Appeared as: E. N. Cinicioglu and P. P. Shenoy, "Arc reversals in hybrid Bayesian networks with deterministic variables," *International Journal of Approximate Reasoning*, 50(5), 2009, 763–777.

Arc Reversals in Hybrid Bayesian Networks with Deterministic Variables

Esma Nur Cinicioglu[a] *and Prakash P. Shenoy*[b]

[a] Istanbul University Faculty of Business Administration, 34320 Avcilar, Istanbul, Turkey

[b] University of Kansas School of Business, 1300 Sunnyside Ave, Summerfield Hall, Lawrence, KS 66045

<esmanurc@istanbul.edu.tr>, <pshenoy@ku.edu>

Abstract

This article discusses arc reversals in hybrid Bayesian networks with deterministic variables. Hybrid Bayesian networks contain a mix of discrete and continuous chance variables. In a Bayesian network representation, a continuous chance variable is said to be deterministic if its conditional distributions have zero variances. Arc reversals are used in making inferences in hybrid Bayesian networks and influence diagrams. We describe a framework consisting of potentials and some operations on potentials that allows us to describe arc reversals between all possible kinds of pairs of variables. We describe a new type of conditional distribution function, called partially deterministic, if some of the conditional distributions have zero variances and some have positive variances, and show how it can arise from arc reversals.

1 Introduction

Hybrid Bayesian networks are Bayesian networks (BNs) containing a mix of discrete (countable) and continuous (real-valued) chance variables. Shenoy [2006] describes a new technique for “exact” inference in hybrid BNs using mixture of Gaussians. This technique consists of transforming a general hybrid BN to a mixture of Gaussians (MoG) BN. Lauritzen and Jensen [2001] have described a fast algorithm for making inferences in MoG BN, and it is implemented in Hugin, a commercial software package.

A MoG BN is a hybrid BN such that all continuous variables have conditional linear Gaussian (CLG) distributions, and there are no discrete variables with continuous parents. If we have a general hybrid BN containing a discrete variable with continuous parents, then one method of transforming such a network to a MoG BN is to do arc reversals. If a continuous variable has a non-CLG distribution, then we can approximate it with a CLG distribution. In the process of doing so, we may create a discrete variable with continuous parents. In this case, arc reversals are again necessary to convert the resulting hybrid BN to a MoG BN.

Arc reversals were pioneered by Olmsted [1984] for solving discrete influence diagrams. They were further studied by Shachter [1986, 1988, 1990] for solving discrete influence diagrams, finding posterior marginals in discrete BNs, and for finding relevant sets of variables for a decision variable in an influence diagram. Kenley [1986] generalized arc reversals in influence diagrams with continuous variables having conditional linear Gaussian distributions (see also Shachter and Kenley [1989]). Poland [1994] further generalized arc reversals in influence diagrams with Gaussian mixture distributions. Recently, Madsen [2008] has described solving a class of Gaussian influence diagrams using arc-reversal theory. Although there are currently no exact algorithms to solve general hybrid influence diagrams (containing a mix of discrete and continuous chance variables), a theory of arc reversals is useful in this endeavor. We believe that Olmsted's arc reversal algorithm for discrete influence diagrams would apply to influence diagrams with a mix of discrete, continuous, and deterministic chance variables using the arc reversal theory described in this paper. This claim, of course, needs further investigation.

Hybrid BNs containing deterministic variables pose a special problem since the joint density for all continuous variables does not exist. Thus, a method for propagating density potentials would need to be modified to account for the non-existence of the joint density [Cobb and Shenoy 2005a, 2006].

A traditional way of handling continuous chance variables in hybrid BNs is to discretize the conditional density functions, and convert continuous nodes to discrete nodes. There are several problems with this method. First, to get a decent approximation, we need to use many bins. This increases the computational effort for computing marginals. Second, even with many bins, based on what evidence is obtained, which may not be easy to forecast, the posterior marginal may result in all mass in one of the bins resulting in an unacceptable discrete approximation of the posterior marginal. One way to mitigate this problem is to do a dynamic discretization as suggested by Kozlov and Koller [1997], but this is not as simple as just dividing the sample space of continuous variables evenly into some number of bins.

Another method of handling continuous variables in hybrid BNs is to use mixtures of truncated exponentials (MTEs) to approximate probability density functions [Moral *et al*. 2001]. MTEs are easy to integrate in closed form. Since the family of mixtures of truncated exponentials is closed under multiplication, addition, and integration, the Shenoy-Shafer architecture [Shenoy and Shafer 1990] can be used to find posterior marginals. Cobb *et al*. [2006] propose using a non-linear optimization technique for finding mixtures of truncated exponentials approximation for the many commonly used distributions.

Cobb and Shenoy [2005a, b] extend this approach to Bayesian networks with linear and non-linear deterministic variables.

Arc reversals involve divisions. MTEs are not closed under the division operation. Thus, we don't see much relevance for MTEs for describing the arc-reversal theory. However, making inferences in hybrid Bayesian networks with deterministic variables can be described without a division operation, i.e., without arc reversals. And in this case, MTEs can be used to ensure that marginalization of density functions can be easily done.

The main goal of this paper is to describe an arc-reversal theory in hybrid Bayesian networks with deterministic variables. While such a theory would be useful in making inferences in hybrid BNs and also in solving influence diagrams with a mix of discrete, continuous, and deterministic variables, the scope of this paper doesn't include either inference in hybrid Bayesian networks nor solving influence diagrams.

Arc reversal is described in terms of functions called potentials with combination, marginalization, and division operations. One advantage of this framework is that it can be easily adapted to make inferences in hybrid BNs and to solve hybrid influence diagrams. For example, if we use the Shenoy-Shafer architecture [Shenoy and Shafer 1990] for making inferences in hybrid Bayesian networks, then the potentials that are generated by combination and marginalization operations do not always have probabilistic semantics. For example, the combination of a probability density function and a deterministic equation (which is represented as a Dirac delta function) does not have probabilistic semantics. Nevertheless, as we will show in this paper, the use of potentials is useful for describing arc reversals. Furthermore, we believe that this framework can be extended further for computing marginals in hybrid Bayesian networks and for solving hybrid influence diagrams.

Shachter [1988] describes how the arc-reversal theory for discrete Bayesian networks can be used for probabilistic inference. Given that we extend arc-reversal theory for continuous and deterministic variables, Shachter's [1988] framework can thus be used for making inferences in hybrid Bayesian networks with deterministic variables. As observed by Madsen [2006], an important advantage of using arc-reversal theory for making inferences is that after arc-reversal, the network remains a Bayesian network, and we can exploit, e.g., *d*-separation, for probabilistic inference.

An outline of the remainder of this paper is as follows. Section 2 describes the framework of potentials used to describe arc reversals. We use Dirac delta functions to represent conditionally deterministic distributions, and we describe some properties of Dirac delta functions in the Appendix. Section 3 describes arc reversals for arcs between all kinds of pairs of variables. Section 4 describes partially deterministic distributions that arise from arc reversals. Finally, in Section 5, we summarize and conclude.

2 The Framework of Potentials

In this section we will describe the notation and definitions used in the paper. Also, we will decribe a framework consisting of potentials and some operations on potentials that will let us describe the arc reversal process in hybrid Bayesian networks with deterministic conditionals.

Variables and States. We are concerned with a finite set V of variables. Each variable $X \in V$ is associated with a set of its possible *states* denoted by Ω_X. If Ω_X is a countable set, finite or infinite, we say X is *discrete*, and depict it by a rectangular node in a graph; otherwise X is said to be *continuous* and is depicted by an oval node.

In a BN, each variable has a conditional distribution function for each state of its parents. A conditional distribution function associated with a continuous variable is said to be *deterministic* if the variances (for each state of its parents) are all zeros. For simplicity, we will refer to continuous variables with non-deterministic conditionals as continuous, and continuous variables with deterministic conditionals as deterministic. Deterministic variables are represented as oval nodes with a double border in a Bayesian network graph.

We will assume that the state space of continuous variables is the set of real numbers (or some subset of it) and that the states of a discrete variable are symbols. If $r \subseteq V$, then $\Omega_r = \times\{\Omega_X \mid X \in r\}$.

Discrete Potentials. In a BN, each variable has a conditional probability function given its parents and these are represented by functions called *potentials*. If X is discrete, it has a *discrete* potential. Formally, suppose r is a set of variables that contains a discrete variable. A discrete potential ρ for r is a function ρ: $\Omega_r \rightarrow [0, 1]$. The values of discrete potentials are probabilities.

Although the domain of the function ρ is Ω_r, for simplicity, we will refer to r as the *domain* of ρ. Thus the domain of a potential representing the conditional probability mass function associated with some discrete variable X in a BN is always the set $\{X\} \cup pa(X)$, where $pa(X)$ denotes the set of parents of X. The set $pa(X)$ may contain continuous variables.

Density Potentials. Continuous non-deterministic variables typically have conditional density functions, which are represented by functions called *density potentials*. Formally, suppose r is a set of variables that contains a continuous variable. A density potential ρ for r is a function $\rho: \Omega_r \to \mathbb{R}^+$, where $\mathbb{R}^+$ is the set of non-negative real numbers. The values of density potentials are probability densities.

Deterministic variables have conditional distributions containing equations. We will represent such conditional distributions using Dirac delta functions [Dirac 1927]. First, we will define Dirac delta functions.

Dirac Delta Functions. $\delta: \mathbb{R} \to [0, 1]$ is called a Dirac delta function if $\delta(x) = 0$ if $x \neq 0$, and $\int \delta(x)\, dx = 1$. Whenever the limits of integration of an integral are not specified, the entire range $(-\infty, \infty)$ is to be understood. δ is not a proper function since the value of the function at 0 doesn't exist (i.e., not finite). It can be regarded as a limit of a certain sequence of functions (such as, e.g., the Gaussian density function with mean 0 and variance σ^2 in the limit as $\sigma \to 0$). However, it can be used as if it were a proper function for practically all our purposes without getting incorrect results. It was first defined by Dirac [1927]. Some properties of Dirac delta functions are described in the Appendix.

Dirac Potentials. Deterministic variables have conditional distributions containing equations. We will represent such functions by *Dirac* potentials. Before we define Dirac potentials formally, we need to define projection of states. Suppose $\boldsymbol{y}$ is a state of variables in r, and suppose $s \subseteq r$. Then the *projection* of $\boldsymbol{y}$ to s, denoted by $\boldsymbol{y}^{\downarrow s}$ is the state of s obtained from $\boldsymbol{y}$ by dropping states of $r \backslash s$. Thus, $(w, x, y, z)^{\downarrow \{W, X\}} = (w, x)$, where $w \in \Omega_W$, and $x \in \Omega_X$. If $s = r$, then $\boldsymbol{y}^{\downarrow s} = \boldsymbol{y}$.

