Marginals of DAG-Isomorphic Independence Models

Peter R. de Waal

Technical Report UU-CS-2008-050
December 2008

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

Marginals of DAG-Isomorphic Independence Models

Peter R. de Waal
Department of Information and Computing Sciences,
Faculty of Sciences,
Universiteit Utrecht,
P.O. Box 80089, 3508TB Utrecht, The Netherlands
waal@cs.uu.nl

Abstract

Probabilistic and graphical independence models both satisfy the semi-graphoid axioms, but their respective modelling powers are not equal. For every graphical independence model that is represented by d-separation in a directed acyclic graph, there exists an isomorphic probabilistic independence model, i.e. it has exactly the same independence statements. The reverse does not hold, as there exist probability distributions for which there are no perfect maps. We investigate if a given probabilistic independence model can be augmented with latent variables to a new independence model that is isomorphic with a graphical independence model of a directed acyclic graph. The original independence model can then be viewed as the marginal of the model with latent variables. We show that for some independence models we need infinitely many latent variables to accomplish this.

1 Introduction

Probabilistic models in artificial intelligence are typically built on the semi-graphoids axioms of independence. These axioms in fact are exploited explicitly in probabilistic graphical models, where independence is captured by topological properties, such as separation of vertices in an undirected graph or d-separation in a directed graph. A graphical representation with directed graphs for use in a decision support system has the advantage that it allows an intuitive interpretation by domain experts in terms of influences between the variables.

Ideally a probabilistic model is represented as a graphical model in a one-to-one way, that is, independence in the one representation implies independence in the other representation. The probabilistic model then is said to be isomorphic with the graphical model, and vice versa. Pearl and Paz [5] established a set of sufficient and necessary conditions under which a probabilistic model is isomorphic with an undirected graph. In this paper we shall not consider representations of independence with undirected graphs, but focus on directed representations. Contrary to undirected graphs directed graphs allow the representation of induced dependencies: if a specific independence has been established given some evidence, it is possible that this independence becomes invalid if more evidence is obtained. Pearl gave a set of necessary conditions for directed graph isomorphism in [6]. To the best of our knowledge there is no known set of sufficient conditions.

Pearl [6] also shows how a particular independence model that is not isomorphic with a directed graphical model, can be made isomorphic by the introduction of an auxiliary variable. In [6] the isomorphism is then established by conditioning on the auxiliary variable. In this paper we choose a different approach. We extend the model with auxiliary variables to a directed graph isomorph and we then take the marginal over the original variables of this extended model. For this we introduce the concept of the marginal of a formal independence model. The model with auxiliary variables can then be considered as a latent perfect map. We show that it is possible to establish isomorphism in this manner, but that we may need an infinite number of auxiliary variables to accomplish this. We also show that there exists a probabilistic independence model that needs infinitely many latent variables.

This paper is organised as follows. In Section 2 we briefly review probabilistic and graphical independence models, and the semi-graphoid properties of these models. In Section 3 we introduce the concept of marginals of an independence model and latent perfect maps. In Section 4 we discuss the existence of latent perfect maps, and in Section 5 we wrap up with conclusions and recommendations.

2 Preliminaries

In this section, we provide some preliminaries on probabilistic independence models as defined by conditional independence for probability distributions, graphical independence models as defined by d-separation in directed acyclic graphs, and formal independence models that capture the properties that probabilistic and graphical models have in common.

2.1 Conditional independence models

We consider a finite set of distinct symbols $V = \{V_1, \ldots, V_N\}$, called the *attributes* or *variable names*. With each variable V_i we associate a finite domain set $\mathcal{V}_i$, which is the set of possible values the variable can take. We define the domain of V as $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_N$, the Cartesian product of the domains of the individual variables.

A *probability measure* over V is defined by the domains $\mathcal{V}_i$, $i = 1, \ldots, N$, and a probability mapping $P : \mathcal{V} \to [0, 1]$ that satisfies the three basic axioms of probability theory [4].

