Potential approach in marginalizing Gibbs models

Enrique Castillo a,*, Juan Ferrándiz b, Pilar Sanmartín c

a Departamento de Matemática Aplicada y Ciencias de las Computación, Universidad de Cantabria, 39005 Santander, Spain
b Department of Statistics and Operations Research, Universitat de València, Doctor Moliner 50, 46100 Burjassot, Spain
c Department of Mathematics, University Jaume I, Campus Riu sec, 12071 Castellón, Spain

Received 1 October 1998; accepted 1 June 1999

Abstract

Given an undirected graph $G$ or hypergraph potential $H$ model for a given set of variables $V$, we introduce two marginalization operators for obtaining the undirected graph $G_A$ or hypergraph $H_A$ associated with a given subset $A \subseteq V$ such that the marginal distribution of $A$ factorizes according to $G_A$ or $H_A$, respectively. Finally, we illustrate the method by its application to some practical examples. With them we show that potential approach allow defining a finer factorization or performing a more precise conditional independence analysis than undirected graph models. Finally, we explain connections with related works. © 1999 Elsevier Science Inc. All rights reserved.

Keywords: Factorization; Gibbs model; Hypergraph models; Marginalization

1. Introduction

In many practical situations the structural relationship among a set of variables $V = \{V_1, \ldots, V_n\}$ can be represented as an undirected graph $G = (V, E)$, where $E$ is the set of edges of $G$. If two variables are independent, the corresponding nodes should not be connected by a path.
Similarly, if the independence between variables $X$ and $Y$ is indirect and mediated by a third variable $Z$ (that is, if $X$ and $Y$ are conditionally independent given $Z$), we display $Z$ as a node that intersects the path between $X$ and $Y$, i.e., $Z$ is a cutset separating $X$ and $Y$.

Dawid [10] constitutes one of the earliest systematic studies of conditional independence, which is treated more formally in [11]. The correspondence between conditional independence and cutset separation in undirected graphs forms the basis of the theory of Markov fields [23,24], [5,36], and has been given axiomatic characterizations [7,31]. Given a graph $\mathcal{G}$ on $V$, a probability distribution is said to be $\mathcal{G}$-markovian if every separation statement in $\mathcal{G}$ corresponds to an independence statement of this distribution.

Unfortunately, not all probabilistic models can be represented by undirected perfect maps. Pearl and Paz [31] characterize the dependency models represented by undirected perfect maps (their theorem refers not only to probabilistic but to general dependency models).

However, in many practical cases we can be interested not in the whole set of variables $V$ but in a subset $A$ of them. In this case the initial graph model is not the most appropriate to work with and we are interested in the graph model induced in $A$.

The interest in stochastic independence statements induced in marginal distributions arose in the context of multi-way contingency tables and log-linear models [4,37], one of the main fields from which graphical models originated. The prime interest was in the preservation of stochastic independence between factors when we collapse the table with respect to some other factor, thus leading to valid inferences from the marginal table. Collapsibility properties (preservation under marginalization) of parametric functions have been studied for binary response in contingency tables [1,8,12,19,30,34,35], linear models [35] conditionally Gaussian graphical association models [17] or hierarchical interaction models [13], and logistic regression [21].

Recent developments in Spatial and spatio-temporal statistics have promoted the use of Gibbs distributions, which can be considered an extension of the previously mentioned models (see e.g. [3,16]). These distributions are characterized by their set of interaction functions which is called the potential. Marginal distributions arise in a natural way when we include hidden latent variables in order to make the model more realistic. Observable data correspond to marginal distributions and their mutual relationships originate from the hidden dependence structure of the complete model.

On the other hand, the independence structure induced in marginal distributions is at the very heart of general dependence models. Every statement of the type ‘the set of variables in $A$ is conditionally independent of the set of variables in $B$ given the set of variables in $C$ involving a proper subset $E = A \cup B \cup C$ of the original collection $V$ is linked to the corresponding marginal distribution. Therefore, any graphical representation of the
dependence structure of the whole set $V$ carries out an implicit collection of independence statements about all its possible marginal distributions.

Some results have been obtained so far in the graphical aspect of collapsibility, that is, characterizing the kind of graphs $\mathcal{G}$ whose subgraphs correspond to the marginals of their markovian distributions [17], or alternatively, characterizing the set of markovian distributions whose marginals are also markovian with respect to an appropriately defined marginal graph [33].

Nevertheless, very often in model-oriented practical situations, suitable models suggested by the application context will not admit a perfect graphical map of all of their independence statements, while the families of distributions with appropriate collapsibility properties could be too general for practical purposes.

Since the resulting marginal independence graphs reveal a lack of sensitivity to detect all independence properties and lack identification of missing $n$th ($n > 2$) order interactions when second order interactions are present, as an alternative, we propose the use of the potential approach based in Gibbs models and hypergraphs.

Given the extreme generality and flexibility of Gibbs distributions, this approach has the advantage of being close to practitioner's practical modeling concerns while allowing the understanding of general independence structures through factorization properties of the involved probability density functions.

In the present paper, based on these factorization properties, we give an algorithm for obtaining the marginal independence graph under very general conditions. To illustrate these concepts, we use some examples. There we show how the marginal graphs can fail in capturing all the independence statements of the probabilistic model. This inadequacy motivates the subsequent consideration of the potential-hypergraph approach. It leads to an exact expression of the dependence structure of the model, from which we can derive necessary and sufficient conditions of collapsibility.

The proposed approach is related to results in linear and log-linear models and their extensions, including the graphical symmetric interaction models, which can be worked out as particular cases of Gibbs models. Similarities and differences are discussed around the paper together with the connection with the graphical collapsibility results mentioned above.

In Section 2 we introduce the main concepts to be used in the rest of the paper with a distinction between those required for the case of graphs and those for hypergraphs. In particular, we introduce the hypergraph models based on Gibbs distributions, and discuss their better adequacy to describe dependence structures than simple graph models. In Section 3 we introduce a marginalization operator for the case of undirected graphs that allows obtaining such a graph in the sense of the marginal model to satisfy the corresponding factorization properties. We also give an algorithm to implement this operator. In Section 4 we follow exactly the same process for the case of
hypergraphs, and the conditions for graphical and parametric collapsibility become apparent. We explain the connection of this approach to related works.

In both sections we illustrate the methods by means of practical examples. Finally, we make some comparisons, and in Section 6 we give some conclusions and recommendations.

2. Background

We divide this section in three parts. The first is devoted to undirected graphs, the second to Gibbs distributions and hypergraphs, and the third to the connection between independence statements and separation criteria of hypergraphs.

2.1. Undirected graphs

The main theorem to be given in Section 3 requires several concepts of undirected graphs which are given below. We illustrate them with some examples.

Edges in undirected graphs are unordered pairs of vertices and, hereafter, we will denote them by the corresponding parenthesized pair, as in \((V_1, V_2)\), with no associated meaning with the particular order in the pair. We think it is easier to read than the usual equivalence symbol \(V_1 \sim V_2\).

**Definition 1 (Path).** Given a graph \(G\) a path of length \(n\) between nodes \(V_r\) and \(V_s\) is a sequence of nodes \(V_0, \ldots, V_n\) such that \((V_i, V_{i+1}); i = 0, \ldots, n - 1\) are edges of \(G\) and \(V_0 = V_r\) and \(V_n = V_s\).

**Definition 2 (Connected nodes).** Given a graph \(G = (V, E)\), two nodes \(V_r, V_s \in V\) are said to be connected if there is a path from \(V_r\) to \(V_s\). They are said to be directly connected iff the path is of length 1.

**Definition 3 (Complete set).** Given a graph \(G = (V, E)\), a set \(A \subseteq V\) is said to be complete if all nodes in \(A\) are mutually and directly connected by edges in \(E\).

**Definition 4 (Clique).** A maximal complete set of nodes is called a clique.

**Definition 5 (Boundary).** Given a graph \(G = (V, E)\) and a subset \(A \subseteq V\) the boundary \(bd(A)\) of \(A\) is the set of nodes \(V_r \notin A\) such that they are directly connected to an element of \(A\), i.e.,

\[
bd(A) = \left\{ V_r \notin A \mid (V_r, V_s) \in E,\ V_s \in A \right\}.
\]
Definition 6 (Connectivity components). Given a graph $G = (V, E)$ its set of nodes $V$ can be partitioned in maximal subsets of nodes which are mutually connected (see [26], p. 6). These sets are called connectivity components of $G$.

Example 1. Consider the set of variables $V = \{V_1, V_2, \ldots, V_{10}\}$ and the graph $G = (V, E)$ shown in Fig. 1, where

$$E = \{(V_1, V_3), (V_1, V_4), (V_1, V_5), (V_2, V_4), (V_3, V_4), (V_5, V_4), (V_6, V_4),$$
$$ (V_7, V_9), (V_7, V_{10}), (V_8, V_{10})\}.$$ 

Some illustrative examples of the above definitions are as follows.

Path. The sequence of nodes $\{V_1, V_4, V_5, V_3\}$ is a path of length 3 between $V_1$ and $V_3$, as it is the sequence $\{V_1, V_3\}$, which has length 1.

Connected nodes. The nodes $V_8$ and $V_9$ are connected nodes because there is a path $\{V_8, V_{10}, V_7, V_9\}$ joining $V_8$ and $V_9$.

Directly connected nodes. Nodes $V_7$ and $V_{10}$ are directly connected nodes because the path $\{V_7, V_{10}\}$ joining them has length 1.

Complete sets. The only complete set of four elements in $G$ is $\{V_1, V_3, V_4, V_5\}$ (all pairs of nodes are directly connected). Obviously, all its subsets are also complete and it contains the only four complete sets of three elements. The remaining complete sets contain one or two elements.

Clique. The sets $\{V_1, V_3, V_4, V_5\}$, $\{V_4, V_2\}$, $\{V_4, V_6\}$, $\{V_7, V_9\}$, $\{V_7, V_{10}\}$, $\{V_8, V_{10}\}$ are the cliques of $G$.

Boundary set. The boundary of the set $\{V_1, V_3, V_4, V_5\}$ is the set $\{V_2, V_6\}$.

Connectivity components. The connectivity components of the graph $G$ are

$$\tau_1 = \{V_1, V_2, V_3, V_4, V_5, V_6\} \quad \text{and} \quad \tau_2 = \{V_7, V_8, V_9, V_{10}\}.$$ 

Definition 7 (Completed edge set). Given a graph $G = (V, E)$ and a subset $A \subseteq V$, the completed edge set $E^*(A)$ of $A$ is the set of all possible edges between nodes in $A$. 

Fig. 1. Undirected graph.
**Definition 8 (Subgraph).** Given a graph $G = (V, E)$ and a subset $A \subset V$, the subgraph $G_A$ is the graph $G_A = (A, E_{\mid A})$, that is, the graph defined over $A$ and containing the edges of $E$ connecting nodes in $A$.

**Definition 9 (Factorization property).** A probability distribution $P$ on $V$, is said to factorize according to an undirected graph $G$ (UDG), if for all complete set, $C$, of vertices there exist non-negative functions $\psi_C$ such that

$$p(v) = \prod_{C \subset V \text{ complete}} \psi_C(c).$$

The above factorization can be done using only cliques. However, this leads to a coarser factorization.

