Inference in Hybrid Bayesian Networks with Deterministic Variables

Prakash P. Shenoy and James C. West

University of Kansas School of Business, 1300 Sunnyside Ave., Summerfield Hall, Lawrence, KS 66045-7585 USA
{pshenoy,cully}@ku.edu

Abstract. The main goal of this paper is to describe an architecture for solving large general hybrid Bayesian networks (BNs) with deterministic variables. In the presence of deterministic variables, we have to deal with non-existence of joint densities. We represent deterministic conditional distributions using Dirac delta functions. Using the properties of Dirac delta functions, we can deal with a large class of deterministic functions. The architecture we develop is an extension of the Shenoy-Shafer architecture for discrete BNs. We illustrate the architecture with some small illustrative examples.

Keywords: Hybrid Bayesian networks, deterministic variables, Dirac delta functions, Shenoy-Shafer architecture.

1 Introduction

Bayesian networks (BNs) and influence diagrams (IDs) were invented in the mid 1980s (see e.g., [19, 9]) to represent and reason with large multivariate discrete probability models and decision problems, respectively. Several efficient algorithms exist to compute exact marginals of posterior distributions for discrete BNs (see e.g., [13, 27]), and to solve discrete influence diagrams exactly (see e.g., [18, 22, 24]).

The state of the art exact algorithm for mixtures of Gaussians hybrid BNs is the Lauritzen-Jensen [14] algorithm. This requires the conditional distributions of continuous variables to be conditional linear Gaussians (CLG), and that discrete variables do not have continuous parents.

If a BN has discrete variables with continuous parents, Murphy [17] uses a variational approach to approximate the product of the potentials associated with a discrete variable and its parents with a CLG. Lerner [15] uses a numerical integration technique called Gaussian quadrature to approximate non-CLG distributions with CLG, and this same technique can be used to approximate the product of potentials associated with a discrete variable and its continuous parents. Murphy's and Lerner's approach is then embedded in the Lauritzen-Jensen [14] algorithm to solve the resulting mixtures of Gaussians BN. Shenoy [26] proposes approximating non-CLG distributions by mixtures of Gaussians using a nonlinear optimization technique, and using arc reversals to ensure discrete variables do not have continuous parents. The resulting mixture of Gaussians BN is then solved using Lauritzen-Jensen [14] algorithm.

C. Sossai and G. Chemello (Eds.): ECSQARU 2009, LNAI 5590, pp. 46–58, 2009.

Moral *et al.* [16] proposes approximating probability density functions (PDFs) by mixtures of truncated exponentials (MTE), which are easy to integrate in closed form. Since the family of MTE is closed under combination and marginalization, the Shenoy-Shafer [27] architecture can be used to solve the MTE BN. Cobb *et al.* [5] proposes using a non-linear optimization technique for finding MTE approximation for the many commonly used PDFs. Cobb and Shenoy [2, 3] extend this approach to BNs with linear and non-linear deterministic variables. In the latter case, they approximate non-linear deterministic functions by piecewise linear ones.

Shenoy and West [29] propose mixtures of polynomials (MOP) to approximate PDFs. Like MTE, MOP are easy to integrate, and are closed under combination and marginalization. Unlike MTE, they can be easily found using the Taylor series expansion of differentiable functions, and they are closed under a larger family of deterministic functions than MTE, which are closed only for linear functions.

For Bayesian decision problems, Kenley [11] (see also [23]) describes the representation and solution of Gaussian IDs that include continuous chance variables with CLG distributions. Poland [20] extends Gaussian IDs to mixtures of Gaussians IDs. Thus, continuous chance variables can have any distributions, and these are approximated by mixtures of Gaussians. Cobb and Shenoy [4] extend MTE BNs to MTE IDs for the special case where all decision variables are discrete.

In this paper, we describe a generalization of the Shenoy-Shafer architecture for discrete BNs so that it applies to hybrid BNs with deterministic variables. The functions associated with deterministic variables do not have to be linear (as in the CLG case) or even invertible. We use Dirac delta functions to represent such functions and also for observations of continuous variables. We use mixed potentials to keep track of the nature of potentials (discrete and continuous). We define combination and marginalization of mixed potentials. Finally, we illustrate our architecture by solving some small examples that include non-linear, non-invertible deterministic variables.

An outline of the remainder of the paper is as follows. In Section 2, we describe our architecture for making inferences in hybrid BNs with deterministic variables. This is the main contribution of this paper. In Section 3, we solve three small examples to illustrate the architecture. In Section 4, we end with a summary and discussion.

2 The Extended Shenoy-Shafer Architecture

In this section, we describe the extended Shenoy-Shafer architecture for representing and solving hybrid BNs with deterministic variables. The architecture and notation is adapted from Cinicioglu and Shenoy [1], and Cobb and Shenoy [2].

Variables and States. We are concerned with a finite set V of *variables*. Each variable $X \in V$ is associated with a set Ω_X of its possible *states*. If Ω_X is a finite set or countably infinite, we say X is *discrete*, otherwise X is *continuous*. We will assume that the state space of continuous variables is the set of real numbers (or some subset of it), and that the state space of discrete variables is a set of symbols (not necessarily real numbers). If $r \subseteq V$, $r \neq \emptyset$, then $\Omega_r = \times\{\Omega_X \mid X \in r\}$. If $r = \emptyset$, we will adopt the convention that $\Omega_\emptyset = \{\blacklozenge\}$. If $\boldsymbol{r} \in \Omega_r$, $\boldsymbol{s} \in \Omega_s$, and $r \cap s = \emptyset$, then $(\boldsymbol{r}, \boldsymbol{s}) \in \Omega_{r \cup s}$. Therefore, $(\boldsymbol{r}, \blacklozenge) = \boldsymbol{r}$.

