

Dependence and Relevance: A probabilistic view

Dan Geiger

Technion—Israel Institute of Technology
Haifa, Israel, 32000

David Heckerman*

Computer Science Department
University of California
Los Angeles, CA 90024

February 2, 1993

Abstract

We examine three probabilistic concepts related to the sentence “two variables have no bearing on each other”. We explore the relationships between these three concepts and establish their relevance to the process of constructing similarity networks—a tool for acquiring probabilistic knowledge from human experts. We also establish a precise relationship between connectedness in Bayesian networks and relevance in probability.

1 Introduction

The notion of relevance between pieces of information plays a key role in the theory of Bayesian networks and in the way they are used for inference.

*Current mailing address: One Microsoft Way, 9S/1024, Redmond, WA 98052-6399

The intuition that guides the construction of Bayesian networks draws from the analogy between “connectedness” in graphical representations and “relevance” in the domain represented, that is, two nodes connected along some path correspond to variables of mutual relevance.

We examine three formal concepts related to the sentence “variables a and b have no bearing on each other”. First, two variables a and b are said to be *mutually irrelevant* if they are conditionally independent given any value of any subset of the other variables in the domain. Second, two variables are said to be *uncoupled* if the set of variables representing the domain can be partitioned into two independent sets one containing a and the other containing b . Finally, two variables a and b are *unrelated* if the corresponding nodes are disconnected in every minimal Bayesian network representation (to be defined).

The three concepts, mutual-irrelevance, uncoupledness, and unrelatedness are not identical. We show that uncoupledness and unrelatedness are always equivalent but sometimes differ from the notion of mutual-irrelevance. We identify a class of models called *transitive* for which all three concepts are equivalent. Strictly positive binary distributions (defined below) are examples of transitive models. We also show that “disconnectedness” in graphical representations and mutual-irrelevance in the domain represented coincide for every transitive model and for none other.

These results have theoretical and practical ramifications. Our analysis uses a qualitative abstraction of conditional independence known as graphoids [Pearl and Paz, 1989], and demonstrates the need for this abstraction in manipulating conditional independence assumptions (which are

an integral part of every probabilistic reasoning engine). Our results also simplify the process of acquiring probabilistic knowledge from domain experts via similarity networks.

This article is organized as follows: A short overview on graphoids and their Bayesian network representation is provided in Section 2. (For more details consult Chapter 3 in Pearl, 1988.) Section 3 and 4 investigate properties of mutual-irrelevance, uncoupledness, and unrelatedness and their relation to each other. Section 5 discusses two definitions of similarity networks. Section 6 shows that for a large class of probability distributions these definitions are equivalent.

2 Graphoids and Bayesian Networks

Since our definitions of mutual-irrelevance, uncoupledness, and unrelatedness all rely on the notion of conditional independence, it is useful to abstract probability distributions to reflect this fact. In particular, every probability distribution is viewed as a list of conditional independence statements with no reference to numerical parameters. This abstraction, called a *graphoid*, was proposed by Pearl and Paz [1989] and further discussed by Pearl [1988] and Geiger [1990].

Throughout the discussion we consider a finite set of variables $U = \{u_1, \dots, u_n\}$ each of which is associated with a finite *set of values* $d(u_i)$ and a probability distribution P having the Cartesian product $\times_{u_i \in U} d(u_i)$ as its sample space.

Definition A probability distribution P is *defined over* U if its sample space is $\times_{u_i \in U} d(u_i)$.

We use lowercase letters possibly subscripted (e.g., a , b , x or u_i) to denote variables, and use uppercase letters (e.g., X , Y , or Z) to denote sets of variables. A bold lowercase or uppercase letter refers to a value (instance) of a variable or of a set of variables, respectively. A value \mathbf{X} of a set of variables X is an element in the Cartesian product $\times_{x \in X} d(x)$, where $d(x)$ is the set of values of x . The notation $X = \mathbf{X}$ stands for $x_1 = \mathbf{x}_1, \dots, x_n = \mathbf{x}_n$, where $X = \{x_1, \dots, x_n\}$ and \mathbf{x}_i is a value of x_i .

Definition The expression $I(X, Y \mid Z)$, where X , Y , and Z are disjoint subsets of U , is called an *independence statement*, or *independency*. Its negation $\neg I(X, Y \mid Z)$ is called a *dependence statement*, or *dependency*. An independence or dependence statement is defined *over* $V \subseteq U$ if it mentions only elements of V .

Definition Let $U = \{u_1, \dots, u_n\}$ be a finite set of variables with $d(u_i)$ and P as above. An independence statement $I(X, Y \mid Z)$ is said to *hold for* P if for every value \mathbf{X} , \mathbf{Y} , and \mathbf{Z} of X , Y , and Z , respectively

$$P(X = \mathbf{X} \mid Y = \mathbf{Y}, Z = \mathbf{Z}) = P(X = \mathbf{X} \mid Z = \mathbf{Z})$$

or $P(Z = \mathbf{Z}) = 0$. Equivalently, P is said to *satisfy* $I(X, Y \mid Z)$. Otherwise, P is said to *satisfy* $\neg I(X, Y \mid Z)$.

Definition When $I(X, Y \mid Z)$ holds for P , then X and Y are *conditionally independent* relative to P and if $Z = \emptyset$, then X and Y are *marginally independent* relative to P .

Every probability distribution defines a *dependency model*.

Definition [Pearl, 1988] A *dependency model* M over a finite set of elements

U is a set of triplets $(X, Y | Z)$, where X , Y and Z are disjoint subsets of U .

The definition of dependency models does not assume any structure on the elements of U . Namely, an element of U could be, for example, a node in some graph or a name of a variable. In particular, if each $u_i \in U$ is associated with a finite set $d(u_i)$, then every probability distribution having the Cartesian product $\times_{u_i \in U} d(u_i)$ as its sample space defines a dependency model via the rule:

$$(X, Y | Z) \in M \text{ if and only if } I(X, Y | Z) \text{ holds in } P, \quad (1)$$

for every disjoint subsets X, Y , and Z of U . Dependency models constructed using Equation 1 have some interesting properties that are summarized in the definition below.

Definition A *Graphoid* M over a finite set U is any set of triplets $(X, Y | Z)$, where X , Y , and Z are disjoint subsets of U such that the following axioms are satisfied:

Trivial Independence

$$(X, \emptyset | Z) \in M \quad (2)$$

Symmetry

$$(X, Y | Z) \in M \Rightarrow (Y, X | Z) \in M \quad (3)$$

Decomposition

$$(X, Y \cup W | Z) \in M \Rightarrow (X, Y | Z) \in M \quad (4)$$

Weak union

$$(X, Y \cup W | Z) \in M \Rightarrow (X, Y | Z \cup W) \in M \quad (5)$$

Contraction

$$\begin{aligned} (X, Y \mid Z) \in M \ \& \ (X, W \mid Z \cup Y) \in M \ \Rightarrow \\ (X, Y \cup W \mid Z) \in M. \end{aligned} \tag{6}$$

The above relations are called the *graphoid axioms*.¹

Using the definition of conditional independence, it is easy to show that each probability distribution defines a graphoid via Equation 1 [Pearl, 1988]. The graphoid axioms have an appealing interpretation. For example, the weak union axiom states: If Y and W are conditionally independent of X , given a knowledge base Z , then Y is conditionally independent of X given W is added to the known knowledge base Z . In other words, the fact that a piece of information W , which is conditionally independent of X , becomes known, does not change the status of Y ; Y remains conditionally independent of X given the new *irrelevant* information W [Pearl, 1988].

Graphoids are suited to represent the qualitative part of a task that requires a probabilistic analysis. For example, suppose an alarm system is installed in your house in order to detect burglaries; and suppose it can be activated by two separate sensors. Suppose also that, when the alarm sound is activated, there is a good chance that a police patrol will show up. We are interested in computing the probability of a burglary given a police car is near your house.

The dependencies in this story can be represented by a graphoid. We consider five binary variables, *burglary*, *sensorA*, *sensorB*, *alarm*, and *patrol*, each

¹This definition differs slightly from that given in [Pearl and Paz, 1989], where axioms (3) through (6) define semi-graphoids. A variant of these axioms was first studied by David [1979] and Spohn [1980].

having two values *yes* and *no*. We know that the outcome of the two sensors are conditionally independent given *burglary*, and that *alarm* is conditionally independent of *burglary* given the outcome of the sensors. We also know that *patrol* is conditionally independent of *burglary* given *alarm*. (Assuming that only the alarm prompts a police patrol.) This qualitative information implies that the following three triplets must be included in a dependency model that describes the above story: $(sensorA, sensorB \mid burglary)$, $(alarm, burglary \mid \{sensorA, sensorB\})$ and $(patrol, \{burglary, sensorA, sensorB\} \mid alarm)$.

The explicit representation of all triplets of a dependency model is often impractical, because there are an exponential number of possible triplets. Consequently, an implicit representation is needed. We will next describe such a representation.

Definition [Pearl, 1988] Let M be a graphoid over U . A directed acyclic graph D is a *Bayesian network of M* if D is constructed from M by the following steps: assign a *construction order* u_1, u_2, \dots, u_n to the elements in U , and designate a node for each u_i . For each u_i in U , identify a set $\pi(u_i) \subseteq \{u_1, \dots, u_{i-1}\}$ such that

$$(\{u_i\}, \{u_1, \dots, u_{i-1}\} \setminus \pi(u_i) \mid \pi(u_i)) \in M. \quad (7)$$

Assign a link from every element in $\pi(u_i)$ to u_i . The resulting network is *minimal* if, for each $u \in U$, no proper subset of $\pi(u)$ satisfies Equation (7).

