CONDITIONAL INDEPENDENCE AND PROBABILISTIC INFLUENCE DIAGRAMS

By Richard E. Barlow[1] and Carlos Alberto de Braganca Pereira[1]

University of California, Berkeley, and Universidade de Sao Paulo, Brazil

A graphical approach to conditional independence is discussed. Some well known results concerning conditional independence are proved using simple influence diagram arguments. This material is, in part, from a book in progress tentatively titled *Applied Bayesian Statistics*, by the present authors.

1. Introduction. Influence diagrams with decision nodes were invented in 1976 by Miller et al. [cf. Howard and Matheson (1984)]. Shachter (1986) further developed methods for analyzing influence diagrams. S. Wright (1934) used diagrams to aid in understanding his "method of path coefficients." Although his diagrams pictorially resemble Gaussian influence diagrams [cf. Shachter and Kenley (1988)], they are not based on the Bayesian paradigm. They are not in any sense influence diagrams. I.J. Good (1961) invented "causal nets" that resemble influence diagrams. He used them to illustrate his ideas of causality and conditional independence. In this respect they are similar to influence diagrams. However he did not develop a comparable methodology for analyzing the diagrams. His diagrams are not influence diagrams as we define them below.

Influence diagrams are useful for **modeling** statistical problems. Construction of the diagram is helpful in understanding the problem and communicating the interdependencies to others. In the process of constructing the influence diagram, a representation of the joint distribution of random quantities related to the problem of interest is developed. Usually one does not start with the joint distribution but uses the influence diagram model to determine a useful representation of the joint distribution. In the case of decision influence diagrams, the diagram can be used to help solve the decision problem(s) of interest. Examples of the use of influence diagrams can be found in Barlow and Zhang (1987) and Lauritzen and Spiegelhalter (1988).

[1]This research was partially supported by the U. S. Army Research Office under Contract DAAG29-85-K-0208 with the University of California and by CNPq, Brasilia, Brazil under Contract No. 20771/85-MA.

AMS 1980 subject classifications. 60–02, 62A15, 62B15.

Key words and phrases. Conditional independence, conditional probability, graphs, influence diagrams.

1.1. Definitions and Basic Results. An influence diagram is, first of all, a **directed graph**. A **graph** is a set, V, of nodes or vertices together with a set, A, of arcs joining the nodes. It is said to be directed if the arcs are arrows (directed arcs). Let $V = \{v_1, \ldots, v_n\}$ and let A be a set of ordered pairs of elements of V, representing the directed arcs. That is, if $[v_i, v_j] \in A$ for $1 \leq i, j \leq n$, then there is a directed arc (arrow) from vertex v_i to vertex v_j (the arrow is directed from v_i to v_j). If $[v_i, v_j] \in A$, v_i is said to be an **adjacent predecessor** of v_j and v_j is said to be an **adjacent successor** of v_i. The direction of arcs is meant to denote influence (or possible dependence).

Circles (or ovals) represent random quantities which may, at some time, be observed and consequently may change to data. Circle nodes are called **probabilistic nodes**. Attached to each circle node is a conditional probability (density) function. This function is a function of the state of the node and also of the states of the adjacent predecessor nodes.

A double circle (or double oval) denotes a **deterministic node** which is a node with only one possible state, given the states of the adjacent predecessor nodes; i.e., it denotes a deterministic function of all adjacent predecessors. Thus, to include the background information, H, in the graph, we would have to use a double circle around H.

The following concepts formalize the ideas used in drawing the diagrams of this paper.

DEFINITION 1.1. A directed graph is cyclic, and is called a **cyclic directed graph**, if there exists a sequence of ordered pairs in A such that the initial and terminal vertices are identical; i.e., there exists an integer $k \leq n$ and a sequence of k arcs of the following type:

$$[v_{i_1}, v_{i_2}], [v_{i_2}, v_{i_3}], \ldots, [v_{i_{k-1}}, v_{i_k}], [v_{i_k}, v_{i_1}].$$

DEFINITION 1.2. An **acyclic directed graph** is a directed graph that is not cyclic.

DEFINITION 1.3. A **root node** is a node with no adjacent predecessors. A **sink node** is a node with no adjacent successors. Note that any acyclic directed graph must have at least one root and one sink node.

DEFINITION 1.4. A **Probabilistic Influence Diagram** is an acyclic directed graph in which

i) nodes represent random quantities while directed arcs indicate possible dependence; and

ii) attached to each node is a conditional probability function (for the node) which depends on the states of adjacent predecessor nodes.

Given a directed acyclic graph together with node conditional probabilities (i.e., a probabilistic influence diagram), there exists a *unique joint probability function* corresponding to the random quantities represented by the nodes of the graph. This is because a directed graph is acyclic if and only if there exists a list ordering of the nodes such that any successor of a node x in the graph follows node x in the list as well. Consequently, following the list ordering and taking the product of all node conditional probabilities we obtain the joint probability of the random quantities corresponding to the nodes in the graph. Note that in a cyclic graph the product of the conditional probability functions attached to the nodes *would not determine* the joint probability function.

