

SCHOOL OF BUSINESS WORKING PAPER NO. 318

Extended Shenoy-Shafer Architecture for
Inference in Hybrid Bayesian Networks with Deterministic Conditionals

Prakash P. Shenoy

and

James C. West

June 2009. Revised February 2011[†]

University of Kansas School of Business
1300 Sunnyside Ave., Summerfield Hall
Lawrence, KS 66045-7601 USA
<pshenoy@ku.edu>, <cully@ku.edu>

[†] Accepted for publication in *International Journal of Approximate Reasoning*.

Table of Contents

Abstract	1
1 Introduction	1
2 Dirac Delta Functions.....	3
3 An Architecture for Computing Marginals.....	5
3.1 Variables and States	5
3.2 Projection of States	6
3.3 Discrete Potentials	6
3.4 Continuous Potentials	7
3.5 Combination of Potentials	8
3.6 Marginalization of Potentials	8
3.7 Normalization of Potentials.....	10
3.8 Solving Hybrid Bayesian Networks	11
4 Some Illustrative Examples	11
4.1 Example 1: Mixture distribution	11
4.2 Example 2: Transformation of variables	13
4.3 Example 3: Mixed distributions	14
4.4 Example 4: Discrete Variable with Continuous Parents	14
5 The Extended Crop Problem.....	16
6 Summary and Discussion	20
Acknowledgments	22
Appendix: Properties of the Dirac Delta Function	22
References	23

Extended Shenoy-Shafer Architecture for Inference in Hybrid Bayesian Networks with Deterministic Conditionals

Prakash P. Shenoy and James C. West

University of Kansas School of Business, 1300 Sunnyside Ave,
Summerfield Hall, Lawrence, KS 66045-7601 USA
<pshenoy@ku.edu>, <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 conditionals for continuous variables using local computation. In the presence of deterministic conditionals for continuous variables, we have to deal with the non-existence of the joint density function for the continuous variables. We represent deterministic conditional distributions for continuous variables 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 extend the definitions of potentials to include conditional probability density functions and deterministic conditionals for continuous variables. We keep track of the units of continuous potentials. Inference in hybrid BNs is then done in the same way as in discrete BNs but by using discrete and continuous potentials and the extended definitions of combination and marginalization. We describe several small examples to illustrate our architecture. In addition, we solve exactly an extended version of the crop problem that includes non-conditional linear Gaussian distributions and non-linear deterministic functions.

1 Introduction

Bayesian networks (BNs) and influence diagrams (IDs) were invented in the mid 1980s (see e.g., [Pearl, 1986], [Howard and Matheson 1984]) 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., [Lauritzen and Spiegelhalter 1988], and [Shenoy and Shafer 1990]) and to solve discrete influence diagrams exactly (see e.g., [Olmsted 1983], [Shachter 1986], [Shenoy 1992]).

The state of the art exact algorithm for mixtures of Gaussians hybrid BNs is Lauritzen-Jensen's [2001] algorithm implemented with Madsen's [2008] lazy propagation technique. This requires the conditional distributions of continuous variables to be conditional linear Gaussians,

and that discrete variables do not have continuous parents. Marginals of multivariate normal distributions can be found easily without the need for integration. The disadvantages are that in the inference process, continuous variables have to be marginalized before discrete ones. In some problems, this restriction can lead to large cliques [Lerner and Parr 2001].

If a BN has discrete variables with continuous parents, Murphy [1999] uses a variational approach to approximate the product of the potentials associated with a discrete variable and its parents with a conditional linear Gaussian. Lerner [2002] uses a numerical integration technique called Gaussian quadrature to approximate non-conditional linear Gaussian distributions with conditional linear Gaussians, 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 Lauritzen-Jensen's [2001] algorithm to solve the resulting mixtures of Gaussians BN.

Shenoy [2006] proposes approximating non-conditional linear Gaussian 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's [2001] algorithm.

Moral *et al.* [2001] proposes approximating probability density functions (PDFs) by mixtures of truncated exponentials (MTE), which are easy to integrate in closed form. Since the family of mixtures of truncated exponentials are closed under combination and marginalization, the Shenoy-Shafer [1990] algorithm can be used to solve a MTE BN. Cobb and Shenoy [2006] and 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 BNs with linear and non-linear deterministic variables. In the latter case, they approximate non-linear deterministic functions by piecewise linear ones. Rumi and Salmeron [2007] describe approximate probability propagation with MTE approximations that have only two exponential terms in each piece. Romero *et al.* [2007] describe learning MTE potentials from data, and Langseth *et al.* [2010] investigate the use of MTE approximations where the coefficients are restricted to integers.

Shenoy and West [2011] have proposed mixtures of polynomials, in the same spirit as MTEs, as a solution to the integration problem. Shenoy [2010] proposes relaxing the hypercube condition of MOP functions, which enables easy representation of two and three-dimensional CLG conditionals by MOP functions. The family of MOP functions is closed under transformations needed for multi-dimensional linear and quotient deterministic functions.

For Bayesian decision problems, Kenley [1986] (see also Shachter and Kenley [1989]) describes the representation and solution of Gaussian IDs that include continuous chance variables with conditional linear Gaussian distributions. Poland [1994] extends Gaussian IDs to

mixture of Gaussians IDs. Thus, continuous chance variables can have any distributions, and these are approximated by mixtures of Gaussians. Cobb and Shenoy [2008] extend MTE BNs to MTE IDs for the special case where all decision variables are discrete. Li and Shenoy [2010] have proposed an architecture that is an extension of the architecture described in this paper for solving hybrid influence diagrams with deterministic variables.

In this paper, we describe a generalization of the Shenoy-Shafer architecture for discrete BNs so that it applies to hybrid BNs with deterministic conditionals for continuous variables. The functions associated with deterministic conditionals do not have to be linear (as in the CLG case) or even invertible. We use Dirac delta functions to represent such functions. We keep track of the units of continuous potentials. This enables us, e.g., to describe the units of the normalization constant, which are often referred to as “probability” of evidence. Finally, we illustrate our architecture using several small examples, and by solving a modified version of the Crop problem initially introduced by Murphy [1999].

An outline of the remainder of the paper is as follows. In Section 2, we define Dirac delta functions and describe some of their properties. In Section 3, we describe our architecture for making inferences in hybrid BNs with deterministic variables. This is the main contribution of this paper. In Section 4, we describe four small examples of hybrid BNs with deterministic variables to illustrate our definitions and our architecture. In Section 5, we describe and solve a modification of the crop problem, initially described by Murphy [1999], and subsequently modified by a number of authors. Finally, in Section 6, we end with a summary and discussion.

2 Dirac Delta Functions

In this section, we define Dirac delta functions. We use Dirac delta functions to represent deterministic conditionals associated with some continuous variables in BNs. Dirac delta functions are also used to represent observations of continuous variables.

$\delta: \mathbb{R} \rightarrow \mathbb{R}^+$ 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. The values of δ are assumed to be in units of density. δ is not a proper function since the value of the function at 0 doesn’t exist (i.e., 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 $\sigma \rightarrow 0$). 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].

As defined above, the value $\delta(0)$ is undefined, i.e., ∞ , in units of 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 $M(t) = e^{\sigma^2 t^2/2}$. In the limit as $\sigma \rightarrow 0$, $M(t) = 1$. Now,

$M(t) = 1$ is the MGF of the distribution $X = 0$ with probability 1. Therefore, we can *interpret* the value $\delta(0)$ (in units of density) as probability 1 at the location $x = 0$.

Some basic properties of the Dirac delta functions that are useful in uncertain reasoning are described in the Appendix. Properties (i)–(iv) are useful in integrating potentials containing Dirac delta functions. Property (v) defines the Heaviside function, which is related to the Dirac delta function. Properties (vi)–(x) are useful in representing deterministic conditionals by Dirac delta functions.