A Bayesian network of the burglary story is shown in Figure 2. We shall see that aside of the triplets that were used to construct the network, the triplet $(patrol, burglary \mid \{sensorA, sensorB\})$ follows from the topology of the

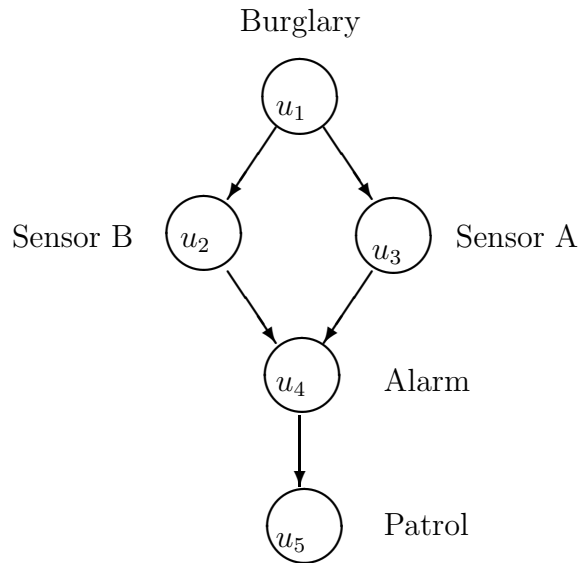


Figure 1: An example of a Bayesian network

network. In the remainder of this section, we describe a general methodology to determine which triplets of a graphoid M are represented in a Bayesian network of M . The criteria of d -separation, defined below, provides the answer. Some preliminary definitions are needed.

Definition The *underlying graph* of a Bayesian network is an undirected graph obtained from the Bayesian network by replacing every link with an undirected edge.

Definition A *trail* in a Bayesian network is a sequence of links that form a simple (cycle-free) path in the underlying graph. Two nodes are *connected* in a Bayesian network if there exists a trail connecting them. Otherwise they are *disconnected*. If $x \rightarrow y$ is a link in a Bayesian network, then x is a *parent*

of y and y is a *child* of x . If there is a directed path of length greater than zero from x to y , then x is an *ancestor* of y and y is a *descendant* of x .

Definition (Pearl, 1988) A node b is called a *head-to-head* node wrt (with respect to) a trail t if there are two consecutive links $a \rightarrow b$ and $b \leftarrow c$ on t .

For example, $u_2 \rightarrow u_4 \leftarrow u_3$ is a trail in Figure 2 and u_4 is a head-to-head node with respect to this trail.

Definition (Pearl, 1988) A trail t is *active* wrt a set of nodes Z if (1) every head-to-head node wrt t either is in Z or has a descendant in Z and (2) every other node along t is outside Z . Otherwise, the trail is said to be *blocked* (or *d-separated*) by Z .

In Figure 2, for example, both trails between $\{u_2\}$ and $\{u_3\}$ are d-separated by $Z = \{u_1\}$; the trail $u_2 \leftarrow u_1 \rightarrow u_3$ is d-separated by Z because node u_1 , which is not a head-to-head node wrt this trail, is in Z . The trail $u_2 \rightarrow u_4 \leftarrow u_3$ is d-separated by Z , because node u_4 and its descendant u_5 are outside Z . In contrast, $u_2 \rightarrow u_4 \leftarrow u_3$ is not d-separated by $Z' = \{u_1, u_5\}$ because u_5 is in Z' .

The theorem below is the major building block for most of the developments presented in this article.

Theorem 1 [Verma and Pearl, 1988] *Let D be a Bayesian network of a graphoid M over U ; and let X , Y , and Z be three disjoint subsets of U . If all trails between a node in X and a node in Y are d-separated by Z , then $(X, Y \mid Z) \in M$.*

For example, in the Bayesian network of Figure 2, all trails between u_1 and u_5 are d-separated by $\{u_2, u_4\}$. Thus, Theorem 1 guarantees that

$(u_5, u_1 \mid \{u_2, u_3\}) \in M$, where M is the graphoid from which this network was constructed.² Furthermore, Geiger and Pearl [1990] show that no other graphical criteria reveals more triplets of M than does d -separation.

Geiger et al. [1990] generalize Theorem 1 to networks that include *deterministic nodes* (i.e., nodes whose value is a *function* of their parents' values). Shachter [1990] obtains related results. Lauritzen et al. [1990] establish another graphical criteria and show that it is equivalent to d -separation.

3 Three Notions of Relevance

We can now define mutual-irrelevance, uncoupledness, and unrelatedness, and study their properties.

Definition Let M be a graphoid over U , and let $x, y \in U$.

- x and y are *uncoupled* if there exist a partition U_1, U_2 of U such that $x \in U_1$, $y \in U_2$, and $(U_1, U_2 \mid \emptyset) \in M$. Otherwise, x and y are *coupled*, denoted *coupled*(x, y).
- x and y are *unrelated* if x and y are disconnected in every minimal Bayesian network of M . Otherwise, x and y are *related*, denoted *related*(x, y).
- x and y are *mutually irrelevant* if $(x, y \mid Z) \in M$ for every $Z \subseteq U \setminus \{x, y\}$. Otherwise, x and y are *mutually-relevant*, denoted *relevant*(x, y).

We could have defined these concepts using conditional independence. By choosing the graphoid framework, however, we gain in two aspects. First,

²Within the expression of independence statements, we often write u_i instead of $\{u_i\}$.

we emphasize that all the properties that we discuss are proved using the graphoid axioms. We do not use any properties of probability theory that are not summarized in these axioms. Second, our results are more general in that they appeal to any graphoid, not necessarily a graphoid that is defined by conditional independence, or even a graphoid defined by a probability distribution. Examples of other types of graphoids are given in Pearl [1988].

Later in this section, we show that if two nodes x and y are disconnected in one minimal Bayesian network of M , then x and y are disconnected in every minimal network of M . Thus, to check whether x and y are unrelated, it suffices to examine whether or not they are connected in one minimal network representation rather than examine all possible minimal networks. This observation, demonstrated by Theorem 6, offers a considerable reduction in complexity. Based on the development that leads to Theorem 6, we also prove that x and y are unrelated if and only if they are uncoupled.

Definition A *connected component* C of a Bayesian network D is a subgraph of D in which every two nodes are connected (by a trail). A connected component is *maximal* if there exists no proper super-graph of C that is a connected component of D .

Lemma 2 *Let D be a Bayesian network of a graphoid M over U , and let A and B be subsets of U . If all nodes in A are disconnected from all nodes in B , then $(A, B \mid \emptyset) \in M$.*

Proof: There is no active trail between a node in A and a node in B . Thus, by Theorem 1, $(A, B \mid \emptyset) \in M$. \square

Lemma 3 *Let D be a Bayesian network of a graphoid M over U , x be in U , Z be the set of x 's parents, and Y be the set of all nodes that are not descendants of x except x 's parents. Then, $(x, Y \mid Z) \in M$.*

Proof: The set Z d-separates all trails between a node in Y and x , because each such trail either passes through a parent of x and therefore is blocked by Z , or each such trail must reach x through one of x 's children and thus must have a head-to-head node w , where neither w nor its descendants are in Z . Thus, by Theorem 1, $(x, Y \mid Z) \in M$. \square

Lemma 4 *Let D be a minimal Bayesian network of a graphoid M over U , x be in U , and $Z_1 \cup Z_2$ be x 's parents. Then $(x, Z_1 \mid Z_2) \notin M$, unless $Z_1 = \emptyset$.*

Proof: Since $Z_1 \cup Z_2$ are the parents of x , by Lemma 3, $(x, Y \mid Z_1 \cup Z_2) \in M$, where Y is the set of x 's non-descendants except its parents. Assume, by contradiction, that $(x, Z_1 \mid Z_2) \in M$ and $Z_1 \neq \emptyset$. The two triplets imply by the contraction axiom that $(x, Z_1 \cup Y \mid Z_2) \in M$. Since $Z_1 \neq \emptyset$, Z_2 is a proper subset of $Z_1 \cup Z_2$, where $Z_1 \cup Z_2$ are the parents of x in D . Hence D is not minimal, because Equation 7 is satisfied by a proper subset of x 's parents—a contradiction. \square

Definition A set $\{X_1, X_2\}$ is a *partition* of X iff $X = X_1 \cup X_2$, $X_1 \cap X_2 = \emptyset$, $X_1 \neq \emptyset$, and $X_2 \neq \emptyset$.

Lemma 5 *Let M be a graphoid over U , D be a minimal Bayesian network of M , and D_X be a connected component of D with a set of nodes X . Then, there exists no partition X_1, X_2 of X such that $(X_1, X_2 \mid \emptyset) \in M$.*

Proof: Suppose X_1, X_2 is a partition of X such that $(X_1, X_2 \mid \emptyset) \in M$. Since X_1 and X_2 are connected in D , there must exist a link between a node in X_1 and a node in X_2 . Without loss of generality, assume it is directed from a node in X_1 to a node u in X_2 .

