Algorithmes de factorisation d'une loi de probabilité jointe en facteurs independents et minimaux

Maxime Gasse Alex Aussem

University of Lyon, CNRS
Université Lyon 1, LIRIS, UMR5205, F-69622, France
{maxime.gasse, alexandre.aussem}@liris.cnrs.fr

Résumé

Nous étudions le problème de la décomposition d'une distribution multivariée $p(\mathbf{v})$*, définie sur un ensemble de variables* $\mathbf{V} = \{V_1, \ldots, V_n\}$*, en un produit de facteurs définis sur des sous-ensembles disjoints* $\{\mathbf{V}_{F_1}, \ldots, \mathbf{V}_{F_m}\}$*. Nous montrons que la décomposition de* $\mathbf{V}$ *en facteurs disjoints irréductibles forme une partition unique qui correspond aux composantes connexes d'un réseau Bayésien ou un champ aléatoire de Markov, dès lors que ces graphes représentent* fidèlement *la distribution* p*. Nous proposons trois procédures génériques pour trouver ces facteurs avec* $O(n^2)$ *tests d'indépendance conditionnelle deux à deux* ($V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$)*, sous des hypothèses beaucoup moins restrictives : i)* p *vérifie la propriété d'Intersection ; ii) de Composition ; iii) ne vérifie aucune hypothèse.*

Mots Clefs

Modèles graphiques probabilistes, indépendance conditionnelle, factorisation de loi de probabilité, graphoides.

Abstract

We study the problem of decomposing a multivariate probability distribution $p(\mathbf{v})$ *defined over a set of random variables* $\mathbf{V} = \{V_1, \ldots, V_n\}$ *into a product of factors defined over disjoint subsets* $\{\mathbf{V}_{F_1}, \ldots, \mathbf{V}_{F_m}\}$*. We show that the decomposition of* $\mathbf{V}$ *into irreducible disjoint factors forms a unique partition, which corresponds to the connected components of a Bayesian or Markov network, given that it is faithful to* p*. Finally, we provide three generic procedures to identify these factors with* $O(n^2)$ *pairwise conditional independence tests* ($V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$) *under much less restrictive assumptions : 1)* p *supports the property of Intersection ; ii) of Composition ; iii) no assumption at all.*

Keywords

Probabilistic graphical models, conditional independence, probability distribution factorization, graphoids.

1 Introduction

The whole point of modeling a multivariate probability distribution p with a probabilistic graphical model, namely a Bayesian or a Markov network, is to encode independence relations into the graphical structure $\mathcal{G}$, thereby factorizing the joint probability distribution into a product of potential functions,

$$p(\mathbf{v}) = \prod_{i=1}^{m} \Phi_i(\mathbf{v}_i).$$

Such a factorization acts as a structural constraint on the expression of p, which reduces the number of free parameters in the model and facilitates both the learning and inference tasks, i.e. estimating p from a set of data samples, and answering probabilistic queries such as $\arg\max_{\mathbf{v}} p(\mathbf{v})$.

The fundamental problem that we wish to address in this paper involves finding a factorization of p into potential functions defined over minimal disjoint subsets, called *irreducible disjoint factors* (IDF as a shorthand). Such a factorization represents a strong structural constraint, and simplifies greatly the expression of p. For example, given two disjoint factors $\mathbf{V}_1$ and $\mathbf{V}_2$, the task of obtaining $\arg\max_{\mathbf{v}} p(\mathbf{v})$ can be decomposed into two independent problems $\arg\max_{\mathbf{v}_1} p(\mathbf{v}_1)$ and $\arg\max_{\mathbf{v}_2} p(\mathbf{v}_2)$. Finding a set of disjoint factors is, for instance, an essential task in Sum-Product network (SPN) structure learning [4], where product nodes correspond exactly to a product between disjoint factors, i.e. $p(\mathbf{v}_1) \times p(\mathbf{v}_2)$. Some of the results presented in this work have been published in a conference paper [3], where we showed that identifying disjoint factors in a conditional distribution $p(\mathbf{y}|\mathbf{x})$ can effectively improve the maximum-a-posteriori estimation $\arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x})$ in multi-label classification.

In Section 2 we define our notations and introduce the concept of irreducible disjoint factors as well as some basic properties of conditional independence on which our theoretical results will heavily rely. In Section 3 we show that irreducible disjoint factors necessarily form a unique partition, which relates to connected components in classical probabilistic graphical models. In Section 4, we establish several theoretical results to characterize the irreducible disjoint factors with pairwise conditional independence tests given several assumptions about p, namely the Intersection and Composition assumption, and then without any assumption. Each of these results establishes a quadratic generic procedure, which can be instantiated with only $O(n^2)$ statistical tests of independence. Finally, we conclude in Section 5.

2 Basic concepts

In this paper, upper-case letters in italics denote random variables (e.g. X, Y) and lower-case letters in italics denote their values (e.g. x, y). Likewise, upper-case bold letters denote random variable sets (e.g. $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$) and lower-case bold letters denote their values (e.g. $\mathbf{x}, \mathbf{y}, \mathbf{z}$). In the following we will consider only the multi-variate random variable $\mathbf{V} = \{V_1, \ldots, V_n\}$ and its subsets. To keep the notation uncluttered, we use $p(\mathbf{v})$ to denote $p(\mathbf{V} = \mathbf{v})$ the joint distribution of $\mathbf{V}$. We recall the definition of conditional independence,

Def. 1. $\mathbf{X}$ *is conditionally independent of* $\mathbf{Y}$ *given* $\mathbf{Z}$*, denoted* $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$*, when* $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ *are disjoint subsets of random variables such that for every value of* $\mathbf{x}, \mathbf{y}, \mathbf{z}$ *the following holds :*

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z})p(\mathbf{z}) = p(\mathbf{x}, \mathbf{z})p(\mathbf{y}, \mathbf{z}).\ [1]$$

We will assume the reader is familiar with the concept of separation in probabilistic graphical models, namely d-separation in directed acyclic graphs (DAGs) for Bayesian networks, and u-separation in undirected graphs (UGs) for Markov networks. These can be found in most books about probabilistic graphical models, e.g. PEARL [9], STUDENY [18] et KOLLER et FRIEDMAN [6].