**Example 2.** Consider again the graph in Example 1.

*Completed edge set.* The completed edge set of the set $\{V_7, V_8, V_9\}$ is

$$\{(V_7, V_8), (V_7, V_9), (V_8, V_9)\}.$$

*Subgraph.* The subgraph associated with the set $\{V_2, V_4, V_5, V_6\}$ is

$$\{\{V_2, V_4, V_5, V_6\}, \{(V_2, V_4), (V_4, V_5), (V_4, V_6)\}\}.$$

*Factorization.* A possible factorization of $p(v)$ is

$$p(v) = \psi(v_1, v_3, v_4, v_5)\psi(v_2, v_4)\psi(v_4, v_6)\psi(v_7, v_9)\psi(v_7, v_{10})\psi(v_8, v_{10}).$$

2.2. Gibbs distributions and hypergraphs

As it is well known undirected graphs do not lead to the finest possible factorization in probabilistic models. This justifies the use of the Gibbs distributions and hypergraph models to be given below.

**Definition 10 (Gibbs distribution).** Given a graph $G = (V, E)$, the set of random variables $V$ is said to follow a Gibbs distribution according to the graph $G$ if its associated probability density function (pdf) can be written in the form

$$p(v) = \exp \left( - \sum_{C \in \mathcal{C}} U_C(c) \right) / K,$$

where $K$ is a normalizing constant and $\mathcal{C}$ is the class of all complete sets of $V$ with respect to $G$. The functions $U_C$ are called interaction functions and some of them can be null. (In order to avoid trivial undeterminations we shall assume hereafter $U_{\emptyset}(\cdot) = 0$.)

The set $U = \{U_C(c) \mid C \in \mathcal{C}\}$ in (1) is called a potential.
Note that Expression (1) shows a characteristic factorization property of the corresponding Gibbs distribution. In fact the density in (1) factorizes as

\[ p(v) = \frac{1}{K} \prod_{c \in \mathcal{C}} \exp \left( - U_C(c) \right) = \frac{1}{K} \prod_{c \in \mathcal{C}} \psi_C(c), \]  

(2)

where the factors in \( \{\psi_C(c) \mid C \in \mathcal{C}\} \) are positive.

The above interpretation of the joint density in terms of the interaction functions is not unique. However, we are interested in the finest possible representation, which is given by the normalized potential.

**Definition 11 (Normalized potential).** A potential \( U \) such that \( U_c(c) = 0 \) whenever some component of \( c \) is null is called a normalized potential.

We assume that the range of every variable is a real set containing the zero, but we could take any other reference element 0, for each \( V_i \in V \) in Definition 11 (see [38]).

It can be shown that this potential is unique for a given probability distribution \( p(v) \) (see [38]). In addition, any given potential \( U^0 \) can be normalized in the sense of leading to the same joint distribution for \( V \), by means of the double sum (\( B \) and \( D \) varying)

\[ U_C(c) = \sum_{B \subseteq C \subseteq D \subseteq V} (-1)^{|C \setminus B|} U^0_D(b, 0_{D \setminus B}). \]  

(3)

This last equation makes evident that the normalized potential produces a finer factorization (2) of the pdf, because for every non-null interaction function \( U_C(c) \) of the normalized potential there is at least one non-null interaction function \( U^0_D(d) \) involving a bigger set of variables.

**Definition 12 (Potential restricted to a set).** Given a potential \( U \) on the set \( V \) and a subset \( A \subset V \) the potential \( U|_A \) restricted to \( A \) is the set

\[ U|_A = \{U_C \mid U_C \in U \text{ and } C \subset A\}. \]

**Example 3.** Consider the set of variables \( V = \{V_1, V_2, V_3, V_4, V_5, V_6\} \) and the graph \( \mathcal{G} = (V, E) \), where

\[ E = \{(V_1, V_2), (V_1, V_3), (V_2, V_3), (V_4, V_5), (V_5, V_6)\} \]

and each \( V_i \) ranges in a finite set containing 0.

**Gibbs Distribution.** Let us assume the following density:

\[ p(v) \propto \exp \left( - \theta_{12}(1 + v_1)v_2 - \theta_{13}v_1v_3 - \theta_{23}v_2v_3 - \theta \phi_0(v_4 - v_5) \\
- \theta \phi_0(v_5 - v_6) \right), \]  

(4)
with associated potential $U^0$ (used in analysis of gray level images, see [22], p. 93):
\[
\{ \theta_{12}(1 + v_1)v_2, \theta_{13}v_1v_3, \theta_{23}v_2v_3, \theta\phi_0(v_4 - v_5), \theta\phi_0(v_5 - v_6) \},
\]
where $\theta$ and $\theta_{ij}$ are constants and $\phi_0(d) = (1 + d^2)^{-1}$.

**Normalized potential.** The corresponding normalized potential $U$ becomes:
\[
\{ \theta_{12}v_2, \theta_{12}v_1v_2, \theta_{13}v_1v_3, \theta_{23}v_2v_3, \theta\phi_1(v_4), 2\theta\phi_1(v_5), \theta\phi_1(v_6), \\
\theta\phi(v_4 - v_5), \theta\phi(v_5 - v_6) \},
\]
with
\[
\phi(v_i - v_j) = \phi_0(v_i - v_j) - \phi_0(v_i) - \phi_0(-v_j) + \phi_0(0), \text{ and} \\
\phi_1(v_i) = \phi_0(v_i) - \phi_0(0),
\]

**Potential restricted to a set.** Given the set $A = \{V_1, V_3, V_5\}$, the potential restricted to $A$ is:
\[
|U|_d = \{ \theta_{13}v_1v_3, 2\theta\phi_1(v_5) \}.
\]

**Definition 13 (Hypergraph).** Given a set $V$, an hypergraph $\mathcal{H}$ is a subset of its parts $\mathcal{P}(V)$. Its elements are called hyperedges and the set $\text{supp}(\mathcal{H}) = \bigcup_{A \in \mathcal{H}} A$ its support.

**Definition 14 (Induced graph).** Given an hypergraph $\mathcal{H}$ on a set $V$, the induced graph $\mathcal{G}(\mathcal{H})$ is defined to be $(V, E)$ with edges all pairs of nodes included in some hyperedge of $\mathcal{H}$, i.e.:
\[
E = \{ (V_r, V_s) \mid \{V_r, V_s\} \subseteq A \in \mathcal{H} \}.
\]

**Definition 15 (Hypergraph associated with a potential).** Given a Gibbs distribution on $V$ with potential $U$, we define the hypergraph $\mathcal{H}(U)$ associated with the potential $U$ as the family of subsets of $V$ with non-null interaction function,
\[
\mathcal{H}(U) = \{ C \in \mathcal{C} \mid U_C \neq 0 \}.
\]

Expressions (1) and (2) show the pdf of a Gibbs distribution factorizing in accordance with the induced graph $\mathcal{G}(\mathcal{H}(U))$. Therefore (see [26], p. 35) the probability distribution verifies all the markovian properties with respect to $\mathcal{G}(\mathcal{H}(U))$ (in the pairwise, local and global sense).

Attending only to this factorization property, we could consider a reduced version of the preceding hypergraph by keeping only maximal hyperedges. This would be the hypergraph associated with a new potential made by adding up in a same term the interaction functions of subsets of $V$ included in every maximal hyperedge. This is the option taken, for instance, in [33]. But the resulting potential would then miss the normalized condition and we would lose the stated bijection between pdf and potential.
Definition 16 (Hypergraph precedence). Given two hypergraphs \( H_1 \) and \( H_2 \) on \( V \), we say that \( H_1 \) precedes \( H_2 \) iff every hyperedge of \( H_1 \) is contained in an hyperedge of \( H_2 \), that is,

\[
H_1 \preceq H_2 \iff \forall H_1 \in H_1 \exists H_2 \in H_2 \quad \text{with} \quad H_1 \subseteq H_2
\]

Reflexivity and transitivity of hypergraph precedence are easy to verify, thus showing the preorder nature of this binary relation. Two different hypergraphs can precede each other if they have the same maximal hyperedges (with regard to set inclusion). Precedence is a partial order in the quotient space of equivalence classes of hypergraphs induced by the binary relation of preceding each other.

As it has been pointed out, a Gibbs distribution can be represented by different potentials, each inducing its own associated hypergraph and density factorization. Given two potentials \( U_1 \) and \( U_2 \) leading to the same Gibbs distribution, it is clear from (2) that precedence \( H(U_1) \preceq H(U_2) \) implies \( U_1 \) producing a finer factorization than \( U_2 \). We can arrange \( U_1 \) factors showing a factorization of \( p(v) \) in terms of \( H(U_2) \). It suffices that, for each hyperedge \( A \in H(U_1) \), we choose one hyperedge \( C \in H(U_2) \) containing \( A \). Let us call \( A(C) \) the subset of \( H(U_1) \) associated in this way with \( C \in H(U_2) \). Then, we can write

\[
p(v) = \prod_{C \in H(U_2)} \left( \prod_{A \in A(C)} \psi_A(a) \right),
\]

where we have to understand any occasional empty product as being equal to 1.

Now we can state the property of normalized potentials producing finer factorizations in the more precise terms of partial ordering of the associated hypergraphs.

Proposition 1. The hypergraph associated with a normalized potential precedes the hypergraph associated with any other potential leading to the same probability distribution.

Proof. It is a consequence of (3) and the comment below it. \( \square \)

The following definition will prove to be appropriate when dealing with hypergraphs of marginal distributions.

Definition 17 (Boundary hypergraph). Let \( H \) be the hypergraph and \( A \subset V \). The boundary hypergraph \( H_A \) of \( V \setminus A \) is the hypergraph of all subsets of \( A \) which are the boundary of some connectivity component of \( G_{V \setminus A} \) in \( G(H) \).
Example 4. Consider again Example 3.

Hypergraph associated with a family of potentials. The hypergraph associated with the potential \( U \) is

\[ H^f = \{ \{ V_2 \}, \{ V_1, V_2 \}, \{ V_1, V_3 \}, \{ V_2, V_3 \}, \{ V_4 \}, \{ V_5 \}, \{ V_4, V_5 \}, \{ V_5, V_6 \} \}. \]

Graph associated with a hypergraph. The graph associated with hypergraph \( H \) is

\[ \{(V_1, V_2, V_3, V_4, V_5, V_6), (V_1, V_2, (V_1, V_3), (V_2, V_5), (V_4, V_5), (V_5, V_6))\}. \]

Boundary hypergraph. Given \( A = \{ V_1, V_3, V_5 \} \), since the connectivity components of \( V_n \mid A \) are

\[ \tau_1 = \{ V_2 \}, \quad \tau_2 = \{ V_4 \}, \quad \tau_3 = \{ V_6 \}, \]

the boundary hypergraph \( H_A \) of \( V \mid A \) is the hypergraph:

\[ \{ \{ V_1, V_3 \}, \{ V_5 \} \}. \]

Definition 18 (Hypergraph models). Given a parametric family of potentials, the hypergraph associated with its normalized potentials \( U^\theta \) is defined as the class of all sets of \( V \) with non-null interaction function \( U^\theta_C \) for at least one element in the family, i.e.:

\[ H = \{ C \subseteq V \mid U^\theta_C \neq 0 \text{ for some } \theta \}. \]

The corresponding model is called an interaction functions hypergraph or simply hypergraph model.