For any subset $X = \{V_{i_1}, \ldots, V_{i_k}\} \subset V$, for some $k \geq 1$, we define the domain $\mathcal{X}$ of X as $\mathcal{X} = \mathcal{V}_{i_1} \times \cdots \mathcal{V}_{i_k}$. For a probability mapping P on V we define its *marginal* mapping over X, denoted by P^X, as the probability measure P^X over $\mathcal{X}$, defined by

$$P^X(x) = \sum \left\{ P(x, y) \,\middle|\, y \in \bigtimes_{\{i \mid V_i \notin X\}} \mathcal{V}_i \right\}$$

for $x \in \mathcal{X}$. By definition $P^V \equiv P$, $P^{\varnothing} \equiv 1$, and $(P^X)^Y = (P^Y)^X = P^{X \cap Y}$, for $X, Y \subset V$.

We denote the set of ordered triplets $(X, Y|Z)$ for disjoint subsets X, Y and Z of V as $\mathcal{T}(V)$. We shall use the notation $\mathcal{I}(X, Y|Z)$ to indicate $(X, Y|Z) \in \mathcal{I}$, for any ternary relation $\mathcal{I}$ on V. For simplicity of notation we will often write XY to denote the union $X \cup Y$, for $X, Y \subset V$. To avoid complicated notation we also allow Xy to denote $X \cup \{y\}$, for $X \subset V$ and $y \in V$.

Definition 2.1 (Conditional independence) *Let X, Y and Z be disjoint subsets of V, with domains $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. The sets X and Y are defined to be* conditionally independent under P given Z*, if for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$, we have*

$$P^{XYZ}(x, y, z) \cdot P^Z(z) = P^{XZ}(x, z) \cdot P^{YZ}(y, z)$$

Definition 2.2 *Let V be a set of variables and P a probability measure over V. The* probabilistic independence model $\mathcal{I}_P$ of P *is defined as the ternary relation $\mathcal{I}_P$ on V for which $\mathcal{I}_P(X, Y|Z)$ if and only if X and Y are conditionally independent under P given Z.*

If no ambiguity can arise we may omit the reference to the probability measure and just refer to the probabilistic independence model.

2.2 Graphical independence models in directed acyclic graphs

We first introduce the standard concepts of blocking and d-separation in directed graphs.

We consider a directed acyclic graph (DAG) $G = (V, A)$, with V the set of vertices and A the set of arcs. A path s in G of length $k - 1$ from a vertex V_{i_1} to V_{i_2} is a k-tuple $s = (W_1, W_2, \ldots, W_k)$ with $W_i \in V$ for $i = 1, \ldots, k$, $W_1 = V_{i_1}$, $W_k = V_{i_2}$ and for each $i = 1, \ldots, k - 1$ either $(W_i, W_{i+1}) \in A$ or $(W_{i+1}, W_i) \in A$. Without loss of generality we assume that a path has no loops, so there are no duplicates in $\{W_1, \ldots, W_k\}$. We define a path s to be *unidirectional* if all the arcs in s point in the same direction. More specifically, we define the unidirectional $s = (W_1, W_2, \ldots, W_k)$ to be a *descending* path if $(W_i, W_{i+1}) \in A$, for all $i = 1, \ldots, k - 1$. A vertex Y is called a descendant of a vertex X if there is a descending path from X to Y.

Definition 2.3 *Let Z be a subset of V. We say that a path s is* blocked *in G by Z, if s contains three consecutive vertices W_{i-1}, W_i, and W_{i+1} for which one of the following conditions hold:*

- *$W_{i-1} \leftarrow W_i \rightarrow W_{i+1}$, and $W_i \in Z$,*
- *$W_{i-1} \rightarrow W_i \rightarrow W_{i+1}$, and $W_i \in Z$,*
- *$W_{i-1} \leftarrow W_i \leftarrow W_{i+1}$, and $W_i \in Z$,*
- *$W_{i-1} \rightarrow W_i \leftarrow W_{i+1}$, and $\sigma(W_i) \cap Z = \varnothing$, where $\sigma(W_i)$ consists of W_i and all its descendants.*

We refer to the first three conditions as blocking by presence*, and the last condition as* blocking by absence*. We refer to node W_i in the last condition as a* converging *or* colliding node *on the path.*

While the concept of blocking is defined for a single path, the d-separation criterion applies to the set of all paths in G.

Definition 2.4 *Let $G = (V, A)$ be a DAG, and let X, Y and Z be disjoint subsets of V. The set Z is said to* d-separate *X and Y in G, if every path s between any variable $x \in X$ to any variable $y \in Y$ is blocked in G by Z.*

Based on the d-separation criterion we can define the notion of a graphical independence model.