Consider a simple Bayesian network consisting of two continuous variables X and Y with X as a parent of Y . Suppose X has PDF $f_X(x)$, and suppose the conditional PDF for Y given $X = x$ is given by $f_{Y|x}(y)$. Then, it follows from probability theory that the marginal for Y can be found by first multiplying the two PDFs to yield the joint PDF of X and Y , and then integrating X from the joint. Thus, if $f_Y(y)$ denotes the marginal of Y ,

$$f_Y(y) = \int f_X(x) f_{Y|x}(y) dx. \quad (2.1)$$

Now suppose that Y has a deterministic conditional given by the equation $Y = g(X)$, i.e., given $X = x$, $Y = g(x)$ with probability 1. In this case, there does not exist a joint PDF for X and Y . However, property (vi) of Dirac delta functions tells us that we can represent the conditional for $Y|x$ by the Dirac delta function $\delta(y - g(x))$, and we can find the marginal for Y in the usual way using (2.1), i.e.,

$$f_Y(y) = \int f_X(x) \delta(y - g(x)) dx. \quad (2.2)$$

The result in equation (2.2) is valid regardless of the nature of the function g . However, the integration in (2.2) is possible only if the function g is differentiable and the real roots of the equation $y - g(x) = 0$ in x can be computed in terms of y . This includes a wide family of functions including non-invertible ones, such as e.g., $Y = X^2$.

We can extend the result in equation (2.2) for deterministic conditionals with several parents. For example, consider a Bayesian network with three continuous variables X_1 , X_2 , and Y , such that X_2 has X_1 as a parent and Y has X_1 and X_2 as parents. Suppose the PDF of X_1 is given by $f_{X_1}(x_1)$, the conditional PDF of X_2 given x_1 is given by $f_{X_2|x_1}(x_2)$, and Y has a deterministic conditional given by $Y = g(X_1, X_2)$. Then, we can represent the deterministic conditional for Y by the Dirac delta function $\delta(y - g(x_1, x_2))$, and property (ix) tells us that we can find the marginal PDF of Y as follows:

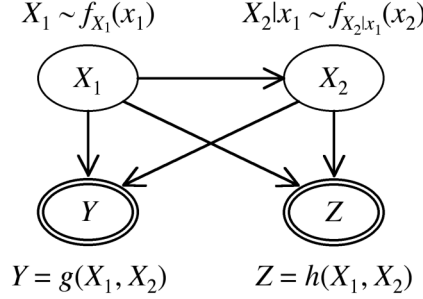
$$f_Y(y) = \iint f_{X_1}(x_1) f_{X_2|x_1}(x_2) \delta(y - g(x_1, x_2)) dx_2 dx_1 \quad (2.3)$$

Finally, consider the Bayesian network consisting of four continuous variables as shown in Figure 1. The continuous potentials associated with deterministic conditionals for variables Y and

Z are $\delta(y - g(x_1, x_2))$ and $\delta(z - h(x_1, x_2))$, respectively. Property (x) tells that the joint PDF of Y and Z , denoted by $f_{Y,Z}(y, z)$, can be computed as follows:

$$f_{Y,Z}(y, z) = \int \int f_{X_1}(x_1) f_{X_2|x_1}(x_2) \delta(y - g(x_1, x_2)) \delta(z - h(x_1, x_2)) dx_2 dx_1. \quad (2.4)$$

Figure 1. A Bayesian network with two deterministic variables



In general, if Y is a continuous variable with a deterministic conditional $Y = g(X_1, \dots, X_n)$, where $\{X_1, \dots, X_n\}$ are the continuous parents of Y , then such a deterministic conditional is represented by the continuous potential $\psi(\mathbf{x}, y) = \delta(y - g(\mathbf{x}))$ for all $\mathbf{x} \in \Omega_{\{X_1, \dots, X_n\}}$, and $y \in \Omega_Y$. If Y is a continuous variable with continuous parents $\{X_1, \dots, X_n\}$, and discrete parents $\{A_1, \dots, A_m\}$, and has a deterministic conditional $Y = g_i(X_1, \dots, X_n)$ if $(A_1, \dots, A_m) = \mathbf{a}_i$ for $i = 1, \dots, |\Omega_{\{A_1, \dots, A_m\}}|$, then such a deterministic conditional is represented by the continuous potential $\psi(\mathbf{x}, \mathbf{a}_i, y) = \delta(y - g_i(\mathbf{x}))$ for all $\mathbf{x} \in \Omega_{\{X_1, \dots, X_n\}}$, $\mathbf{a}_i \in \Omega_{\{A_1, \dots, A_m\}}$, and $y \in \Omega_Y$.

3 An Architecture for Computing Marginals

In this section, we will describe an extended Shenoy-Shafer architecture for representing and solving hybrid BNs with deterministic variables. The Shenoy-Shafer architecture [Shenoy and Shafer 1990] was initially proposed for computing marginals in discrete Bayesian networks. It was extended by Moral *et al.* [2001] to include continuous variables for propagation of mixtures of truncated exponentials. Cobb and Shenoy [2005a] extended it further to include linear deterministic variables. Cinicioglu and Shenoy [2009] extended it further to include linear and non-linear deterministic functions to define arc reversals. They propose the use of Dirac delta functions for representing conditionals of deterministic variables.

3.1 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 measurable 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 $\Omega_\emptyset = \{\diamond\}$.

In a BN, each variable has a conditional distribution function for each state of its parents. A conditional distribution function associated with a variable is said to be *deterministic* if the variances (for each state of its parents) are all zeros. Deterministic conditionals for discrete variables pose no computational problems as the joint probability mass function for all discrete variables exists. However, deterministic conditionals for continuous variables pose a computational challenge, as the joint density function for all continuous variables does not exist. Henceforth, when we speak of deterministic conditionals, we are referring to continuous variables, and to avoid convoluted language, we will loosely refer to a continuous variable with a deterministic conditional as a deterministic variable. In a BN, discrete variables are denoted by rectangular-shaped nodes, continuous variables by oval-shaped nodes, and deterministic variables by oval-shaped nodes with a double border.

3.2 Projection of States

If $\mathbf{x} \in \Omega_r$, $\mathbf{y} \in \Omega_s$, and $r \cap s = \emptyset$, then $(\mathbf{x}, \mathbf{y}) \in \Omega_{r \cup s}$. Thus, $(\mathbf{x}, \diamond) = \mathbf{x}$. Suppose $\mathbf{x} \in \Omega_r$, and suppose $s \subseteq r$. Then the *projection* of $\mathbf{x}$ to s , denoted by $\mathbf{x}^{\downarrow s}$, is the state of s obtained from $\mathbf{x}$ 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 $\mathbf{x}^{\downarrow s} = \mathbf{x}$. If $s = \emptyset$, then $\mathbf{x}^{\downarrow s} = \diamond$.

3.3 Discrete Potentials

In a BN, the conditional probability function associated with each variable is 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*. Formally, suppose $r \subseteq V$. A discrete potential α for r is a function $\alpha: \Omega_r \rightarrow [0, 1]$ such that its values are in units of probability, which are dimension-less numbers in the interval $[0, 1]$. By dimension-less, we mean they do not have physical units (such as, e.g., feet/meters, pounds/grams, seconds, or some combination of these).

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.

For an example of a discrete potential, suppose β is a discrete potential for $\{B, P\}$, where B is a discrete variable with states $\{b, nb\}$ and P is a continuous variable, such that $\beta(b, p) =$

$$\frac{1}{1 + e^{-6.5+p}}, \text{ and } \beta(nb, p) = \frac{e^{-6.5+p}}{1 + e^{-6.5+p}}. \text{ The values of } \beta \text{ are in units of probability.}$$

Another example of a discrete potential is the identity discrete potential for the empty set, denoted by ι_d , such that $\iota_d(\diamond) = 1$. The sole value 1 of ι_d is in units of probability.

3.4 Continuous Potentials

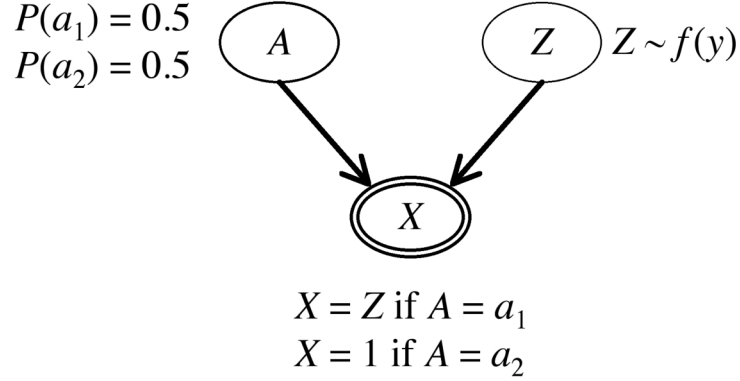
If X is a continuous variable in a BN, it is associated with a conditional distribution that is represented by a function called a continuous potential. Formally, suppose $x \subseteq V$. Then, a *continuous potential* ξ for x is a function $\xi: \Omega_x \rightarrow \mathbb{R}^+$ such that its values are in units of density.

