Appeared in: International Journal of Approximate Reasoning, 41(3), 257--286.

Inference in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials

Barry R. Cobb [a,*] Prakash P. Shenoy [b]

[a] *Department of Economics and Business, Virginia Military Institute, Lexington, Virginia, 24450, U.S.A*

[b] *School of Business, University of Kansas, 1300 Sunnyside Ave., Summerfield Hall, Lawrence, Kansas, 66045-7585, U.S.A*

Abstract

Mixtures of truncated exponentials (MTE) potentials are an alternative to discretization for solving hybrid Bayesian networks. Any probability density function (PDF) can be approximated with an MTE potential, which can always be marginalized in closed form. This allows propagation to be done exactly using the Shenoy-Shafer architecture for computing marginals, with no restrictions on the construction of a join tree. This paper presents MTE potentials that approximate an arbitrary normal PDF with any mean and a positive variance. The properties of these MTE potentials are presented, along with examples that demonstrate their use in solving hybrid Bayesian networks. Assuming that the joint density exists, MTE potentials can be used for inference in hybrid Bayesian networks that do not fit the restrictive assumptions of the conditional linear Gaussian (CLG) model, such as networks containing discrete nodes with continuous parents.

Key words: Hybrid Bayesian networks, mixtures of truncated exponentials, Shenoy-Shafer architecture, conditional linear Gaussian models.

1 Introduction

Bayesian networks model knowledge about propositions in uncertain domains using graphical and numerical representations [26]. 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.

* Corresponding author.

Email addresses: cobbbr@vmi.edu (Barry R. Cobb), pshenoy@ku.edu (Prakash P. Shenoy).

Hybrid Bayesian networks contain both discrete and continuous conditional probability distributions as numerical inputs. An example of a business domain with continuous and discrete variables is a capital investment project where the outcome of an uncertain continuous variable, such as cash flows or customer demand, affects the probability that a business will invest (a discrete variable).

A commonly used type of hybrid Bayesian network is the conditional linear Gaussian (CLG) model [6,12,13]. In CLG models, the distribution of a continuous variable is a linear Gaussian function of its continuous parents. One limitation of CLG models is that discrete nodes cannot have continuous parents. Lerner *et al.* [16] introduces augmented CLG networks where discrete variables may depend on continuous parents. Several algorithms for performing approximate inference in such networks have been developed (see, e.g., [2,10,15,16,21]).

Discretization of continuous distributions can allow approximate inference in a hybrid Bayesian network without limitations on relationships among continuous and discrete variables. Discretization of continuous chance variables is equivalent to approximating a probability density function (PDF) with mixtures of uniform distributions. Discretization with a small number of states can lead to poor accuracy, while discretization with a large number of states can lead to excessive computational effort. Kozlov and Koller [11] improve discretization accuracy by using a non-uniform partition across all variables represented by a distribution and adjusting the discretization for evidence. However, the increased accuracy requires an iterative algorithm and is still problematic for continuous variables whose posterior marginal PDF can vary widely depending on the evidence for other related variables. A classification tree structure is used by Davies and Moore [7] to construct discrete approximations that may vary depending on the values of the parent variables.

To use general purpose algorithms, such as the Shenoy–Shafer architecture, to compute marginal distributions in hybrid Bayesian networks, the operations of combination and marginalization must always preserve potentials within a closed class of functions that can be integrated in closed form. An alternative to discretization fulfilling this requirement is suggested by Moral *et al.* [18] and Rumí [23]. They propose using mixtures of truncated exponentials (MTE) potentials to approximate PDF's in hybrid Bayesian networks. The main goal of this paper is to describe an implementation of MTE potentials in hybrid Bayesian networks where continuous distributions are conditional linear Gaussian distributions. We demonstrate propagation in such networks using two examples. Also, an MTE solution of an augmented CLG network containing a discrete variable with a continuous parent is presented.

The remainder of this paper is organized as follows. Section 2 introduces notation and definitions used throughout the paper, including a description of the CLG model. Section 3 introduces MTE potentials and defines the properties of MTE approximations for an arbitrary normal PDF. Section 4 reviews the operations required for propagation in hybrid Bayesian networks with MTE potentials using the Shenoy-Shafer architecture. Section 5 contains three examples that demonstrate propagation of MTE potentials. Section 6 summarizes and states some directions for future research.

2 Notation and Definitions

This section contains notation and definitions used throughout the paper.

2.1 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 $\Omega_{\mathbf{X}}$.

MTE probability potentials and discrete probability potentials are denoted by lower-case greek letters, e.g., α, β, γ. Subscripts are used for fragments of MTE potentials or conditional probability tables when different parameters or values are required for each configuration of a variable's discrete parents, e.g., α_1, β_2, γ_3 [1] . Discrete probabilities for a specific element of the state space are denoted as an argument to a discrete potential, e.g. $\delta(0) = P(D = 0)$.

In graphical representations, continuous nodes in hybrid Bayesian networks are represented by double-border ovals, whereas discrete nodes are represented by single-border ovals.

2.2 Conditional Linear Gaussian (CLG) Models

Let X be a continuous node in a hybrid Bayesian network, $\mathbf{Y} = (Y_1, \ldots, Y_d)$ be its discrete parents, and $\mathbf{Z} = (Z_1, \ldots, Z_c)$ be its continuous parents. *Conditional linear Gaussian (CLG)* potentials [6,12,13] in hybrid Bayesian networks have the form

$$\pounds(X \mid \mathbf{y},\, \mathbf{z}) \sim N(w_{\mathbf{y},0} + \sum_{i=1}^{c} w_{\mathbf{y},i} z_i, \sigma^2_{\mathbf{y}}), \qquad (1)$$

where $\mathbf{y}$ and $\mathbf{z}$ are a combination of discrete and continuous states of the parents of X. In this formula, $\sigma^2_{\mathbf{y}} > 0$, $w_{\mathbf{y},0}$ and $w_{\mathbf{y},i}$ are real numbers, and $w_{\mathbf{y},i}$ is defined as the i-th component of a vector of the same dimension as the continuous part $\mathbf{Z}$ of the parent variables. This assumes that the mean of a potential depends linearly on the continuous parent variables and that the variance does not depend on the continuous parent variables. For each configuration of the discrete parents of a variable X, a linear function of the continuous parents is specified

[1] The term *potential* was introduced by Lauritzen and Spiegelhalter [14] to describe conditional probability tables since the values in a conditional probability table do not sum to one, but to the sum of the number of configurations of parent variables. The term *fragment* was introduced by Demirer and Shenoy [8] to describe a portion of a potential that is defined over a subset of the domain of the parent variables

as the mean of the conditional distribution of X given its parents, and a positive real number is specified for the variance of the distribution of X given its parents.