The hypergraph \( H \) associated with a family of distributions is the union of the hypergraphs associated with each distribution and, consequently, it is preceded by all of them. Thus, the factorization

\[ p_\theta(v) = \frac{1}{K_\theta} \prod_{C \in H} \exp\left(-U^\theta_C(c)\right) = \frac{1}{K_\theta} \prod_{C \in H} \psi^\theta_C(c), \]

is valid for all members of the family.

Note that hypergraphs are more capable to distinguish models than undirected graphs. For example, let us compare the models with potentials \( U^1 \), the one in Eq. (5), and \( U^2 = U^1 \cup \{ \theta_{123}v_1v_2v_3 \} \). We can say that the hypergraph associated with \( U^1 \) precedes the hypergraph associated with \( U^2 \), but not conversely, although both induce the same graph.

Probability distributions on finite sets \( V \) with positive pdf can always be expressed as Gibbs distributions according to the complete graph \( (V, E(V)) \), through the normalized potential (see [38], pp. 57–59):
This is the general framework. In order to devise interesting models for practical purposes, we usually reduce the hyperedges of $\mathcal{H}$ adequately, stating the appropriate independence statements and parameterizations.

**Example 5** (Graphical and hierarchical interaction models). Graphical interaction models constitute a joint generalization of log–linear models for contingency tables and multivariate Gaussian models for continuous variables (see [7,17,24–26,28,32]). They are particular cases of Hypergraph models on a finite set of vertices $V$. They consider $V$ partitioned in the set of discrete variables $D$, assumed to have finite ranges, and that of continuous variables $C$, with $V = D \cup C$. The marginal distribution of variables in $D$ has no restriction other than positivity (this is the log-linear part of the model), while the conditional distributions of variables in $C$, given those in $D$, follow multivariate Gaussian distributions. The corresponding interaction functions $U^0_{\mathcal{A}}(a)$ are completely general if $A \subseteq D$. When $A \cap C \neq \emptyset$ they are restricted to involve a maximum of two continuous variables, adopting the form

$$U^0_{\mathcal{A}}(a) = U^0_{\mathcal{A}}(a_{A \cap D}) \prod_{v_i \in A \cap D} v_i.$$ 

Therefore, $U^0_{\mathcal{A}}(a) \equiv 0$ if $|A \cap C| > 2$.

A subclass of graphical interaction models are the hierarchical mixed interaction models (see [13,14]). As in the case of hierarchical log–linear models, they stipulate a hypergraph of interaction functions such that all subsets of its maximal hyperedges are present. The set $\mathcal{A}$ of maximal hyperedges is called its generating class, and it determines the whole hypergraph of permissible interactions. If the family of cliques of the graph induced by $\mathcal{A}$ equals $\mathcal{A}$, then the model is said to be graphical.

Gibbs models arose in a more general context than contingency tables and covariance selection models just outlined in the preceding example. They have been widely used, for instance, in image analysis (since the seminal paper by Geman and Geman [18]), as well as in spatial statistics and ecological analysis (see e.g. [6,15,16]), genetics (see e.g. [20]), etc.

**Example 6** (Besag’s spatial auto-models). Gibbs distributions are determined by the set of their complete conditional distributions $\{p(v_i | v_{\neq i}) : V_i \in V\}$, as a consequence of expression (1) and the general statement $p(v_i | v_{\neq i}) \propto p(v)$. This is one of their main appealing properties because it allows modeling joint distributions piecewise, considering one variable at a time. Besag [2] proposed

$$U^0(c) = -\sum_{B \subseteq C} (-1)^{|C \setminus B|} \log(p_0(c, 0_{V \setminus C})).$$
the spatial auto-models as those Gibbs distributions with complete conditionals in the exponential family of probability distributions and interaction functions involving no more than two variables. They inherit their name from that of the corresponding conditional distribution. For instance, we have:

**auto-Binomial:** \( v_i | v_{V \setminus i} \sim \text{Binomial}(n_i, \pi_i) \)

\[
\logit(\pi_i) = \alpha_i + \sum_{V_j : \{V_i, V_j\} \in \mathcal{E}} \beta_{ij} v_j;
\]

\[
p(v) \propto \exp \left( \sum_{V_i \in V} \left( \log \left( \frac{n_i}{v_i} \right) - \alpha_i v_i \right) - \sum_{\{V_i, V_j\} \in \mathcal{E}} \beta_{ij} v_i v_j \right);
\]

**auto-Poisson:** \( v_i | v_{V \setminus i} \sim \text{Poisson}(\lambda_i) \)

\[
\log(\lambda_i) = \alpha_i + \sum_{V_j : \{V_i, V_j\} \in \mathcal{E}} \beta_{ij} v_j;
\]

\[
p(v) \propto \exp \left( \sum_{V_i \in V} \left( \log(v_i!) - \alpha_i v_i \right) - \sum_{\{V_i, V_j\} \in \mathcal{E}} \beta_{ij} v_i v_j \right);
\]

**auto-Gaussian:** \( v_i | v_{V \setminus i} \sim \text{Gaussian}(\mu_i, \tau_i) \) conditional mean \( \mu_i \), conditional precision \( \tau_i \)

**conditional mean:** \( \mu_i = v_i + \sum_{V_j : \{V_i, V_j\} \in \mathcal{E}} \beta_{ij} (v_j - v_i) \),

**joint precision:** \( \mathcal{C}_{ij} = \begin{cases} 
\tau_i & \text{if } V_i = V_j, \\
-\beta_{ij} \tau_i & \text{if } V_i \neq V_j \text{ and } \{V_i, V_j\} \in \mathcal{E}, \\
0 & \text{if } V_i \neq V_j \text{ and } \{V_i, V_j\} \notin \mathcal{E},
\end{cases} \)

\[
p(v) \propto \exp \left( \sum_{V_i \in V} \left( \sum_{V_j \in V} \left( \mathcal{C}_{ij} v_j - \frac{1}{2} \mathcal{C}_{ij} v_i^2 \right) v_i \right) - \sum_{\{V_i, V_j\} \in \mathcal{E}} \mathcal{C}_{ij} v_i v_j \right). \tag{9}
\]

In the present paper we approach the marginalization problem for Gibbs models in its general formulation, and we base our results in the corresponding parametric families of potentials described by their interaction functions hypergraph. We use normalized potentials in order to achieve uniqueness in model representation.
2.3. Hypergraph separation and independence statements

Independence statements involving disjoint subsets \( A, B \) and \( C \) of variables in \( V \) are related to factorization properties of the marginal distribution of \( A \cup B \cup C \). The basic connection between both concepts comes from the expression (see p. 29 of [26])

\[
A \perp B \mid C \iff p(a, b, c) = f(a, c)g(b, c),
\]

where the marginal pdf \( p \) is decomposed in at least two factors \( f \) and \( g \) which separate variables in \( A \) from variables in \( B \).

The marginal pdf of (10) is obtained from the joint pdf \( p(v) \) by integrating out the variables in \( U = V \setminus (A \cup B \cup C) \). If we pay attention only to the factorization properties of \( p \) and not to the current values of the involved functions, the only possible way to guarantee (10) is the decomposition of \( U \) in two disjoint subsets of variables, \( U_1 \) linked to \( A \) and \( U_2 \) linked to \( B \), through the factorization,

\[
p(v) = f(a, c, u_1)g(b, c, u_2).
\]

Integrating out \( U \) in (11) leads to

\[
p(a, b, c) = \int p(a, b, c, u) \, du = \int f(a, c, u_1) \, du_1 \int g(b, c, u_2) \, du_2.
\]

From the Definition 10 and Expressions (2) and (11) the following hypergraph separation criterion seems natural.

**Definition 19 (Hypergraph separation).** Given a hypergraph \( \mathcal{H} \in \mathcal{P}(V) \), let \( A, B, C \in \mathcal{P}(V) \) be disjoint subsets of \( V \). We say that \( A \) and \( B \) are \( \mathcal{H} \)-separated by \( C \) iff we can find a partition \( \mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \) such that

(i) \( A \cap \text{supp}(\mathcal{H}_1) = \emptyset \),
(ii) \( B \cap \text{supp}(\mathcal{H}_2) = \emptyset \), and
(iii) \( \text{supp}(\mathcal{H}_1) \cap \text{supp}(\mathcal{H}_2) \subseteq C \).

Studeny [33] introduced a related notion of connection in hypergraphs. He says that \( A, B \in \mathcal{H} \) are connected outside \( C \subset V \), \( C \neq \emptyset \), if there exists a sequence \( A = L_1, \ldots, L_n = B \), \( n \geq 1 \), in \( \mathcal{H} \) such that \( L_i \cap L_{i+1} \setminus C \neq \emptyset \) for \( i = 1, \ldots, n - 1 \). It can be verified that \( A \) and \( B \) are not connected outside \( C \) iff they are \( \mathcal{H} \)-separated by \( C \) in the sense of Definition 19.

The following proposition states hypergraph separation as a sufficient condition for conditional independence in the appropriate marginal distribution.
Proposition 2 (Independence statements and hypergraph separation). Given a potential $U$ associated with a Gibbs distribution, let $\mathcal{H}$ be the corresponding interaction function hypergraph. If $A$ and $B$ are $\mathcal{H}$-separated by $C$ then the set $A$ is independent of the set $B$ given $C$.

Proof. Variables outside $\text{supp}(\mathcal{H})$ have null interaction functions according to Definition 15. From factorization (2) each one is independent of the remaining variables in $V$, so that we can incorporate their marginal densities in the appropriate factors of (11).

Now, if we call $A_1 = A \cap \text{supp}(\mathcal{H})$, $A_2 = A \setminus \text{supp}(\mathcal{H})$, $B_1 = B \cap \text{supp}(\mathcal{H})$, $B_2 = B \setminus \text{supp}(\mathcal{H})$, $U_1 = \text{supp}(\mathcal{H}_2) \setminus (A \cup C)$, $U_2 = \text{supp}(\mathcal{H}_1) \setminus (B \cup C)$, by arranging factors in (2) we get

$$p(v) = f_1(a_1)f_2(a_2,c,u_1)g_1(b_1)g_2(b_2,c,u_2)$$

which has the desired form (11). □

We use the hypergraph separation criterion to prove that a normalized potential satisfies all the independence statements that can be deduced from any other potential leading to the same distribution. This is another reason justifying its use in the Definition 18 of hypergraph models. The interesting result comes as a corollary of the following theorem.

Theorem 1 (Hypergraph precedence and separation). Let $\mathcal{H}$ and $\mathcal{H}^0$ be two hypergraphs defined on $V$ such that $\mathcal{H} \preceq \mathcal{H}^0$. If $A$, $B$, and $C$ are disjoint subsets of $V$ verifying that $A$ and $B$ are $\mathcal{H}^0$-separated by $C$, then $A$ and $B$ are also $\mathcal{H}$-separated by $C$.

