Informational Independence: Models and Normal Forms

L.C. van der Gaag & J.-J.Ch. Meyer
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

Abstract

The concept of informational independence plays a key role in most knowledge-based systems. J. Pearl and his co-researchers have analysed the basic properties of the concept and have formulated an *axiomatic system for informational independence*. This axiomatic system focuses on independences among mutually disjoint sets of variables. We show that in the context of probabilistic independence a focus on disjoint sets of variables can hide various interesting properties. To capture these properties, we enhance Pearl's axiomatic system with two additional axioms. We investigate the set of *models* of the thus enhanced system and show that it provides a better characterisation of the concept of probabilistic independence than Pearl's system does. In addition, we observe that both Pearl's axiomatic system and our enhanced system offer inference rules for deriving new independences from an initial set of independence statements and as such allow for a *normal form* for representing independence. We address the normal forms ensuing from the two axiomatic systems for informational independence.

1 Introduction

The concept of *informational independence* pervades most knowledge-based systems. The concept is used for example for demarcating a system's scope. But more importantly, the concept of informational independence is essential for reasoning purposes: it is making effective use of knowledge about independences that renders knowledge-based systems capable of dealing with the computational complexity of their problem-solving tasks. Although informational independence plays a key role in knowledge-based systems, it generally is defined in different terms in different contexts. For example, in systems built on probability theory, informational independence is identified with statistical independence among sets of variables; in constraint-satisfaction systems, informational independence is defined in terms of constraints — two variables are said to be independent if restricting the domain of one variable leaves the other one's domain unaltered. Despite these definitional differences, the various perspectives on informational independence share the same basic properties.

The basic properties of informational independence have been identified and taken to constitute an *axiomatic system for informational independence* by J. Pearl and his co-researchers [Pearl, 1988, Pearl *et al.*, 1990]. There are many advantages to an axiomatic system for informational independence. Such a system for example offers inference rules for computing new independence statements from an initial set of statements. As such, it allows for a *normal form* for independence and, hence, provides for a concise representation of a set of independences. An axiomatic system may further be used for verifying whether a new statement logically follows from a set of independence statements and for studying inconsistency among

independences. These advantages fall in with the advantages of axiomatisation in general. The impact of Pearl's axiomatic system for informational independence is most clearly demonstrated in the context of probability theory, as it underlies the formalism of *Bayesian belief networks* for reasoning with uncertainty in knowledge-based systems [Pearl, 1988].

Pearl's axiomatic system for informational independence focuses on independences among *mutually disjoint* sets of variables. In the context of probability theory, however, the concept of informational independence is not restricted to mutually disjoint sets of variables. In this paper, we examine the consequences of Pearl's restricted focus and show that focusing on disjoint sets of variables can hide various interesting properties pertaining to the functional dependences holding in a joint probability distribution. To capture these properties, we identify two new axioms to be included in Pearl's axiomatic system. A comparison of the sets of models of Pearl's system and our thus enhanced axiomatic system shows that the enhanced system provides the better characterisation of the concept of probabilistic independence in the sense that it has fewer models that are not embedded in any joint probability distribution.

Since the introduction of Pearl's axiomatic system for informational independence, several new axioms have been identified. In fact, it has been shown that a finite axiomatisation of the concept of probabilistic independence does not exist [Studený, 1992]. The importance of our enhanced axiomatic system in view of this result arises from its allowing for a new normal form for informational independence. This normal form elucidates the necessity of explicitly representing the functional dependences holding in a joint probability distribution, for example in addition to its representation in a Bayesian belief network.

The paper is structured as follows. We review Pearl's axiomatic system for informational independence in Section 2 and discuss the extension Pearl has proposed for his system in Section 3. The two additional axioms that we have identified from an analysis of the restricted focus on disjoint sets of variables, are described in Section 4. In Section 5 we address the normal forms ensuing from Pearl's axiomatic system for informational independence and our enhanced system, respectively. The paper is rounded off with some conclusions in Section 6.

2 Independence Revisited

The concept of informational independence has been studied in various contexts. Especially in the context of *probability theory* has independence been a subject of extensive studies, see for example [Dawid, 1979, Lauritzen, 1982]. The main objective of the early statistical studies was to identify and express in algebraic form independence relations embedded in joint probability distributions to allow for comparison and classification. J. Pearl and his co-researchers were among the first to formalise properties of independence relations in an *axiomatic system* and to develop a logic for *informational independence* [Pearl & Paz, 1985, Pearl & Verma, 1987, Geiger & Pearl, 1988]. In this section, we review Pearl's axiomatic system.

2.1 Pearl's Axiomatic System for Informational Independence

In the context of probability theory, the concept of independence is generally introduced in terms of numerical quantities: the independence relation of a joint probability distribution is taken to be implicitly embedded in the probabilities involved. A definition of independence in terms of numbers suggests that, in order to determine whether two sets of variables are

(conditionally) independent, several conditional probabilities have to be computed and several equalities have to be tested; moreover, such a definition suggests that for determining independence a joint probability distribution on the variables discerned has to be explicitly available. In contrast, humans tend to be able to state directly, often with conviction and consistency, whether or not two sets of variables are independent. Such statements of independence typically are issued qualitatively, without any reference to numerical manipulation of exact probabilities. Based on these observations, Pearl argues that the concept of independence is far more basic to human reasoning than its numerical definition suggests and that in fact the definition of independence in terms of probabilities may be looked upon as a *quantitative* way of capturing the basic concept which is *qualitative* in nature [Pearl, 1988]. Pearl's aim in designing an axiomatic system now is to provide an explicit qualitative definition of independence.

We begin our review of Pearl's axiomatic system for informational independence by introducing some notational convention.

Definition 2.1 *Let V be a finite set of (discrete) variables and let* $\Pr$ *be a joint probability distribution on V. Then, the* independence relation $I_{\Pr} \subseteq 2^V \times 2^V \times 2^V$ *of* $\Pr$ *is defined by* $(X, Z, Y) \in I_{\Pr}$ *if and only if* $\Pr(X = x \mid Y = y \wedge Z = z) = \Pr(X = x \mid Z = z)$ *for all value assignments x, y, z to the sets of variables $X, Y, Z \subseteq V$, respectively.*

Note that in the previous definition we have expressed the concept of (conditional) independence from probability theory in the *Bayesian formalism*, thereby closely following Pearl [Pearl, 1988]. From now on, we will write $I_{\Pr}(X, Z, Y)$ to denote $(X, Z, Y) \in I_{\Pr}$ and $\neg I_{\Pr}(X, Z, Y)$ to denote $(X, Z, Y) \notin I_{\Pr}$, for short. A statement $I_{\Pr}(X, Z, Y)$ of a joint probability distribution Pr's independence relation $I_{\Pr}$ is termed an *independence statement*. In qualitative terms, an independence statement $I_{\Pr}(X, Z, Y)$ expresses that in the context of information about Z information about Y is irrelevant with respect to X. In the following, we will use $\mathcal{P}$ to denote the set of all independence relations that are embedded in a joint probability distribution on a set of variables V. We will want to single out independence relations that are embedded in a *strictly positive* probability distribution, since these relations exhibit some interesting properties that do not hold for independence relations in general. The set of all such independence relations will be denoted as $\mathcal{P}^+$. Note that $\mathcal{P}^+ \subset \mathcal{P}$.

In designing his axiomatic system for informational independence, Pearl builds on a set of properties that are satisfied by any joint probability distribution's independence relation; Theorem 2.2 reviews these properties.

Theorem 2.2 *Let V be a finite set of variables. Let* $\Pr$ *be a joint probability distribution on V and let $I_{\Pr}$ be its independence relation as defined in Definition 2.1. Then, $I_{\Pr}$ satisfies the properties*

- $I_{\Pr}(X, Z, Y) \rightarrow I_{\Pr}(Y, Z, X)$;
- $I_{\Pr}(X, Z, Y \cup W) \rightarrow I_{\Pr}(X, Z, Y) \wedge I_{\Pr}(X, Z, W)$;
- $I_{\Pr}(X, Z, Y \cup W) \rightarrow I_{\Pr}(X, Z \cup W, Y)$;
- $I_{\Pr}(X, Z, Y) \wedge I_{\Pr}(X, Z \cup Y, W) \rightarrow I_{\Pr}(X, Z, Y \cup W)$;

for all mutually disjoint sets of variables $X, Y, Z, W \subseteq V$. If the distribution $\Pr$ *is strictly positive, then $I_{\Pr}$ satisfies the additional property*

- $I_{\Pr}(X, Z \cup W, Y) \wedge I_{\Pr}(X, Z \cup Y, W) \rightarrow I_{\Pr}(X, Z, Y \cup W)$;

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

The properties stated in Theorem 2.2 are easily verified from the basic axioms of probability theory. We would like to note that we have stated the properties in the theorem to hold for *mutually disjoint* sets of variables only [Pearl, 1988]. These properties, however, also hold for overlapping sets of variables; we will return to this observation in the following section.