Let Z_1, Z_2 be the parents of u in X_1 and X_2 , respectively. The triplet $(X_1, X_2 \mid \emptyset)$, which we assumed to be in M , implies—using symmetry and decomposition—that $(u \cup Z_2, Z_1 \mid \emptyset)$ is in M . By symmetry and weak union, $(u, Z_1 \mid Z_2)$ is in M as well. Thus, by Lemma 4, the network D is not minimal, unless $Z_1 = \emptyset$. We have assumed, however, that u has a parent in X_1 . Hence, $Z_1 \neq \emptyset$. Therefore, D is not minimal, contrary to our assumption. \square

Theorem 6 *Let M be a graphoid over U . If two elements of U are disconnected in some minimal Bayesian network of M , then they are disconnected in every minimal Bayesian network of M .*

Proof: It suffices to show that any two minimal Bayesian networks of M share the same maximal connected components. Let D_A and D_B be two minimal Bayesian networks of M . Let C_A and C_B be maximal connected components of D_A and D_B , respectively. Let A and B be the nodes of C_A and C_B , respectively. We show that either $A = B$ or $A \cap B = \emptyset$. This demonstration will complete the proof, because for an arbitrary maximal connected component C_A in D_A there must exist a maximal connected component in D_B that shares at least one node with C_A . Thus, by the above claim, it must have exactly the same nodes as C_A . Therefore, each maximal connected component of D_A shares the same nodes with exactly one maximal connected component of D_B . Hence, D_A and D_B share the same maximal connected components.

Since D_A is a minimal Bayesian network of M and C_A is a maximal connected component of D_A , by Lemma 2, $(A, U \setminus A \mid \emptyset) \in M$. Using symmetry and decomposition, $(A \cap B, B \setminus A \mid \emptyset) \in M$. Thus, by Lemma 5, for C_B to be a maximal connected component, either $A \cap B$ or $B \setminus A$ must be empty, lest D_B would not be minimal. Similarly, for C_A to be a maximal connected component, $A \cap B$ or $A \setminus B$ must be empty. Thus, either $A = B$ or $A \cap B = \emptyset$. \square

Theorem 7 *Two variables x and y of a graphoid M over U are unrelated iff they are uncoupled.*

Proof: If x and y are unrelated, then let U_1 be the variables connected to x in some minimal network of M , and U_2 be the rest of the variables in U . By Lemma 2, $(U_1, U_2 \mid \emptyset) \in M$. Thus, x and y are uncoupled.

If x and y are uncoupled, then there exist a partition U_1, U_2 of U such that $x \in U_1$, $y \in U_2$, and $(U_1, U_2 \mid \emptyset) \in M$. We show that in every minimal Bayesian network D of M , nodes x and y do not reside in the same connected component. Thus, x and y are unrelated. Assume, to the contrary, that x and y reside in the same maximal connected component of a minimal Bayesian network D of M , and that X are the nodes in that component. The statement $(U_1 \cap X, U_2 \cap X \mid \emptyset)$ follows from $(U_1, U_2 \mid \emptyset) \in M$ by the symmetry and decomposition axioms. Moreover, $U_1 \cap X$ and $U_2 \cap X$ are not empty, because they include x and y , respectively. Since U_1 and U_2 are disjoint, the two sets $U_1 \cap X$, $U_2 \cap X$ partition X . Therefore, by Lemma 5, D cannot be minimal, contrary to our assumption. \square

Theorem 7 shows that x and y are related if and only if they are coupled.

4 Transitive Graphoids

In this section, we show that if x and y are coupled, then x and y are mutually-relevant. Then, we identify conditions under which the converse holds, and provide an example in which these conditions are not met.

Theorem 8 *Let M be a graphoid over U . Then, for every $x, y \in U$,*

$$\text{relevant}(x, y) \Rightarrow \text{coupled}(x, y).$$

Proof: Suppose x and y are not coupled. Let U_1, U_2 be a partition of U such that $x \in U_1, y \in U_2$ and $(U_1, U_2 \mid \emptyset) \in M$. We show that x and y must be mutually irrelevant. Let Z be an arbitrary subset of $U \setminus \{x, y\}$. Let $Z_1 = Z \cap U_1$ and $Z_2 = Z \cap U_2$. The statement $(U_1, U_2 \mid \emptyset) \in M$ implies—by decomposition and symmetry axioms—that $(\{x\} \cup Z_1, \{y\} \cup Z_2 \mid \emptyset) \in M$. By symmetry and weak union, $(x, y \mid Z_1 \cup Z_2) \in M$. Thus, $(x, y \mid Z) \in M$ for every $Z \subseteq U \setminus \{x, y\}$. Hence, x and y are mutually irrelevant. \square

The converse of Theorem 8 does not hold in general; if x and y are mutually irrelevant, it does not imply that x and y are uncoupled. For example, assume M is a graphoid over $U = \{x, y, z\}$ that consists of $(x, y \mid \emptyset)$, $(x, y \mid z)$ and the statements implied from them by the graphoid axioms. Then, x and y are mutually irrelevant, yet x and y are coupled because $(x, \{y, z\} \mid \emptyset) \notin M$ and $(\{x, z\}, y \mid \emptyset) \notin M$.

To see that there is a probability distribution that induces this graphoid, suppose x and y are the outcomes of two independent fair coins. In addition, suppose that z is a variable whose domain is $\{head, tail\} \times \{head, tail\}$ and

whose value is (i, j) if and only if the outcome of x is i and the outcome of y is j . Then x and y are mutually irrelevant, because x and y are marginally independent and independent given z . Nevertheless, they are coupled, because neither $I(x, \{y, z\} \mid \emptyset)$ nor $I(\{x, z\}, y \mid \emptyset)$ hold for P .

A necessary and sufficient condition for the converse of Theorem 8 to hold, as we shall see, is that the graphoid M is transitive.

Definition A graphoid M over U is *transitive* if for every $x, y, z \in M$,

$$\text{relevant}(x, y) \ \& \ \text{relevant}(y, z) \Rightarrow \text{relevant}(x, z). \quad (8)$$

First, we show that transitivity is necessary.

Theorem 9 *Let M be a graphoid over U such that for every $x, y \in U$, $\text{coupled}(x, y)$ implies $\text{relevant}(x, y)$. Then M is a transitive graphoid.*

Proof: By Lemma 8, $\text{relevant}(x, y)$ if and only if $\text{coupled}(x, y)$. Also, by Theorem 7, $\text{coupled}(x, y)$ if and only if $\text{related}(x, y)$. Since related is a transitive relation, so is relevant . Thus, M is transitive. \square

Some preliminaries are needed before we show that transitivity is a sufficient condition as well.

Definition Let M be a graphoid over U , and A, B be two disjoint subsets of U . Then A and B are *mutually irrelevant*, if $(A, B \mid Z) \in M$ for every Z that is a subset of $U \setminus A \cup B$.

Lemma 10 *Let M be a graphoid over U , and A, B , and C be three disjoint subsets of U . If A and B are mutually irrelevant, and A and C are mutually irrelevant, then A and $B \cup C$ are mutually irrelevant as well.*

Proof: Denote the sentence “ X and Y are mutually irrelevant” with $J(X, Y)$. By definition, $J(A, B)$ implies $(A, B \mid Z) \in M$ and $J(A, C)$ implies $(A, C \mid Z \cup B) \in M$, where Z is an arbitrary subset of $U \setminus A \cup B \cup C$. Together, these statements imply by the contraction axiom that $(A, B \cup C \mid Z) \in M$. Since Z is arbitrary, $J(A, B \cup C)$ holds. \square

As is well known from probability theory, if A and B are independent, and A and C are independent, then, contrary to our intuition, A is not necessarily independent of $B \cup C$. Lemma 10, on the other hand, shows that if A and B are mutually irrelevant, and A and C are mutually irrelevant, then A and $B \cup C$ must also be mutually irrelevant.

Theorem 11 *If M is a transitive graphoid over U , then for every $x, y \in U$,*

$$\text{coupled}(x, y) \Rightarrow \text{relevant}(x, y).$$

Proof: Let M be a transitive graphoid over U , and x, y be two arbitrary elements in U such that x and y are mutually irrelevant. We will show by induction on $|U|$ that if *relevant* is transitive, then there exists a Bayesian network D of M where x and y are disconnected. Consequently, x and y are uncoupled (Theorem 7).

We construct D in the ordering $u_1(\equiv x), u_2(\equiv y), u_3, \dots, u_{n-1}, u_n(\equiv e)$ of U . Assume $n = 2$. Variables x and y are mutually irrelevant. Thus, $(x, y \mid \emptyset) \in M$. Hence, x and y are not connected. Otherwise, $n > 2$.

Let M_e be a dependency model over $U \setminus \{e\}$ formed from M by removing all triplets involving e . The model M_e is a graphoid, because whenever the left hand side of one of the graphoid axioms does not mention e , then neither does the right hand side. Let D_e be a minimal Bayesian network of M_e

formed from M_e by the construction order u_1, \dots, u_{n-1} . Let A be the set of nodes connected to x , let B be the set of nodes connected to y , and let C be the rest of the nodes in D_e . The Bayesian network D of M is formed from D_e by adding the last node e as a sink and letting its parents be a minimal set that makes e independent of all the rest of the variables in U (following the definition of minimal Bayesian networks).