2.1 Disjoint factorization

We shall now introduce the concept of disjoint factors of random variables that will play a pivotal role in the factorization of the distribution $p(\mathbf{v})$.

Def. 2. *A disjoint factor of random variables is a subset* $\mathbf{V}_F \subseteq \mathbf{V}$ *such that* $\mathbf{V}_F \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$*. Additionally, an irreducible disjoint factor is non-empty and has no other non-empty disjoint factor as proper subset.*

As we will show next, irreducible disjoint factors necessarily form a partition of $\mathbf{V}$, which we denote $\mathcal{F}_I$. The key idea is then to decompose the joint distribution of the variables into a product of marginal distributions,

$$p(\mathbf{v}) = \prod_{\mathbf{V}_F \in \mathcal{F}_I} p(\mathbf{v}_F).$$

This paper aims to obtain theoretical results for the characterization of the irreducible disjoint factors with only pairwise conditional independence relations in the form $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$.

1. Note that most definitions from the literature present the condition $p(\mathbf{x}, \mathbf{y}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})$, which rely on the positivity condition $p(\mathbf{z}) > 0$.

2.2 Conditional independence properties

Consider four mutually disjoint random variables, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, and p the underlying probability distribution. As shown in [2, 16], the properties of *Symmetry*, *Decomposition*, *Weak Union* and *Contraction* hold for any p, that is

$$\begin{gathered}
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \iff \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \text{ (sym.)},\\
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ (dec.)},\\
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \text{ (w.u.)},\\
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \text{ (con.)}.
\end{gathered}$$

Any independence model that respects these four properties is called a *semi-graphoid* [11]. A fifth property holds in strictly positive distributions ($p > 0$), i.e. the *Intersection* property

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z}.$$

Any independence model that respects these five properties is called a *graphoid*. The term "graphoid" was proposed by PEARL et PAZ [10] who noticed that these properties had striking similarities with vertex separation in graphs. Finally, a sixth property will be of particular interest in this work, that is the *Composition* property

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \wedge \mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z}.$$

The composition property holds in particular probability distributions, such as regular multivariate Gaussian distributions. Any independence model that respects these six properties is called a *compositional graphoid* [14]. As shown in [15], independence models induced by classic probabilistic graphical models are compositional graphoids.

3 Problem analysis

Let us now develop further the notion of irreducible disjoint factors, and derive a first general graphical characterization. All proofs of the Theorems and Lemmas presented in this Section are deferred to the Appendix.

3.1 Disjoint factors algebraic structure

We first show that disjoint factors can be characterized as an algebraic structure satisfying certain axioms. Let $\mathcal{F}$ denote the set of all disjoint factors (DFs for short) defined over $\mathbf{V}$, and $\mathcal{F}_I \subset \mathcal{F}$ the set of all irreducible disjoint factors (IDFs for short). It is easily shown that $\{\mathbf{V}, \emptyset\} \subseteq \mathcal{F}$. More specifically, the collection of all DFs in $\mathcal{F}$ can be ordered via subset inclusion to obtain a lattice bounded by $\mathbf{V}$ itself and the null set.

Thm. 1. *If* $\mathbf{V}_{F_i}, \mathbf{V}_{F_j} \in \mathcal{F}$*, then* $\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j} \in \mathcal{F}$ *and* $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} \in \mathcal{F}$*. Moreover, the decomposition of* $\mathbf{V}$ *into irreducible disjoint factors is unique.*

It follows from Definition 2 and Theorem 1 that the set of all IDFs $\mathcal{F}_I$ forms a partition of $\mathbf{V}$.

3.2 Graphical characterization

Irreducible disjoint factors will be conveniently represented as connected components in a graph, as we will see. Let us first introduce an important intermediary result.

Lem. 2. *Two distinct variables V_i and V_j belong to the same irreducible disjoint factor if there exists $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ such that $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}$.*

Note that the converse is not true, so Lemma 2 does not provide a complete characterization of irreducible disjoint factors. The following example illustrates that point.

Ex. 1. *Consider $\mathbf{V} = \{V_1, V_2, V_3\}$, with V_3 a quaternary random variable in $\{00, 01, 10, 11\}$ and V_1, V_2 two binary variables respectively equal to the first and the second binary digit of V_3. Then, we have that $V_1 \perp\!\!\!\perp V_2$ and $V_1 \perp\!\!\!\perp V_2 \mid V_3$, and yet V_1 and V_2 belong to the same IDF $\{V_1, V_2, V_3\}$ due to $V_1 \not\perp\!\!\!\perp V_3$ and $V_2 \not\perp\!\!\!\perp V_3$.*

We now expand on Lemma 2 to propose a complete characterization of the irreducible disjoint factors using graph properties.