The following basic result shows that the absence of an arc connecting two nodes in the influence diagram denotes the judgment that the unknown quantities associated with these nodes are conditionally independent given the states of all *adjacent* predecessor nodes.

REMARK 1.5. Let x_i and x_j represent two nodes in a probabilistic influence diagram. If there is no arc connecting x_i and x_j, then x_i and x_j are conditionally independent given the states of the adjacent predecessor nodes; i.e.,

$$p(x_i, x_j \mid w_i, w_j, w_{ij}) = p(x_i \mid w_i, w_j, w_{ij})p(x_j \mid w_i, w_j, w_{ij})$$

where $w_i(w_j)$ denotes the set of adjacent predecessor nodes to only $x_i(x_j)$ while w_{ij} denotes the set of adjacent predecessor nodes to both x_i and x_j.

REMARK 1.6. In a probabilistic influence diagram, if two nodes, x_i and x_j, are root nodes then they are independent.

EXAMPLE 1.7. (Forensic Science). A robbery has been committed and a suspect, a young man, is on trial. In the course of the robbery, a window pane was broken. The robber had apparently cut himself and a blood stain was left at the scene of the crime. Let x represent the blood type of the suspect, y the blood type of the blood stain found at the scene of the crime, and θ the quantity of interest, "the state of culpability" (guilt or innocence) of the suspect. Formally, and before using the actual values of the observable quantities, we have:

$$x = \begin{cases} 1 & \text{if the suspect's} \\ & \text{blood type is } A, \\ 0 & \text{otherwise.} \end{cases} \quad y = \begin{cases} 1 & \text{if the blood} \\ & \text{stain type is } A, \\ 0 & \text{otherwise.} \end{cases} \quad \theta = \begin{cases} 1 & \text{if the suspect} \\ & \text{is guilty,} \\ 0 & \text{otherwise.} \end{cases}$$

The following diagram is a probability model constructed for this case. Note that the actual values of x and y that are known at the time of the analysis are not yet used. In fact, the diagram describes the dependence relations among the quantities and the conditional probabilities to be used.

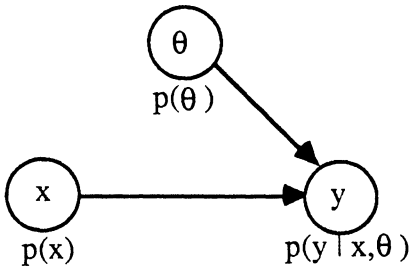

Figure 1.1. Influence Diagram for a Problem in Forensic Science

If p represents the proportion of people in the population with blood type A and if, for a jury member that happens to be interested in probability, q represents his probability that the suspect is guilty before the juror has learned about the blood evidence, then a reasonable probability model is:

$$p(\theta) = \begin{cases} q & \text{if } \theta = 1 \\ 1-q & \text{if } \theta = 0. \end{cases} \quad p(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0. \end{cases}$$

$$p(y \mid x, \theta) = \begin{cases} p & \text{if } \theta \neq y = 1 \\ 1-p & \text{if } \theta = y = 0 \\ 1 & \text{if } \theta = 1 \text{ and} \\ & y = x \\ 0 & \text{otherwise.} \end{cases}$$

The objective of the jury member is to obtain the probability of guilt ($\theta = 1$) after observing the evidence ($x = y = 1$) namely that the blood type of the suspect is the same as that of the stain. That is, the jury member needs to obtain $p(\theta \mid x, y)$ evaluated at $\{\theta = x = y = 1\}$.

2. Probabilistic Influence Diagram Operations. The Bayesian approach to statistics is based on probability judgments and as such follows the laws of probability. You are said to be **coherent** if i) you use probability to measure your uncertainty about quantities of interest and ii) you do not violate the laws of probability when stating your measurements (probabilities). Probabilistic influence diagrams (and influence diagrams in general) are helpful in assuring coherence. Clearly, from coherence, any operation to be performed in a probabilistic influence diagram must not violate the laws of probability. The three basic probabilistic influence diagram operations that we discuss next are based on the addition and product laws. These operations are: 1) Splitting Nodes, 2) Merging Nodes, and 3) Arc Reversal.

2.1. Splitting Nodes. In general a node in a probabilistic influence diagram can denote a vector random quantity. It is always possible to split such a node into other nodes corresponding to the elements of the vector random quantity. To illustrate ideas, suppose that a node corresponds to a vector of two random quantities, x and y, with joint probability function $p(x,y)$. From the product law we know that

$$p(x,y) = p(x)p(y \mid x) = p(x \mid y)p(y).$$

Hence, Figure 2.1 presents the 3 possible probabilistic influence diagrams that can be used in this case showing the two ways of splitting node (x,y).