Definition 2.5 *Let $G = (V, A)$ be a DAG. The* graphical independence model $\mathcal{I}_G$ defined by *G is a ternary relation on V such that $\mathcal{I}_G(X, Y|Z)$ if and only if Z d-separates X and Y in G.*

2.3 Formal independence models

Both a probabilistic independence model on a set of variables V and a graphical independence model on a DAG $G = (V, A)$ define a ternary relation on V. In fact we can capture this in a formal construct of an independence model.

Definition 2.6 *A* formal independence model *on a set V is a ternary relation on V.*

Both probabilistic and graphical independence models satisfy a set of axioms of independence. A special class within the set of formal independence models is defined based on these axioms.

Definition 2.7 *A ternary relation $\mathcal{I}$ on V is a* semi-graphoid independence model*, or semi-graphoid for short, if it satisfies the following four axioms:*

A1: $\mathcal{I}(X, Y|Z) \Rightarrow \mathcal{I}(Y, X|Z)$,

A2: $\mathcal{I}(X, YW|Z) \Rightarrow \mathcal{I}(X, Y|Z) \wedge \mathcal{I}(X, W|Z)$,

A3: $\mathcal{I}(X, YW|Z) \Rightarrow \mathcal{I}(X, Y|ZW)$,

A4: $\mathcal{I}(X, Y|Z) \wedge \mathcal{I}(X, W|ZY) \Rightarrow \mathcal{I}(X, YW|Z)$.

for all disjoint sets of variables $W, X, Y, Z \subset V$.

The axioms convey the idea that learning irrelevant information does not alter the relevance relationships among the other variables discerned. The four axioms are termed the *symmetry* (A1), *decomposition* (A2), *weak union* (A3) and the *contraction axiom* (A4), respectively.

The axioms were first introduced in [1] for probabilistic conditional independence. The properties were later recognised in artificial intelligence as properties of separation in graphs [5, 6], and are since known as the *semi-graphoid* axioms.

In the formulation that we have used so far we can allow X and Y to be empty, which leads to the so-called *trivial independence* axiom:

A0: $\mathcal{I}_P(X, \varnothing|Z)$,

This axiom trivially holds for both probabilistic independence and graphical independence.

An axiomatic representation allows us to derive qualitative statements about conditional independence that may not be immediate from a numerical representation of probabilities. It also enables a parsimonious specification of an independence model, since it is sufficient to enumerate the so-called dominating independence statements, from which all other statements can be derived by application of the axioms [9].

2.4 Graph-isomorph

Probabilistic independence models and graphical independence models both satisfy the semi-graphoid axioms, so it is interesting to investigate whether they have equal modelling power. Can any probabilistic independence model be represented by a graphical model, and vice versa? For this we introduce the notions of I-maps and P-maps.

Definition 2.8 *Let $\mathcal{I}$ be an formal independence model on V, and $G = (V, A)$ a DAG that defines a graphical independence model $\mathcal{I}_G$ through d-separation.*

1. *The graph G is called an* independence map*, or* I-map *for short, for $\mathcal{I}$, if for all disjoint X, Y, $Z \subset V$ we have: $\mathcal{I}_G(X, Y|Z) \Rightarrow \mathcal{I}(X, Y|Z)$. If G is an I-map for $\mathcal{I}$, and deleting any arc makes G cease to be an I-map for $\mathcal{I}$, then G is called a* minimal I-map *for $\mathcal{I}$.*
2. *The graph G is called a* perfect map*, or* P-map *for short, for $\mathcal{I}$, if for all disjoint X, Y, $Z \subset V$ we have: $\mathcal{I}_G(X, Y|Z) \Leftrightarrow \mathcal{I}(X, Y|Z)$,*

Definition 2.9 (DAG-isomorph) *An independence model $\mathcal{I}$ on V is said to be a* DAG-isomorph*, if there exists a graph $G = (V, A)$ that is a perfect map for $\mathcal{I}$.*

Since a graphical independence model satisfies the semi-graphoid axioms, a DAG-isomorph has to be a semi-graphoid itself. Being a semi-graphoid is not a sufficient condition for DAG-isomorphism, however. To the best of our knowledge there does not exists a sufficient set of conditions, although [6] presents a set of necessary conditions.