Thm. 3. *Let $\mathcal{G}$ be an undirected graph whose nodes correspond to the random variables in $\mathbf{V}$, in which two nodes V_i and V_j are adjacent* iff *there exists $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ such that $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}$. Then, two variables V_i and V_j belong to the same IDF* iff *a path exists between them in $\mathcal{G}$.*

Theorem 3 offers an elegant graphical approach to characterize the IDFs, by mere inspection of the connected components in a graph. The problem of identifying all these connected components can be solved efficiently using a breadth-first search algorithm. Despite the desirable simplicity of this graphical characterization, deciding upon whether $\exists \mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ such that $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}$ remains a challenging combinatorial problem, an exhaustive search for $\mathbf{Z}$ being computationally infeasible even for moderate amounts of variables. Moreover, a second issue is that performing a statistical test of independence conditioned on a large $\mathbf{Z}$ can become problematic ; in the discrete, case the sample size required for high-confidence grows exponentially in the size of the conditioning set. We show next that it is possible to overcome these limitations by considering restrictive assumptions regarding p.

3.3 IDFs and PGM structures

Note that, due to the d-separation criterion for DAGs and the u-separation criterion for UGs, it is possible to read off the IDFs directly from a Bayesian network or Markov network structure, given that it is faithful to p.

Cor. 4. *Let $\mathcal{G}$ be a Bayesian or Markov network structure that is faithful to p. Then, two variables V_i and V_j belong to the same IDF* iff *there is a path between them in $\mathcal{G}$.*

Corollary 4 bridges the gap between the notion of irreducible disjoint factors and classical probabilistic graphical models. Still, the problem of structure learning for Bayesian and Markov networks is known to be NP-hard in general [1, 8], and we have no guarantee that the probability distribution underlying the data is faithful to a Bayesian network or a Markov network structure. In the next section we consider practical procedures inspired from constraint-based structure learning algorithms, which allow us to extract the IDFs without relying on a particular PGM structure.

4 Generic procedures

In this section, we address the problem of identifying the irreducible disjoint factors from pairwise conditional independence tests. Finding a sound and efficient algorithmic procedure for general distributions is not completely trivial as we shall see, so we may consider several (reasonable) assumptions about the underlying distribution p, namely the Intersection and Composition properties. All proofs of the Theorems and Lemmas presented in this Section are deferred to the Appendix.

4.1 Under the Intersection assumption

Let us present first a simplified characterization of the IDFs for distributions satisfying the Intersection property.

Thm. 5. *Let $\mathcal{G}$ be an undirected graph whose nodes correspond to the random variables in $\mathbf{V}$, in which two nodes V_i and V_j are adjacent* iff *$V_i \not\perp\!\!\!\perp Y_j \mid \mathbf{V} \setminus \{V_i, V_j\}$. Then, when p supports the Intersection property, two variables V_i and V_j belong to the same IDF* iff *there is a path between them in $\mathcal{G}$.*

Theorem 5 is appealing compared to Theorem 3, as it greatly reduces computational expense incurred in obtaining the irreducible disjoint factors, with only a quadratic number of conditional independence tests. Note that the resulting graph may not be the same, though under the Intersection property the connected components are identical. Still, the size of the conditioning set $\mathbf{V} \setminus \{V_i, V_j\}$ is problematic as it can be very high for large variable sets, which greatly reduces the confidence of a statistical test with limited samples. However, under the Intersection assumption the problem of performing that statistical test can be translated into a Markov boundary discovery problem, which can be solved with any off-the-shelf minimal feature subset selection algorithm.

Lem. 6. *Consider $V_i, V_j \in \mathbf{V}$ two distinct variables, and $\mathbf{M}_i$ a Markov boundary of V_i in $\mathbf{V}$. Then, $V_j \notin \mathbf{M}_i$ implies $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$, and the converse holds when p supports the Intersection property.*

Note that the Intersection assumption might be too restrictive in many practical scenarios. In fact, many real-life distributions (e.g. engineering systems such as digital circuits and engines that contain deterministic components) violate the Intersection property. As noted in [17], high-throughput molecular data, known as the "multiplicity" of molecular signatures (i.e., different gene/biomarker sets perform equally well in terms of predictive accuracy of phenotypes) also suggests existence of multiple Markov boundaries, which violates Intersection. It is usually unknown to what degree the Intersection assumption holds in distributions encountered in practice. The following example provides a particular case where the Intersection property does not hold.

Ex. 2. *Consider* $\mathbf{V} = \{V_1, V_2, V_3, V_4\}$ *four binary random variables such that* $V_1 = V_2$, $V_3 = V_4$ *and* $p(V_2 = V_3 \mid V_2) = \alpha$, $0.5 < \alpha < 1$. *Clearly here* p *is not strictly positive and the Intersection property does not hold. If we apply Theorem 5 along with Lemma 6 then we have* $\mathbf{M}_1 = \{V_2\}$, $\mathbf{M}_2 = \{V_1\}$, $\mathbf{M}_3 = \{V_4\}$ *and* $\mathbf{M}_4 = \{V_3\}$, *which results in two connected components* $\{V_1, V_2\}$ *and* $\{V_3, V_4\}$ *in* $\mathcal{G}$. *These are clearly not disjoint factors since* $\{V_1, V_2\} \not\perp\!\!\!\perp \{V_3, V_4\}$.

4.2 Under the Composition assumption

Second, we consider an even simpler characterization of the IDFs for distributions satisfying the Composition property.

Thm. 7. *Let* $\mathcal{G}$ *be an undirected graph whose nodes correspond to the random variables in* $\mathbf{V}$, *in which two nodes* V_i *and* V_j *are adjacent* iff $V_i \not\perp\!\!\!\perp V_j$. *Then, when* p *supports the Composition property, two variables* V_i *and* V_j *belong to the same IDF* iff *there is a path between them in* $\mathcal{G}$.

