SCHOOL OF BUSINESS WORKING PAPER NO. 314

Propagation in Hybrid Bayesian Networks with Linear Deterministic Variables

Barry R. Cobb[1] Prakash P. Shenoy[2]
cobbbr@vmi.edu pshenoy@ku.edu

[1]Department of Economics and Business
Virginia Military Institute
Lexington, VA 24450

[2]University of Kansas School of Business
1300 Sunnyside Ave., Summerfield Hall
Lawrence, KS 66045–7585

April 21, 2007

Abstract

This paper extend exacts inference for hybrid Bayesian networks to allow continuous variables with any conditional density functions, discrete variables with continuous parents, and conditionally deterministic continuous variables that are linearly dependent on their continuous parents. We introduce a mixed distribution representation of potentials and derive operations from the method of convolutions in probability theory to determine distributions for linear functions of random variables. Mixtures of truncated exponentials (MTE) potentials are used to approximate probability density functions in the representation so that probability density functions can be easily marginalized in closed form. The Shenoy-Shafer architecture is used to calculate marginals and variables can be marginalized in any order using any join tree structure.

Key Words: Hybrid Bayesian networks, MTE potentials, Shenoy-Shafer architecture, CLG models, deterministic variables

Contents

1 Introduction

Bayesian networks model knowledge about propositions in uncertain domains using graphical and numerical representations (Spiegelhalter *et al.* 1993). At the qualitative level, a Bayesian network is a directed acyclic graph where nodes represent variables and the (missing) edges represent conditional independence relations among the variables. At the numerical level, a Bayesian network consists of a factorization of a joint probability distribution into a set of conditional distributions, one for each variable in the network. Hybrid Bayesian networks contain both discrete and continuous probability distributions in the numerical representation.

An important class of Bayesian networks with discrete and continuous variables are those that have conditionally deterministic variables (a variable that is a deterministic function of its parents). Conditional linear Gaussian (CLG) models (Lauritzen 1992, Lauritzen and Jensen 2001) can handle such cases when continuous variables have a multi-variate Gaussian distribution and discrete nodes do not have continuous parents. In models with non-Gaussian distributions and deterministic variables, Monte Carlo methods may be required to obtain an approximate solution. General purpose solution algorithms, e.g., the Shenoy-Shafer architecture, have not been adapted to such models, primarily because the joint density for the variables in models with deterministic variables does not exist and these methods involve propagation of probability densities.

An example of a Bayesian network representation of a conditionally deterministic variable is given below.

1.1 Example

Consider a real-valued random variable Z_2 with a mixed distribution where $P(Z_2 = 1) = 0.5$, $P(Z_2 = 2) = 0.3$, and which has probability density $0.2 \cdot \phi(z_2)$, where $\phi(z_2)$ is a normal probability density function (PDF) with mean 3 and variance 1, i.e. $N(3, 1)$. We can represent this mixed distribution by a *mixed potential* ζ_3 for Z_2 as follows:

$$\zeta_3(z_2) = (1, (0.5 \otimes [z_2 = 1])(z_2) + (0.3 \otimes [z_2 = 2])(z_2) + 0.2 \cdot \phi_2(z_2)) \quad .$$

We can verify ζ_3 is a probability distribution for Z_2 by calculating: $0.5 + 0.3 + \int_{\Omega_{Z_2}} 0.2 \cdot \phi(z_2)\, dz_2 = 1$. The expected value and variance of this mixed distribution are calculated as

$$E(Z_2) = 0.5 \cdot 1 + 0.3 \cdot 2 + \int_{\Omega_{Z_2}} z_2 \cdot 0.2 \cdot \phi(z_2)\, dz_2 = 1.7 \ ,$$

$$Var(Z_2) = 0.5 \cdot (1 - E(Z_2))^2 + 0.3 \cdot (2 - E(Z_2))^2 + \int_{\Omega_{Z_2}} (z_2 - E(Z_2))^2 \cdot 0.2 \cdot \phi(z_2) = 0.81 \ .$$

This mixed distribution is shown graphically in the left panel of Figure 1.

Mixed distributions result from the computation of marginal distributions for variables in hybrid Bayesian networks, such as the network shown in the right panel of Figure 1. In this example, Y_1 is a discrete variable with state space $\Omega_{Y_1} = \{A, B, C\}$, and Z_1 and Z_2 are real-valued variables. Mixed potentials are assigned to each variable in the network. Let θ_1 be a potential for Y_1, let ζ_1 be a potential for Z_1, and let ζ_2 be a potential for $\{Y_1, Z_1, Z_2\}$ as follows:

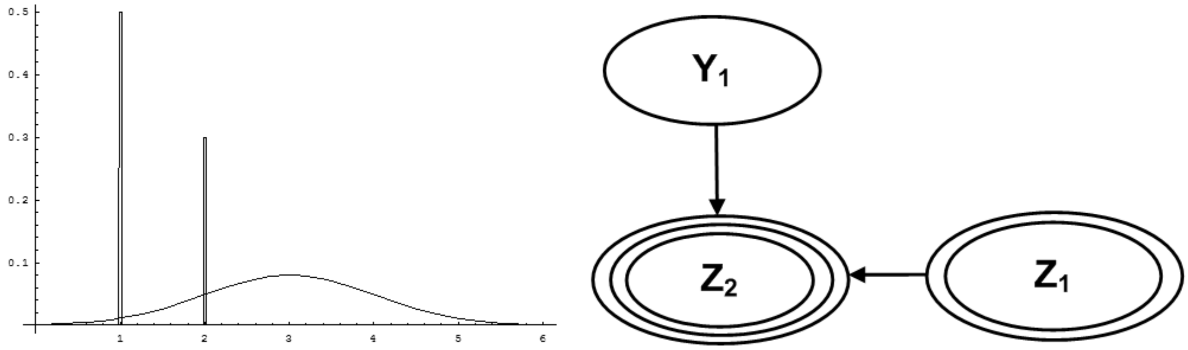


Figure 1: A Graphical (left) and Bayesian network (right) representation of the mixed distribution for Z_2.

$$\begin{aligned}
&\theta_1(Y_1 = A) = (0.5, \iota),\ \ \theta_1(Y_1 = B) = (0.3, \iota),\ \ \theta_1(Y_1 = C) = (0.2, \iota)\ \ ,\\
&\zeta_1(z_1) = (1, \phi_1(z_1))\ \ ,\\
&\zeta_2(z_1, z_2, Y_1 = A) = (1, [0z_1 + z_2 = 1]),\ \ \zeta_2(z_1, z_2, Y_1 = B) = (1, [0z_1 + z_2 = 2]),\\
&\qquad\qquad \zeta_2(z_1, z_2, Y_1 = C) = (1, [-z_1 + z_2 = 0])\ \ .
\end{aligned}$$

The first component of the value of a mixed potential is called a mass part and the second component is called a density part. The symbol ι represents a vacuous density or an absence of a probability density. The potential θ_1 represents the probability mass function (PMF) of Y_1, the potential ζ_1 represents the PDF of Z, and the potential ζ_2 represents the conditional distribution of Z_2 given Y_1 and Z_1. For instance, $\zeta_2(z_1, z_2, Y_1 = A)$ denotes that $Z_2 \mid \{z_1, Y_1 = A\} = 1$ with probability 1, whereas $\zeta_2(z_1, z_2, Y_1 = C)$ denotes that the conditional density function for $Z_2 \mid \{z_1, Y_1 = C\}$ is a degenerate linear function of z_1. Removing Y_1 and Z_1 from the combination of θ_1, ζ_1, and ζ_2 (an operation denoted as $(\theta_1 \otimes \zeta_1 \otimes \zeta_2)^{-\{Y_1, Z_1\}}$) results in the potential defined previously for Z_2. The mechanics of computing this marginal are described later in this paper.

Potentials of the type described above are used in this paper to represent distributions of variables in hybrid Bayesian networks where continuous variables are (possibly) linear deterministic functions of their continuous parents. This paper develops inference methods using mixed potentials that allow inference in hybrid Bayesian networks without placing limitations on the types of continuous distributions allowed and without limitations on the placement of continuous and discrete variables. We use the Shenoy-Shafer architecture to compute marginals and we can use any join tree to do the propagation.

In the following section, we briefly define notation that will be used throughout the remainder of the paper.

1.2 Notation

Random variables in a hybrid Bayesian network will be denoted by capital letters, e.g., A, B, C. Sets of variables will be denoted by boldface capital letters, $\mathbf{Y}$ if all variables are discrete, $\mathbf{Z}$ if all variables are continuous, or $\mathbf{X}$ if some of the components are discrete and some are continuous. If $\mathbf{X}$ is a set of variables, $\mathbf{x}$ is a configuration of specific states of those variables. The discrete, continuous, or mixed state space of $\mathbf{X}$ is denoted by $\Omega_{\mathbf{X}}$.

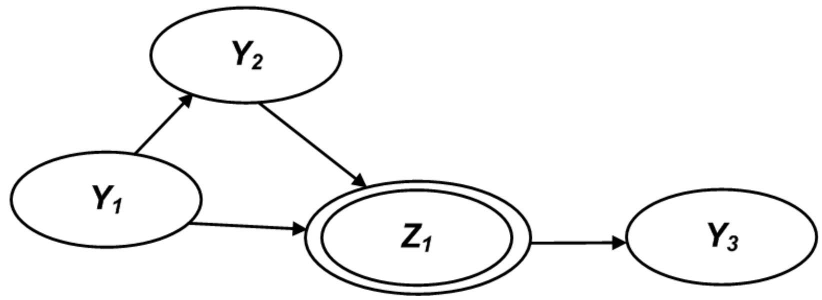


Figure 2: The Bayesian network described in Example 1.

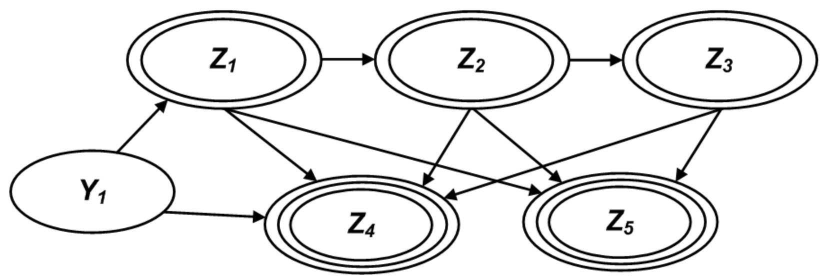


Figure 3: The Bayesian network described in Example 1.2.

Probability potentials and mixed potentials are denoted by lower-case greek letters, e.g., α, β, γ, possibly with subscripts as an index. Deterministic potentials are denoted by $\mathcal{M}$, with components ω for a constant and $\mathbf{M}$ for a matrix of coefficients, possibly with subscripts as an index. Values of potentials for countable elements in the state space of a set of variables may be denoted by either subscripts or as an argument to a function, e.g., $\alpha_{1,A} = \alpha_1(Y_1 = A) = P(Y_1 = A)$, $\phi_{1,A}(z_1) = \phi(z_1, Y_1 = A) = f_{Z1|\{Y_1=A\}}(z_1)$, etc.

In graphical representations, continuous variables in hybrid Bayesian networks are represented by double-border ovals, whereas discrete variables are represented by single-border ovals. Variables that are conditionally deterministic functions of their parents are represented by triple-border ovals. A variable is discrete if its state space is countable, and continuous otherwise.

Example 1 Consider the Bayesian network in Figure 2. In this model, Y_1, Y_2, and Y_3 are discrete variables with $\Omega_{Y_1} = \Omega_{Y_2} = \Omega_{Y_3} = \{A, B\}$. The variable Z_1 is a continuous variable whose state space is the set of real numbers.

Example 2 Consider the Bayesian network in Figure 3. In this model, Y_1 is a discrete variable with $\Omega_{Y_1} = \{A, B\}$. The variables $Z_1, \ldots, Z_5$ are continuous variables whose state spaces are the set of real numbers. The variables Z_4 and Z_5 are conditionally deterministic given values of Z_1, Z_2, and Z_3.

This paper expands the notation and operations first described by Cobb and Shenoy (2005b, 2006b) into a complete propagation algorithm for hybrid Bayesian networks. The remainder of this paper is organized as follows. Section 2 defines probability potentials, including mass and density potentials. Section 3 defines combination and marginalization operations for probability potentials and gives an example of propagation of these potentials in a hybrid Bayesian network. Section 4 defines deterministic potentials, which define deterministic relationships between real-valued variables, and outlines combination and marginalization operations for these potentials. Section 5 defines mixed potentials, which can accomodate the distributions of conditionally deterministic and mixed random variables, and outlines combination and marginalization operations for these potentials. Section 6 uses the operations described in previous sections to solve a hybrid Bayesian network. Section 7 concludes the paper. In Sections 4 and 5 we demonstrate that the operations for deterministic and mixed potentials satisfy the Shenoy-Shafer axioms. All proofs appear in the appendix.

2 Probability Potentials

This section describes mass and density potentials (either of which may be termed a probability potential) and gives examples of propagation of these potentials in a hybrid Bayesian network.

2.1 Mass Potentials

Definition 1 *A* **mass potential** α *for* $\mathbf{X}$ *assigns a positive real number to some countable elements in* $\Omega_{\mathbf{X}}$.

Since $\mathbf{X}$ has some continuous variables, $\Omega_{\mathbf{X}}$ has an uncountable number of states and it is only possible to list the values for the countable number of states that have positive mass.

Example 3 In the Bayesian network of Example 1, $P(Y_1 = A) = 0.7$ and $P(Y_1 = B) = 0.3$, which is defined by the mass potential α_1 where $\alpha_1(Y_1 = A) = 0.7$ and $\alpha_1(Y_1 = B) = 0.3$. The mass potential α_2 for $\{Y_1, Y_2\}$ has values $\alpha_2(Y_1 = A, Y_2 = A) = 0.6$, $\alpha_2(Y_1 = A, Y_2 = B) = 0.4$, $\alpha_2(Y_1 = B, Y_2 = A) = 0.2$, and $\alpha_2(Y_1 = B, Y_2 = B) = 0.8$ which represent conditional probabilities; for instance, $\alpha_2(Y_1 = A, Y_2 = A) = P(Y_2 = A \mid Y_1 = A)$.