Projection of States. Suppose $\boldsymbol{r} \in \Omega_r$, and suppose $s \subseteq r$. Then the *projection* of $\boldsymbol{r}$ to s, denoted by $\boldsymbol{r}^{\downarrow s}$, is the state of s obtained from $\boldsymbol{r}$ by dropping states of $r \setminus s$. Thus, $(w, x, y, z)^{\downarrow \{W, X\}} = (w, x)$, where $w \in \Omega_W$, and $x \in \Omega_X$. If $s = r$, then $\boldsymbol{r}^{\downarrow s} = \boldsymbol{r}$. If $s = \emptyset$, then $\boldsymbol{r}^{\downarrow s} = \blacklozenge$.

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 zeros. For simplicity, henceforth, we will refer to continuous variables with non-deterministic conditionals as *continuous*, and continuous variables with deterministic conditionals as *deterministic*. In a BN, discrete variables are denoted by rectangular shaped nodes, continuous variables by oval shaped nodes, and deterministic variables by oval nodes with a double border.

Discrete Potentials. In a BN, the conditional probability functions associated with the variable are represented by functions called *potentials*. If A is discrete, it is associated with conditional probability mass functions, one for each state of its parents. The conditional probability mass functions are represented by functions called *discrete potentials*. Suppose $r \subseteq V$ is such that it contains a discrete variable. A discrete potential α for r is a function α: $\Omega_r \to [0, 1]$. The values of discrete potentials are probabilities. We will sometimes write the range of α as $[0, 1](m)$ to denote that the values in $[0, 1]$ are probability masses.

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

Density Potentials. If Z is continuous, then it is associated with a *density* potential. Suppose $r \subseteq V$ is such that it contains a continuous variable. A density potential ζ for r is a function ζ: $\Omega_r \to \mathbb{R}^+$, where $\mathbb{R}^+$ is the set of non-negative real numbers. The values of density potentials are probability densities. We will sometimes write the range of ζ as $\mathbb{R}^+(d)$ to denote that the values in $\mathbb{R}^+$ are densities.

Dirac Delta Functions. δ: $\mathbb{R} \to \mathbb{R}^+(d)$ 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., $\delta(0)$ is 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 [6].

As defined above, the value $\delta(0)$ is undefined, i.e., $\delta(0) = \infty$, when considered as density. We argue that we can *interpret* the value $\delta(0)$ as probability 1. Consider the normal PDF with mean 0 and variance σ^2. Its moment generating function (MGF) is $M(t) = e^{\sigma^2 t}$. In the limit as $\sigma \to 0$, $M(t) = 1$. Now, $M(t) = 1$ is the MGF of the distribution $X = 0$

with probability 1. Therefore, we can interpret the value $\delta(0)$ as probability 1. This is strictly for interpretation only.

Some basic properties of the Dirac delta function are as follows [6, 7, 8, 10, 21, 12].

(i) If $f(x)$ is any function that is continuous in the neighborhood of a, then $f(x)\,\delta(x-a) = f(a)\,\delta(x-a)$, and $\int f(x)\,\delta(x-a)\,dx = f(a)$.

(ii) $\int \delta(x-h(u,v))\,\delta(y-g(v,w,x))\,dx = \delta(y-g(v,w,h(u,v)))$. This follows from (i).

(iii) 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 $\delta(g(x)) = \Sigma_i\,\delta(x - a_i)/|g'(a_i)|$. For example, $\delta(ax) = \delta(x)/|a|$, if $a \neq 0$. Therefore, $\delta(-x) = \delta(x)$.

(iv) Suppose continuous variable X has PDF $f_X(x)$ and $Y = g(X)$. Then Y has PDF $f_Y(y) = \int f_X(x)\,\delta(y-g(x))\,dx$. The function g does not have to be invertible.

A more extensive list of properties of the Dirac delta function that is relevant for uncertain reasoning is stated in [1].

Dirac Potentials. Deterministic variables have conditional distributions containing functions. We will represent such functions by *Dirac* potentials. Suppose $x = r \cup s$ 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 ξ: $\Omega_x \rightarrow \mathbb{R}^+(d)$ such that $\xi(\boldsymbol{r}, \boldsymbol{s})$ is of the form $\Sigma\{p_{\boldsymbol{r},i}\,\delta(z - g_{\boldsymbol{r},i}(\boldsymbol{s}^{\downarrow(s\setminus\{Z\})})) \mid i = 1, \ldots, n$, and $\boldsymbol{r} \in \Omega_r\}$, where $\boldsymbol{s} \in \Omega_s$, $Z \in s$ is a continuous or deterministic variable, $z \in \Omega_Z$, $\delta(z - g_{\boldsymbol{r},i}(\boldsymbol{s}^{\downarrow(s\setminus\{Z\})}))$ are Dirac delta functions and $p_{\boldsymbol{r},i}$ are probabilities for all $i = 1, \ldots, n$, and $\boldsymbol{r} \in \Omega_r$, and n is a positive integer. Here, we are assuming that Z is a weighted sum of functions $g_{\boldsymbol{r},i}(\boldsymbol{s}^{\downarrow(s\setminus\{Z\})})$ of the other continuous variables in s, weighted by $p_{\boldsymbol{r},i}$, and that the nature of the functions and weights may depend on $\boldsymbol{r} \in \Omega_r$, and/or on some latent index i.