Since x and y are mutually irrelevant in M , it follows that they are also mutually irrelevant in M_e . Thus, by the induction hypothesis, x and y are disconnected in D_e . After node e is added, a trail through e might exist in D that connects a node in A and a node in B . We will show that there is none; if the parent set of e is minimal, then either e has no parents in A or it has no parents in B , rendering x and y disconnected in D .

Since x and y are mutually irrelevant, it follows that either x and e are mutually irrelevant or y and e are mutually irrelevant, lest M would not be transitive. Without loss of generality, assume that x and e are mutually irrelevant. Let x' be an arbitrary node in A . By transitivity it follows that either x and x' are mutually irrelevant or e and x' are mutually irrelevant, lest x and e would not be mutually irrelevant, contrary to our selection of x . If x and x' are mutually irrelevant, then by the induction hypothesis, A can be partitioned into two marginally independent subsets. Thus, by Lemma 5, A would not be connected in the Bayesian network D_e , contradicting our selection of A . Thus, every element $x' \in A$ and e are mutually irrelevant. It follows that the entire set A and e are mutually irrelevant (Lemma 10). Thus, in particular, $(e, A \mid \hat{B} \cup \hat{C}) \in M$, where \hat{B} are the parents of e in B , and \hat{C} are the parents of e in C . Assume \hat{A} is the set of parents of e in

A. By decomposition, $(e, A \mid \hat{B} \cup \hat{C}) \in M$ implies $(e, \hat{A} \mid \hat{B} \cup \hat{C}) \in M$. By Theorem 4, D is not minimal, unless \hat{A} is empty. \square

Theorems 8, 9 and 11 show that the relations *coupled* and *relevant* are identical for every transitive graphoid and for none other. We emphasize that these results apply also to every probability distribution that defines a transitive graphoid. In section 6, we show that many probability distributions indeed define transitive graphoids. First, however, we pause to demonstrate the relationship of these results to knowledge acquisition and knowledge representation.

5 Similarity Networks

Similarity networks were invented by Heckerman [1990] as a tool for constructing large Bayesian networks from domain experts judgements. Heckerman used them to construct a large diagnosis system for lymph-node pathology. The main advantage of similarity networks is their ability to utilize statements of conditional independence that are not represented in a Bayesian network, in order to reduce more drastically the number of parameters a domain expert needs to specify. Furthermore, the construction of a large Bayesian network is divided into several stages each of which involves the construction of a small local Bayesian network. This divide and conquer approach helps to elicit reliable expert judgements. At the diagnosis stage, the local networks are combined into one global Bayesian network that represents the entire domain.

In [Geiger and Heckerman, 1993], we show how to use the local networks directly for inference without converting them to a global Bayesian network,

and remove several technical restrictions imposed by the original development. Also, we develop two simple definitions of similarity networks which we present here informally. In the next section, we show that although the two definitions are conceptually distinct they often coincide.

A Bayesian network of a probability distribution $P(u_1, \dots, u_n)$ is constructed as defined in Section 2 with an important addition. After the topology of the network is set, we also associate with each node a conditional probability distribution: $P(u_i | \pi(u_i))$. By the chaining rule it follows that

$$P(u_1, \dots, u_n) = \prod P(u_i | u_1, \dots, u_{i-1})$$

and by the definition of $I(\{u_i\}, \{u_1, \dots, u_{i-1}\} \setminus \pi(u_i) | \pi(u_i))$ we further obtain

$$P(u_1, \dots, u_n) = \prod P(u_i | \pi(u_i))$$

Thus, the joint distribution is represented by the network and can be used for computing the posterior probability of every variable, given a specific value for some other variables. For example, for the network of the burglary story (Figure 2), we need to specify the following conditional distributions: $P(burglary)$, $P(sensorA | burglary)$, $P(sensorB | burglary)$, $P(alarm | sensorA, sensorB)$, and $P(patrol | alarm)$. From these numbers, we can now compute any probability involving these variables.

A similarity network is a set of Bayesian networks, called the local networks, each constructed under a different set of hypotheses H_i . In each local network D_i , only those variables that “help to distinguish” between the hypotheses in H_i are depicted. The success of this model stems from the fact that only a small portion of variables helps to distinguish between the

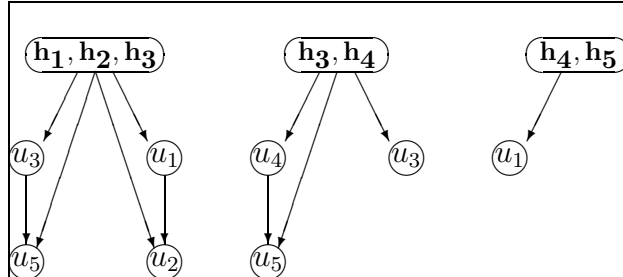


Figure 2: A similarity network

carefully chosen set of hypotheses $H_i, i = 1 \dots k$. Thus, the model usually includes several small networks instead of one large Bayesian network.

For example, Figure 2 is an example of a similarity network representation of $P(h, u_1, \dots, u_5)$ where h is a distinguished variable that represents five hypotheses $\mathbf{h}_1, \dots, \mathbf{h}_5$. In this similarity network, variable u_1 is the only one that helps to discriminate between \mathbf{h}_4 and \mathbf{h}_5 , and variable u_4 is the only variable that does not help to discriminate among $\{\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3\}$.

At the heart of the definition of similarity networks lies the notion of discrimination. The study of the relations *coupled*, *related* and *relevant* presented in the previous sections, enables us to formulate this notion in two ways, yielding two types of similarity networks.

Definition [Geiger and Heckerman, 1993] A similarity network constructed by including in each local network D_i every variable x , such that x and h are *related* given that h draws its values from H_i , is of *type 1*. A similarity network constructed by including in each local network D_i every variable x , such that x and h are *relevant* given that h draws its values from H_i , is of *type 2*.

In [Geiger and Heckerman, 1993], we show that type 1 similarity networks are diagnostically complete. That is, although some variables are removed from each local network, the posterior probability of every hypothesis, given any value combination for the variables in U , can still be computed. This result is reassuring because it guarantees that the computation we strive to achieve—namely, the computation of the posterior probability of the hypothesis— can be performed. The caveat of this result is that a knowledge engineer uses a type 1 similarity network to determine whether a variable “helps to discriminate” the values in H_i , by asking a domain expert whether the node corresponding to this variable is connected to h in the local Bayesian network associated with H_i . This query might be too hard for a domain expert to answer, because a domain expert does not necessarily understand the properties of Bayesian networks.

On the other hand, a knowledge engineer uses a type 2 similarity networks to determine whether a node “does not help to discriminate” the values in H_i , by asking an expert whether this variable can ever help to distinguish the values of h , given that h draws its values from H_i . This query concerns the subject matter of the domain; and therefore a domain expert can more reliably answer the query. In fact, this is the actual query Heckerman used in constructing his lymph-node pathology diagnosis system.

Next, we show that these two definitions coincide for large families of probability distributions.

6 Transitive Distributions

We show that the relation *relevant* is transitive whenever it is defined by a probability distribution that belongs to one of the following two families: strictly positive binary distributions and regular Gaussian distributions. Hence, for these two classes of distributions, type 1 and type 2 similarity networks are identical. Currently, we are working to show that transitivity holds for other families.

Definition A *strictly positive binary* distribution is a probability distribution where every variable has a domain of two values—say, 0 and 1—and every combination of the variables’ values has a probability greater than zero. A *regular Gaussian* distribution is a multivariate normal distribution with finite nonzero variances and with finite means.

Theorem 12 Let $P(u_1, \dots, u_n, e)$ be a strictly positive binary distribution or a regular Gaussian distribution. Let $\{X_1, X_2\}$, $\{Y_1, Y_2\}$ and $\{Z_1, Z_2\}$ be three partitions of $U = \{u_1, \dots, u_n\}$. Let R_1 be $X_1 \cap Y_1 \cap Z_1$, and R_2 be $X_2 \cap Y_2 \cap Z_2$. Then,

$$I(X_1, X_2 \mid \emptyset) \ \& \ I(Y_1, Y_2 \mid e = \mathbf{e}') \ \& \ I(Z_1, Z_2 \mid e = \mathbf{e}'') \Rightarrow \\ I(R_1, \{e\} \cup U \setminus R_1 \mid \emptyset) \ \vee \ I(R_2, \{e\} \cup U \setminus R_2 \mid \emptyset) \quad (9)$$

where \mathbf{e}' and \mathbf{e}'' are two distinct values of e .

When all three partitions are identical, the above theorem can be phrased as follows. If two sets of variables A and B are marginally independent, and

if $I(A, B | e)$ holds as well, then either A is marginally independent of $\{e\} \cup B$ or B is marginally independent of $\{e\} \cup A$. This special case has been stated in the literature [Dawid, 1979, Pearl, 1988].

The proof of Theorem 12 is given in Appendices A and B. Theorem 12 and the theorem below state together that strictly positive binary distributions and regular Gaussian distributions are transitive. Our assumptions of strict positiveness and regularity were added to obtain a simpler proof. We conjecture that both theorems still hold when these restrictions are omitted.

Theorem 13 *Every probability distribution that satisfies Equation 9 is transitive.*

Proof: Let $P(u_1, \dots, u_{n+1})$ be a probability distribution, let $U = \{u_1, \dots, u_{n+1}\}$; and let x, y be two arbitrary variables in U such that x and y are mutually irrelevant. We will show by induction on $|U|$ that if P satisfies Equation 9, then x and y are uncoupled. Thus, according to Theorem 9, P is transitive.