The scheme originally developed by Lauritzen [12] allowed exact computation of means and variances in CLG networks; however, this algorithm did not always compute the exact marginal densitites of continuous variables. A new computational scheme for CLG models was later developed by Lauritzen and Jensen [13]. This scheme allows calculation of full local marginals and also permits conditionally deterministic linear variables, i.e. distributions where $\sigma^2_{\mathbf{y}} = 0$ in (1). To find full local marginals, restrictions are placed on construction and initialization of the junction tree.

The CLG model has the property that for any assignment of values for the discrete variables, the distribution for the continuous variables is multivariate Gaussian. This is because, given an assignment of the discrete variables, the conditional probability distributions for the continuous nodes are simple linear univariate Gaussians. When these simple linear Gaussians are combined, they produce a multivariate Gaussian. The joint distribution of all continuous variables in the network is a mixture of Gaussians. CLG models cannot accommodate continuous random variables whose conditional distribution is not Gaussian.

2.3 Logistic function

CLG models cannot accomodate discrete nodes with continuous parents because of the assumption that the joint distribution is a mixture of Gaussians. One model for representing the conditional distribution of a discrete variable given continuous parents is the *logistic* or *softmax* distribution.

Let A be a discrete variable with $\Omega_A = \{a_1, \ldots, a_m\}$ and let $\mathbf{Z} = \{Z_1, \ldots, Z_k\}$ be its continuous parents. The logistic function is defined as

$$P(A = a_i \mid \mathbf{z}) = \frac{\exp\{g_i + \sum_{n=1}^{k} w_{i,n} z_n\}}{\sum_{j=1}^{m} \exp\{g_j + \sum_{n=1}^{k} w_{j,n} z_n\}}, \tag{2}$$

where the magnitude of $w_{i,n}$ determines the steepness of the threshold and g is the offset from 0. A large magnitude of $w_{i,n}$ corresponds to a hard threshold and a small magnitude of $w_{i,n}$ corresponds to a soft threshold. If a discrete variable has discrete and continuous parents, a different logistic function can be defined for each combination of its discrete parents.

If A is binary with $\Omega_A = \{a_1, a_2\}$ and has continuous parents $\mathbf{Z} = \{Z_1, \ldots, Z_k\}$, the logistic function can be simplified to a *sigmoid* function as follows

$$P(A = a_1 \mid \mathbf{z}) = \frac{1}{1 + \exp\{g + \sum_{n=1}^{k} w_n z_n\}}. \tag{3}$$

Thus, in the binary case, $P(A = a_2 \mid \mathbf{z}) = 1 - P(A = a_1 \mid \mathbf{z})$.

Methods of estimating parameters for logistic functions are discussed by McCullagh and Nelder [17], and Jordan and Jacobs [9].

3 Mixtures of Truncated Exponentials

3.1 Definition

A mixture of truncated exponentials (MTE) [18,23] potential has the following definition.

MTE potential. Let $\mathbf{X}$ be a mixed n-dimensional random variable. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)$ and $\mathbf{Z} = (Z_1, \ldots, Z_c)$ be the discrete and continuous parts of $\mathbf{X}$, respectively, with $c + d = n$. A function $\phi : \Omega_{\mathbf{X}} \mapsto \mathcal{R}^+$ is an MTE potential if one of the next two conditions holds:

(1) The potential ϕ can be written as

$$\phi(\mathbf{x}) = \phi(\mathbf{y}, \mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{ \sum_{j=1}^{d} b_i^{(j)} y_j + \sum_{k=1}^{c} b_i^{(d+k)} z_k \right\} \tag{4}$$

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

(2) There is a partition $\Omega_1, \ldots, \Omega_k$ of $\Omega_{\mathbf{X}}$ verifying that the domain of continuous variables, $\Omega_{\mathbf{Z}}$, is divided into hypercubes, the domain of the discrete variables, $\Omega_{\mathbf{Y}}$, is divided into arbitrary sets, and such that ϕ is defined as

$$\phi(\mathbf{x}) = \phi_i(\mathbf{x}) \qquad \text{if } \mathbf{x} \in \Omega_i, \tag{5}$$

where each $\phi_i, i = 1, ..., k$ can be written in the form of equation (4) (i.e. each ϕ_i is an MTE potential on Ω_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 potential. In this paper, all MTE potentials are equal to zero in unspecified regions.

Moral *et al.* [19] proposes an iterative algorithm based on least squares approximation to estimate MTE potentials from data. Moral *et al.* [20] describes a method to approximate conditional MTE potentials using a mixed tree structure. Cobb *et al.* [3] describes a nonlinear optimization procedure used to fit MTE parameters for approximations to standard PDF's, including the uniform, exponential, gamma, beta, and lognormal distributions. Romero *et al.* [22] introduces a structural learning algorithm for Bayesian networks where the conditional distribution for each variable is an MTE potential.