Example 4 Consider the variable Y_3 from Example 1 with continuous parent Z_1. Values for a conditional mass potential γ for Y_3 given Z_1 are as follows:

$$\gamma_{A|\{-\infty<z_1\leq 0\}}(z_1) = P(Y_3 = A \mid z_1) = 0.2\ ,\ \gamma_{B|\{-\infty<z_1\leq 0\}}(z_1) = P(Y_3 = B \mid z_1) = 0.8\ ,$$
$$\gamma_{A|\{0<z_1<\infty\}}(z_1) = P(Y_3 = A \mid z_1) = 0.7\ ,\ \gamma_{B|\{0<z_1<\infty\}}(z_1) = P(Y_3 = B \mid z_1) = 0.3\ .$$

2.2 Density Potentials

Definition 2 *A* **density potential** ϕ *for* $\mathbf{Z}$ *is a function* $\phi : \Omega_{\mathbf{Z}} \to \mathcal{R}^+$, *where* $\mathcal{R}^+$ *denotes the set of non-negative real numbers.*

Example 5 In the Bayesian network in Figure 2, Z_1 is a normal random variable whose mean and variance depends on the values of its discrete parents Y_1 and Y_2 as follows:

$$Z_1 \mid \{Y_1 = A, Y_2 = A\} \sim N(0,1)\ \ , Z_1 \mid \{Y_1 = A, Y_2 = B\} \sim N(1,4)$$
$$Z_1 \mid \{Y_1 = B, Y_2 = A\} \sim N(4,4)\ \ , Z_1 \mid \{Y_1 = B, Y_2 = B\} \sim N(4,16)\ \ .$$

The functions $\varphi_{A,A}$, $\varphi_{A,B}$, $\varphi_{B,A}$, and $\varphi_{B,B}$ constitute the density potential φ for $\{Y_1, Y_2, Z_1\}$. Specific values of these functions are denoted by, e.g., $\varphi_{A,A}(z_1)$.

2.3 Mixtures of Truncated Exponentials (MTE) Potentials

Definition 3 *A* **mixture of truncated exponentials (MTE) potential** (Moral *et al.* 2001, Rumí 2003) *has the following definition.*

Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ *be an n-dimensional random variable. A function* $\phi : \Omega_{\mathbf{Z}} \mapsto \mathcal{R}^+$ *is an MTE potential if one of the next two conditions holds:*

1. *The potential* ϕ *can be written as*

$$\phi(\mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \ exp\left\{ \sum_{j=1}^{n} b_i^{(j)} z_j \right\} \tag{1}$$

for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$, *where* $a_i, i = 0, \ldots, m$ *and* $b_i^{(j)}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ *are real numbers.*

2. *The domain of the variables,* $\Omega_{\mathbf{Z}}$, *is partitioned into hypercubes* $\{\Omega_{\mathbf{Z}_1}, \ldots, \Omega_{\mathbf{Z}_k}\}$ *such that* ϕ *is defined as*

$$\phi(\mathbf{z}) = \phi_i(\mathbf{z}) \qquad \textit{if } \mathbf{z} \in \Omega_{\mathbf{Z}_i} \ , \ \ i = 1, \ldots, k \ ,$$

where each $\phi_i, i = 1, ..., k$ *can be written in the form of equation (1) (i.e. each* ϕ_i *is an MTE potential on* $\Omega_{\mathbf{Z}_i}$*).*

In the definition above, k is the number of *pieces* and m is the number of exponential *terms* in each piece of the MTE density potential. In this paper, all MTE density potentials are equal to zero in unspecified regions. Since the terms in the exponent are a linear function of the arguments, MTE potentials can be easily marginalized in closed form (without having to do computationally expensive numerical integration).

Example 6 A 2-piece, 3-term un-normalized MTE density potential $\phi'_{A,A}$ which approximates the standard normal PDF is defined (Cobb and Shenoy 2006a) as

$$\phi'_{A,A}(z_1) = \begin{cases} -0.0105643 + 197.0557202 \ \exp\{2.2568434 z_1\} & \\ -461.4392506 \ \exp\{2.3434117 z_1\} & \\ +264.7930371 \ \exp\{2.4043270 z_1\} & \text{if } -3 \leq z_1 < 0 \\ & \\ -0.0105643 + 197.0557202 \ \exp\{-2.2568434 z_1\} & \\ -461.4392506 \ \exp\{-2.3434117 z_1\} & \\ +264.7930371 \ \exp\{-2.4043270 z_1\} & \text{if } 0 \leq z_1 \leq 3 \ . \end{cases}$$

A normalized version of the 2-piece, 3-term MTE approximation to the $N(0, 1)$ PDF has values

$$\phi_{A,A}(z_1) = (1/0.997306) \cdot \phi'_{A,A}(z_1) \ .$$

The PDF $\phi_{A,A}$ approximates the PDF $\varphi_{A,A}$ for $\{Y_1, Y_2, Z_1\}$ in the Bayesian network of Figure 2 for the case where $Y_1 = A$ and $Y_2 = A$. Similar MTE density potentials $\phi_{A,B}$, $\phi_{B,A}$, and $\phi_{B,B}$ can also be used to approximate the remainder of the component potentials in the density potential φ. These potentials are shown graphically in the left panel of Figure 4, overlayed on the corresponding normal PDF's.

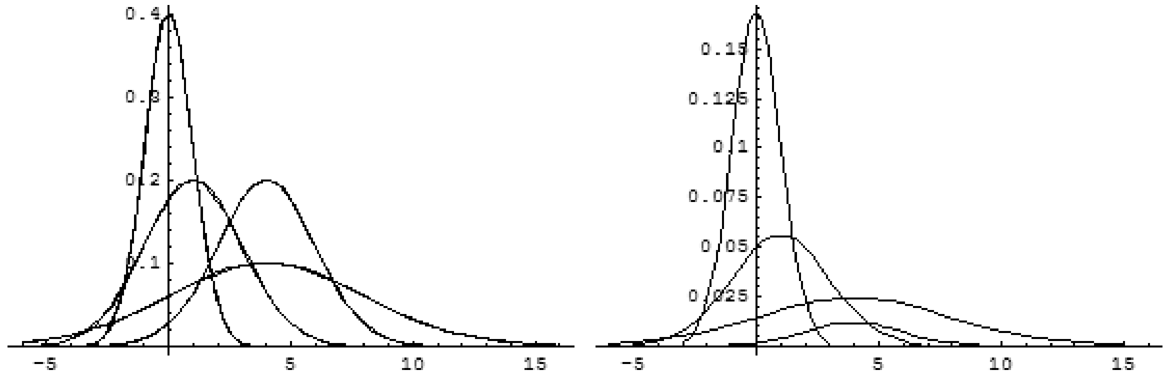


Figure 4: The MTE approximations to the normal PDF's for Z_1 given $\{Y_1, Y_2\}$ (left) and the components of the density potential ψ resulting from the combination in Example 8 (right) from the Bayesian network from Example 1.

While the example above uses an MTE potential to approximate a density potential, MTE potentials can also be used to model mass potentials for discrete variables with continuous variables. For example, Cobb and Shenoy (2006a) define an MTE approximation to the binary sigmoid function.

3 Operations with Probability Potentials

This section defines the operations that are used to perform propagation of mass and density potentials in hybrid Bayesian networks. Definitions are preceded by one or more examples that illustrate the operation.

3.1 Combination

Combination of probability potentials is pointwise multiplication of functions.

Example 7 The combination of the mass potentials α_1 and α_2 from Example 3, denoted by $\alpha_3 = (\alpha_1 \otimes \alpha_2)$, is a mass potential with values $\alpha_3(Y_1 = A, Y_2 = A) = 0.42$, $\alpha_3(Y_1 = A, Y_2 = B) = 0.28$, $\alpha_3(Y_1 = B, Y_2 = A) = 0.06$, and $\alpha_3(Y_1 = B, Y_2 = B) = 0.24$.

Example 8 The combination of the mass potentials α_3 from Example 7 and the density potential ϕ formed from the functions $\phi_{A,A}$, $\phi_{A,B}$, $\phi_{B,A}$, and $\phi_{B,B}$ from Example 6, denoted by $\psi = (\alpha_3 \otimes \phi)$, is a density potential with values calculated as

$$\psi_{A,A}(z_1) = \alpha_3(Y_1 = A, Y_2 = A) \cdot \phi_{A,A}(z_1) \ , \psi_{A,B}(z_1) = \alpha_3(Y_1 = A, Y_2 = B) \cdot \phi_{A,B}(z_1) \ ,$$
$$\psi_{B,A}(z_1) = \alpha_3(Y_1 = B, Y_2 = A) \cdot \phi_{B,A}(z_1) \ , \psi_{B,B}(z_1) = \alpha_3(Y_2 = B, Y_2 = B) \cdot \phi_{B,B}(z_1) \ .$$

These four components of the density potential ψ are shown graphically in the right panel of Figure 4.

Definition 4 (Combination of Probability Potentials) *If φ_1 and φ_2 are mass or density potentials for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, then $\varphi_1 \otimes \varphi_2$ is a potential for $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2$ defined as follows:*

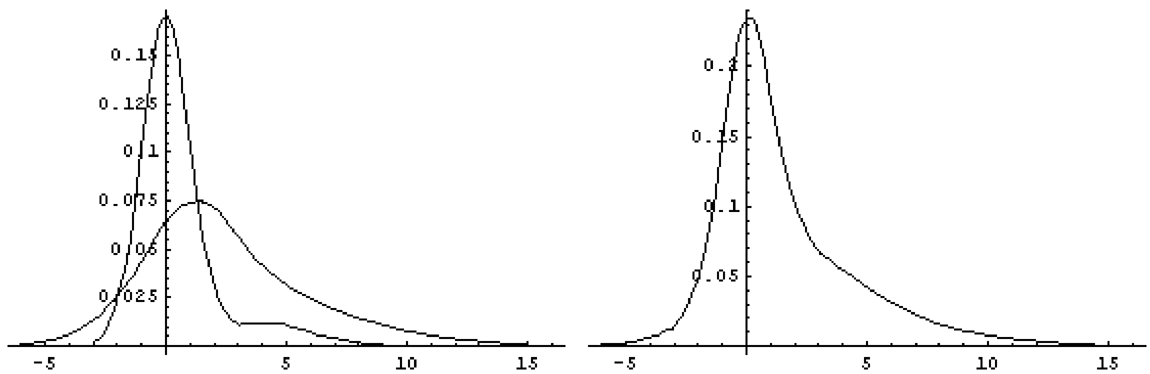


Figure 5: The components of the density potentials ϑ for $\{Y_2, Z_1\}$ (left) and ξ for Z_1 (right) resulting from the marginalization operations in Example 9.

$$(\varphi_1 \otimes \varphi_2)(\mathbf{x}) = \varphi_1\left(\mathbf{x}^{\downarrow \mathbf{X}_1}\right) \cdot \varphi_2\left(\mathbf{x}^{\downarrow \mathbf{X}_2}\right) \quad \textit{for all } \mathbf{x} \in \Omega_{\mathbf{X}} \ .$$

If φ_1 and φ_2 are both mass potentials, the result of the combination is a mass potential; otherwise, the result is a density potential. If φ_1 and φ_2 are both MTE mass and/or density potentials, the result of the combination operation is an MTE potential (Moral *et al.* 2001).

3.2 Marginalization

Example 9 The marginalization of discrete variable Y_1 from the potential ψ for $\{Y_1, Y_2, Z_1\}$ from Example 8, denoted by $\vartheta = \psi^{-Y_1}$, is a density potential with values calculated as

$$\vartheta_A(z_1) = \psi_A^{-Y_1}(z_1) = \psi_{A,A}(z_1) + \psi_{B,A}(z_1) \quad \text{and} \quad \vartheta_B(z_1) = \psi_B^{-Y_1}(z_1) = \psi_{B,A}(z_1) + \psi_{B,B}(z_1) \ .$$

The components of the density potential ϑ for $\{Y_2, Z_1\}$ are shown graphically in the left panel of Figure 5. Removing Y_2 from the potential ϑ by calculating $\xi(z_1) = \vartheta_B^{-Y_2}(z_1) = \vartheta_A(z_1) + \vartheta_B(z_1)$ yields the marginal distribution for Z_1, which is shown in the right panel of Figure 5. We can also calculate the marginal distribution for Y_2 from the density potential ϑ by integrating over the state space of Z_1 as

$$\alpha_4(Y_2 = A) = \int_{\Omega_{Z_1}} \vartheta_A(z_1)\, dz_1 = 0.48 \quad \text{and} \quad \alpha_4(Y_2 = B) = \int_{\Omega_{Z_1}} \vartheta_B(z_1)\, dz_1 = 0.52 \ .$$

Marginalization of a set of variables from a mass or density potential corresponds to summing over discrete variables and integrating over continuous variables.

Definition 5 (Marginalization of Probability Potentials) *Let φ be a mass or density potential for $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$. The marginal of φ for a set of variables $\mathbf{X}' = \mathbf{Y}' \cup \mathbf{Z}' \subseteq \mathbf{X}$ is a potential computed as*

$$\varphi_{\mathbf{y}'}^{\downarrow \mathbf{X}'}(\mathbf{z}') = \sum_{\mathbf{y} \in \Omega_{\mathbf{Y} \setminus \mathbf{Y}'}} \left(\int_{\Omega_{\mathbf{Z}''}} \varphi_{\mathbf{y}}(\mathbf{z})\, d\mathbf{z}'' \right) \tag{2}$$

where $\mathbf{z} = (\mathbf{z}', \mathbf{z}'')$, and $(\mathbf{y}', \mathbf{z}') \in \Omega_{\mathbf{X}'}$.