If $n = 1$, then the variables x and y are mutually irrelevant. Thus, $I(x, y | \emptyset)$ holds for P . Consequently, x and y are uncoupled. Otherwise, assume without loss of generality that x is u_1 and y is u_2 , and denote u_{n+1} by e . Since x and y are mutually irrelevant with respect to $P(u_1, \dots, u_{n+1})$, x and y are also mutually irrelevant with respect to $P(u_1, \dots, u_n)$, $P(u_1, \dots, u_n | e = \mathbf{e}')$, and $P(u_1, \dots, u_n | e = \mathbf{e}'')$, where \mathbf{e}' and \mathbf{e}'' are two distinct values of u_{n+1} . Thus, by applying the the induction hypothesis three times, we conclude that there are three partitions $\{X_1, X_2\}$, $\{Y_1, Y_2\}$, and $\{Z_1, Z_2\}$ of $U = \{u_1, \dots, u_n\}$ such that x is in X_1 , Y_1 , and Z_1 , and y is in X_2 , Y_2 , and Z_2 . Hence, the antecedents of Equation 9 are

satisfied. Consequently, $\{u_1, \dots, u_{n+1}\}$ can be partitioned into two marginally independent sets: either R_1 and $U \setminus R_1$, or R_2 and $U \setminus R_2$, where R_1 is $X_1 \cap Y_1 \cap Z_1$ and R_2 is $X_2 \cap Y_2 \cap Z_2$. Because, in both cases, one set contains x and the other contains y , it follows that x and y are uncoupled. \square

The practical ramification of this theorem is that our concern of how to define discrimination via the relation *related* or via *relevant* is not critical. In many situations the two concepts coincide.

From a mathematical point of view, our proof demonstrates that using an abstraction of conditional independence—namely, the ternary relation I combined with a set of axioms—we are able to prove properties of very distinct classes of distributions: strictly positive binary distributions and regular Gaussian distributions.

7 Summary

We have examined the notion of unrelatedness of variables in a probabilistic framework, introduced three formulations of this notion, and explored their interrelationships. From a practical point of view, these results legitimize prevailing decomposition techniques of knowledge acquisition. These results permit an expert to decompose the construction of a complex Bayesian network into a set of Bayesian networks of manageable size.

Our proofs use the qualitative notion of independence as captured by the axioms of graphoids. These proofs would have been harder to obtain had we used the usual definitions of conditional independence. This axiomatic approach enables us to identify a common property—Equation 9—shared by two distinct classes of probability distributions (regular Gaussian and strictly

positive binary), and to use this property without attending to the detailed characteristics of these classes.

Acknowledgments

An earlier version of this article has been presented in the sixth conference on uncertainty in artificial intelligence [Geiger and Heckerman, 1990].

References

- [Dawid, 1979] Dawid, A.P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society B* 41, 1-31.
- [Geiger, 1990] Geiger D. 1990. Graphoids: A Qualitative Framework for Probabilistic Inference. PhD. Diss., Computer Science Dept., UCLA.
- [Geiger and Heckerman, 1993] Knowledge representation and inference in similarity networks and Bayesian multinets. Submitted.
- [Geiger and Heckerman, 1990] Geiger D., and Heckerman, D. (1990). Separable and transitive graphoids. Proc. Sixth Conference on Uncertainty in Artificial Intelligence.
- [Geiger et al., 1990] Geiger, D.; Verma, T.S.; and Pearl, J. 1990. Identifying Independence in Bayesian Networks. *Networks*, 20:507-534.
- [Geiger and Pearl, 1990] Geiger D.; and Pearl J. 1990. On the logic of Causal Models. In *Uncertainty in Artificial Intelligence 4*, ed-

- s. Shachter R. D.; Levitt T.S.; Kanal L.N.; and Lemmer J.F., 3–12. Elsevier Science Publishers B.V. (North-Holland).
- [Heckerman, 1990] Heckerman, D. 1990. Probabilistic Similarity Networks. MIT press.
- [Lauritzen et al., 1992] Lauritzen, S.L.; Dawid, A. P.; Larsen B. N. and Leimer H. G. Independence properties of directed Markov fields. *Networks*, 20: 491-506.
- [Pearl, 1988] Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
- [Pearl and Paz, 1989] Pearl, J. and Paz, A. 1989. Graphoids: A Graph-Based Logic for Reasoning about Relevance Relations. In *Advances in Artificial Intelligence-II*, eds. Du Boulay, B. et al., 357–363. Amsterdam: North Holland.
- [Shachter, 1992] An ordered examination of influence diagrams. *Networks*, 20:535-564.
- [Spohn, 1980] Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *J. of Philosophical logic* 9 73-99.
- [Verma and Pearl, 1988] Verma, T. and Pearl, J. (1988). Causal networks: Semantics and expressiveness. In *Proceedings of Fourth Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN, pages 352–359. Association for Uncertainty in Artificial Intelligence, Mountain View, CA. Also, Technical Report R-65

by Verma, Cognitive Systems Laboratory, University of California, Los Angeles, 1986.

Appendix A: Strictly Positive Binary Distributions

Below, we prove Theorem 12 for strictly positive binary distributions. First, we phrase the theorem differently.

Theorem 14 *Strictly positive binary distributions satisfy the following axiom.³*

$$\begin{aligned}
 & I(A_1A_2A_3A_4, B_1B_2B_3B_4 \mid \emptyset) \ \& \\
 & I(A_1A_2B_3B_4, B_1B_2A_3A_4 \mid e = \mathbf{e}') \ \& \\
 & I(A_1A_3B_2B_4, B_1B_3A_2A_4 \mid e = \mathbf{e}'') \Rightarrow \\
 & I(A_1, eA_2A_3A_4B_1B_2B_3B_4 \mid \emptyset) \vee I(B_1, eA_1A_2A_3A_4B_2B_3B_4 \mid \emptyset) \quad (10)
 \end{aligned}$$

where all sets mentioned are pairwise disjoint and do not contain e , and \mathbf{e}' and \mathbf{e}'' are distinct values of e .

To obtain the original theorem, we set $A_1A_2A_3A_4$, $B_1B_2B_3B_4$, $A_1A_2B_3B_4$, $B_1B_2A_3A_4$, $A_1A_3B_2B_4$, and $B_1B_3A_2A_4$ to be equal to X_1, X_2, Y_1, Y_2, Z_1 , and Z_2 of the original theorem, respectively.

Denote the three antecedents of Equation 10 by I_1 , I_2 , and I_3 . We need the following two Lemmas.

Lemma 15 *Let X and Y be two disjoint sets of variables, and let \mathbf{e} be an instance of a single binary variable e not in $X \cup Y$. Let P be a probability distribution over the variables $X \cup Y \cup \{e\}$. If $I(X, Y \mid e = \mathbf{e})$ holds for P ,*

³In complicated expressions, A_1A_2 is used as a shorthand notation for $A_1 \cup A_2$ and eA_1 denotes $\{e\} \cup A_1$.

then for every pair of instances $\mathbf{X}', \mathbf{X}''$ of X and $\mathbf{Y}', \mathbf{Y}''$ of Y , the following equation must hold:

$$\frac{P(\mathbf{e}|\mathbf{X}'\mathbf{Y}')P(\mathbf{X}'\mathbf{Y}')}{P(\mathbf{e}|\mathbf{X}''\mathbf{Y}')P(\mathbf{X}''\mathbf{Y}')} = \frac{P(\mathbf{e}|\mathbf{X}'\mathbf{Y}'')P(\mathbf{X}'\mathbf{Y}'')}{P(\mathbf{e}|\mathbf{X}''\mathbf{Y}'')P(\mathbf{X}''\mathbf{Y}'')}$$

Proof: Bayes' theorem states that

$$P(\mathbf{X}'|\mathbf{e}\mathbf{Y}') = \frac{P(\mathbf{e}|\mathbf{X}'\mathbf{Y}')P(\mathbf{X}'\mathbf{Y}')}{P(\mathbf{e}\mathbf{Y}')}$$

Thus,

$$\frac{P(\mathbf{e}|\mathbf{X}'\mathbf{Y}')P(\mathbf{X}'\mathbf{Y}')}{P(\mathbf{e}|\mathbf{X}''\mathbf{Y}')P(\mathbf{X}''\mathbf{Y}')} = \frac{P(\mathbf{X}'|\mathbf{e}, \mathbf{Y}')}{P(\mathbf{X}''|\mathbf{e}, \mathbf{Y}')} = \frac{P(\mathbf{X}'|\mathbf{e}, \mathbf{Y}'')}{P(\mathbf{X}''|\mathbf{e}, \mathbf{Y}'')} = \frac{P(\mathbf{e}|\mathbf{X}'\mathbf{Y}'')P(\mathbf{X}'\mathbf{Y}'')}{P(\mathbf{e}|\mathbf{X}''\mathbf{Y}'')P(\mathbf{X}''\mathbf{Y}'')}$$

The middle equality follows from the fact that $I(X, Y | e = \mathbf{e})$ holds for P .