Pearl now takes the properties stated in Theorem 2.2 as *axioms* for the qualitative concept of informational independence [Pearl, 1988].

Definition 2.3 *The* axiomatic system for informational independence $\mathcal{A}$ *is composed of the four axioms*

- $I(X, Z, Y) \rightarrow I(Y, Z, X)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z, W)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y)$;
- $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ *on a finite set of variables* V*, for all mutually disjoint sets of variables* $X, Y, Z, W \subseteq V$*; the set of all models of the system* $\mathcal{A}$ *will be denoted as* $[\mathcal{A}]$*. The axiomatic system* $\mathcal{A}^+$ *comprises the four axioms from the system* $\mathcal{A}$ *and the additional axiom*

- $I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ *on* V*, for all mutually disjoint sets of variables* $X, Y, Z, W \subseteq V$*; the set of all models of* $\mathcal{A}^+$ *will be denoted as* $[\mathcal{A}^+]$.

The axioms described in Definition 2.3 have been proven logically independent [Pearl, 1988].

The axioms of the axiomatic system for informational independence $\mathcal{A}$ with each other convey the idea that learning irrelevant information does not alter the independences among the variables discerned. We consider the qualitative meanings of the various axioms separately. The axiom

$$I(X, Z, Y) \rightarrow I(Y, Z, X)$$

states that if information about Y is deemed irrelevant with respect to X in the context of information about Z, then information about X must be irrelevant with respect to Y in this context; this axiom is called the *symmetry axiom*. The axiom

$$I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z, W)$$

asserts that if information about both Y and W is judged irrelevant with respect to X, then both information about Y and information about W must be irrelevant with respect to X separately; this axiom is known as the *decomposition axiom*. We would like to note that the decomposition axiom may be reformulated as $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y)$; we have chosen to use Pearl's original formulation because it conveys the idea of decomposition more clearly. The axiom

$$I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y)$$

states that learning information about W that is known to be irrelevant with respect to X cannot help irrelevant information about Y to become relevant with respect to X; this axiom is known as the *weak union axiom*. The axiom

$$I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$$

states that if we judge information about W to be irrelevant with respect to X after learning some irrelevant information about Y, then the information about W must have been irrelevant with respect to X before we learned Y; this axiom is known as the *contraction axiom*. Note that the contraction axiom can be reformulated as $I(X, Z, Y) \rightarrow (I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W))$. From this reformulation, it is seen that the axiom can be looked upon as a *conditional* reverse of the weak union axiom. To conclude, we consider the axiom

$$I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$$

This axiom states that if, in the context of information about Z, learning information about W renders information about Y irrelevant with respect to X and learning Y renders W irrelevant with respect to X, then the information about both Y and W must be irrelevant with respect to X given Z; this axiom is known as the *intersection axiom*.

The axiomatic system $\mathcal{A}$ introduced in Definition 2.3 will be referred to as Pearl's *restricted* axiomatic system for informational independence. Any model in the set of models $[\mathcal{A}]$ of the system $\mathcal{A}$ is termed a *semi-graphoid independence relation*; any model in $[\mathcal{A}^+]$ is termed a *graphoid independence relation*. The terms *graphoid* and *semi-graphoid* refer to the representation of independence relations in graphical structures [Pearl & Paz, 1985, Pearl, 1988].

Pearl's restricted axiomatic system for informational independence $\mathcal{A}$ serves to capture (at least) all independence relations that are embedded in a joint probability distribution; this property is stated more formally in the following lemma.

Lemma 2.4 *Let $\mathcal{P}$ be the set of all independence relations that are embedded in a joint probability distribution and let $\mathcal{P}^+$ be the set of independence relations that are embedded in a strictly positive distribution. Furthermore, let $[\mathcal{A}]$ be the set of models of the axiomatic system $\mathcal{A}$ and let $[\mathcal{A}^+]$ be the set of models of the system $\mathcal{A}^+$ as defined in Definition 2.3. Then,*

- $\mathcal{P} \subseteq [\mathcal{A}]$;
- $\mathcal{P}^+ \subseteq [\mathcal{A}^+]$.

Proof. We will only prove the first property stated in the lemma; the proof of the second property is analogous. Let $I \in \mathcal{P}$ be an independence relation. By definition of $\mathcal{P}$, we have that there exists a joint probability distribution Pr such that I is the independence relation of Pr. From Theorem 2.2 and Definition 2.3 we have that I satisfies all four axioms from the axiomatic system $\mathcal{A}$. So, $I \in [\mathcal{A}]$. We conclude that $\mathcal{P} \subseteq [\mathcal{A}]$. $\square$

Although any probability distribution's independence relation is a model of the axiomatic system $\mathcal{A}$, the reverse property does *not* hold, that is, there exist models in the set of models of $\mathcal{A}$ for which there do not exist joint probability distributions embedding them. This observation is due to M. Studený who has been the first to identify an additional, logically independent axiom for semi-graphoid independence relations [Studený, 1989]. From Studený's observation we have the following lemma.

Lemma 2.5 *Let $\mathcal{P}$ and $\mathcal{P}^+$ be as before. Furthermore, let $[\mathcal{A}]$ and $[\mathcal{A}^+]$ be the sets of models of the axiomatic systems $\mathcal{A}$ and $\mathcal{A}^+$, respectively. Then,*

- $[\mathcal{A}] \not\subseteq \mathcal{P}$;
- $[\mathcal{A}^+] \not\subseteq \mathcal{P}^+$.

We would like to note that it has been shown that a finite axiomatisation of the concept of probabilistic independence does not exist, that is, there does not exist a finite set of axioms $\mathcal{X}$ such that $[\mathcal{X}] = \mathcal{P}$ [Studený, 1992].

2.2 Properties of Independence Relations

Using Pearl's axiomatic system for informational independence, we derive some convenient properties of (semi-graphoid and graphoid) independence relations. The following lemma shows that the symmetry and contraction axioms from the axiomatic system $\mathcal{A}$ are easily generalised to bi-implications.

Lemma 2.6 *Let $[\mathcal{A}]$ be the set of models of the axiomatic system for informational independence $\mathcal{A}$. Let $I \in [\mathcal{A}]$ be an independence relation on a finite set of variables V. Then,*

- $I(X, Z, Y) \leftrightarrow I(Y, Z, X)$;
- $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \leftrightarrow I(X, Z, Y \cup W)$;

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

Proof. We begin our proof by observing that since $I \in [\mathcal{A}]$, it obeys the four axioms from the axiomatic system $\mathcal{A}$. The first property stated in the lemma now follows directly from the symmetry axiom. For the second property, we observe that $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$ coincides with the contraction axiom and therefore trivially holds for the independence relation I. We will now prove that $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z \cup Y, W)$. We have

$$\begin{aligned} I(X, Z, Y \cup W) &\Rightarrow I(X, Z, Y) \wedge I(X, Z, W) \Rightarrow \\ &\Rightarrow I(X, Z, Y) \end{aligned}$$

by the decomposition axiom. In addition, we have

$$I(X, Z, Y \cup W) \Rightarrow I(X, Z \cup Y, W)$$

by weak union. The property stated in the lemma now follows directly. □

For graphoid independence relations we further have that the intersection axiom from the axiomatic system $\mathcal{A}^+$ can be generalised to a bi-implication.

Lemma 2.7 *Let $[\mathcal{A}^+]$ be the set of models of the axiomatic system for informational independence $\mathcal{A}^+$. Let $I \in [\mathcal{A}^+]$ be an independence relation on a finite set of variables V. Then,*

$$I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \leftrightarrow I(X, Z, Y \cup W)$$

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

Proof. We will only prove that $I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W)$; the reverse property coincides with the intersection axiom from the axiomatic system $\mathcal{A}^+$ and therefore trivially holds for the independence relation I. We have

$$I(X, Z, Y \cup W) \Rightarrow I(X, Z \cup W, Y)$$

and

$$I(X, Z, Y \cup W) \Rightarrow I(X, Z \cup Y, W)$$

by the weak union axiom. The property stated in the lemma now follows directly. □

3 Pearl's Extended Axiomatic System

Pearl's restricted axiomatic system for informational independence involves axioms that are stated to apply to *mutually disjoint* sets of variables only. The probabilistic concept of independence, however, is not restricted to mutually disjoint sets of variables: a joint probability distribution's independence relation typically includes independence statements involving overlapping sets of variables. To fully capture the basic concept of independence, an axiomatic system should therefore provide axioms, not just for mutually disjoint sets, but for overlapping sets of variables as well. In this section, we review the extension Pearl has proposed for this purpose to his axiomatic system.