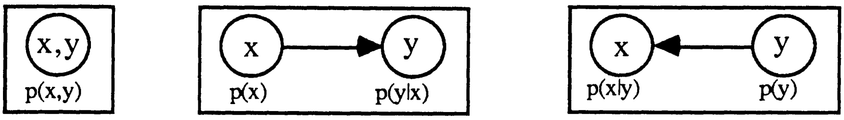

Figure 2.1.
Probabilistic Influence Diagrams for Two Random Quantities

The following property is also a direct consequence of the laws of probability and it is of special interest for statistical applications.

PROPERTY 2.1. Let x be a random quantity represented by a node of a probabilistic influence diagram and let $f(x)$ be a (deterministic) function of x. Suppose we connect to the original diagram a deterministic node representing $f(x)$ using a directed arc from x to $f(x)$. Then, the joint probability distributions for the two diagrams are equal. (See Figure 2.2 for illustration.)

PROOF. Let w and y represent the sets of random quantities that precede and succeed x, respectively, in a list ordering. Note that $p(f(x) \mid w,x) = p(f(x) \mid x) = 1$ and consequently from the product law $p(x, f(x) \mid w) = p(x \mid w)$. That is, node x may be replaced by node $(x, f(x))$ without changing the joint probability of the graph nodes. Using the splitting node operation in node $(x, f(x))$ with x preceding $f(x)$, we obtain the original graph with the additional deterministic node $f(x)$ and a directed arc from x to $f(x)$. Note also that no other arc is necessary since $f(x)$ is determined by x and $p(y \mid w, f(x), x) = p(y \mid w, x)$. ||

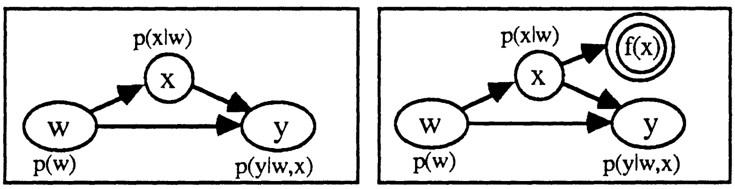

Figure 2.2. Addition of a Deterministic Node

2.2. Merging Nodes. The second probabilistic influence diagram operation is the merging of nodes. Consider first a probabilistic influence diagram with two nodes, x and y, with a directed arc from x to y. The product law states that $p(x,y) = p(x)p(y \mid x)$. Hence, without changing the joint probability of x and y, the original diagram can be replaced by a single node diagram representing the vector (x,y). The first two diagrams of Figure 2.1 in the reverse order illustrate this operation. In general, two nodes, x and y, can be replaced by a single node, representing the vector (x,y), if there is a list ordering such that x is an immediate predecessor or successor of y.

It is not always possible to merge two adjacent nodes in a probabilistic influence diagram. Note that two adjacent nodes may not be neighbors in any list ordering. For example, consider the first diagram of Figure 2.3. Note that all pairs of nodes in this diagram constitute adjacent nodes.

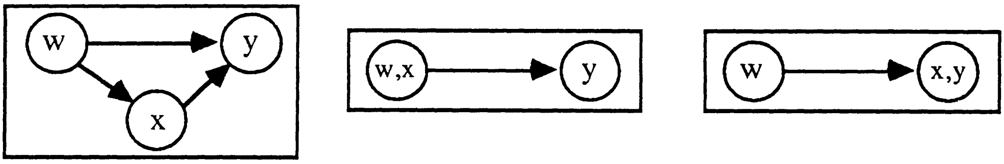

Figure 2.3. Diagram with Adjacent Nodes, w and y, Not Allowed to Be Merged

However, w and y cannot be merged into a node representing (w,y). Clearly the only list ordering here is $w < x < y$ and w and y are not immediate neighbors in this ordering. The problem here is that to merge w and y we would need an arc from (w,y) to x and another from x to (w,y). The reason for this is the existence of arcs $[w,x]$ and $[x,y]$ in the original graph. If we were to have arcs in both directions between (w,y) and x, we would not obtain, in general, the joint probability function from the diagram since $p(w,x,y) \neq p(x \mid w,y)p(w,y \mid x)$. Also it can be seen from the first diagram of Figure 2.3 that there exist two paths from w to y. This is the graphical way to see that w and y cannot be merged into a single node. To construct a graphical technique to check if two nodes can be merged, we need the following definition and theorem.

DEFINITION 2.2. A **directed path** from node x_i to node x_j is a chain of ordered pairs

$$([x_i, x_{k_1}], [x_{k_1}, x_{k_2}], \ldots, [x_{k_{t-1}}, x_{k_t}], [x_{k_t}, x_j])$$

corresponding to directed arcs which lead from x_i to x_j.