Proof. If $A$ and $B$ are $\mathcal{H}^0$-separated by $C$ then there is a partition of $\mathcal{H}^0$ in $\mathcal{H}_1^0$ and $\mathcal{H}_2^0$ such that, $A \cap \text{supp}(\mathcal{H}^0_1) = \emptyset$, $B \cap \text{supp}(\mathcal{H}^0_2) = \emptyset$, and $\text{supp}(\mathcal{H}^0_1) \cap \text{supp}(\mathcal{H}^0_2) \subseteq C$.

From Definition 16, precedence $\mathcal{H}(U) \preceq \mathcal{H}(U^0)$ implies that

$$\mathcal{H}_1 = \{D \in \mathcal{H}(U) \mid D \subseteq E, E \in \mathcal{H}_1^0\},$$
$$\mathcal{H}_2 = \{D \in \mathcal{H}(U) \mid D \subseteq E, E \in \mathcal{H}_2^0\},$$

constitute a partition of $\mathcal{H}$ verifying $\text{supp}(\mathcal{H}_1) \subseteq \text{supp}(\mathcal{H}^0_1)$, and $\text{supp}(\mathcal{H}_2) \subseteq \text{supp}(\mathcal{H}^0_2)$.

Therefore,

$A \cap \text{supp}(\mathcal{H}^0_1) = \emptyset$ implies $A \cap \text{supp}(\mathcal{H}_1) = \emptyset$,
$B \cap \text{supp}(\mathcal{H}^0_2) = \emptyset$ implies $B \cap \text{supp}(\mathcal{H}_2) = \emptyset$,
$\text{supp}(\mathcal{H}_1) \cap \text{supp}(\mathcal{H}_2) \subseteq \text{supp}(\mathcal{H}^0_1) \cap \text{supp}(\mathcal{H}^0_2) \subseteq C$. □
Corollary 1 (Normalized potential and hypergraph separation). Let \( U \) be the normalized potential and \( \mathcal{H}(U) \) its associated hypergraph. Let \( U^0 \) be another potential leading to the same distribution and \( \mathcal{H}(U^0) \), its associated hypergraph. Let \( A, B \) and \( C \) be disjoint subsets of \( V \). If \( A \) and \( B \) are \( \mathcal{H}(U^0) \)-separated by \( C \) then \( A \) and \( B \) are \( \mathcal{H}(U) \)-separated by \( C \).

Proof. It is a consequence of Proposition 1 and the previous theorem. □

The hypergraph separation criterion coincides with the independence graph separation criterion applied to \( \mathcal{G}(\mathcal{H}) \), the graph associated with \( \mathcal{H} \). It is easily verified from Definition 19 that \( A \) and \( B \) are \( \mathcal{H} \)-separated by \( C \) iff every path with origin in \( A \) and destination in \( B \) meets the set \( C \).

We can conclude that, in order to represent conditional independence statements of the probabilistic models, connection properties of interaction function hypergraphs do not improve the cutset separation criterion of the induced independence graphs.

Nevertheless, as it has been pointed out in the preceding paragraphs, conditional independence statements are related to the marginalization process because they involve proper subsets of the original set of variables \( V \).

In the following sections of this paper we develop a precise algorithm which allows to check all independent statements in the probabilistic model. Using normalized potentials and the appropriate hypergraphs, we obtain the independent graph of any marginal distribution.

3. The marginal graph

Theorem 2 (Marginal graph). Let \( \mathcal{G} \) be the undirected graph \( (V, E) \), and \( P \) the probability distribution over \( V \). If \( A \subseteq V \) and \( P_A \) is the marginal distribution associated with \( A \), we have that if \( P \) factorizes according to the graph \( \mathcal{G} \), then, the marginal distribution \( P_A \) factorizes according to the graph \( \mathcal{G}_A^{ma} = (A, E_A^{ma}) \), where

\[
E_A^{ma} = E|_A \bigcup_{\tau \in \mathcal{F}} E_a^{bd} (\tau),
\]

and \( \mathcal{F} \) is the set of connectivity components of \( \mathcal{G}_{V \setminus A} \).

Proof. The marginal distribution is obtained by integration over the range of \( Z = V \setminus A \), that is:

\[
p_A(a) = \int p(a, z) \, dz. \tag{13}
\]

Replacing the value of \( p \) in terms of its factors and assuming that \( C \) varies in the class of all complete sets \( \mathcal{C} \), we get:
Thus,
\[ p_A(a) = \psi_{c_0}(c_0) \prod_{C \subseteq A} \psi_C(c), \]
where
\[ \psi_{c_0}(c_0) = \int \prod_{C \subseteq A} \psi_C(c) \, dz, \quad (14) \]
\[ C_0 = \left( \bigcup_{C \subseteq A} C \right) \cap A. \quad (15) \]

Let \( \mathcal{F} \) be the set of connectivity components of the subgraph \( G_{V \setminus A} \). Obviously, there are no elements in \( C \) with indices in more than one of these different components. Thus, the integration over \( V \setminus A \) in (14) factorizes in integrals, each on a connectivity component, as:
\[ \psi_{c_0}(c_0) = \prod_{\tau \in \mathcal{F}} \int \prod_{C \cap \tau \neq \emptyset} \psi_C(c) \, d\tau, \quad (16) \]
where each factor is of the form:
\[ \psi_{bd(\tau)}(v_{bd(\tau)}) = \int \prod_{C \cap \tau \neq \emptyset} \psi_C(c) \, d\tau, \]
a function of the set of locations in \( A \) which are neighbors of some location in the connectivity component \( \tau \), that is, the set \( bd(\tau) \). Then, \( C_0 = \bigcup_{\tau \in \mathcal{F}} bd(\tau) \).

We shall write (14) as:
\[ \psi_{c_0}(c_0) = \prod_{\tau} \psi_{bd(\tau)}^{\tau}(v_{bd(\tau)}). \quad (17) \]
Consequently, the marginal pdf can be written as:
\[ p_A(a) = \prod_{C \subseteq A} \psi_C(c) \prod_{\tau} \psi_{bd(\tau)}^{\tau}(v_{bd(\tau)}), \quad (18) \]
where we can see that the distribution \( P_A \) satisfies the factorization property with respect to \( G_A^{ma} = (A, E_A^{ma}) \), as was to be proven.

The computation of the marginal graph reminds us, in a certain way, the moralization of chain graphs, the difference being that this applies to chain
graphs (with the existence of arrows) to obtain an undirected graph, by “marrying” the parents of each chain component. This new operation applies to undirected graphs and what get married are the elements in the boundaries of the connectivity components of the locations associated with variables disappearing during the marginalization process.

Related work can be found in [27]. They considered the marginalization procedure on graphical structures and their applications to expert systems. From the representation of probability distribution through evidence potentials they suggested a result similar to (14) and (15) as the factorization of the marginal potential. However, they did not pursued its further decomposition in factors depending on the connectivity components of the boundary $bd(V \setminus A)$, which is crucial in the computation of the exact marginal graph to be developed next in the following section.

From a purely graphical point of view Frydenberg [17] stated a sufficient condition for the subgraph $G_A$ to be the independence graph of the marginal distribution $P_A$. His argument is based on a previous result for log–linear models by Asmussen and Edwards [1]. He proves that $P_A$ is $G_A$-markovian if $P$ is $G$-markovian and all the boundaries of the connectivity components of $V \setminus A$ are complete in the graph $G$.

Theorem 2 could be proved using this result, because $P$ will factorize according to any graph obtained by adding edges to the graph $G$. We can make the boundaries $\{bd(\tau) \mid \tau \in T\}$ of (16) complete and then the new subgraph corresponding to $A$ will coincide with the marginal graph $G_{ma}$ of Theorem 2. We have adopted the direct argument through factorizations of the involved probability density functions for completeness within the potential approach and because some expressions in the preceding proof will be useful in the following sections.

Nevertheless, this graphical argument reveals how the marginal graph is based on a conservative criterion. Adding edges to $G$ will hide independence statements, and the resulting marginal graph may fail to detect them.

Studeny [33] defines the marginal graph $G^d$ by connecting directly two vertices $V_1, V_2 \in A$ if they are connected in $G$ by a path outside $A \setminus \{V_1, V_2\}$. This condition is the same as saying that both are directly connected in $G$ or both belong to the same boundary of some connectivity component in $G \setminus A$. Consequently, this definition agrees with the marginal graph obtained in Theorem 2.

The above theorem suggests the following algorithm for marginalization.

**Algorithm 1 (Marginalization).**

**Input:** A graph $G = (V, E)$ and a subset $A \subset V$.

**Output:** A graph $G_{ma} = (A, E_{ma})$ such that the $A$-marginal of the graphical model associated with the graph $G$ factorizes according to $G_{ma}$. 


Step 1. Obtain the set \(E_\mathcal{A}|_\mathcal{A}\) (edges in \(\mathcal{G}_\mathcal{A}\)).

Step 2. Obtain the subgraph \(\mathcal{G}_{V\setminus A}\).

Step 3. Obtain connectivity components \(\mathcal{T}\) of \(\mathcal{G}_{V\setminus A}\).

Step 4. Determine the set \(bd(\tau)\) in \(\mathcal{G}\) for each \(\tau \in \mathcal{T}\).

Step 5. Obtain the completed edge sets \(E^*_\mathcal{A}(bd(\tau))\) for each \(\tau \in \mathcal{T}\).

Step 6. Return the graph \(\mathcal{G}_{\mathcal{A}^*\mathcal{A}}\) where \(E_{\mathcal{A}^*\mathcal{A}}\) is the union of \(E_\mathcal{A}|_\mathcal{A}\) and \(\bigcup_{\tau \in \mathcal{T}} E^*_\mathcal{A}(bd(\tau))\).

Example 7. Assume the graph \(\mathcal{G} = (V, E)\), where

\[
V = \{V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}, V_{11}, V_{12}\}
\]
\[
E = \{(V_1, V_2), (V_1, V_4), (V_2, V_3), (V_3, V_4), (V_4, V_5), (V_5, V_6), (V_6, V_7),
(V_6, V_9), (V_5, V_{10}), (V_{10}, V_{11}), (V_{10}, V_{12})\}
\]

and the set \(A = \{V_2, V_4, V_5, V_7, V_8, V_{11}, V_{12}\}\) (see dotted regions in Fig. 2).

If we apply Algorithm 1, we obtain:

Step 1. \(E_\mathcal{A}|_\mathcal{A} = \{(V_4, V_5), (V_5, V_6)\}\).

Step 2. \(\mathcal{G}_{V\setminus A} = \{\{V_1, V_2, V_6, V_9, V_{10}\}, \emptyset\}\).

Step 3. \(\mathcal{T} = \{\tau_1, \tau_2, \tau_3, \tau_4\} = \{\{V_1\}, \{V_3\}, \{V_6\}, \{V_9, V_{10}\}\}\).

