative Markov properties for chain graphs”. *Scand. J. Statist.*, **28**, 33-85.
- [5] Andersson, S. A., and Perlman, M.D. (2006). Characterizing Markov equivalence classes for AMP chain graph models. *The Annals of Statistics*, **34**(2), 939-972.
- [6] 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.
- [7] Bergsma, W.P., Rudas, T. (2002). “Marginal models for categorical data”. *Ann. Statist.* **30**,140-159.
- [8] Bergsma, W.P., Rudas, T. (2002a). “Variation independent parametrizations of multivariate categorical distributions”. *Distributions with given marginals and related topics*, 21-28.
- [9] Bergsma, W.P., Rudas, T., Németh, R., (2010). “Marginal conditional independence models with application to graphical modeling”. *Biometrika*, **97**, 4, 1006-1012.
- [10] Carey, V., Long, L., Gentleman, R., “RBGL: An interface to the BOOST graph library”. *R package version 1.32.0*

- [11] Cazzaro, M., and Colombi, R. (2008). "Modelling two way contingency tables with recursive logits and odds ratios". *Statistical Methods and Applications, Scand. J. Statist*(4), 435-453.
- [12] Colombi, R. (1995). "A class of log-linear models with constrained marginal distributions". *J. Ital. Statist. Soc.* 4, 147-65.
- [13] Colombi, R. (2008). "Hierarchical Multinomial Marginal Models". Atti della XLI riunione scientifica dell SIS, Università della Calabria, Cleup, Cosenza.
- [14] Colombi, R., Forcina, A. (2001), "Marginal regression models for the analysis of positive association of ordinal response variables", *Biometrika*, **88**, 1007-1019.
- [15] Colombi, R., Giordano, S., Cazzaro, M., R Development Core Team (2013). "hmmm: Hierarchical Multinomial Marginal Models". *R package version 1.1.0*.
- [16] Cox, D. R. (1972). "Regression models and life-tables. Journal of the Royal Statistical Society". *Series B (Methodological)*, 187-220.
- [17] Cox, D. R. and Wermuth, N. (1993) "Linear dependencies represented by chain graph". *Statist. Sci.*,8, 204-218,247-277.
- [18] Cox, D.R. and Wermuth, N. (1996). *Multivariate Dependencies*. London: Chapman and Hall.
- [19] Darroch, John N., Steffen L. Lauritzen, and T. P. Speed.,(1980) "Markov fields and log-linear interaction models for contingency tables." *The Annals of Statistics* : 522-539.
- [20] Dethlefsen, C., Højsgaard,S. (2005). "A Common Platform for Graphical Models in R: The gRbase Package". *Journal of Statistical Software*, **14**(17), 1-12.
- [21] Douglas, R., Fienberg, S., Lee, M. T., Sampson, A. and Whitaker, R. L. (1990). "Positive dependence concepts for ordinal contingency tables. In Topics in Statistical Dependence, Lecture Notes, Monograph series, 16, Ed. W. Block, A. R. Sampson and T. H. Savits, 189-202. Haywood, CA: *Institute of Mathematical Statistics*
- [22] Drton, M. (2009). "Discrete chain graph models". *Bernoulli* **15**(3), 736-7553.

- [23] Drton, M. (2009a). “likelihood ratio tests and singularities”. *The Annals of Statist.* **37**(2), 979-1012.
- [24] EVS (2010):European Values Study 2008, 4th wave, Italy. GESIS Data Archive, Cologne, Germany, ZA4755 Data File Version 1.0.0 (2010-11-30), doi: 10.4232/1.10031.
- [25] Fitzmaurice, Garrett M., N. M. Laird, and A. G. Rotnitzky. “Regression models for discrete longitudinal responses”. *Statistical Science*, 284-299.
- [26] Forcina, A. (2011). “Smoothness of conditional independence models for discrete data”. *Journal of Multivariate Analysis*.
- [27] Forcina, A., Lupparelli, M., and Marchetti, G. M. (2010). “Marginal parametrizations of discrete models defined by a set of conditional independencies”. *Journal of Multivariate Analysis*, **101**(10), 2519-2527
- [28] Frydenberg, M. (1990). “The chain Markov property”. *Scand. J. Statist.* **17** 333-353.
- [29] Geiger, D., and Meek, C. (1998). Graphical models and exponential families. In Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence 156-165. *Morgan Kaufmann Publishers Inc.*
- [30] Geiger, D., Heckerman, D., King, H., and Meek, C. (2001). Stratified exponential families: graphical models and model selection. *The Annals of statistics*, **29**(2), 505-529.
- [31] Glonek, G. F., and McCullagh, P. (1995). “Multivariate logistic models. Journal of the royal statistical society”. *Series B (Methodological)*, 533-546.
- [32] Goodman, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, **65**(329), 226-256.
- [33] Goodman, L. A. (1971). “The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications”. *Technometrics*, **13**(1), 33-61.
- [34] Goodman, L.A. (1979). “Simple models for the analysis of association in cross-classifications having ordered categories”. *J. Amer Statist. Assoc.* **74** 537-552.