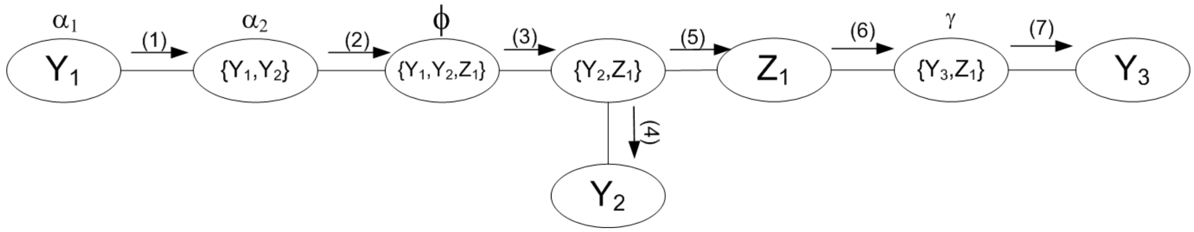


Figure 6: A join tree for the Bayesian network in Figure 2.

Although we show the continuous variables being marginalized before the discrete variables in (2), the variables can be marginalized in any sequence, resulting in the same final potential. The definition above allows for discrete variables whose probability values are given as a function of continuous or discrete parents. If φ is an MTE potential, the result of the operation above is an MTE potential (Moral *et al.* 2001).

3.3 Propagation in Hybrid Bayesian Networks

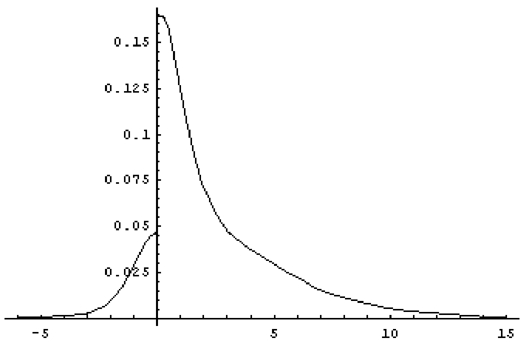
A Bayesian network model can be used to calculate marginal probability distributions for variables of interest by message passing in a join tree structure. To calculate marginal probability distributions in a hybrid Bayesian network where probability potentials are approximated by MTE potentials, we utilize the Shenoy-Shafer architecture (Shenoy and Shafer 1990), since only combination and marginalization operations are required and the class of MTE potentials is closed under these operations.

The Shenoy-Shafer architecture relies on three axioms—consonance of marginalization, commutativity and associativity of combination, and distributivity of marginalization over combination—that enable efficient local computation of marginals of the joint distribution of variables in a Bayesian network. To complete the algorithm, each node in the join tree sends a message to each of its neighbors that is the combination of its own potential and all incoming messages—except the message from the receiving node—followed by marginalization to the intersection with the receiving node. The normalized combination of a node's own potential and all incoming messages is the posterior distribution of the variables in the node conditioned on the evidence.

Example 10 A join tree constructed from the Bayesian network in Figure 2 is shown in Figure 6. The messages required to calculate marginal probability distributions for the four variables are numbered in the join tree above the arrows between the nodes.

The potential α_1 is sent as Message 1, then combined with the potential α_2 at node $\{Y_1, Y_2\}$ (the details are shown in Example 7) and sent to node $\{Y_1, Y_2, Z_1\}$ as Message 2. Message 3 sends the potential ψ resulting from the marginalization operation in Example 9. Messages 4 and 5 are the marginal distributions for Y_2 and Z_1, respectively.

At node $\{Y_3, Z_1\}$, the potential ξ (the marginal distribution for Z_1) is combined with the potential γ described in Example 4, an operation denoted by $\chi = (\gamma \otimes \xi)$. The result is a density potential with components χ_A and χ_B, which correspond to the cases $Y_3 = A$ and $Y_3 = B$. These potentials are shown graphically in Figure 7. Message 7 removes Z_1 from χ to calculate the marginal probability distribution for Y_3 as follows:



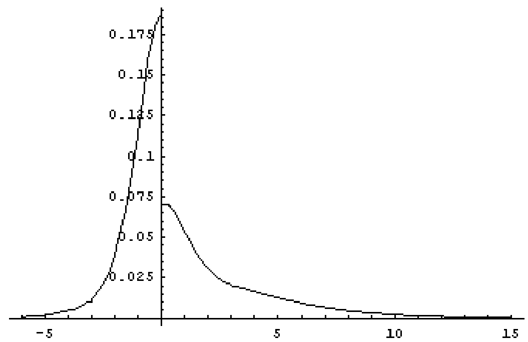


Figure 7: The result of the combination of the potentials γ and ξ in Example 10 for the cases where $Y_3 = A$ (left) and $Y_3 = B$ (right).

$$\alpha_5(Y_3 = A) = \int_{\Omega_{Z_1}} \chi_A(z_1)\, dz_1 = 0.5325 \quad \text{and} \quad \alpha_5(Y_3 = B) = \int_{\Omega_{Z_1}} \chi_B(z_1)\, dz_1 = 0.4675 \ .$$

The preceding example demonstrates how MTE potentials can be used in hybrid Bayesian networks to calculate marginal probability distributions. If evidence is available for a variable, a restriction (or substitution) operation can be used to instantiate the evidence and calculate posterior marginal distributions considering the evidence (for details, see (Cobb and Shenoy 2006a)).

The algorithm demonstrated above relies on the existence of the joint density function for the continuous variables. In cases where a Bayesian network has conditionally deterministic variables with continuous parents, the joint density function for the continuous variables does not exist. The next section describes a deterministic potential representation for conditionally deterministic variables.

4 Deterministic Potentials

This section describes deterministic potentials, which are used to express conditionally deterministic relationships between real-valued variables, and defines combination and marginalization operations needed for propagation with these potentials. Each definition is preceded by one or more illustrative examples.

4.1 Definition

Example 11 In the Bayesian network of Example 3, $Z_4 = 3Z_1 + 5Z_2 + 4Z_3$ if $Y_1 = A$, $Z_4 = 3Z_1 + Z_2 + 4Z_3$ if $Y_1 = B$, and $Z_5 = 10Z_1 + 2Z_2 + 3Z_3$. These conditionally deterministic relationships are represented by the deterministic potentials $\mathcal{M}_{1,A}$ and $\mathcal{M}_{1,B}$ for $\{Z_1, Z_2, Z_3, Z_4\}$, and $\mathcal{M}_2$ for $\{Z_1, Z_2, Z_3, Z_5\}$, where, for example,

$$\mathcal{M}_{1,A} = \omega_{1,A} \cdot \mathbf{M}_{1,A} = 1 \cdot \left[\begin{array}{cccc|c} z_1 & z_2 & z_3 & z_4 & \\ \hline -3 & -5 & -4 & 1 & 0 \end{array} \right] \quad \text{and}$$

$$\mathcal{M}_2 = \omega_2 \cdot \mathbf{M}_2 = 1 \cdot \left[\begin{array}{cccc|c} z_1 & z_2 & z_3 & z_5 & \\ \hline -10 & -2 & -3 & 1 & 0 \end{array} \right] \ .$$

Definition 6 *A* **deterministic potential** *describes the linear deterministic relationship(s) between a set of variables* $\mathbf{Z} = \{Z_1, \ldots, Z_n\}$*. A deterministic potential* $\mathcal{M}$ *for* $\mathbf{Z}$ *is comprised of a constant* ω *and a matrix of real numbers* $\mathbf{M}$ *as follows:*

$$\mathcal{M} = \omega \cdot \mathbf{M} = \omega \cdot \left[\begin{array}{ccccc|c} z_1 & z_2 & \ldots & z_{n-1} & z_n & \\ \hline a_{11} & a_{12} & \ldots & a_{1,n-1} & a_{1n} & b_1 \\ a_{21} & b_{22} & \ldots & a_{2,n-1} & a_{2n} & b_2 \\ \vdots & \ddots & & \vdots & \vdots & \\ a_{K1} & a_{K2} & \ldots & a_{K,n-1} & a_{Kn} & b_K \end{array} \right] ,$$

The constant ω *is a positive real number that can be interpreted as a weight on the matrix.*

Definition 7 (Extension of Deterministic Potentials) *A deterministic potential* $\mathcal{M}'$ *for* $\mathbf{Z}'$ *is extended to form a deterministic potential* $\mathcal{M} = \mathcal{M}'^{\uparrow \mathbf{Z}}$ *for* $\mathbf{Z} \supset \mathbf{Z}'$ *by inserting zero coefficients in the equations for the variables in* $\mathbf{Z} \setminus \mathbf{Z}'$*.*

Example 12 Suppose evidence exists that the variables Z_4 and Z_5 in the Bayesian network of Example 3 are equal to 76.5 and 78, respectively. This evidence is expressed by the deterministic potentials $\mathcal{M}_3$ for Z_4 and $\mathcal{M}_4$ for Z_5, which each include one equation: $Z_4 = 76.5$ and $Z_5 = 78$. The details of the deterministic potential $\mathcal{M}_3$ are

$$\mathcal{M}_3 = 1 \cdot \left[\begin{array}{c|c} z_4 & \\ \hline 1 & 76.5 \end{array} \right] \text{ or } \mathcal{M}_3^{\uparrow \{Z_1, Z_2, Z_3, Z_4\}} = 1 \cdot \left[\begin{array}{cccc|c} z_4 & z_1 & z_2 & z_3 & \\ \hline 1 & 0 & 0 & 0 & 76.5 \end{array} \right] .$$

4.2 Combination

To propagate deterministic potentials in hybrid Bayesian networks, we must define the operations of combination and marginalization. We define combination of a deterministic potential with both another deterministic potential, and with a probability potential.

4.2.1 Deterministic Potentials

Example 13 Consider the deterministic potentials $\mathcal{M}_{1,A}$, $\mathcal{M}_{1,B}$, and $\mathcal{M}_3$ from Examples 11 and 12, respectively. The combination of $\mathcal{M}_{1,A}$ and $\mathcal{M}_3$ is denoted by $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$, and the combination of $\mathcal{M}_{1,B}$ and $\mathcal{M}_3$ is denoted by $(\mathcal{M}_{1,B} \otimes \mathcal{M}_3)$. No further combination occurs and we maintain, for instance, $\mathcal{M}_{1,A}$ and $\mathcal{M}_3$, as decomposed components of $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$. The domain of $\mathcal{M}_{1,A}$ and $\mathcal{M}_{1,B}$ is $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$ and the domain of $\mathcal{M}_3$ is Z_4, while the domain of $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$ and $(\mathcal{M}_{1,B} \otimes \mathcal{M}_3)$ is $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$.

Definition 8 (Combination of Deterministic Potentials) *Let* $\mathcal{M}_1$ *be a deterministic potential for* $\mathbf{Z}_1$ *and* $\mathcal{M}_2$ *be a deterministic potential for* $\mathbf{Z}_2$ *with* $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2$*. The combination of* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *is a deterministic potential* $(\mathcal{M}_1 \otimes \mathcal{M}_2)$ *for* $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2$*.*

We refer to $(\mathcal{M}_1 \otimes \mathcal{M}_2)$ as a deterministic potential with *component potentials* $\mathcal{M}_1$ and $\mathcal{M}_2$. Component potentials are themselves deterministic potentials and are maintained as decomposed weighted matrices.

Proposition 1 (Commutativity and Associativity of Combination of Deterministic Potentials) *If* $\mathcal{M}_1$*,* $\mathcal{M}_2$*, and* $\mathcal{M}_3$ *are deterministic potentials, then*

$$\mathcal{M}_1 \otimes \mathcal{M}_2 = \mathcal{M}_2 \otimes \mathcal{M}_1 \quad \textit{and}$$
$$(\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3 = \mathcal{M}_1 \otimes (\mathcal{M}_2 \otimes \mathcal{M}_3) \quad .$$

4.2.2 Deterministic and Probability Potentials

Example 14 Consider the deterministic potentials $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$ and $(\mathcal{M}_{1,B} \otimes \mathcal{M}_3)$ from Examples 11 and 13. Suppose ϕ_5 is a density potential for $\{Y_1, Z_1, Z_2\}$ with fragments $\phi_{5,A}$ and $\phi_{5,B}$ for the cases $Y_1 = A$ and $Y_1 = B$, respectively. The potentials $\phi_{5,A}$ and $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$ cannot be combined by pointwise multiplication, because the joint density for $\{Z_1, Z_2, Z_3, Z_4\}$ does not exist. Thus, we simply denote the combination of these potentials by $(\phi_{5,A} \otimes (\mathcal{M}_{1,A} \otimes \mathcal{M}_3))$ and maintain $\phi_{5,A}$ and $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$ as decomposed potentials. Similarly, the combination of $\phi_{5,B}$ and $(\mathcal{M}_{1,B} \otimes \mathcal{M}_3)$ is denoted by $(\phi_{5,B} \otimes (\mathcal{M}_{1,B} \otimes \mathcal{M}_3))$. These are fragments of the potential $(\phi_5 \otimes (\mathcal{M}_1 \otimes \mathcal{M}_3))$ for $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$.

Definition 9 (Combination of Deterministic and Probability Potentials) *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be deterministic potentials for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, and let φ_1 and φ_2 be mass or density potentials for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, with $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2$. The combination of these potentials is*

$$\varphi_1 \otimes \varphi_2 \otimes \mathcal{M}_1 \otimes \mathcal{M}_2 = (\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2)) = ((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \varphi)$$

where $\varphi = (\varphi_1 \otimes \varphi_2)$ is calculated as in Defintion 4.

Once the probability potentials and deterministic potentials are combined using the appropriate definitions, no further combination is performed. We will see later that the result of the marginalization operation in Definition 11 for the combination of a deterministic potential and a probability potential is itself the combination of a deterministic potential and a probability potential which contains additive factors. Note that we may need to combine this with another potential according to Defintion 9, in which case the additive factors of the two potentials are combined pointwise.