THEOREM 2.3. *(Merging Nodes Theorem) In a probabilistic influence diagram, nodes x and y can be merged if either*

1) the only directed path between x and y is a directed arc connecting x and y; or

2) there is no directed path connecting x and y.

PROOF. To be definite, suppose that x precedes y in an associated list ordering corresponding to a probabilistic influence diagram. Let $w_x(w_y)$ be the set of adjacent predecessors of $x(y)$ but not of $y(x)$ and let w_{xy} be the set of node which are adjacent predecessors of both x and y. Since there is no directed path from x to y except, possibly, for a directed arc from x to y, we may add arcs from each node in w_x to y and from each node in w_y to x without creating any cycles. This is possible because directed arcs indicate *possible dependence* not necessarily strict dependence. We have of course lost some graph information as a result of these arc additions.

In the associated list ordering of nodes for our modified diagram, the family of nodes $\{w_x, w_y, w_{xy}\}$ precede both x and y. Since there is no other directed path from x to y other than possibly a directed arc from x to y, there exists an associated list ordering of nodes for which x is an immediate predecessor of y in this list ordering. The product

$$p(x \mid w_x, w_y, w_{xy})p(y \mid x, w_x, w_y, w_{xy})$$

must appear in the representation for the joint probability function for all probabilistic nodes based on the list ordering. Since

$$p(y, x \mid w_x, w_y, w_{xy}) = p(x \mid w_x, w_y, w_{xy})p(y \mid x, w_x, w_y, w_{xy})$$

by the product law, we can merge x and y.

Finally, suppose that there is a directed path from x to y other than a directed arc from x to y. In this case it is not difficult to see that merging x and y would create a cycle which is not allowed. ||

The above result is related to arc reversal, an important operation discussed next.

2.3. Reversing Arcs. The probabilistic influence diagram operation corresponding to Bayes' formula is that of arc reversal. Consider the diagram on the left in Figure 2.4. Using the merging nodes operation we obtain the single node diagram in the center where the probability function of the node (x, y) is obtained from the first diagram as $p(x, y) = p(x)p(y \mid x)$. Using the splitting nodes operation we can obtain the diagram on the right of Figure 2.4. Note that to obtain the corresponding probability functions we use

1) the theorem of total probability for $p(y) = \Sigma_x p(y \mid x)p(x)$, where Σ_x is the sum (or integral) over all possible values of x, and

2) the multiplication law for $p(x \mid y) = p(x, y)/p(y)$ since $p(y)p(x \mid y) = p(x, y)$.

By substituting the appropriate expression in $p(x \mid y)$ we obtain Bayes' formula. That is,

$$p(x \mid y) = \{p(x)p(y \mid x)\}/\{\Sigma_x p(x)p(y \mid x)\}.$$

Hence, by using the theorem of total probability and Bayes' formula when performing an arc reversal operation, we can go directly from the left diagram to the right one in Figure 2.4 without having to consider the one in the center.

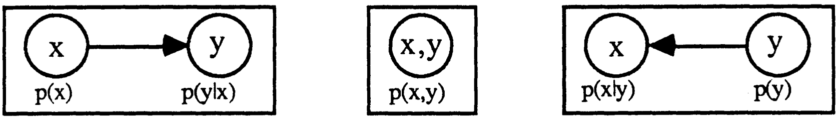

Figure 2.4. Reversing Arc Operation in a Two Node Probabilistic Influence Diagram

Although the diagrams are different they have the same joint probability function for node random quantities. This fact is formalized in the following definition.

DEFINITION 2.4. Two probabilistic influence diagrams are said to be **equivalent in probability** if they have the same joint probability function for node random quantities.

Consider the diagram of Figure 2.5 where w_x, w_y, and $w_{x,y}$ are sets of adjacent predecessors of x and (or) y as indicated by the figure. If arc $[x, y]$ is the *only* directed path from node x to node y, we may add arcs $[w_x, y]$ and $[w_y, x]$ to the diagram without introducing any cycles. (See left diagram of Figure 2.6.) Remember that a directed arc only indicates *possible* dependence.

The following result introduces the conditions under which arc reversal operations can be performed.

THEOREM 2.5. *(Reversing Arcs Theorem) Suppose that arc $[x, y]$ connects nodes x and y in a probabilistic influence diagram. $[x, y]$ can be reversed to $[y, x]$, without changing the joint probability function of the diagram if*

1) there is no other directed path from x to y,

2) all the adjacent predecessors of $x(y)$, in the original diagram, become also adjacent predecessors of $y(x)$, in the modified diagram, and

3) the conditional probability functions attached to nodes x and y are also modified in accord with the laws of probability.