Some results from literature describe the modelling power of the independence models of the previous sections. Concerning the relationship between probabilistic and graphical models Geiger and Pearl show that for every DAG graphical model there exists a probability model for which that particular DAG is a perfect map [3]. The reverse does not hold, there exist probability models for which there is no DAG perfect map [6].

In [7] it is shown that the semi-graphoid axioms are not complete for probabilistic independence models. Studený derives a new axiom for probabilistic independence models that is not implied by the semi-graphoid axioms. He also shows in [8] that probabilistic independence models cannot be characterised by a finite set of inference rules.

3 Marginal of an independence model

A set of necessary conditions for a formal independence model to be a DAG-isomorph is known from [6]. The conditions are based on properties of d-separation in DAG's. One of the conditions that is not already implied by the semi-graphoid axioms is the so-called *chordality condition*:

$$\mathcal{I}(x,y|zw) \wedge \mathcal{I}(z,w|xy) \Rightarrow \mathcal{I}(x,y|z) \vee \mathcal{I}(x,y,w)$$

for all $x,y,z,w \in V$. Pearl shows in [6, Section 3.3.3] by example how conditioning on an auxiliary variable can be used to dispose of this chordality condition. In his example the independence model is not DAG-isomorph, but there exists a DAG with one extra variable, that, when conditioned on the auxiliary variable, is isomorphic with the independence model.

In this paper we take a different approach as we introduce an auxiliary variable without conditioning to create a DAG that is a P-map for an independence model. We formulate this in the following definition.

Definition 3.1 *Let $\mathcal{I}$ be an independence model on a set of variables V, and let A be a subset of V. We define the* marginal of $\mathcal{I}$ on A, *denoted by $\mathcal{I}^A$, as $\mathcal{I}^A = \mathcal{I} \cap \mathcal{T}(A)$.*

The following lemma follows immediately from this definition. It implies that taking the marginal of a probabilistic model is equivalent to taking the probabilistic model of the marginal of a probability measure. As such it justifies our use of the phrase "marginal of an independence model".

Lemma 3.2 *Let V be a set of variables, P a probability measure on V with probabilistic independence model $\mathcal{I}_P$. For every subset $A \subseteq V$ we have $\mathcal{I}_P^A = \mathcal{I}_{P^A}$.*

Proof. Let X, Y and Z be disjoint subsets of A, then the following holds:

$$\mathcal{I}_{P^A}(X,Y|Z)$$

$$\Leftrightarrow (P^A)^{XYZ}(x,y,z)(P^A)^Z(z) = (P^A)^{XZ}(x,z)(P^A)^{YZ}(y,z)$$

$$\Leftrightarrow P^{XYZ}(x,y,z)P^Z(z) = P^{XZ}(x,z)P^{YZ}(y,z)$$

$$\Leftrightarrow \mathcal{I}_P(X,Y|Z)$$

$$\Leftrightarrow \mathcal{I}_P^A(X,Y,Z)$$

The first and third steps follow from Definition 2.2, the second step from the observation that $(P^A)^W = P^W$ for any $W \subseteq A$, and the final step follows from $(X,Y|Z) \in \mathcal{T}(A)$ and Definition 3.1. □

We can now extend the definition of DAG-isomorphism.

Definition 3.3 (DAG-isomorph marginal) *Let V be a set of variables, and $\mathcal{I}$ an independence model on V. We say that $\mathcal{I}$ is a* DAG-isomorph marginal, *if there exists a finite set of variables $\overline{V} \supseteq V$, an independence model $\overline{\mathcal{I}}$ on $\overline{V}$ and a DAG $\overline{G} = (\overline{V}, \overline{A})$, such that $\overline{G}$ is a P-map for $\overline{\mathcal{I}}$ and $\overline{\mathcal{I}}^V = \mathcal{I}$. We then say that $\overline{G}$ is a* latent P-map *of $\mathcal{I}$.*

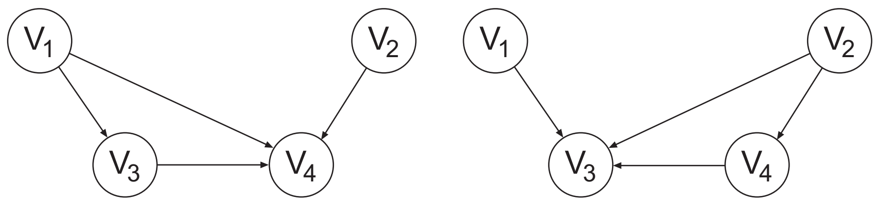

Figure 1: G_1 (left) and G_2 (right), minimal I-maps for $\mathcal{I}$.