3.2 MTE Approximation to the Normal PDF

Any continuous PDF can be approximated by an MTE potential. For instance, consider a normally distributed random variable X with mean μ and variance $\sigma^2 > 0$. The PDF for the normal distribution is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\}. \tag{6}$$

A general formulation for a 4-piece, 1-term normalized MTE potential which approximates a normal PDF is as follows:

$$\phi(x) = \begin{cases} \frac{-0.017203}{\sigma} + \frac{0.930964}{\sigma} \exp\{1.27(\frac{x-\mu}{\sigma})\} & \text{if } \mu - 3\sigma \leq x < \mu - \sigma \\ \frac{0.442208}{\sigma} - \frac{0.038452}{\sigma} \exp\{-1.64(\frac{x-\mu}{\sigma})\} & \text{if } \mu - \sigma \leq x < \mu \\ \frac{0.442208}{\sigma} - \frac{0.038452}{\sigma} \exp\{1.64(\frac{x-\mu}{\sigma})\} & \text{if } \mu \leq x < \mu + \sigma \\ \frac{-0.017203}{\sigma} + \frac{0.930964}{\sigma} \exp\{-1.27(\frac{x-\mu}{\sigma})\} & \text{if } \mu + \sigma \leq x < \mu + 3\sigma. \end{cases} \tag{7}$$

In this formulation, the mean, μ, of X may be represented by a linear function of its continuous parents, as in (1). Details of the method used to determine the parameters (constants) for the MTE approximation in (7) and others in this paper are available in [3].

The MTE potential in (7) has the following properties:

(1) $\int_{\mu-3\sigma}^{\mu+3\sigma} \phi(x)\, dx = 1$

(2) $\phi(x) \geq 0$

(3) $\phi(x)$ is symmetric around μ

(4) $\int_{\mu-3\sigma}^{\mu+3\sigma} x \cdot \phi(x) dx = \mu$

(5) $\int_{\mu-3\sigma}^{\mu+3\sigma} (x-\mu)^2 \cdot \phi(x) dx = 0.989532\sigma^2$.

Additionally, areas in the four regions of the MTE potential in (7) are

$$\int_{\mu-3\sigma}^{\mu-\sigma} \phi(x)\, dx = \int_{\mu+\sigma}^{\mu+3\sigma} \phi(x)\, dx = 0.15522,$$

$$\int_{\mu-\sigma}^{\mu} \phi(x)\, dx = \int_{\mu}^{\mu+\sigma} \phi(x)\, dx = 0.344784.$$

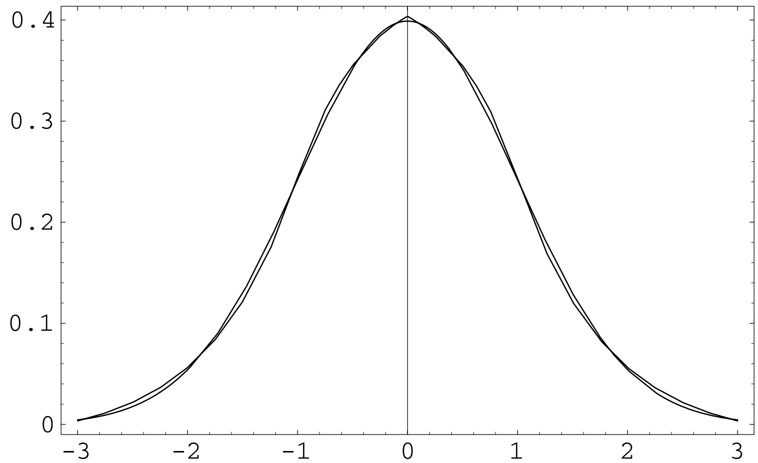

Fig. 1. 4-piece, 1-term MTE approximation overlayed on the standard normal distribution.

The corresponding areas for the normal PDF are

$$\int_{\mu-3\sigma}^{\mu-\sigma} f_X(x)\,dx = \int_{\mu+\sigma}^{\mu+3\sigma} f_X(x)\,dx = 0.157305,$$

$$\int_{\mu-\sigma}^{\mu} f_X(x)\,dx = \int_{\mu}^{\mu+\sigma} f_X(x)\,dx = 0.341345.$$

Area in the extreme tails of the normal distribution is re-assigned to the four regions of the MTE potential in proportion to the areas in the related regions of the normal distribution. Fig. 1 shows a graph of the 4-piece, 1-term MTE approximation overlayed on the actual normal PDF for the case where $\mu = 0$ and $\sigma^2 = 1$ over the domain $[-3, 3]$.

Multiple MTE approximations can be created for a given PDF. For instance, by using three exponential terms in each piece, the normal PDF can be fit with two pieces. A general formulation for a 2-piece, 3-term un-normalized MTE potential which approximates the normal PDF is as follows:

$$\psi'(x) = \begin{cases} \sigma^{-1}(-0.010564 + 197.055720 \exp\{2.2568434(\frac{x-\mu}{\sigma})\} \\ -461.439251 \exp\{2.3434117(\frac{x-\mu}{\sigma})\} \\ +264.793037 \exp\{2.4043270(\frac{x-\mu}{\sigma})\}) & \text{if } \mu - 3\sigma \le x < \mu \\ \sigma^{-1}(-0.010564 + 197.055720 \exp\{-2.2568434(\frac{x-\mu}{\sigma})\} \\ -461.439251 \exp\{-2.3434117(\frac{x-\mu}{\sigma})\} \\ +264.793037 \exp\{-2.4043270(\frac{x-\mu}{\sigma})\}) & \text{if } \mu \le x \le \mu + 3\sigma. \end{cases} \tag{8}$$

The MTE potential in (8) has the following properties:

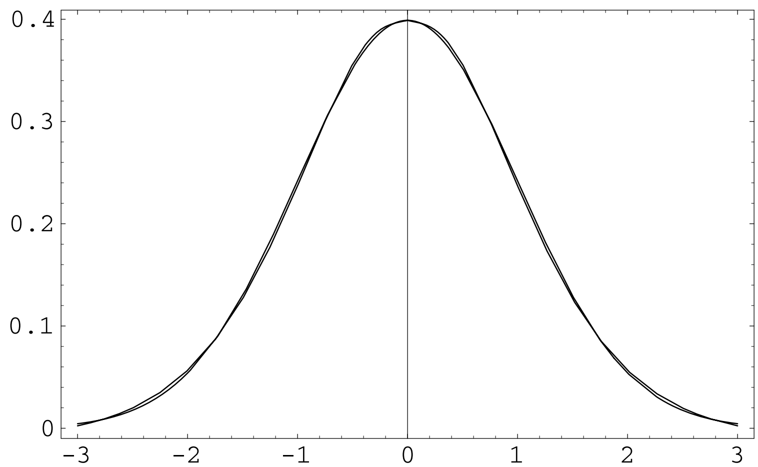

Fig. 2. 2-piece, 3-term MTE approximation overlayed on the standard normal distribution.