S uppose $x = r \cup s$ is a set of variables containing some discrete variables r and some continuous variables s. We assume $s \neq \emptyset$. A Dirac potential ξ for x is a function $\xi: \Omega_x \to [0, 1]$ such that $\xi(\boldsymbol{r}, \boldsymbol{s})$ is of the form $\Sigma\{p_{\boldsymbol{r},i}\, \delta(z - g_{\boldsymbol{r},i}(\boldsymbol{s}^{\downarrow \{s \backslash \{Z\}\}})) \mid i = 1, \ldots, n, \text{ and } \boldsymbol{r} \in \Omega_r\}$, where $\boldsymbol{r} \in \Omega_r$, $\boldsymbol{s} \in \Omega_s$, $Z \in s$ is a continuous variable, $z \in \Omega_Z$, $\delta(z - g_{\boldsymbol{r},i}(\boldsymbol{s}^{\downarrow \{s \backslash \{Z\}\}}))$ are Dirac delta functions, $p_{\boldsymbol{r},i}$ are probabilities for all $i = 1, \ldots, n$, and n is a positive integer. Here, we are assuming that continuous variable Z is a deterministic function $g_{\boldsymbol{r},i}(\boldsymbol{s}^{\downarrow \{s \backslash \{Z\}\}})$ of the other continuous variables in s, and that the nature of the deterministic function may be indexed by the states of the discrete variables $\boldsymbol{r}$, and/or by some latent index i.

S uppose X is a deterministic variable with continuous parent Z, and suppose that the deterministic relationship is $X = Z^2$. This conditional distribution is represented by the Dirac potential $\delta(x - z^2)$ for $\{Z, X\}$.

A more general example of a Dirac potential for $\{Z, X\}$ is $\zeta(z, x) = (½)\ \delta(x - z) + (½)\ \delta(x - 1)$. Here, X is a continuous variable with continuous parent Z. The conditional distribution of X is as follows: $X = Z$ with probability ½, and $X = 1$ with probability ½. Notice that X is not deterministic (since the variances of its conditional distributions are not all zeros). Also, notice that strictly speaking, the values of the Dirac potential $\zeta(z, x)$ are either 0 or undefined (when $x = z$ or $x = 1$). For interpretation reasons only, we can follow the convention that the "values" of $\zeta(z, x)$ are ½ when $x = z$, ½ when $x = 1$, and 0 otherwise, which is consistent with the conditional distribution the potential is representing. Thus, as per this convention, the values of Dirac potentials are probabilities in the unit interval [0, 1].

Both density and Dirac potentials are special instances of a broader class of potentials called continuous potentials. Suppose $\mathbf{x}$ is a set of variables containing a continuous variable. Then a *continuous potential* ξ for $\mathbf{x}$ is a function ξ: $\Omega_{\mathbf{x}} \to [0, 1] \cup \mathbb{R}^+$. The values of ξ can be probabilities (in [0, 1]) or densities (in $\mathbb{R}^+$). If some of the values of ξ are probabilities and some are densities, then ξ is a continuous potential that is neither a Dirac potential nor a density potential. For example, consider a continuous variable X with a mixed distribution: a probability of 0.5 at $X = 1$, and a probability density of 0.5 f, where f is a PDF. This mixed distribution can be represented by the continuous potential ξ for $\{X\}$ as follows: $\xi(x) = 0.5\ \delta(x - 1) + 0.5\ f(x)$. Notice that $\xi(1) = 0.5\ \delta(0) + 0.5\ f(1)$. The first term can be interpreted as probability of 0.5 and the second term is a probability density. The distribution $\xi(x)$ is a well-defined distribution since $\int \xi(x)\ dx = 0.5 \int \delta(x - 1)\ dx + 0.5 \int f(x)\ dx = 0.5 + 0.5 = 1$.

As we will see shortly, the combination of two density potentials is a density potential, the combination of two Dirac potentials is a Dirac potential, and the combination of two continuous potentials is a continuous potential. Also, continuous potentials can result from the combination, marginalization and division operations. These operations will be defined shortly.

Consider the BN given in Figure 1. Let α denote the discrete potential for $\{A\}$ associated with A. Then, $\alpha(a_1) = 0.5$ and $\alpha(a_2) = 0.5$, Let ζ be the density potential for $\{Z\}$ associated with Z. Then, $\zeta(z) = f(z)$. Let ξ denote the Dirac potential for $\{A, Z, X\}$ associated with X. Then, $\xi(a_1, z, x) = \delta(x - z)$ and $\xi(a_2, z, x) = \delta(x - 1)$.

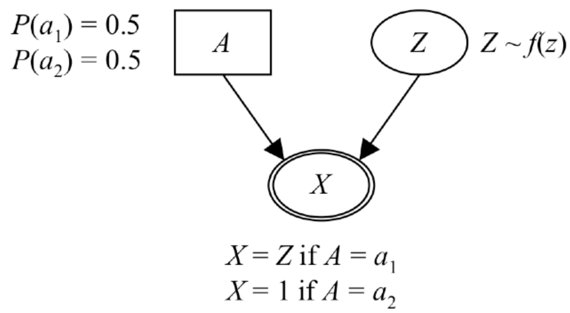

Figure 1: A BN with a discrete, a continuous and a deterministic variable

Next, we define three operations associated with potentials, combination, marginalization, and division.

Combination. Suppose α is a potential (discrete or continuous) for a, and β is a potential (discrete or continuous) for b. Then the *combination* of α and β, denoted by $\alpha\otimes\beta$, is the potential for $a\cup b$ obtained from α and β by point-wise multiplication, i.e., $(\alpha\otimes\beta)(\boldsymbol{x}) = \alpha(\boldsymbol{x}^{\downarrow a})\ \beta(\boldsymbol{x}^{\downarrow b})$ for all $\boldsymbol{x} \in \Omega_{a\cup b}$. If α and β are both discrete potentials, then $\alpha\otimes\beta$ is a discrete potential. If α and β are both density potentials, then $\alpha\otimes\beta$ is a density potential. If α and β are both Dirac potentials, then $\alpha\otimes\beta$ is a Dirac potential. And if α and β are both continuous potentials, then $\alpha\otimes\beta$ is a continuous potential.

Combination of potentials (discrete or continuous) is commutative ($\alpha\otimes\beta = \beta\otimes\alpha$) and associative ($(\alpha\otimes\beta)\otimes\gamma = \alpha\otimes(\beta\otimes\gamma)$). A potential for r such that its values are identically one is called an *identity* potential, and denoted by ι_r. The identity potential ι_r for r has the property that given any potential α for $s \supseteq r$, $\alpha\otimes\iota_r = \alpha$.

Marginalization. The definition of marginalization of potentials (discrete or continuous) depends on the nature of the variable being marginalized out. Suppose χ is a potential (discrete or continuous) for c, and suppose A is a discrete variable in c. Then the *marginal* of χ by removing A, denoted by χ^{-A}, is the potential for $c\backslash\{A\}$ obtained from χ by addition over the states of A, i.e., $\chi^{-A}(\boldsymbol{x}) = \sum\{\chi(a, \boldsymbol{x}) \mid a \in \Omega_A\}$ for all $\boldsymbol{x} \in \Omega_{c\backslash\{A\}}$.

Suppose χ is a potential (discrete or continuous) for c and suppose X is a continuous variable in c. Then the *marginal* of χ by removing X, denoted by χ^{-X}, is the potential for $c\backslash\{X\}$ obtained from χ by integration over the states of X, i.e., $\chi^{-X}(\boldsymbol{y}) = \int \chi(x, \boldsymbol{y})\ dx$ for all $\boldsymbol{y} \in \Omega_{c\backslash\{X\}}$. If χ contains no Dirac delta functions, then the integral is the usual Riemann integral, and integration is done over Ω_X. If χ contains a Dirac delta function, then the integral has to follow the properties of Dirac delta functions. Some examples of integrals with Dirac delta functions are as follows (these examples follow from properties of Dirac delta functions described in the Appendix).

(i) $\int \delta(x - a)\, dx = 1$.

(ii) $\int \delta(x - a)\, f(x)\, dx = f(a)$, assuming f is continuous in a neighborhood of a.

(iii) $\int \delta(y - g(x))\, \delta(z - h(x))\, dx = |(d/dy)(g^{-1}(y))|\, \delta(z - h(g^{-1}(y)))$, assuming g is invertible and differentiable on Ω_X.

(iv) $\int \delta(y - g(x))\, f(x)\, dx = |(d/dy)(g^{-1}(y))|\, f(g^{-1}(y))$, assuming g is invertible and differentiable on Ω_X.

(v) $\int \delta(y - g(x))\, \delta(z - h(x))\, f(x)\, dx = |(d/dy)(g^{-1}(y))|\, \delta(z - h(g^{-1}(y)))\, f(g^{-1}(y))$, assuming g is invertible and differentiable on Ω_X.

If α is a conditional associated with A, and its domain is a (i.e., $a = \{A\} \cup pa(A)$), then α^{-A} is an identity potential for $a \backslash \{A\} = pa(A)$, i.e., if β is any potential whose domain contains $a \backslash \{A\}$, then $\alpha^{-A} \otimes \beta = \beta$.

To reverse an arc (X, Y) in a BN, we compute the marginal $(\xi \otimes \psi)^{-X}$, where ξ is the conditional associated with X, and ψ is the conditional associated with Y. The potential $(\xi \otimes \psi)^{-X}$ represents the conditional for Y given $pa(X) \cup pa(Y) \backslash \{X\}$, and its nature (discrete or continuous) depends on Y. Thus, if Y is discrete, then $(\xi \otimes \psi)^{-X}$ is a discrete potential, and if Y is continuous or deterministic, then $(\xi \otimes \psi)^{-X}$ is a continuous potential.

Divisions. Arc reversals involve divisions of potentials, and the potential in the denominator is always a marginal of the potential in the numerator. Suppose (X, Y) is a reversible arc in a BN, suppose ξ is a potential for $\{X\} \cup pa(X)$ associated with X, and suppose ψ is a potential for $\{Y\} \cup pa(Y)$ associated with Y. After reversing the arc (X, Y), the revised potential associated with X is $(\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X}$. The definition of $(\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X}$ is as follows. $(\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X}$ is a potential for $\{Y\} \cup pa(X) \cup pa(Y)$ obtained from $(\xi \otimes \psi)$ and $(\xi \otimes \psi)^{-X}$ by point-wise division, i.e., $((\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X})(x, y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) = (\xi \otimes \psi)(x, y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) \,/\, ((\xi \otimes \psi)^{-X})(y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t})$ for all $x \in \Omega_X$, $y \in \Omega_Y$, $\boldsymbol{r} \in \Omega_{pa(X) \backslash pa(Y)}$, $\boldsymbol{s} \in \Omega_{pa(X) \cap pa(Y)}$, $\boldsymbol{t} \in \Omega_{pa(Y) \backslash (\{X\} \cup pa(X))}$. Notice that if $((\xi \otimes \psi)^{-X})(y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) = 0$, then $(\xi \otimes \psi)(x, y, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}) = 0$. In this case, we will simply define 0/0 as 0.

The quotient $(\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X}$ represents the conditional for X given $pa(X) \cup pa(Y) \cup \{Y\}$, and its nature depends on X. Thus, if X is discrete, then $(\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X}$ is a discrete potential (whose values are probabilities), and if X is continuous, then $(\xi\otimes\psi)\oslash(\xi\otimes\psi)^{-X}$ is a continuous potential (whose values are either probability densities or probabilities or both).