PROOF. Let $w_x(w_y)$ be the set of adjacent predecessors of $x(y)$ but not of $y(x)$ and $w_{x,y}$ be the set of adjacent predecessors of both x and y. Since arcs represent possible dependence, we can add arcs to the diagram in order to make the set $(w_x, w_x, w_{x,y})$ an adjacent predecessor of both x and y. Since there is no other directed path connecting x and y, there is a list ordering such that x is an immediate predecessor of y in the list. Note also that the elements of the set $(w_x, w_x, w_{x,y})$ are all predecessors of both x and y in the list ordering. To obtain the joint probability function corresponding to the first diagram we consider the product, following the list ordering, of all node conditional probability functions. As a factor of this product we have

$$p(x \mid w_x, w_{x,y})p(y \mid x, w_y, w_{x,y}) = p(x \mid w_x, w_y, w_{x,y})p(y \mid x, w_x, w_y, w_{x,y}) = p(x, y \mid w_x, w_y, w_{x,y}) = p(y \mid w_x, w_y, w_{x,y})p(x \mid y, w_x, w_y, w_{x,y}).$$

The first equality is due to the fact that x and w_y are conditionally independent given $(w_x, w_{x,y})$ and y and w_x are conditionally independent given $(w_y, w_{x,y})$. [See Figure 2.5.] The other two equalities follow from the product law.

Replacing $p(x \mid w_x, w_{x,y})p(y \mid x, w_y, w_{x,y})$ in the product of the conditional probability functions for the original diagram by $p(y \mid w_x, w_y, w_{x,y})p(x \mid y, w_x, w_y, w_{x,y})$ we obtain the product of the conditional probability functions for the second diagram. This proves that the joint probability functions of the two diagrams are equal. Finally, we notice that if there were another directed path from x to y, we would create a cycle by reversing arc $[x, y]$, which is not allowed. ||

In general, reversing an arc corresponds to applying Bayes' formula and the theorem of total probability. However, it may also involve the addition of arcs and such arcs, in some cases represent only pseudo dependencies. In this sense, some relevant information may have been lost after arc reversal.

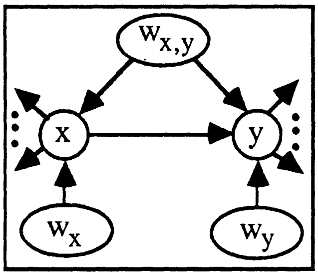

Figure 2.5.

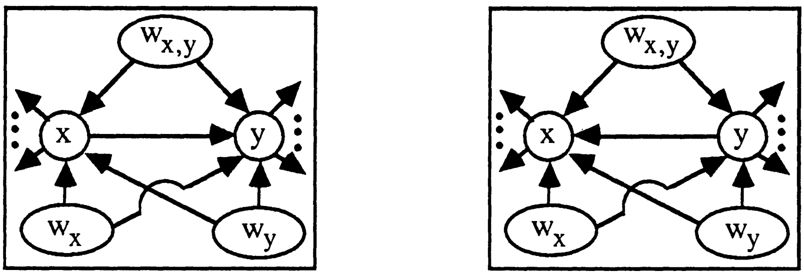

Figure 2.6. Equivalent Probabilistic Influence Diagrams. Probability Nodes in the Right Diagram are Obtained From the Left Diagram by Using Bayes' Formula and the Theorem of Total Probability

3. Conditional Independence. The objective of this section is to study the concept of conditional independence and introduce its basic properties. We believe that the simplest and most intuitive way that this study can be performed is by using all the visual force of the probabilistic diagram.

We now introduce the two most common definitions of conditional independence.

DEFINITION 3.1. (Intuitive) Given random quantities x, y, and z, we say that y is conditionally independent of x given z if the conditional distribution of y given (x, z) is equal to the conditional distribution of y given z.

The interpretation of this concept is that, if z is given, no additional information about y can be extracted from x. The influence diagram representing this statement is presented in Figure 3.1.

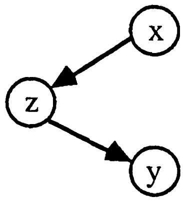

Figure 3.1. Intuitive Definition of Conditional Independence

DEFINITION 3.2. (Symmetric) Given random quantities x, y, and z, we say that x and y are conditionally independent given z if the conditional distribution of (x, y) given z is the product of the conditional distributions of x given z and that of y given z.

The interpretation is that, if z is given, x and y share no additional information. The influence diagram representing this statement is displayed in Figure 3.2.