Step 4. \(bd(\tau_1) = \{V_2, V_4\}, bd(\tau_2) = \{V_2, V_4\}, bd(\tau_3) = \{V_5, V_7\}, bd(\tau_4) = \{V_8, V_{11}, V_{12}\}\).

Step 5. \(E^*(bd(\tau_1)) = E^*(bd(\tau_2)) = \{(V_2, V_4)\}, E^*(bd(\tau_3)) = \{(V_5, V_7)\}, E^*(bd(\tau_4)) = \{(V_8, V_{11}), (V_8, V_{12}), (V_{11}, V_{12})\}\).

Fig. 2. Graph associated with the hypergraph in Example 8 showing the connectivity components \(\tau_1, \tau_2, \tau_3\) of the subgraph associated with \(V \setminus A\) (sets corresponding to Step 3), and boundary hypergraph \(\mathcal{H}_\mathcal{A}\) (sets in dotted regions).
Step 6. We return the graph $\mathcal{G}_A^{ma} = (A, E_A^{ma})$, $E_A^{ma} = \{(V_2, V_4), (V_4, V_5), (V_5, V_7), (V_5, V_8), (V_8, V_{11}), (V_8, V_{12}), (V_{11}, V_{12})\}$. □

Assume that now we add the edges $(V_2, V_4)$ and $(V_{11}, V_{12})$ to $E$. If we apply the Algorithm 1, we obtain the same results except for the Step 1 in which $E_A = \{(V_2, V_4), (V_4, V_5), (V_5, V_8), (V_8, V_{11}), (V_8, V_{12})\}$.

As we pointed out before, the graph approach is not sufficient in order to characterize all independence statements contained in a probability distribution. This motivates the marginal hypergraph approach developed in the next section.

4. The marginal hypergraph

In this section we analyze the marginalization problem in hypergraphs models. Given the subset $A \subset V$, we find the hypergraph associated with the potential $U_A^{H}$ corresponding to the family of marginal distributions $P_A^{h}$ of an hypergraph model. First we study the changes suffered by the original potential $U$ and how they induce the marginal hypergraph.

4.1. Marginal potential and hypergraph

From the proof of Theorem 2 the role of the connectivity components $\tau$ of the subgraph $\mathcal{G}_{V \setminus A}$ when integrating out the variables in $V \setminus A$ have been seen. Their contributions to the marginal potential could be called the innovations of $U_A$. The following lemma shows how this innovations are. It involves $H_A = \{B \in H \mid B \subseteq A\}$, the hypergraph associated with the potential $U$ restricted to $A$ (see Definition 12), and $H_A$, the boundary hypergraph introduced in Definition 17.

Lemma 1 (Marginal potential). The marginal distribution $P_A$ can be expressed by means of the non-normalized potential:

$$\hat{U}^A_D(d) = U_D(d) + U^*_D(d),$$

where $U^*_D$ is null unless $D$ is an hyperedge of $H_A$. In this last case it can be computed as

$$U^*_D(d) = \sum_{\tau : bd(\tau) = D} U_{bd(\tau)}^*(d),$$

$$U_{bd(\tau)}^*(d) = -\ln \left( \int \exp \left( - \sum_{C \cap \tau \neq \emptyset} U_C(c) \right) \, dv_\tau \right),$$

where $\tau$ ranging in the set $\mathcal{T}$ of connectivity components of $V \setminus A$. 

Proof. From expression (18) in the proof of Theorem 2, calling

$$U_C(c) = -\ln(\psi_C(c)), \quad C \in \mathcal{H}|_A,$$

$$U_{bd(t)}^\tau(bd(\tau)) = -\ln(\psi_{bd(t)}^{\tau}(v_{bd(\tau)})), \quad \tau \in \mathcal{T},$$

$$U_D^e(d) = \sum_{\tau: bd(\tau) = D} U_{bd(t)}^\tau(d), \quad D \in \mathcal{H}^A,$$

we can write the marginal pdf as

$$p_A(a) = K^{-1} \exp \left( -\sum_{C \subseteq A} U_C(c) - \sum_{D \in \mathcal{H}^A} U_D^e(d) \right).$$

If we extend the definition of $U_D^e$ to encompass all subsets of $A$ by making $U_D^e \equiv 0$ whenever $D \notin \mathcal{H}^A$, the potential defined in (19) verifies,

$$p_A(a) = \exp \left( -\sum_{B \subseteq A} U_B^e(b) \right) / K_A,$$

$$K_A = K \exp(U_0^e(0)),$$

as was to be proven. \(\square\)

Theorem 3 (Marginal normalized potential). The normalized potential of the marginal distribution $P_A$ can be expressed as

$$U_B^A(b) = U_B(b) + V_B^A(b), \quad (22)$$

where $B$ ranges in the union of the restricted hypergraph $\mathcal{H}|_A$ and the family of non-empty subsets of hyperedges in $\mathcal{H}^A$. Moreover, the innovation $V_B^A(b)$ can be computed as the double sum

$$V_B^A(b) = \sum_{B \subseteq D \in \mathcal{H}^A} \sum_{E \subseteq B} (-1)^{|E|} U_D^e(e, 0_{D\setminus E}), \quad (23)$$

when $B$ is a non-empty subset of some hyperedge $D$ in $\mathcal{H}^A$, and $0$ otherwise. In (23) we have to understand $U_D^e$ as in (20).

Proof. We can obtain the marginal normalized potential $U^A$ by applying expression (3) to the potential $\hat{U}^A$ of Lemma 1,

$$U_B^A(b) = \sum_{E \subseteq B \subseteq D \subseteq A} (-1)^{|E|} \hat{U}_D^A(e, 0_{D\setminus E})$$

$$= \sum_{E \subseteq B \subseteq D \subseteq A} (-1)^{|E|} U_D(e, 0_{D\setminus E}) + \sum_{E \subseteq B \subseteq D \subseteq A} (-1)^{|E|} U_D^e(e, 0_{D\setminus E}). \quad (24)$$

Being $U$ a normalized potential, $U_D(e, 0_{D\setminus E}) = 0$ if $D \setminus E \neq \emptyset$. Moreover $U_D^e(e, 0_{D\setminus E}) = 0$ if $D \notin \mathcal{H}^A$. Expression (24) becomes:
\[ U_B(b) + \sum_{B \subseteq D \in \mathcal{X}^A} \sum_{E \subseteq B} (-1)^{|B \setminus E|} U_D^*(e, 0_{D \setminus E}). \tag{25} \]

If \( B = \emptyset, E \subseteq B \) is also empty so that
\[ U_\emptyset^A(a) = \sum_{D \in \mathcal{X}^A} (-1)^0 U_D^*(0_D) = \text{Constant}. \tag{26} \]

We can include (26) in the normalizing constant \( K_A \) and define \( U_\emptyset^A \equiv 0 \).

Finally, by introducing the function
\[ V_B^A(b) = \begin{cases} \sum_{B \subseteq D \in \mathcal{X}^A} \sum_{E \subseteq B} (-1)^{|B \setminus E|} U_D^*(e, 0_{D \setminus E}) & B \subseteq D \in \mathcal{X}^A, \\ 0 & \text{otherwise,} \end{cases} \]
we can write the marginal normalized potential as in (22).

We consider next the hypergraph associated with the marginal normalized potential of a hypergraph model. We add the \( \theta \) superscript to all the functions derived from \( U^0 \) whenever their parametric condition is to be emphasized.

**Theorem 4** (Marginal hypergraph). Consider a hypergraph model on \( V \), with interaction functions hypergraph \( \mathcal{H} \), and let \( P^0_A \) be the corresponding family of marginal distributions over \( A \). Then, the interaction function hypergraph \( \mathcal{H}_A \) of the family \( P^0_A \) can be expressed as:
\[ \mathcal{H}_A = \mathcal{H}^\theta_A \cup \mathcal{H}^-_A \setminus \mathcal{H}^+_A, \tag{27} \]
where
1. \( \mathcal{H}^\theta_A \) is the restriction of \( \mathcal{H} \) to \( A \), that is, the set of elements in \( \mathcal{H} \) which are subsets of \( A \). (These are the complete sets that will remain after marginalization).
2. \( \mathcal{H}^-_A \) is the family of subsets \( B \subseteq A \) not in \( \mathcal{H}^\theta_A \) and such that \( V_B^0(b) \) is a non-null function for some \( \theta \). (These are the new complete sets that will appear after marginalization).
3. \( \mathcal{H}^+_A \) is the set of complete sets in \( \mathcal{H}^\theta_A \) such that they are a subset of some set in \( \mathcal{H}^A \) and satisfy the equation:
\[ U_B^0(b) = -V_B^\theta_A(b); \quad \forall \theta. \tag{28} \]
(These are the complete sets that will disappear after marginalization).
And, for \( B \in \mathcal{H}_A \), the \( A \)-marginal potential is \( U_B^{0A}(b) = U_B^0(b) + V_B^{0A}(b) \).

**Proof.** Applying Theorem 3 to each distribution of the hypergraph model yields,
\[ U_B^{0A}(b) = U_B^0(b) + V_B^{0A}(b). \tag{29} \]
Now, consider the following cases:

1. $U^{\theta}_{B}(\cdot) = -V^{\theta}_{B}(\cdot) \forall \theta$. In this case, $U^{\theta}_{B} \equiv 0$ and $B \not\in \mathcal{H}_{A}$. This functional equality can arise in the following situations:
   1.1. $B \in \mathcal{H}$ and $V^{\theta}_{B}(\cdot) \not\equiv 0$ for at least one value of $\theta$. In this case, $B \in \mathcal{H}_{A}^{-}$. 
   1.2. $B \not\in \mathcal{H}$ and $V^{\theta}_{B}(\cdot) \equiv 0 \forall \theta$.

2. $U^{\theta}_{B}(\cdot) \not= -V^{\theta}_{B}(\cdot)$ for at least one value of $\theta$. In this case $U^{\theta}_{B} \not\equiv 0$ for at least one element in the family. Thus $B \in \mathcal{H}_{A}$. This condition could happen in two circumstances:
   2.1. $B \in \mathcal{H}$ whether $V^{\theta}_{B}(\cdot) \equiv 0$ or not. Then $B \in \mathcal{H}_{A}^{-}$. 
   2.2. $B \not\in \mathcal{H}$ and $V^{\theta}_{B}(\cdot) \not\equiv 0$ for some $\theta$. Thus $B \in \mathcal{H}_{A}^{+}$.

Being the previous enumeration an exhaustive and exclusive list of possibilities for $B \subseteq A$, expression (27) is substantiated.  

This theorem suggests the algorithm below for marginalizing a hypergraph. In it we clarify the meaning of the sets and functions appearing in expressions (21) and (23), which are not easy to understand. With the same purpose we also include a simple example.