For an example of division, consider the BN shown in Figure 2 consisting of two continuous variables X and Y, where X has PDF $f(x)$, and Y is a deterministic function of X, say $Y = g(X)$, where g is invertible and differentiable in Ω_X. Let ξ and ψ denote the density and Dirac potentials associated with X and Y, respectively. Then $\xi(x) = f(x)$, and $\psi(x, y) = \delta(y - g(x))$. After reversal of the arc (X, Y), the revised potential associated with Y is $\psi'(y) = (\xi\otimes\psi)^{-X}(y) = \int f(x)\, \delta(y - g(x))\, dx = |(d/dy)(g^{-1}(y))|\, f(g^{-1}(y))$. After arc reversal, the revised potential associated with X is $\xi' = (\xi\otimes\psi)\oslash(\xi\otimes\psi)^{-X}$. Thus,

$$\xi'(x, y) = (\xi\otimes\psi)(x, y)/(\xi\otimes\psi)^{-X}(y) = f(x)\, \delta(y - g(x)) \,/\, (|(d/dy)(g^{-1}(y))|\, f(g^{-1}(y))) = \delta(x - g^{-1}(y)),$$

which is a Dirac potential. Notice that after arc reversal, X is deterministic. This is a consequence of the deterministic function at Y being invertible. As we will show later, if the deterministic function is not invertible, then after arc reversal, X may not be a deterministic variable. Notice that $\xi\otimes\psi = \xi'\otimes\psi'$, i.e., $f(x)\, \delta(y - g(x)) = \delta(x - g^{-1}(y))\, |(d/dy)(g^{-1}(y))|\, f(g^{-1}(y))$.

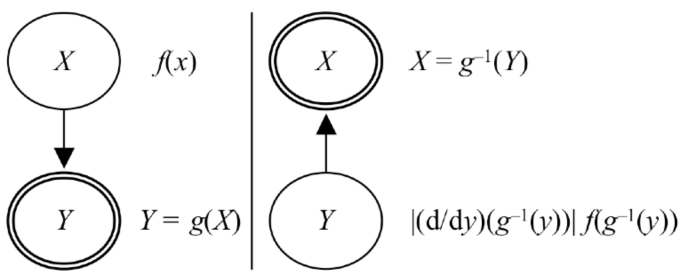

Figure 2. Arc reversal between a continuous and a deterministic variable with an invertible and differentiable function.

3 Arc Reversals

This section describes arc reversals between every possible kinds of pairs of variables. As we mentioned in the introduction, arc reversals were pioneered by Olmsted [1984] and studied extensively by Shachter [1986, 1988, 1990] for discrete Bayesian networks and influence diagrams. Here we draw on the literature and extend the arc-reversal theory to the case where we have continuous and detrministic variables in addition to discrete ones.

Given a BN graph, i.e., a directed acyclic graph, there always exists a sequence of variables such that whenever there is an arc (X, Y) in the network, X precedes Y in the sequence. An arc (X, Y) can be reversed only if there exists a sequence such that X and Y are adjacent in the sequence.

In a BN, each variable is associated with a conditional potential representing the conditional distribution for it given its parents. A fundamental assumption of the BN theory is that the combination of all the conditional potentials is the joint distribution of all variables in the network. Suppose (X, Y) is an arc in a BN such that X and Y are adjacent, and suppose ξ and ψ are the potentials associated with X and Y, respectively. Let $pa(X) = r\cup s$, and $pa(Y) = \{X\}\cup s\cup t$. Since X and Y are adjacent, the variables in $r\cup s\cup t$ precede X and Y in a sequence compatible with the arcs. Then $\xi\otimes\psi$ represents the conditional joint distributions of $\{X, Y\}$ given $r\cup s\cup t$, $(\xi\otimes\psi)^{-X}$ represents the conditional distributions of Y given $r\cup s\cup t$, and $(\xi\otimes\psi)\oslash(\xi\otimes\psi)^{-X}$ represents the conditional distributions of X given $r\cup s\cup t\cup\{Y\}$. If the arc (X, Y) is reversed, the potentials ξ and ψ associated with X and Y are replaced by potentials $\xi' = (\xi\otimes\psi)\oslash(\xi\otimes\psi)^{-X}$, and $\psi' = (\xi\otimes\psi)^{-X}$, respectively. This general case is illustrated in Figure 3. Although X and Y are shown as continuous nodes, they can each be discrete or deterministic.

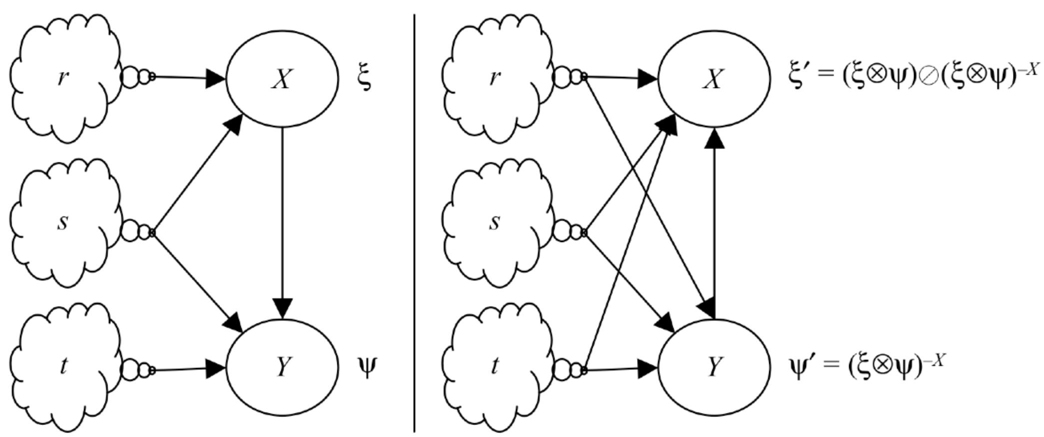

Figure 3. Reversal of arc (X, Y)

Some observations about the arc reversal process are as follows. First, arc reversal is a local operation that affects only the potentials associated with the two variables defining the arc. The potentials associated with the other variables remain unchanged.

Second, notice that $\xi\otimes\psi = \xi'\otimes\psi'$. Thus, the joint conditional distributions of $\{X, Y\}$ given $r\cup s\cup t$ remain unchanged by arc reversal. Also, since the other potentials for $r\cup s\cup t$ do not change, the joint distribution of all variables in a BN remains unchanged.

Third, for any potential α, let $dom(\alpha)$ denote the domain of α. Notice that the $dom(\xi') = dom(\xi) \cup dom(\psi) = r \cup s \cup t \cup \{X\} \cup \{Y\}$, and the $dom(\psi') = dom(\xi) \cup dom(\psi) \setminus \{X\} = r \cup s \cup t \cup \{Y\}$. Thus after arc reversal, X and Y inherit each other's parents, Y loses X as a parent, and X gains Y as a parent. As we will see, there are exceptions to this general rule when either X or Y (or both) are deterministic.

Fourth, suppose we reverse the arc (Y, X) in the revised BN. Let ξ'' and ψ'' denote the potentials associated with X and Y after reversal of arc (Y, X). Then

$$\xi'' = (\xi' \otimes \psi')^{-Y} = ((\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X} \otimes (\xi \otimes \psi)^{-X})^{-Y} = (\xi \otimes \psi)^{-Y} = \xi \otimes (\psi^{-Y}) = \xi \otimes \iota_{pa(Y)}, \text{ and}$$

$$\psi'' = (\xi' \otimes \psi') \oslash (\xi' \otimes \psi')^{-Y} = (\xi \otimes \psi) \oslash (\xi \otimes \iota_{pa(Y)}) = \psi \otimes \iota_{\{X\} \cup pa(X)}.$$

If we ignore the identity potentials (since these have no effect on the joint distribution), ξ'' and ψ'' are the same as ξ and ψ, what we started with.

In the remainder of this section, we will describe arc reversals between all kinds of pairs of variables. For each pair, we will assume that their parents are all continuous and that they have a parent in common. Also, we will assume that prior to arc reversal, continuous variables have conditional probability density functions represented by density potentials, deterministic variables are associated with equations represented by Dirac potentials, and discrete variables have conditional probability mass functions represented by discrete potentials. Thus, the nine cases described in this section should be viewed as examples rather than an exhaustive list of cases of arc reversals. The framework described in Section 2 should allow us to describe arc reversals between any pair of variables assuming that a closed form exists for describing the results of arc reversals.

3.1 Two Discrete Variables

In this section we describe reversal of an arc between two discrete nodes. This is the standard case and we discuss it here only for completeness.

Consider the BN given on the left-hand side of Figure 4. Let α and β denote the discrete potentials associated with variables A and B, respectively, before arc reversal, and α' and β' after arc reversal. Then, for all $b_j \in \Omega_B$, and $a_i \in \Omega_A$,

$$\alpha(u, v, a_i) = P(a_i | u, v),$$

$$\beta(v, w, a_i, b_j) = P(b_j | v, w, a_i),$$

$$\beta'(u, v, w, b_j) = (\alpha \otimes \beta)^{-A}(u, v, w, b_j) = \Sigma\{P(a_i | u, v)\, P(b_j | v, w, a_i) \mid a_i \in \Omega_A\}, \text{ and}$$

$$\alpha'(u, v, w, a_i, b_j) = ((\alpha\otimes\beta)\oslash(\alpha\otimes\beta)^{-A})(u, v, w, a_i, b_j)$$

$$= P(a_i|u, v)\, P(b_j|v, w, a_i) \,/\, \Sigma\{P(a_i|u, v)\, P(b_j|v, w, a_i) \mid a_i \in \Omega_A\}.$$

The resulting BN is given on the right-hand side of Figure 4.

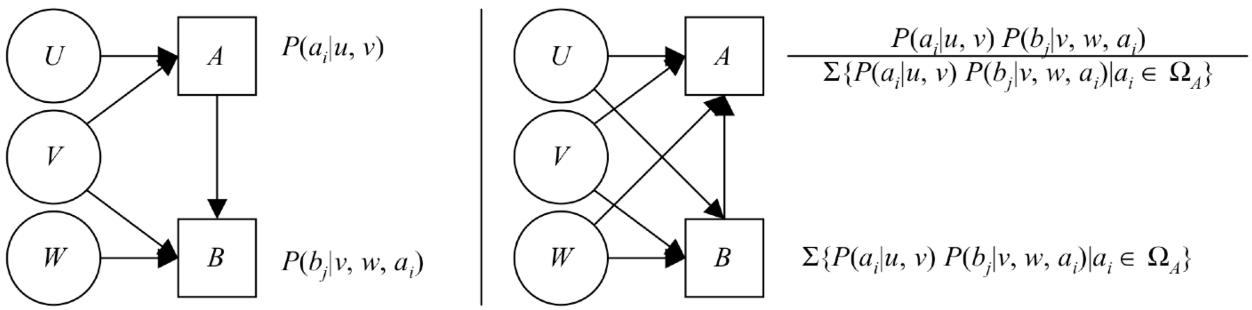

Figure 4. Arc reversal between two discrete nodes.