For example, if Z is a continuous variable with the standard normal distribution, then the values of the continuous potential for Z , $\zeta(z) = (1/\sqrt{2\pi})e^{-z^2/2}$, are in units of density. More precisely, the values $\zeta(z)$ are in units of probability/unit of Z , which is denoted by $(\text{unit } Z)^{-1}$.

For another example, suppose X is a deterministic variable with parents A and Z , where A is discrete with states a and na , and Z is continuous. Suppose the deterministic function defining X is as follows: $X = 1$ if $A = a$, and $X = Z$ if $A = na$. Then, this conditional is represented by a continuous potential ξ for $\{A, Z, X\}$ such that $\xi(a, z, x) = \delta(x - 1)$, and $\xi(na, z, x) = \delta(x - z)$. The values of ξ are in units of $(\text{unit } X)^{-1}$.

For yet another example, consider a continuous variable X with a mixed distribution: a probability of 0.5 at $X = 1$, and a probability density of $0.5 f(x)$, where $f(x)$ 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 the values of ξ are in units of $(\text{unit } X)^{-1}$, and that $\int \xi(x) dx = 0.5 \int \delta(x - 1) dx + 0.5 \int f(x) dx = 0.5 + 0.5 = 1$ (in units of probability), so that it is a proper distribution function.

Consider the BN in Figure 2. A is discrete (with two states, a_1 and a_2), Z and X are continuous, and X has a deterministic conditional. Let α denote the discrete potential for $\{A\}$. Then $\alpha(a_1) = 0.5$, $\alpha(a_2) = 0.5$. Let ζ denote the continuous potential for $\{Z\}$ in $(\text{unit } Z)^{-1}$. Then $\zeta(z) = f(z)$. Let ξ denote the continuous potential for $\{A, Z, X\}$ in $(\text{unit } X)^{-1}$. Then $\xi(a_1, z, x) = \delta(x - z)$, and $\xi(a_2, z, x) = \delta(x - 1)$. This BN will be analyzed further in Example 4.3 in Section 4.

Figure 2. A hybrid BN with a discrete, a continuous and a deterministic variable

3.5 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.,

$$(\alpha \otimes \beta)(x) = \alpha(x^{\downarrow a}) \beta(x^{\downarrow b}) \text{ for all } x \in \Omega_{a \cup b}. \quad (3.1)$$

The units of $\alpha \otimes \beta$ are the product of the units of α and β . Thus, if α and β are both discrete potentials, then $\alpha \otimes \beta$ is a discrete potential (since the product of two probabilities is a probability). In all other cases, $\alpha \otimes \beta$ is a continuous potential (since the product of probability and density or the product of two densities are densities).

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 ($((\alpha \otimes \beta) \otimes \gamma = \alpha \otimes (\beta \otimes \gamma))$).

3.6 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 properties of Dirac delta functions (see properties (i)–(iv) in the Appendix). Also, after marginalization of a continuous variable, the nature of a potential could change from continuous to discrete.

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.,

$$\alpha^{-X}(y) = \sum \{ \alpha(x, y) \mid x \in \Omega_X \} \text{ for all } y \in \Omega_{a \setminus \{X\}}. \quad (3.2)$$

The nature of α^{-X} depends on the nature of α . If α is discrete, then α^{-X} is discrete, and if α is continuous, then α^{-X} is continuous. This follows from equation (3.2) since marginalization is addition, and sums of probabilities are probabilities and sums of densities are densities.

Suppose α is a continuous potential for a , and suppose 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.,

$$\alpha^{-X}(\mathbf{y}) = \int \alpha(x, \mathbf{y}) dx \text{ for all } \mathbf{y} \in \Omega_{a \setminus \{X\}}. \quad (3.3)$$

In this case, the nature of α^{-X} is slightly more complex. First, before we marginalize a variable X , we combine all potentials that include X in their domains. Since X is continuous, there is always a conditional for X , which is a continuous potential whose domain contains X (in units of density). Since the product of probability and density is density, the potential α that includes the conditional for X will always be a continuous potential. Second, if the units of α are $(\text{unit } X)^{-1}$, then α^{-X} will be in units of probability since integrating α with respect to X is tantamount to multiplying the units of α by unit X . However, if the units of α include other units such as, e.g. $(\text{unit } X)^{-1}(\text{unit } Y)^{-1}$, then α^{-X} will be in units of $(\text{unit } Y)^{-1}$, and, thus, units of density.

For example, if $\xi(x) = f_X(x)$ is the PDF of X in units of $(\text{unit } X)^{-1}$, and $\psi(x, y) = f_{Y|x}(y)$ is the conditional PDF of Y given $X = x$ in units of $(\text{unit } Y)^{-1}$, then $\xi^{-X} = \mathbf{1}_d$ (identity discrete potential for the empty set), ψ^{-Y} is the identity discrete potential for $\{X\}$, i.e., $\psi^{-Y}(x) = 1$ (in units of probability) for all x , whereas $(\xi \otimes \psi)^{-X}$ is a density potential for $\{Y\}$ in units of $(\text{unit } Y)^{-1}$. The same is true if we have deterministic conditionals represented by Dirac delta functions. Thus if $\psi(x, y) = \delta(y - g(x))$ is the deterministic conditional of Y given x , then the values of ψ are in units of $(\text{unit } Y)^{-1}$. As before, ψ^{-Y} is the identity discrete potential for $\{X\}$ (from property (ii) of Dirac delta functions, $\int \delta(y - g(x)) dy = 1$), and $(\xi \otimes \psi)^{-X}$ is a density potential for $\{Y\}$ in units of $(\text{unit } Y)^{-1}$.

If α contains Dirac delta functions, then we have to use properties of Dirac delta functions (described in the Appendix) in doing the integration. The two most important properties are the sampling property (properties (i) and (ii)) and the re-scaling property (property (iv)). For example, if we consider $y - x^2$ as a function of x , then by using the rescaling property we have:

$$\delta(y - x^2) = \frac{1}{2\sqrt{y}} \left(\delta\left(x - \sqrt{y}\right) + \delta\left(x + \sqrt{y}\right) \right) \text{ if } y \geq 0,$$

and by using the sampling property we have:

$$\begin{aligned} \int f_X(x) \delta(y - x^2) dx &= \frac{1}{2\sqrt{y}} \int f_X(x) \left(\delta\left(x - \sqrt{y}\right) + \delta\left(x + \sqrt{y}\right) \right) dx, \text{ if } y \geq 0, \\ &= \frac{1}{2\sqrt{y}} \left(f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right), \text{ if } y \geq 0. \end{aligned}$$

The Dirac delta function is implemented in Mathematica® and Maple®, so the properties (i)–(iv) described in the Appendix can be implemented on a computer. However, not all deterministic functions can be handled using Dirac delta functions. Some limitations are as follows. First, it must be possible to find the real zeroes of the function in closed form as a function of other variables. Second, to enable the computation of the derivative in the re-scaling property, the function must be differentiable, and the value of the derivative at the real zeroes of the function must be non-zero. Thus, Dirac delta functions can be used, e.g., with linear functions ($W = X + Y$), products ($W = X \cdot Y$), and quotients ($W = X/Y$). However, they cannot be used, e.g., with functions such as $W = \max\{X, Y\}$.

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 will refer to this property as *local computation*.

3.7 Normalization of Potentials

The Shenoy-Shafer [1990] 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 by a constant.

Suppose ξ is a discrete or continuous potential for $\{X\}$ representing the un-normalized posterior marginal for X . To normalize ξ , we divide all values of ξ by the constant $\xi^{-X}(\diamond)$, i.e., if ξ' denotes the normalized potential for $\{X\}$, then

$$\xi'(x) = \xi(x) / \xi^{-X}(\diamond), \text{ for all } x \in \Omega_X. \quad (3.4)$$

If ξ is a discrete potential (in units of probability), and X is a discrete variable, then $\xi^{-X}(\diamond)$ is in units of probability, and the normalized potential ξ' is a discrete potential for $\{X\}$. If ξ is a continuous potential, say in units of $(\text{unit } X)^{-1}$, and X is a discrete variable, then $\xi^{-X}(\diamond)$ is in units of $(\text{unit } X)^{-1}$, and consequently, the normalized potential ξ' is a discrete potential for $\{X\}$ (since the units of the values of ξ' are now dimension-less, i.e., units of probability). Finally, if ξ is a continuous potential, say in units of $(\text{unit } X)^{-1}$, and X is a continuous variable, then $\xi^{-X}(\diamond)$ is in units of probability, and consequently, ξ' is a continuous potential for $\{X\}$ in units of $(\text{unit } X)^{-1}$.