□

Lemma 16 *Let $A_1, A_2, A_3, A_4, B_1, B_2, B_3,$ and B_4 be disjoint sets of variables, and e be a single binary variable not contained in any of these sets. Let P be a probability distribution over the union of these variables. If the antecedents $I_1, I_2,$ and I_3 of Equation 10 hold for P , then the following conditions must also hold:*

$$I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \Rightarrow I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4B_1B_2\mathbf{B}'_3\mathbf{B}'_4) \quad (11)$$

$$I(B_1, e | \mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \Rightarrow I(B_1, e | A_1A_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \quad (12)$$

$$I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \Rightarrow I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4B_1\mathbf{B}'_2B_3\mathbf{B}'_4) \quad (13)$$

$$I(B_1, e | \mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \Rightarrow I(B_1, e | A_1\mathbf{A}'_2A_3\mathbf{A}'_4\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \quad (14)$$

$$I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \Rightarrow I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3A_4B_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \quad (15)$$

$$I(B_1, e | \mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \Rightarrow I(B_1, e | A_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_2\mathbf{B}'_3B_4) \quad (16)$$

where each \mathbf{A}'_i and \mathbf{B}'_i denote a specific value for A_i and B_i , respectively. (In words, Equation 11 states that if A_1 and e are conditionally independent for one specific value \mathbf{B}'_1 of B_1 and \mathbf{B}'_2 of B_2 , then they are conditionally independent given every value of B_1 and B_2 , provided the values of the other variables remain unaltered. The other five equations have a similar interpretation.)

Proof: First, we prove Equation 11. Then we show that the proofs of Equations 12 through 14 are symmetric. Finally, we will prove Equations 15 and 16. Let $X = A_1A_2B_3B_4$ and $Y = B_1B_2A_3A_4$. Then, Lemma 15 and I_2 yield the following equation:

$$\begin{aligned} & \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \\ &= \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \end{aligned}$$

where \mathbf{A}'_1 , \mathbf{B}'_1 , and \mathbf{B}'_2 are arbitrary instances of A_1 , B_1 , and B_2 , respectively. Applying I_1 and cancelling equal terms yields

$$\frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} = \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \quad (17)$$

Furthermore, $I(A_1, e \mid \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)$ (the antecedent of Equation 11) implies that

$$P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) = P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)$$

Thus, from Equation 17, it follows that

$$P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) = P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4) \quad (18)$$

Subtracting each side of Equation 18 from 1 yields

$$P(\bar{e}|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}'_3\mathbf{B}'_4) = P(\bar{e}|\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}'_3\mathbf{B}'_4) \quad (19)$$

Thus, $I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}'_3\mathbf{B}'_4)$ holds for P . Because \mathbf{B}_1^* and \mathbf{B}_2^* are arbitrary instances, $I(A_1, e | \mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4B_1B_2\mathbf{B}'_3\mathbf{B}'_4)$ also holds for P . Thus, Equation 11 is proved.

Equation 12 is symmetric with respect to Equation 11 by switching the role of A_1 with that of B_1 and the role of A_2 with that of B_2 . Equation 13 is symmetric with respect to Equation 11 by switching the roles of B_2 and B_3 . Equation 14 is symmetric with respect to Equation 12 by switching the roles of A_2 and A_3 .

Now we prove Equation 15. Equation 16 is symmetric with respect to Equation 15 by switching the role of A_1 with that of B_1 and the role of A_4 with that of B_4 .

Let $X = A_1A_2B_3B_4$ and $Y = B_1B_2A_3A_4$. Applying Lemma 15 and I_2 and then using I_1 to cancel equal terms, yields the following equation:

$$\begin{aligned} & \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4)}{P(e|\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4)} \\ &= \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}_4^*\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}_4^*)}{P(e|\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}_4^*\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}_4^*)} \end{aligned} \quad (20)$$

where \mathbf{A}_1^* , \mathbf{B}_1^* , and \mathbf{A}_4^* are arbitrary instances of A_1 , B_1 , and A_4 , respectively. Similarly, let $X = A_1A_3B_2B_4$ and $Y = B_1B_3A_2A_4$. Then, applying Lemma 15 and I_3 and using I_1 to cancel equal terms, yields the following equation:

$$\frac{P(\bar{e}|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4)}{P(\bar{e}|\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}_1^*\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4)}$$

$$= \frac{P(\bar{e} | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) P(\mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4)}{P(\bar{e} | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) P(\mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4)} \quad (21)$$

Now $I(A_1, e | \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)$ implies the following two conditions:

$$P(e | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) = P(e | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (22)$$

$$P(\bar{e} | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) = P(\bar{e} | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (23)$$

After using Equation 22 to cancel equal terms in Equation 20 and using Equation 23 to cancel equal terms in Equation 21, we compare Equations 20 and 21 and obtain

$$\frac{P(e | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)}{P(e | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)} = \frac{P(\bar{e} | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)}{P(\bar{e} | \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)} \quad (24)$$

Equation 24 has the form:

$$\frac{x}{y} = \frac{1-x}{1-y}$$

which yields $x = y$.

Consequently, we obtain $I(A_1, e | \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)$. Furthermore, because \mathbf{A}'_4 and \mathbf{B}'_1 are arbitrary instances, $I(A_1, e | \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_1 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4)$ holds for P . \square

Next, we prove Theorem 12. Let $C = A_2 A_3 A_4$ and $D = B_2 B_3 B_4$. We will see that I_1 , I_2 , and I_3 imply the following four properties:

$$I(A_1, e | CDB_1) \text{ or } I(B_1, e | CDA_1) \quad (25)$$

$$I(A_1, e | CDB_1) \Rightarrow I(A_1, A_4 | A_2 A_3) \quad (26)$$

$$I(A_1, e | CDB_1) \& I(A_1, A_4 | A_2 A_3) \Rightarrow I(A_1, A_3 | A_2) \quad (27)$$

$$I(A_1, e | CDB_1) \& I(A_1, A_4 | A_2 A_3) \& I(A_1, A_3 | A_2) \Rightarrow I(A_1, A_2 | \emptyset) \quad (28)$$

First, we prove Equation 10, using these four properties. Then, we will show that these properties are valid. From Equation 25, there are two symmetric cases to consider. Without loss of generality, assume $I(A_1, e \mid CDB_1)$ holds. (Otherwise, we switch the roles of subscripted A 's with subscripted B 's in Equations 26 through 28.) By a single application of each of Equations 26, 27, and 28, the following independence statements are proved to hold for P :

$$I(A_1, A_2 \mid \emptyset), I(A_1, A_3 \mid A_2), I(A_1, A_4 \mid A_2A_3)$$

These three statements yield $I(A_1, A_2A_3A_4 \mid \emptyset)$ ($\equiv I_4$) by two applications of contraction. Consider Equation 10. The statement $I(A_1A_2A_3A_4, B_1B_2B_3B_4 \mid \emptyset)$ (i.e., I_1) implies $I(A_1, B_1B_2B_3B_4 \mid A_2A_3A_4)$ using weak union, which together with I_4 imply using contraction $I(A_1, A_2A_3A_4B_1B_2B_3B_4 \mid \emptyset)$. This statement together with the statement $I(A_1, e \mid CB_1D)$ imply, using contraction, the statement $I(A_1, eA_2A_3A_4B_1B_2B_3B_4 \mid \emptyset)$, thus completing the proof.

It remains to prove Equations 25 through 28. First, we prove Equation 25. Let \mathbf{A}' , \mathbf{A}'' , \mathbf{B}' , \mathbf{B}'' , \mathbf{C}^* , and \mathbf{D}^* be arbitrary instances of A_1 , B_1 , C , and D , respectively. Let $X = AC$ and $Y = BD$. Then, Lemma 15 and I_2 yield the following equation:

$$\frac{P(e|\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}')P(\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}')}{P(e|\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}')P(\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}')} = \frac{P(e|\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')P(\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')}{P(e|\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')P(\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')} \quad (29)$$

From I_1 , we obtain $P(ACDB) = P(AC)P(DB)$. Consequently, Equation 29 yields

$$\frac{P(e|\mathbf{A}'\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)}{P(e|\mathbf{A}''\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)} = \frac{P(e|\mathbf{A}'\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)}{P(e|\mathbf{A}''\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)} \quad (30)$$

Equation 30 has the following algebraic form, where subscripted X s replace the corresponding terms:

$$\frac{X_{\mathbf{A}'\mathbf{B}'}}{X_{\mathbf{A}''\mathbf{B}'}} = \frac{X_{\mathbf{A}'\mathbf{B}''}}{X_{\mathbf{A}''\mathbf{B}''}} \quad (31)$$

Using Lemma 15 and I_3 , we obtain a relationship similar to Equation 30, where the only change is that \mathbf{e} is replaced with $\bar{\mathbf{e}}$:

$$\frac{P(\bar{\mathbf{e}}|\mathbf{A}'\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)}{P(\bar{\mathbf{e}}|\mathbf{A}''\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)} = \frac{P(\bar{\mathbf{e}}|\mathbf{A}'\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)}{P(\bar{\mathbf{e}}|\mathbf{A}''\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)} \quad (32)$$

We rewrite Equation 32 in terms of X s, and then use Equation 31 to obtain

$$\frac{1 - X_{\mathbf{A}'\mathbf{B}'}}{1 - X_{\mathbf{A}''\mathbf{B}'}} = \frac{1 - kX_{\mathbf{A}'\mathbf{B}'}}{1 - kX_{\mathbf{A}''\mathbf{B}'}} \quad (33)$$

where $k = \frac{X_{\mathbf{A}'\mathbf{B}''}}{X_{\mathbf{A}'\mathbf{B}'}}$. Equation 33 implies that either $X_{\mathbf{A}'\mathbf{B}''} = X_{\mathbf{A}'\mathbf{B}'}$ (i.e., $k = 1$) or $X_{\mathbf{A}'\mathbf{B}'} = X_{\mathbf{A}''\mathbf{B}'}$. Because the choice of instances for A_1 and B_1 is arbitrary, at least one of the following two sequences of equalities must hold:

- For every instance \mathbf{B} of B_1 , $X_{\mathbf{A}^1\mathbf{B}} = X_{\mathbf{A}^2\mathbf{B}} = \dots = X_{\mathbf{A}^m\mathbf{B}}$
- For every instance \mathbf{A} of A_1 , $X_{\mathbf{A}\mathbf{B}^1} = X_{\mathbf{A}\mathbf{B}^2} = \dots = X_{\mathbf{A}\mathbf{B}^n}$

where $\mathbf{A}^1, \dots, \mathbf{A}^m$ are the instances of A_1 and $\mathbf{B}^1, \dots, \mathbf{B}^n$ are the instances of B_1 .