3.2 Two Continuous Variables

In this section, we describe arc reversals between two continuous variables. Consider the BN given on the left-hand side of Figure 5. In this BN, X has conditional PDF $f(x \mid u, v)$ and Y has conditional PDF $g(y \mid v, w, x)$. Let ξ and ψ denote the continuous potentials at X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then,

$$\xi(u, v, x) = f(x \mid u, v),$$

$$\psi(v, w, x, y) = g(y \mid v, w, x),$$

$$\psi'(u, v, w, y) = (\xi\otimes\psi)^{-X}(u, v, w, y) = \int f(x \mid u, v)\, g(y \mid v, w, x)\, \mathrm{d}x,$$

$$\xi'(u, v, w, x, y) = ((\xi\otimes\psi)\oslash(\xi\otimes\psi)^{-X})(u, v, w, x, y) = f(x \mid u, v)\, g(y \mid v, w, x) \,/\, (\int f(x \mid u, v)\, g(y \mid v, w, x)\, \mathrm{d}x).$$

The resulting BN is shown on the right-hand side of Figure 5.

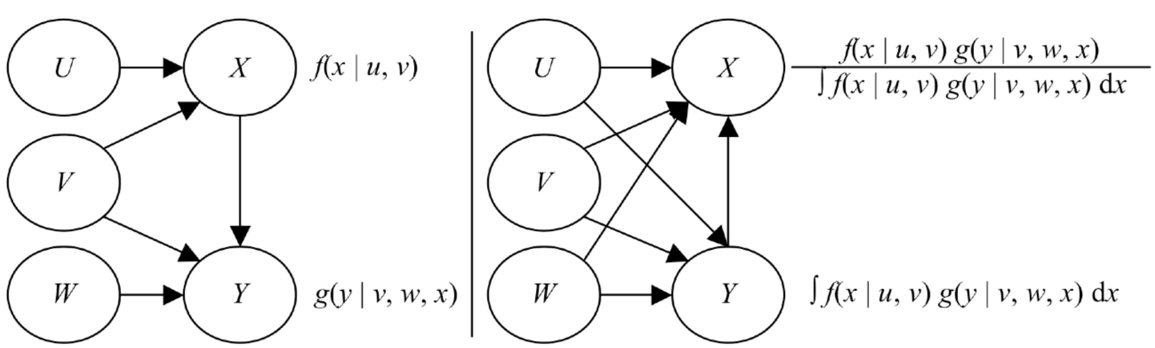

Figure 5. Arc reversal between two continuous nodes.

3.3 Continuous to Deterministic

As we have already discussed, the arc reversal between a continuous and a deterministic variable is slightly different from the arc reversal between two continuous variables since their joint PDF does not exist. After arc reversal, we transfer the density from the continuous node to the deterministic node, which results in the deterministic node being continuous and the continuous node having a Dirac potential.

Consider the situation shown in Figure 6. In this BN, X has continuous parents U and V, and Y has continuous parents V and W in addition to X. The density at X is f and the equation at Y is $Y = h(V, W, X)$. We assume h is invertible in X and differentiable on Ω_X. The potentials before and after arc reversals are as follows.

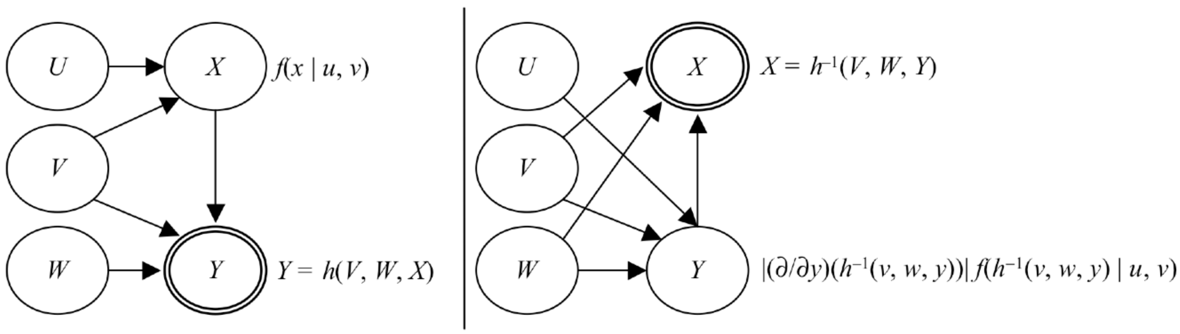

Figure 6. Arc reversal between a continuous and a deterministic variable.

$$
\begin{aligned}
\xi(u, v, x) &= f(x \mid u, v), \\
\psi(v, w, x, y) &= \delta(y - h(v, w, x)), \\
\psi'(u, v, w, y) &= (\xi \otimes \psi)^{-X}(u, v, w, y) = \int f(x \mid u, v)\, \delta(y - h(v, w, x))\, dx \\
&= |(\partial/\partial y)(h^{-1}(v, w, y))|\, f(h^{-1}(v, w, y) \mid u, v), \text{ and} \\
\xi'(u, v, w, x, y) &= ((\xi \otimes \psi) \oslash (\xi \otimes \psi)^{-X})(u, v, w, x, y) \\
&= f(x \mid u, v)\, \delta(y - h(v, w, x)) \,/\, (|(\partial/\partial y)(h^{-1}(v, w, y))|\, f(h^{-1}(v, w, y) \mid u, v)) \\
&= \delta(x - h^{-1}(v, w, y)).
\end{aligned}
$$

After we reverse the arc (X, Y), both X and Y inherit each other's parents, but X loses U as a parent. Also, Y has a density function and X has a deterministic conditional distribution. The determinism of the conditional for X after arc reversal is a consequence of the invertibility of the relationship at Y before arc reversal. The resulting BN is given on right-hand side of Figure 6. Also, some of the qualitative conclusions here, namely X loses U as a parent, Y has a density function, and X has a deterministic conditional distribution, are based on the assumption that U, V, W are continuous. If any of these are discrete, the conclusion can change, as we will demonstrate in Section 4.

As an example, consider the BN consisting of two continuous variables and a deterministic variable whose function is the sum of its two parents as shown in Figure 7. $X \sim f(x)$, $Y|x \sim g(y \mid x)$, and $Z = X + Y$. Let ξ, ψ, ζ denote the potentials associated with X, Y, and Z, respectively, before arc reversal, and ψ' and ζ' denote the revised potentials associated with Y and Z, respectively, after reversal of arc (Y, Z). Then,

$$\xi(x) = f(x),$$

$$\psi(x, y) = g(y \mid x),$$

$$\zeta(x, y, z) = \delta(z - x - y),$$

$$\zeta'(x, z) = (\psi \otimes \zeta)^{-Y}(x, z) = \int g(y \mid x)\, \delta(z - x - y)\, dy = \int g(y \mid x)\, \delta(y - (z - x))\, dy = g(z - x \mid x), \text{ and}$$

$$\psi'(x, y, z) = ((\psi \otimes \zeta) \oslash (\psi \otimes \zeta)^{-Y})(x, y, z) = g(y \mid x)\, \delta(z - x - y) \,/\, g(z - x \mid x) = \delta(y - (z - x)).$$

If we reverse the arc (X, Z) in the revised BN, we obtain the marginal distribution of Z,

$$\zeta''(z) = (\xi \otimes \zeta')^{-X}(z) = \int f(x)\, g(z - x \mid x)\, dx,$$

which is the convolution formula for Z. The revised potential at X,

$$\xi'(x, z) = ((\xi \otimes \zeta') \oslash (\xi \otimes \zeta')^{-X})(x, z) = f(x)\, g(z - x \mid x)/(\int f(x)\, g(z - x \mid x)\, dx),$$

represents the conditional distribution of X given z.

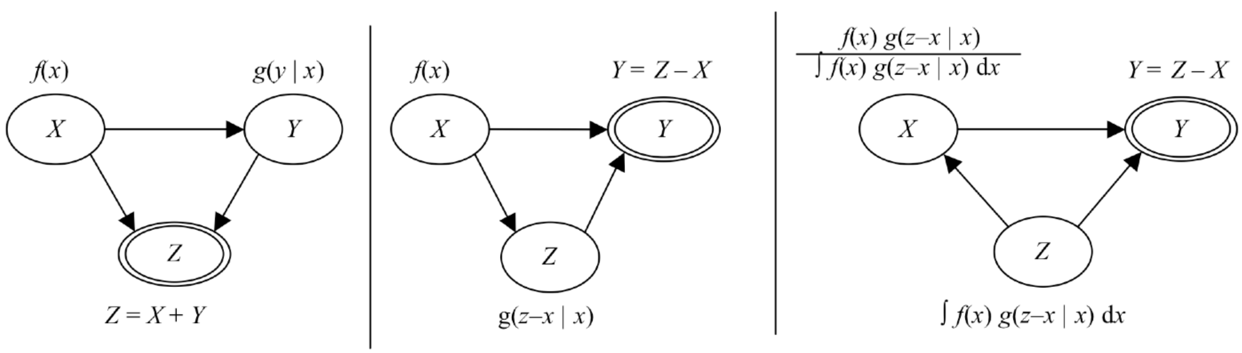

Figure 7. A continuous BN with a deterministic variable.

We have assumed that the function describing the deterministic variable is invertible and differentiable. Let us consider the case where the function is not invertible, but has known simple zeros, and is differentiable. For example, consider a BN with two continuous variables X and Y, where X has PDF $f(x)$ and Y is a deterministic function of X described by the function $Y = X^2$ as shown in Figure 8.

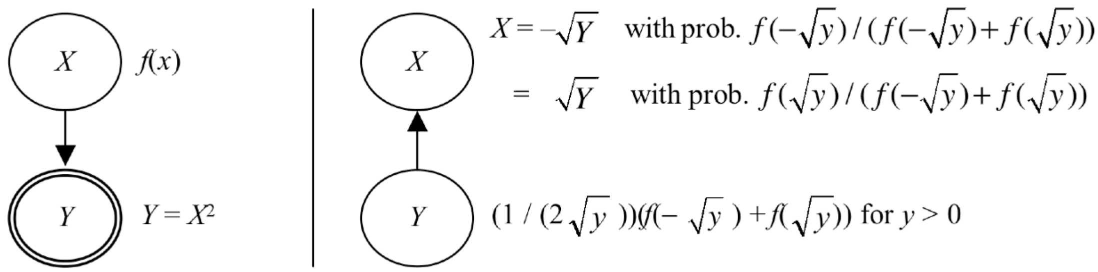

Figure 8. Arc reversal between a continuous node and a deterministic node with a non-invertible function.