In Section 2 we have already mentioned that the properties from Theorem 2.2, although stated to hold for mutually disjoint sets of variables only, hold for overlapping sets as well. The axioms from Pearl's restricted axiomatic system for informational independence therefore are generalised straightforwardly to apply to overlapping sets of variables. In addition to generalising the symmetry, decomposition, weak union, contraction, and intersection axioms, Pearl proposes including an extra axiom in his system. This additional axiom has its motivational foundation in the property

$$I_{\mathrm{Pr}}(X, Z, Y) \leftrightarrow I_{\mathrm{Pr}}(X - Z, Z, Y - Z)$$

for all sets of variables $X, Y, Z \subseteq V$, first identified by A.P. Dawid to hold for any joint probability distribution Pr's independence relation I_{Pr} [Dawid, 1979]. To capture Dawid's property for overlapping sets, Pearl introduces the axiom $I(X, Z, Z)$ for all sets of variables $X, Z \subseteq V$ [Pearl, 1988 (pp. 84 – 85)]. This axiom asserts that, once information about Z is known, learning information about Z becomes irrelevant with respect to any set of variables X. The validity of the axiom in the context of probabilistic independence is easily verified; moreover, it is readily seen that the new axiom is logically independent from the generalised symmetry, decomposition, weak union, contraction, and intersection axioms. The axiom will be termed the *overlap axiom*.

Pearl's extended axiomatic system for informational independence $\mathcal{E}$ is summarised in the following definition.

Definition 3.1 *The* axiomatic system for informational independence $\mathcal{E}$ *is composed of the five axioms*

- $I(X, Z, Y) \rightarrow I(Y, Z, X)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z, W)$;

- $I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y)$;
- $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;
- $I(X, Z, Z)$;

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ *on a finite set of variables* V*, for* all *sets of variables* $X, Y, Z, W \subseteq V$*; the set of all models of the system* $\mathcal{E}$ *will be denoted as* $[\mathcal{E}]$*. The axiomatic system* $\mathcal{E}^+$ *comprises the five axioms from the system* $\mathcal{E}$ *and the additional axiom*

- $I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ *on* V*, for* all *sets of variables* $X, Y, Z, W \subseteq V$*; the set of all models of* $\mathcal{E}^+$ *will be denoted as* $[\mathcal{E}^+]$*.*

Any model in the set of models $[\mathcal{E}]$ of the axiomatic system $\mathcal{E}$ will be termed an *extended semi-graphoid independence relation*; a model in $[\mathcal{E}^+]$ will be called an *extended graphoid independence relation.*

The set of all models of Pearl's extended axiomatic system for informational independence $\mathcal{E}$ is a proper subset of the set of models of the restricted axiomatic system $\mathcal{A}$, yet includes all independence relations that are embedded in a joint probability distribution. This property is stated more formally in the following lemma.

Lemma 3.2 *Let* $\mathcal{P}$ *and* $\mathcal{P}^+$ *be as before. Furthermore, let* $[\mathcal{A}]$ *and* $[\mathcal{A}^+]$ *be the sets of models of the axiomatic systems* $\mathcal{A}$ *and* $\mathcal{A}^+$*, respectively. Let* $[\mathcal{E}]$ *be the set of models of the axiomatic system for informational independence* $\mathcal{E}$ *and let* $[\mathcal{E}^+]$ *be the set of models of the system* $\mathcal{E}^+$ *as defined in Definition 3.1. Then,*

- $\mathcal{P} \subset [\mathcal{E}] \subset [\mathcal{A}]$;
- $\mathcal{P}^+ \subset [\mathcal{E}^+] \subset [\mathcal{A}^+]$.

To conclude our review of Pearl's extension to the restricted system, the following lemma confirms that Dawid's property for overlapping sets of variables can indeed be derived from the extended axiomatic system for informational independence $\mathcal{E}$.

Lemma 3.3 *Let* $[\mathcal{E}]$ *be the set of models of the axiomatic system* $\mathcal{E}$*. Let* $I \in [\mathcal{E}]$ *be an independence relation on a finite set of variables* V*. Then,*

$$I(X, Z, Y) \leftrightarrow I(X - Z, Z, Y - Z)$$

for all sets of variables $X, Y, Z \subseteq V$*.*

Proof. We begin our proof by observing that since $I \in [\mathcal{E}]$, it obeys the five axioms from the axiomatic system $\mathcal{E}$. We first prove the property $I(X, Z, Y) \rightarrow I(X - Z, Z, Y - Z)$. We have that

$$\begin{aligned} I(X, Z, Y) \quad & \Rightarrow I(X, Z, Y - Z) \Rightarrow \\ & \Rightarrow I(Y - Z, Z, X) \Rightarrow \\ & \Rightarrow I(Y - Z, Z, X - Z) \Rightarrow \\ & \Rightarrow I(X - Z, Z, Y - Z) \end{aligned}$$

by the (generalised) decomposition and symmetry axioms.

We now prove the property $I(X - Z, Z, Y - Z) \rightarrow I(X, Z, Y)$. For this purpose, we assume that $I(X - Z, Z, Y - Z)$ for some subsets of variables $X, Y, Z \subseteq V$. From the overlap axiom, we furthermore have that $I(X - Z, Z, Z)$. From this independence statement we find $I(X - Z, Z, Z \cap Y)$ by the decomposition axiom. In conjunction with our assumption $I(X - Z, Z, Y - Z)$, we now find

$$\begin{aligned} I(X - Z, Z, Z \cap Y) \wedge I(X - Z, Z, Y - Z) \quad & \Rightarrow I(X - Z, Z, Y) \Rightarrow \\ & \Rightarrow I(Y, Z, X - Z) \end{aligned}$$

by the contraction and symmetry axioms. Now observe that the relation I includes the independence statement $I(Y, Z, Z)$. From this statement, we have $I(Y, Z, Z \cap X)$ by the decomposition axiom. In conjunction with the independence statement $I(Y, Z, X - Z)$, we find

$$\begin{aligned} I(Y, Z, Z \cap X) \wedge I(Y, Z, X - Z) \quad & \Rightarrow I(Y, Z, X) \Rightarrow \\ & \Rightarrow I(X, Z, Y) \end{aligned}$$

by the contraction and symmetry axioms. We conclude that $I(X - Z, Z, Y - Z) \rightarrow I(X, Z, Y)$. □

4 Enhancing Pearl's Axiomatic System

Pearl's extended axiomatic system for informational independence includes axioms that address independences among overlapping sets of variables. As such, it is a better characterisation of the concept of probabilistic independence than the restricted axiomatic system in the sense that it allows fewer independence relations that are not embedded in any joint probability distribution. Independences among overlapping sets of variables, however, induce various interesting properties that are not yet captured by Pearl's extended axiomatic system. In this section, we identify several properties pertaining to functional dependences among variables. To capture these properties, we include two additional axioms in Pearl's extended axiomatic system for informational independence.

4.1 Strictly Positive Probability Distributions

In designing his axiomatic system for informational independence, Pearl has singled out strictly positive joint probability distributions [Pearl, 1988]. This special attention for strictly positive distributions has resulted in the concept of a graphoid independence relation as a special type of semi-graphoid independence relation. In this section, we reconsider the independence relation of a strictly positive joint probability distribution and examine its independence statements that involve overlapping sets of variables.

Strictly positive joint probability distributions are well-known for the property that they do not embed any functional dependences among their variables: in a strictly positive distribution there are no variables whose value is completely determined by some other variables' values. This lack of functional dependences is reflected in the independence relations of strictly positive distributions.

Proposition 4.1 *Let V be a finite set of variables. Let* Pr *be a strictly positive joint probability distribution on V and let $I_{\Pr}$ be its independence relation. Then,*

$$\neg I_{\Pr}(\{V_i\}, Z, \{V_i\})$$

for each variable $V_i \in V$ and all sets of variables $Z \subseteq V - \{V_i\}$.