Depending on the units of $\xi^{-X}(\diamond)$, it represents either the probability of the evidence if $\xi^{-X}(\diamond)$ is in units of probability, or it represents the density of the evidence if $\xi^{-X}(\diamond)$ is in units of density. One advantage of keeping track of the units of continuous potentials is that it allows us

to determine the units of the normalization constant, whether it is probability or density. Thus, for methods that are based on analysis of the normalization constant (see, e.g., [Nielsen and Jensen 2007]), it is crucial to distinguish between probability and density of evidence.

3.8 Solving Hybrid Bayesian Networks

We have all the definitions needed to solve hybrid BNs with deterministic variables. The solution algorithm is essentially the same as described in Shenoy and Shafer [1990] and Shenoy [1997], i.e., we use the Shenoy-Shafer architecture to propagate the potentials in a binary join tree.

A major issue in solving hybrid Bayesian networks is marginalizing continuous variables, which involves integration. In general, there is no guarantee that we can always find the result of integration in closed form. One solution is to approximate all PDFs by MTE functions [Moral *et al.* 2001]. The family of MTE functions is closed under combination, marginalization, and transformations needed for one-dimensional linear deterministic functions. For one-dimensional non-linear deterministic functions, Cobb and Shenoy [2005b] propose such functions by piecewise linear ones.

Another solution is to approximate all PDFs by mixtures of polynomials [Shenoy and West 2011, Shenoy 2010]. The family of mixture of polynomials functions are closed under combination, marginalization, and transformations needed for multi-dimensional linear and quotient deterministic functions. Like MTEs, non-linear deterministic functions can be approximated by piecewise linear functions.

In this paper, the focus is on the architecture for making inferences in hybrid Bayesian networks without concerning ourselves explicitly with the problem of integration. Of course, to be useful in practice, we need to address also the problem of integration. By combining the research on MTE and mixture of polynomials functions with the architecture described here, we can now solve hybrid Bayesian networks that were not solvable before.

4 Some Illustrative Examples

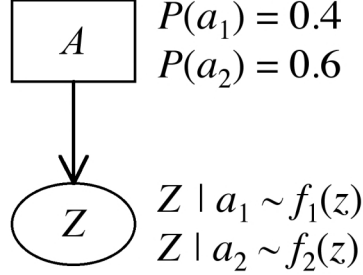
In this section, we will illustrate our framework and definitions using several small illustrative examples. For each continuous potential, we keep track of its units.

4.1 Example 1: Mixture distribution

Consider a hybrid BN with a discrete variable and a continuous variable as shown in Figure 3. A is discrete and Z is continuous. What is the prior marginal distribution of Z ? Suppose we observe

$Z = c$ (where c is such that the marginal density of Z at c is positive). What is the posterior marginal distribution of A ?

Figure 3. A hybrid BN with a discrete and a continuous variable



Let α denote the discrete potential for A (in units of probability), ζ_1 denote the continuous potential for $\{A, Z\}$ in units of $(\text{unit } Z)^{-1}$. Then,

$$\begin{aligned}\alpha(a_1) &= 0.4, \\ \alpha(a_2) &= 0.6; \\ \zeta_1(a_1, z) &= f_1(z), (\text{unit } Z)^{-1} \\ \zeta_1(a_2, z) &= f_2(z), (\text{unit } Z)^{-1}.\end{aligned}$$

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

$$((\alpha \otimes \zeta_1)^{-A})(z) = 0.4 f_1(z) + 0.6 f_2(z), (\text{unit } Z)^{-1}$$

Thus, Z has a mixture PDF weighted by the probabilities of A . Let $f_Z(z)$ denote $0.4 f_1(z) + 0.6 f_2(z)$. Let ζ_2 denote the observation potential for Z . We assume the constant c is such that $f_Z(c) = 0.4 f_1(c) + 0.6 f_2(c) > 0$, i.e., either $f_1(c) > 0$ or $f_2(c) > 0$ or both. To find the posterior marginal for A , first we combine ζ_1 and ζ_2 , next we marginalize Z from the combination, and finally we combine the result with α .

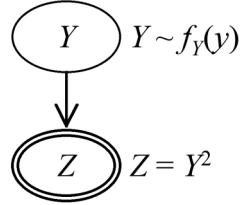
$$\begin{aligned}((\zeta_1 \otimes \zeta_2)^{-Z} \otimes \alpha)(a_1) &= 0.4 f_1(c), (\text{unit } Z)^{-1} \\ ((\zeta_1 \otimes \zeta_2)^{-Z} \otimes \alpha)(a_2) &= 0.6 f_2(c), (\text{unit } Z)^{-1}.\end{aligned}$$

The normalization constant is $0.4 f_1(c) + 0.6 f_2(c) = f_Z(c)$, in $(\text{unit } Z)^{-1}$, representing density of evidence. After normalization, the posterior marginal distribution of A is $0.4 f_1(c) / (0.4 f_1(c) + 0.6 f_2(c))$ at a_1 , and $0.6 f_2(c) / (0.4 f_1(c) + 0.6 f_2(c))$ at a_2 , both in units of probability.

4.2 Example 2: Transformation of variables

Consider a BN with continuous variable Y and deterministic variable Z as shown in Figure 4. Notice that the function defining the deterministic variable is not invertible. What is the prior marginal distribution of Z ? If we observe $Z = c$, what is the posterior marginal distribution of Y ?

Figure 4. A continuous BN with a deterministic variable with a non-invertible function.



Let ψ denote the continuous potential for $\{Y\}$ (in (unit Y)⁻¹) and let ζ_1 denote the deterministic conditional for Y (in (unit Z)⁻¹). Then,

$$\begin{aligned}\psi(y) &= f_Y(y), \text{ (unit } Y)^{-1}; \\ \zeta_1(y, z) &= \delta(z - y^2), \text{ (unit } Z)^{-1}.\end{aligned}$$

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

$$((\psi \otimes \zeta_1)^{-Y})(z) = \left(\frac{1}{2\sqrt{z}} \right) (f_Y(\sqrt{z}) + f_Y(-\sqrt{z})) \text{ for } z > 0, \text{ (unit } Z)^{-1} \quad (4.1)$$

The result in (4.1) follows from properties (iv) and (ii) of Dirac delta functions. Let $f_Z(z)$ denote

$\left(\frac{1}{2\sqrt{z}} \right) (f_Y(\sqrt{z}) + f_Y(-\sqrt{z}))$. 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 continuous potential for Z , $\zeta_2(z) = \delta(z - c)$, (unit Z)⁻¹. Then, the un-normalized posterior marginal distribution of Y is computed as follows:

$$\begin{aligned}(\zeta_1 \otimes \zeta_2)^{-Z}(y) &= \int \delta(z - y^2) \delta(z - c) dz = \delta(y^2 - c), \text{ (unit } Z)^{-1} \\ (\psi \otimes (\zeta_1 \otimes \zeta_2)^{-Z})(y) &= \frac{f_Y(\sqrt{c})\delta(y - \sqrt{c}) + f_Y(-\sqrt{c})\delta(y + \sqrt{c})}{2\sqrt{c}}, \text{ (unit } Y)^{-1} \text{ (unit } Z)^{-1}\end{aligned}$$

The normalization constant is $(f_Y(\sqrt{c}) + f_Y(-\sqrt{c})) / (2\sqrt{c}) = f_Z(c)$ is in units of (unit Z)⁻¹.

Therefore the normalized posterior distribution of Y is

$(f_Y(\sqrt{c}) \delta(y - \sqrt{c}) + f_Y(-\sqrt{c}) \delta(y + \sqrt{c})) / (f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))$, in units of (unit Y)⁻¹. This can be

interpreted as follows: $Y = \sqrt{c}$ with probability $f_Y(\sqrt{c})/(f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))$, and $Y = -\sqrt{c}$ with probability $f_Y(-\sqrt{c})/(f_Y(\sqrt{c}) + f_Y(-\sqrt{c}))$.