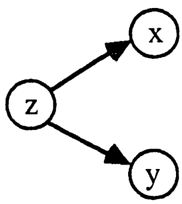

Figure 3.2. Symmetric Definition of Conditional Independence

Using the arc reversal operation, we can easily prove that the probabilistic influence diagrams in Figures 3.1 and 3.2 are equivalent. Thus, Definitions 3.1 and 3.2 are equivalent, which means that in a specific problem we can use either one. To represent the conditional independence described by both Figures 3.1 and 3.2

we can write either $x \perp\!\!\!\perp y \mid z$ or $y \perp\!\!\!\perp x \mid z$. This is a very general notation since x, y, and z are general random quantities (scalars, vectors, events, etc.). If in place of $\perp\!\!\!\perp$ we use $\not\perp\!\!\!\perp$, then x and y are said to be strictly dependent given z. We obtain independence (dependence) and write $x \perp\!\!\!\perp y$ ($x \not\perp\!\!\!\perp y$) if z is an event which occurs with probability one. It is important to notice that the symbol $\perp\!\!\!\perp$ corresponds to the absence of an arc in a probabilistic influence diagram. However, the existence of an arc only indicates possible dependence. Although $\not\perp\!\!\!\perp$ is the negation of $\perp\!\!\!\perp$, the "absence of an arc" is included in the "presence of an arc."

The following proposition introduces the essence of the DROP/ADD principles for conditional independence which are briefly discussed in the sequel.

PROPOSITION 3.3. *If $x \perp\!\!\!\perp y \mid z$ then, for every $f = f(x)$, we have:*

(i) $f \perp\!\!\!\perp y \mid z$; *and*

(ii) $x \perp\!\!\!\perp y \mid (z, f)$.

The proof of this property is the sequence of diagrams of Figure 3.3. First note that (by Property 2.1) to obtain the second diagram from the first we can connect to x a deterministic node f using arc $[x, f]$ without changing the joint probability function. Consequently, by reversing arc $[x, f]$ we obtain the third diagram. To obtain the last diagram from the third we use the merging nodes operation. Relations i) and ii) of Proposition 3.3 are represented by the second and the third diagrams of Figure 3.3.

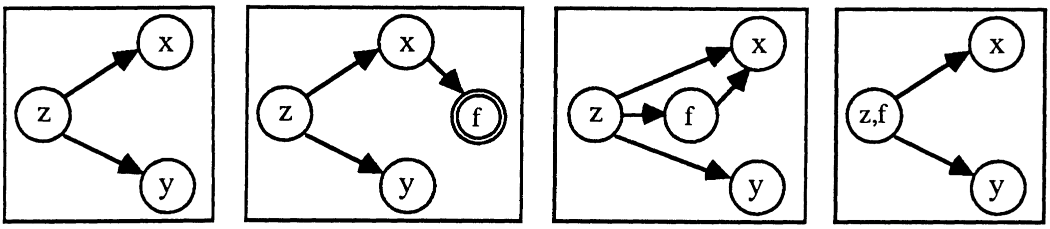

Figure 3.3. Proof of Proposition 3.3

As direct consequences of Proposition 3.3 we have:

$C1-$ If $g = g(z)$ then $x \perp\!\!\!\perp y \mid z$ if and only if $x \perp\!\!\!\perp (y, g) \mid z$.

$C2-$ Let $f = f(x, z)$ and $g = g(y, z)$. If $x \perp\!\!\!\perp y \mid z$ then, $f \perp\!\!\!\perp g \mid z$ and $x \perp\!\!\!\perp y \mid (z, f, g)$.

The concept of conditional independence gives rise to many questions. Among them are the ones involving the DROP/ADD principles that we describe next. Suppose that x, y, z, w, f, and g are random objects such that $x \perp\!\!\!\perp y \mid z$, $f = f(x)$ and $g = g(z)$. What can be said about the relation $\perp\!\!\!\perp$ if f is substituted for x, g for z, (y, w) for y or (z, w) for z? In other words, can x, y, and z be reduced or enlarged without destroying the $\perp\!\!\!\perp$ relation? In general, the answer is no. However, for special kinds of reductions or enlargements the conditional independence relation is preserved.

First we present two simple examples to show that arbitrary enlargements of x, y, or z may destroy the $\perp\!\!\!\perp$ relation. The forensic science example shows that $\theta \perp\!\!\!\perp y$ but $\theta \not\perp\!\!\!\perp (x, y)$ or, in the present notation, considering z a sure event and $w = x$, $\theta \perp\!\!\!\perp y \mid z$ but $\theta \not\perp\!\!\!\perp (y, w) \mid z$. Consider now that w_1 and w_2 are two independent standard normal random variables; i.e., $w_1 \sim w_2 \sim N(0,1)$, and $w_1 \perp\!\!\!\perp w_2$. If $x = w_1 - w_2$ and $y = w_1 + w_2$, then $x \perp\!\!\!\perp y$ but certainly $x \not\perp\!\!\!\perp y \mid w_2$. Note that if z is a constant and $w = w_1$, we conclude that $x \perp\!\!\!\perp y \mid z$ but $x \not\perp\!\!\!\perp y \mid (z, w)$.