(1) $\int_{\mu-3\sigma}^{\mu+3\sigma} \psi'(x)\, dx = \int_{\mu-3\sigma}^{\mu+3\sigma} f_X(x)\, dx = 0.9973$

(2) $\int_{\mu-3\sigma}^{\mu} \psi'(x)\, dx = \int_{\mu}^{\mu+3\sigma} \psi'(x)\, dx = \int_{\mu-3\sigma}^{\mu} f_X(x)\, dx = \int_{\mu}^{\mu+3\sigma} f_X(x)\, dx = 0.49865$

(3) $\psi'(x) \geq 0$

(4) $\psi'(x)$ is symmetric around μ.

A normalized version of the 2-piece, 3-term MTE approximation to the normal PDF is

$$\psi(x) = (1/0.9973) \cdot \psi'(x). \tag{9}$$

The normalized MTE potential in (9) has the following properties:

(1) $\int_{\mu-3\sigma}^{\mu+3\sigma} \psi(x)\, dx = 1$

(2) $\int_{\mu-3\sigma}^{\mu+3\sigma} x \cdot \psi(x) dx = \mu$

(3) $\int_{\mu-3\sigma}^{\mu+3\sigma} (x-\mu)^2 \cdot \psi(x) dx = 0.98187\sigma^2.$

Fig. 2 shows a graph of the 2-piece, 3-term MTE approximation overlayed on the actual normal PDF for the case where $\mu = 0$ and $\sigma^2 = 1$ over the domain $[-3, 3]$. The un-normalized potential in (8) shows the constants obtained using the method in [3]. Normalized potentials, such as those shown in (7) and (9), are used in the initialization phase of the examples later in the paper to solve hybrid Bayesian networks.

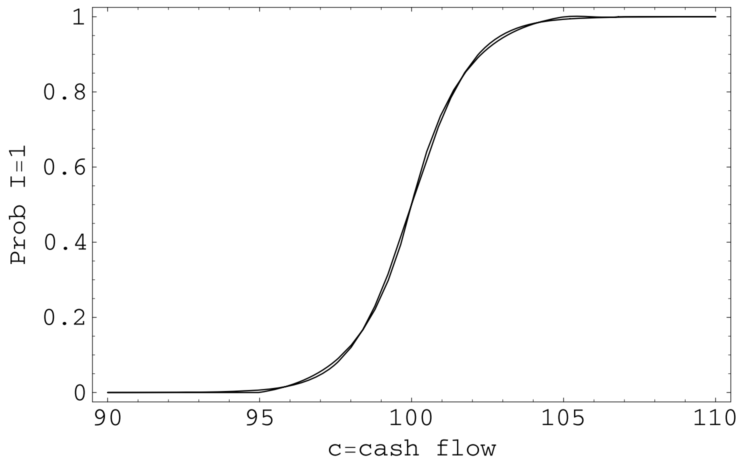

Fig. 3. A sigmoid function with parameters $w = -1$ and $g = 100$ and its MTE approximation.

3.3 MTE Approximation to the Binary Sigmoid Function

Consider the univariate, binary case of the sigmoid function in (3) where g is any real number and $-\infty < w < 0$ for a variable, A, with one continuous parent, Z. The general formulation for a 2-piece MTE potential which approximates this function is

$$\tau(z) = P(A = a_1 \mid z) = \begin{cases} 0 & \text{if } z < \frac{(5-g)}{w} \\ -0.021704 + 0.521704b \exp\{-0.635w(z - g(w+1))\} & \text{if } \frac{(5-g)}{w} \leq z < \frac{(-g)}{w} \\ 1.021704 - 0.521704b^{-1} \exp\{0.635w(z - g(w+1))\} & \text{if } \frac{(-g)}{w} \leq z \leq \frac{(-5-g)}{w} \\ 1 & \text{if } z > \frac{(-5-g)}{w} \end{cases}, \quad (10)$$

where $b = 0.529936^{g(w^2+w+1)}$.

Suppose the probability of a company making an investment in new equipment (I) depends on continuous variable cash flow (C). The investment variable is binary with states $I = 1$ (invest) and $I = 0$ (do not invest). If the company establishes a soft threshold of \$100 in cash flows for making an investment, this can be represented by the parameter $g = 100$ in the sigmoid function. With $w = -1$, the sigmoid function represents a family of discrete conditional probability distributions for $P(I = 1 \mid c)$. This sigmoid function is shown graphically in Fig. 3, overlayed with the corresponding MTE approximation.

4 Propagation in MTE Networks

This section describes the operations required to carry out propagation of MTE potentials in a hybrid Bayesian network. The definitions in this section are described in [18].

4.1 Restriction

Restriction—or entering evidence—involves dropping coordinates to define a potential on a smaller set of variables. During propagation, restriction is performed by substituting values for known variables into the appropriate MTE potentials and simplifying the potentials accordingly.

Let ϕ be an MTE potential for $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$. Assume a set of variables $\mathbf{X}' = \mathbf{Y}' \cup \mathbf{Z}' \subseteq \mathbf{X}$, whose values $\mathbf{x}^{\downarrow \Omega_{\mathbf{X}'}}$ are fixed to values $\mathbf{x}' = (\mathbf{y}', \mathbf{z}')$. The restriction of ϕ to the values $(\mathbf{y}', \mathbf{z}')$ is a new potential defined on $\Omega_{\mathbf{X} \setminus \mathbf{X}'}$ according to the following expression:

$$\phi^{R(\mathbf{X}'=\mathbf{x}')}(\mathbf{w}) = \phi^{R(\mathbf{Y}'=\mathbf{y}', \mathbf{Z}'=\mathbf{z}')}(\mathbf{w}) = \phi(\mathbf{x}) \tag{11}$$

for all $\mathbf{w} \in \Omega_{\mathbf{X} \setminus \mathbf{X}'}$ such that $\mathbf{x} \in \Omega_{\mathbf{X}}$, $\mathbf{x}^{\downarrow \Omega_{\mathbf{X} \setminus \mathbf{X}'}} = \mathbf{w}$ and $\mathbf{x}^{\downarrow \Omega_{\mathbf{X}'}} = \mathbf{x}'$. In this definition, each occurrence of $\mathbf{X}'$ in ϕ is replaced with $\mathbf{x}'$.