4.3 Example 3: Mixed distributions

Consider the hybrid BN shown in Figure 2 with three variables. A is discrete with state space $\Omega_A = \{a_1, a_2\}$, Z and X are continuous, and the conditional associated with 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 denote the discrete potential for $\{A\}$ (in units of probability), ζ the continuous potential for Z (in units of $(\text{unit } Z)^{-1}$), and ξ_1 the conditional for X (in units of $(\text{unit } X)^{-1}$). Then:

$$\begin{aligned}\alpha(a_1) &= 0.5, \\ \alpha(a_2) &= 0.5; \\ \zeta(z) &= f_Z(z), (\text{unit } Z)^{-1}; \\ \xi_1(a_1, z, x) &= \delta(x - z), (\text{unit } X)^{-1}, \\ \xi_1(a_2, z, x) &= \delta(x - 1), (\text{unit } X)^{-1}.\end{aligned}$$

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

$$(((\alpha \otimes \xi_1)^{-A} \otimes \zeta)^{-Z})(x) = 0.5 f_Z(x) + 0.5 \delta(x - 1), (\text{unit } X)^{-1}$$

The normalization constant is 1 (in units of probability). Thus the prior marginal distribution of X is mixed with PDF $0.5 f_Z(x)$ and a probability of 0.5 at $X = 1$.

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

$$\begin{aligned}(\alpha \otimes (\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z})(a_1) &= 0.5 f_Z(1), (\text{unit } X)^{-1} \\ (\alpha \otimes (\zeta \otimes (\xi_1 \otimes \xi_2)^{-X})^{-Z})(a_2) &= 0.5 \delta(0), (\text{unit } X)^{-1}.\end{aligned}$$

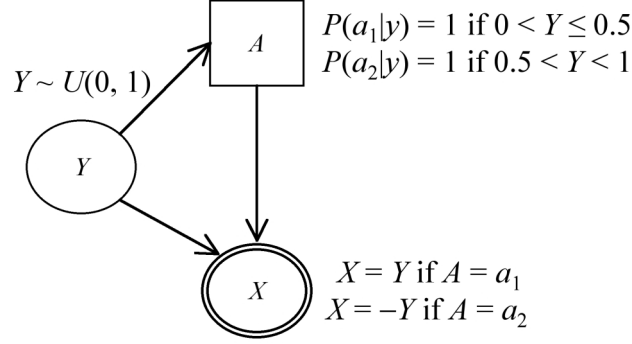
The normalization constant is $0.5(f_Z(1) + \delta(0))$, $(\text{unit } X)^{-1}$. Thus, after normalization, the posterior probability of a_1 is 0, and the posterior probability of a_2 is 1, both in units of probability. The normalization constant can be interpreted as 0.5 in units of probability.

4.4 Example 4: Discrete Variable with Continuous Parents

Consider the hybrid BN consisting of continuous variables X and Y , a discrete variable A , and a deterministic conditional associated with X as shown in Figure 5. 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 of A ?

Figure 5. A hybrid BN with a continuous, a discrete, and a deterministic variable.



Let ψ denote the continuous potential for $\{Y\}$, α the discrete potential for $\{Y, A\}$, and ξ_1 the continuous potential for $\{Y, A, X\}$.

$\psi(y) = f_Y(y)$, (unit Y) $^{-1}$, where $f_Y(y) = 1$ if $0 < y < 1$, $= 0$ otherwise;

$\alpha(a_1, y) = H(y) - H(y - 0.5)$, where $H(\cdot)$ is the *Heaviside* function (defined in property (v) of Dirac Delta functions in the Appendix),

$\alpha(a_2, y) = H(y - 0.5) - H(y - 1)$;

$\xi_1(a_1, y, x) = \delta(x - y)$, (unit X) $^{-1}$

$\xi_1(a_2, y, x) = \delta(x + y)$, (unit X) $^{-1}$.

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.

$$(((\alpha \otimes \xi_1)^{-A}) \otimes \psi)^{-Y}(x) = H(x) - H(x - 0.5) + H(-x - 0.5) - H(-x - 1), \text{ (unit } X)^{-1} \quad (4.2)$$

Thus, the prior marginal distribution of X in (4.2) is uniform in the interval $(-1, -0.5) \cup (0, 0.5)$.

Let ξ_2 be the continuous potential denoting the observation that $X = 0.25$. Thus, $\xi_2 = \delta(x - 0.25)$, (unit X) $^{-1}$. The un-normalized posterior marginal of Y is given by $(\xi_1 \otimes (\xi_2 \otimes \alpha))^{-\{A, X\}} \otimes \psi$.

$$(((\xi_1 \otimes (\xi_2 \otimes \alpha))^{-\{A, X\}}) \otimes \psi)(y) = f_Y(0.25) \delta(y - 0.25), \text{ (unit } X)^{-1} \text{ (unit } Y)^{-1}.$$

The normalization constant is $f_Y(0.25)$ in units of (unit X) $^{-1}$. The normalized posterior marginal for Y is $\delta(y - 0.25)$, (unit Y) $^{-1}$, i.e., $Y = 0.25$ with probability 1. The un-normalized posterior distribution of A is given by $((\xi_1 \otimes (\xi_2 \otimes \alpha))^{-X} \otimes \psi)^{-Y}$.

$$(((\xi_1 \otimes (\xi_2 \otimes \alpha))^{-X} \otimes \psi)^{-Y}(a_1) = f_Y(0.25), \text{ (unit } Y)^{-1}$$

$$(((\xi_1 \otimes (\xi_2 \otimes \alpha))^{-X}) \otimes \psi)^{-Y}(a_2) = 0, (\text{unit } Y)^{-1}$$

The normalization constant is $f_Y(0.25)$, $(\text{unit } Y)^{-1}$, the same as that for the marginal of Y . After normalization the posterior probability of a_1 is 1, and the posterior probability of a_2 is 0.

5 The Extended Crop Problem

In this section, we describe a modification of the Crop problem initially described by Murphy [1999], and extended by Lerner [2002]. Here we extend it further to include deterministic variables and we describe its exact solution using the extended Shenoy-Shafer framework described in Section 3.

The hybrid Bayesian network of the extended crop network is shown in Figure 6. *Policy* (Po) is a discrete variable and describes the nature of the policy in place, liberal (l) or conservative (c). *Rain* (R) is discrete and has three states: *drought* (d), *average* (a), or *flooding* (f). *Subsidy* (S), with states *subsidy* (s) or *no subsidy* (ns), is a discrete variable whose conditional distribution depends on *Policy* and *Rain*. *Crop* (C) is a continuous variable that denotes the size of the crop yield (in million bushels (mB)). It is dependent on *Rain*, and anything other than average lowers expected yield. *Price* (Pr) (in \$/B) is a continuous variable, and is dependent on *Subsidy* and *Crop*. For a given state of the variable *Subsidy*, the expected value of *Price* decreases as the yield increases. Similarly for a given crop yield, the price will be lower if there is a subsidy. *Buy* (B) is a discrete variable with states *buy* (b) and *not buy* (nb) whose conditional distribution depends on *Price*, and denotes whether a prospective buyer will buy the entire crop yield or not. It depends on *Price*, and as the price increases, the probability that the crop will be bought decreases. *Revenue*₁ (R_1) is a deterministic variable, and it denotes the portion of revenue (in m\$) the farmer will receive from selling the crop. It depends on *Buy*, *Price* and *Crop*. $Revenue_1 = Crop \cdot Price$ if $Buy = b$, $Revenue_1 = 0$ if $Buy = nb$. *Revenue*₂ (R_2) is also a deterministic variable, and represents the portion of revenue (in m\$) the farmer will receive due to the subsidy, if any. $Revenue_2 = 2$ if $Subsidy = s$, $Revenue_2 = 0$ if $Subsidy = ns$. *Revenue*₃ (R_3) is another deterministic variable, and $Revenue_3 = Revenue_1 + Revenue_2$ (in m\$).