**Algorithm 2 (Marginalization of hypergraphs).**

- **Input.** A set $V$, a parametric family of normalized potentials $U^{\theta}$ over $V$, and a subset $A \subseteq V$.
- **Output.** The $A$-marginal potential $U^{\theta}_{A}$, together with its associated hypergraph $\mathcal{H}_{A}$ and graph $\mathcal{G}(\mathcal{H}_{A})$.
- **Step 1.** Obtain the hypergraph $\mathcal{H}$ associated with the given potential $U$.
- **Step 2.** Obtain the graph $\mathcal{G}(\mathcal{H})$ associated with the hypergraph.
- **Step 3.** Determine the connectivity components of the subgraph associated with $V \setminus A$.
- **Step 4.** Obtain the boundary hypergraph $\mathcal{H}_{A}^{\text{b}}$, as the collection of the boundaries in $\mathcal{G}(\mathcal{H})$ of the connectivity components of $V \setminus A$.
- **Step 5.** For each element $B \in \mathcal{H}_{A}^{\text{b}}$ and each $\tau$ verifying $bd(\tau) = B$ in (21) calculate the functions $U_{B}^{\tau}(b)$.
- **Step 6.** For each element $B \in \mathcal{H}_{A}^{\text{b}}$ calculate the functions $U_{B}(b)$ (see (20)).
- **Step 7.** Using (23), calculate $V_{B}(b)$ for each non-empty subset $B$ of the sets in $\mathcal{H}_{A}^{\text{b}}$.
- **Step 8.** Calculate the $A$-marginal potential $U^{A}$ by adding $V^{A}$ to the initial potential $U$ restricted to $A$.
- **Step 9.** Obtain the hypergraph $\mathcal{H}_{A}$ associated with $U^{A}$.
- **Step 10.** Obtain the graph $\mathcal{G}(\mathcal{H}_{A})$ associated with $U^{A}$.
- **Step 11.** Return $U^{A}$, $\mathcal{H}_{A}$ and $\mathcal{G}(\mathcal{H}_{A})$.

**Example 8.** Assume the set

$$V = \{V_{1}, V_{2}, V_{3}, V_{4}, V_{5}, V_{6}, V_{7}, V_{8}, V_{9}, V_{10}, V_{11}, V_{12}\},$$
of binary \((0, 1)\) variables, and the normalized potential

\[ U = \{ x_{12}v_1v_2, x_{14}v_1v_4, x_{23}v_2v_3, x_{34}v_3v_4, x_{45}v_4v_5, x_{56}v_5v_6, x_{58}v_5v_8, x_{67}v_6v_7, \\
   x_{89}v_8v_9, x_{9,10}v_9v_{10}, x_{10,11}v_9v_{11}, x_{10,12}v_9v_{12}, x_{11,12}v_9v_{12} \} \]

with \(x_{ij} \neq 0\), and \(A = \{ V_2, V_4, V_5, V_7, V_8, V_{11}, V_{12} \} \).

**Step 1.** The hypergraph \( \mathcal{H} \) associated with the given potential \( U \) is

\[ \mathcal{H} = \{ \{ V_1, V_2 \}, \{ V_1, V_4 \}, \{ V_2, V_3 \}, \{ V_3, V_4 \}, \{ V_4, V_5 \}, \{ V_5, V_8 \}, \{ V_6, V_7 \}, \{ V_8, V_9 \}, \{ V_9, V_{10} \}, \{ V_{10}, V_{11} \}, \{ V_{10}, V_{12} \}, \{ V_{11}, V_{12} \} \} \]

**Step 2.** The graph associated with the hypergraph is the one used in Example 7 and is shown in Fig. 2.

**Step 3.** The connectivity components of the subgraph associated with \( V \setminus A \) are those given in Step 3 of Example 7 and are shown in Fig. 2.

**Step 4.** The boundary hypergraph coincides with that in Step 4 of Example 7 and it is shown in Fig. 2 (dotted regions). Note that \( bd(\tau_1) = bd(\tau_2) = \{ V_2, V_4 \} \).

**Step 5.** In order to avoid a lengthy exposition, we compute only \( U_B^T \) for \( B = \{ V_8, V_{11}, V_{12} \} \in \mathcal{H}^A \), as an example. From (21) we get:

\[ U_{\{ V_8, V_{11}, V_{12} \}}^T = -\ln \left( \sum_{i=1}^{10} \sum_{j=1}^{10} \exp \left\{ \sum_{k=0}^{10} x_{9,10}v_9v_{10} - x_{89}v_8v_{9} - x_{10,11}v_9v_{11} - x_{10,12}v_9v_{12} \right\} \right) \]

**Step 6.** Similarly, from (20) we get:

\[ U_{\{ V_2, V_4 \}}^T = U_{\{ V_2, V_4 \}}^T + U_{\{ V_2, V_4 \}}^T \]

\[ U_{\{ V_8, V_{11}, V_{12} \}}^T = U_{\{ V_8, V_{11}, V_{12} \}}^T \]

**Step 7.** We have to consider all non-void subsets of the hyperedges in \( \mathcal{H}^A = \{ \{ V_2, V_4 \}, \{ V_5, V_7 \}, \{ V_8, V_{11}, V_{12} \} \} \). As an example, taking \( B = \{ V_{11}, V_{12} \} \) we obtain

\[ V_{\{ V_{11}, V_{12} \}}^A(v_{11}, v_{12}) = -\ln \left( 3 + \exp \left\{ -x_{9,10} \right\} \right) \]

\[ + \ln \left( 2 + \exp \left\{ -x_{10,11}v_{11} \right\} \exp \left\{ -x_{9,10} - x_{10,11}v_{11} \right\} \right) \]

\[ + \ln \left( 2 + \exp \left\{ -x_{10,12}v_{12} \right\} \exp \left\{ -x_{9,10} - x_{10,12}v_{12} \right\} \right) \]

\[ - \ln \left( 2 + \exp \left\{ -x_{10,11}v_{11} - x_{10,12}v_{12} \right\} \right) \]

\[ + \exp \left\{ -x_{9,10} - x_{10,11}v_{11} - x_{10,12}v_{12} \right\} \].

**Step 8.** If we are interested in the non-null interaction functions we must check whether the candidate functions are non-null. Equating \( V_B^A \) to zero for all \( B \subseteq D \in \mathcal{H}^A \), we get only the trivial solution \( x_{ij} = 0 \), which
contradicts the assumptions made at the beginning of this example for the potential \( U \). Then \( U^A \) becomes in this special case:

\[
U^A = V^A \cup U|_A. 
\]

**Step 9.** The marginal hypergraph becomes

\[
\mathcal{H}_A = \{(V_2), (V_4), (V_5), (V_7), \{V_8\}, \{V_{11}\}, \{V_{12}\}, \{V_2, V_4\}, \{V_4, V_5\}, \\
\{V_5, V_7\}, \{V_5, V_8\}, \{V_8, V_{11}\}, \{V_8, V_{12}\}, \{V_{11}, V_{12}\}, \{V_8, V_{11}, V_{12}\}\}.
\]

**Step 10.** Finally, the associated marginal graph becomes:

\[
\mathcal{G}(\mathcal{H}_A) = (A, \{(V_2, V_4), (V_4, V_5), (V_5, V_7), \\
(V_5, V_8), (V_8, V_{11}), (V_8, V_{12}), (V_{11}, V_{12})\}.
\]

**Step 11.** Return \( U^A, \mathcal{H}_A \) and \( \mathcal{G}(\mathcal{H}_A) \).

Note that we have obtained the same marginal graph as the one we derived by means of the undirected graph algorithm (see Step 6 in Example 7).

However, if the potential \( U \) includes the interaction functions \( z_{11,12}v_{11}v_{12} \) and \( z_0v_8v_{11}v_{12} \), we can get different results. Potentials verifying:

\[
V^A_{\{v_{11}, v_{12}\}}(v_{11}, v_{12}) + z_{11,12}v_{11}v_{12} \equiv 0, 
\]

\[
V^A_{\{v_8, v_{11}, v_{12}\}} + z_0v_8v_{11}v_{12} \equiv 0, 
\]

or equivalently,

\[
z_{11,12} = -V^A_{\{v_{11}, v_{12}\}}(1, 1), 
\]

\[
z_0 = -V^A_{\{v_8, v_{11}, v_{12}\}}(1, 1, 1), 
\]

lead to the same marginal potential of \( (30) \) but with \( V^A_{\{v_{11}, v_{12}\}} \) and \( V^A_{\{v_8, v_{11}, v_{12}\}} \) removed.

In this case the marginal hypergraph will be \( \mathcal{H}_A \setminus \{(V_{11}, V_{12}), \{V_8, V_{11}, V_{12}\}\} \), and the induced graph, compared with \( (32) \), will lose the edge \( (V_{11}, V_{12}) \).

On the other hand, potentials verifying \( (33) \) but not \( (34) \) (or vice versa), add the \( z \)-term of \( (34) \) (the \( z \)-term of \( (33) \)) to the marginal potential considered above. The marginal hypergraph will be as \( \mathcal{H}_A \) without hyperedges \( \{V_{11}, V_{12}\} \) (or \( \{V_8, V_{11}, V_{12}\}\)). Finally, the induced graph is again \( \mathcal{G}(\mathcal{H}_A) \) in both cases.

If neither of both conditions are true, the marginal potential has the \( z \)-terms given in \( (33) \) and \( (34) \) as interaction functions for \( \{V_{11}, V_{12}\} \) and \( \{V_8, V_{11}, V_{12}\} \) respectively. The marginal hypergraph will be \( \mathcal{H}_A \) and the marginal graph \( \mathcal{G}(\mathcal{H}_A) \).

Let \( \mathcal{G}_{\text{ma}}^A \) be the marginal graph (Algorithm 1) and \( \mathcal{G}(\mathcal{H}_A) \) the graph associated with the marginal hypergraph (Algorithm 2). Then, from the preceding discussion we can conclude:

1. In case the edge \( (V_{11}, V_{12}) \) is not in \( \mathcal{G} \equiv \mathcal{G}(\mathcal{H}) \), then \( (V_{11}, V_{12}) \) is an edge of both \( \mathcal{G}_{\text{ma}}^A \) and \( \mathcal{G}(\mathcal{H}_A) \).
2. In case the edge \((V_{11}, V_{12})\) is in \(\mathcal{G} \equiv \mathcal{G}(\mathcal{H})\), then \((V_{11}, V_{12})\) is an edge of \(\mathcal{G}^\text{ma}_A\). 

The absence of \((V_{11}, V_{12})\) as an edge of a graph over \(A\) implies the stochastic independence of both variables conditioned to \(A\). This independence statement is included in the model which has \((V_{11}, V_{12})\) as an edge of \(\mathcal{G}(\mathcal{H}_A)\) and verifies (33) and (34). This shows that, in this case, the undirected graph representation of the model is not able to capture this separating statement while the hypergraph model is. Note that if condition (33) applies while (34) does not, there is no conditional independence for \(\lbrace V_{11}, V_{12}\rbrace\). The same can be said when (33) is true and (34) false.

A similar discussion could be applied to the edge \((V_{2}, V_{4})\) appearing and disappearing in the marginal graph. In this case, to produce these changes, we only need to introduce second order interactions in the original potential.