Proof. From the conditions of the proposition, we have that the probability distribution Pr is strictly positive, that is, we have that $\Pr(X = x) > 0$ for all (non-empty) sets of variables $X \subseteq V$ and all value assignments x to X. We now prove the property stated in the proposition by contradiction. Suppose that $I_{\Pr}(\{V_i\}, Z, \{V_i\})$ for some variable $V_i \in V$ and some set of variables $Z \subseteq V$ with $V_i \notin Z$. Then, by definition we have that

$$\Pr(V_i = v_i \mid V_i = v_i \wedge Z = z) = \Pr(V_i = v_i \mid Z = z)$$

for all value assignments v_i, z to V_i, Z, respectively. From the property that Pr is strictly positive, we have that $\Pr(V_i = v_i \wedge Z = z) > 0$ and, therefore, that

$$\Pr(V_i = v_i \mid V_i = v_i \wedge Z = z) = \frac{\Pr(V_i = v_i \wedge Z = z)}{\Pr(V_i = v_i \wedge Z = z)} = 1$$

for all value assignments v_i, z to V_i, Z, respectively. Consequently, we have that $\Pr(V_i = v_i \mid Z = z) = 1$ for all assignments v_i, z. But then, we have by marginalisation that

$$\Pr(V_i = v_i' \mid Z = z) + \Pr(V_i = v_i'' \mid Z = z) = 2 \leq 1$$

for any value assignment z to Z and any two *different* value assignments v_i' and v_i'' to V_i. From the contradiction, we conclude that $\neg I_{\Pr}(\{V_i\}, Z, \{V_i\})$. $\square$

The property

$$\neg I_{\Pr}(\{V_i\}, Z, \{V_i\})$$

for each variable $V_i \in V$ and all sets of variables $Z \subseteq V - \{V_i\}$, stated in the previous proposition for the independence relation $I_{\Pr}$ of a strictly positive joint probability distribution Pr, expresses that in this distribution Pr information about a variable V_i *cannot* be irrelevant with respect to V_i itself *as long as no information about V_i is available as yet.* This property conveys that indeed there do not exist any functional dependences among the variables of the distribution. Note that, if information about V_i *is* available, that is, if $V_i \in Z$, we have that $I_{\Pr}(\{V_i\}, Z, \{V_i\})$.

Pearl's extended axiomatic system for informational independence $\mathcal{E}^+$ has been designed to capture independence relations of strictly positive joint probability distributions [Pearl, 1988]. Unfortunately, the axiomatic system $\mathcal{E}^+$ does not reflect the property stated in Proposition 4.1. An independence relation I in the set of models of $\mathcal{E}^+$ may very well include statements of the form $I(\{V_i\}, Z, \{V_i\})$ for some $V_i \in V$ and some $Z \subseteq V - \{V_i\}$; we will return to this observation shortly. To exclude such statements from models of informational independence, we add to Pearl's extended axiomatic system $\mathcal{E}^+$ the axiom $\neg I(\{V_i\}, Z, \{V_i\})$ for each variable $V_i \in V$ and all sets of variables $Z \subseteq V - \{V_i\}$. The axiom will be termed the *no-functions axiom*. Note that, in contrast with all other axioms in Pearl's extended axiomatic system, the no-functions axiom is *negative* in the sense that it explicitly states which

independence statements do *not* hold. In fact, the axiom expresses a dependence and may be looked upon as providing for a *dependence closure*.

The following definition summarises the enhanced axiomatic system for informational independence $\mathcal{F}^+$.

Definition 4.2 *Let $\mathcal{E}^+$ be the axiomatic system for informational independence as defined in Definition 3.1. The* axiomatic system for informational independence *$\mathcal{F}^+$ is composed of the six axioms from $\mathcal{E}^+$ and the additional axiom*

$$\neg I(\{V_i\}, Z, \{V_i\})$$

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ on a finite set of variables V, for each variable $V_i \in V$ and all sets of variables $Z \subseteq V - \{V_i\}$; the set of all models of the system $\mathcal{F}^+$ will be denoted as $[\mathcal{F}^+]$.

The enhanced axiomatic system for informational independence $\mathcal{F}^+$ allows fewer models of independence than the axiomatic system $\mathcal{E}^+$, yet still serves to capture all independence relations that are embedded in a strictly positive joint probability distribution. This property is stated more formally in the following lemma.

Lemma 4.3 *Let $\mathcal{P}^+$ be as before. Furthermore, let $[\mathcal{E}^+]$ be the set of models of the axiomatic system $\mathcal{E}^+$. Let $[\mathcal{F}^+]$ be the set of all models of the axiomatic system for informational independence $\mathcal{F}^+$ as defined in Definition 4.2. Then,*

$$\mathcal{P}^+ \subset [\mathcal{F}^+] \subset [\mathcal{E}^+]$$

Proof. The property $\mathcal{P}^+ \subset [\mathcal{F}^+]$ follows from Proposition 4.1 and our observations in Section 2. We now prove that $[\mathcal{F}^+] \subset [\mathcal{E}^+]$. From Definition 4.2, we have that $[\mathcal{F}^+] \subseteq [\mathcal{E}^+]$. To show that $[\mathcal{E}^+] \not\subseteq [\mathcal{F}^+]$, it suffices to give an example independence relation that is included in the set of models $[\mathcal{E}^+]$ and not in $[\mathcal{F}^+]$. For this purpose, we consider the independence relation $I = 2^V \times 2^V \times 2^V$ on some set of variables V. It will be evident that for this independence relation I we have that $I \in [\mathcal{E}^+]$. We further observe that the relation I includes for example the statement $I(\{V_i\}, \varnothing, \{V_i\})$, for any variable $V_i \in V$. From this observation, we have that $I \notin [\mathcal{F}^+]$. We conclude that $[\mathcal{F}^+] \subset [\mathcal{E}^+]$. $\square$

From the no-functions axiom that we have included in the axiomatic system $\mathcal{F}^+$, we have that an independence relation I in the set of models of $\mathcal{F}^+$ does not comprise any statement of the form $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$. In fact, the independence relation does not even include statements of the form $I(X, Z, Y)$ with $X \cap Y \not\subseteq Z$. This property is stated more formally in the following lemma.

Lemma 4.4 *Let $[\mathcal{F}^+]$ be the set of models of the axiomatic system $\mathcal{F}^+$. Let $I \in [\mathcal{F}^+]$ be an independence relation on a finite set of variables V. Then,*

$$\neg I(X, Z, Y)$$

for all sets of variables $X, Y, Z \subseteq V$ with $X \cap Y \not\subseteq Z$.

Proof. From $I \in [\mathcal{F}^+]$, we have that $\neg I(\{V_i\}, Z, \{V_i\})$ for every variable $V_i \in V$ and all sets of variables $Z \subseteq V - \{V_i\}$; furthermore, we have that the independence relation I obeys the axioms from Pearl's extended axiomatic system $\mathcal{E}^+$. We now prove the property stated in the lemma by contradiction. Suppose that $I(X, Z, Y)$ for some sets of variables $X, Y, Z \subseteq V$ with $X \cap Y \not\subseteq Z$. From $X \cap Y \not\subseteq Z$ we have that there exists a variable $V_j \in V$ such that $V_j \in X \cap Y$ and $V_j \notin Z$. For this variable V_j, we find that

$$\begin{aligned} I(X, Z, Y) \quad & \Rightarrow I(X, Z, \{V_j\}) \Rightarrow \\ & \Rightarrow I(\{V_j\}, Z, X) \Rightarrow \\ & \Rightarrow I(\{V_j\}, Z, \{V_j\}) \end{aligned}$$

by the decomposition and symmetry axioms. For the variable V_j, we therefore have both $I(\{V_j\}, Z, \{V_j\})$ and $\neg I(\{V_j\}, Z, \{V_j\})$. From the contradiction, we conclude the property stated in the lemma. □

From Lemma 4.4 we have the following corollary.

Corollary 4.5 *Let $[\mathcal{F}^+]$ be the set of models of the axiomatic system $\mathcal{F}^+$. Let $I \in [\mathcal{F}^+]$ be an independence relation on a finite set of variables V. If $I(X, Z, Y)$, then $X \cap Y \subseteq Z$, for all sets of variables $X, Y, Z \subseteq V$.*

4.2 General Probability Distributions

While strictly positive joint probability distributions do not embed any functional dependences among their variables, probability distributions in general may very well do so: a joint probability distribution may involve one or more deterministic variables whose value is completely determined by some other variables' values. A well-known example of a joint probability distribution embedding a functional dependence from the literature on Bayesian belief networks is the probability distribution of the *Visit-to-Asia* example first introduced by S.L. Lauritzen and D.J. Spiegelhalter [Lauritzen & Spiegelhalter, 1988]. In this section, we reconsider independence relations of joint probability distributions that may embed functional dependences; in doing so, we once more examine their independence statements that involve overlapping sets of variables.