Note that if $\mathcal{I}$ is a DAG-isomorph, then it is by Definition 3.3 also a DAG-isomorph marginal.

As an example we present the variable set $V = \{V_1, V_2, V_3, V_4\}$ and the formal independence model $\mathcal{I}$ on V defined by the following non-trivial independence statements (and their symmetric equivalents):

$$\begin{array}{llll} (S1): & \mathcal{I}(V_1, V_2|\varnothing) & (S2): & \mathcal{I}(V_1, V_2|V_3) \\ (S3): & \mathcal{I}(V_2, V_3|\varnothing) & (S4): & \mathcal{I}(V_1, V_4|\varnothing) \\ (S5): & \mathcal{I}(V_1, V_2|V_4) & (S6): & \mathcal{I}(V_2, V_3|V_1) \\ (S7): & \mathcal{I}(V_1, V_4|V_2) & & \end{array}$$

The DAG $G_1 = (V, A)$ defined on the variables V as depicted on the left-hand side in Figure 1, is a minimal I-map for $\mathcal{I}$, since the non-trivial graphical independence statement that can be derived from the DAG correspond to the statements (S1), (S2), (S3), and (S6). It is not a P-map for $\mathcal{I}$, since the statements (S4), (S5), and (S7) are not reflected as graphical independence statements in G_1. An alternative minimal I-map is G_2, as depicted on the right-hand side in Figure 1. According [2, Lemma 5.1] there does not exist a P-map for $\mathcal{I}$ on V, although $\mathcal{I}$ satisfies the necessary conditions for DAG-isomorphism of [6].

We can, however, construct a DAG $\overline{G}$ on a superset $\overline{V}$ of V for which the corresponding graphical independence model $\mathcal{I}_{\overline{G}}$ satisfies all the independence statements (S1)–(S7). This DAG is depicted in Figure 2. It has an extra, latent, variable V_0. The graphical independence model $\mathcal{I}_{\overline{G}}$ satisfies more independence statements than (S1)–(S7), like for instance $\mathcal{I}_{\overline{G}}(V_1, V_2|V_0)$. There are, however, no new independence statements $\mathcal{I}_{\overline{G}}(X, Y|Z)$ in $\mathcal{I}_{\overline{G}}$ for subsets X, Y, $Z \subset V$, other than (S1)–(S7). All new independence statements involve the latent variable V_0 in one of the arguments. By Definition 3.1 $\mathcal{I}$ in the example above is the marginal of $\mathcal{I}_{\overline{G}}$ on V, and $\overline{G}$ is a latent P-map of $\mathcal{I}$.

For the example we have from [3] that there exists a probability distribution $\overline{P}$ on $\overline{V}$ that has $\overline{G}$ of Figure 2 as a perfect map. The structure of $\overline{G}$ implies that $\overline{P}$ factorises as:

$$\begin{aligned} &\overline{P}(v_0, v_1, v_2, v_3, v_4) = \\ &\quad p_0(v_0)\, p_1(v_1)\, p_2(v_2)\, p_3(v_3|v_0 v_1) p_4(v_4|v_0 v_2) \end{aligned}$$

for some functions $p_1, \ldots, p_4$. It can be shown that the DAG's G_1 and G_2 are minimal I-maps for the marginal distribution of $\overline{P}$ on V. G_1 corresponds to a factorisation of P

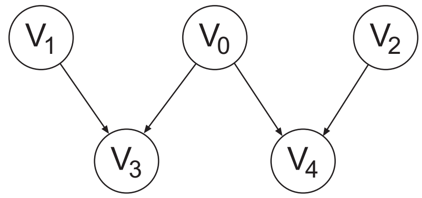

Figure 2: $\overline{G}$, a latent P-map for $\mathcal{I}$

as:

$$P(v_1, v_2, v_3, v_4) = \\ p_1(v_1)\, p_2(v_2)\, p'_3(v_3|v_1)\, p'_4(v_4|v_1 v_2 v_3) \tag{1}$$

and G_2 corresponds to a factorisation of P as:

$$P(v_1, v_2, v_3, v_4) = \\ p_1(v_1)\, p_2(v_2)\, p''_3(v_3|v_1 v_2 v_4)\, p''_4(v_4|v_2) \tag{2}$$

In the example we thus have a probability distribution P and the corresponding independence model $\mathcal{I}_P$ on V that is not DAG-isomorphic, but it is the marginal of a distribution $\overline{P}$ that corresponds to a DAG-isomorphic probabilistic independence model.