This function is not invertible, but $y - x^2$ has two simple zeros at $x = \pm\sqrt{y}$. Suppose ξ and ψ denote the continuous potentials at X and Y, respectively, before arc reversal, and ξ′ and ψ′ after arc reversal. Then

$$\xi(x) = f(x),$$

$$\psi(x, y) = \delta(y - x^2) = \delta(x^2 - y) = (\delta(x + \sqrt{y}) + \delta(x - \sqrt{y}))/(2\sqrt{y})$$

$$\psi'(y) = (\xi \otimes \psi)^{-X}(y) = \int f(x)\,(\delta(x + \sqrt{y}) + \delta(x - \sqrt{y}))/(2\sqrt{y})\,\mathrm{d}x$$

$$= (f(-\sqrt{y}) + f(\sqrt{y}))\,/\,(2\sqrt{y}), \text{ for all } y > 0.$$

$$\xi'(x, y) = f(x)\,(\delta(x + \sqrt{y}) + \delta(x - \sqrt{y}))\,/\,(f(-\sqrt{y}) + f(\sqrt{y}))$$

$$= (f(-\sqrt{y})\,\delta(x + \sqrt{y}) + f(\sqrt{y})\,\delta(x - \sqrt{y})\,/\,(f(-\sqrt{y}) + f(\sqrt{y})), \text{ for all } y > 0.$$

Notice that the revised conditional for X is not deterministic if $f(-\sqrt{y}) > 0$ and $f(\sqrt{y}) > 0$, but it is a Dirac potential. The revised potential for Y is a density potential.

If the deterministic function is such that it's zeros are not so easily determined or it is not differentiable, then we would not be able to write a closed form expression for the distributions of X and Y after arc reversal.

3.4 Deterministic to Continuous

In this subsection, we describe arc reversal between a deterministic and a continuous variable. Consider a BN as shown on the left-hand side of Figure 9. X is a deterministic variable associated with a function, $X = h(U, V)$, and Y is a continuous variable and the conditional distribution of $Y|(v, w, x)$ is distributed as $g(y\,|v, w, x)$. Suppose we wish to reverse the arc (X, Y). Since there is no density potential at X, Shenoy [2006] suggests to first reverse arc (U, X) or (V, X) (resulting in a density potential at X), and then reverse arc (X, Y) using the rules for arc reversal between two continuous nodes. However, here we show that it is possible to reverse an arc between a deterministic node and a continuous node directly without having to reverse other arcs.

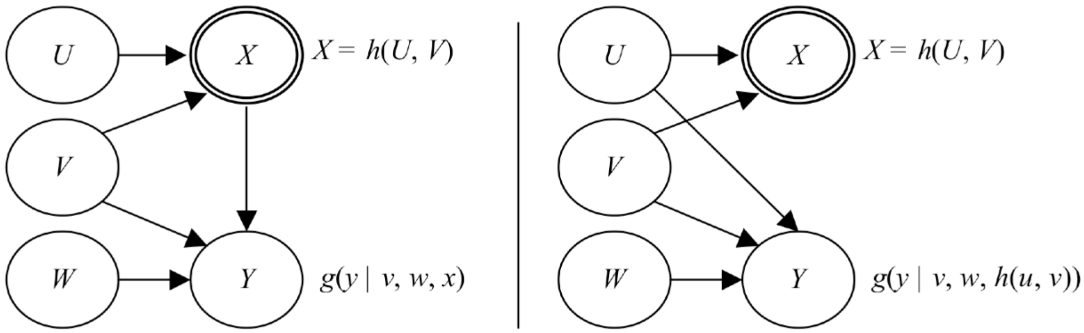

Figure 9. Arc reversal between a deterministic and a continuous node.

Consider again the BN given on left-hand side of Figure 9. Suppose we wish to reverse the arc (X, Y). Let ξ and ψ denote the continuous potentials at X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then,

$$\xi(u, v, x) = \delta(x - h(u, v)),$$

$$\psi(v, w, x, y) = g(y \mid v, w, x),$$

$$\psi'(u, v, w, y) = ((\xi \otimes \psi)^{-X})(u, v, w, y) = \int \delta(x - h(u, v))\, g(y \mid v, w, x)\, \mathrm{d}x = g(y \mid v, w, h(u, v)), \text{ and}$$

$$\xi'(u, v, w, x, y) = (\xi \otimes \psi) \oslash ((\xi \otimes \psi)^{-X})(u, v, w, x, y) = \delta(x - h(u, v))\, g(y \mid v, w, x) \,/\, g(y \mid v, w, h(u, v))$$

$$= \delta(x - h(u, v)).$$

N otice that ξ' does not depend on either W or Y. Thus, after arc reversal, there is no arc from Y to X, i.e., the arc being reversed disappears, and X does not inherit an arc from W. The resulting BN is shown on the right-hand side of Figure 9.

3.5 Deterministic to Deterministic

In this subsection, we describe arc reversal between two deterministic variables. Consider the BN on the left-hand side of Figure 10. X is a deterministic function of its parents $\{U, V\}$, and Y is also a deterministic function of its parents $\{X, V, W\}$. Suppose we wish to reverse the arc (X, Y). Let ξ and ψ denote the potentials associated with X and Y, respectively, before arc reversal, and ξ' and ψ' after arc reversal. Then,

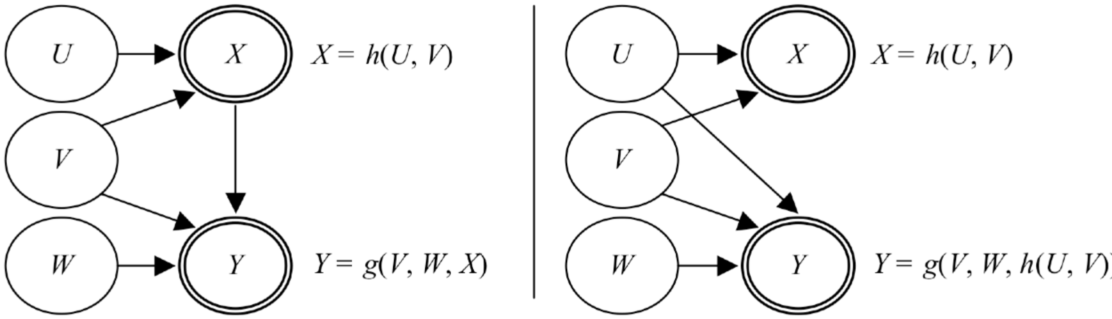

Figure 10. Arc reversal between two deterministic nodes.

$$\xi(u, v, x) = \delta(x - h(u, v)),$$

$$\psi(v, w, x, y) = \delta(y - g(v, w, x)),$$

$$\psi'(u, v, w, y) = (\xi\otimes\psi)^{-X}(u, v, w, y) = \int \delta(x - h(u, v))\, \delta(y - g(v, w, x))\, \mathrm{d}x$$

$$= \delta(y - g(v, w, h(u, v))), \text{ and}$$

$$\xi'(u, v, w, x, y) = ((\xi\otimes\psi)\oslash(\xi\otimes\psi)^{-X})(u, v, w, x, y)$$

$$= \delta(x - h(u, v))\, \delta(y - g(v, w, x)) / \delta(y - g(v, w, h(u, v))) = \delta(x - h(u, v)).$$

Notice that ξ' does not depend on either Y or W. The arc being reversed disappears, and X does not inherit a parent of Y.

3.6 Continuous to Discrete

In this section, we will describe arc reversal between a continuous and a discrete node. Consider the BN as shown in Figure 11. X is a continuous node with conditional PDF $f(x \mid u, v)$, and A is a discrete node with conditional masses $P(a_i \mid v, w, x)$ for each $a_i \in \Omega_A$. Let ξ and α denote the density and discrete potentials associated with X and A, respectively, before arc reversal, and ξ' and α' after arc reversal. Then

$$\xi(u, v, x) = f(x \mid u, v),$$

$$\alpha(v, w, x, a_i) = P(a_i \mid v, w, x),$$

$$\alpha'(u, v, w, a_i) = (\xi\otimes\alpha)^{-X}(u, v, w, a_i) = \int f(x \mid u, v)\, P(a_i \mid v, w, x)\, \mathrm{d}x, \text{ and}$$

$$\xi'(u, v, w, x, a_i) = ((\xi\otimes\alpha)\oslash(\xi\otimes\alpha)^{-X})(u, v, w, x, a_i)$$

$$= f(x \mid u, v)\, P(a_i \mid v, w, x) / (\int f(x \mid u, v)\, P(a_i \mid v, w, x)\, \mathrm{d}x).$$

The BN on the RHS of Figure 11 depicts the results after arc reversal.

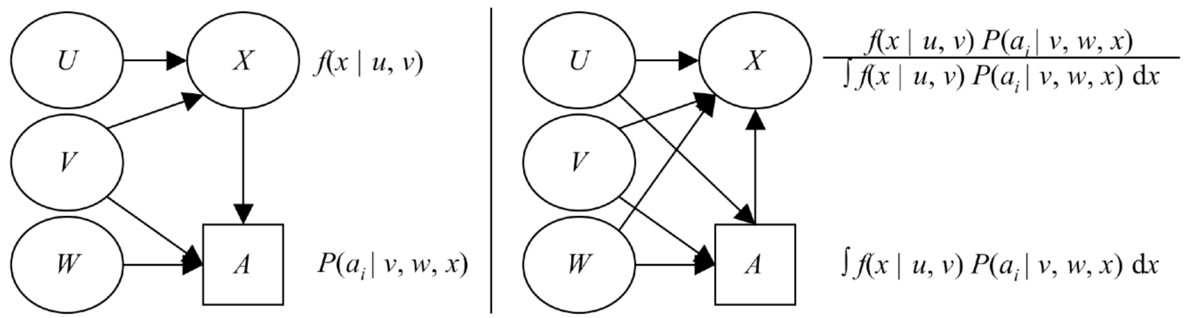

Figure 11. Arc reversal between a continuous and a discrete node.

For a concrete example, consider the simpler hybrid BN shown on the LHS of Figure 12. X is a continuous variable, distributed as $N(0, 1)$. A is a discrete variable with two states $\{a_1, a_2\}$. The conditional probability mass functions of A are as follows: $P(a_1 \mid x) = 1/(1 + e^{-2x})$ and $P(a_2 \mid x) = e^{-2x}/(1 + e^{-2x})$. Let α and ξ denote the potentials associated with A and X, respectively, before arc reversal, and α' and ξ' after arc reversal. Then,

$$\alpha(a_1, x) = 1/(1 + e^{-2x}),$$

$$\alpha(a_2, x) = e^{-2x}/(1 + e^{-2x}),$$

$$\xi(x) = \varphi_{0,1}(x), \text{ where } \varphi_{0,1}(x) \text{ is the PDF of the standard normal distribution,}$$

$$\alpha'(a_1) = (\alpha\otimes\xi)^{-X}(a_1) = \int (1/(1 + e^{-2x}))\, \varphi_{0,1}(x)\, dx = 0.5,$$

$$\alpha'(a_2) = (\alpha\otimes\xi)^{-X}(a_2) = \int (e^{-2x}/(1 + e^{-2x})\, \varphi_{0,1}(x)\, dx = 0.5,$$

$$\xi'(a_1, x) = ((\alpha\otimes\xi)\oslash(\alpha\otimes\xi)^{-X})(a_1, x) = (1/(1 + e^{-2x}))\, \varphi_{0,1}(x))/0.5 = (2/(1 + e^{-2x}))\, \varphi_{0,1}(x),$$

$$\xi'(a_2, x) = (\alpha\otimes\xi)\oslash(\alpha\otimes\xi)^{-X}(a_2, x) = (e^{-2x}/(1 + e^{-2x}))\, \varphi_{0,1}(x))/0.5 = (2e^{-2x}/(1 + e^{-2x}))\, \varphi_{0,1}(x),$$

The resulting BN after the arc reversal is given on the RHS of Figure 12.