Theorem 7 is very similar to Theorem 5. Again, the number of independence tests required is quadratic, and the resulting graph may not be the same as the one from Theorem 3. The connected components though, are the same when p supports the Composition property. Moreover, a desirable property of this characterization is that the conditioning set vanishes, which ensures high confidence when performing a statistical test from finite samples.
Still, it is usually unknown to what degree the Composition assumption holds in distributions encountered in practice. Some special distributions are known to satisfy the Composition property, for example multivariate Gaussian distributions [18, Corollary 2.4] and the symmetric binary distributions used in [19]. The following example provides a case where the Composition property does not hold.

Ex. 3. *Consider* $\mathbf{V} = \{V_1, V_2, V_3\}$ *three binary variables such that* V_2 *and* V_3 *are independent and uniformly distributed, and* $p(V_1 = V_2 \oplus V_3 \mid V_2 \oplus V_3) = \alpha$, $0.5 < \alpha < 1$ *(*$\oplus$ *denotes the exclusive OR operator). If we apply Theorem 7 we have that every pair of variables is mutually independent, which results in three connected components* $\{V_1\}$, $\{V_2\}$ *and* $\{V_3\}$ *in* $\mathcal{G}$. *These are clearly not disjoint factors since* $\{V_1\} \not\perp\!\!\!\perp \{V_2, V_3\}$, $\{V_2\} \not\perp\!\!\!\perp \{V_3, V_1\}$ *and* $\{V_3\} \not\perp\!\!\!\perp \{V_1, V_2\}$.

4.3 For any probability distribution

Finally, we present a less trivial characterization of the IDFs that also loosens the computational burden by orders of magnitude compared to Theorem 3, and yet does not require any assumption about p.

Thm. 8. *Consider* $<$ *a strict total order of* $\mathbf{V}$. *Let* $\mathcal{G}$ *be an undirected graph whose nodes correspond to the random variables in* $\mathbf{V}$, *obtained from the following procedure :*

1: $\mathcal{G} \leftarrow (\mathbf{V}, \emptyset)$ *(empty graph)*
2: ***for all*** $V_i \in \mathbf{V}$ ***do***
3: $\quad \mathbf{V}^i_{ind} \leftarrow \emptyset$
4: $\quad$ ***for all*** $V_j \in \{V | V > V_i\}$ ***do***
5: $\quad\quad$ ***if*** $V_i \perp\!\!\!\perp V_j \mid \{V | V < V_i\} \cup \mathbf{V}^i_{ind}$ ***then***
6: $\quad\quad\quad \mathbf{V}^i_{ind} \leftarrow \mathbf{V}^i_{ind} \cup \{V_j\}$
7: $\quad\quad$ ***else***
8: $\quad\quad\quad$ *Insert a new edge* (i, j) *in* $\mathcal{G}$

Then, two variables V_i *and* V_j *belong to the same IDF* iff *there is a path between them in* $\mathcal{G}$.

Here again, the number of conditional independence tests required in Theorem 8 is quadratic in the number of variables. Compared to the previous results under the Intersection and the Composition properties, this new characterization has the desirable advantage of requiring no assumption about the underlying distribution p. However, it suffers from two limitations : i) the conditioning set at line 5 ranges from $\emptyset$ in the first iteration to $\mathbf{V} \setminus \{V_i, V_j\}$ in the last iteration, which is is problematic in high-dimensional data ; and ii) the whole procedure is prone to a propagation of error, since each iteration depends on the result of the previous tests to constitute the $\mathbf{V}^i_{ind}$ set. Also, the procedure can not be fully run in parallel, contrary to the procedures in Theorems 5 and 7.

5 Discussion

We presented three procedures based on pairwise conditional independence tests to identify the irreducible disjoint factors of a multivariate probability distribution $p(\mathbf{v})$. These procedures require only a quadratic number of independence tests. The first one is only correct under the assumption that p supports the Intersection property, and involves conditional independence tests in the form $V_i \perp\!\!\!\perp V_j \mid \mathbf{V} \setminus \{V_i, V_j\}$ between each pair of variables $V_i, V_j \in \mathbf{V}$. The second one is correct under the assumption that p supports the Composition property, and involves conditional independence tests in the form $V_i \perp\!\!\!\perp V_j \mid \emptyset$. Finally, the third procedure we propose is correct for any probability distribution p, and involves conditional independence tests in the form $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$, where $\mathbf{Z}$ is updated iteratively from the outcome of the previous tests, and ranges from $\emptyset$ to $\mathbf{V} \setminus \{V_i, V_j\}$.
Among the three generic procedures presented in Theorems 5, 7 and 8, the second procedure (under Composition) is the more appealing, in our view, since it relies on low-order conditional independence test, which are more robust

in practice. Moreover, the Composition property is usually considered as a reasonable assumption, and often tacitly assumed. For example, linear models rely on the Composition property. In the context of feature subset selection, it is often argued that forward selection is computationally more efficient than backward elimination [5]. In fact such a statement tacitly supposes that the Composition property holds [12]. Interestingly, the procedure used for SPN structure learning in [4] to "*partition* $\mathbf{V}$ *into approximately independent subsets* $\mathbf{V}_j$" can be seen as a direct instantiation of Theorem 7 with a G-test of pairwise independence. We proved therefore that this particular procedure in [4] is in fact correct and optimal (i.e., it yields independent and minimal subsets) when p supports the Composition property. Finally, the problem of decomposing p into irreducible disjoint factors seems closely related to the so-called "all-relevant" feature subset selection problem discussed in [13, 7], where a variable V_i is said to be relevant to another variable V_j *iff* there exists $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ such that $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}$. The graph in Theorem 3 provides a straightforward solution to this problem, therefore it may be interesting to investigate further how the graphs in Theorems 5, 7 and 8 may solve the "all-relevant" feature selection problem. This is left for future work..