For a probability measure we can now present a refined definition of DAG-isomorph marginal based on the probabilistic notion of a marginal.

Definition 3.4 (P-DAG-isomorph marginal) *Let V be a set of variables and P a probability measure on V. We say that P is a* P-DAG-isomorph marginal*, if there exists a finite set of variables $\overline{V} = \{\overline{V}_1, \ldots, \overline{V}_{\overline{N}}\} \supseteq V$ with domains $\overline{\mathcal{V}}_i$, $i = 1, \ldots, \overline{N}$, a DAG $\overline{G} = (\overline{V}, \overline{A})$, and a probability measure $\overline{P}$ on $\overline{V}$, such that*

- *The domains of the variables V_i in V for $\overline{P}$ are the same as for P ,*
- *The marginal distribution $\overline{P}^V$ of $\overline{P}$ over V is equal to P,*
- *$\overline{G}$ is a perfect map for $\overline{P}$.*

4 Existence of a latent perfect map

Weak transitivity and chordality are necessary conditions for DAG-isomorphism. Assume that we have an independence model $\mathcal{I}$ on the set of variables V that does not satisfy either of these two conditions. For any independence model $\overline{\mathcal{I}}$ on a superset $\overline{V} \supseteq V$ the conflicting conditions remain unsatisfied, since the independence statements in $\mathcal{I}$ that violated the conditions will also be in $\overline{\mathcal{I}}$. This implies that any independence model that does not satisfy any of these two properties, is not a DAG-isomorph marginal.

For the example of the previous section, which satisfies weak transitivity and chordality, there does not exist a P-map, but we were able to construct a latent perfect map. In this section we show that a latent perfect map does not always exists, even if the independence model satisfies all the necessary conditions for a DAG-isomorph. The main result is captured in the following theorem.

Theorem 4.1 *There exists an independence model satisfying the necessary conditions for DAG-isomorphism that neither has a P-map nor a latent P-map.*

We shall prove Theorem 4.1 by showing that there is no latent perfect map for the following independence model.

Definition 4.2 *Let $V = \{B, C, D, E\}$ and let $\mathcal{I}^*$ be the independence model on V, that consists of the following three non-trivial independent statements (and their symmetric equivalents):*

$$
\begin{aligned}
(T1): &\quad \mathcal{I}^*(B, E|CD) \\
(T2): &\quad \mathcal{I}^*(C, E|\varnothing) \\
(T3): &\quad \mathcal{I}^*(C, D|B)
\end{aligned}
$$

It is a straight-forward exercise to verify that $\mathcal{I}^*$ is indeed a semi-graphoid. Application of the semi-graphoid axioms on (T1)–(T3) does not yield any new non-trivial independence statements. Moreover, $\mathcal{I}^*$ satisfies Pearl's necessary conditions for DAG-isomorphism.

We prove by contradiction that $\mathcal{I}^*$ is not a DAG-isomorph marginal. The steps in the proof are summarised in the following four lemmas.

Lemma 4.3 *Assume that $\mathcal{I}^*$, as defined in Definition 4.2, is a DAG-isomorph marginal and $\overline{G}$ is a latent P-map for $\mathcal{I}^*$, then there exists at least one path in $\overline{G}$ from C to E that is neither blocked by B nor by D.*

Proof. By contradiction: assume that there are no paths in $\overline{G}$ between C and E. C and E are then d-separated by any subset of $\overline{V}$, which contradicts, for instance, $\neg\mathcal{I}^*(C, E|BD)$.