Let π denote the discrete potential for *Policy*, ρ denote the discrete potential for *Rain*, σ denote the discrete potential for *Subsidy*, χ denote the continuous potential for *Crop* (in mB^{-1}), ϕ denote the continuous potential for *Price* (in $(\$/B)^{-1}$), and β denote the discrete potential for *Buy*. The details of these potentials are shown in Table 1. To avoid problems with integrating density functions, we have assumed beta densities for crop and price instead of the normal distribution. Suppose X is a continuous variable, $m > 0$, $n > 0$, and $a < b$. We say $X \sim \text{Beta}[m, n]$ on $[a, b]$ if the PDF of X is as follows:

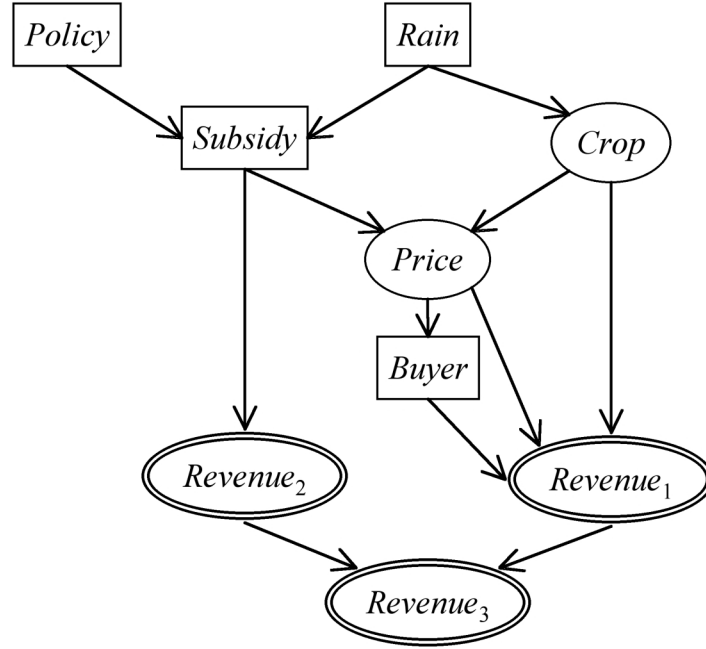
$$f_X(x) = \frac{1}{\beta(m,n)(b-a)} \left(\frac{x-a}{b-a} \right)^{m-1} \left(1 - \frac{x-a}{b-a} \right)^{n-1} \text{ if } a \leq x \leq b, \quad (5.1)$$

where $\beta(m, n)$ is a constant such that $\int f_X(x) dx = 1$. The function β is called Euler's beta function, and is defined as follows. If $m > 0$, and $n > 0$, then

$$\beta(m, n) = \int_0^1 t^{m-1} (1-t)^{n-1} dt.$$

Notice that if $a = 0$, and $b = 1$, then the PDF in (5.1) reduces to the standard Beta[m, n] on $[0, 1]$ PDF.

Figure 6. The hybrid BN for the extended crop problem



Let τ_1, τ_2, τ_3 denote the Dirac potentials at $Revenue_1, Revenue_2$, and $Revenue_3$, respectively. Then,

$$\tau_1(b, p, c, r_1) = \delta(r_1 - p c), \tau_1(nb, p, c, r_1) = \delta(r_1), \text{ in } (\text{m\$})^{-1};$$

$$\tau_2(s, r_2) = \delta(r_2 - 2), \tau_2(ns, r_2) = \delta(r_2), \text{ in } (\text{m\$})^{-1};$$

$$\tau_3(r_1, r_2, r_3) = \delta(r_3 - r_1 - r_2), (\text{m\$})^{-1}.$$

We will describe the computation of the marginal for R_3 . Suppose we delete R_2 first. R_2 is in the domain of τ_2 and τ_3 . Let τ_4 denote the Dirac potential $(\tau_2 \otimes \tau_3)^{-R_2}$ (in $(\text{m\$})^{-1}$). Then,

$$\tau_4(s, r_1, r_3) = \int \delta(r_2 - 2) \delta(r_3 - r_1 - r_2) dr_2 = \delta(r_3 - r_1 - 2),$$

$$\tau_4(ns, r_1, r_3) = \int \delta(r_2) \delta(r_3 - r_1 - r_2) dr_2 = \delta(r_3 - r_1), (\text{m\$})^{-1}.$$

Table 1. The discrete and density potentials for the variables in the extended crop problem

<i>Policy</i>			π			
<i>liberal (l)</i>			0.5			
<i>conservative (c)</i>			0.5			
<i>Rain</i>			ρ			
<i>drought (d)</i>			0.35			
<i>average (a)</i>			0.60			
<i>flood (f)</i>			0.05			
σ	<i>l</i>			<i>c</i>		
<i>Subsidy</i>	<i>d</i>	<i>a</i>	<i>f</i>	<i>d</i>	<i>a</i>	<i>f</i>
<i>subsidy (s)</i>	0.4	0.95	0.5	0.3	0.95	0.2
<i>no subsidy (ns)</i>	0.6	0.05	0.5	0.7	0.05	0.8
<i>Crop</i> <i>Rain</i>		χ (in (mB) ⁻¹)				
<i>drought (d)</i>		<i>Beta</i> [2, 2] over the range (1.5, 4.5)				
<i>average (a)</i>		<i>Beta</i> [2, 2] over the range (2, 8)				
<i>flooding (f)</i>		<i>Beta</i> [2, 2] over the range (0.5, 3.5)				
<i>Price</i> (<i>Subsidy</i> , <i>Crop</i> = <i>c</i>)		φ (in (\$/B) ⁻¹)				
<i>subsidy (s)</i>		<i>Beta</i> [2, 2] over the range (8.5 − <i>c</i> , 11.5 − <i>c</i>)				
<i>no subsidy (ns)</i>		<i>Beta</i> [2, 2] over the range (10.5 − <i>c</i> , 13.5 − <i>c</i>)				
<i>Buy</i> <i>Price</i> = <i>p</i>		β				
<i>buy (b)</i>	$= 1 + 0.001212\, p - 0.01202\, p^2$					if $0 \leq p \leq 6.5$
	$= -0.001212\, (13 - p) + 0.01202\, (13 - p)^2$					if $6.5 < p \leq 13$
	$= 0$					if $p > 13$
<i>not buy (nb)</i>		$1 - \beta(b, p)$				

Next, suppose we delete R_1 next. R_1 is in the domain of τ_1 and τ_4 . Let τ_5 denote the Dirac potential $(\tau_1 \otimes \tau_4)^{-R_1}$ (in (m\$)⁻¹). Then,

$$\tau_5(s, b, p, c, r_3) = \int \delta(r_1 - p c) \delta(r_3 - r_1 - 2) dr_1 = \delta(r_3 - p c - 2),$$

$$\tau_5(s, nb, p, c, r_3) = \int \delta(r_1) \delta(r_3 - r_1 - 2) dr_1 = \delta(r_3 - 2),$$

$$\tau_5(ns, b, p, c, r_3) = \int \delta(r_1 - p c) \delta(r_3 - r_1) dr_1 = \delta(r_3 - p c),$$

$$\tau_5(ns, nb, p, c, r_3) = \int \delta(r_1) \delta(r_3 - r_1) dr_1 = \delta(r_3).$$

Next, suppose we delete *Po* next. *Po* is in the domain of π and σ . Let σ_2 denote the discrete potential $(\pi \otimes \sigma)^{-Po}$. Then, σ_2 is shown in Table 2.

Table 2. The details of discrete potential σ_2

σ_2	d	a	f
s	0.35	0.95	0.35
ns	0.65	0.05	0.65

Next, we delete *Rain* (R). R is in the domain of ρ , σ_2 , and χ . Let χ_2 denote the continuous potential $(\rho \otimes \sigma_2 \otimes \chi)^{-R}$ (in (mB) $^{-1}$). Then,

$$\chi_2(s, c) = 0.1225 \chi(d, c) + 0.57 \chi(a, c) + 0.0175 \chi(f, c),$$

$$\chi_2(ns, c) = 0.2275 \chi(d, c) + 0.03 \chi(a, c) + 0.0325 \chi(f, c).$$

Next, we delete *Price* (Pr). Pr is in the domain of ϕ , β , and τ_5 . Let τ_6 denote $(\phi \otimes \beta \otimes \tau_5)^{-Pr}$ (in (m\$) $^{-1}$). Then,

$$\tau_6(s, b, c, r_3) = \int \phi(s, c, p) \beta(b, p) \delta(r_3 - p c - 2) dp,$$

$$\tau_6(s, nb, c, r_3) = \delta(r_3 - 2) \int \phi(s, c, p) \beta(nb, p) dp,$$

$$\tau_6(ns, b, c, r_3) = \int \phi(ns, c, p) \beta(b, p) \delta(r_3 - p c) dp,$$