Suppose 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 $\xi(z, x) = \delta(x - z^2)$ for $\{Z, X\}$. In this case, $n = 1$, and $r = \emptyset$.

A more general example of a Dirac potential for $\{Z, X\}$ is $\xi(z, x) = (\frac{1}{2})\,\delta(x - z) + (\frac{1}{2})\,\delta(x - 1)$. Here, X is a continuous variable with continuous parent Z. As argued before, we can interpret the value $\xi(x, x)$ as $\frac{1}{2}(m)$, and the value $\xi(1, x)$ as $\frac{1}{2}(m)$. All other values are equal to zero. 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 zeros.

Continuous Potentials. Both density and Dirac potentials are special instances of a broader class of potentials called *continuous* potentials. Suppose $x \subseteq V$ is such that it contains a continuous or deterministic variable. Then a *continuous potential* ξ for x is a function ξ: $\Omega_x \rightarrow \mathbb{R}^+(d)$. 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 a continuous potential ξ for $\{X\}$ as follows: $\xi(x) = 0.5\,\delta(x-1) + 0.5\,f(x)$. Notice that $\int \xi(x)\,dx = 0.5 \int \delta(x-1)\,dx + 0.5\int f(x)\,dx = 0.5 + 0.5 = 1$.

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

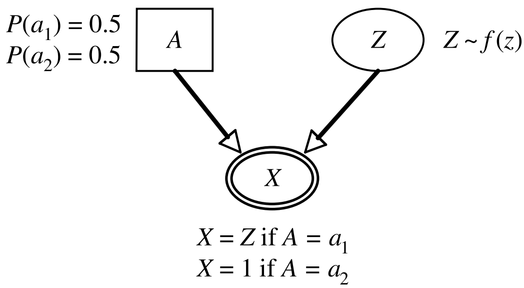

Fig. 1. A hybrid BN with a discrete, a continuous, and a deterministic variable

Mixed Potentials. In reasoning with hybrid models, we need to define mixed potentials. A mixed potential has two parts, the first part is a discrete potential and the second part is a continuous potential. Formally, suppose α is a discrete potential for r. Then a mixed potential representation of α is $\mu_1 = (\alpha, \iota)$ for r, where ι denotes the *identity* potential for the empty set, $\iota(\blacklozenge) = 1$. Suppose ζ is a continuous potential for s. Then, a mixed potential representation of ζ is $\mu_2 = (\iota, \zeta)$ for s. Mixed potentials can have non-vacuous discrete and continuous parts. Thus $\mu_3 = (\alpha, \zeta)$ is a mixed potential for $r \cup s$. Such a mixed potential would be the result of combining μ_1 and μ_2, which we will define next. The main idea behind mixed potentials is to represent the nature (discrete or continuous) of potentials.

Combination of Potentials. Suppose α is a discrete or continuous potential for some subset a of variables and β is a discrete or continuous potential 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 pointwise multiplication, i.e.,

$$(\alpha \otimes \beta)(\boldsymbol{x}) = \alpha(\boldsymbol{x}^{\downarrow a})\, \beta(\boldsymbol{x}^{\downarrow b}) \text{ for all } \boldsymbol{x} \in \Omega_{a \cup b}. \tag{2.1}$$

If α and β are both discrete potentials, then $\alpha \otimes \beta$ is a discrete potential, and if α and β are both continuous potentials, then $\alpha \otimes \beta$ is a continuous potential. The definition of combination in (2.1) is valid also if α is discrete and β is continuous and vice-versa, and will be used when we define marginalization of mixed potentials. However, the nature of the potential $\alpha \otimes \beta$ when α is discrete and β is continuous (or vice-versa) will not arise in the combination operation since we will used mixed potentials to represent the potentials, and as we will see, combination of mixed potentials avoids such combinations.

The identity potential ι_r for r has the property that given any potential ξ for $s \supseteq r$, $\xi \otimes \iota_r = \xi$. If $r = \emptyset$, then we will let ι denote $\iota_\emptyset$.

Combination of Mixed Potentials. Suppose $\mu_1 = (\alpha_1, \zeta_1)$, and $\mu_2 = (\alpha_2, \zeta_2)$ are two mixed potentials with discrete parts α_1 for r_1 and α_2 for r_2, respectively, and continuous parts ζ_1 for s_1 and ζ_2 for s_2, respectively. Then, the combination $\mu_1 \otimes \mu_2$ is a mixed potential for $r_1 \cup s_1 \cup r_2 \cup s_2$ given by

$$\mu_1 \otimes \mu_2 = (\alpha_1 \otimes \alpha_2, \zeta_1 \otimes \zeta_2). \tag{2.2}$$

Since $\alpha_1 \otimes \alpha_2$ is a discrete potential and $\zeta_1 \otimes \zeta_2$ is a continuous potential, the definition of combination of mixed potentials in (2.2) is consistent with the definition of mixed potentials.

If $\mu_1 = (\alpha, \iota)$ represents the discrete potential α for r, and $\mu_2 = (\iota, \zeta)$ represents the continuous potential for s, then $\mu_1 \otimes \mu_2 = (\alpha, \zeta)$ is a mixed potential for $r \cup s$.