Proposition 2 (Commutativity and Associativity of Combination of Deterministic and Probability Potentials) *If $\mathcal{M}_1$ and $\mathcal{M}_2$ are deterministic potentials and φ_1 and φ_2 are mass or density potentials, then*

$$\begin{aligned} \mathcal{M}_1 \otimes \varphi_1 &= \varphi_1 \otimes \mathcal{M}_1 \quad , \\ (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \varphi_1 &= \mathcal{M}_1 \otimes (\mathcal{M}_2 \otimes \varphi_1) \quad , \quad \textit{and} \\ (\varphi_1 \otimes \varphi_2) \otimes \mathcal{M}_1 &= \varphi_1 \otimes (\varphi_2 \otimes \mathcal{M}_1) \quad . \end{aligned}$$

4.3 Marginalization

In this section, marginalization is defined for a deterministic potential, and for a deterministic potential previously combined with a probability potential. The latter is defined separately for discrete and continuous variables.

4.3.1 Deterministic Potential

Example 15 Consider the potential $(\phi_{5,A} \otimes (\mathcal{M}_{1,A} \otimes \mathcal{M}_3))$ that resulted from the combination in Example 14. The marginalization of Z_4 from this potential involves only the potential $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)$ because Z_4 is not in the domain of ϕ_5. This operation, denoted by $(\mathcal{M}_{1,A} \otimes \mathcal{M}_3)^{-Z_4}$, is performed by first creating the following matrix:

$$\mathbf{M}_{5,A} = \left[\begin{array}{cccc|c} z_4 & z_1 & z_2 & z_3 & \\ \hline 1 & -3 & -5 & -4 & 0 \\ 1 & 0 & 0 & 0 & 76.5 \end{array}\right] .$$

From the matrix $\mathbf{M}_{5,A}$, we create the matrices

$$\mathbf{R}_{5,A} = \left[\begin{array}{cccc|c} z_4 & z_1 & z_2 & z_3 & \\ \hline 1 & 0 & 0 & 0 & 76.5 \\ 0 & 1 & 1.667 & 1.333 & 25.5 \end{array}\right] \quad \text{and} \quad \mathbf{P}_{5,A} = \left[\begin{array}{cc} z_4 & z_1 \\ \hline 1 & -3 \\ 1 & 0 \end{array}\right] .$$

The matrix $\mathbf{M}_{5,A}$ contains two lead variables (Z_4 and Z_1) and two free variables (Z_2 and Z_3). The matrix $\mathbf{R}_{5,A}$ is the reduced row echelon form (RREF) of $\mathbf{M}_{5,A}$ and the matrix $\mathbf{P}_{5,A}$ is the first two columns of $\mathbf{M}_{5,A}$. The result of the operations is the following deterministic potential:

$$\mathcal{M}_{6,A} = \omega_{6,A} \cdot \mathbf{M}_{6,A} = \left(1 \cdot \frac{1}{\left|det\left(\mathbf{P}_{5,A}\right)\right|}\right) \cdot \left[\begin{array}{ccc|c} z_1 & z_2 & z_3 & \\ \hline 1 & 1.667 & 1.333 & 25.5 \end{array}\right] .$$

Similarly, the result of the marginalization of Z_4 from $(\mathcal{M}_{1,B} \otimes \mathcal{M}_3)$ is denoted by $\mathcal{M}_{6,B} = (\mathcal{M}_{1,B} \otimes \mathcal{M}_3)^{-Z_4}$.

Marginalization of a continuous variable Z_i from a deterministic potential is accomplished by arranging the component potentials whose domains contain Z_i in a matrix with the coefficients on Z_i in the first column, reducing this matrix to RREF, then eliminating the first column and first row of the RREF matrix. The weights on the component potentials are multiplied, then updated using the absolute value of the inverse of the determinant of the matrix of coefficients on lead variables. Formally, the marginalization operation for a deterministic potential is stated in the following definition.

Definition 10 (Marginalization of Deterministic Potentials)
Let $\mathcal{M} = (\mathcal{M}_1 \otimes \mathcal{M}_2)$ be a deterministic potential on $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2$, where $\mathcal{M}_1$ and $\mathcal{M}_2$ are deterministic potentials for $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively. Define $\mathbf{M}_1^+$ as the matrix formed by rearranging $\mathbf{M}_1$ with Z_i in the first column, let $\mathbf{P}_1$ denote the square matrix formed from the coefficients on the lead variables in $\mathbf{M}_1^+$, and assume that $\det(\mathbf{P}_1) \neq 0$. The marginal of $\mathcal{M}$ for $\mathbf{Z}' = \mathbf{Z} \setminus Z_i$, where $Z_i \in \mathbf{Z}_1$ and $Z_i \notin \mathbf{Z}_2$, is computed as

$$\mathcal{M}^{-Z_i} = \left(\mathcal{M}_1^{-Z_i} \otimes \mathcal{M}_2\right) ,$$

where $\mathcal{M}_1^{-Z_i} = \omega_1' \cdot \mathbf{R}_1'$, with $\omega_1' = \omega_1 \cdot \frac{1}{|det(\mathbf{P_1})|}$ and $\mathbf{R}'_1$ is obtained by eliminating the first column and first row from the reduced row echolon form (RREF) $\mathbf{R}_1$ of $\mathbf{M}_1^+$. If the matrix $\mathbf{M}_1$ has only one row, the result of the operation is an identity potential.

Proposition 3 (Consonance of Marginalization for Deterministic Potentials) *If $\mathcal{M}$ is a deterministic potential for $\mathbf{Z}$ with $(Z_i, Z_j) \in \mathbf{Z}$, then $\left(\mathcal{M}^{-Z_i}\right)^{-Z_j} = \left(\mathcal{M}^{-Z_j}\right)^{Z_i}$.*

Proposition 4 (Distributivity of Marginalization over Combination for Deterministic Potentials) *If $\mathcal{M}_1$ and $\mathcal{M}_2$ are deterministic potentials for $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively with $Z_i \in \mathbf{Z}_1$ and $Z_i \notin \mathbf{Z}_2$, then $(\mathcal{M}_1 \otimes \mathcal{M}_2)^{-Z_i} = \mathcal{M}_1^{-Z_i} \otimes \mathcal{M}_2$.*

4.3.2 Deterministic and Probability Potentials—Discrete Variables

Example 16 Consider the potential $(\phi_5 \otimes (\mathcal{M}_1 \otimes \mathcal{M}_3))$ resulting from the combination in Example 14. The values of the potential resulting from the marginalization of Y_1 from $(\phi_5 \otimes (\mathcal{M}_1 \otimes \mathcal{M}_3))$ are determined as

$$\begin{aligned}&(\phi_5 \otimes (\mathcal{M}_1 \otimes \mathcal{M}_3))^{-Y_1}(z_1, z_2, z_3, z_4) = \\ &\quad (\phi_{5,A} \otimes (\mathcal{M}_{1,A} \otimes \mathcal{M}_3))(z_1, z_2, z_3, z_4) + (\phi_{5,B} \otimes (\mathcal{M}_{1,B} \otimes \mathcal{M}_3))(z_1, z_2, z_3, z_4) \ .\end{aligned}$$

Note from Example 15 that removing Z_4 from this potential will result in the potential $(\phi_5 \otimes \mathcal{M}_6)$ for $\{Z_1, Z_2, Z_3\}$.

Marginalization of a discrete variable from the combination of a deterministic and probability potential involves summing the fragments of the potential over the state space of the variable being removed. The results are maintained as a decomposed sum of factors. This is formalized in the following definition.

Definition 11 (Marginalization of a Discrete Variable from the Combination of a Deterministic Potential and Probability Potential) *Let $(\varphi \otimes \mathcal{M})$ be a potential for $Y \cup \mathbf{Y} \cup \mathbf{Z}$ where φ is a mass or density potential and $\mathcal{M}$ is a deterministic potential. The marginal of $(\varphi \otimes \mathcal{M})$ for $\mathbf{Y} \cup \mathbf{Z}$ is computed as*

$$(\varphi \otimes \mathcal{M})^{-Y_i}(\mathbf{z}, \mathbf{Y} = \mathbf{y}) = \sum_{y_i \in \Omega_Y} (\varphi_{y_i,\mathbf{y}} \otimes \mathcal{M}_{y_i,\mathbf{y}})(\mathbf{z}) \ ,$$

for all $\mathbf{y} \in \Omega_{\mathbf{Y}}$.

Proposition 5 (Consonance of Marginalization of Discrete Variables from Deterministic and Probability Potentials) *If $(\varphi \otimes \mathcal{M})$ is a potential for $\mathbf{Y} \cup \mathbf{Z}$ with $(Y_i, Y_j) \in \mathbf{Y}$, then*

$$\left((\varphi \otimes \mathcal{M})^{-Y_i}\right)^{-Y_j} = \left((\varphi \otimes \mathcal{M})^{-Y_j}\right)^{-Y_i} \ .$$

Proposition 6 (Distributivity of Marginalization of Discrete Variables over Combination from Deterministic and Probability Potentials) *If $(\varphi_1 \otimes \mathcal{M}_1)$ and $(\varphi_2 \otimes \mathcal{M}_2)$ are potentials for $\mathbf{Y}_1 \cup \mathbf{Z}$ and $\mathbf{Y}_2 \cup \mathbf{Z}$, respectively, with $Y_i \in \mathbf{Y}_1$ and $Y_i \notin \mathbf{Y}_2$ and where $\mathbf{Y} = \mathbf{Y}_1 \cup \mathbf{Y}_2$ and $\mathbf{Y}' = \mathbf{Y} \setminus Y_i$, then*

$$((\varphi_1 \otimes \mathcal{M}_1) \otimes (\varphi_2 \otimes \mathcal{M}_2))^{-Y_i} = (\varphi_1 \otimes \mathcal{M}_1)^{-Y_i} \otimes (\varphi_2 \otimes \mathcal{M}_2) \ .$$

4.3.3 Deterministic and Probability Potentials—Continuous Variables

Example 17 Consider the potential $(\phi_5 \otimes \mathcal{M}_6)$ described in Examples 15 and 16. Suppose this potential is combined with $\mathcal{M}_2$ for $\{Z_1, Z_2, Z_3, Z_5\}$ and $\mathcal{M}_4$ for Z_5 as described in Examples 11 and 12. To marginalize Z_3 from $(\phi_5 \otimes (\mathcal{M}_2 \otimes \mathcal{M}_4 \otimes \mathcal{M}_6))$, we first rearrange the matrices in deterministic potentials $\mathcal{M}_2$ and $\mathcal{M}_6^{\uparrow\{Z_1,Z_2,Z_3,Z_5\}}$ (the potentials with Z_1 in their domains) with the coefficients on Z_1 in the first column:

$$\mathbf{M}^{+}_{7,A} = \begin{array}{c} \begin{array}{ccccc} z_1 & z_2 & z_3 & z_5 & \end{array} \\ \left[\begin{array}{cccc|c} 1 & 1.667 & 1.333 & 0 & 25.5 \\ -10 & -2 & -3 & 1 & 0 \end{array}\right] \end{array} \quad \text{and} \quad \mathbf{M}^{+}_{7,B} = \begin{array}{c} \begin{array}{ccccc} z_1 & z_2 & z_3 & z_5 & \end{array} \\ \left[\begin{array}{cccc|c} 1 & 0.333 & 1.333 & 0 & 25.5 \\ -10 & -2 & -3 & 1 & 0 \end{array}\right] \end{array} \quad .$$

The RREF matrices are calculated as

$$\mathbf{R}_{7,A} = \begin{array}{c} \begin{array}{ccccc} z_1 & z_2 & z_3 & z_5 & \end{array} \\ \left[\begin{array}{cccc|c} 1 & 0 & 0.159 & -0.114 & -3.477 \\ 0 & 1 & 0.705 & 0.068 & 17.386 \end{array}\right] \end{array} \quad \text{and} \quad \mathbf{R}_{7,B} = \begin{array}{c} \begin{array}{ccccc} z_1 & z_2 & z_3 & z_5 & \end{array} \\ \left[\begin{array}{cccc|c} 1 & 0 & -1.25 & -0.25 & -38.25 \\ 0 & 1 & 7.75 & 0.75 & 191.25 \end{array}\right] \end{array} \quad .$$

Matrices $\mathbf{P}_{7,A}$ and $\mathbf{P}_{7,B}$ include the first two columns from $\mathbf{M}_{7,A}$ and $\mathbf{M}_{7,B}$, respectively, with $1/\left|det(\mathbf{P}_{7,A})\right| = 0.068$ and $1/\left|det(\mathbf{P}_{7,B})\right| = 0.75$. The first rows of $\mathbf{R}_{7,A}$ and $\mathbf{R}_{7,B}$ are used to transform the potential ϕ_5 into a new density potential for $\{Z_2, Z_3, Z_5\}$ as follows:

$$\phi_{6,A}(z_2) = 1 \cdot \tfrac{1}{3} \cdot \phi_{5,A}\left(-3.477 - 0.159z_3 + 0.114z_5, z_2\right)$$
$$\phi_{6,B}(z_2) = 1 \cdot \tfrac{1}{3} \cdot \phi_{5,B}\left(-38.25 + 1.25z_3 + 0.25z_5, z_2\right) \quad .$$

The second rows of $\mathbf{R}_{7,A}$ and $\mathbf{R}_{7,B}$ are used to form the deterministic potentials

$$\mathcal{M}_{8,A} = 0.068 \cdot \begin{array}{c} \begin{array}{cccc} z_5 & z_2 & z_3 & \end{array} \\ \left[\begin{array}{ccc|c} 0.068 & 1 & 0.705 & 17.386 \end{array}\right] \end{array} \quad \text{and} \quad \mathcal{M}_{8,B} = 0.75 \cdot \begin{array}{c} \begin{array}{cccc} z_5 & z_2 & z_3 & \end{array} \\ \left[\begin{array}{ccc|c} 0.75 & 1 & 7.75 & 191.25 \end{array}\right] \end{array} \quad .$$

The result of the operation is the following potential for $\{Z_2, Z_3, Z_5\}$:

$$\left(\phi_5 \otimes (\mathcal{M}_2 \otimes \mathcal{M}_4 \otimes \mathcal{M}_6)\right)^{-Z_1} = \left(\phi_6 \otimes (\mathcal{M}_4 \otimes \mathcal{M}_8)\right) \quad .$$

Marginalization of a continuous variable Z_i from the combination of a deterministic potential and a probability potential is substitution of the first row of the RREF matrix corresponding with the deterministic potential into the probability potential, with the result multiplied by the weights on the deterministic potential. The matrix resulting after eliminating the first row and column from the RREF matrix remains as a deterministic potential. The result of the operation is a new probability potential combined with a new deterministic potential. This operation is derived from the method of convolutions in probability theory, as shown in (Cobb and Shenoy 2006b). Mao *et al.* (2004) also discusses convolution operations in the context of probabilistic models. The formal definition of this operation is stated as follows.