$$\tau_6(ns, nb, c, r_3) = \delta(r_3) \int \phi(ns, c, p) \beta(nb, p) dp.$$

Next, we delete *Crop* (C). C is in the domain of τ_6 , and χ_2 . Let τ_7 denote $(\tau_6 \otimes \chi_2)^{-C}$ (in (m\$) $^{-1}$). Then,

$$\tau_7(s, b, r_3) = \int \tau_6(s, b, c, r_3) \chi_2(s, c) dc,$$

$$\tau_7(s, nb, r_3) = \int \tau_6(s, nb, c, r_3) \chi_2(s, c) dc,$$

$$\tau_7(ns, b, r_3) = \int \tau_6(ns, b, c, r_3) \chi_2(ns, c) dc,$$

$$\tau_7(ns, nb, r_3) = \int \tau_6(ns, nb, c, r_3) \chi_2(ns, c) dc.$$

Next, we delete *Buy* (B). B is in the domain of τ_7 . Let τ_8 denote τ_7^{-B} (in (m\$) $^{-1}$). Then,

$$\tau_8(s, r_3) = \tau_7(s, b, r_3) + \tau_7(s, nb, r_3),$$

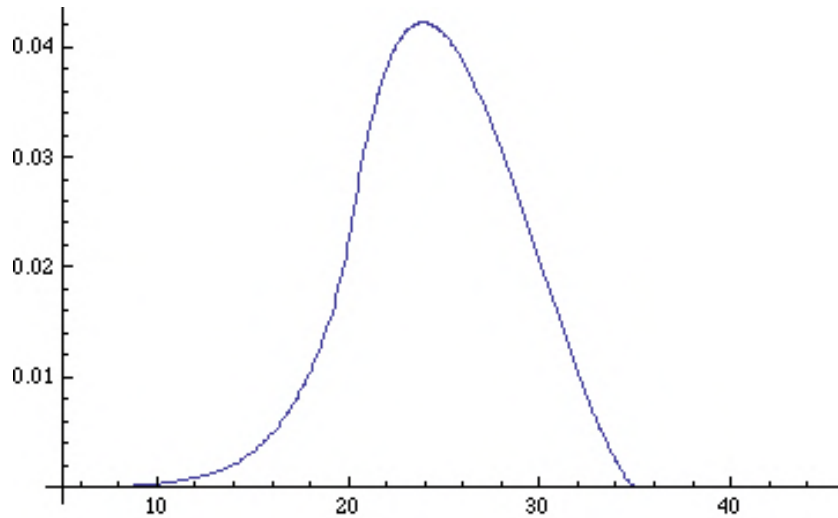
$$\tau_8(ns, r_3) = \tau_7(ns, b, r_3) + \tau_7(ns, nb, r_3).$$

Finally, we delete *Subsidy* (S). S is in the domain of τ_8 . Let τ_9 denote τ_8^{-S} (in (m\$) $^{-1}$). Then,

$$\tau_9(r_3) = \tau_8(s, r_3) + \tau_8(ns, r_3).$$

τ_9 represents the marginal distribution for R_3 . An implementation in Mathematica shows that τ_9 is as follows:

$\tau_9(r_3) = 0.228 \delta(r_3) + 0.263 \delta(r_3 - 2) + 0.509 f(r_3)$ (in (m\$) $^{-1}$), where $f(r_3)$ is a PDF as shown in Figure 7. Thus the marginal distribution of R_3 is a mixed distribution with probability masses of 0.228 at $R_3 = 0$, 0.263 at $R_3 = 2$, and a weighted density function $0.509 f(r_3)$.

Figure 7. The PDF portion of the mixed distribution of marginal for R_3 .

6 Summary and Discussion

We have described an extension of the Shenoy-Shafer architecture for discrete BNs so it applies to hybrid BNs with deterministic variables. We use Dirac delta functions to represent deterministic conditionals of continuous variables. We use discrete and continuous potentials, and we keep track of the units of continuous potentials. Marginalization of discrete variables is done using addition and marginalization of continuous variables is done using integration. We illustrate our architecture by solving some small examples of hybrid BNs. We also solve exactly a modified version of the extended crop problem that has non-conditional linear Gaussian conditionals, and non-linear functions for deterministic variables.

The extended architecture described in this paper is different from the architectures described by Moral *et al.* [2001] and by Cobb and Shenoy [2005a,b]. Moral *et al.* [2001] do not consider deterministic conditionals. Also, they use a restriction operation to incorporate observations of continuous variables. In our framework, this operation is unnecessary. We represent observations of continuous variables by Dirac delta functions, and the restriction operation is equivalent to marginalization of the observed continuous variable. Cobb and Shenoy [2005a] use an equation potential to represent linear deterministic conditionals. This framework is unable to directly represent non-linear deterministic conditionals. Cinicioglu and Shenoy [2009] introduce Dirac delta functions to represent deterministic conditionals. But the framework in Cinicioglu and Shenoy [2009] is designed for describing arc reversals rather than inference. While arc reversals can be used for making inferences in hybrid BNs (see, e.g., [Shachter 1988] for the case of discrete BNs), it is not as computationally efficient as using the extended architecture described in this paper.

We have ignored the computational problem of integrating density potentials. In many cases, e.g., Gaussian density functions, there does not exist a closed form solution of the integral of the Gaussian density functions.

One way around this problem is to use mixtures of truncated exponentials (MTEs) to approximate density functions [Moral *et al.* 2001, Cobb *et al.* 2006]. MTEs are easy to integrate and are closed under combination and marginalization. They are also closed under transformations needed for a one-dimensional linear deterministic functions [Cobb and Shenoy 2005a], but not non-linear ones. One solution for non-linear functions of a single variable is to approximate them by piecewise linear functions [Cobb and Shenoy 2005b]. However, many issues remain unsolved. For example, the family of MTE functions is not closed under transformations needed by linear deterministic functions involving two or more continuous parent variables [Shenoy 2010]. Also, finding an MTE approximation of a high-dimensional conditional (with two or more continuous parent variables) is not easy.

Another way around the problem of integration of density functions is to approximate them using mixtures of polynomials (MOP) [Shenoy and West 2011, Shenoy 2010]. MOP functions are closed under a bigger class of functions for deterministic variables (including linear and quotient functions) than MTE functions. In the extended crop problem discussed in the previous section, we have a product function for one of the deterministic variables, and we can compute a closed form solution for the marginal (although it is not a mixture of polynomials function). The use of MTE and MOP functions for inference in hybrid BNs needs further investigation.

The use of Dirac delta functions for representing deterministic functions is practical only for differentiable functions. If the function is not differentiable, then there doesn't always exist a closed form solution for the integral of such Dirac delta functions. For example, if $Z = \max\{X, Y\}$, then we are unable to compute the marginal of Z even if the densities of X and Y are easily integrable, i.e., there is no closed form solution for the integral:

$$f_Z(z) = \int \int f_X(x) f_{Y|x}(y) \delta(z - \max\{x, y\}) dy dx,$$

where $f_Z(z)$ denotes the marginal PDF of Z , $f_X(x)$ denotes the PDF of X , and $f_{Y|x}(y)$ denotes the conditional PDF of Y given x . However, we can convert the *max* deterministic function to a differentiable one as follows: $Z = X$ if $X \geq Y$, and $Z = Y$ if $X < Y$. We introduce a discrete variable, say A , with two states, a and na , with X and Y as parents, where a denotes that $X \geq Y$, and make A a parent of Z . This hybrid BN can then be solved using the extended Shenoy-Shafer architecture. A solution of this problem for the case where X and Y are independent with normal distributions is described in Shenoy and West [2011].

Acknowledgments

We are grateful to Barry Cobb for many discussions, to the reviewers of ECSQARU-09, and to the two reviewers of this journal. A short version of this paper appeared as Shenoy and West [2009].