Since combination is pointwise multiplication, and multiplication is commutative, combination of potentials (discrete or continuous) is commutative ($\alpha \otimes \beta = \beta \otimes \alpha$) and associative ($(\alpha \otimes \beta) \otimes \gamma = \alpha \otimes (\beta \otimes \gamma)$). Since the combination of mixed potentials is defined in terms of combination of discrete and continuous potentials, each of which is commutative and associative, combination of mixed potentials is also commutative and associative.

Marginalization of Potentials. The definition of marginalization depends on whether the variable being marginalized is discrete or continuous. We marginalize discrete variables by addition, and continuous variables by integration. Integration of potentials containing Dirac delta functions is done using the properties of Dirac delta functions. Also, after marginalization, the nature of a potential could change, e.g., from continuous to discrete (if the domain of the marginalized potential contains only discrete variables) and from discrete to continuous (if the domain of the marginalized potential contains only continuous variables). We will make this more precise when we define marginalization of mixed potentials.

Suppose α is a discrete or continuous potential for a, and suppose X is a discrete variable in a. Then the *marginal* of α by deleting X, denoted by α^{-X}, is the potential for $a \setminus \{X\}$ obtained from α by addition over the states of X, i.e.,

$$\alpha^{-X}(\boldsymbol{y}) = \Sigma\{\alpha(x, \boldsymbol{y}) \mid x \in \Omega_X\} \text{ for all } \boldsymbol{y} \in \Omega_{a \setminus \{X\}}. \tag{2.3}$$

If X is a continuous variable in a, then the marginal of α by deleting X is obtained by integration over the state space of X, i.e.,

$$\alpha^{-X}(\boldsymbol{y}) = \int \alpha(x, \boldsymbol{y})\, dx \text{ for all } \boldsymbol{y} \in \Omega_{a \setminus \{X\}}. \tag{2.4}$$

If α contains Dirac delta functions, then we have to use the properties of Dirac delta functions in doing the integration.

If ξ is a discrete or continuous potential for $\{X\} \cup pa(X)$ representing the conditional distribution for X in a BN, then ξ^{-X} is an identity potential for $pa(A)$.

If we marginalize a discrete or continuous potential by deleting two (or more) variables from its domain, then the order in which the variables are deleted does not matter, i.e., $(\alpha^{-A})^{-B} = (\alpha^{-B})^{-A} = \alpha^{-\{A, B\}}$.

If α is a discrete or continuous potential for *a,* β is a discrete or continuous potential for b, $A \in a$, and $A \notin b$, then $(\alpha \otimes \beta)^{-A} = (\alpha^{-A}) \otimes \beta$. This is a key property of combination and marginalization that allows local computation [Shenoy and Shafer 1990]. We call this property *local computation.*

Marginalization of Mixed Potentials. Mixed potentials allow us to represent the nature of potentials, and marginalization of mixed potentials allows us to represent the nature of the marginal. Suppose $\mu = (\alpha, \zeta)$ is a mixed potential for $r \cup s$ with discrete part α for r, and continuous part ζ for s. Let C denote the set of continuous variables, and let D denote the set of discrete variables. The marginal of μ by deleting $X \in r \cup s$, denoted by μ^{-X}, is defined as follows.

$$\mu^{-X} = \begin{cases} (\alpha^{-X}, \zeta), & \text{if } X \in r, X \notin s, \text{ and } r \setminus \{X\} \not\subseteq C, \quad (2.5) \\ (\iota, \alpha^{-X} \otimes \zeta), & \text{if } X \in r, X \notin s, \text{ and } r \setminus \{X\} \subseteq C, \quad (2.6) \\ (\alpha, \zeta^{-X}), & \text{if } X \notin r, X \in s, \text{ and } s \setminus \{X\} \not\subseteq D, \quad (2.7) \\ (\alpha \otimes \zeta^{-X}, \iota), & \text{if } X \notin r, X \in s, \text{ and } s \setminus \{X\} \subseteq D, \quad (2.8) \\ ((\alpha \otimes \zeta)^{-X}, \iota), & \text{if } X \in r, X \in s, \text{ and } (r \cup s) \setminus \{X\} \subseteq D, \text{ and} \quad (2.9) \\ (\iota, (\alpha \otimes \zeta)^{-X}), & \text{if } X \in r, X \in s, \text{ and } (r \cup s) \setminus \{X\} \not\subseteq D. \quad (2.10) \end{cases}$$

Some comments about the definition of marginalization of mixed potentials are as follows. First, if the variable being deleted belongs only to one part (discrete or continuous, as in cases (2.5)–(2.8)), then the local computation property allow us to delete the variable from that part only leaving the other part unchanged. If the variable being deleted belongs to both parts (as in cases (2.9)–(2.10)), then we first need to combine the two parts before deleting the variable. Second, when we have only continuous variables left in a discrete potential after marginalization, we move the potential to the continuous part (2.6), and when we only have discrete variables left, we move the potential to the discrete part (2.8), otherwise we don’t change the nature of the marginalized potentials ((2.5) and (2.7)). In cases (2.9)–(2.10), when we have to combine the discrete and continuous potentials before marginalizing X, if only discrete variables are left, then we have to classify it as a discrete potential (2.9), and if we have only continuous variables left, then we have to classify it as a continuous potential (2.10). However, if it has discrete and continuous variables, it could be classified as either discrete or continuous, and the definition above has chosen to classify it as continuous (2.10). It makes no difference one way or the other.