Definition 12 (Marginalization of a Continuous Variable from the Combination of a Deterministic Potential and Probability Potential) *Let* $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))$ *be a potential on* $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2 \cup \mathbf{Z}_3$*, where* φ *is a mass or density potential on* $\mathbf{Z}_1$*, and* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *are deterministic potentials on* $\mathbf{Z}_2$ *and* $\mathbf{Z}_3$*, respectively. To marginalize* $Z_i \in (\mathbf{Z}_1 \cup \mathbf{Z}_2)$*,* $Z_i \notin \mathbf{Z}_3$*, perform the marginalization operation in Definition 10 on* $\mathcal{M}_1$ *to obtain* $\mathbf{R}_1$*,* $\mathbf{R}'_1$*, and* $\det(\mathbf{P}_1)$*. The marginal of* $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))$ *for* $\mathbf{Z} \backslash Z_i$ *is denoted by* $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))^{-Z_i} = (\varphi' \otimes (\mathcal{M}'_1 \otimes \mathcal{M}_2))$*, with values of the potentials determined as*

$$\varphi'(\mathbf{z}'_1, \mathbf{z}'_2) = \omega_1 \cdot \varphi\left(h\left(\mathbf{z}'_2\right), \mathbf{z}'_1\right) \quad \text{and} \quad \mathcal{M}'_1 = \frac{1}{\left|\det\left(\mathbf{P}_1\right)\right|} \cdot \mathbf{R}'_1 \quad ,$$

for all $\mathbf{z}' \in \Omega_{\mathbf{Z}'}$*, where* $\mathbf{Z}' = \mathbf{Z} \setminus Z_i$*,* $\mathbf{z}_1 = (\mathbf{z}'_1, z_i)$*,* $\mathbf{z}_2 = (\mathbf{z}'_2, z_i)$*, and*

$$h(\mathbf{z}'_2) = -a_1^{(1)} z_1 - \cdots - a_{i-1}^{(1)} z_{i-1} - a_{i+1}^{(1)} z_{i+1} - \cdots - a_n^{(1)} z_n + b^{(1)} \ .$$

The constant $a_j^{(1)}$ *represents the coefficient for variable* Z_j *in the first row of* $\mathbf{R}_1$*. If* $\mathbf{Z}_2 \backslash Z_i = \emptyset$*, the result of the operation includes only a probability potential. If* Z_i *is not in the domain of both the deterministic and probability potential, we simply marginalize* Z_i *from the potential that contains* Z_i *in its domain according to either Definition 5 or 10, leave the other potential unchanged, and maintain the result as a decomposed sum of potentials.*

The normalization constant $1/\left|det\left(\mathbf{P}_1\right)\right|$ is justified as follows. If we have a joint PDF for Z_1 and Z_2, and we define $Z_3 = g_1(z_1, z_2)$ and $Z_4 = g_2(z_1, z_2)$, then the joint PDF of Z_3 and Z_4 has a normalization constant equal to the inverse of the absolute value of the Jacobian of g_1 and g_2 (Hogg and Craig 1978). The Jacobian of g_1 and g_2 is the 2×2 matrix of partial derivatives of g_1 and g_2 with respect to Z_1 and Z_2. For the case of linear functions, the Jacobian translates to the determinant of the matrix $\mathbf{P}_1$.

A potential that represents the combination of a probability potential and a deterministic potential may have additive factors resulting from the previous marginalization of a discrete variable, as detailed in Definition 11. In this case, the operation in Definition 12 is applied to each additive factor and the result remains a decomposed sum of potentials. This is justified for the same reason that the marginalization of discrete and continuous variables can occur in any order in the operation of Definition 5.

Proposition 7 (Consonance of Marginalization of Continuous Variables from Deterministic and Probability Potentials) *If* $(\varphi \otimes \mathcal{M})$ *is a potential for* $\mathbf{Z}$ *with* $(Z_i, Z_j) \in \mathbf{Z}$*, then*

$$\left((\varphi \otimes \mathcal{M})^{-Z_i}\right)^{-Z_j} = \left((\varphi \otimes \mathcal{M})^{-Z_j}\right)^{-Z_i} \ .$$

Proposition 8 (Distributivity of Marginalization of Continuous Variables over Combination from Deterministic and Probability Potentials) *If* $(\varphi_1 \otimes \mathcal{M}_1)$ *and* $(\varphi_2 \otimes \mathcal{M}_2)$ *are potentials for* $\mathbf{Z}_1$ *and* $\mathbf{Z}_2$*, respectively, with* $Z_i \in \mathbf{Z}_1$ *and* $Z_i \notin \mathbf{Z}_2$ *and where* $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2$ *and* $\mathbf{Z}' = \mathbf{Z} \setminus Z_i$*, then*

$$((\varphi_1 \otimes \mathcal{M}_1) \otimes (\varphi_2 \otimes \mathcal{M}_2))^{-Z_i} = (\varphi_1 \otimes \mathcal{M}_1)^{-Z_i} \otimes (\varphi_2 \otimes \mathcal{M}_2) \ .$$

The class of MTE density potentials is closed under combination and marginalization of probability potentials (Moral *et al.* 2001). The following theorem states that the class of MTE probability potentials is closed under Definition 12.

Theorem 1 (Marginalization of MTE density potentials and deterministic potentials). *If* $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))$ *is a potential on* $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2 \cup \mathbf{Z}_3$*, where* φ *is an MTE mass or density potential on* $\mathbf{Z}_1$*, and* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *are deterministic potentials on* $\mathbf{Z}_2$ *and* $\mathbf{Z}_3$*, respectively, then the probability potential* φ' *resulting from the calculation of* $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))^{-Z_i} = (\varphi' \otimes (\mathcal{M}'_1 \otimes \mathcal{M}_2))$ *is an MTE probability potential.*

Now that we have introduced deterministic potentials and operations for deterministic potentials, we can discuss a mixed potential representation that allows mixed distributions to be represented in hybrid Bayesian networks.

5 Mixed Potentials

This section defines mixed potentials, which can accomodate the distributions of conditionally deterministic and mixed random variables, and outlines combination and marginalization for these potentials. Each definition is preceded by one or more illustrative examples.

5.1 Definition

Example 18 Consider the Bayesian network in Figure 3 where $P(Y_1 = A) = P(Y_1 = B) - 0.5$, which is represented by the mass potential α with values $\alpha_1(Y_1 = A) = \alpha_1(Y_1 = B) = 0.5$. Continuous variables are normally distributed as follows: $Z_1 \mid \{Y_1 = A\} \sim N(1, 100)$, $Z_1 \mid \{Y_1 = B\} \sim N(4, 64)$, $Z_2 \mid z_1 \sim N(2z_1, 100)$, and $Z_3 \mid z_2 \sim N(3z_2, 100)$. These conditional PDF's are approximated by MTE approximations to the normal PDF (Cobb and Shenoy 2006a) with the appropriate mean and variance, with the density potentials denoted by ϕ_1 for $\{Y_1, Z_1\}$ (with fragments $\phi_{1,A}$ and $\phi_{1,B}$), ϕ_2 for $\{Z_1, Z_2\}$, and ϕ_3 for $\{Z_2, Z_3\}$.

Mixed potentials for Y_1 and $\{Y_1, Z_1\}$ are defined as

$$\begin{aligned}&\theta_1(Y_1 = A) = (\alpha_1(Y_1 = A), \iota) \quad \text{and} \quad \theta_1(Y_1 = A) = (\alpha_1(Y_1 = B), \iota) \ , \\ &\zeta_1(z_1, Y_1 = A) = (1, \phi_{1,A}(z_1)) \quad \text{and} \quad \zeta_1(z_1, Y_1 = B) = (1, \phi_{1,B}(z_1)) \ .\end{aligned}$$

Mixed potentials for $\{Z_1, Z_2\}$ and $\{Z_2, Z_3\}$ are defined as $\zeta_2(z_1, z_2) = (1, \phi_2(z_1, z_2))$ and $\zeta_3(z_2, z_3) = (1, \phi_3(z_2, z_3))$, respectively. The mixed potentials for $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$ and $\{Z_1, Z_2, Z_3, Z_5\}$ are defined as

$$\begin{aligned}&\zeta_4(z_1, z_2, z_3, z_4, Y_1 = A) = (1, \mathcal{M}_{1,A}) \quad \text{and} \quad \zeta_4(z_1, z_2, z_3, z_4, Y_1 = B) = (1, \mathcal{M}_{1,B}), \\ &\zeta_5(z_1, z_2, z_3, z_5) = (1, \mathcal{M}_2) \ ,\end{aligned}$$

where $\mathcal{M}_{1,A}$, $\mathcal{M}_{1,B}$, and $\mathcal{M}_2$ are defined as in Example 11. The evidence that $Z_4 = 76.5$ and $Z_5 = 78$ is represented by the mixed potentials $\eta_1(z_4) = (1, \mathcal{M}_3)$ and $\eta_2(z_5) = (1, \mathcal{M}_4)$, with $\mathcal{M}_3$ and $\mathcal{M}_4$ defined as in Example 12.

Definition 13 *A* **mixed potential** ζ *in a hybrid Bayesian network for a multi-dimensional variable* $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z} = \mathbf{Y} \cup \mathbf{Z}_1 \cup \mathbf{Z}_2$ *is a pair*

$$\zeta(\mathbf{x}) = \zeta(\mathbf{y}, \mathbf{z}) = (\alpha(\mathbf{y}, \mathbf{z}_1), (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z}_2)) \ ,$$

where α is a mass potential, $\varphi_{\mathbf{y}}$ is a mass or density potential, and $\mathcal{M}_{\mathbf{y}}$ is a deterministic potential.

We refer to the first part of the mixed potential as the *mass* part and the second part of the mixed potential as the *density* part. The absence of probability mass is represented by a 1 in the mass part and an absence of density is represented by the identity potential ι in the density part. The potential $(\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})$ is defined as in Definition 9 to allow the density part to contain the combination of a deterministic potential and probability potential, but either $\varphi_{\mathbf{y}}$ or $\mathcal{M}_{\mathbf{y}}$ may be empty.

5.2 Combination

Example 19 Consider the potentials θ_1, ζ_1, ζ_2, and ζ_4 defined in Example 18. The combination of these mixed potentials is the following mixed potential for $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$,

$$(\theta_1 \otimes \zeta_1 \otimes \zeta_2 \otimes \zeta_4) = (\alpha_1, (\phi_4 \otimes \mathcal{M}_1)) \quad ,$$

where $\phi_4 = (\phi_1 \otimes \phi_2)$ as calculated by Definition 4. Specific values of this mixed potential can be defined, for example, as,

$$(\theta_1 \otimes \zeta_1 \otimes \zeta_2 \otimes \zeta_4)\,(z_1, z_2, z_3, z_4, Y_1 = A) = (\alpha_1(Y_1 = A), (\phi_{4,A} \otimes \mathcal{M}_{1,A})\,(z_1, z_2, z_3, z_4)) \quad ,$$

where values of $\phi_{4,A}$ are calculated as $\phi_{4,A}(z_1, z_2) = \phi_{1,A}(z_1) \cdot \phi_2(z_1, z_2)$.

Combination of mixed potentials is combination of the mass and density parts, according to the previously defined combination operations. This is formalized in the following definition.

Definition 14 (Combination of Mixed Potentials) *Let* $\zeta_1 = (\alpha_1, (\varphi_1 \otimes \mathcal{M}_1))$ *and* $\zeta_2 = (\alpha_2, (\varphi_2 \otimes \mathcal{M}_2))$ *be two mixed potentials. The combination of* ζ_1 *and* ζ_2 *is a mixed potential* ζ *for defined as*

$$\zeta = \zeta_1 \otimes \zeta_2 = (\alpha_1 \otimes \alpha_2, (\varphi_1 \otimes \varphi_2) \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2)) \quad .$$

Proposition 9 (Commutativity and Associativity of Combination of Mixed Potentials) *If* $\zeta_1 = (\alpha_1, (\varphi_1 \otimes \mathcal{M}_1))$, $\zeta_2 = (\alpha_2, (\varphi_2 \otimes \mathcal{M}_2))$, *and* $\zeta_3 = (\alpha_3, (\varphi_3 \otimes \mathcal{M}_3))$ *are mixed potentials, then*

$$\zeta_1 \otimes \zeta_2 = \zeta_2 \otimes \zeta_1 \quad \textit{and}$$
$$\zeta_1 \otimes (\zeta_2 \otimes \zeta_3) = (\zeta_1 \otimes \zeta_2) \otimes \zeta_3 \quad .$$

5.3 Marginalization

Marginalization of mixed potentials is defined separately for discrete and continuous variables.

5.3.1 Discrete Variables

Example 20 Consider the potential $(\theta_1 \otimes \zeta_1 \otimes \zeta_2 \otimes \zeta_4)$ for $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$ resulting from the combination operation in Example 18. Discrete variable Y_1 is in the domain of the mass and density parts of this potential, as both the values of the mass potential α_1, the density potential ϕ_4, and the deterministic potential $\mathcal{M}_1$ depend on Y_1. The marginalization of Y_1 from this mixed potential is the following mixed potential for $\{Z_1, Z_2, Z_3, Z_4\}$,

$$\zeta_8 = \left(1, (\phi_5 \otimes \mathcal{M}_1)^{-Y_1}\right) \quad ,$$

where $\phi_5 = \alpha_1 \otimes \phi_4$ is calculated as in Definition 4. Specific values of this potential can be defined as,

$$\zeta_8(z_1, z_2, z_3, z_4) = (1, (\phi_{5,A} \otimes \mathcal{M}_{1,A})(z_1, z_2, z_3, z_4) + (\phi_{5,B} \otimes \mathcal{M}_{1,B})(z_1, z_2, z_3, z_4)) \quad .$$

The results of the marginalization of a discrete variable Y_i from a mixed potential depends on the domain of the mass and density parts of the potential. The following definition considers three possible cases.