Example 1. Consider the following MTE potential:

$$\delta_1(d,e) = \begin{cases} -0.0086015 + 0.465482 \exp\{0.635(d-e-5)\} & \text{if } e-1 \leq d < e+3 \\ 0.221104 - 0.019226 \exp\{-0.82(d-e-5)\} & \text{if } e+3 \leq d < e+5 \\ 0.221104 - 0.019226 \exp\{0.82(d-e-5)\} & \text{if } e+5 \leq d < e+7 \\ -0.0086015 + 0.465482 \exp\{-0.635(d-e-5)\} & \text{if } e+7 \leq d < e+11. \end{cases}$$

If evidence is received that $D = 12$, the restriction operation is used to create a new MTE potential as follows

$$\rho(e) = \delta_1(12,e) = \begin{cases} -0.0086015 + 0.465482 \exp\{0.635(7-e)\} & \text{if } 9 < e \leq 13 \\ 0.221104 - 0.019226 \exp\{-0.82(7-e)\} & \text{if } 7 < e \leq 9 \\ 0.221104 - 0.019226 \exp\{0.82(7-e)\} & \text{if } 5 < e \leq 7 \\ -0.0086015 + 0.465482 \exp\{-0.635(7-e)\} & \text{if } 1 < e \leq 5. \end{cases}$$

4.2 Combination

Combination of MTE potentials is pointwise multiplication. Let ϕ_1 and ϕ_2 be MTE potentials for $\mathbf{X}_1 = \mathbf{Y}_1 \cup \mathbf{Z}_1$ and $\mathbf{X}_2 = \mathbf{Y}_2 \cup \mathbf{Z}_2$. The combination of ϕ_1 and ϕ_2 is a new MTE potential for $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2$ defined as follows

$$\phi(\mathbf{x}) = \phi_1(\mathbf{x}^{\downarrow \mathbf{X}_1}) \cdot \phi_2(\mathbf{x}^{\downarrow \mathbf{X}_2}) \text{ for all } \mathbf{x} \in \Omega_{\mathbf{X}}. \tag{12}$$

Example 2. Consider the following MTE potential:

$$\varepsilon_1(e) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27(e-5)\} & \text{if } 2 \le e < 4 \\ 0.442208 - 0.038452 \exp\{-1.64(e-5)\} & \text{if } 4 \le e < 5 \\ 0.442208 - 0.038452 \exp\{1.64(e-5)\} & \text{if } 5 \le e < 6 \\ -0.017203 + 0.930964 \exp\{-1.27(e-5)\} & \text{if } 6 \le e < 8. \end{cases}$$

The combination of the potential $\rho(e)$ from *Example 1* and $\varepsilon_1(e)$ produces the following MTE potential

$$\varepsilon_1(e) \otimes \rho(e) = \begin{cases} 0.0001 - 0.0001 \exp\{0.635e\} - 0.00001 \exp\{1.27e\} + (8.8844\text{E}-06) \exp\{1.27e\} \\ \qquad \text{if } 2 \le e < 4 \\ -0.0038 + 1.2042 \exp\{-1.64e\} - 0.7649 \exp\{-1.005e\} + 0.0024 \exp\{0.635e\} \\ \qquad \text{if } 4 \le e < 5 \\ 0.0978 - 2.6446 \exp\{-0.82e\} + 0.0001 \exp\{0.82e\} - (2.3351\text{E}-06) \exp\{1.64e\} \\ \qquad \text{if } 5 \le e < 6 \\ -0.0038 - 3187.4407 \exp\{-2.09e\} + 117.8418 \exp\{-1.27e\} + 0.1029 \exp\{-0.82e\} \\ \qquad \text{if } 6 \le e < 7 \\ -0.0038 + 117.8418 \exp\{-1.27e\} - 0.0329 \exp\{-0.45e\} - (1.0633\text{E}-06) \exp\{0.82e\} \\ \qquad \text{if } 7 \le e < 8. \end{cases}$$

When two 4-piece MTE potentials are combined, the result can be an MTE potential with up to 16 pieces. In this example, however, the domains of the two functions being combined only intersect in five regions, so the resulting potential has five pieces.

4.3 Marginalization

Marginalization in a network with MTE potentials corresponds to summing over discrete variables and integrating over continuous variables. Let ϕ be an MTE 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 an MTE potential computed as

$$\phi^{\downarrow \mathbf{X}'}(\mathbf{y}', \mathbf{z}') = \sum_{\mathbf{y} \in \Omega_{\mathbf{Y} \setminus \mathbf{Y}'}} \left(\int_{\Omega_{\mathbf{Z}''}} \phi(\mathbf{y}, \mathbf{z}) \, d\mathbf{z}'' \right) \tag{13}$$

where $\mathbf{z} = (\mathbf{z}', \mathbf{z}'')$, and $(\mathbf{y}', \mathbf{z}') \in \Omega_{\mathbf{X}'}$. Although we show the continuous variables being marginalized before the discrete variables in (13), the variables can be marginalized in any sequence, resulting in the same final MTE potential.