Acknowledgments

This work was funded by both the French state trough the Nano 2017 investment program and the European Community through the European Nanoelectronics Initiative Advisory Council (ENIAC Joint Undertaking), under grant agreement no 324271 (ENI.237.1.B2013).

Références

[1] D. M. Chickering, D. Heckerman et C. Meek. Large-Sample Learning of Bayesian Networks is NP-Hard. *Journal of Machine Learning Research* 5 (2004), p. 1287–1330.

[2] A. P. Dawid. Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society, Series B* 41 (1979), p. 1–31.

[3] M. Gasse, A. Aussem et H. Elghazel. On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property. *ICML*. T. 37. 2015, p. 2531–2539.

[4] R. Gens et P. M. Domingos. Learning the Structure of Sum-Product Networks. *ICML (3)*. T. 28. 2013, p. 873–880.

[5] I. Guyon et A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 3 (2003), p. 1157–1182.

[6] D. Koller et N. Friedman. Probabilistic Graphical Models - Principles and Techniques. MIT Press, 2009, p. I–XXXV, 1–1231.

[7] R. Nilsson et al. Consistent Feature Selection for Pattern Recognition in Polynomial Time. *Journal of Machine Learning Research* 8 (2007), p. 589–612.

[8] S. Ordyniak et S. Szeider. Algorithms and Complexity Results for Exact Bayesian Structure Learning. *UAI 2010, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, July 8-11, 2010*. 2010, p. 401–408.

[9] J. Pearl. Probabilistic reasoning in intelligent systems - networks of plausible inference. Morgan Kaufmann, 1989, p. I–XIX, 1–552.

[10] J. Pearl et A. Paz. Graphoids : Graph-Based Logic for Reasoning about Relevance Relations or When would x tell you more about y if you already know z ? *ECAI*. 1986, p. 357–363.

[11] J. Pearl et T. Verma. The Logic of Representing Dependencies by Directed Graphs. *AAAI*. 1987, p. 374–379.

[12] J. M. Peña et al. Towards scalable and data efficient learning of Markov boundaries. *Int. J. Approx. Reasoning* 45.2 (2007), p. 211–232.

[13] W. R. Rudnicki, M. Wrzesien et W. Paja. All Relevant Feature Selection Methods and Applications. In : *Feature Selection for Data and Pattern Recognition*. T. 584. 2015, p. 11–28.

[14] K. Sadeghi et S. Lauritzen. Markov properties for mixed graphs. *Bernoulli* 20.2 (2014), p. 676–696.

[15] K. Sadeghi et S. L. Lauritzen. Unifying Markov properties in graphical models. Unpublished manuscript (2015).

[16] W. Spohn. Stochastic independence, causal independence, and shieldability. *J. Philosophical Logic* 9.1 (1980), p. 73–99.

[17] A. R. Statnikov, J. Lemeire et C. F. Aliferis. Algorithms for discovery of multiple Markov boundaries. *Journal of Machine Learning Research* 14.1 (2013), p. 499–566.

[18] M. Studeny. Probabilistic Conditional Independence Structures. 1re éd. Springer, 2005.

[19] N. Wermuth, G. M. Marchetti et D. R. Cox. Triangular systems for symmetric binary variables. *Electron. J. Statist.* 3 (2009), p. 932–955.

A Proofs

Proof of Theorem 1. First, we prove that $\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j} \in \mathcal{F}$. From the DF assumption for $\mathbf{V}_{F_i}$ and $\mathbf{V}_{F_j}$ we have $\mathbf{V}_{F_i} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_{F_i}$ and $\mathbf{V}_{F_j} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_{F_j}$. Using the Weak Union property we obtain that $\mathbf{V}_{F_i} \perp\!\!\!\perp \mathbf{V} \setminus (\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j}) \mid \mathbf{V}_{F_j} \setminus \mathbf{V}_{F_i}$, and similarly with the Decomposition property we get $\mathbf{V}_{F_j} \setminus \mathbf{V}_{F_i} \perp\!\!\!\perp \mathbf{V} \setminus (\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j})$. We may now apply the Contraction property to show that $\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j} \perp\!\!\!\perp \mathbf{V} \setminus (\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j})$. Therefore, $\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j}$ is a DF by definition. Second, we prove that $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} \in \mathcal{F}$. From the DF assumption for $\mathbf{V}_{F_i}$ and $\mathbf{V}_{F_j}$ we have $\mathbf{V}_{F_i} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_{F_i}$ and $\mathbf{V}_{F_j} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_{F_j}$. Using the Weak Union property we obtain $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} \perp\!\!\!\perp (\mathbf{V} \setminus (\mathbf{V}_{F_i} \cup \mathbf{V}_{F_j})) \cup (\mathbf{V}_{F_j} \setminus \mathbf{V}_{F_i}) \mid \mathbf{V}_{F_i} \setminus \mathbf{V}_{F_j}$, and similarly with the Decomposition property we get $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} \perp\!\!\!\perp \mathbf{V}_{F_i} \setminus \mathbf{V}_{F_j}$. We may now apply the Contraction property to show that $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} \perp\!\!\!\perp \mathbf{V} \setminus (\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j})$. Therefore, $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j}$ is a DF by definition. Third, we prove by contradiction that the decomposition of $\mathbf{V}$ into IDFs is unique. Suppose it is not the case, then there exists two distinct and overlapping IDFs $\mathbf{V}_{F_i}$ and $\mathbf{V}_{F_j}$, i.e. $\mathbf{V}_{F_i} \neq \mathbf{V}_{F_j}$ and $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} \neq \emptyset$. As $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j}$ is also a distinct factor, then due to the IDF assumption we have either $\mathbf{V}_{F_i} = \mathbf{V}_{F_j}$ or $\mathbf{V}_{F_i} \cap \mathbf{V}_{F_j} = \emptyset$ which shows the desired result. □