Definition 15 (Marginalization of Discrete Variables from Mixed Potentials) *Let $\zeta = (\alpha, (\varphi \otimes \mathcal{M}))$ be a mixed potential on $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z} = \mathbf{Y}_1 \cup \mathbf{Y}_2 \cup \mathbf{Z}$, where α is a mass potential for $\mathbf{Y}_1 \cup \mathbf{Z}$, and where φ and $\mathcal{M}$ are probability and deterministic potentials, respectively, for $\mathbf{Y}_2 \cup \mathbf{Z}$. Define $\mathbf{Y}' = \mathbf{Y} \setminus Y_i$ and $\mathbf{y} = (\mathbf{y}', y_i)$. The marginalization of a discrete variable Y_i from ζ results in a mixed potential for $\mathbf{Y}' \cup \mathbf{Z}$ defined according to one of the following:*

1. *If $Y_i \notin \mathbf{Y}_2$, then $(\varphi_{\mathbf{y}'} \otimes \mathcal{M}_{\mathbf{y}'})(\mathbf{z}) = (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z})$, and the result is computed as*
$$\zeta^{-Y_i}(\mathbf{y}', \mathbf{z}) = \left(\alpha_{\mathbf{y}'}^{-Y_i}(\mathbf{z}), (\varphi_{\mathbf{y}'} \otimes \mathcal{M}_{\mathbf{y}'})(\mathbf{z})\right) \text{ for all } (\mathbf{y}', \mathbf{z}) \in \Omega_{\mathbf{Y}', \mathbf{Z}} \ .$$
2. *If $Y_i \in \mathbf{Y}_2$ and $Y_i \notin \mathbf{Y}_1$, then $\alpha_{\mathbf{y}'}(\mathbf{z}) = \alpha_{\mathbf{y}}(\mathbf{z})$, and the result is computed as*
$$\zeta^{-Y_i}(\mathbf{y}', \mathbf{z}) = \left(\alpha_{\mathbf{y}'}(\mathbf{z}), (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})^{-Y_i}(\mathbf{z})\right) \text{ for all } (\mathbf{y}', \mathbf{z}) \in \Omega_{\mathbf{Y}', \mathbf{Z}} \ .$$
3. *If $Y_i \in (\mathbf{Y}_1 \cap \mathbf{Y}_2)$, the result is computed as*
$$\zeta^{-Y_i}(\mathbf{y}', \mathbf{z}) = \left(1, ((\alpha_{\mathbf{y}} \otimes \varphi_{\mathbf{y}}) \otimes \mathcal{M}_{\mathbf{y}})^{-Y_i}(\mathbf{z})\right) \text{ for all } (\mathbf{y}', \mathbf{z}) \in \Omega_{\mathbf{Y}', \mathbf{Z}} \ .$$

Note that in the last case, it is possible that $\mathbf{Y}' = \emptyset$ and the result is a potential for $\mathbf{Z}$.

Proposition 10 (Consonance of Marginalization of Discrete Variables from Mixed Potentials) *If $\zeta = (\alpha, (\varphi \otimes \mathcal{M}))$ is a mixed potential for $\mathbf{Y} \cup \mathbf{Z}$, with $(Y_i, Y_j) \in \mathbf{Y}$, then $\left(\zeta^{-Y_i}\right)^{-Y_j} = \left(\zeta^{-Y_j}\right)^{-Y_i}$.*

Proposition 11 (Distributivity of Marginalization of Discrete Variables over Combination from Mixed Potentials) *If $\zeta_1 = (\alpha_1, (\varphi_1 \otimes \mathcal{M}_1))$ and $\zeta_2 = (\alpha_2, (\varphi_2 \otimes \mathcal{M}_2))$ are mixed potentials for $\mathbf{Y}_1 \cup \mathbf{Z}$ and $\mathbf{Y}_2 \cup \mathbf{Z}$, respectively, with $Y_i \in \mathbf{Y}_1$ and $Y_i \notin \mathbf{Y}_2$ and where $\mathbf{Y} = \mathbf{Y}_1 \cup \mathbf{Y}_2$ and $\mathbf{Y}' = \mathbf{Y} \setminus Y_i$, then $(\zeta_1 \otimes \zeta_2)^{-Y_i} = \zeta_1^{-Y_i} \otimes \zeta_2$.*

5.3.2 Continuous Variables

Example 21 Consider the potential $(\phi_6 \otimes (\mathcal{M}_6 \otimes \mathcal{M}_8))$ for $\{Z_2, Z_3, Z_5\}$ resulting from the operation in Example 17. A mixed potential represenation of this potential is

$$\zeta_{10}(z_2, z_3, z_5) = (1, (\phi_{6,A} \otimes (\mathcal{M}_4 \otimes \mathcal{M}_{8,A}))(z_2, z_3, z_5) + (\phi_{6,B} \otimes (\mathcal{M}_4 \otimes \mathcal{M}_{8,B}))(z_2, z_3, z_5)) \ .$$

Marginalization of Z_5 from the mixed potential ζ_{10} involves marginalizing Z_5 from $(\phi_6 \otimes (\mathcal{M}_6 \otimes \mathcal{M}_8))$ according to Definition 12. The result is denoted by

$$\zeta_{10}^{-Z_5}(z_2, z_3) = (1, (\phi_{7,A} \otimes \mathcal{M}_{10,A})(z_2, z_3) + (\phi_{7,B} \otimes \mathcal{M}_{10,B})(z_2, z_3)) \ ,$$

where $(\phi_6 \otimes (\mathcal{M}_6 \otimes \mathcal{M}_8))^{-Z_5} = (\phi_7 \otimes \mathcal{M}_{10})$.

The results of the marginalization of a continuous variable Z_i from a mixed potential depends on the domain of the mass and density parts of the potential. The following definition considers four possible cases.

Definition 16 (Marginalization of Continuous Variables from Mixed Potentials) *Let $\zeta = (\alpha, (\varphi \otimes \mathcal{M}))$ be a mixed potential on $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z} = \mathbf{Y} \cup \mathbf{Z}_1 \cup \mathbf{Z}_2$, where α is a mass potential for $\mathbf{Y} \cup \mathbf{Z}_1$, and where φ and $\mathcal{M}$ are probability and deterministic potentials, respectively, for $\mathbf{Y} \cup \mathbf{Z}_2$. Define $\mathbf{Z}' = \mathbf{Z} \setminus Z_i$ and $\mathbf{z} = (\mathbf{z}', z_i)$. The marginalization of a continuous variable Z_i from ζ results in a mixed potential for $\mathbf{Y} \cup \mathbf{Z}'$ defined according to one of the following:*

1. *If $Z_i \notin \mathbf{Z}_2$ and $\mathbf{Z}' \neq \emptyset$, then $(\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z}') = (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z})$, and the result is computed as*
$$\zeta^{-Z_i}(\mathbf{y}, \mathbf{z}') = \left(\alpha_{\mathbf{y}}^{-Z_i}(\mathbf{z}'), (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z}')\right) \;\; \textit{for all} \;\; (\mathbf{y}, \mathbf{z}') \in \Omega_{\{\mathbf{Y},\mathbf{Z}'\}} \;\; .$$

2. *If $Z_i \in \mathbf{Z}_2$, $Z_i \notin \mathbf{Z}_1$, and $\mathbf{Z}' \neq \emptyset$, then $\alpha_{\mathbf{y}}(\mathbf{z}') = \alpha_{\mathbf{y}}(\mathbf{z})$, and the result is computed as*
$$\zeta^{-Y_i}(\mathbf{y}, \mathbf{z}') = \left(\alpha_{\mathbf{y}}(\mathbf{z}'), (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})^{-Z_i}(\mathbf{z}')\right) \;\; \textit{for all} \;\; (\mathbf{y}, \mathbf{z}') \in \Omega_{\{\mathbf{Y},\mathbf{Z}'\}} \;\; .$$

3. *If $Z_i \in (\mathbf{Z}_1 \cap \mathbf{Z}_2)$ and $\mathbf{Z}' \neq \emptyset$, the result is computed as*
$$\zeta^{-Z_i}(\mathbf{y}, \mathbf{z}') = \left(1, ((\alpha_{\mathbf{y}} \otimes \varphi_{\mathbf{y}}) \otimes \mathcal{M}_{\mathbf{y}})^{-Z_i}(\mathbf{z}')\right) \;\; \textit{for all} \;\; (\mathbf{y}, \mathbf{z}') \in \Omega_{\{\mathbf{Y},\mathbf{Z}'\}} \;\; .$$

4. *If $\mathbf{Z}' = \emptyset$, the result is computed as*
$$\zeta^{-Z_i}(\mathbf{y}) = \left(((\alpha_{\mathbf{y}} \otimes \varphi_{\mathbf{y}}) \otimes \mathcal{M}_{\mathbf{y}})^{-Z_i}, \iota\right) \;\; \textit{for all} \;\; \mathbf{y} \in \Omega_{\mathbf{Y}} \;\; .$$

Proposition 12 (Consonance of Marginalization of Continuous Variables from Mixed Potentials) *If $\zeta = (\alpha, (\varphi \otimes \mathcal{M}))$ is a mixed potential for $\mathbf{Y} \cup \mathbf{Z}$, with $(Z_i, Z_j) \in \mathbf{Z}$, then $\left(\zeta^{-Z_i}\right)^{-Z_j} = \left(\zeta^{-Z_j}\right)^{-Z_i}$.*

Proposition 13 (Distributivity of Marginalization of Continuous Variables over Combination from Mixed Potentials) *If $\zeta_1 = (\alpha_1, (\varphi_1 \otimes \mathcal{M}_1))$ and $\zeta_2 = (\alpha_2, (\varphi_2 \otimes \mathcal{M}_2))$ are mixed potentials for $\mathbf{Y} \cup \mathbf{Z}_1$ and $\mathbf{Y} \cup \mathbf{Z}_2$, respectively, with $Z_i \in \mathbf{Z}_1$ and $Z_i \notin \mathbf{Z}_2$ and where $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2$ and $\mathbf{Z}' = \mathbf{Z} \setminus Z_i$, then $(\zeta_1 \otimes \zeta_2)^{-Z_i} = \zeta_1^{-Z_i} \otimes \zeta_2$.*

Additional illustrations of specific cases of the marginalization operations in Definitions 15 and 16 are provided in the following examples, which are based on propagation in the hybrid Bayesian network in Example 3 with evidence on Z_3, Z_4, and Z_5.

Example 22 Consider the potentials θ_1 and ζ_1 defined in Example 18, and the potential ζ_{18} defined as follows:

$$\begin{aligned} &\zeta_{18}(z_1, Y_1 = A) = (1, (\phi_{13,A} \otimes \mathcal{M}_{16,A})\,(z_1)) \;\; \text{and} \\ &\zeta_{18}(z_1, Y_1 = B) = (1, (\phi_{13,B} \otimes \mathcal{M}_{16,B})\,(z_1)) \;\; , \end{aligned}$$

where

$$\mathcal{M}_{16,A} = \left[\begin{array}{c|c} & z_1 \\ \hline 1 & 2.920 \end{array}\right] \text{ and } \mathcal{M}_{16,B} = \left[\begin{array}{c|c} & z_1 \\ \hline 1 & 0.625 \end{array}\right] .$$

The combination of these three potentials is defined as

$$\begin{aligned}(\theta \otimes \zeta_1 \otimes \zeta_{18})(z_1, Y_1 = A) &= (\alpha_1(Y_1 = A), (\phi_{14,A} \otimes \mathcal{M}_{16,A})(z_1)) \quad \text{and} \\ (\theta \otimes \zeta_1 \otimes \zeta_{18})(z_1, Y_1 = B) &= (\alpha_1(Y_1 = B), (\phi_{14,B} \otimes \mathcal{M}_{16,B})(z_1)) \quad ,\end{aligned}$$

where $\phi_{14,A} = \phi_{1,A} \otimes \phi_{13,A}$ and $\phi_{14,B} = \phi_{1,B} \otimes \phi_{13,B}$. Removing Y_1 according to the third case in Definition 15 results in the potential

$$(\theta \otimes \zeta_1 \otimes \zeta_{18})^{-Y_1}(z_1) = (1, (\phi_{15,A} \otimes \mathcal{M}_{16,A})(z_1) + (\phi_{15,A} \otimes \mathcal{M}_{16,A})(z_1)) \quad ,$$

where $\phi_{15,A}(z_1) = \alpha_1(Y_1 = A) \cdot \phi_{14,A}(z_1)$ and $\phi_{15,B}(z_1) = \alpha_1(Y_1 = B) \cdot \phi_{14,B}(z_1)$. Note that Z_1 is the last remaining variable in the potential, but we cannot sum the factors in the density part of the potential because the probability and deterministic potentials remain decomposed. The continuous variable Z_1 is now restricted to two values and the remaining mixed potential defines a probability distribution as follows:

z_1	$P(Z_1 = z_1)$
2.920	$\phi_{15,A}(2.920)/\kappa_1$
0.625	$\phi_{15,B}(0.625)/\kappa_1$

The normalization constant, $\kappa_1 = \phi_{15,A}(2.920) + \phi_{15,B}(0.625)$, represents the probability of the observed evidence. This probability distribution is the result of marginalizing Z_1 from $(\theta \otimes \zeta_1 \otimes \zeta_{18})^{-Y_1}$ according to the fourth case in Definition 16.