Another point to be highlighted in the previous example is the use of the normalized potential. Consider again the case in which condition (34) applies and let us compare the normalized potential with the non-normalized one of Lemma 1. The interaction functions involving variables \(V_{11}\) and \(V_{12}\) are,

\[
\dot{U}^A_{\lbrace V_{11}, V_{12}\}}(v_{11}, v_{12}) = U^*_{V_{11}, V_{12}} + \alpha_{11,12}v_{11}v_{12} = U^A_{\lbrace V_{11}, V_{12}\}}(v_{11}) + U^A_{\lbrace V_{11}\}}(v_{11}) + U^A_{\lbrace V_{12}\}}(v_{12}).
\]

If we nullify \(\dot{U}^A_{\lbrace V_{11}, V_{12}\}}(v_{11}, v_{12})\), then variables \(V_{11}\) and \(V_{12}\) will not appear in the potential and will not contribute to the pdf (each one will be independent of all remaining variables and uniformly distributed). But we can cancel \(U^A_{\lbrace V_{11}, V_{12}\}}(v_{11})\) in the normalized potential, making \(V_{11}\) independent of \(V_{12}\) given the remaining variables, and both will still contribute to the potential with interaction functions \(U^A_{\lbrace V_{11}\}}(v_{11})\) and \(U^A_{\lbrace V_{12}\}}(v_{12})\). This illustrates how the normalized potential allows a more gradual way to incorporate independence statements.

**Example 9** (Besag’s spatial auto-models, continued). As we have seen in Example 6, the general expression for spatial auto-models can be written

\[
p(v) \propto \exp \left( \sum_{v_{i} \in \Omega} (\alpha_i v_i + f_i(v_i)) - \sum_{\{v_i, v_j\} \in \mathcal{E}} \beta_{ij} v_i v_j \right),
\]

where only a maximum of two variables are involved in each interaction function. Applying Theorem 3 in order to obtain the marginal normalized potential relative to subset \(A\), we get

\[
U^A_B(b) = U^A_B(b) + V^A_B(b) \quad \text{if} \ |B| \leq 2
\]

\[
U^A_B(b) = V^A_B(b) \quad \text{if} \ |B| > 2,
\]

i.e., interaction functions with more than two variables could appear only as innovations. Let us assume for the moment that \(U^*_{B}\) of (20) has the following form:
\[ U^*_D(d) = \sum_{C \subseteq D} F^D_C(c) + L_D \] (38)

for some constant \( L_D \) and functions \( F^D_C(c) \) verifying \( F^D_C(c) = 0 \) whenever some component of \( c \) is 0.

Computing the innovations (37) for \( B \subseteq D \) according to Theorem 3, we get

\[ V^A_B(b) = \sum_{B \subseteq D \in \mathcal{A}} \sum_{E \subseteq B} (-1)^{|B \setminus E|} U^*_D(e, 0_{D \setminus E}), \] (39)

\[ = \sum_{B \subseteq D \in \mathcal{A}} \sum_{E \subseteq B} (-1)^{|B \setminus E|} \sum_{C \subseteq D} (F^D_C(c) + L_D), \] (40)

\[ = \sum_{B \subseteq D \in \mathcal{A}} \sum_{C \subseteq B} (F^D_C(c) + L_D) \sum_{H \subseteq B \setminus C} (-1)^{|H|}, \] (41)

\[ = \sum_{B \subseteq D \in \mathcal{A}} (F^D_B(b) + L_D), \] (42)

where equality (40) results from \( F_C(c) = 0 \) if \( C \not\subseteq E \) because of the hypothesized properties of functions \( F_C \), equality (41) is obtained by rearranging terms putting \( E = B \setminus H \), and equality (42) is a consequence of

\[ \sum_{H \subseteq B} (-1)^{|H|} = (1 - 1)^{|B|} \neq 0 \quad \text{iff } B = \emptyset. \]

Eq. (42) looks much simpler than (39). Auto-Gaussian models are a clear example of this advantage. According to (9), the terms \( U^*_{bd(c)}(d) \) adding up to \( U^*_D \) are of the form

\[ U^*_{bd(c)}(d) = -\ln \left( \int_{\mathbb{R}^{|d|}} \exp \left\{ \sum_{v_r \in \tau} \left( v_r d v_r - \frac{1}{2} \sum_{r,s} \gamma_{rs} v_r v_s \right) \right\} d\tau \right), \]

\[ = \kappa + \sum_{v_i \in D} \omega_i v_i + \sum_{v_i, v_j \in \tau} \gamma_{ij} v_i v_j \] (43)

by putting

\[ \kappa_c(d) = \sum_{v_j \in \tau \setminus D} \gamma_{rj} v_j - \sum_{v_j \in D} \gamma_{rj} v_j \]

and noting that the integrand in (43) is the kernel of a Gaussian probability density. The symbols \( \kappa_c \), \( \omega_i \) and \( \gamma_{ij} \) stand for appropriate functions of the original parameters. We can see from (43) that \( U^*_D \) will adopt the form (38). In particular, all innovations will be sums of functions involving at most two variables. The study of whether these conditions could be met in other auto-models deserves more consideration, but we shall not pursue it any further here.
4.2. Collapsibility and precollapsibility

Collapsibility concerns arose originally in the context of log-linear models in contingency tables (see for example [1,4,37]). The main interest focussed in conditions for preserving parameter values or independence between classifying factors, when we collapse a contingency table and produce marginal tables of lower dimension.

Generally speaking, a probabilistic model on $V$ is collapsible onto $A \subset V$ (over $V \setminus A$) with regard to some specified property if the $A$-marginal model (which is obtained by integrating out the variables in $V \setminus A$) verifies the same property.

Due to its origin in contingency tables, collapsibility has had two main concerns which we can designate as parametric and graphical. By parametric collapsibility we understand preservation of the values of parameters or parametric functions such as association measures in contingency tables, or regression coefficients, etc. This is the approach in [8,9,12,19,21,29,34,35].

Graphical collapsibility is concerned with the stability of the model structure. A model possesses graphical collapsibility if its independence-graph structure is preserved by marginalization. This property depends on the flexibility of the model as well as on the particular subset of variables whose marginals we are looking for.

Frydenberg [17] stated a necessary and sufficient condition for graphical collapsibility of graphical models (see Example 5 for a brief definition). A graphical model on $V$ is collapsible onto $A \subset V$ iff every connected component of $V \setminus A$ is strongly simplicial. $B \subset V$ is strongly simplicial iff its boundary is complete and (i) all variables in $B$ are continuous, or (ii) the variables in $bd(B)$ are all discrete. The marginal graph coincides with the subgraph corresponding to the subset $A$.

Studeny [33] proves that for any undirected graph $\mathcal{G}$ on $V$, the class of $\mathcal{G}$-markovian discrete distributions on $V$ is closed under marginalization onto any subset $A \subset V$, i.e. their $A$-marginals constitute the class of $\mathcal{G}^A$-markovian discrete distributions for an appropriately defined marginal graph $\mathcal{G}^A$. He states a similar result for hypergraphs but in the framework of strictly positive discrete distributions. He calls this property precollapsibility of undirected graphs and hypergraphs. The price to pay for this generality is the need to include all possible discrete distributions in the mentioned class, without any restriction on the range of values of each variable in $V$. This is necessary in order to find the appropriate $\mathcal{G}$-markovian distribution corresponding to each $\mathcal{G}^A$-markovian one.

The two cited results establish graphical collapsibility by coupling model characteristics with marginal graph definitions. Frydenberg [17] considers graphical models and subgraphs while Studeny [33] circumscribes the
framework to discrete distributions and defines the marginal graph as explained in the comments after Theorem 2.

Theorem 4 shows instead the marginal hypergraph of any model with positive pdf without conditions on the kind of variables (discrete or continuous, bounded or not), the complexity of the interaction functions (maximum number of variables involved) or any distributional assumption (Gaussianity of continuous variables, etc.). It does not state a collapsibility condition, but a procedure to compute the marginal graph and hypergraph of any hypergraph model. Nevertheless, from (27) we can understand better the conditions leading to collapsibility.

With regard to graphical collapsibility, it becomes apparent from (27) that the graph $\mathcal{G}(|_A)$ associated with the marginal hypergraph coincides with the subgraph $\mathcal{G}(|_A)$ iff $\mathcal{H}_A \preceq \mathcal{H} |_A$ and $\mathcal{H}_A \preceq (\mathcal{H} |_A \setminus \mathcal{H}_A)$.

**Example 10 (Example 8 continued).** Let us consider again the modified model in the second part of Example 8, where the initial potential $U$ was enlarged with the terms $\alpha_{11} v_{11} v_{12}$ and $\alpha_{0} v_{8} v_{11} v_{12}$.

Now, if conditions (33) and (34) are not satisfied, no innovation will cancel the corresponding $U^A$ term, and $\mathcal{H}_A = \emptyset$.

If condition (33) holds but (34) does not, there will not be an $U^A_{\{v_{11} v_{12}\}}$ interaction, but the term $U^A_{\{v_{8} v_{11} v_{12}\}}$ will not be null, and $\mathcal{H}_A = \{\{v_{11}, v_{12}\}\} \preceq \mathcal{H} |_A \setminus \mathcal{H}_A$ because $\{v_{8}, v_{11}, v_{12}\} \in \mathcal{H} |_A \setminus \mathcal{H}_A$.

Both models will be collapsible onto $A$.

On the other hand, parametric collapsibility with regard to the symmetric measure of association given by the interaction function $U^A_b \in U^A$, requires that the corresponding innovation $V^A_b (b) \in (23)$ be null (this is obviously a sufficient condition). Collapsibility with regard to more involved parametric functions needs further elaboration and we will not pursue it any more here.

**5. Example of application**

In this example, the objective is to assess the damage of reinforced concrete structures of buildings. This example, which is taken from Liu and Li (1994) (see also [5]), is slightly modified for illustrative purposes. The goal variable (the damage of a reinforced concrete beam) is denoted by $X_1$. A civil engineer initially identifies 16 variables ($X_9, \ldots, X_{24}$) as the main variables influencing the damage of reinforced concrete structures. In addition, the engineer identifies seven intermediate unobservable variables ($X_2, \ldots, X_8$) that define some partial states of the structure. Table 1 shows the list of variables and their definitions.

In our example, the engineer specifies the following cause–effect relationships, as depicted in Fig. 3(a). The goal variable $X_1$, is related primarily to three
factors: $X_0$, the weakness of the beam available in the form of a damage factor; $X_{10}$, the deflection of the beam; and $X_2$, its cracking state. The cracking state, $X_2$, is related to four variables: $X_3$, the cracking state in the shear domain; $X_9$, the evaluation of the shrinkage cracking; $X_4$, the evaluation of the steel corrosion; and $X_5$, the cracking state in the flexure domain. Shrinkage cracking, $X_6$, is related to shrinkage, $X_23$, and the corrosion state, $X_8$. Steel corrosion, $X_4$, is related to $X_8$, $X_{24}$, and $X_5$. The cracking state in the shear domain, $X_3$, is related to four factors: $X_{11}$, the position of the worst shear crack; $X_{12}$, the breadth of the worst shear crack; $X_{21}$, the number of shear cracks; and $X_9$. The cracking state in the flexure domain, $X_5$ is affected by three variables: $X_{13}$, the position of the worst flexure crack; $X_{22}$, the number of flexure cracks; and $X_7$, the worst cracking state in the flexure domain. The variable $X_{13}$ is influenced by $X_4$. The variable $X_7$ is a function of five variables: $X_{14}$, the breadth of the worst flexure crack; $X_{15}$, the length of the worst flexure crack; $X_{16}$, the cover; $X_{17}$, the structure age; and $X_8$, the corrosion state. The variable $X_8$ is related to three variables: $X_{18}$, the humidity; $X_{19}$, the PH value in the air; and $X_{20}$, the content of chlorine in the air.