- [35] Haberman, S. J. (1974). "Log-linear models for frequency tables with ordered classifications". *Biometrics*, 589-600.
- [36] Hammersley, J., and Clifford, P. (1971). "Markov fields on finite graphs and lattices". Unpublished manuscript.
- [37] Hunter, MS Handcock (2004), "Inference in curved exponential family models for networks", *Penn State Department of Statistics technical report* number 04-02 DR
- [38] Jensen, F. V., Lauritzen, S. L., and Olesen, K. G. (1990). "Bayesian updating in causal probabilistic networks by local computations". *Computational statistics quarterly*, **4**, 269-282.
- [39] Kauermann, G. (1997). "A note on multivariate logistic models for contingency tables". *Australian Journal of Statistics* **39**(3), 261-276.
- [40] Lang, J. B., and Agresti, A. (1994). "Simultaneously modeling joint and marginal distributions of multivariate categorical responses". *Journal of the American statistical association*, **89**(426), 625-632.
- [41] Lauritzen, S.L. (1996). *Graphical Models*. New York: Oxford Univ. Press.
- [42] Lauritzen, S.L. (2002). "Chain graph models and their casual interpretations". *J. R. Statist. Soc. B*, **64**(3),321-361.
- [43] Lauritzen, S. L., and Wermuth, N. (1989). "Graphical models for associations between variables, some of which are qualitative and some quantitative". *The Annals of Statistics*, 31-57.
- [44] Lauritzen, S. L., and Spiegelhalter, D. J. (1988). "Local computations with probabilities on graphical structures and their application to expert systems". *Journal of the Royal Statistical Society. Series B (Methodological)*, 157-224.
- [45] Liang, K.Y., Zeger, S.L. and Qaqish, B. (1992). "Multivariate regression analysis for categorical data (with discussion)". *J. Roy. Statist. Soc. Ser. B* **54** 3-40.
- [46] Marchetti, G.M., Lupparelli, M.,Bergsma, W.P,(2009). "Parametrizations and fitting of bi-directed graph models to categorical data". *Scand. J. Statist.* **36**, 559-576.
- [47] Marchetti, G.M., Lupparelli, M. (2011). "Chain graph models of multivariate regression type for categorical data". *Bernoulli* **17**(3),827-844.

- [48] McCullagh, P., Nelder, J.A. (1989). *Generalized linear models. Monographs on Statistics and Applied Probability 37*. Chapman Hall, London.
- [49] Molenberghs, G. and Lesaffre, E. (1994). "Marginal modelling of correlated ordinal data using a multivariate Plackett distribution.", *J. Amer Statist. Assoc.* **89** 633-644
- [50] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- [51] R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [52] Richardson, T. S. (2009, June). "A factorization criterion for acyclic directed mixed graphs". In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence 462-470. AUAI Press.
- [53] Richardson, T. S., and Spirtes, P. (2003). "Causal inference via ancestral graph models". *Oxford Statistical Science Series*, 83-105.
- [54] Shachter, R., and Peot, M. (1989). Evidential reasoning using likelihood weighting. Personal Communication.
- [55] Studeny, M. (2004). *Probabilistic conditional independence structures*. Springer.
- [56] Studeny, M., and Bouckaert, R. R. (1998). "On chain graph models for description of conditional independence structures". *Annals of Statistics*, 1434-1495.
- [57] Wermuth, N., and Cox, D. R. (2004). "Joint response graphs and separation induced by triangular systems". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(3), 687-717.
- [58] Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. New York: Jhon Wiley & Son.