Division of Potentials. The Shenoy-Shafer [27] architecture requires only the combination and marginalization operations. However, at the end of the propagation, we need to normalize the potentials, and this involves division. Divisions are also involved in doing arc reversals [1].

Suppose ρ is a discrete or continuous potential for r, and suppose $X \in r$. Then the *division* of ρ by ρ^{-X}, denoted by $\rho \oslash (\rho^{-X})$, is the potential for r obtained by pointwise division of ρ by ρ^{-X}, i.e.,

$$(\rho \oslash (\rho^{-X}))(x, \boldsymbol{y}) = \rho(x, \boldsymbol{y})/\rho^{-X}(\boldsymbol{y}) \tag{2.11}$$

for all $\boldsymbol{y} \in \Omega_{r \setminus \{X\}}$ and $x \in \Omega_X$. Notice that if $\rho^{-X}(\boldsymbol{y}) = 0$, then $\rho(x, \boldsymbol{y}) = 0$. In this case, we will simply define 0/0 as 0. Also, notice that $(\rho \oslash (\rho^{-X})) \otimes \rho^{-X} = \rho$.

Suppose ξ is a discrete or continuous potential for $\{X\}$ representing the unnormalized posterior marginal for X. To normalize ξ, we divide ξ by ξ^{-X}. Thus the normalized posterior marginal for X is $\xi \oslash (\xi^{-X})$. The value $\xi^{-X}(\blacklozenge)$ represents the probability of the evidence, and is the same regardless of variable X for which we are computing the marginal.

3 Some Illustrative Examples

In this section, we illustrate the extended Shenoy-Shafer architecture using several small examples. More examples can be found in [28].

Example 1 (*Transformation of variables*)**.** Consider a BN with continuous variable Y and deterministic variable Z as shown in Fig. 2. Notice that the function defining the deterministic variable is not invertible.

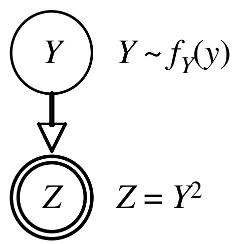

Fig. 2. A BN with a deterministic variable with a non-invertible, non-linear function

Let ψ and ζ_1 denote the mixed potentials for $\{Y\}$ and $\{Y, Z\}$, respectively. Then,

$$\psi(y) = (1, f_Y(y)) \tag{4.1}$$

$$\zeta_1(y, z) = (1, \delta(z - y^2)) \tag{4.2}$$

To find the prior marginal distribution of Z, first we combine ψ and ζ_1, and then we marginalize Y from the combination.

$$(\psi \otimes \zeta_1)(y, z) = (1, f_Y(y)\ \delta(z - y^2)) \tag{4.3}$$

$$((\psi \otimes \zeta_1)^{-Y})(z) = (1, \int f_Y(y)\ \delta(z - y^2)\ dy) = (1, (1/(2\sqrt{z}))(f_Y(\sqrt{z}) + f_Y(-\sqrt{z})))$$
$$= (1, f_Z(z)) \text{ for } z > 0, \text{ where } f_Z(z) = (1/(2\sqrt{z}))(f_Y(\sqrt{z}) + f_Y(-\sqrt{z})) \tag{4.4}$$

The result in (4.4) follows from (2.7) and properties (iii) and (iv) of Dirac delta functions. Now suppose we observe $Z = c$, where c is a constant such that $f_Z(c) > 0$, i.e., $c > 0$ and $f_Y(\sqrt{c}) > 0$ or $f_Y(-\sqrt{c}) > 0$ or both. This observation is represented by the mixed potential for Z, $\zeta_2(z) = (1, \delta(z - c))$. Then, the un-normalized posterior marginal distribution of Y is computed as follows:

$$((\zeta_1 \otimes \zeta_2)^{-Z})(y) = (1, \int \delta(z - y^2)\, \delta(z - c))\, dz) = (1, \delta(c - y^2)) = (1, \delta(y^2 - c)) \quad (4.5)$$

$$\begin{aligned}(\psi \otimes (\zeta_1 \otimes \zeta_2)^{-Z})(y) &= (1, f_Y(y)\, \delta(y^2 - c)) \\ &= (1, f_Y(y)\, (\delta(y - \sqrt{c}) + \delta(y + \sqrt{c}))/(2\sqrt{c})) \\ &= (1, (f_Y(\sqrt{c})\, \delta(y - \sqrt{c}) + f_Y(-\sqrt{c})\, \delta(y + \sqrt{c}))/(2\sqrt{c})) \quad (4.6)\end{aligned}$$

The normalization constant is $(f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))/(2\sqrt{c})$. Therefore the normalized posterior distribution of Y is $(f_Y(\sqrt{c})\delta(y - \sqrt{c}) + f_Y(-\sqrt{c})\, \delta(y + \sqrt{c}))\, /\, (f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))$, i.e., $Y = \sqrt{c}$ with probability $f_Y(\sqrt{c})/(f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))$, and $Y = -\sqrt{c}$ with probability $f_Y(-\sqrt{c})/(f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))$.

Example 2 (*Mixed distributions*)**.** Consider the hybrid BN shown in Fig. 1 with three variables. A is discrete with state space $\Omega_Y = \{a_1, a_2\}$, Z is continuous, and X is deterministic. What is the prior marginal distribution of X? Suppose we observe $X = 1$. What is the posterior marginal distribution of A?