Proof of Lemma 2. By contradiction, suppose V_i and V_j do not belong to the same IDF, and let $\mathbf{V}_{F_i}$ denote the irreducible disjoint factor to which Y_i belongs. From the DF definition we have $\mathbf{V}_{F_i} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_{F_i}$. Let $\mathbf{Z}$ denote any arbitrary subset of $\mathbf{V} \setminus \{V_i, V_j\}$, we can apply the Weak Union property to obtain $\mathbf{V}_{F_i} \setminus \mathbf{Z} \perp\!\!\!\perp \mathbf{V} \setminus (\mathbf{V}_{F_i} \cup \mathbf{Z}) \mid \mathbf{Z}$. Then, from the Decomposition property we have $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$. This is true for every such $\mathbf{Z}$ subset, which concludes the proof. □

We now introduce Lemma 9 which will prove very useful to our subsequent demonstrations.

Lem. 9. *Let $\mathbf{V}_F$ be an IDF. Then, for every nonempty proper subset $\mathbf{Z}$ of $\mathbf{V}_F$, we have $\mathbf{Z} \not\perp\!\!\!\perp \mathbf{V}_F \setminus \mathbf{Z} \mid \mathbf{V} \setminus \mathbf{V}_F$. Additionally, if p satisfies the Composition property, then we have $\mathbf{Z} \not\perp\!\!\!\perp \mathbf{V}_F \setminus \mathbf{Z}$.*

Proof of Lemma 9. By contradiction, suppose such a $\mathbf{Z}$ exists with $\mathbf{Z} \perp\!\!\!\perp \mathbf{V}_F \setminus \mathbf{Z} \mid \mathbf{V} \setminus \mathbf{V}_F$. From the DF assumption of $\mathbf{V}_F$, we also have that $\mathbf{V}_F \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$, and therefore $\mathbf{Z} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$ due to the Decomposition property. We may now apply the Contraction property on these two statements to obtain $\mathbf{Z} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{Z}$ which contradicts the IDF assumption for $\mathbf{V}_F$. When p satisfies the Composition property, we proceed in the same way. Suppose such a $\mathbf{Z}$ exists with $\mathbf{Z} \perp\!\!\!\perp \mathbf{V}_F \setminus \mathbf{Z}$. We also have that $\mathbf{Z} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{V}_F$ and therefore $\mathbf{Z} \perp\!\!\!\perp \mathbf{V} \setminus \mathbf{Z}$ due to the Composition property, which again contradicts the IDF assumption of $\mathbf{V}_F$. This concludes the proof. □

Proof of Theorem 3. If a path exists between V_i and V_j in $\mathcal{G}$ then owing to Lemma 2 all pairs of successive variables in the path are in the same IDF, and by transitivity V_i and V_j necessarily belong to the same IDF. We may now prove the converse. Suppose that V_i and V_j belong to the same IDF, denoted $\mathbf{V}_F$. Define $\{\mathbf{X}, \mathbf{Y}\}$ a partition of $\mathbf{V}$ such that $V_i \in \mathbf{X}$ and $V_j \in \mathbf{Y}$. Then, owing to Lemma 9, we have that $\mathbf{X} \cap \mathbf{V}_F \not\perp\!\!\!\perp \mathbf{Y} \cap \mathbf{V}_F \mid \mathbf{V} \setminus \mathbf{V}_F$. Using the Weak Union property, we obtain $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}$. Consider X_1 an arbitrary variable from $\mathbf{X}$. Using the Contraction property, we have that either $\{X_1\} \not\perp\!\!\!\perp \mathbf{Y}$ or $\mathbf{X} \setminus \{X_1\} \not\perp\!\!\!\perp \mathbf{Y} \mid \{X_1\}$. Consider X_2 another arbitrary variable from $\mathbf{X} \setminus \{X_1\}$, we can apply the Contraction property again on the second expression to obtain that either $\{X_2\} \not\perp\!\!\!\perp \mathbf{Y} \mid \{X_1\}$ or $\mathbf{X} \setminus \{X_1, X_2\} \not\perp\!\!\!\perp \mathbf{Y} \mid \{X_1, X_2\}$. If we proceed recursively, we will necessarily find a variable $X_k \in \mathbf{X}$ such that $\{X_k\} \not\perp\!\!\!\perp \mathbf{Y} \mid \{X_1, \ldots, X_{k-1}\}$. Likewise, we can proceed along the same line to exhibit a variable $Y_l \in \mathbf{Y}$ such that $\{X_k\} \not\perp\!\!\!\perp \{Y_l\} \mid \{X_1, \ldots, X_{k-1}\} \cup \{Y_1, \ldots, Y_{l-1}\}$. In other words, for every partition $\{\mathbf{X}, \mathbf{Y}\}$ of $\mathbf{V}$ such that $V_i \in \mathbf{X}$ and $V_j \in \mathbf{Y}$, there exists at least one variable X in $\mathbf{X}$, one variable Y in $\mathbf{Y}$ and one subset $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$, such that $X \not\perp\!\!\!\perp Y \mid \mathbf{Z}$. So there necessarily exists a path between V_i and V_j in $\mathcal{G}$. This concludes the proof. □

Proof of Corollary 4. From the d-separation criterion in DAGs (resp. the u-separation criterion in UGs), if $\mathcal{G}$ is faithful to p, then $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}$ *iff* there is an open path between V_i and V_j given $\mathbf{Z}$. Since every path can be made open by conditioning on its collider nodes (resp. on the empty set), then for every pair of distinct variables $V_i, V_j \in \mathbf{V}$ connected by a path, there exists a subset $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$ such that $V_i \not\perp\!\!\!\perp V_j \mid \mathbf{Z}$. Conversely, if there exists no path between V_i and V_j, then $V_i \perp\!\!\!\perp V_j \mid \mathbf{Z}$ for every such a $\mathbf{Z}$ subset. This concludes the proof. □