Secondly, we present an example to show that an arbitrary reduction of z, the conditioning quantity, can destroy the $\perp\!\!\!\perp$ relation. Let w_1, w_2, and w be three mutually independent standard normal random quantities; i.e., $w_1 \perp\!\!\!\perp (w_2, w)$, $(w_1, w_2) \perp\!\!\!\perp w$, $w_2 \perp\!\!\!\perp (w_1, w)$, $w_1, \perp\!\!\!\perp w_2$, $w_1 \perp\!\!\!\perp w$, $w_2 \perp\!\!\!\perp w$, and $w_1 \sim w_2 \sim w \sim N(0,1)$. Define $x = w_1 - w_2 + w$ and $y = w_1 + w_2 + w$, and note that $x \perp\!\!\!\perp y \mid w$ but $x \not\perp\!\!\!\perp y$. As before, if z is a constant we can conclude that $x \perp\!\!\!\perp y \mid (z, w)$ but $x \not\perp\!\!\!\perp y \mid z$.

The destruction of the $\perp\!\!\!\perp$ relation by reducing or enlarging its arguments is known as Simpson's paradox (for more details, see Lindley and Novick, 1981). The paradox, however, is much stronger since highly positively correlated random variables could be highly negatively correlated after some Drop/Add operations. For instance, let z and w be two independent normal random variables with zero means. Define $x = z + w$ and $y = z - w$ and note that the correlation between x and y is given by correlation$(x, y) = (l - r)(l + r)^{-1}$ where r is equal to the variance of w divided by the variance of z. Also, if z is given it is clear that the conditional correlation is -1. In order to make $\text{cor}(x, y)$ close to 1 we can consider r arbitrarily small. This shows that we can have cases where x and y are strongly positive (negative) dependent but, when z is given, x and y turn to be strongly negative (positive) conditionally dependent.

The following is another important property of conditional independence. It is presented in Dawid (1979).

PROPOSITION 3.4. *The following statements are equivalent:*

(i) $x \perp\!\!\!\perp y \mid z$ *and* $x \perp\!\!\!\perp w \mid (y, z)$*;*

(ii) $x \perp\!\!\!\perp (w, y) \mid z$*; and*

(iii) $x \perp\!\!\!\perp w \mid z$ *and* $x \perp\!\!\!\perp y \mid (w, z)$.

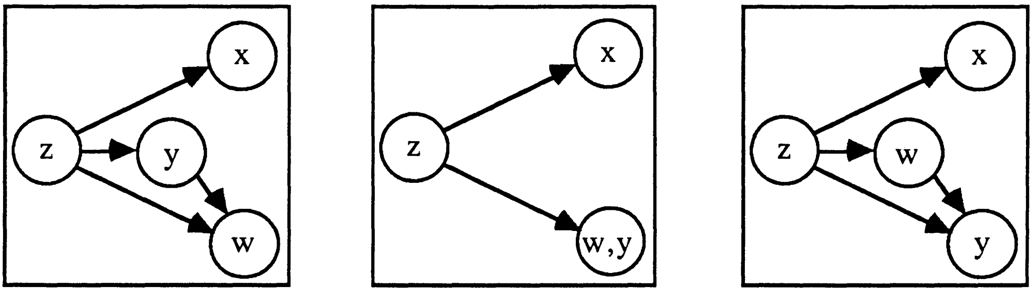

Figure 3.4. Proof of Proposition 3.4

Figure 3.4 is the proof of Proposition 3.4. Again, only the basic probabilistic influence diagrams operations are used. The second graph is obtained from the first by merging nodes w and y. The third graph is obtained from the second by splitting node (w, y) and the first is obtained from the third by reversing arc $[w, y]$.

The above simple properties are very useful in some statistical applications and they are related to the concept of sufficient statistic. In the context of comparisons of experiments a very general concept of sufficiency was introduced by Blackwell (1953). We next discuss Blackwell's concept of sufficiency using probabilistic influence diagrams.

3.1. Blackwell Sufficiency. Suppose that we can perform either one of two experiments to learn about a random quantity θ. In the first experiment, we observe x, knowing $p(x \mid \theta)$. In the second experiment, we observe y, knowing $p(y \mid \theta)$. If, furthermore, there exists a random quantity x' such that $\theta \perp\!\!\!\perp x' \mid y$ and $p(x' \mid \theta) = p(x \mid \theta)$, then we say that y is **Blackwell sufficient** for x relative to θ.

In terms of probabilistic influence diagrams, we construct two diagrams, the first with nodes θ and x connected by arc $[\theta, x]$ and the second with three nodes θ, y, and x' connected by arcs $[\theta, y]$ and $[y, x']$. If in the second diagram, after eliminating node y, we obtain a diagram having only two nodes, θ and x', equivalent to the first diagram, then we have Blackwell's concept of sufficiency. See Figure 3.5. In this sense x' is a "garbling" of y. If we cannot observe both x and y, it is better to observe y and use $p(y \mid \theta)$ to make inferences about θ.