Example 23 Consider the potential $(\theta \otimes \zeta_1 \otimes \zeta_{18})$ resulting from the combination in the previous example. Removing Z_1 (the last continuous variable) according to the fourth case of Definition 16 results in the potential

$$\begin{aligned}(\theta \otimes \zeta_1 \otimes \zeta_{18})^{-Z_1}(Y_1 = A) &= (\phi_{15,A}(2.920), \iota) \quad \text{and} \\ (\theta \otimes \zeta_1 \otimes \zeta_{18})^{-Z_1}(Y_1 = B) &= (\phi_{15,B}(0.625), \iota) \quad ,\end{aligned}$$

where $\phi_{15,A}(z_1) = \alpha_1(Y_1 = A) \cdot \phi_{14,A}(z_1)$ and $\phi_{15,B}(z_1) = \alpha_1(Y_1 = B) \cdot \phi_{14,B}(z_1)$. The result is a mass potential for Y_1. Normalizing this potential gives a probability distribution for Y_1 given the evidence on the continuous variables Z_3, Z_4, and Z_5.

5.4 Restriction of Discrete Variables

Restriction is an operation used to enter evidence. During propagation, restriction is performed by zeroing out potential values for unobserved states. Restriction of a mixed potential to an observed state for a discrete variable is illustrated in the following example.

Example 24 Suppose a mixed potential is defined for $\{Y_1, Z_1, Z_2\}$ as

$$\zeta(z_1, z_2, Y_1 = A) = (0.6, \phi_A(z_1, z_2)) \text{ and } \zeta(z_1, z_2, Y_1 = B) = (0.4, \phi_B(z_1, z_2)) \ .$$

A potential $\xi(Y_1 = B) = (1, \iota)$ for Y_1 characterizes the evidence $Y_1 = B$. The restriction of ζ to $Y_1 = B$, denoted by $\zeta' = (\zeta \otimes \xi)^{-Y_1}$, results in a potential ζ' for $\{Z_1, Z_2\}$ defined as $\zeta'(z_1, z_2) = (0.4, \phi_B(z_1, z_2))$.

Restriction of a discrete variable to an observed state that has positive mass leads to zeroing out all masses for unobserved states. The restriction operation illustrated in the previous example is defined as follows.

Definition 17 (Restriction of Discrete Variables) *Let ζ be a mixed potential for $\mathbf{X} = \{Y\} \cup \mathbf{Y} \cup \mathbf{Z}$, with $\zeta(\mathbf{x}) = (\alpha(y, \mathbf{y}, \mathbf{z}), (\varphi_{y,\mathbf{y}} \otimes \mathcal{M}_{y,\mathbf{y}})(\mathbf{z}))$. Suppose we observe the state y_i of Y. Let ξ be the potential for Y that represents the observation $Y = y_i$, i.e. $\xi(Y = y_i) = (1, \iota)$. The restriction of ζ to $Y = y_i$ is the mixed potential $\zeta^{R(Y=y_i)} = (\zeta \otimes \xi)^{-Y}$ for $\mathbf{Y} \cup \mathbf{Z}$ defined by*

$$\zeta^{R(Y=y_i)}(\mathbf{y}, \mathbf{z}) = (\alpha(y_i, \mathbf{y}, \mathbf{z}), (\varphi_{y,\mathbf{y}} \otimes \mathcal{M}_{y,\mathbf{y}})(\mathbf{z})) \quad \text{for all} \quad (\mathbf{y}, \mathbf{z}) \in \Omega_{\mathbf{Y} \cup \mathbf{Z}} \ .$$

Restriction of a continuous variable is accomplised by assigning a deterministic potential for the continuous variable that, when combined with other probability or deterministic potentials, effectively restricts the variable to a specific value. This restriction operation trivially satisfies the Shenoy-Shafer axioms, so no proof is presented. Example 12 illustrates this type of potential, which will also be used in the example of the next section.

Because all combination and marginalization operations for probability, deterministic, and mixed potentials satisfy the Shenoy-Shafer axioms, we can propagate mixed potentials in any join tree structure to calculate marginal distributions for variables of interest in any hybrid Bayesian network. Although we approximate probability potentials with MTE potentials, the propagation itself is exact.

6 Example

Consider the Bayesian network in Figure 3 where mixed potentials are defined as in Example 18 and initialized to the nodes in the join tree shown in Figure 8. In this section, we consider the calculation of marginals for this network with evidence on variables Z_4 and Z_5.

6.1 Messages

Examples of messages (numbered corresponding to the messages in the join tree of Figure 8) required to calculate marginals for Z_2 and Z_3 considering the evidence are shown below.

(4) $\{Y_1, Z_1, Z_2\}$ to $\{Y_1, Z_1, Z_2, Z_3\}$

The combination of potentials θ_1 for Y_1, ζ_1 for $\{Y_1, Z_1\}$, and ζ_2 for $\{Z_1, Z_2\}$, denoted by $\zeta_7 = (\theta_1 \otimes \zeta_1 \otimes \zeta_2)$, is determined as

$$\begin{aligned}
\zeta_7(z_1, z_2, Y_1 = A) &= (\theta_1 \otimes \zeta_1)(z_1, Y_1 = A) = (\alpha_1(A), \phi_{4,A}(z_1, z_2)) \ , \\
\zeta_7(z_1, z_2, Y_1 = B) &= (\theta_1 \otimes \zeta_1)(z_1, Y_1 = A) = (\alpha_1(B), \phi_{4,B}(z_1, z_2)) \ .
\end{aligned}$$

The new density potentials $\phi_{4,A}$ and $\phi_{4,B}$ are determined as $\phi_{4,A} = (\phi_{1,A} \otimes \phi_2)$ and $\phi_{4,B} = (\phi_{1,B} \otimes \phi_2)$.

(5) $\{Y_1, Z_1, Z_2, Z_3\}$ to $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$

The computation of this message is demonstrated in Example 19.

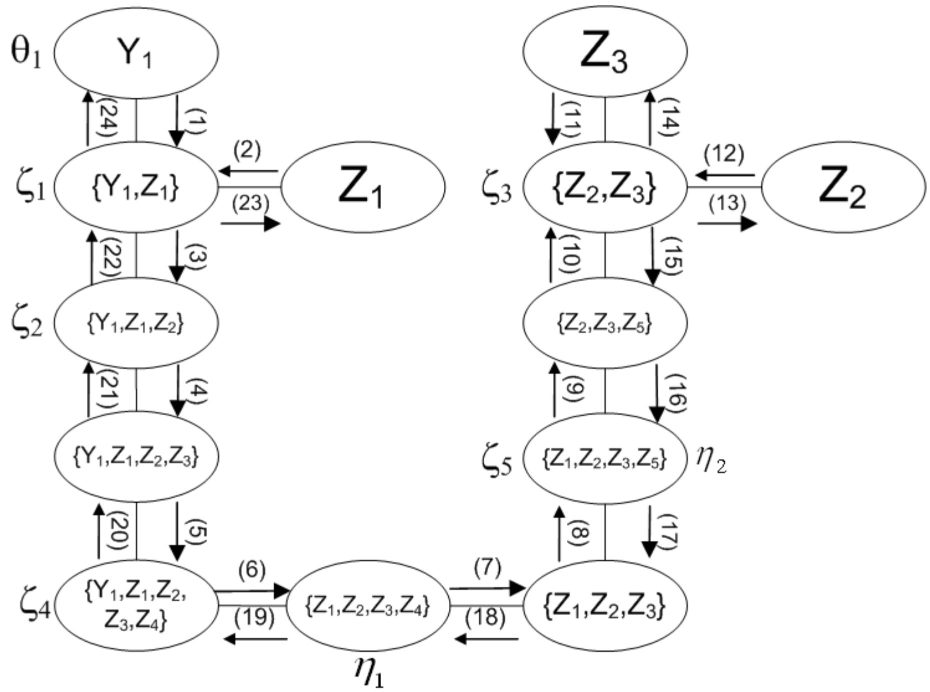


Figure 8: The binary join tree for the Bayesian network in Figure 3.

(6) $\{Y_1, Z_1, Z_2, Z_3, Z_4\}$ to $\{Z_1, Z_2, Z_3, Z_4\}$

The computation of this message is demonstrated in Example 20 and results in the mixed potential ζ_8 for $\{Z_1, Z_2, Z_3, Z_4\}$.

(7) $\{Z_1, Z_2, Z_3, Z_4\}$ to $\{Z_1, Z_2, Z_3\}$

Computation of this message results in the following mixed potential for $\{Z_1, Z_2, Z_3\}$:

$$\zeta_9 = (1, (\phi_{5,A} \otimes \mathcal{M}_{6,A}) + (\phi_{5,B} \otimes \mathcal{M}_{6,B})) \quad .$$

Example 15 illustrates the calculation of the density part of this potential.

(9) $\{Z_1, Z_2, Z_3, Z_5\}$ to $\{Z_2, Z_3, Z_5\}$

Removing Z_1 from the combination of the potentials ζ_5, ζ_9, and η_2 by applying Definition 16 results in the mixed potential $\zeta_{10} = (1, (\phi_{6,A} \otimes \mathcal{M}_{8,A}) + (\phi_{6,B} \otimes \mathcal{M}_{8,B}))$ as calculated in Example 17.

(10) $\{Z_2, Z_3, Z_5\}$ to $\{Z_2, Z_3\}$

This message is also calculated (as shown in Example 16) according to Definition 16 and results in the mixed potential

$$\zeta_{11} = \zeta_{10}^{-Z_5} = (1, (\phi_{7,A} \otimes \mathcal{M}_{10,A}) + (\phi_{7,B} \otimes \mathcal{M}_{10,B})) \quad .$$

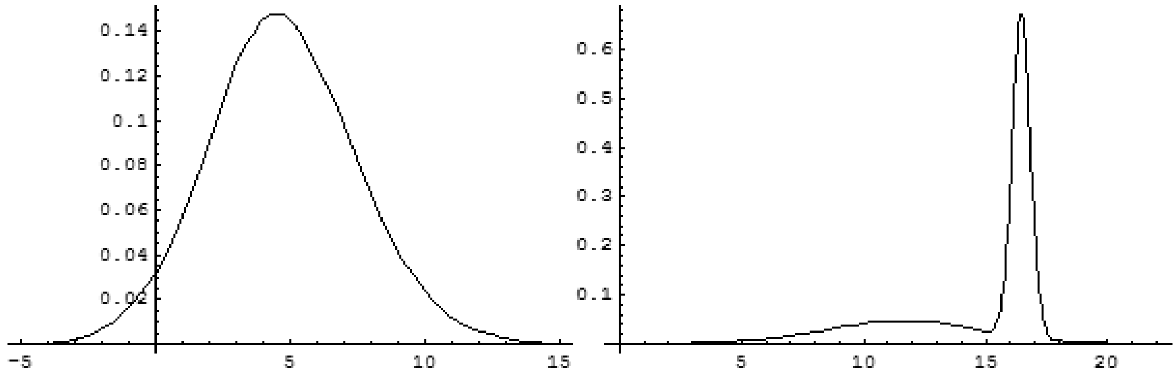


Figure 9: The posterior distributions for Z_2 (left) and Z_3 (right) given the evidence for the Bayesian network in Figure 3.

6.2 Posterior Marginals

Posterior marginal distributions for Z_2 and Z_3 are calculated from messages (13) and (14), respectively. Message (13) removes Z_3 from the combination of the potentials ζ_3 and ζ_{11} for $\{Z_2, Z_3\}$. The result is a density potential for Z_3 which has a mean and variance of 4.7353 and 7.5681, respectively. This compares to a mean of 4.7338 and a variance of 7.6715 calculated using Hugin software. Message (14) removes Z_2 from the combination of the potentials ζ_3 and ζ_{11}. The result is a density potential for Z_2 which has a mean and variance of 14.6266 and 9.4264, respectively. This compares to a mean of 14.6291 and a variance of 9.5006 calculated using Hugin software. The posterior distributions for Z_2 and Z_3 given the evidence are shown graphically in Figure 9.

Completing messages (15) through (24) gives the posterior marginal distributions for Y_1 and Z_1. The posterior marginal distribution for Z_1 has a mean and variance of 2.4650 and 0.9315, respectively. This compares to a mean of 2.4645 and a variance of 0.9344 calculated in Hugin. The posterior proabilities for the discrete variable Y_1 are calculated as $P(Y_1 = A) = 0.3727$ and $P(Y_1 = B) = 0.6273$ using the algorithm in this paper, and compare to $P(Y_1 = A) = 0.3720$ and $P(Y_1 = B) = 0.6280$ calculated in Hugin.

The calculations in Examples 22 and 23 are required to calculate marginal distributions for Y_1, Z_1, and Z_2 when evidence exists that Z_3, Z_4, and Z_5 are limited to values 15.5, 76.5, and 78, respectively. As we can observe from the examples, the continuous variable Z_1 is now limited to the values 2.920 and 0.625. The probabilities of observing these two values are 0.4184 and 0.5816, respectively, which means the variable Z_1 has a mean of 1.5854 and a variance of 1.2822. This compares to a mean of 1.6372 and a variance of 1.2989 calculated in Hugin.

The examples presented previously have included normal distributions assigned to continuous random variables. This has been done to compare the results of the propagation algorithm with those using the method of Lauritzen and Jensen (2001), which is programmed in Hugin software. The numerical representation of the normal PDF's is approximate, but the propagation algorithm is exact. The only errors that occur in the marginal distributions are due to the approximation of the normal PDF's with MTE potentials.

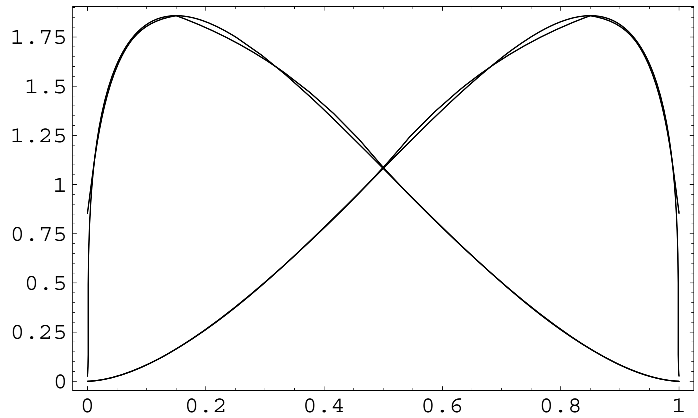


Figure 10: The MTE potential approximations to the beta PDF's for Z_1 given Y_1 in the revised example, overlayed on the actual $Beta(1.3, 2.7)$ and $Beta(2.7, 1.3)$ PDF's.