In the foregoing, we have seen that *lack* of functional dependences in a joint probability distribution gives rise to the property $\neg I(\{V_i\}, Z, \{V_i\})$ for each variable $V_i \in V$ and all sets of variables $Z \subseteq V - \{V_i\}$. In the presence of functional dependences, this property no longer holds. The functional dependences are reflected in the independence relation of the distribution at hand; the independence relation of a joint probability distribution that embeds one or more functional dependences typically includes statements of the form $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$. Note that an independence statement $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$ expresses that in the context of information about Z, information about V_i is irrelevant with respect to V_i itself, *even if no information about V_i is available as yet*. From a qualitative point of view, independence statements of this form may seem counterintuitive: intuitively, no variable can be independent from itself. Recall, however, that we build on the probabilistic concept of independence which is defined in terms of numerical quantities. The statement $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$ reflects that the variable V_i's value is *completely* determined by the available information about Z: once the value of V_i is determined by the information about Z, then the marginal, updated probability distribution on V_i is a degenerate $0 - 1$

distribution and learning the true value of V_i cannot alter this distribution, that is, $\Pr(V_i = v_i \mid V_i = v_i \wedge Z = z) = \Pr(V_i = v_i \mid Z = z)$ for all value assignments v_i, z to V_i, Z, respectively.

The presence of independence statements of the form $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$, or more in general, statements of the form $I(X, Z, X)$ with $X \not\subseteq Z$, in a joint probability distribution's independence relation gives rise to the property stated in the following proposition.

Proposition 4.6 *Let V be a finite set of variables. Let $\Pr$ be a joint probability distribution on V and let $I_{\Pr}$ be its independence relation. Then,*

$$I_{\Pr}(X, Z, X) \rightarrow I_{\Pr}(X, Z, Y)$$

for all sets of variables $X, Y, Z \subseteq V$.

Proof. To prove the property stated in the proposition, we assume that $I_{\Pr}(X, Z, X)$ for some sets of variables $X, Z \subseteq V$. Using this independence statement, we now first show that $I_{\Pr}(X \cup Y, Z, X)$ for all sets of variables $Y \subseteq V$. To this end, we consider the probabilities $\Pr(X = x \wedge Y = y \mid X = x \wedge Z = z)$ for all value assignments x, y, z to the sets of variables X, Y, Z, respectively. By the chain rule, we find

$$\begin{aligned}
&\Pr(X = x \wedge Y = y \mid X = x \wedge Z = z) = \\
&\quad = \Pr(Y = y \mid X = x \wedge X = x \wedge Z = z) \cdot \Pr(X = x \mid X = x \wedge Z = z) \\
&\quad = \Pr(Y = y \mid X = x \wedge Z = z) \cdot \Pr(X = x \mid X = x \wedge Z = z)
\end{aligned}$$

for all value assignments x, y, z to X, Y, Z, respectively. From our assumption $I_{\Pr}(X, Z, X)$, we have by definition that $\Pr(X = x \mid X = x \wedge Z = z) = \Pr(X = x \mid Z = z)$ for all value assignments x, z to X, Z. Using this property, we find

$$\begin{aligned}
&\Pr(X = x \wedge Y = y \mid X = x \wedge Z = z) = \\
&\quad = \Pr(Y = y \mid X = x \wedge Z = z) \cdot \Pr(X = x \mid Z = z) = \\
&\quad = \Pr(X = x \wedge Y = y \mid Z = z)
\end{aligned}$$

for all value assignments x, y, z to X, Y, Z, by once more using the chain rule. We conclude that $I_{\Pr}(X \cup Y, Z, X)$. Now observe that the independence relation $I_{\Pr}$ of the probability distribution $\Pr$ is a model of Pearl's extended axiomatic system for informational independence $\mathcal{E}$ and therefore obeys the five axioms from $\mathcal{E}$. From $I_{\Pr}(X \cup Y, Z, X)$, we now find

$$\begin{aligned}
I_{\Pr}(X \cup Y, Z, X) \quad &\Rightarrow \quad I_{\Pr}(X, Z, X \cup Y) \Rightarrow \\
&\Rightarrow \quad I_{\Pr}(X, Z, Y)
\end{aligned}$$

by the symmetry and decomposition axioms. We conclude that $I_{\Pr}(X, Z, X) \rightarrow I_{\Pr}(X, Z, Y)$ for all sets of variables $X, Y, Z \subseteq V$. $\square$

The property

$$I_{\Pr}(X, Z, X) \rightarrow I_{\Pr}(X, Z, Y)$$

for all sets of variables $X, Y, Z \subseteq V$, stated in the previous proposition, expresses that, if information about X is irrelevant with respect to X itself in the context of information about Z, then information about *any* set of variables is irrelevant with respect to X in this context.

Pearl's extended axiomatic system for informational independence $\mathcal{E}$ has been designed to characterise independence relations that may be embedded in a joint probability distribution [Pearl, 1988]. Unfortunately, the axiomatic system $\mathcal{E}$ does not reflect the property

stated in Proposition 4.6. An independence relation I in the set of models of the system $\mathcal{E}$ may very well include a statement $I(X, Z, X)$ for some sets of variables $X, Z \subseteq V$ and yet lack the statement $I(X, Z, Y)$ for some set $Y \subseteq V$; we will return to this observation shortly. To enforce independence statements of the form $I(X, Z, Y)$ to be included in models of informational independence in which $I(X, Z, X)$ holds, we add to Pearl's extended axiomatic system $\mathcal{E}$ the axiom $I(X, Z, X) \rightarrow I(X, Z, Y)$ for all sets of variables $X, Y, Z \subseteq V$. The newly included axiom will be termed the *determination axiom*.

The following definition summarises the thus enhanced axiomatic system for informational independence $\mathcal{I}$.

Definition 4.7 *Let $\mathcal{E}$ be as before. The* axiomatic system for informational independence $\mathcal{I}$ *is composed of the five axioms from $\mathcal{E}$ and the additional axiom*

$$I(X, Z, X) \rightarrow I(X, Z, Y)$$

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ on a finite set of variables V, for all sets of variables $X, Y, Z \subseteq V$; the set of all models of the system $\mathcal{I}$ will be denoted as $[\mathcal{I}]$.

The enhanced axiomatic system for informational independence $\mathcal{I}$ provides a better characterisation of the concept of probabilistic independence than the axiomatic system $\mathcal{E}$: it allows fewer models of independence that are not embedded in any joint probability distribution and, yet, serves to capture all probabilistic independence relations. This property is stated more formally in the following lemma.

Lemma 4.8 *Let $\mathcal{P}$ be as before. Furthermore, let $[\mathcal{E}]$ be the set of models of the axiomatic system $\mathcal{E}$. Let $[\mathcal{I}]$ be the set of all models of the axiomatic system for informational independence $\mathcal{I}$ as defined in Definition 4.7. Then,*

$$\mathcal{P} \subset [\mathcal{I}] \subset [\mathcal{E}]$$

Proof. The property $\mathcal{P} \subset [\mathcal{I}]$ follows from Proposition 4.6 and our observations in Section 2. We now prove that $[\mathcal{I}] \subset [\mathcal{E}]$. From Definition 4.7, we have that $[\mathcal{I}] \subseteq [\mathcal{E}]$. To show that $[\mathcal{E}] \not\subseteq [\mathcal{I}]$, it suffices to give an example independence relation I that is included in the set of models $[\mathcal{E}]$ and not in $[\mathcal{I}]$. For this purpose, we consider the independence relation $I \subseteq 2^V \times 2^V \times 2^V$ on the set of variables $V = \{V_0, V_1, V_2\}$ that is defined by the (single) statement $I(\{V_0\}, \{V_1\}, \{V_0\})$, that is, the relation I includes $I(\{V_0\}, \{V_1\}, \{V_0\})$ and all independence statements that can be derived from this statement by the five axioms of the system $\mathcal{E}$. It will be evident that for this independence relation I we have that $I \in [\mathcal{E}]$. We will now show that $\neg I(\{V_0\}, \{V_1\}, \{V_2\})$, from which we will conclude that $I \notin [\mathcal{I}]$.

We begin by showing that for any independence statement $I(\{V_0\}, Z, Y)$ with $V_2 \in Y$ that is derived from the statement $I(\{V_0\}, \{V_1\}, \{V_0\})$ by the decomposition, weak union, contraction, and overlap axioms, we have that $V_2 \in Z$. We will prove this property by induction on the length of the derivation.

Induction Basis
The property stated above trivially holds for every independence statement that is derived in zero derivation steps.

Induction Hypothesis
For some $n \geq 0$, we assume that for all independence statements $I(\{V_0\}, Z, Y)$ with $V_2 \in Y$ that are derived from the statement $I(\{V_0\}, \{V_1\}, \{V_0\})$ by application of the four axioms mentioned above in $k \leq n$ derivation steps, we have that $V_2 \in Z$.

Induction Step
Now, consider an independence statement $I(\{V_0\}, Z, Y)$ with $V_2 \in Y$ that is derived from $I(\{V_0\}, \{V_1\}, \{V_0\})$ in $n+1$ derivation steps. We distinguish between all possible derivations of $I(\{V_0\}, Z, Y)$.