Let α, ζ, and ξ_1 denote the mixed potentials for $\{A\}$, $\{Z\}$, and $\{A, Z, X\}$, respectively. Then:

$$\alpha(a_1) = (0.5, 1), \alpha(a_2) = (0.5, 1); \quad (4.7)$$

$$\zeta(z) = (1, f_Z(z)); \quad (4.8)$$

$$\xi_1(a_1, z, x) = (1, \delta(x - z)), \xi_1(a_2, z, x) = (1, \delta(x - 1)). \quad (4.9)$$

The prior marginal distribution of X is given by $(\alpha \otimes \zeta \otimes \xi_1)^{-\{A, Z\}} = ((\alpha \otimes \xi_1)^{-A} \otimes \zeta)^{-Z}$.

$$(\alpha \otimes \xi_1)(a_1, z, x) = (0.5, \delta(x - z)), (\alpha \otimes \xi_1)(a_2, z, x) = (0.5, \delta(x - 1)); \quad (4.10)$$

$$((\alpha \otimes \xi_1)^{-A})(z, x) = (1, 0.5\, \delta(x - z) + 0.5\, \delta(x - 1)); \quad (4.11)$$

$$((\alpha \otimes \xi_1)^{-A} \otimes \zeta)(z, x) = (1, 0.5\, \delta(x - z)\, f_Z(z) + 0.5\, \delta(x - 1)\, f_Z(z)); \quad (4.12)$$

$$\begin{aligned}(((\alpha \otimes \xi_1)^{-A} \otimes \zeta)^{-Z})(x) &= (1, \int 0.5\, \delta(x - z)\, f_Z(z)\, dz + 0.5\, \delta(x - 1) \int f(z)\, dz) \\ &= (1, 0.5\, f_Z(x) + 0.5\, \delta(x - 1)). \quad (4.13)\end{aligned}$$

Thus the prior marginal distribution of X is mixed with PDF $0.5\, f_Z(x)$ and a mass of 0.5 at $X = 1$. (4.11) results from use of (2.9) since Y is in the domain of discrete and continuous parts. (4.13) follows from (2.7).

Let ξ_2 denote the observation $X = 1$. Thus, $\xi_2(x) = (1, \delta(x - 1))$. The (unnormalized) posterior marginal of A is given by $(\alpha \otimes \zeta \otimes \xi_1 \otimes \xi_2)^{-\{Z, X\}} = \alpha \otimes (\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z}$.

$$(\xi_1 \otimes \xi_2)(a_1, z, x) = (1, \delta(x - z)\, \delta(x - 1)),$$

$$(\xi_1 \otimes \xi_2)(a_2, z, x) = (1, \delta(x - 1)\, \delta(x - 1)) = (1, \delta(x - 1)); \tag{4.14}$$

$$(\xi_1 \otimes \xi_2)^{-X}(a_1, z) = (1, \int \delta(x - z)\, \delta(x - 1)\, \mathrm{d}x) = (1, \delta(1 - z)) = (1, \delta(z - 1)),$$

$$(\xi_1 \otimes \xi_2)^{-X}(a_2, z) = (1, \int \delta(x - 1)\, \mathrm{d}x) = (1, 1); \tag{4.15}$$

$$(\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})(a_1, z) = (1, \delta(z - 1)\, f_Z(z)),$$

$$((\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})(a_2, z) = (1, f_Z(z)); \tag{4.16}$$

$$(\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z}(a_1) = (\int \delta(z - 1)\, f_Z(z)\, \mathrm{d}z, 1) = (f_Z(1), 1),$$

$$((\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z}(a_2) = (\int f(z)\, \mathrm{d}z, 1) = (1, 1); \tag{4.17}$$

$$(\alpha \otimes (\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z})(a_1) = (0.5\, f_Z(1), 1),$$

$$(\alpha \otimes (\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z})(a_2) = (0.5, 1). \tag{4.18}$$

Notice that the un-normalized posterior marginal for A is in units of density for $A = a_1$, and in units of probability for $A = a_2$. Thus, after normalization, the posterior probability of a_1 is 0, and the posterior probability of a_2 is 1.

Example 3 (*Discrete variable with a continuous parent*)**.** Consider the hybrid BN consisting of a continuous variable Y, a discrete variable A, and a deterministic variable X as shown in Fig. 3. A is an indicator variable with states $\{a_1, a_2\}$ such that $A = a_1$ if $0 < Y \leq 0.5$, and $A = a_2$ if $0.5 < Y < 1$. What is the prior marginal distribution of X? If we observe $X = 0.25$, what is the posterior marginal distribution of Y?

Let ψ, α, and ξ_1 denote the mixed potentials for $\{Y\}$, $\{Y, A\}$, and $\{Y, A, X\}$, respectively. The *Heaviside* function $H(\cdot)$ is: $H(y) = 0$ if $y < 0$, and $= 1$ if $y > 0$.