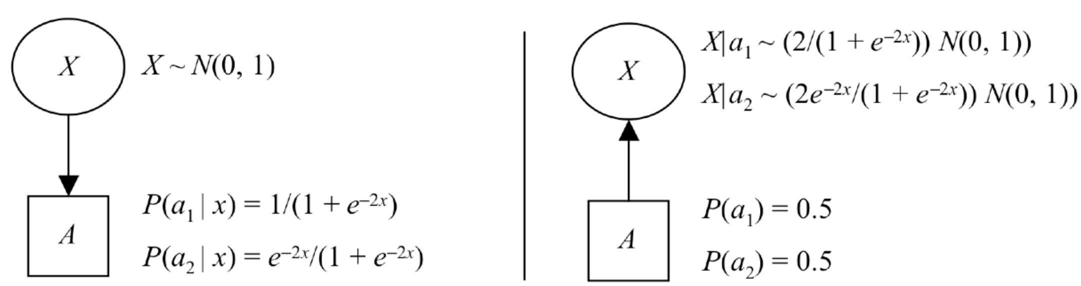

Figure 12. Arc reversal between a continuous and a discrete node.

3.7 Deterministic to Discrete

In this subsection, we describe reversal of an arc between a deterministic and a discrete variable. Consider the hybrid BN shown on the left-hand side of Figure 13. Let ξ and α denote the potentials at X and A, respectively, before arc reversal, and let ξ' and α' denote the potentials after arc reversal. Then,

$$\xi(u, v, x) = \delta(x - h(u, v)),$$

$$\alpha(v, w, x, a_i) = P(a_i \mid v, w, x),$$

$$\alpha'(u, v, w, a_i) = \int \delta(x - h(u, v))\, P(a_i \mid v, w, x)\, \mathrm{d}x = P(a_i \mid v, w, h(u, v)), \text{ and}$$

$$\xi'(u, v, w, x, a_i) = \delta(x - h(u, v))\, P(a_i \mid v, w, x)/P(a_i \mid v, w, h(u, v)) = \delta(x - h(u, v)).$$

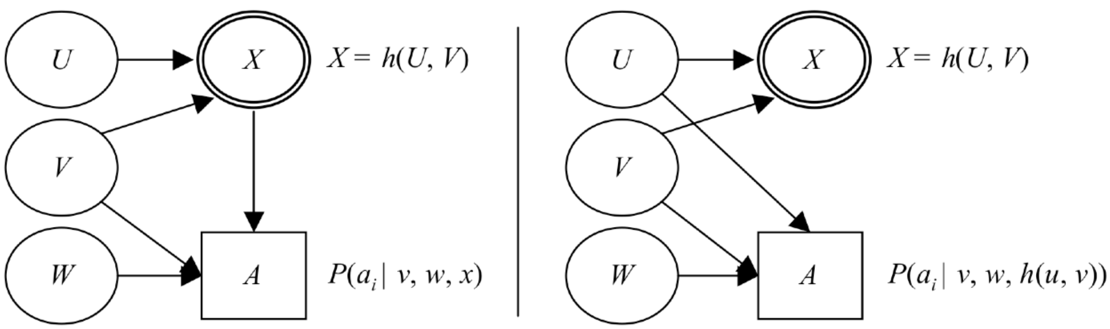

Figure 13. Arc reversal between a deterministic and a discrete variable.

Notice that ξ' depends on neither A nor W. The illustration of an arc reversal between a deterministic and discrete node with parents is given in Figure 13.

For a concrete example, consider the BN given on the LHS of Figure 14. The continuous variable $V \sim$ U[0, 2], deterministic variable $X = V^2$, and discrete variable A with two states $\{a_1, a_2\}$ has the conditional distribution $P(a_1 \mid v, x) = 1$ if $v \le x$, and $P(a_1 \mid v, x) = 0$ if $v > x$. Let ξ and α denote the potentials associated with X and A, respectively, before arc reversal, and ξ' and α' after arc reversal. Then,

$$\xi(v, x) = \delta(x - v^2)$$

$$\alpha(a_1, v, x) = P(a_1 \mid v, x) \begin{array}{l} = 1 \text{ if } v \le x \\ = 0 \text{ if } v > x, \end{array}$$

$$\alpha(a_2, v, x) = P(a_2 \mid v, x) \begin{array}{l} = 0 \text{ if } v \le x \\ = 1 \text{ if } v > x, \end{array}$$

$$\alpha'(a_1, v) = \int \delta(x - v^2)\, \alpha(a_1, v, x)\, \mathrm{d}x = \alpha(a_1, v, v^2) = P(a_1 \mid v) \begin{array}{l} = 1 \text{ if } v \le v^2, \text{ and} \\ = 0 \text{ if } v > v^2, \end{array}$$

$$\alpha'(a_2, v) = \int \delta(x - v^2)\, \alpha(a_2, v, x)\, \mathrm{d}x = \alpha(a_2, v, v^2) = P(a_2 \mid v) \begin{array}{l} = 0 \text{ if } v \le v^2, \text{ and} \\ = 1 \text{ if } v > v^2, \end{array}$$

$$\xi'(a_1, v, x) = \delta(x - v^2)\ \alpha(a_1, v, x)/\alpha(a_1, v, v^2) = \delta(x - v^2)$$
$$\xi'(a_2, v, x) = \delta(x - v^2)\ \alpha(a_2, v, x)/\alpha(a_2, v, v^2) = \delta(x - v^2)$$

The situation after arc-reversal is shown in the RHS of Figure 14.

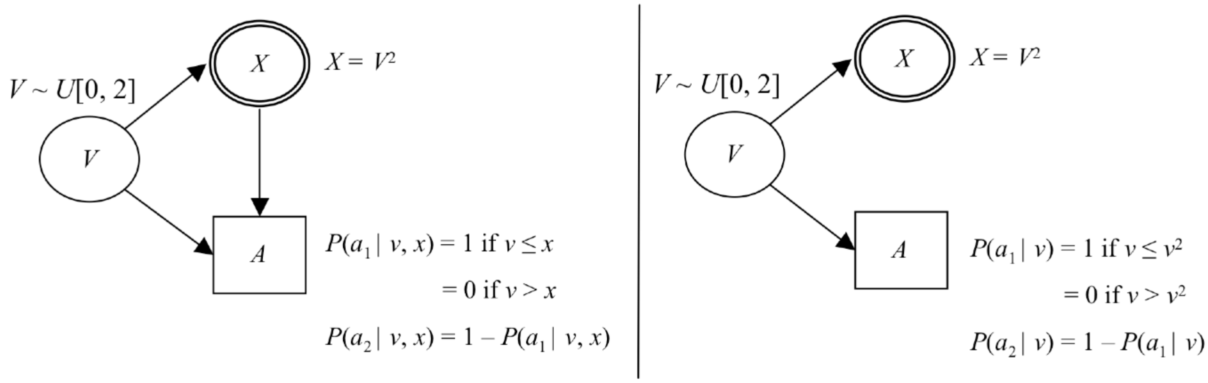

Figure 14. An example of arc-reversal between a deterministic and a discrete variable.

3.8 Discrete to Continuous

In this subsection, we describe reversal of an arc from a discrete to a continuous variable. Consider the hybrid BN shown on the LHS of Figure 15. Let α and ξ denote the potentials associated with A and X, respectively, before arc reversal, and α' and ξ' after arc reversal. Then,

$$\alpha(u, v, a_i) = P(a_i \mid u, v),$$
$$\xi(v, w, x, a_i) = f_i(x \mid v, w),$$
$$\xi'(u, v, w, x) = (\alpha\otimes\xi)^{-A}(u, v, w, x) = \Sigma\{P(a_i \mid u, v)\ f_i(x \mid v, w) \mid a_i \in \Omega_A\},$$
$$\alpha'(u, v, w, x, a_i) = ((\alpha\otimes\xi)\oslash(\alpha\otimes\xi)^{-A})(u, v, w, x, a_i)$$
$$= P(a_i \mid u, v)\ f_i(x \mid v, w)/\Sigma\{P(a_i \mid u, v)\ f_i(x \mid v, w) \mid a_i \in \Omega_A\}.$$

The density at X after arc reversal is a mixture density.

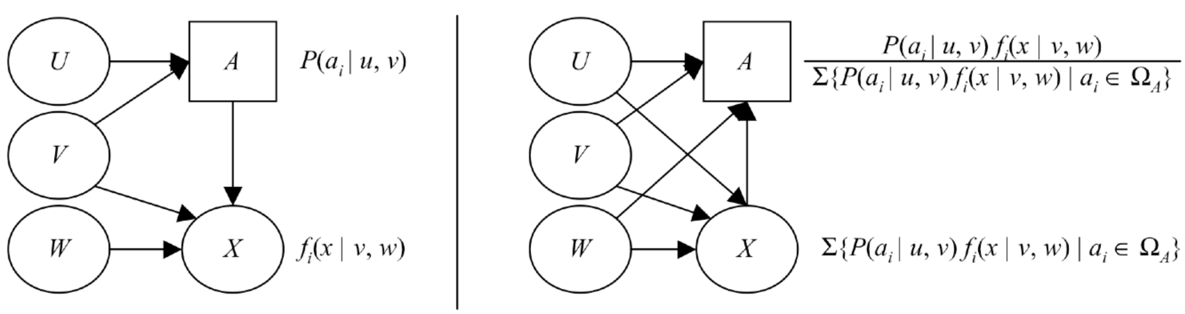

Figure 15. Arc reversal between a discrete and a continuous variable.

For a concrete example, consider the BN given on the LHS of Figure 16. The discrete variable A has two states $\{a_1, a_2\}$ with $P(a_1) = 0.5$ and $P(a_2) = 0.5$. X is a continuous variable whose conditional distributions are $X|a_1 \sim N(0, 1)$ and $X|a_2 \sim N(2, 1)$. Let α and ξ denote the potentials associated with A and X, respectively, before arc reversal, and α' and ξ' after arc reversal. Then,

$$\alpha(a_1) = 0.5,$$

$$\alpha(a_2) = 0.5,$$

$$\xi(a_1, x) = \varphi_{0,1}(x),$$

$$\xi(a_2, x) = \varphi_{2,1}(x),$$

$$\xi'(x) = (\alpha \otimes \xi)^{-A}(x) = 0.5\ \varphi_{0,1}(x) + 0.5\ \varphi_{2,1}(x),$$

$$\alpha'(a_1, x) = ((\alpha \otimes \xi) \oslash (\alpha \otimes \xi)^{-A})(a_1, x) = (0.5\ \varphi_{0,1}(x))/(0.5\ \varphi_{0,1}(x) + 0.5\ \varphi_{2,1}(x)),$$

$$\alpha'(a_2, x) = ((\alpha \otimes \xi) \oslash (\alpha \otimes \xi)^{-A})(a_2, x) = (0.5\ \varphi_{2,1}(x))/(0.5\ \varphi_{0,1}(x) + 0.5\ \varphi_{2,1}(x)),$$

The resulting BN after the arc reversal is given on the RHS of Figure 16.