Assume that all paths in $\overline{G}$ between C and E are blocked by B or D. Since there is at least one path in $\overline{G}$ from C to E, this again contradicts $\neg\mathcal{I}^*(C, E|BD)$. □

Lemma 4.4 *Assume that $\mathcal{I}^*$, as defined in Definition 4.2, is a DAG-isomorph marginal, $\overline{G}$ is a latent P-map for $\mathcal{I}^*$, and s is a path in $\overline{G}$ from C to E, then s has at least one converging node.*

Proof. Let s be a path from C to E. Due to $\mathcal{I}^*(C, E|\varnothing)$ s must be blocked by $\varnothing$, which implies that s has a converging node. □

Lemma 4.5 *Assume that $\mathcal{I}^*$, as defined in Definition 4.2, is a DAG-isomorph marginal, $\overline{G}$ is a latent P-map of $\mathcal{I}^*$, s a path in $\overline{G}$ from C to E that is neither blocked by B nor by D, and let F be a converging node on s, then $D \in \sigma(F)$ and $B \in \sigma(F)$. Moreover every descending path from F to D is blocked by B.*

Proof. If there exists a converging node F on s for which $B \notin \sigma(F)$ or $D \notin \sigma(F)$, then the path s would be blocked by B or D , which is in contradiction with the definition of s.

Let F be a converging node on s. Since $D \in \sigma(F)$, there exists a descending path s_1 from F to D. We now construct a new path s_2 from C to D by concatenating the subpath of s between C and F with s_1 (see Figure 4). Due to $\mathcal{I}^*(C, D|B)$ this path must be blocked by B. It cannot be blocked by B on the segment between C and F, since then also the original path s would be blocked by B. Therefore s_2 must be blocked by B on the subpath s_1. Since s_1 is descending, it is unidirectional. Therefore B must lie on s_1 and s_1 is blocked by B. □

Lemma 4.6 *Assume that $\mathcal{I}^*$, as defined in Definition 4.2, is a DAG-isomorph marginal, $\overline{G}$ is a latent P-map of $\mathcal{I}^*$, s is a path in $\overline{G}$ from C to E that is neither blocked by B nor by D. For any converging node F on s there is also a second converging node on the subpath of s between F and E.*

Proof. Let F be a converging node on s, which exists due to Lemma 4.4. From Lemma 4.5 we have that any descending path from F to D has B on it. At least one such path, say s_1, must exist, since $B \neq D$, and thus D cannot be equal to the converging node F. We now construct a path s_3 from B to E by concatenating the reverse of the part of subpath s_1 between B and F with the subpath of s between F and E (see also Figure 4).

Now s_3 is a path from B to E via F. Due to $\mathcal{I}^*(B, E|CD)$, this path s_3 must be blocked by CD. Since s_1 is descending and thus unidirectional, the first part of s_3 between B and F is unidirectional. D is not on this subpath, so it cannot be blocked by D. The second part of s_3 between F and E cannot be blocked by D, since it is part of the original path s and s is not blocked by D. In path s_3 the node F, where the two subpaths join, is not a converging node, so we conclude that s_3 cannot be blocked by D. This implies that s_3 must be blocked by C.

There are two possibilities for C to block s_3. The first possibility is that C blocks s_3 by presence on the (unidirectional) subpath s_1 between B and F. If this is the case, then we can construct a new path s_4 from C to E, by dropping from s_3 the first part between B and C. This new path s_4 consists of a unidirectional path between C and F, that has neither B nor D on it. The second part of the path, between F and E, is the segment of the original path s. Since F is not a converging node on s_4 and s is not blocked by B nor D, we conclude that s_4 is also not blocked by B nor by D. From Lemma 4.4 we conclude that s_4 must have a converging node, which can lie only between F and E. Therefore this converging node must also lie on the original path s.

The second possibility for C to block s_3 is through absence, if there is a converging node on s_3 that does not have C as a descendant. Since the first part of s_3 between B and F is unidirectional, and F is not a converging node on s_3, this converging node must lie on the segment of s_3 strictly between F and E and therefore also on s. □

Proof.(Of Theorem 4.1) Let $\mathcal{I}^*$ be as defined in Definition 4.2. Due to Lemma 4.3 we know that there is at least one path s between C and E that is not blocked by B nor by D. According to Lemma 4.4 this path s must have at least one converging node (Lemma 4.4) and due to Lemma 4.6 we can conclude that s must have an infinite number of converging nodes. Therefore $\overline{V}$ cannot be finite, and $\mathcal{I}^*$ is not a DAG-isomorph marginal. □

The next theorem shows that there is also a probabilistic independence model without a latent perfect map.