Appendix: Properties of the Dirac Delta Function

Some properties of δ are as follows [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, $f(x) \delta(x) = f(0) \delta(x)$. If $f(x)$ is continuous in the neighborhood of 0, then $\int f(x) \delta(x) dx = f(0) \int \delta(x) dx = f(0)$. The range of integration need not be from $-\infty$ to ∞ , but cover any domain containing 0.
- (ii) (*Change of Origin*) $\int \delta(x - a) dx = 1$, and $f(x) \delta(x - a) = f(a) \delta(x - a)$. If $f(x)$ is any function which is continuous in the neighborhood of a , then $\int f(x) \delta(x - a) dx = f(a)$.
- (iii) $\int \delta(x - h(u, v)) \delta(y - g(v, w, x)) dx = \delta(y - g(v, w, h(u, v)))$. This follows from property (ii) of Dirac delta functions if we regard $\delta(y - g(v, w, x))$ as a function of x .
- (iv) (*Rescaling*) If $g(x)$ has real (non-complex) zeros at $a_1, \dots, a_n$, and is differentiable at these points, and $g'(a_i) \neq 0$ for $i = 1, \dots, n$, then $\delta(g(x)) = \sum_i \delta(x - a_i)/|g'(a_i)|$. 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) 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., $(d/dx)H(x) = \delta(x)$. $H(x)$ 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 $\sigma \rightarrow 0$). Then, the generalized derivative of $H(x)$ is the limit of the derivative of these functions.
- (vi) 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. To show the validity of this formula, let $F_Y(y)$ denote the cumulative distribution function of Y . Then, $F_Y(y) =$

$P(g(X) \leq y) = \int f_X(x) H(y - g(x)) dx$, where $H(\cdot)$ is the Heaviside function defined in (vii). Then, $f_Y(y) = (d/dy)(F_Y(y)) = \int f_X(x) (d/dy)(H(y - g(x))) dx = \int f_X(x) \delta(y - g(x)) dx$.

(vii) 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 $f_Y(y) = \int f_X(x) \delta(y - g(x)) dx = |(d/dy)(g^{-1}(y))| \int f_X(x) \delta(x - g^{-1}(y)) dx = |(d/dy)(g^{-1}(y))| f_X(g^{-1}(y))$.

(viii) The definition of δ can be extended to R^n , the n -dimensional Euclidean space. Thus, if $\mathbf{x} \in R^n$, $\delta(\mathbf{x}) = 0$ if $\mathbf{x} \neq \mathbf{0}$, and $\int \dots \int \delta(\mathbf{x}) d\mathbf{x} = 1$, where $d\mathbf{x} = dx_1 \dots dx_n$. Thus, e.g., $\int \dots \int f(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} = f(\mathbf{x}_0)$.

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

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

References

- Cinicioglu, E. N. and P. P. Shenoy (2009), "Arc reversals in hybrid Bayesian networks with deterministic variables," *International Journal of Approximate Reasoning*, **50**(5), 763–777.
- Cobb, B. R. and P. P. Shenoy (2005a), "Hybrid Bayesian networks with linear deterministic variables," *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-First Conference (UAI-05)*, F. Bacchus and T. Jaakkola (eds.), Corvallis, OR, AUAI Publishers: 136–144.
- 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), "Inference in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials," *International Journal of Approximate Reasoning*, **41**(3), 2006, 257–286.
- Cobb, B. R. and P. P. Shenoy (2008), "Decision making with hybrid influence diagrams using mixtures of truncated exponentials," *European Journal of Operational Research*, **186**(1), 261–275.

- 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.
- Howard, R. A. and J. E. Matheson (1984), “Influence diagrams,” in *The Principles and Applications of Decision Analysis*, R. A. Howard and J. E. Matheson (eds.), Menlo Park, CA, Strategic Decisions Group, **2**: 719–762.
- 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, Department of Engineering-Economic Systems, Stanford University, Stanford, CA.
- 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.
- Langseth, H., T. D. Nielsen, R. Rumi, and A. Salmeron (2010), “Parameter estimation and model selection for mixtures of truncated exponentials,” *International Journal of Approximate Reasoning*, **51**(5), 485–498.
- Lauritzen, S. L. and D. J. Spiegelhalter (1988), “Local computations with probabilities on graphical structures and their application to expert systems (with discussion),” *Journal of Royal Statistical Society, Series B*, **50**(2): 157–224.
- Lauritzen, S. L. and F. Jensen (2001), “Stable local computation with conditional Gaussian distributions,” *Statistics and Computing*, **11**: 191–203.
- Lerner, U. N. (2002), “Hybrid Bayesian networks for reasoning about complex systems,” PhD thesis, Department of Computer Science, Stanford University.
- Lerner, U. N. and R. Parr (2001), “Inference in hybrid networks: Theoretical limits and practical algorithms,” in *Uncertainty in Artificial Intelligence: Proceedings of the Seventeenth Conference (UAI-01)*, J. Breese and D. Koller (eds.), Morgan Kaufmann, San Francisco, CA: 310–318.
- Li, Y. and P. P. Shenoy (2010), “Solving hybrid influence diagrams with deterministic variables,” in *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-sixth Conference*, P. Grünwald and P. Spirtes (eds.), AUAI Press, Corvallis, OR: 322–331.
- Madsen, A. L. (2008), “Belief update in CLG Bayesian networks with lazy propagation,” *International Journal of Approximate Reasoning*, **49**(2), 503–521.

- 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.
- Murphy, K. (1999), "A variational approximation for Bayesian networks with discrete and continuous latent variables," in *Uncertainty in Artificial Intelligence: Proceedings of the Fifteenth Conference* (UAI-99), K. B. Laskey and H. Prade (eds.), Morgan Kaufmann, San Francisco, CA: 457–466.
- Nielsen, T. D. and F. V. Jensen (2007), "On-line alert systems for production plants: A conflict based approach," *International Journal of Approximate Reasoning*, **45**(2), 255–270.
- Olmsted, S. M. (1983), "On representing and solving decision problems," PhD dissertation, Department of Engineering-Economic Systems, Stanford University, Stanford, CA.
- Pearl, J. (1986), "Fusion, propagation and structuring in belief networks," *Artificial Intelligence*, **29**: 241–288.
- Poland III, W. B. (1994), "Decision analysis with continuous and discrete variables: A mixture distribution approach," PhD dissertation, Department of Engineering-Economic Systems, Stanford University, Stanford, CA.
- Romero, V., R. Rumi, and A. Salmeron (2007), "Learning hybrid Bayesian networks using mixtures of truncated exponentials," *International Journal of Approximate Reasoning*, **42**(1–2), 54–68.
- Rumi, R. and A. Salmeron (2007), "Approximate probability propagation with mixtures of truncated exponentials," *International Journal of Approximate Reasoning*, **45**(2), 191–200.
- 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**(6): 871– 882.
- Shachter, R. D. (1988), "Probabilistic inference and influence diagrams," *Operations Research*, **36**(4), 589–604.
- Shachter, R. D. and C. R. Kenley (1989), "Gaussian influence diagrams," *Management Science*, **35**(5): 527–550.
- Shenoy, P. P. (1992), "Valuation-based systems for Bayesian decision analysis," *Operations Research*, **40**(3): 463–484.
- Shenoy, P. P. (1997), "Binary join trees for computing marginals in the Shenoy-Shafer architecture," *International Journal of Approximate Reasoning*, **17**(2–3), 239–263.
- Shenoy, P. P. (2006), "Inference in hybrid Bayesian networks using mixtures of Gaussians," in *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Second Conference* (UAI-06), R. Dechter and T. Richardson (eds.), Corvallis, OR, AUAI Press: 428–436.

- Shenoy, P. P. (2010), “Some issues in using mixtures of polynomials for inference in hybrid Bayesian networks,” Working Paper No. 323, School of Business, University of Kansas, Lawrence, KS.
- Shenoy, P. P. and G. Shafer (1990), “Axioms for probability and belief-function propagation,” in *Uncertainty in Artificial Intelligence (UAI-88)*, R. D. Shachter, T. S. Levitt, J. F. Lemmer and L. N. Kanal (eds.), Amsterdam, North-Holland, **4**:169–198.
- Shenoy, P. P. and J. C. West (2009), “Inference in hybrid Bayesian networks with deterministic variables,” in *Symbolic and Quantitative Approaches to Reasoning under Uncertainty—10th ECSQARU*, C. Sossai and G. Chemello (eds.), Berlin, Springer-Verlag, Lecture Notes in Artificial Intelligence, **5590**: 46–58.
- Shenoy, P. P. and J. C. West (2011), “Inference in hybrid Bayesian networks using mixtures of polynomials,” *International Journal of Approximate*, in press, <<http://www.ku.edu/home/pshenoy/Papers/IJAR11b.pdf>>.