Proof of Theorem 5. If a path exists between V_i and V_j in $\mathcal{G}$ then owing to Lemma 2 all pairs of successive variables in the path are in the same IDF, and by transitivity V_i and V_j necessarily belong to the same IDF. We may now prove the converse. Suppose that V_i and V_j belong to the same IDF, denoted $\mathbf{V}_F$. Define $\{\mathbf{X}, \mathbf{Y}\}$ a partition of $\mathbf{V}$ such that $V_i \in \mathbf{X}$ and $V_j \in \mathbf{Y}$. Then, owing to Lemma 9, we have that $\mathbf{X} \cap \mathbf{V}_F \not\perp\!\!\!\perp \mathbf{Y} \cap \mathbf{V}_F \mid \mathbf{V} \setminus \mathbf{V}_F$. Using the Weak Union property, we obtain $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}$. Consider X_1 an arbitrary variable from $\mathbf{X}$. Using the Intersection property, we have that either $\{X_1\} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X} \setminus \{X_1\}$ or $\mathbf{X} \setminus \{X_1\} \not\perp\!\!\!\perp \mathbf{Y} \mid \{X_1\}$. Consider X_2 another arbitrary variable from $\mathbf{X} \setminus \{X_1\}$, we can apply the Intersection property again on the second expression to obtain that either $\{X_2\} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X} \setminus \{X_2\}$ or $\mathbf{X} \setminus \{X_1, X_2\} \not\perp\!\!\!\perp \mathbf{Y} \mid \{X_1, X_2\}$. If we proceed recursively, we will necessarily find a variable $X_k \in \mathbf{X}$ such that $\{X_k\} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X} \setminus \{X_k\}$. Likewise, we can proceed along the same line to exhibit a variable $Y_l \in \mathbf{Y}$ such that $\{X_k\} \not\perp\!\!\!\perp \{Y_l\} \mid \mathbf{V} \setminus \{X_k, Y_l\}$. In other words, for every partition $\{\mathbf{X}, \mathbf{Y}\}$ of $\mathbf{V}$ such that $V_i \in \mathbf{X}$ and $V_j \in \mathbf{Y}$, there exists at least one variable X in $\mathbf{X}$ and one variable Y in $\mathbf{Y}$ such that $X \not\perp\!\!\!\perp Y \mid \mathbf{V} \setminus \{X, Y\}$. So there necessarily exists a path between V_i and V_j in $\mathcal{G}$. This concludes the proof. □

Proof of Lemma 6. First, if $\{V_i\} \not\perp\!\!\!\perp \{V_j\} \mid \mathbf{V} \setminus \{V_i, V_j\}$ then from the Weak Union property we have that $\{V_i\} \not\perp\!\!\!\perp \mathbf{V} \setminus (\{V_i\} \cup \mathbf{Z}) \mid \mathbf{Z}$ for every $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_i, V_j\}$. In other words, there exists no Markov blanket (neither

Markov boundary) of V_i in $\mathbf{V}$ which does not contain V_j. Second, we prove the converse. Suppose the Markov boundary $\mathbf{M}_i$ of V_i in $\mathbf{V}$ contains V_j, then we have $\{V_i\} \not\perp\!\!\!\perp \{V_j\} \cup \mathbf{V} \setminus (\{V_i\} \cup \mathbf{M}_i) \mid \mathbf{M}_i \setminus \{V_j\}$. We can apply the Intersection property to obtain that either $\{V_i\} \not\perp\!\!\!\perp \{V_j\} \mid \mathbf{V} \setminus \{V_i, V_j\}$ or $\{V_i\} \not\perp\!\!\!\perp \mathbf{V} \setminus (\{V_i\} \cup \mathbf{M}_i) \mid \mathbf{M}_i$. The second statement contradicts the Markov blanket assumption for $\mathbf{M}_i$, so we necessarily have that $\{V_i\} \not\perp\!\!\!\perp \{V_j\} \mid \mathbf{V} \setminus \{V_i, V_j\}$. This concludes the proof. □

Proof of Theorem 7. If a path exists between V_i and V_j in $\mathcal{G}$ then owing to Lemma 2 all pairs of successive variables in the path are in the same IDF, and by transitivity V_i and V_j necessarily belong to the same IDF. We may now prove the converse. Suppose that V_i and V_j belong to the same IDF, denoted $\mathbf{V}_F$. Define $\{\mathbf{X}, \mathbf{Y}\}$ a partition of $\mathbf{V}$ such that $V_i \in \mathbf{X}$ and $V_j \in \mathbf{Y}$. Then, owing to Lemma 9, we have that $\mathbf{X} \cap \mathbf{V}_F \not\perp\!\!\!\perp \mathbf{Y} \cap \mathbf{V}_F \mid \mathbf{V} \setminus \mathbf{V}_F$. Using the Weak Union property, we obtain $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}$. Consider X_1 an arbitrary variable from $\mathbf{X}$. Using the Composition property, we have that either $\{X_1\} \not\perp\!\!\!\perp \mathbf{Y}$ or $\mathbf{X} \setminus \{X_1\} \not\perp\!\!\!\perp \mathbf{Y}$. Consider X_2 another arbitrary variable from $\mathbf{X} \setminus \{X_1\}$, we can apply the Composition property again on the second expression to obtain that either $\{X_2\} \not\perp\!\!\!\perp \mathbf{Y}$ or $\mathbf{X} \setminus \{X_1, X_2\} \not\perp\!\!\!\perp \mathbf{Y}$. If we proceed recursively, we will necessarily find a variable $X_k \in \mathbf{X}$ such that $\{X_k\} \not\perp\!\!\!\perp \mathbf{Y}$. Likewise, we can proceed along the same line to exhibit a variable $Y_l \in \mathbf{Y}$ such that $\{X_k\} \not\perp\!\!\!\perp \{Y_l\}$. In other words, for every partition $\{\mathbf{X}, \mathbf{Y}\}$ of $\mathbf{V}$ such that $V_i \in \mathbf{X}$ and $V_j \in \mathbf{Y}$, there exists at least one variable X in $\mathbf{X}$ and one variable Y in $\mathbf{Y}$ such that $X \not\perp\!\!\!\perp Y$. So there necessarily exists a path between V_i and V_j in $\mathcal{G}$. This concludes the proof. □