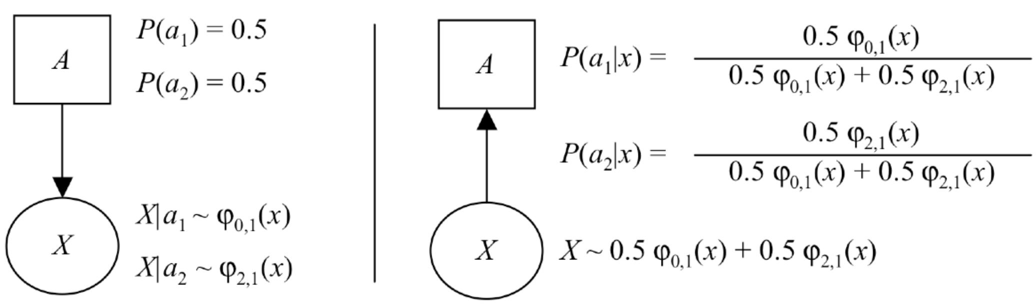

Figure 16. An example of an arc reversal between a discrete and a continuous variable.

3.9 Discrete to Deterministic

In this subsection, we describe reversal of an arc between a discrete and a deterministic variable. Consider the hybrid BN as shown on the left-hand side of Figure 17. Suppose that $\Omega_A = \{a_1, \ldots, a_k\}$. Let α and ξ denote the potentials associated with A and X, respectively, before arc reversal, and α' and ξ' after arc reversal. Then,

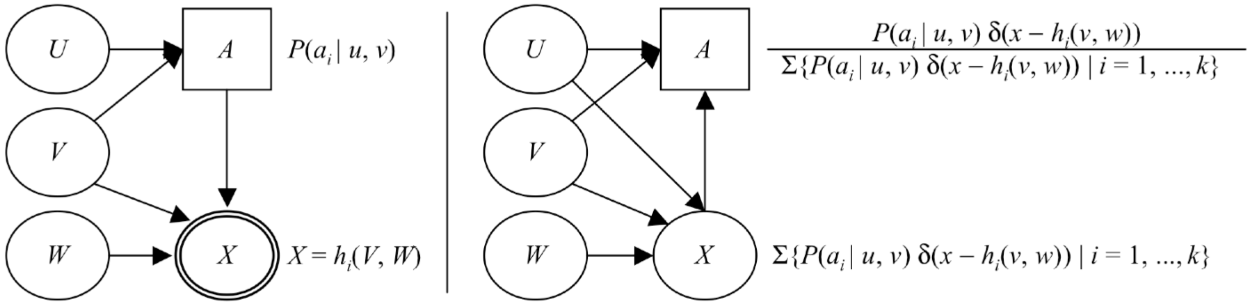

Figure 17. Arc reversal between a discrete and a deterministic variable.

$$\alpha(u, v, a_i) = P(a_i \mid u, v),$$

$$\xi(v, w, x, a_i,) = \delta(x - h_i(v, w)),$$

$$\xi'(u, v, w, x) = (\alpha \otimes \xi)^{-A}(u, v, w, x) = \Sigma\{P(a_i \mid u, v)\ \delta(x - h_i(v, w)) \mid i = 1, \ldots, k\},$$

$$\alpha'(u, v, w, x, a_i) = P(a_i \mid u, v)\ \delta(x - h_i(v, w)) \,/\, \Sigma\{P(a_i \mid u, v)\ \delta(x - h_i(v, w)) \mid i = 1, \ldots, k\}.$$

The situation after arc reversal is shown on the right-hand side of Figure 17. Notice that after arc reversal, X has a weighted sum of Dirac delta functions. Since the variances of the conditional for X after arc reversal may not be zeros, X may not be deterministic after arc reversal.

For a concrete example, consider the simpler hybrid BN shown on the LHS of Figure 18. V has the uniform distribution on (0, 1). A has two states $\{a_1, a_2\}$ with $P(a_1 \mid v) = 1$ if $0 < v \leq 0.5$, and = 0 otherwise, and $P(a_2 \mid v) = 1 - P(a_1 \mid v)$. X is deterministic with equations $X = V$ if $A = a_1$, and $X = -V$ if $A = a_2$. After arc reversal, the conditional distributions at A and X are as shown in the RHS of Figure 17 (these are special cases of the general formulae given in Figure 16). Let ϖ denote the density potential at V. Then $\varpi(v) = 1$ if $0 < v < 1$. We can find the marginal of X from the BN on the RHS of Figure 17 by reversing arc (V, X) as follows.

$$(\varpi \otimes \xi')^{-V}(x) = \int \varpi(v)\ \mathrm{P}(a_1 \mid v)\ \delta(x - v)\ \mathrm{d}v + \int \varpi(v)\ \mathrm{P}(a_2 \mid v)\ \delta(x + v)\ \mathrm{d}v = 1 \text{ if } 0 < x \leq 0.5 \text{ or } -1 < x < -0.5.$$

Thus, the marginal distribution of X is uniform on the interval $(-1, -0.5) \cup (0, 0.5)$.

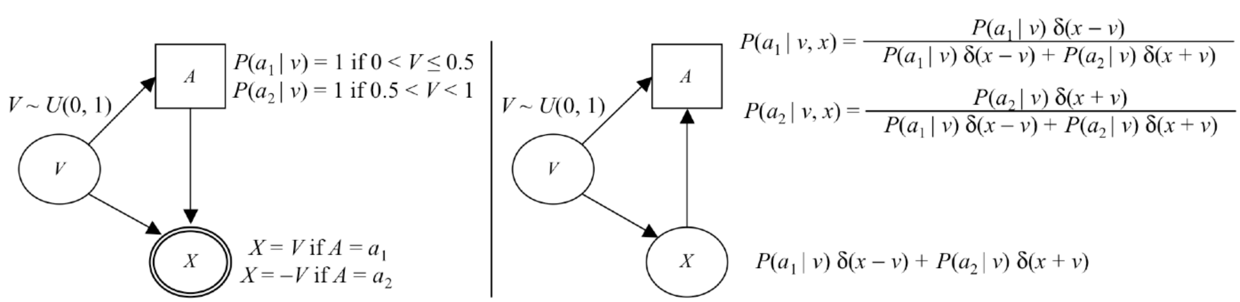

Figure 18. An example of arc reversal between a discrete and deterministic variable.

4 Partially Deterministic Distributions

In this section, we describe a new kind of conditional distribution called partially deterministic. Partially deterministic distributions arise in the process of arc reversals in hybrid BNs.

The conditional distributions associated with a deterministic variable have zero variances. If some of the conditional distributions have zero variances and some have positive variances, we say that the distribution is *partially deterministic*.

We get such distributions during the process of the arc reversals between a continuous node and a deterministic node with discrete and continuous parents. Consider the BN shown on the left-hand side of Figure 19. Let ξ and ζ denote the continuous potentials at X and Z, respectively, before arc reversal, and ξ' and ζ' after arc reversal. Then,

$$\xi(x) = f(x),$$

$$\zeta(x, y, z, a_1) = \delta(z - x) = \delta(x - z),$$

$$\zeta(x, y, z, a_2) = \delta(z - y) = \delta(y - z),$$

$$\zeta'(y, z, a_1) = (\xi\otimes\zeta)^{-X}(y, z, a_1) = \int f(x)\, \delta(x - z)\, dx = f(z),$$

$$\zeta'(y, z, a_2) = (\xi\otimes\zeta)^{-X}(y, z, a_2) = \delta(y - z) \int f(x)\, dx = \delta(y - z),$$

$$\xi'(x, y, z, a_1) = (\xi\otimes\zeta)\oslash(\xi\otimes\zeta)^{-X}(x, y, z, a_1) = f(x)\, \delta(x - z) / f(z) = f(z)\, \delta(x - z) / f(z) = \delta(z - x),$$

$$\xi'(x, y, z, a_2) = (\xi\otimes\zeta)\oslash(\xi\otimes\zeta)^{-X}(x, y, z, a_2) = f(x)\, \delta(y - z) / \delta(y - z) = f(x).$$

Thus, after arc reversal, both X and Z have partially deterministic distributions.

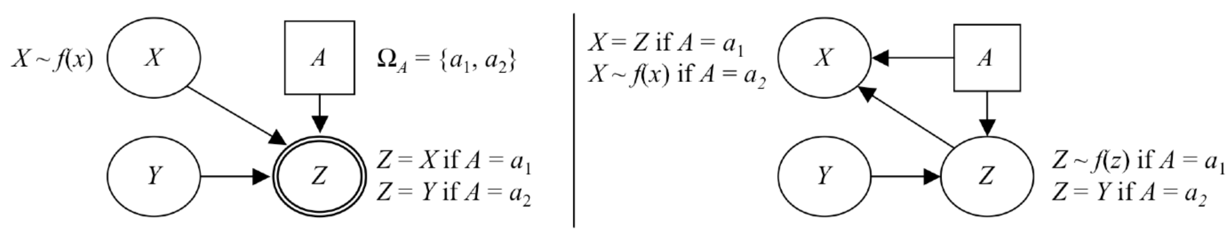

Figure 19. Arc reversal leading to partially deterministic distributions.

The significance of partially deterministic distributions is as follows. If we have a Bayesian network with all continuous variables such that each continuous variable is associated with a density potential, then we can propagate the density potentials similar to discrete potentials in a discrete Bayesian network. The only difference is that we use integration for marginalizing continuous variables (instead of summation for discrete variables). This assumes that the joint potential obtained by combining all density potentials represents the joint density for all variables in the Bayesian network. However, if even a single variable has a deterministic or a partially deterministic conditional, then the joint potential (obtained by combining all conditionals associated with the variables) no longer represents the joint density as the joint density does not exist. Thus, one cannot assume that in a Bayesian network with no deterministic variables (as in the case of RHS of Figure 19), that the joint density exists for all continuous variables in the network. It is clear from the Bayesian network in the LHS of Figure 19, that the joint density for $\{X, Y, Z\}$ (conditioned on the states of A) does not exist. And since the joint distributions of the two Bayesian networks are the same, the joint density for $\{X, Y, Z\}$ does not exist also for the Bayesian network in the RHS of Figure 19.

5 Conclusions and Summary

We have described arc reversals in hybrid BNs with deterministic variables between all possible kinds of pairs of variables. In some cases, there is no closed form for the distributions after arc reversals. For example, if a deterministic variable has a function that is not differentiable, then we cannot describe the distributions after arc reversal in closed form. We do believe, however, that the framework described in Section 2 is sufficient to describe arc reversals in those cases where there is a closed form for the revised distributions. Also, we have described a new kind of conditional distribution called partially deterministic that can arise after arc reversals.

The arc-reversal theory facilitates the task of approximating general BNs with mixture of Gaussians BNs. Also, the arc-reversal theory is potentially useful in solving hybrid influence diagrams, i.e., influence diagrams with discrete, continuous, and deterministic chance variables. We conjecture that Olmsted's arc-reversal algorithm for solving discrete influence diagrams would apply to hybrid influence diagrams also. The arc-reversal theory described here would make this possible. Of course, this is a topic that needs further investigation.