Example 3. Consider the following MTE potential:

$$
\upsilon(g,h) =
\begin{cases}
0.448 - 163.28\ \exp\{-1.64g\} + 59449.91\ \exp\{-1.64(g+h)\} - 163.28\ \exp\{-1.64h\} \\
\hfill \text{if } (4 \le g < 5) \cap (4 \le h < 5) \\
0.448 - 163.28\ \exp\{-1.64g\} - 0.000012\ \exp\{-1.64h\} + 0.0045\ \exp\{-1.64(g-h)\} \\
\hfill \text{if } (4 \le g < 5) \cap (5 \le h < 6) \\
0.448 - 0.000012\ \exp\{1.64g\} + 0.0045\ \exp\{-1.64(g-h)\} - 163.28\ \exp\{-1.64h\} \\
\hfill \text{if } (5 \le g < 6) \cap (4 \le h < 5) \\
0.448 - 0.000012\ \exp\{1.64g\} - 0.000012\ \exp\{-1.64h\} + (3.38\text{E}-10)\ \exp\{1.64(g+h)\} \\
\hfill \text{if } (5 \le g < 6) \cap (5 \le h < 6).
\end{cases}
$$

The marginal of υ for g is

$$
\upsilon^{\downarrow g}(g) = \begin{cases}
0.334835 - 121.912\ \exp\{-1.64g\} & \text{if } 4 \le g < 5 \\
0.442208 - (8.879\text{E}-6)\ \exp\{1.64g\} & \text{if } 5 \le g < 6.
\end{cases}
$$

4.4 Normalization

Let $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$ be a set of variables with discrete and continuous elements, and let ϕ' be an un-normalized MTE potential for $\mathbf{X}$. A normalization constant, K, for ϕ' is calculated as

$$
K = \sum_{\mathbf{y} \in \Omega_{\mathbf{Y}}} \left(\int_{\Omega_{\mathbf{Z}}} \phi'(\mathbf{y}, \mathbf{z})\ d\mathbf{z} \right). \tag{14}
$$

Normalization is implicit in the definition of combination (in the sense that instead of normalizing every time combination is done, we omit it and normalize just once at the end of propagation). When using the Shenoy-Shafer architecture in a hybrid Bayesian network with MTE potentials, a normalization constant, K, is determined by integrating a marginal potential (that is computed at any node in the join tree by combining the input potential with the incoming messages) over the continuous variables in its domain and summing over the discrete variables in its domain. Each marginal potential in the network can then be multiplied by the same constant K^{-1} to obtain a normalized MTE potential. In the case where no evidence is observed, the normalization constant equals one, assuming the join tree is initialized with normalized potentials.

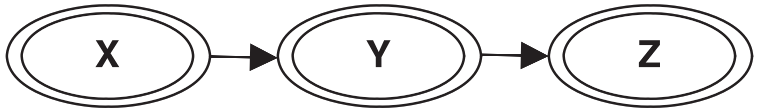

Fig. 4. The Bayesian network for the simple Gaussian example.

4.5 Shenoy-Shafer Architecture

Moral *et al.* [18] shows that the class of MTE potentials is closed under restriction, marginalization, and combination when the domains of MTE potentials are hypercubes. In this paper, we allow the domains of MTE potentials to be linearly dependent on parent variables, such as in the potential δ_1 in Section 4.1. In this case, the result of the marginalization operation may include linear terms in the remaining variables because the limits of integration of some pieces may include linear functions. To ensure that the result of marginalization is an MTE potential, we replace the variable in the linear terms with an MTE approximation. Specifically, for a linear term x defined over the domain $[x_{min}, x_{max}]$, we replace x with

$$(x_{max} - x_{min}) \cdot \left(x_{min} - 13.507018 + 13.512870 \cdot \exp\left\{ \frac{0.071387}{x_{max} - x_{min}} \cdot (x - x_{min}) \right\} \right) . \quad (15)$$

With replacement of the linear terms, MTE potentials as defined in this paper are closed under marginalization. Thus, MTE potentials can be propagated using the Shenoy-Shafer architecture [25], since only restrictions, combinations and marginalizations are perfomed. Normalization involves multiplication of an MTE potential by a real number (the reciprocal of the normalization constant), so this operation is also closed under the class of MTE potentials. In all examples that follow, the Shenoy-Shafer architecture is used for propagation.

5 Examples

5.1 Simple Gaussian Example

A simple Gaussian example [6] is depicted in Fig. 4. This CLG network contains three continuous variables with Gaussian potentials, thus the joint distribution for the three variables is trivariate Gaussian.

5.1.1 Binary Join Tree Initialization

A binary join tree representation [24] of the simple Gaussian example is shown in Fig. 5. The three potentials defined in this example are conditional linear Gaussian distributions.

The probability distribution for X is a standard normal PDF, i.e. $\pounds(X) \sim N(0,1)$. The PDF

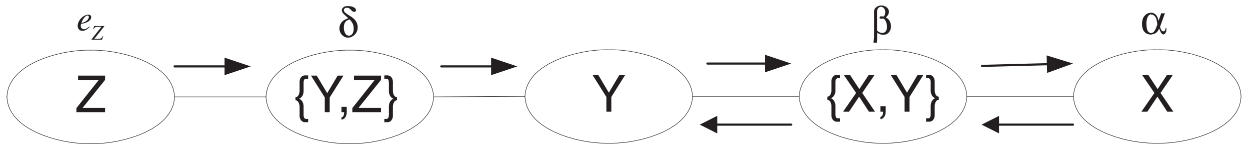

Fig. 5. A binary join tree representation of the simple Gaussian example.

for X is approximated by an MTE potential as follows:

$$\alpha(x) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27x\} & \text{if } -3 \le x < -1 \\ 0.442208 - 0.038452 \exp\{-1.64x\} & \text{if } -1 \le x < 0 \\ 0.442208 - 0.038452 \exp\{1.64x\} & \text{if } 0 \le x < 1 \\ -0.017203 + 0.930964 \exp\{-1.27x\} & \text{if } 1 \le x < 3. \end{cases}$$

The probability distribution for Y is defined as $\pounds(Y \mid x) \sim N(x, 1)$. The conditional PDF for Y is approximated by an MTE potential as follows:

$$\beta(y,x) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27(y-x)\} & \text{if } x-3 \le y < x-1 \\ 0.442208 - 0.038452 \exp\{-1.64(y-x)\} & \text{if } x-1 \le y < x \\ 0.442208 - 0.038452 \exp\{1.64(y-x)\} & \text{if } x \le y < x+1 \\ -0.017203 + 0.930964 \exp\{-1.27(y-x)\} & \text{if } x+1 \le y < x+3. \end{cases}$$