Thus, by definition of the X s, we obtain

$$\forall \mathbf{C}^*\mathbf{D}^* \text{ instances of } CD \quad [I(e, A_1 \mid \mathbf{C}^*\mathbf{D}^*B_1) \text{ or } I(e, B_1 \mid \mathbf{C}^*\mathbf{D}^*A_1)] \quad (34)$$

On the other hand, Equation 25, which we are now proving, states

$$[\forall \mathbf{C}^*\mathbf{D}^* I(e, A_1 \mid \mathbf{C}^*\mathbf{D}^*B_1)] \text{ or } [\forall \mathbf{C}^*\mathbf{D}^* I(e, B_1 \mid \mathbf{C}^*\mathbf{D}^*A_1)] \quad (35)$$

which is stronger than Equation 34. Equation 25 can also be written as follows:

$$\neg I(B_1, e \mid CDA_1) \Rightarrow I(A_1, e \mid CDB_1) \quad (36)$$

We prove Equation 36. The statement $\neg I(B_1, e \mid CDA_1)$ implies that there exists instances $\mathbf{A}'_1, \mathbf{A}'_2, \mathbf{A}'_3, \mathbf{A}'_4, \mathbf{B}'_1, \mathbf{B}'_2, \mathbf{B}'_3, \mathbf{B}'_4$, and \mathbf{e}' of $A_1, A_2, A_3, A_4, B_1, B_2, B_2, B_3, B_4$, and e , respectively, such that

$$\neg I(\mathbf{B}'_1, \mathbf{e}' \mid \mathbf{A}'_1 \mathbf{A}'_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (37)$$

Hence,

$$\neg I(B_1, e \mid A_1 A_2 \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (38)$$

From Lemma 16 (contrapositive form of Equation 12), Equation 38 implies

$$\neg I(B_1, e \mid \mathbf{A}_1^* \mathbf{A}_2^* \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (39)$$

where \mathbf{A}_1^* and \mathbf{A}_2^* are arbitrary instances of A_1 and A_2 , respectively. Hence, in particular, if $\mathbf{A}_1^* = \mathbf{A}'_1$, we have

$$\neg I(B_1, e \mid \mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}'_3 \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (40)$$

Similarly, from Lemma 16 (Equation 14), Equation 40 implies

$$\neg I(B_1, e \mid \mathbf{A}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}'_4) \quad (41)$$

where \mathbf{A}_3^* is an arbitrary instance of A_3 . Also, from Lemma 16 (Equation 16), Equation 41 implies

$$\neg I(B_1, e \mid \mathbf{A}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}_4^*) \quad (42)$$

where \mathbf{B}_4^* is an arbitrary instance of B_4 . Examine Equation 34. Equation 42 states that the second disjunct cannot be true for every instance of

$A_1A_2A_3B_1B_4$ and e . Hence, for each of these instances the other disjunct must hold. That is,

$$\forall \mathbf{A}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{B}_1^* \mathbf{B}_4^* e^* I(\mathbf{A}_1^*, e^* \mid \mathbf{B}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 \mathbf{B}_4^*) \quad (43)$$

or, equivalently,

$$I(A_1, e \mid B_1 A_2 A_3 \mathbf{A}'_4 \mathbf{B}'_2 \mathbf{B}'_3 B_4) \quad (44)$$

Applying Equation 44 to Equation 11 yields,

$$I(A_1, e \mid B_1 A_2 A_3 \mathbf{A}'_4 B_2 \mathbf{B}'_3 B_4) \quad (45)$$

Similarly, applying Equation 45 to Equations 13, and 15 yields the statement

$$I(A_1, e \mid A_2 A_3 A_4 B_1 B_2 B_3 B_4) \quad (46)$$

which is the desired consequence of Equation 36. Thus, we have proved Equation 25.

Next, we show that Equation 26 must hold. Lemma 15 and I_2 yield the following equation:

$$\begin{aligned} & \frac{P(e \mid \mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) P(\mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{P(e \mid \mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) P(\mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}'_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \\ &= \frac{P(e \mid \mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}''_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) P(\mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}''_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{P(e \mid \mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}''_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) P(\mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}''_4 \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}. \end{aligned}$$

Incorporating I_1 and $I(e, A_1 \mid A_2 A_3 A_4 B_1 B_2 B_3 B_4)$ (Equation 46), and cancelling some equal terms yields

$$\begin{aligned} & \frac{P(\mathbf{A}'_4 \mid \mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{P(\mathbf{A}'_4 \mid \mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \\ &= \frac{P(\mathbf{A}''_4 \mid \mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{A}'_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{P(\mathbf{A}''_4 \mid \mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{A}''_1 \mathbf{A}_2^* \mathbf{A}_3^*) P(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}. \end{aligned}$$

Further cancelation of equal terms yields

$$\frac{P(\mathbf{A}'_4|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3)}{P(\mathbf{A}''_4|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3)} = \frac{P(\mathbf{A}'_4|\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}'_3)}{P(\mathbf{A}''_4|\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}'_3)}.$$

Thus, $P(\mathbf{A}'_4|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3) = P(\mathbf{A}'_4|\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}'_3)$ for every instance $\mathbf{A}'_1, \mathbf{A}''_1$, and \mathbf{A}'_4 . That is, $I(A_4, A_1 | \mathbf{A}'_2\mathbf{A}'_3)$ holds. Because \mathbf{A}'_2 and \mathbf{A}'_3 are arbitrary instances, $I(A_4, A_1 | A_2A_3)$ follows.

Next, we show that Equation 27 must hold. Lemma 15 and I_2 yield the following equation:

$$\begin{aligned} & \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(e|\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \\ &= \frac{P(e|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}''_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}''_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(e|\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}''_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}''_1\mathbf{A}'_2\mathbf{A}''_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \end{aligned}$$

Incorporating I_1 , $I(A_1, A_4 | A_2A_3)$, and $I(e, A_1 | A_2A_3A_4B_1B_2B_3B_4)$ and cancelling some equal terms yields

$$\begin{aligned} & \frac{P(\mathbf{A}'_4|\mathbf{A}'_2\mathbf{A}'_3)P(\mathbf{A}'_3|\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(\mathbf{A}''_4|\mathbf{A}'_2\mathbf{A}'_3)P(\mathbf{A}'_3|\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{A}''_1\mathbf{A}'_2)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \\ &= \frac{P(\mathbf{A}'_4|\mathbf{A}'_2\mathbf{A}''_3)P(\mathbf{A}''_3|\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(\mathbf{A}'_4|\mathbf{A}'_2\mathbf{A}''_3)P(\mathbf{A}''_3|\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{A}''_1\mathbf{A}'_2)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \end{aligned}$$

Further cancelation of equal terms yields

$$\frac{P(\mathbf{A}'_3|\mathbf{A}'_1\mathbf{A}'_2)}{P(\mathbf{A}''_3|\mathbf{A}'_1\mathbf{A}'_2)} = \frac{P(\mathbf{A}'_3|\mathbf{A}''_1\mathbf{A}'_2)}{P(\mathbf{A}''_3|\mathbf{A}''_1\mathbf{A}'_2)}$$

Thus, $P(\mathbf{A}'_3|\mathbf{A}'_1\mathbf{A}'_2) = P(\mathbf{A}'_3|\mathbf{A}''_1\mathbf{A}'_2)$ for every instance $\mathbf{A}'_1, \mathbf{A}''_1$ and \mathbf{A}'_3 . That is, $I(A_3, A_1 | \mathbf{A}'_2)$ holds. Because \mathbf{A}'_2 is an arbitrary instance, $I(A_3, A_1 | A_2)$ follows.