References

Cobb, B. R. and P. P. Shenoy (2005a), "Hybrid Bayesian networks with linear deterministic variables," in F. Bacchus and T. Jaakkola (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-First Conference* (UAI-05), 136–144, AUAI Press, Corvallis, OR.

Cobb, B. R. and P. P. Shenoy (2005b), "Nonlinear deterministic relationships in Bayesian networks," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (ECSQARU-05), L. Godo (ed.), Berlin, Springer-Verlag: 27–38.

Cobb, B. R. and P. P. Shenoy (2006), "Operations for inference in continuous Bayesian networks with linear deterministic variables," *International Journal of Approximate Reasoning*, **42**(1–2), 21–36.

Cobb, B. R., P. P. Shenoy, and R. Rumi (2006), "Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials," *Statistics and Computing*, **16**(3): 293–308.

Dirac, P. A. M. (1927), "The physical interpretation of the quantum dynamics," *Proceedings of the Royal Society of London*, *series A*, **113**(765), 621–641.

Dirac, P. A. M. (1958), *The Principles of Quantum Mechanics*, 4th ed., Oxford University Press, London.

Hoskins, R. F. (1979), *Generalised Functions*, Ellis Horwood, Chichester.

Kanwal, R. P. (1998), *Generalized Functions: Theory and Technique*, 2nd ed., Birkhäuser, Boston.

Kenley, C. R. (1986), "Influence diagram models with continuous variables," PhD dissertation, Dept. of Engineering-Economic Systems, Stanford University.

Khuri, A. I. (2004), "Applications of Dirac's delta function in statistics," *International Journal of Mathematical Education in Science and Technology*, **32**(2), 185–195.

Kozlov, A. V. and D. Koller (1997), "Nonuniform dynamic discretization in hybrid networks," in D. Geiger and P. P. Shenoy (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Thirteenth Conference* (UAI-97), 310–318, Morgan Kaufmann, San Francisco.

Lauritzen, S. L. and F. Jensen (2001), "Stable local computation with conditional Gaussian distributions," *Statistics and Computing*, **11**, 191–203.

Madsen, A. L. (2006), "Variations Over the Message Computation Algorithm of Lazy Propagation," *IEEE Transactions on Systems, Man, and Cybernetics*, part B: Cybernetics, **36**(3), 636–648.

Madsen, A. L. (2008), "Solving CLQG Influence Diagrams Using Arc-Reversal Operations in a Strong Junction Tree," in *Proceedings of the 4th European Workshop on Probabilistic Graphical Models*, 201–208.

Moral, S., R. Rumi, A. Salmeron (2001), "Mixtures of truncated exponentials in hybrid Bayesian networks," in *Symbolic and Quantitative Approaches to Reasoning under Uncertainty* (ECSQARU-2001), P. Besnard and S. Benferhat (eds.), Berlin, Springer-Verlag, **2143**: 156–167.

Olmsted, S. O. (1984), "On representing and solving decision problems," PhD dissertation, Dept. of Engineering-Economic Systems, Stanford University.

Poland III, W. B. (1994), "Decision analysis with continuous and discrete variables: A mixture distribution approach," PhD dissertation, Dept. of Engineering-Economic Systems, Stanford University.

Saichev, A. I. and W. A. Woyczyński (1997), *Distributions in the Physical and Engineering Sciences*, **1**, Birkhäuser, Boston.

Shachter, R. D. (1986), "Evaluating influence diagrams," *Operations Research*, **34**, 871−882.

Shachter, R. D. (1988), "Probabilistic inference and influence diagrams," *Operations Research*, **36**, 589−604.

Shachter, R. D. (1990), "An ordered examination of influence diagrams," *Networks*, **20**, 535−563.

Shachter, R. D. and C. R. Kenley (1989), "Gaussian influence diagrams," *Management Science*, **35**(5), 527−550.

Shenoy, P. P. (2006), "Inference in hybrid Bayesian networks using mixture of Gaussians," in R. Dechter and T. Richardson (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Second Conference* (UAI-06), 428−436, AUAI Press, Corvallis, OR.

Shenoy, P. P. and G. Shafer (1990), "Axioms for Probability and belief-function propagation," in R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence 4*, 169−198, North-Holland, Amsterdam.

Appendix: Properties of Dirac Delta Functions

In this appendix, we describe some basic properties of Dirac delta functions [Dirac 1927, Dirac 1958, Hoskins 1979, Kanwal 1998, Saichev and Woyczynski 1997, Khuri 2004]. We attempt to justify most of the properties. These justifications should not be viewed as formal mathematical proofs, but rather as examples of the use of Dirac delta functions that lead to correct conclusions.

(i) (*Sampling*) If $f(x)$ is any function, $f(x)\,\delta(x) = f(0)\,\delta(x)$. Thus, if $f(x)$ is continuous in the neighborhood of 0, then $\int f(x)\,\delta(x)\,\mathrm{d}x = f(0) \int \delta(x)\,\mathrm{d}x = f(0)$. The range of integration need not be from $-\infty$ to ∞, but can cover any domain containing 0.

(ii) (*Change of Origin*) If $f(x)$ is any function which is continuous in the neighborhood of a, then $\int f(x)\,\delta(x-a)\,\mathrm{d}x = f(a)$.

(iii) $\int \delta(x - h(u, v))\,\delta(y - g(v, w, x))\,\mathrm{d}x = \delta(y - g(v, w, h(u, v)))$. This follows from property (ii) of Dirac delta functions.

(iv) (*Rescaling*) If $g(x)$ has real (non-complex) zeros at $a_1, \ldots, a_n$, and is differentiable at these points, and $g'(a_i) \neq 0$ for $i = 1, \ldots, n$, then $\delta(g(x)) = \Sigma_i\, \delta(x - a_i)/|g'(a_i)|$. In particular, if $g(x)$ has only one real zero at a_0, and $g'(a_0) \neq 0$, then $\delta(g(x)) = \delta(x - a_0)/|g'(a_0)|$.

(v) $\delta(ax) = \delta(x)/|a|$ if $a \neq 0$. $\delta(-x) = \delta(x)$, i.e., δ is symmetric about 0.

(vi) Suppose $Y = g(X)$, where g is invertible and differentiable on Ω_X. Then $\delta(y) = \delta(g(x)) =$ $\delta(x - a_0)\,/\,|g'(a_0)|$, where $a_0 = g^{-1}(0)$. Also, $\delta(y - g(x)) = \delta(g(x) - y) =$
$\delta(x - g^{-1}(y))\,/\,|(\mathrm{d}/\mathrm{d}x)(g(g^{-1}(y)))| = \delta(x - g^{-1}(y))\,/\,|\mathrm{d}y/\mathrm{d}x| = \delta(x - g^{-1}(y))\,|\mathrm{d}x/\mathrm{d}y|$
$= \delta(x - g^{-1}(y))\,|(\mathrm{d}/\mathrm{d}y)(g^{-1}(y))|$.

(vii) Consider the Heaviside function $H(x) = 0$ if $x < 0$, $H(x) = 1$ if $x \geq 0$. Then, $\delta(x)$ can be regarded as the "generalized" derivative of $H(x)$ with respect to x, i.e., $(\mathrm{d}/\mathrm{d}x)H(x) = \delta(x)$. $H(x)$ can be regarded as the limit of certain differentiable functions (such as, e.g., the cumulative distribution functions (CDF) of the Gaussian random variable with mean 0 and variance σ^2 in the limit as $\sigma \to 0$). Then, the generalized derivative of $H(x)$ is the limit of the derivative of these functions.

(viii) Suppose continuous variable X has probability density function (PDF) $f_X(x)$ and $Y = g(X)$. Then Y has PDF $f_Y(y) = \int f_X(x)\ \delta(y - g(x))\ \mathrm{d}x$. The function g does not have to be invertible. To show the validity of this formula, let $F_Y(y)$ denote the cumulative distribution function of Y. Then, $F_Y(y) = P(g(X) \le y) = \int f_X(x)\ H(y - g(x))\ \mathrm{d}x$, where $H(\cdot)$ is the Heaviside function defined in (vii). Then, $f_Y(y) = (\mathrm{d}/\mathrm{d}y)(F_Y(y)) = \int f_X(x)\ (\mathrm{d}/\mathrm{d}y)(H(y - g(x)))\ \mathrm{d}x =$ $\int f_X(x)\ \delta(y - g(x))\ \mathrm{d}x$.

(ix) Suppose continuous variable X has pdf $f_X(x)$ and $Y = g(X)$, where g is invertible and differentiable on Ω_X. Then the pdf of Y is $f_Y(y) = \int f_X(x)\ \delta(y - g(x))\ \mathrm{d}x =$ $|(\mathrm{d}/\mathrm{d}y)(g^{-1}(y))| \int f_X(x)\ \delta(x - g^{-1}(y))\ \mathrm{d}x = |(\mathrm{d}/\mathrm{d}y)(g^{-1}(y))|\ f_X(g^{-1}(y))$. Also, $f_X(x)\ \delta(y - g(x))\ /\ (|(\mathrm{d}/\mathrm{d}y)(g^{-1}(y))|\ f_X(g^{-1}(y))) = \delta(x - g^{-1}(y))$. This is because if we consider the left-hand side as a function of x, say $\phi(x)$, it is equal to 0 if $x \ne g^{-1}(y)$, and $\int \phi(x)\ \mathrm{d}x = 1$. Therefore, by definition, $\phi(x) = \delta(x - g^{-1}(y))$. Finally, $f_X(x)\ \delta(y - g(x)) =$ $(|(\mathrm{d}/\mathrm{d}y)(g^{-1}(y))|\ f_X(g^{-1}(y)))\ \delta(x - g^{-1}(y))$.

(x) The definition of δ can be extended to $\mathbb{R}^n$, the n-dimensional Euclidean space. Thus, if $\boldsymbol{x} \in \mathbb{R}^n$, $\delta(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \ne \boldsymbol{0}$, and $\int \ldots \int \delta(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x} = 1$, where $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1 \ldots \mathrm{d}x_n$. Thus, e.g., $\int \ldots \int f(\boldsymbol{x})\ \delta(\boldsymbol{x} - \boldsymbol{x}_0)\ \mathrm{d}\boldsymbol{x} = f(\boldsymbol{x}_0)$.

(xi) Suppose $X_1, \ldots, X_n$ are continuous variables with joint PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then, the deterministic variable $Y = g(X_1, \ldots, X_n)$ has PDF $f_Y(y) = \int \ldots \int f_{\boldsymbol{X}}(\boldsymbol{x})\ \delta(y - g(\boldsymbol{x}))\ \mathrm{d}\boldsymbol{x}$. The function g does not have to be invertible.

(xii) Suppose $X_1, \ldots, X_n$ are continuous variables with joint PDF $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then the joint PDF of deterministic variables $Y = g(X_1, \ldots, X_n)$ and $Z = h(X_1, \ldots, X_n)$ is given by $f_{Y,Z}(y, z) = \int \ldots \int f_{\boldsymbol{X}}(\boldsymbol{x})\ \delta(y - g(\boldsymbol{x}))\ \delta(z - h(\boldsymbol{x}))\ \mathrm{d}\boldsymbol{x}$. The functions g and h do not have to be invertible.