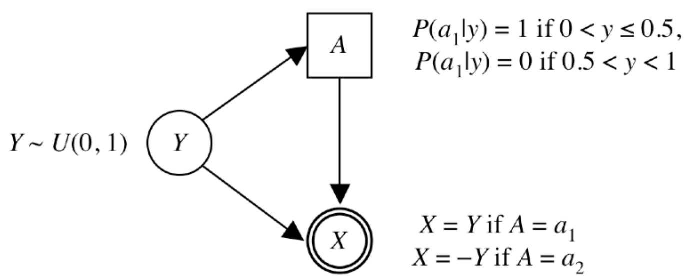

Fig. 3. A hybrid BN with a discrete variable that has a continuous parent

$$\psi(y) = (1, f_Y(y)), \text{ where } f_Y(y) = 1 \text{ if } 0 < y < 1, = 0 \text{ otherwise}; \tag{4.19}$$

$$\alpha(a_1, y) = (H(y) - H(y - 0.5), 1),$$

$$\alpha(a_2, y) = (H(y - 0.5) - H(y - 1), 1); \tag{4.20}$$

$$\xi_1(a_1, y, x) = (1, \delta(x - y)),$$

$$\xi_1(a_2, y, x) = (1, \delta(x + y)). \tag{4.21}$$

To find the marginal distribution of X, first we combine α and ξ_1 and marginalize A from the combination, next we combine the result with ψ and marginalize Y from the combination.

$$(\alpha\otimes\xi_1)(a_1, y, x) = (H(y) - H(y - 0.5), \delta(x - y)),$$

$$(\alpha\otimes\xi_1)(a_2, y, x) = (H(y - 0.5) - H(y - 1), \delta(x + y)); \tag{4.22}$$

$$(\alpha\otimes\xi_1)^{-A}(y, x) = (1, (H(y) - H(y - 0.5))\, \delta(x - y) + (H(y - 0.5) - H(y - 1))\, \delta(x + y)); \tag{4.23}$$

$$(((\alpha\otimes\xi_1)^{-A})\otimes\psi)(y, x) = (1, f_Y(y)[((H(y) - H(y - 0.5))\, \delta(x - y) + (H(y - 0.5) - H(y - 1))\, \delta(x + y))]); \tag{4.24}$$

$$\begin{aligned}(((\alpha\otimes\xi_1)^{-A})\otimes\psi)^{-Y}(x) &= (1, f_Y(x)((H(x) - H(x - 0.5)) + f_Y(-x)(H(-x - 0.5) - H(-x - 1))\\ &= (1, H(x) - H(x - 0.5) + H(-x - 0.5) - H(-x - 1)).\end{aligned} \tag{4.25}$$

Thus, the prior marginal distribution of X is uniform in the interval $(-1, -0.5)\cup(0, 0.5)$. Let ξ_2 be the mixed potential denoting the observation that $X = 0.25$. Thus, $\xi_2(x) = (1, \delta(x - 0.25))$. The (unnormalized) posterior marginal of Y is given by $(\xi_1\otimes(\xi_2\otimes\alpha))^{-\{A, X\}})\otimes\psi$.

$$(\xi_2\otimes\alpha)(a_1, y, x) = (H(y) - H(y - 0.5), \delta(x - 0.25)),$$

$$(\xi_2\otimes\alpha)(a_2, y, x) = (H(y - 0.5) - H(y - 1), \delta(x - 0.25)); \tag{4.26}$$

$$(\xi_1\otimes(\xi_2\otimes\alpha))(a_1, y, x) = (H(y) - H(y - 0.5), \delta(x - 0.25)\, \delta(x - y)),$$

$$(\xi_1\otimes(\xi_2\otimes\alpha))(a_2, y, x) = (H(y - 0.5) - H(y - 1), \delta(x - 0.25)\, \delta(x + y)); \tag{4.27}$$

$$\begin{aligned}((\xi_1\otimes(\xi_2\otimes\alpha))^{-\{A, X\}})(y) &= (1, [(H(y) - H(y - 0.5)) \textstyle\int \delta(x - 0.25)\, \delta(x - y)\, dx]\\ &\quad + [(\mathrm{H}(y - 0.5) - \mathrm{H}(y - 1)) \textstyle\int \delta(x - 0.25)\, \delta(x + y)\, dx]\\ &= (1, (H(y) - H(y - 0.5))\, \delta(y - 0.25));\end{aligned} \tag{4.28}$$

$$\begin{aligned}(((\xi_1\otimes(\xi_2\otimes\alpha))^{-\{A, X\}})\otimes\psi)(y) &= (1, f_Y(y)([(\mathrm{H}(y) - \mathrm{H}(y - 0.5))\, \delta(y - 0.25)]))\\ &= (1, \delta(y - 0.25)).\end{aligned} \tag{4.29}$$

The posterior marginal for Y is $Y = 0.25$ with probability 1.

4 Summary and Discussion

We have described a generalization of the Shenoy-Shafer architecture for discrete BNs so it applies to hybrid BNs with deterministic variables. We use Dirac delta functions to represent conditionals of deterministic variables, and observations of continuous variables. We use mixed potentials to keep track of the discrete and continuous nature of potentials. Marginalization of discrete variables is using addition and marginalization of continuous variables is by integration. We define marginalization of mixed potentials to keep track of the nature of marginalized potentials.

We have ignored the computational problem of integration of density potentials. In some cases, e.g., Gaussian density functions, there does not exist a closed form solution of the integral of the Gaussian density. We assume that we can somehow work around such problems by approximating such density functions by mixtures of truncated exponentials [16] or mixtures of polynomials [29]. In any case, this needs further investigation.

Acknowledgments. We are grateful to Barry Cobb for many discussions.