- by the *decomposition axiom*, we have that

  $$I(\{V_0\}, Z, Y') \Rightarrow I(\{V_0\}, Z, Y)$$

  for some set of variables $Y' \subseteq V$ with $Y \subseteq Y'$. Note that from $V_2 \in Y$ and $Y \subseteq Y'$, we have that $V_2 \in Y'$. From the independence statement $I(\{V_0\}, Z, Y')$ having been derived from the statement $I(\{V_0\}, \{V_1\}, \{V_0\})$ in at most n derivation steps and $V_2 \in Y'$, we conclude by the induction hypothesis that $V_2 \in Z$.

- by the *weak union axiom*, we have that

  $$I(\{V_0\}, Z', Y') \Rightarrow I(\{V_0\}, Z, Y)$$

  for some sets of variables $Y', Z' \subseteq V$ with $Y \subseteq Y'$, $Z' \subseteq Z$, and $Y' \cup Z' = Y \cup Z$. Note that from $V_2 \in Y$ and $Y \subseteq Y'$, we have that $V_2 \in Y'$. From the independence statement $I(\{V_0\}, Z', Y')$ having been derived from the statement $I(\{V_0\}, \{V_1\}, \{V_0\})$ in at most n derivation steps and $V_2 \in Y'$, we conclude by the induction hypothesis that $V_2 \in Z'$. From $Z' \subseteq Z$, it follows that $V_2 \in Z$.

- by the *contraction axiom*, we have that

  $$I(\{V_0\}, Z, Y') \wedge I(\{V_0\}, Z', Y'') \Rightarrow I(\{V_0\}, Z, Y)$$

  for some sets of variables $Y', Y'', Z' \subseteq V$ with $Y' \cup Y'' = Y$ and $Z' = Z \cup Y'$. From $V_2 \in Y$ and $Y' \cup Y'' = Y$, we have that $V_2 \in Y'$ or $V_2 \in Y''$.

  - we suppose that $V_2 \in Y'$. From the independence statement $I(\{V_0\}, Z, Y')$ having been derived from $I(\{V_0\}, \{V_1\}, \{V_0\})$ in at most n derivation steps and $V_2 \in Y'$, we conclude by the induction hypothesis that $V_2 \in Z$.
  - we now suppose that $V_2 \in Y''$ and $V_2 \notin Y'$. From the independence statement $I(\{V_0\}, Z', Y'')$ having been derived from $I(\{V_0\}, \{V_1\}, \{V_0\})$ in at most n derivation steps and $V_2 \in Y''$, we conclude by the induction hypothesis that $V_2 \in Z'$. From $Z' = Z \cup Y'$ and $V_2 \notin Y'$, we find that $V_2 \in Z$.

- by the *overlap axiom*, we have that

  $$I(\{V_0\}, Z, Y)$$

  with $Z = Y$. From $V_2 \in Y$ and $Z = Y$, we have that $V_2 \in Z$.

We conclude that for any independence statement $I(\{V_0\}, Z, Y)$ with $V_2 \in Y$ that is derived from the statement $I(\{V_0\}, \{V_1\}, \{V_0\})$ by application of the decomposition, weak union, contractions, and overlap axioms, we have that $V_2 \in Z$. From this property, it now follows that $\neg I(\{V_0\}, \{V_1\}, \{V_2\})$. We would like to note that, so far, the symmetry axiom has not been taken into consideration; analogous observations, however, hold when the symmetry axiom is applied. $\square$

From the determination axiom that we have included in the axiomatic system $\mathcal{I}$, we have that, if in the context of information about Z information about X is irrelevant with respect to X itself, then it can never become relevant with respect to X no matter which further information may be learned. This property is stated more formally in the following lemma.

Lemma 4.9 *Let $[\mathcal{I}]$ be the set of models of the axiomatic system $\mathcal{I}$. Let $I \in [\mathcal{I}]$ be an independence relation on a finite set of variables V. Then,*

$$I(X, Z, X) \rightarrow I(X, Z \cup Y, X)$$

for all sets of variables $X, Y, Z \subseteq V$.

Proof. From $I \in [\mathcal{I}]$, we have that the independence relation I obeys the six axioms from the axiomatic system $\mathcal{I}$. We now find that

$$\begin{aligned} I(X, Z, X) &\Rightarrow I(X, Z, X \cup Y) \Rightarrow \\ &\Rightarrow I(X, Z \cup Y, X) \end{aligned}$$

by the determination axiom and the weak union axiom. $\square$

The determination axiom pertains to the independences holding for the set of functionally dependent variables in the context of their determinants. The following lemma now shows that the presence of functional dependences does not interfere with the independences among all other variables discerned.

Lemma 4.10 *Let $[\mathcal{I}]$ be the set of models of the axiomatic system $\mathcal{I}$. Let $I \in [\mathcal{I}]$ be an independence relation on a finite set of variables V. Then,*

$$I(X, Z, X) \wedge I(Y, Z, W) \rightarrow I(X \cup Y, Z, X \cup W)$$

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

Proof. We begin our proof by observing that since $I \in [\mathcal{I}]$, it obeys all six axioms from the axiomatic system $\mathcal{I}$. To prove the property stated in the lemma, we assume that $I(X, Z, X)$ and $I(Y, Z, W)$ for some sets of variables $X, Y, Z, W \subseteq V$. The basic idea of the proof now is to apply the contraction axiom twice, once to construct the argument $X \cup Y$ for the statement $I(X \cup Y, Z, X \cup W)$ and a second time to construct the argument $X \cup W$. From our assumption $I(X, Z, X)$, we find

$$\begin{aligned} I(X, Z, X) &\Rightarrow I(X, Z \cup Y, X) \Rightarrow \\ &\Rightarrow I(X, Z \cup Y, W) \Rightarrow \\ &\Rightarrow I(W, Z \cup Y, X) \end{aligned}$$

by Lemma 4.9, the determination axiom, and the symmetry axiom. From our assumption $I(Y, Z, W)$, we find

$$I(Y, Z, W) \quad \Rightarrow \quad I(W, Z, Y)$$

by the symmetry axiom. From the independence statements $I(W, Z \cup Y, X)$ and $I(W, Z, Y)$, we now find

$$\begin{aligned} I(W, Z, Y) \wedge I(W, Z \cup Y, X) \quad &\Rightarrow \quad I(W, Z, X \cup Y) \Rightarrow \\ &\Rightarrow \quad I(X \cup Y, Z, W) \end{aligned}$$

by the contraction and symmetry axioms.

From our assumption $I(X, Z, X)$, we further find

$$\begin{aligned} I(X, Z, X) \quad &\Rightarrow \quad I(X, Z \cup W, X) \Rightarrow \\ &\Rightarrow \quad I(X, Z \cup W, X \cup Y) \Rightarrow \\ &\Rightarrow \quad I(X \cup Y, Z \cup W, X) \end{aligned}$$

by Lemma 4.9, the determination axiom, and the symmetry axiom. From the independence statements $I(X \cup Y, Z, W)$ and $I(X \cup Y, Z \cup W, X)$, we find

$$I(X \cup Y, Z, W) \wedge I(X \cup Y, Z \cup W, X) \quad \Rightarrow \quad I(X \cup Y, Z, X \cup W)$$

by the contraction axiom. □

The property stated in the previous lemma is easily generalised to a bi-implication.

4.3 Summary

In the foregoing, we have reviewed Pearl's extended axiomatic system for informational independence $\mathcal{E}$ and have included two additional axioms: the no-functions axiom and the determination axiom. We summarise the resulting axiomatic system $\mathcal{I}$.

Definition 4.11 *The* axiomatic system for informational independence $\mathcal{I}$ *is composed of the six axioms*

- $I(X, Z, Y) \rightarrow I(Y, Z, X)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z, W)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y)$;
- $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;
- $I(X, Z, X) \rightarrow I(X, Z, Y)$;
- $I(X, Z, Z)$;

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ on a finite set of variables V, for all sets of variables $X, Y, Z, W \subseteq V$; the set of all models of the system $\mathcal{I}$ will be denoted as $[\mathcal{I}]$. The axiomatic system $\mathcal{I}^+$ comprises the six axioms from the system $\mathcal{I}$ and the two additional axioms

- $I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;
- $\neg I(\{V_i\}, Z', \{V_i\})$;

for any ternary relation $I \subseteq 2^V \times 2^V \times 2^V$ on V, for all sets of variables $X, Y, Z, W \subseteq V$, and for each variable $V_i \in V$ and all sets of variables $Z' \subseteq V - \{V_i\}$; the set of all models of the system $\mathcal{I}^+$ will be denoted as $[\mathcal{I}^+]$.