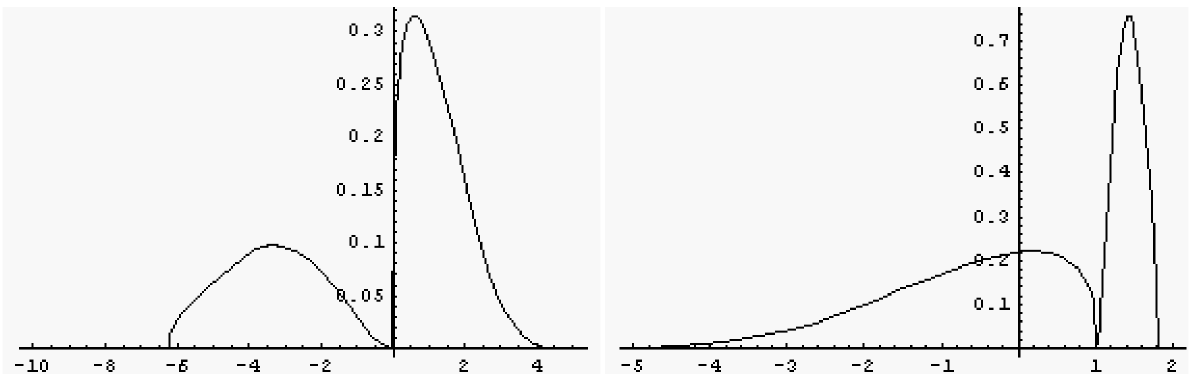


Figure 11: The posterior marginal distributions for Z_2 (left) and Z_3 (right) in the non-Gaussian example.

6.3 Non-Gaussian distributions

MTE potenteials can be used to approximate any PDF. Suppose the PDF's for Z_1 given Y_1 in the hybrid Bayesian network are beta PDF's with parameters as follows: $Z_1 \mid Y_1 = A \sim Beta(1.3, 2.7)$ and $Z_1 \mid Y_1 = B \sim Beta(2.7, 1.3)$. These beta PDF's are approximated by MTE potentials using the method in (Cobb *et al.* 2006). These PDF's are shown graphically in Figure 10.

Suppose evidence now exists that $Z_4 = 4$ and $Z_5 = 3$ and we perform the same propagation as in the previous example. The posterior marginal distributions for Z_2 and Z_3 that result are shown graphically in Figure 11. The posterior marginal distribution for Z_2 has a mean and variance of -0.4898 and 6.0547, respectively. The posterior marginal distribution for Z_3 has a mean and variance of 0.0684 and 1.9606, respectively.

7 Conclusion

The main contribution of this paper is to extend exact inference in hybrid Bayesian networks in which continuous variables may have any conditional density functions (not necessarily conditional linear Gaussian distributions), discrete variables may have continuous parents, and we may have conditionally deterministic continuous variables that are linearly dependent on their continuous parents. The scheme uses a mixed distribution representation of potentials and derives operations from the method of convolutions in probability theory to determine distributions for linear functions of random variables. MTE potentials are used to approximate probability density functions in the representation so that probability density functions can be easily marginalized (without resorting to computationally expensive numerical integration). The Shenoy-Shafer architecture is used to calculate marginals. One advantage of our approach is that we can marginalize variables in any order. This is in contrast with the Lauritzen and Jensen (2001) algorithm in which continuous variables have to marginalized before discrete ones. This restriction can lead to very large cliques.

Two practical limitations of our algorithm are as follows. As in Lauritzen and Jensen (2001), the conditionally deterministic variables must be described by a *linear* function of its continuous parents and the conditional probability density functions for continuous variables must be approximated by MTE potentials. Cobb and Shenoy (2005a) show how the first limitation can be overcome by approximating nonlinear functions by piecewise linear functions. Cobb *et al.* (2006) describes an algorithm for approximating any probability density function by MTE potentials.

Future research in this area will focus on determining the complexity of operations with MTE potentials and exploring the advantages and disadvantages of the MTE representation as compared to other approaches, e.g. mixtures of Gaussians.

References

[1] Cobb, B.R. and P.P. Shenoy (2005a), “Nonlinear Deterministic Relationships in Bayesian Networks,” in L. Godo (ed.), *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Lecture Notes in Artificial Intelligence, 3571, 27–38.

[2] Cobb, B.R. and P.P. Shenoy (2005b), “Hybrid Bayesian networks with linear deterministic variables,” in F. Bacchus and T. Jaakkola (eds.), *Uncertainty in Artificial Intelligence*, 21, 136–144.

[3] Cobb, B.R. and P.P. Shenoy (2006a), “Inference in hybrid Bayesian networks with mixtures of truncated exponentials,”*International Journal of Approximate Reasoning*, 41(3), 257–286.

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

[5] Cobb, B.R., Rumí, R., and P.P. Shenoy (2006), “Approximating Probability Density Functions in hybrid Bayesian networks with Mixtures of Truncated Exponentials,” *Statistics and Computing*, 16(3), 293–308.

[6] Hogg, R.V. and A.T. Craig (1978), *Introduction to Mathematical Statistics*, New York, Macmillan.

[7] Lauritzen, S.L. (1992), "Propagation of probabilities, means and variances in mixed graphical association models," *Journal of the American Statistical Association*, 87, 1098–1108.

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

[9] Mao ,Y., Kschischang, F.R. and B.J. Frey (2004), "Convolutional factor graphs as probabilistic models," in M. Chickering and J. Halpern (eds.), *Uncertainty in Artificial Intelligence*, 20, 374–381.

[10] Moral, S., Rumí, R. and A. Salmerón (2001), "Mixtures of truncated exponentials in hybrid Bayesian networks," in P. Besnard and S. Benferhart (eds.), *Symbolic and Quantitative Approaches to Reasoning under Uncertainty*, Lecture Notes in Artificial Intelligence, 2143, 156–167.

[11] Rumí, R. (2003), *Modelos de Redes Bayesianas con Variables Discretas y Continuas*, Doctoral Thesis, Universidad de Almería, Departamento de Estadística y Matemática Aplicada, Almería, Spain.

[12] Shenoy, P.P. and G. Shafer (1990), "Axioms for probability and belief function propagation," in R.D. Shachter, T.S. Levitt, J.F. Lemmer and L.N. Kanal (eds.), *Uncertainty in Artificial Intelligence*, 4, 169–198.

[13] Spiegelhalter, D.J.,Dawid, A.P.,Lauritzen, S.L. and R.G. Cowell (1993), "Bayesian analysis in expert systems," *Statistical Science*, 8(3) 219–283.

Appendix

Proof of Proposition 1. Follows directly from Definition 8.

Proof of Proposition 2. Follows directly from Definition 9.

Proof of Proposition 3. When creating $\mathbf{M}^+$ to marginalize Z_i, we place coefficients on Z_i in the first column and coefficients on Z_j in the second column, since we are free to choose any order for the remaining variables. The matrix in the resulting deterministic potential, $\omega' \cdot \mathbf{R}'$, already has coefficients on Z_j in the first column so $\mathbf{R}' = \mathbf{R}'^+$, which has $RREF(\mathbf{R}') = \mathbf{R}'$ with $det\,(\mathbf{P}') = 1$, and $(\mathbf{R}')'$ has the same coefficients for $\mathbf{Z} \setminus \{Z_i, Z_j\}$ as $\mathbf{R}'$, which proves the proposition.

Proof of Proposition 4. Follows directly from Definition 10.

Proof of Proposition 5. We are simply summing the potential fragments over the state space of the variables being removed, thus

$$\sum_{y_j \in \Omega_{Y_j}} \left(\sum_{y_i \in \Omega_{Y_i}} (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z}) \right) = \sum_{y_i \in \Omega_{Y_i}} \left(\sum_{y_j \in \Omega_{Y_j}} (\varphi_{\mathbf{y}} \otimes \mathcal{M}_{\mathbf{y}})(\mathbf{z}) \right) ,$$

so the result holds.

Proof of Proposition 6. We prove the proposition for the case where $\Omega_{Y_1} = \Omega_{Y_2} = \{A, B\}$, $\mathbf{Y}_1 = Y_1$, $\mathbf{Y}_2 = Y_2$, $\mathbf{Y} = \{Y_1, Y_2\}$, and $Y_i = Y_1$. Extension to a domain with a larger number of discrete variables is straightforward. Define component potentials as follows: $\varphi_{1,A}$, $\varphi_{1,B}$, $\varphi_{2,A}$, $\varphi_{2,B}$, $\mathcal{M}_{1,A}$, $\mathcal{M}_{1,B}$, $\mathcal{M}_{2,A}$, and $\mathcal{M}_{2,A}$. The indices A and B refer to Y_1 for φ_1 and $\mathcal{M}_1$ and to Y_2 for φ_2 and $\mathcal{M}_2$.

Consider the expression on the left side of the equality. By Definition 9, this expression can be rewritten as

$$((\varphi_1 \otimes \varphi_2) \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))^{-Y_1} = (\varphi \otimes \mathcal{M})^{-Y_1} \ .$$

Where by Definitions 4 and 8, $\varphi = (\varphi_1 \otimes \varphi_2)$ and $\mathcal{M} = (\mathcal{M}_1 \otimes \mathcal{M}_2)$. The fragments of these potentials are, for instance, $\varphi_{A,A} = \varphi_{1,A} \otimes \varphi_{2,A}$ and $\mathcal{M}_{A,A} = \mathcal{M}_{1,A} \otimes \mathcal{M}_{2,A}$. Applying Definition 11 for the case where $Y_2 = A$, we calculate

$$(\varphi \otimes \mathcal{M})^{-Y_1} = (\varphi_{A,A} \otimes \mathcal{M}_{A,A}) + (\varphi_{B,A} \otimes \mathcal{M}_{B,A}) \,.$$

Now consider the expression on the right side of the equality for the case where $Y_2 = A$. By Definition 11, this expression can be rewritten as:

$$((\varphi_1 \otimes \mathcal{M}_1)^{-Y_1} \otimes (\varphi_2 \otimes \mathcal{M}_2)) = ((\varphi_{1,A} \otimes \mathcal{M}_{1,A}) + (\varphi_{1,B} \otimes \mathcal{M}_{1,B})) \otimes (\varphi_{2A} \otimes \mathcal{M}_{2,A}).$$

Combining this result with $(\varphi_2 \otimes \mathcal{M}_2)$ using Definition 9 for the case where $Y_2 = A$ gives

$$((\varphi \otimes \mathcal{M})^{-Y_1} \otimes (\varphi_2 \otimes \mathcal{M}_2)) = (\varphi_{1,A} \otimes \mathcal{M}_{1,A}) \otimes (\varphi_{2,A} \otimes \mathcal{M}_{2,A}) + (\varphi_{1,B} \otimes \mathcal{M}_{1,B}) \otimes (\varphi_{2,A} \otimes \mathcal{M}_{2,A}).$$

By Definition 9, this can be simplified to

$$((\varphi \otimes \mathcal{M})^{-Y_1} \otimes (\varphi_2 \otimes \mathcal{M}_2)) = (\varphi_{A,A} \otimes \mathcal{M}_{A,A}) \otimes (\varphi_{B,A} \otimes \mathcal{M}_{B,A}) \ ,$$

which proves the result.

Proof of Proposition 7. Follows directly from the proof of Proposition 6. If we always choose the coefficients on Z_i and Z_j as the first two columns of $\mathbf{M}^+$ when marginalizing Z_i, the matrix in the resulting deterministic potential will be in RREF. Thus, the same two equations will be substituted into the potential φ regardless of the order in which Z_i and Z_j are marginalized.

Proof of Proposition 8. According to Definition 9, performing the combination inside the parentheses on the left side of the equation results in $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))$, where $\varphi = (\varphi_1 \otimes \varphi_2)$. Applying Definition 12, results in $(\varphi \otimes (\mathcal{M}_1 \otimes \mathcal{M}_2))^{-Z_1} = (\varphi' \otimes (\mathcal{M}'_1 \otimes \mathcal{M}_2))$.

Performing the marginalization on the right-hand side of the equation according to Definition 12 results in $(\varphi'_1 \otimes \mathcal{M}'_1) \otimes (\varphi \otimes \mathcal{M}_2)$. Applying Definition 9, gives $(\varphi'_1 \otimes \varphi_2) \otimes (\mathcal{M}'_1 \otimes \mathcal{M}_2)$. Since φ_2 does not contain Z_i, the linear equation from the first row of $\mathbf{R}_1$ was not substituted into the terms from φ_2 when they were substituted into φ, thus $\varphi' = \varphi'_1 \otimes \varphi_2$, so $(\varphi'_1 \otimes \varphi_2) \otimes (\mathcal{M}'_1 \otimes \mathcal{M}_2) = (\varphi' \otimes (\mathcal{M}'_1 \otimes \mathcal{M}_2))$, which proves the proposition.

Proof of Theoremr̃efmtetheorem. To calculate φ', instances of $c\cdot$ $\exp\{d\cdot z_i\}$ in φ are replaced by

$$c \cdot exp\left\{d \cdot \left(-a_1^{(1)} z_1 - \cdots - a_{i-1}^{(1)} z_{i-1} - a_{i+1}^{(1)} z_{i+1} - \cdots - a_n^{(1)} z_n + b^{(1)}\right)\right\} \quad .$$

Since all the exponential terms contain linear functions of the independent variables, the result is an MTE potential.

Proof of Proposition 9. Follows directly from Definitions 4, 8, and 14.

Proofs of Propositions 10, 11, 12, and 13 follow directly from the proofs of the other propositions.