The probability distribution for Z is defined as $\pounds(Z \mid y) \sim N(y, 1)$. The conditional PDF for Z is approximated by an MTE potential as follows:

$$\delta(z,y) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27(z-y)\} & \text{if } y-3 \le z < y-1 \\ 0.442208 - 0.038452 \exp\{-1.64(z-y)\} & \text{if } y-1 \le z < y \\ 0.442208 - 0.038452 \exp\{1.64(z-y)\} & \text{if } y \le z < y+1 \\ -0.017203 + 0.930964 \exp\{-1.27(z-y)\} & \text{if } y+1 \le z < y+3. \end{cases}$$

5.1.2 Entering Evidence

Assume evidence exists that $Z = -1.5$ and define $e_Z = -1.5$. Using the Shenoy-Shafer architecture, the following messages must be passed through the join tree in Fig. 5 to calculate the posterior marginal distributions for variables X and Y:

1) $e_Z = -1.5$ from $\{Z\}$ to $\{Y, Z\}$
2) $(\delta \otimes e_Z)$ from $\{Y, Z\}$ to $\{X, Y\}$
3) $(\delta \otimes e_Z \otimes \beta)^{\downarrow X}$ from $\{X, Y\}$ to $\{X\}$
4) α from $\{X\}$ to $\{X, Y\}$

5) $(\alpha \otimes \beta)^{\downarrow Y}$ from $\{X, Y\}$ to $\{Y\}$.

The details of the required messages are shown below:

(1) $\{Z\}$ `to` $\{Y, Z\}$
The evidence potential $e_Z = -1.5$ is sent from $\{Z\}$ to $\{Y, Z\}$.

(2) $\{Y, Z\}$ `to` $\{X, Y\}$
Given the evidence, the message restricts δ to $\delta(-1.5, y)$, which amounts to substituting $z = -1.5$ into $\delta(z, y)$—an operation denoted as $\eta = (\delta \otimes e_Z)$—then restating the domain in terms of Y as follows:

$$\eta(y) = \begin{cases} -0.017203 + 0.930964 \exp\{1.905 + 1.27y\} & \text{if } -4.5 < y \leq -2.5 \\ 0.442208 - 0.038452 \exp\{-2.46 - 1.64y\} & \text{if } -2.5 < y \leq -1.5 \\ 0.442208 - 0.038452 \exp\{2.46 + 1.64y\} & \text{if } -1.5 < y \leq -0.5 \\ -0.017203 + 0.930964 \exp\{-1.905 - 1.27y\} & \text{if } -0.5 < y \leq 1.5. \end{cases}$$

(3) $\{X, Y\}$ `to` $\{X\}$
The combination $(\eta \otimes \beta)$ is performed, then this message is calculated as

$$\int_{\Omega_Y} (\eta(y) \cdot \beta(y, x))\, dy.$$

(4) $\{X\}$ `to` $\{X, Y\}$
The potential α is sent from $\{X\}$ to $\{X, Y\}$.

(5) $\{X, Y\}$ `to` $\{Y\}$
This message is calculated by performing the combination $(\alpha \otimes \beta)$, then integrating this combination over the domain of X. This results in the un-normalized potential below:

$$\int_{\Omega_X} (\alpha(x) \cdot \beta(y, x))\, dy = \begin{cases} -0.0164 + 2.076 \exp\{1.27y\} & \text{if } x - 3 \leq y < x - 1 \\ 0.4374 - 0.139 \exp\{-1.64y\} & \text{if } x - 1 \leq y < x \\ 0.4374 - 0.139 \exp\{1.64y\} & \text{if } x \leq y < x + 1 \\ -0.0164 + 2.076 \exp\{-1.27y\} & \text{if } x + 1 \leq y < x + 3. \end{cases}$$

5.1.3 Posterior Marginals

(1) `Posterior Marginal for` X
The posterior marginal distribution for X is determined by combining the message sent from $\{X, Y\}$ to $\{X\}$ with the original potential for X. The combination is performed as

$$\xi'(x) = \left(\int_{\Omega_Y} (\eta(y) \cdot \beta(y, x))\, dy \right) \cdot \alpha(x).$$

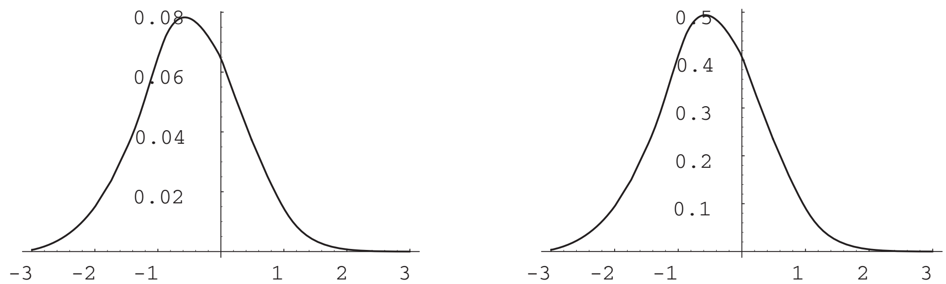

Fig. 6. The un-normalized (left) and normalized (right) posterior distribution for X in the simple Gaussian example.

The un-normalized MTE potential appears in Fig. 6. A normalization constant K is calculated by evaluating the integral in the un-normalized potential over the domain in each region. In this case,

$$K = \int_{\Omega_X} \xi\,'(x)\; dx = 0.1584 \;\; .$$

The normalized posterior MTE potential for X is $\xi(x) = K^{-1} \cdot \xi\,'(x)$. The expected value and variance of this MTE potential are calculated as:

$$E(X) = \int_{\Omega_X} x \cdot \xi(x)\; dx = -0.5000 \;\; ,$$

$$Var(X) = \int_{\Omega_X} (x - E(X))^2 \cdot \xi(x)\; dx = 0.6664 \;\; .$$

These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of -0.5 and 0.6667, respectively. The normalized posterior MTE potential for X is displayed in Fig. 6.