Table 1
Definitions of the variables related to damage assessment of reinforced concrete structures

<table>
<thead>
<tr>
<th>$X_i$</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>Damage assessment</td>
</tr>
<tr>
<td>$X_2$</td>
<td>Cracking state</td>
</tr>
<tr>
<td>$X_3$</td>
<td>Cracking state in shear domain</td>
</tr>
<tr>
<td>$X_4$</td>
<td>Steel corrosion</td>
</tr>
<tr>
<td>$X_5$</td>
<td>Cracking state in flexure domain</td>
</tr>
<tr>
<td>$X_6$</td>
<td>Shrinkage cracking</td>
</tr>
<tr>
<td>$X_7$</td>
<td>Worst cracking in flexure domain</td>
</tr>
<tr>
<td>$X_8$</td>
<td>Corrosion state</td>
</tr>
<tr>
<td>$X_9$</td>
<td>Weakness of the beam</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>Deflection of the beam</td>
</tr>
<tr>
<td>$X_{11}$</td>
<td>Position of the worst shear crack</td>
</tr>
<tr>
<td>$X_{12}$</td>
<td>Breadth of the worst shear crack</td>
</tr>
<tr>
<td>$X_{13}$</td>
<td>Position of the worst flexure crack</td>
</tr>
<tr>
<td>$X_{14}$</td>
<td>Breadth of the worst flexure crack</td>
</tr>
<tr>
<td>$X_{15}$</td>
<td>Length of the worst flexure cracks</td>
</tr>
<tr>
<td>$X_{16}$</td>
<td>Cover</td>
</tr>
<tr>
<td>$X_{17}$</td>
<td>Structure age</td>
</tr>
<tr>
<td>$X_{18}$</td>
<td>Humidity</td>
</tr>
<tr>
<td>$X_{19}$</td>
<td>PH value in the air</td>
</tr>
<tr>
<td>$X_{20}$</td>
<td>Content of chlorine in the air</td>
</tr>
<tr>
<td>$X_{21}$</td>
<td>Number of shear cracks</td>
</tr>
<tr>
<td>$X_{22}$</td>
<td>Number of flexure cracks</td>
</tr>
<tr>
<td>$X_{23}$</td>
<td>Shrinkage</td>
</tr>
<tr>
<td>$X_{24}$</td>
<td>Corrosion</td>
</tr>
</tbody>
</table>

A graphical representation of the damage problem is shown in Fig. 3(a).
Suppose that we are interested in suppressing all the nodes related to the
flection of the beam and keep the remaining nodes (Set $A$), that is (see
Fig. 3(b)):

$$
V \setminus A = \{X_5, X_7, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{22}, X_{23}\}.
$$

5.1. Graph approach

In this case, to marginalize over $A$, we can apply Algorithm 1.

**Step 1.** The set $E|_A$, i.e., the set of edges in the subgraph $G_A$ is shown in
Fig. 3(b) (the continuous edges in the region $A$).

**Step 2.** The subgraph $G_{V \setminus A}$ appears in Fig. 3(b) (region $A$ with continuous
edges).

**Step 3.** The connectivity components $\mathcal{T}$ of $G_{V \setminus A}$:

$$
\tau_1 = \{X_5, X_7, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{22}\};
$$

$$
\tau_2 = \{X_{23}\},
$$

are shown in Fig. 4(a) as white regions.

**Step 4.** The boundaries of the two connectivity components are
$bd(\tau_1) = \{X_2, X_4, X_8\}$ and $bd(\tau_2) = \{X_6\}$, as shown in Fig. 4(a) where they
have been shadowed with dots.

**Step 5.** To complete the set $bd(\tau_1)$ we need to add the edge $(X_2, X_8)$ to the
already two existing edges $(X_2, X_4)$ and $(X_4, X_8)$.

**Step 6.** We return the graph in Fig. 4(b), which incorporates the edge $(X_2, X_3)$
to the subgraph $G_A$, thus, showing that the graph $A$ is not collapsible with
respect to $A$. 
5.2. Hypergraph approach

When applying Algorithm 2, the differences with the preceding results could only appear in the boundaries of the connectivity components of $V \setminus A$, that is, $bd(\tau_1) = \{X_2, X_4, X_8\}$ and $bd(\tau_2) = \{X_6\}$. The non-null innovations (23) could only arise for subsets of variables contained in these sets. As $bd(\tau_2)$ has only one variable, our problem of exploring possible differences between $\mathcal{G}_A^{ma}$ and $\mathcal{G}(\mathcal{M}_A)$ reduce to those edges connecting variables in $bd(\tau_1)$.

To illustrate, let us assume a Gaussian distribution with mean $\mu$ and dispersion matrix $\Sigma$ for the 24 variables in $V$. To express this distribution as a hypergraph model, it is easier to work with the precision matrix $\mathcal{Y} = \Sigma^{-1}$. In fact its pdf can be written as

$$p(v) \propto \exp \left( -\frac{1}{2}(v - \mu)'\mathcal{Y}(v - \mu) \right)$$

$$\propto \exp \left( \sum_i v_i \left( \sum_j Y_{ij}\mu_j - \frac{1}{2} Y_{ij}v_i \right) - \sum_{i \neq j} Y_{ij}v_iv_j \right),$$

(44)

corresponding to expression (1) with normalized potential. Eq. (44) shows the relationship between edges in $\mathcal{G}$ and non-null elements of the matrix $\mathcal{Y}$.

It is a well known fact that the marginal distribution of a multivariate Gaussian model is again multivariate Gaussian, with precision matrix

$$\mathcal{Y}^d = \mathcal{Y}_{AA} - \mathcal{Y}_{AA,V\setminus A}(\mathcal{Y}_{V\setminus A,V\setminus A})^{-1}\mathcal{Y}_{V\setminus A,A},$$

(45)

where the subscripts of $\mathcal{Y}$ stand for the appropriate partition.

Fig. 4. (a) Set $V \setminus A$ with its connectivity components and their completed boundaries (doted regions), and (b) the resulting marginal graph $\mathcal{G}_A^m$ on $A$. 

Eq. (45) shows the decomposition of the precision matrix $\mathcal{T}^A$ related to the marginal normalized potential $U^A$ in two components:

- the matrix $\mathcal{T}^A_{\mathcal{A}}$ corresponding to the restricted potential $U|_{\mathcal{A}}$, and
- the matrix $\mathcal{T}^A = \mathcal{T}^A_{\mathcal{V}\setminus\mathcal{A}}(\mathcal{T}^A_{\mathcal{V}\setminus\mathcal{A},\mathcal{V}\setminus\mathcal{A}})^{-1}\mathcal{T}^A_{\mathcal{V}\setminus\mathcal{A}}$ corresponding to innovations $\mathcal{V}^A_B(b)$ of (23).

In particular, the innovation (23) for two variables $V_i$ and $V_j$ in $A$ is

$$V^A_{(V_i, V_j)}(v_i, v_j) = - \sum_{r,s \in \mathcal{V}\setminus\mathcal{A}} \rho_{rs} \mathcal{T}_{ir} \mathcal{T}_{sj} v_i v_j,$$

where $E$ stands for the set of edges of the graph $\mathcal{G}$ and $\rho_{rs}$ is the $rs$-element of the matrix $(\mathcal{T}^A_{\mathcal{V}\setminus\mathcal{A}})^{-1}$.

Particularizing to our example, the only edges subject to change when applying Algorithm 2 are $\{(X_2, X_4), (X_2, X_8), (X_4, X_8)\}$.

The edge $(X_2, X_8)$, which was not present in the original graph $\mathcal{G}$, arises as a consequence of the innovation $\mathcal{T}^A_{2,8} = \rho_{5,7} \mathcal{T}_{2,5} \mathcal{T}_{7,8}$, and it is null only if $\rho_{5,7}$ vanishes. Matrix $\mathcal{T}$, being a precision matrix, is definite positive, implying $D_{\mathcal{A}} = (\mathcal{T}^A_{\mathcal{V}\setminus\mathcal{A},\mathcal{V}\setminus\mathcal{A}})^{-1} > 0$. After some algebra, $\rho_{5,7}$ can be written as

$$\mathcal{T}_{13,13} \mathcal{T}_{14,14} \mathcal{T}_{15,15} \mathcal{T}_{16,16} \mathcal{T}_{17,17} \mathcal{T}_{22,22} \mathcal{T}_{23,23} \mathcal{T}_{5,7} \mathcal{T}_{7,8} / D,$$

which cannot be null unless one or more of the parameters $\mathcal{T}_{2,5}$, $\mathcal{T}_{5,7}$ and $\mathcal{T}_{7,8}$ vanish. But this would contradict the initial specification of $\mathcal{G}$. Then, the edge $(X_2, X_8)$ will always be present in $\mathcal{G}^{ma}$ and $\mathcal{H}(\mathcal{A})$.

Conditions for $(X_2, X_4)$ and $(X_4, X_8)$ to disappear in $\mathcal{G}(\mathcal{A})$ are $\mathcal{T}_{2,4} = \mathcal{T}_{4,2}$ and $\mathcal{T}_{4,8} = \mathcal{T}_{8,4}$, respectively.

They state functional relationships between the parameters $\mathcal{T}_{2,4}$, $\mathcal{T}_{4,8}$ and those in $\mathcal{T}^A_{\mathcal{V}\setminus\mathcal{A},\mathcal{V}\setminus\mathcal{A}}$. These relationships are compatible with the initial graph $\mathcal{G}$. Thus, the marginal graphs $\mathcal{G}^{ma}$ and $\mathcal{G}(\mathcal{A})$ could differ in edges $(X_2, X_4)$ and $(X_4, X_8)$, according to these conditions.

Thus, the example illustrates clearly the advantages of hypergraph models over the usual graph models.

6. Conclusions and recommendations

Hypergraph models have been shown to be a powerful alternative to undirected graph models. The main advantage consists of its capability to
produce finer factorizations and to catch a more complete set of conditional independence statements. Given a set of variables and an undirected graph or hypergraph model, two algorithms have been given for obtaining the corresponding marginal graph and hypergraph, such that the marginal distribution factorizes according to them. The examples have shown that in some cases the hypergraph is able to capture conditional independence statements that the graph fails to detect. In addition, Theorem 4 states a general framework to understand the necessary and sufficient conditions of graphical and parametric collapsibility.

Acknowledgements

The authors are grateful to Iberdrola, the Universities of Cantabria, Jaume I and Valencia, and the Dirección General de Investigación Científica y Técnica (DGICYT) (projects TIC96-0580 and PB96-0776) for partial support of this research.

References