Any model in the set of models $[\mathcal{I}]$ of the axiomatic system $\mathcal{I}$ will be termed an *enhanced semi-graphoid independence relation*; a model in $[\mathcal{I}^+]$ will be called an *enhanced graphoid independence relation.*

The axiomatic system for informational independence $\mathcal{I}$ reviewed in the previous definition coincides with the system $\mathcal{I}$ introduced in Definition 4.7. The axiomatic system $\mathcal{I}^+$, however, differs from the system $\mathcal{F}^+$ introduced in Definition 4.2 in its including the determination axiom. The following lemma shows, however, that the sets of models of the two axiomatic systems do not differ.

Lemma 4.12 *Let $[\mathcal{F}^+]$ be the set of models of the axiomatic system $\mathcal{F}^+$. Let $[\mathcal{I}^+]$ be the set of models of the axiomatic system for informational independence $\mathcal{I}^+$ as defined in Definition 4.11. Then,*

$$[\mathcal{F}^+] = [\mathcal{I}^+]$$

Proof. From the Definitions 4.2 and 4.11, we have that $[\mathcal{I}^+] \subseteq [\mathcal{F}^+]$. We now prove that $[\mathcal{F}^+] \subseteq [\mathcal{I}^+]$. Let $I \in [\mathcal{F}^+]$ be an independence relation on a finite set of variables V. From $I \in [\mathcal{F}^+]$, we have that I obeys all axioms from the axiomatic system $\mathcal{F}^+$; more in specific, we have that I satisfies the no-functions axiom. We now assume that $I(X, Z, X)$ for some sets of variables $X, Z \subseteq V$. Note that from Corollary 4.5 we have that $X \subseteq Z$. From the overlap axiom, we further have that $I(Y, Z, Z)$. From the latter independence statement, we find

$$\begin{aligned} I(Y, Z, Z) \quad &\Rightarrow \quad I(Y, Z, X) \Rightarrow \\ &\Rightarrow \quad I(X, Z, Y) \end{aligned}$$

by the decomposition and symmetry axioms, using $X \subseteq Z$. So,

$$I(X, Z, X) \quad \Rightarrow \quad I(X, Z, Y)$$

for the sets of variables $X, Y, Z \subseteq V$. We conclude that $I \in [\mathcal{I}^+]$ and, hence, that $[\mathcal{F}^+] \subseteq [\mathcal{I}^+]$. □

Figure 1 summarises the inclusion relationships among the various sets of models of informational independence that we have introduced.

5 On Normal Forms

In his work on graphical representations of independence relations, J. Pearl builds on the *restricted* axiomatic system for informational independence $\mathcal{A}$, that is, on the system that involves axioms for mutually disjoint sets of variables only [Pearl, 1988]. In this section, we address Pearl's restricted focus on disjoint sets of variables in view of such representations. To

$$
\begin{array}{ccccccccc}
\mathcal{P}^+ & \subset & [\mathcal{I}^+] & = & [\mathcal{F}^+] & \subset & [\mathcal{E}^+] & \subset & [\mathcal{A}^+] \\
\subset & & \subset & & & & \subset & & \subset \\
\mathcal{P} & \subset & [\mathcal{I}] & & \subset & & [\mathcal{E}] & \subset & [\mathcal{A}]
\end{array}
$$

Figure 1: Inclusion Relationships among Sets of Models of Independence.

this end, we examine from an independence relation the statements that involve overlapping sets of variables. We would like to note that building on the restricted axiomatic system $\mathcal{A}$ presupposes that any such statement can be derived from the relation's set of statements that involve mutually disjoint sets of variables only, by the axioms of the extended system $\mathcal{E}$. In fact, the extended axiomatic system is then looked upon as allowing for a *normal form* for informational independence that warrants building on the restricted system; this normal form will be termed *Pearl normal form.*

Definition 5.1 *Let $[\mathcal{E}]$ be the set of models of the axiomatic system $\mathcal{E}$. Let $I \in [\mathcal{E}]$ be an independence relation on a finite set of variables V. A statement $I(X, Z, Y)$ with $X, Y, Z \subseteq V$ is said to be in* Pearl normal form *if the sets of variables X, Y, and Z are mutually disjoint.*

In the following sections, we show that, unfortunately, Pearl normal form does not suffice for faithfully representing extended independence relations. In doing so, we separately address graphoid and semi-graphoid independence relations.

5.1 A Normal Form for Graphoid Independence Relations

We begin by examining the set of models of the extended axiomatic system for informational independence $\mathcal{E}^+$. The following lemma states that in this set of models there exists an extended graphoid independence relation I that includes independence statements that cannot be derived from I's statements in Pearl normal form by the axioms of the system $\mathcal{E}^+$.

Lemma 5.2 *Let $[\mathcal{E}^+]$ be the set of models of the axiomatic system $\mathcal{E}^+$. For each independence relation $I \in [\mathcal{E}^+]$, let $I_{PNF} \subseteq I$ be the set of statements from I that are in Pearl normal form. Then, there exists an independence relation $I \in [\mathcal{E}^+]$ such that I includes a statement $s \in I - I_{PNF}$ for which there do not exist statements $s'_i \in I_{PNF}$, $i = 1, \ldots, k$, $k \geq 1$, such that $\bigwedge_{i=1}^{k} s'_i \Leftrightarrow s$.*

Proof. We consider the independence relation $I = 2^V \times 2^V \times 2^V$ on some set of variables V. For this independence relation I we have that $I \in [\mathcal{E}^+]$. We observe that the relation I includes the independence statement $I(\{V_i\}, \varnothing, \{V_i\})$, for any variable $V_i \in V$. This statement is not in Pearl normal form, that is, $I(\{V_i\}, \varnothing, \{V_i\}) \in I - I_{PNF}$. The statement, however, cannot be derived from the relation's independence statements in Pearl normal form by the axioms from $\mathcal{E}^+$. $\square$

From the previous lemma we have that Pearl normal form does not suffice for faithfully representing any extended graphoid independence relation. Pearl normal form, however, does

allow for capturing enhanced graphoid independence relations, as is stated in the following lemma.

Lemma 5.3 *Let $[\mathcal{I}^+]$ be the set of models of the axiomatic system $\mathcal{I}^+$. Let $I \in [\mathcal{I}^+]$ be an independence relation and let $I_{PNF} \subseteq I$ be the set of statements from I that are in Pearl normal form. Then, for each independence statement $s \in I - I_{PNF}$, there exists a statement $s' \in I_{PNF}$ such that $s \Leftrightarrow s'$.*

Proof. We assume that the independence relation I is defined on the set of variables V. We now consider an independence statement $I(X, Z, Y)$, for some sets of variables $X, Y, Z \subseteq V$, that is not in Pearl normal form, that is, the sets X, Y, and Z are not mutually disjoint. From $I \in [\mathcal{I}^+]$ and from Lemma 4.12 and Corollary 4.5, we have that $X \cap Y \subseteq Z$. We now find that

$$I(X, Z, Y) \quad \Leftrightarrow \quad I(X - Z, Z, Y - Z)$$

by Lemma 3.3. From $X \cap Y \subseteq Z$, we conclude that the statement $I(X - Z, Z, Y - Z)$ is in Pearl normal form. □

The property stated in Lemma 5.3 is depicted schematically in Figure 2. In the figure, I represents the set of all statements from the independence relation at hand; I_{PNF} represents the relation's subset of statements that are in Pearl normal form.

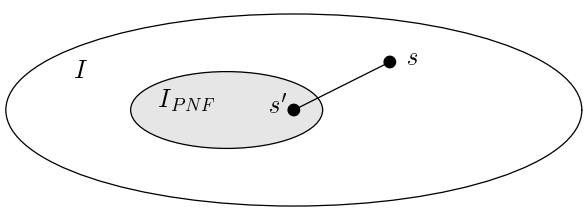

Figure 2: The Derivability of Statements in $I - I_{PNF}$ for Enhanced Graphoid Independence Relations.

From Lemma 5.3, we conclude that in view of enhanced graphoid independence relations, Pearl's focus on mutually disjoint sets of variables is not essential: any statement from such an independence relation that is not in Pearl normal form can be derived by our enhanced axiomatic system $\mathcal{I}^+$ from the relation's set of statements that are in Pearl normal form.

5.2 A Normal Form for Semi-graphoid Independence Relations

We now address Pearl normal form in view of representing semi-graphoid independence relations more in general. To this end, we begin by examining the set of models of the extended axiomatic system for informational independence $\mathcal{E}$. From Lemma 5.2 we have that in this set of models there exists a semi-graphoid independence relation I that includes independence statements that cannot be derived from I's statements in Pearl normal form by the axioms of the system $\mathcal{E}$. Unfortunately, similar observations also apply to enhanced semi-graphoid independence relations, that is, there exists an enhanced graphoid independence relation I that includes independence statements that cannot be derived from I's statements in Pearl normal form by the axioms of the system $\mathcal{I}$. This property is stated more formally in the following lemma.