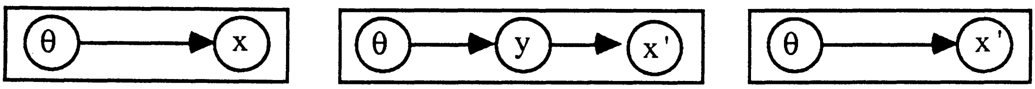

Figure 3.5. Blackwell Sufficiency When the Right and Left Diagrams are Equivalent

DEFINITION 3.5. (Blackwell Sufficiency). A random quantity, y, is sufficient for a random quantity, x, relative to a random quantity, θ, if there exists another random quantity, x', such that

(i) $\theta \perp\!\!\!\perp x' \mid y$ and

(ii) $p(x' \mid \theta) = p(x \mid \theta)$.

To conclude, we present the following example which shows the usefulness of Blackwell sufficiency in comparing experiments.

EXAMPLE 3.6. Let x and y be two Bernoulli quantities such that, given a parameter $\theta, \Pr\{x = 1 \mid \theta\} = \theta/2$ and $\Pr\{y = 1 \mid \theta\} = \theta$. Suppose that we want to learn more about the parameter θ, but we can only observe one of the random quantities x or y, but not both. The question of which one to observe involves the cost of observation and other considerations. For the moment let us suppose they have the same cost. If we can prove that y is Blackwell sufficient for x relative to θ, we must prefer y since it is at least as good as x for learning about θ. We now prove that y is in fact Blackwell sufficient for x.

Suppose that we toss a fair coin and record $r = 1$ if we obtain a tail and $r = 0$ if we obtain a head. Define now the random quantity $x' = yr$. Figure 3.6 shows, on the left, a diagram relating θ, y, r, and x'. After eliminating node r we obtain the diagram in the center of Figure 3.6. The right diagram of Figure 3.6 is obtained after the elimination of node y. This last diagram is equivalent to the probabilistic diagram relating x and θ. Hence, y is Blackwell sufficient for x relative to θ.

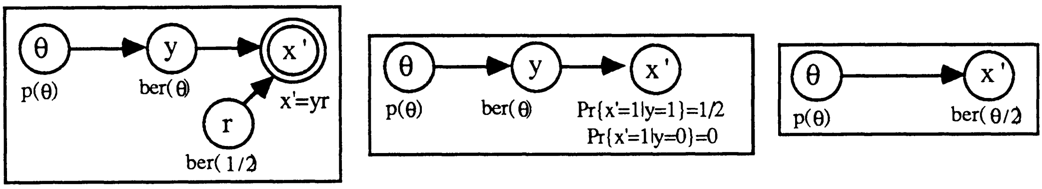

Figure 3.6. Proof of Blackwell Sufficiency

REFERENCES

BARLOW, R.E. and PEREIRA, C.A. DE B. (1987). The Bayesian operation and probabilistic influence diagrams. University of California Engineering Systems Research Center Technical Report ESRC 87-7.

BARLOW, R.E. and ZHANG, X. (1987). Bayesian analysis of inspection sampling procedures discussed by Deming. *J. of Stat. Planning and Inference* **16** 285–296.

BASU, D. (1975). Statistical information and likelihood (with discussion). *Sankhyā* (Ser. A) **37** 1–71.

BASU, D. and PEREIRA, C.A. DE B. (1983). Conditional independence in statistics. *Sankhyā* (Ser. A) **45** 324–337.

BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

DAWID, A.P. (1979). Conditional independence in statistical theory. *J.R. Statist. Soc.* B **41** 1–31.

GOOD, I.J. (1961). A causal calculus I and II. *Br. J. Phil. Sci.* **XI**, 305–318 and **XII**, 43–51.

HOWARD, R.A. and MATHESON, J.E. (1984). Influence diagrams. In *Readings in the Principles and Applications of Decisions Analysis.* Two volumes. Howard and Matheson, eds., Strategic Decision Group, Menlo Park, CA.

LAURITZEN, S.L. and SPIEGELHALTER, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Roy. Statist. Soc.* B **50** (to appear).

LINDLEY, D.V. and NOVICK, M.R. (1981). The role of exchangeability in inference. *Ann. Statist.* **9** 45–58.

SHACHTER, R. (1986). Evaluating influence diagrams. *Oper. Res.* **34** 871–882.

SHACHTER, R. and KENLEY, C.R. (1988). Gaussian influence diagrams. *Man. Sci.* To appear.

WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5** 161–215.

4177 ETCHEVERRY HALL
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CA 94720

INSTITUTO DE MATHEMATICA
E ESTATISTICA
UNIVERSIDADE DE SAO PAULO
CAIXA POSTAL 20570 (AG. IGUATEMI)
CEP 05508 SAO PAULO, S.P.
BRASIL