Finally, we must show that Equation 28 holds. Lemma 15 and I_3 yield the following equation:

$$\begin{aligned} & \frac{P(\bar{e}|\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}'_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(\bar{e}|\mathbf{A}''_1\mathbf{A}''_2\mathbf{A}''_3\mathbf{A}''_4\mathbf{B}''_1\mathbf{B}''_2\mathbf{B}''_3\mathbf{B}''_4)P(\mathbf{A}''_1\mathbf{A}''_2\mathbf{A}''_3\mathbf{A}''_4\mathbf{B}''_1\mathbf{B}''_2\mathbf{B}''_3\mathbf{B}''_4)} \\ &= \frac{P(\bar{e}|\mathbf{A}'_1\mathbf{A}''_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}'_1\mathbf{A}''_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(\bar{e}|\mathbf{A}''_1\mathbf{A}''_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}''_1\mathbf{B}''_2\mathbf{B}'_3\mathbf{B}'_4)P(\mathbf{A}''_1\mathbf{A}''_2\mathbf{A}'_3\mathbf{A}'_4\mathbf{B}''_1\mathbf{B}''_2\mathbf{B}'_3\mathbf{B}'_4)} \end{aligned}$$

Incorporating $I(e, A_1 | A_2A_3A_4B_1B_2B_3B_4)$, I_1 , I_3 , $I(A_1, A_4 | A_2A_3)$, and $I(A_1, A_3 | A_2)$ and cancelling some equal terms yields

$$\begin{aligned} & \frac{P(\mathbf{A}'_4|\mathbf{A}'_2 * \mathbf{A}'_3)P(\mathbf{A}'_3|\mathbf{A}'_1\mathbf{A}'_2)P(\mathbf{A}'_2|\mathbf{A}'_1)P(\mathbf{A}'_1)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(\mathbf{A}''_4|\mathbf{A}''_2 * \mathbf{A}''_3)P(\mathbf{A}''_3|\mathbf{A}''_1\mathbf{A}''_2)P(\mathbf{A}''_2|\mathbf{A}''_1)P(\mathbf{A}''_1)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \\ &= \frac{P(\mathbf{A}'_4|\mathbf{A}''_2 * \mathbf{A}'_3)P(\mathbf{A}'_3|\mathbf{A}'_1\mathbf{A}''_2)P(\mathbf{A}''_2|\mathbf{A}'_1)P(\mathbf{A}'_1)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)}{P(\mathbf{A}''_4|\mathbf{A}''_2 * \mathbf{A}''_3)P(\mathbf{A}''_3|\mathbf{A}''_1\mathbf{A}''_2)P(\mathbf{A}''_2|\mathbf{A}''_1)P(\mathbf{A}''_1)P(\mathbf{B}'_1\mathbf{B}'_2\mathbf{B}'_3\mathbf{B}'_4)} \end{aligned}$$

Further cancelation of equal terms yields

$$\frac{P(\mathbf{A}'_2|\mathbf{A}'_1)}{P(\mathbf{A}''_2|\mathbf{A}'_1)} = \frac{P(\mathbf{A}'_2|\mathbf{A}''_1)}{P(\mathbf{A}''_2|\mathbf{A}''_1)}$$

Thus, $P(\mathbf{A}'_2|\mathbf{A}'_1) = P(\mathbf{A}'_2|\mathbf{A}''_1)$ for every instance $\mathbf{A}'_1, \mathbf{A}''_1$ and \mathbf{A}'_2 . That is, $I(A_2, A_1 | \emptyset)$ holds. \square

We conjecture that Theorem 12 holds also for binary distributions that are not strictly positive.

Appendix B: Regular Gaussian distributions

We show that Equation 10 holds for regular Gaussian distributions. Our proof is based on three properties of regular Gaussian distributions:

$$I(X, Y | Z) \ \& \ I(X, W | Z) \ \Rightarrow \ I(X, YW | Z) \tag{47}$$

$$I(X, Y | \mathbf{Z}) \Rightarrow I(X, Y | Z) \quad (48)$$

$$I(X, Y | \emptyset) \& I(X, Y | e) \Rightarrow I(X, e | \emptyset) \text{ or } I(e, Y | \emptyset) \quad (49)$$

where e is a single variable not contained in XY .

The first two properties can be verified trivially from the definition of regular Gaussian distributions, whereas the third requires some algebra on the determinant of a covariance matrix. These considerations are left to the reader.

Equation 48 is an interesting property. It states that, for Gaussian distributions, if X and Y are conditionally independent given a specific value \mathbf{Z} of Z , then X and Y are conditionally independent for every value of Z . The truth of this property rests on the fact that whether or not $I(X, Y | Z)$ holds for a Gaussian distribution is determined solely by its covariance matrix, which does not depend on the values given to X, Y and Z .

The three statements in the antecedents of Equation 10 are

$$I(A_1 A_2 A_3 A_4, B_1 B_2 B_3 B_4 | \emptyset)$$

$$I(A_1 A_2 B_3 B_4, B_1 B_2 A_3 A_4 | \mathbf{e}')$$

$$I(A_1 A_3 B_2 B_4, B_1 B_3 A_2 A_4 | \mathbf{e}'')$$

Applying Equation 48 yields the following statements, where \mathbf{e}' and \mathbf{e}'' are replaced with their variable name e :

$$I(A_1 A_2 A_3 A_4, B_1 B_2 B_3 B_4 | \emptyset) \quad (I_1)$$

$$I(A_1 A_2 B_3 B_4, B_1 B_2 A_3 A_4 | e) \quad (I_2)$$

$$I(A_1 A_3 B_2 B_4, B_1 B_3 A_2 A_4 | e) \quad (I_3)$$

The following three Equations are also needed for the proof:

$$I(A_1 A_2, B_1 B_2 | \emptyset) \& I(A_1 A_2, B_1 B_2 | e) \Rightarrow$$

$$I(A_1A_2, e \mid \emptyset) \text{ or } I(e, B_1B_2 \mid \emptyset) \quad (50)$$

$$I(A_1A_2, A_3A_4 \mid e) \ \& \ I(A_1A_2, e \mid \emptyset) \Rightarrow I(A_1A_2, eA_3A_4 \mid \emptyset) \quad (51)$$

$$I(A_1A_2, B_1B_2B_3B_4 \mid \emptyset) \ \& \ I(A_1A_2, eA_3A_4 \mid \emptyset) \Rightarrow \\ I(A_1A_2, eA_3A_4B_1B_2B_3B_4 \mid \emptyset) \quad (52)$$

Equations 50, 51, and 52 are special cases of Equations 49, 6, and 47, respectively. Next, we prove the theorem by showing that the right hand side of Equation 10, namely

$$I(A_1, eA_2A_3A_4B_1B_2B_3B_4 \mid \emptyset) \text{ or } I(B_1, eA_1A_2A_3A_4B_2B_3B_4 \mid \emptyset) \quad (53)$$

follows from I_1 , I_2 , and I_3 .

First, note that the two antecedents of Equation 50 follow, using decomposition on I_1 and I_2 , respectively. Thus, Equation 50 yields

$$I(A_1A_2, e \mid \emptyset) \text{ or } I(B_1B_2, e \mid \emptyset) \quad (54)$$

Assume the first disjunct holds. This is the second antecedent of Equation 51. The first antecedent of Equation 51 follows from I_2 , using decomposition. Thus, Equation 51 yields $I(A_1A_2, eA_3A_4 \mid \emptyset)$, which is the second antecedent of Equation 52. The first antecedent of Equation 52 follows from I_1 using decomposition. Thus, Equation 52 yields

$$I(A_1A_2, eA_3A_4B_1B_2B_3B_4 \mid \emptyset) \quad (J_1)$$

Now assume the second disjunct of Equation 54 holds. A similar derivation, where the roles of A s and B s are switched in Equations 50 through 52, yields

$$I(B_1B_2, eB_3B_4A_1A_2A_3A_4 \mid \emptyset) \quad (J_2)$$

Consequently, we have shown that I_1 and I_2 imply J_1 or J_2 .

Similarly, by switching the role of A_2 with that of A_3 and the role of B_2 with that of B_3 , and using I_3 instead of I_2 , we obtain

$$I(A_1A_3, eA_2A_4B_1B_2B_3B_4 \mid \emptyset) \quad (J_3)$$

or

$$I(B_1B_3, eB_2B_4A_1A_2A_3A_4 \mid \emptyset) \quad (J_4)$$

Thus, there are four cases to consider, by choosing one statement of each of the two disjunctions above. The four cases are J_1 and J_3 , J_2 and J_4 , J_1 and J_4 , and J_2 and J_3 .

If J_1 and J_3 hold, then $I(A_1, eA_3A_4B_1B_2B_3B_4 \mid \emptyset)$ follows from J_1 and $I(A_1, A_2 \mid \emptyset)$ follows from J_3 , using decomposition. Together, using Equation 50, the statement $I(A_1, eA_2A_3A_4B_1B_2B_3B_4 \mid \emptyset)$ is implied. Similarly, when J_2 and J_4 hold, $I(B_1, eA_1A_2A_3A_4B_2B_3B_4 \mid \emptyset)$ must hold. If J_1 and J_4 hold, then, using decomposition on J_4 , $I(B_3, e \mid \emptyset)$ is obtained. The statement $I(B_3, B_1 \mid e)$ is implied from I_2 using decomposition. Together, the two statements yield, using contraction and decomposition, $I(B_3, B_1 \mid \emptyset)$. This statement combined with $I(B_1, eB_2B_4A_1A_2A_3A_4 \mid \emptyset)$, which follows from J_4 using decomposition, imply, using symmetry and Equation 50, that the statement $I(B_1, eA_1A_2A_3A_4B_2B_3B_4 \mid \emptyset)$ holds. The case where J_2 and J_3 hold is symmetric to the case where J_1 and J_4 hold (by switching the roles of the A 's with those of the B 's), thus yielding $I(A_1, eA_2A_3A_4B_1B_2B_3B_4 \mid \emptyset)$.

We have shown that for each of the four possible cases at least one of the disjuncts of Equation 53 is implied. Thus, Equation 10 holds. \square