Lemma 5.4 *Let $[\mathcal{I}]$ be the set of models of the axiomatic system $\mathcal{I}$. For each independence relation $I \in [\mathcal{I}]$, let $I_{PNF} \subseteq I$ be the set of statements from I that are in Pearl normal form. Then, there exists an independence relation $I \in [\mathcal{I}]$ such that I includes a statement $s \in I - I_{PNF}$ for which there do not exist statements $s'_i \in I_{PNF}$, $i = 1, \ldots, k$, $k \geq 1$, such that $\bigwedge_{i=1}^{k} s'_i \Leftrightarrow s$.*

Proof. The proof of the property stated in the lemma is analogous to the proof of Lemma 5.2. □

To allow for faithfully representing enhanced semi-graphoid independence relations, we supplement Pearl normal form by an additional normal form called *functional form*.

Definition 5.5 *Let $[\mathcal{I}]$ be the set of models of the axiomatic system $\mathcal{I}$. Let $I \in [\mathcal{I}]$ be an independence relation on a finite set of variables V. A statement $I(X, Z, Y)$ with $X, Y, Z \subseteq V$ is said to be in* functional form *if $X = Y$.*

The following lemma now states that Pearl normal form and functional form with each other allow for capturing enhanced semi-graphoid independence relations.

Lemma 5.6 *Let $[\mathcal{I}]$ be the set of models of the axiomatic system $\mathcal{I}$. Let $I \in [\mathcal{I}]$ be an independence relation; let $I_{PNF} \subseteq I$ be the set of statements from I that are in Pearl normal form and let $I_{FF} \subseteq I$ be the set of statements from I that are in functional form. Then, for each independence statement $s \in I - (I_{PNF} \cup I_{FF})$, there exists a statement $s' \in I_{PNF}$ and a statement $s'' \in I_{FF}$ such that $s' \wedge s'' \Leftrightarrow s$.*

Proof. We assume that the independence relation I is defined on the set of variables V. We now consider an independence statement $I(X, Z, Y)$ for some sets of variables $X, Y, Z \subseteq V$ that are not mutually disjoint. Note that for this statement we have either $X \cap Y \subseteq Z$ or $X \cap Y \not\subseteq Z$. To an independence statement $I(X, Z, Y)$ with $X \cap Y \subseteq Z$ similar observations apply as to the independences in a model of $\mathcal{I}^+$: this statement can be derived from independence statements that are in Pearl normal form. We now consider the case where $X \cap Y \not\subseteq Z$. Without loss of generality we assume that $X \cap Z = \varnothing$ and $Y \cap Z = \varnothing$. From the independence statement $I(X, Z, Y)$, we find that

$$\begin{aligned} I(X, Z, Y) \quad & \Rightarrow I(X, Z, Y - X) \Rightarrow \\ & \Rightarrow I(Y - X, Z, X) \Rightarrow \\ & \Rightarrow I(Y - X, Z, X - Y) \Rightarrow \\ & \Rightarrow I(X - Y, Z, Y - X) \end{aligned}$$

by the decomposition and symmetry axioms; similarly, we find that

$$\begin{aligned} I(X, Z, Y) \quad & \Rightarrow I(X, Z, X \cap Y) \Rightarrow \\ & \Rightarrow I(X \cap Y, Z, X) \Rightarrow \\ & \Rightarrow I(X \cap Y, Z, X \cap Y) \end{aligned}$$

So,

$$I(X, Z, Y) \Rightarrow I(X - Y, Z, Y - X) \wedge I(X \cap Y, Z, X \cap Y)$$

Note that from $X \cap Z = \varnothing$ and $Y \cap Z = \varnothing$, we have that the statement $I(X - Y, Z, Y - X)$ is in Pearl normal form; the statement $I(X \cap Y, Z, X \cap Y)$ is in functional form. Conversely, we have that

$$I(X - Y, Z, Y - X) \wedge I(X \cap Y, Z, X \cap Y) \Rightarrow I(X, Z, Y)$$

by Lemma 4.10. We conclude that, for any independence statement s in I that is not in Pearl's normal form nor in functional form, there exist in I a statement s' that is in Pearl's normal form and a statement s'' that is in functional form such that $s' \wedge s'' \Leftrightarrow s$. $\square$

The property stated in Lemma 5.6 is depicted schematically in Figure 3. In the figure, I represents the set of all statements from the independence relation at hand; I_{FF} represents the relation's subset of statements that are in functional form; I_{PNF} represents the statements from I that are in Pearl normal form.

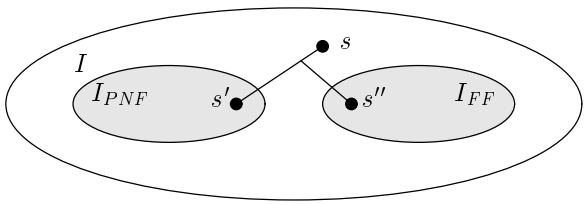

Figure 3: The Derivability of Statements in $I - (I_{PNF} \cup I_{FF})$ for Enhanced Semi-graphoid Independence Relations.

From the above observations we conclude that in view of independence relations embedding functional dependences, Pearl's focus on mutually disjoint sets of variables hides an interesting and important set of independence statements, namely the statements arising from functional dependences, as these cannot be derived by the extended axiomatic system $\mathcal{E}$ from an independence relation's set of statements in Pearl normal form. By supplementing Pearl normal form by functional form, however, the independence statements arising from functional dependences are faithfully represented. We would like to note that in relation with graphical representations of independence Pearl and his co-researchers have, informally, pointed out the necessity of explicitly modelling functional dependences [Pearl *et al.*, 1990]; our new normal form falls in with and extends on their observations.

6 Conclusions

The concept of informational independence has been axiomatised by J. Pearl and his co-researchers. Pearl's axiomatic system for informational independence is restricted in the sense that it applies to independences among mutually disjoint sets of variables only. We have shown that this focus on disjointness can hide a set of interesting properties of informational independence, namely properties arising from functional dependences among the variables discerned. To capture these properties we have enhanced Pearl's axiomatic system with several new axioms. We have shown that the thus enhanced axiomatic system provides a better characterisation of the concept of probabilistic independence than Pearl's system does. We have also shown that the enhanced axiomatic system provides for a normal form for informational independence that elucidates the necessity of explicitly representing the functional dependences among the variables of an independence relation.

Remark

This paper is a revised and extended version of [Van der Gaag & Meyer, 1996].

References

[Dawid, 1979] A.P. Dawid (1979). Conditional independence in statistical theory, *Journal of the Royal Statistical Society (Series B)*, vol. 41, pp. 1 – 31.

[Geiger & Pearl, 1988] D. Geiger and J. Pearl (1988). On the logic of causal models, *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, pp. 136 – 147.

[Lauritzen, 1982] S.L. Lauritzen (1982). *Lectures on Contingency Tables*, 2nd. ed., University of Aalborg Press, Aalborg, Denmark.

[Lauritzen & Spiegelhalter, 1988] S.L. Lauritzen and D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society, Series B*, vol. 50, pp. 157 – 224.

[Pearl, 1988] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Morgan Kaufmann, Palo Alto.

[Pearl *et al.*, 1990] J. Pearl, D. Geiger, and T. Verma (1990). The logic of influence diagrams, in: R.M. Oliver and J.Q. Smith (eds). *Influence Diagrams, Belief Nets and Decision Analysis*, John Wiley & Sons, pp. 67 – 86.

[Pearl & Paz, 1985] J. Pearl and A. Paz (1985). GRAPHOIDS: a graph-based logic for reasoning about relevance relations, in: B. Du Boulay, D. Hogg, and L. Steels (eds). *Advances in Artificial Intelligence 2*, North-Holland.

[Pearl & Verma, 1987] J. Pearl and T.S. Verma (1987). The logic of representing dependencies by directed acyclic graphs, *Proceedings of the Sixth National Conference on Artificial Intelligence*, pp. 374 – 379.

[Studený, 1989] M. Studený (1989). Multiinformation and the problem of characterization of conditional-independence relations, *Problems Control Information Theory*, vol. 18, pp. 3 – 16.

[Studený, 1992] M. Studený (1992). Conditional independence relations have no finite complete characterization, in: S. Kubik, J.A. Visek (eds). *Information Theory, Statistical Decision Functions and Random Processes*, Kluwer, Dordrecht, pp. 377 – 396.

[Van der Gaag & Meyer, 1996] L.C. van der Gaag and J.-J.Ch. Meyer (1996). Characterising normal forms for informational independence, *Proceedings of the Sixth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 973 – 978.