Proof of Theorem 8. First, let $\mathbf{V}^{i,j}_{ind}$ denote $\mathbf{V}^{i}_{ind}$ in its intermediary state at line 5 when V_j is being processed, while $\mathbf{V}^{i}_{ind}$ denotes its state at the end of the procedure.
We start by proving that V_i and V_j are in the same IDF if V_i and V_j are connected in $\mathcal{G}$. If two variables V_p and V_q (with $V_p < V_q$) are adjacent in $\mathcal{G}$, then there exists a set $\mathbf{V}^{p,q}_{ind}$ such that $\{V_p\} \not\perp\!\!\!\perp \{V_q\} \mid \{V|V < V_p\} \cup \mathbf{V}^{p,q}_{ind}$. So there exists a subset $\mathbf{Z} \subseteq \mathbf{V} \setminus \{V_p, V_q\}$ such that $\{V_p\} \not\perp\!\!\!\perp \{V_q\} \mid \mathbf{Z}$. From Lemma 2 we conclude that V_p and V_q belong to the same IDF. Now, if a path exists between V_i and V_j in $\mathcal{G}$, then all pairs of successive variables in the path are in the same IDF, and by transitivity V_i and V_j necessarily belong to the same IDF.
To show the converse, we shall prove by contradiction that if V_i and V_j belong to the same IDF, then there exists a path between V_i and V_j in $\mathcal{G}$. Suppose there is no such path, then there exists an ordering of the variables $\{V_1, \ldots, V_n\}$ and a partition $\{\mathbf{X}, \mathbf{Y}\}$ of $\mathbf{V}$ such that $V_i \in \mathbf{X}$, $V_j \in \mathbf{Y}$, and every variable $X \in \mathbf{X}$ is non-adjacent in $\mathcal{G}$ to every variable in $Y \in \mathbf{Y}$. Equivalently, for every variable $V_k \in \mathbf{V}$ we have $\{X|X > V_k\} \subseteq \mathbf{V}^{k}_{ind}$ if $V_k \in \mathbf{Y}$, and $\{Y|Y > V_k\} \subseteq \mathbf{V}^{k}_{ind}$ if $V_k \in \mathbf{X}$. To proceed, we shall first prove by induction that

$$\forall k > i, \;\; \{V_i\} \perp\!\!\!\perp \mathbf{V}^{i,k}_{ind} \mid \{V|V < V_i\}.$$

For $k = i + 1$, we have that $\mathbf{V}^{i,k}_{ind} = \emptyset$ so the result holds trivially. Suppose that $\{V_i\} \perp\!\!\!\perp \mathbf{V}^{i,k}_{ind} \mid \{V|V < V_i\}$ holds for some k. If $\{V_i\} \perp\!\!\!\perp \{V_k\} \mid \{V|V < V_i\} \cup \mathbf{V}^{i,k}_{ind}$, then $\mathbf{V}^{i,k+1}_{ind} = \mathbf{V}^{i,k}_{ind} \cup \{V_k\}$ and $\{V_i\} \perp\!\!\!\perp \mathbf{V}^{i,k+1}_{ind} \mid \{V|V < V_i\}$ due to the Contraction property. Otherwise, $\mathbf{V}^{i,k+1}_{ind} = \mathbf{V}^{i,k}_{ind}$ and we end up with the same result. Therefore, the result holds for every $k > i$. Now, we prove a second result by induction :

$$\forall k, \;\; \{X|X \geq V_k\} \perp\!\!\!\perp \{Y|Y \geq V_k\} \mid \{V|V < V_k\}.$$

The expression holds trivially for $k = n$. Consider the next variable, V_{k-1}, and suppose it belongs to $\mathbf{X}$, then we have $\{V_{k-1}\} \perp\!\!\!\perp \mathbf{V}^{k-1}_{ind} \mid \{V|V < V_{k-1}\}$. Since $\{Y|Y > V_{k-1}\} \subseteq \mathbf{V}^{k-1}_{ind}$, we may apply the Decomposition property to obtain $\{V_{k-1}\} \perp\!\!\!\perp \{Y|Y \geq V_{k-1}\} \mid \{V|V < V_{k-1}\}$. Combining that last expression with $\{X|X \geq V_k\} \perp\!\!\!\perp \{Y|Y \geq V_k\} \mid \{V|V < V_k\}$ yields $\{X|X \geq V_{k-1}\} \perp\!\!\!\perp \{Y|Y \geq V_{k-1}\} \mid \{V|V < V_{k-1}\}$ due to the Contraction property. The same demonstration holds if $V_{k-1} \in \mathbf{Y}$. Therefore, the results holds for every k by induction. Setting $k = 1$ in the expression above yields $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, therefore V_i and V_j belong to distinct IDFs. This concludes the proof. □