Theorem 4.7 *There exists a set of variables V and a probability distribution on V that is not a P-DAG-isomorph marginal.*

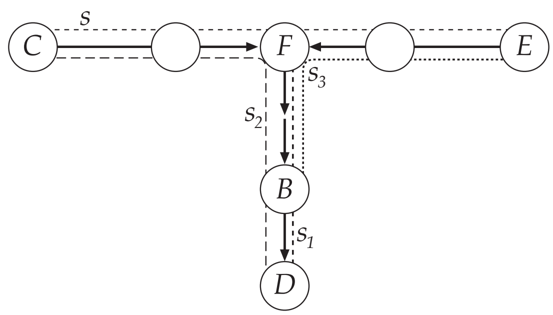

Figure 3: The paths used in the proofs of Lemmas 4.5 and 4.6

Proof. Consider the set of binary variables $V = \{B, C, D, E\}$. Define the probability measure P^* on V as follows:

B	C	D	E	$P^*(B,C,D,E)$	B	C	D	E	$P^*(B,C,D,E)$
0	0	0	0	48/1357	1	0	0	0	96/1357
0	0	0	1	48/1357	1	0	0	1	96/1357
0	0	1	0	144/1357	1	0	1	0	192/1357
0	0	1	1	48/1357	1	0	1	1	64/1357
0	1	0	0	48/1357	1	1	0	0	27/1357
0	1	0	1	96/1357	1	1	0	1	54/1357
0	1	1	0	240/1357	1	1	1	0	90/1357
0	1	1	1	48/1357	1	1	1	1	18/1357

It can be verified that the probabilistic independence model $\mathcal{I}_{P^*}$ of P^* has exactly the same independence statements as $\mathcal{I}^*$ as defined in Definition 4.2. □

5 Conclusions

In this paper we have introduced the concept of the marginal of an formal independence model. We have shown that some independence models are in fact the marginals of models that are DAG-isomorphs, while the marginals themselves are not DAG-isomorphs. We have also proved that there exist some independence models for which we need to introduce an infinite number of auxiliary variables to obtain a latent perfect map. In examples for both cases the marginal independence models satisfy the sufficient conditions of [6] for DAG-isomorphism. It is an interesting topic for future research to investigate if necessary and sufficient conditions can be established to guarantee the existence of a latent perfect map.

It is also worthwhile to investigate if existence results for latent P-maps can be established for other types of graphical model. We can show that this is not true for undirected graphs, and we plan to investigate if it is possible for chain graphs.

6 Acknowledgements

We would like to thank Richard D. Gill for suggesting this problem in the context of probabilistic graphical models.

References

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

[2] A.P. Dawid and M. Studený. Conditional Products: An Alternative Approach to Conditional Independence, in *Artificial Intelligence and Statistics 99, Proceedings of the 7th workshop*, J. Whittaker and D. Heckerman (eds), Morgan Kaufmann, pp. 32-40.

[3] D. Geiger and J. Pearl. 1990. On the Logic of Causal Models, in *Uncertainty in Artificial Intelligence 4*, R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (eds), Elsevier Science Publishers, 3–14.

[4] A.N. Kolmogorov. 1950. *Foundation of the Theory of Probability*, New York, Chelsea Publishing.

[5] J. Pearl, A. Paz. 1987. Graphoids, Graph-based Logic for Reasoning About Relevance Relations, in *Advances in Artificial Intelligence II*, B. du Boulay, D. Hogg, L. Steele (eds), 357–363.

[6] J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems - Networks of Plausible Inference*, Morgan Kaufman.

[7] M. Studený. 1989. Multi-information and the Problem of Characterization of Conditional Independence Relations, *Problems of Control and Information Theory*, 18 (1), 3–16.

[8] M. Studený. 1992. Conditional Independence Relations Have No Finite Complete Characterization, in *Proc. of the 11th Prague conf. on Information Theory, Statistical Decision Functions and Random Processes*, S. Kubik, J.A. Visek (eds), 377–396.

[9] M. Studený. 1998. Complexity of structural models. In: *Proceedings of the Joint Session of the 6th Prague Conference on Asymptotic Statistics and the 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, vol. II, Prague, pp. 521–528.