References

1. Cinicioglu, E.N., Shenoy, P.P.: Arc Reversals in Hybrid Bayesian Networks with Deterministic Variables. Int. J. Approx. Reas. 50(5), 763–777 (2009)
2. Cobb, B.R., Shenoy, P.P.: Hybrid Bayesian Networks with Linear Deterministic Variables. In: Bacchus, F., Jaakkola, T. (eds.) UAI 2005, pp. 136–144. AUAI Publishers, Corvallis (2005)
3. Cobb, B.R., Shenoy, P.P.: Nonlinear Deterministic Relationships in Bayesian networks. In: Godo, L. (ed.) ECSQARU 2005. LNCS, vol. 3571, pp. 27–38. Springer, Heidelberg (2005)
4. Cobb, B.R., Shenoy, P.P.: Decision Making with Hybrid Influence Diagrams Using Mixtures of Truncated Exponentials. Eur. J. Oper. Res. 186(1), 261–275 (2008)
5. Cobb, B.R., Shenoy, P.P., Rumi, R.: Approximating Probability Density Functions in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials. Stat. & Comp. 16(3), 293–308 (2006)
6. Dirac, P.A.M.: The Physical Interpretation of the Quantum Dynamics. Proc. Royal Soc. London. Series A 113(765), 621–641 (1927)
7. Dirac, P.A.M.: The Principles of Quantum Mechanics, 4th edn. Oxford Univ. Press, London (1958)
8. Hoskins, R.F.: Generalised Functions. Ellis Horwood, Chichester (1979)
9. Howard, R.A., Matheson, J.E.: Influence Diagrams. In: Howard, R.A., Matheson, J.E. (eds.) The Principles and Applications of Decision Analysis, vol. 2, pp. 719–762. Strategic Decisions Group, Menlo Park (1984)
10. Kanwal, R.P.: Generalized Functions: Theory and Technique, 2nd edn. Birkhäuser, Boston (1998)
11. Kenley, C.R.: Influence Diagram Models with Continuous Variables. PhD thesis, Dept. of Engg.-Econ. Sys. Stanford University, Stanford (1986)
12. Khuri, A.I.: Applications of Dirac's Delta Function in Statistics. Int. J. Math. Edu. in Sci. & Tech. 32(2), 185–195 (2004)
13. Lauritzen, S.L., Spiegelhalter, D.J.: Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems. J. Royal Stat. Soc. B 50(2), 157–224 (1988)
14. Lauritzen, S.L., Jensen, F.: Stable Local Computation with Conditional Gaussian Distributions. Stat. & Comp. 11, 191–203 (2001)
15. Lerner, U.N.: Hybrid Bayesian Networks for Reasoning About Complex Systems. PhD thesis, Dept. of Comp. Sci. Stanford University, Stanford (2002)

16. Moral, S., Rumi, R., Salmeron, A.: Mixtures of Truncated Exponentials in Hybrid Bayesian Networks. In: Benferhat, S., Besnard, P. (eds.) ECSQARU 2001. LNCS (LNAI), vol. 2143, pp. 156–167. Springer, Heidelberg (2001)
17. Murphy, K.: A Variational Approximation for Bayesian Networks with Discrete and Continuous Latent Variables. In: Laskey, K.B., Prade, H. (eds.) UAI 1999, pp. 457–466. Morgan Kaufmann, San Francisco (1999)
18. Olmsted, S.M.: On Representing and Solving Decision Problems. PhD thesis, Dept. of Engg.-Econ. Sys. Stanford University, Stanford (1983)
19. Pearl, J.: Fusion, Propagation and Structuring in Belief Networks. Art. Int. 29, 241–288 (1986)
20. Poland III, W.B.: Decision Analysis with Continuous and Discrete Variables: A Mixture Distribution Approach. PhD thesis, Dept. of Engg.-Econ. Sys., Stanford University, Stanford (1994)
21. Saichev, A.I., Woyczyński, W.A.: Distributions in the Physical and Engineering Sciences, vol. 1. Birkhäuser, Boston (1997)
22. Shachter, R.D.: Evaluating Influence Diagrams. Oper. Res. 34(6), 871–882 (1986)
23. Shachter, R.D., Kenley, C.R.: Gaussian Influence Diagrams. Mgmt. Sci. 35(5), 527–550 (1989)
24. Shenoy, P.P.: Valuation-Based Systems for Bayesian Decision Analysis. Oper. Res. 40(3), 463–484 (1992)
25. Shenoy, P.P.: A New Method for Representing and Solving Bayesian Decision Problems. In: Hand, D.J. (ed.) Artificial Intelligence Frontiers in Statistics: AI and Statistics III, pp. 119–138. Chapman & Hall, London (1993)
26. Shenoy, P.P.: Inference in Hybrid Bayesian Networks Using Mixtures of Gaussians. In: Dechter, R., Richardson, T. (eds.) UAI 2006, pp. 428–436. AUAI Press, Corvallis (2006)
27. Shenoy, P.P., Shafer, G.: Axioms for Probability and Belief-Function Propagation. In: Shachter, R.D., Levitt, T.S., Lemmer, J.F., Kanal, L.N. (eds.) UAI 1988, vol. 4, pp. 169–198. North-Holland, Amsterdam (1990)
28. Shenoy, P.P., West, J.C.: Inference in Hybrid Bayesian Networks with Deterministic Variables. Working Paper No. 318. School of Business, University of Kansas (2009)
29. Shenoy, P.P., West, J.C.: Inference in Hybrid Bayesian Networks with Mixtures of Polynomials. Working Paper No. 321. School of Business, University of Kansas (2009)