(2) `Posterior Marginal for` Y

The posterior marginal distribution for Y is determined by combining the messages sent to Y on both the inward and outward passes through the binary join tree. This combination is performed as

$$\psi\,'(y) = \eta(y) \cdot \int_{\Omega_X} (\alpha(x) \cdot \beta(y, x))\; dx\,.$$

The same normalization constant, $K = 0.1584$, used to determine the posterior marginal distribution for X can be used to determine the posterior marginal distribution for Y. The normalized posterior MTE potential for Y is $\psi(y) = K^{-1} \cdot \psi\,'(y)$. The expected value and variance of this distribution are calculated as

$$E(Y) = \int_{\Omega_Y} y \cdot \psi(y)\; dy = -1.000 \;\; ,$$

$$Var(Y) = \int_{\Omega_Y} (y - E(Y))^2 \cdot \psi(y)\; dy = 0.6664 \;\; .$$

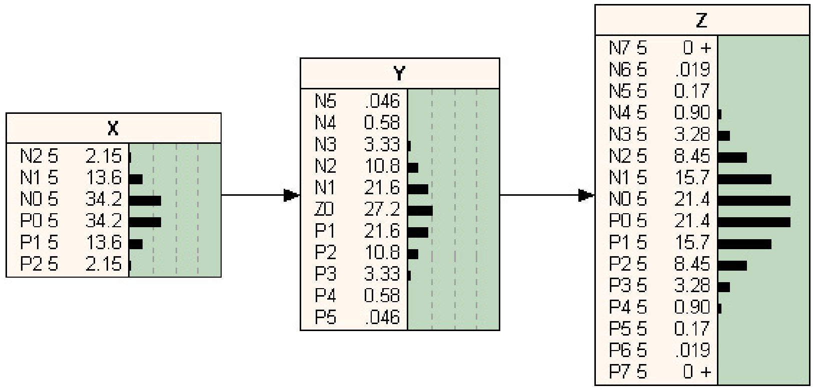

Fig. 7. The marginal distributions created using the discretization in Table 1.

These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of -1.0 and 0.6667, respectively.

5.1.4 A Weakness of Discretization

One alternative for making inferences in the simple Gaussian example is to discretize the three variables. The domain of X in the simple Gaussian example is $[-3, 3]$, considering only area within 3 standard deviations of the mean. Since $\pounds(Y|x) \sim N(x, 1)$ and $\pounds(Z|y) \sim N(y, 1)$, the domain of Y is $[-6, 6]$ and the domain of Z is $[-9, 9]$.

The values for a six-bin discrete approximation to the PDF for X are shown in the body of Table 1, with the probabilities assigned to each bin listed in the column headings. Given the six potential discrete values for X, $\{Y|\ x\}$ can be discretized into 11 values. The discrete approximation of the conditional PDF for Y given each discrete value of X is also shown in Table 1, along with 11 different conditional probability mass functions (PMF's) which include 16 potential discrete values for Z.

The prior discrete approximations can be used to obtain a reasonable approximation of the marginal distributions for the three variables, as shown in Fig. 7, which was produced using Netica software.

Suppose we observe evidence that $Z = -7.5$. Using the discrete approximation in Table 1 and Bayes rule, the posterior marginal distributions for X and Y are degenerate, with $P(X = -2.5) = 1$ and $P(Y = -5) = 1$ (see Fig. 8). This implies that the marginal density for X is uniform from -3 to -2 and the marginal density for Y is uniform from -5.5 to -4.5.

When using the discrete approximation for inference, X is restricted to a uniform density. However, since X actually has a domain of $[z-6, z+6]$, X can take on a value from $[-3, -1.5]$ (considering only area within three standard deviations of the mean). The posterior marginal distribution of X using the model with MTE potentials appears in Fig. 9, which shows that X is clearly not uniform. This distribution has an expected value of -2.5. Thus, discretizing the prior distributions does not guarantee that the posterior distributions given the evidence will provide a reasonable approximation.

Table 1
Discretization of the variables in the simple Gaussian example. Each bin extends from Value−0.5 to Value+0.5.

Value	Bin 1 $P = 0.022$	Bin 2 $P = 0.136$	Bin 3 $P = 0.342$	Bin 4 $P = 0.342$	Bin 5 $P = 0.136$	Bin 6 $P = 0.022$
X	-2.5	-1.5	-0.5	0.5	1.5	2.5
$(Y\| x = -2.5)$	-5	-4	-3	-2	-1	0
$(Y\| x = -1.5)$	-4	-3	-2	-1	0	1
$(Y\| x = -0.5)$	-3	-2	-1	0	1	2
$(Y\| x = 0.5)$	-2	-1	0	1	2	3
$(Y\| x = 1.5)$	-1	0	1	2	3	4
$(Y\| x = 2.5)$	0	1	2	3	4	5
$(Z\| y = -5)$	-7.5	-6.5	-5.5	-4.5	-3.5	-2.5
$(Z\| y = -4)$	-6.5	-5.5	-4.5	-3.5	-2.5	-1.5
$(Z\| y = -3)$	-5.5	-4.5	-3.5	-2.5	-1.5	-0.5
$(Z\| y = -2)$	-4.5	-3.5	-2.5	-1.5	-0.5	0.5
$(Z\| y = -1)$	-3.5	-2.5	-1.5	-0.5	0.5	1.5
$(Z\| y = 0)$	-2.5	-1.5	-0.5	0.5	1.5	2.5
$(Z\| y = 1)$	-1.5	-0.5	0.5	1.5	2.5	3.5
$(Z\| y = 2)$	-0.5	0.5	1.5	2.5	3.5	4.5
$(Z\| y = 3)$	0.5	1.5	2.5	3.5	4.5	5.5
$(Z\| y = 4)$	1.5	2.5	3.5	4.5	5.5	6.5
$(Z\| y = 5)$	2.5	3.5	4.5	5.5	6.5	7.5

5.2 Simple Waste Example

This example is derived from [6] and will provide an example of using MTE potentials for inference in a hybrid Bayesian network. Some parameters have been changed from the original example to make the domains of the MTE potentials easier to interpret, but all relationships between variables are unchanged. The hybrid Bayesian network and join tree for the problem are shown in Figs. 10 and 11, respectively. In this problem, the emission (E) from a waste incinerator differs because of the type of waste (W) and the waste burning regime (B). The filter efficiency (E) depends on the technical state of the electrofilter (F) and the type of waste (W). The emission of dust (D) depends on the type of waste (W), the

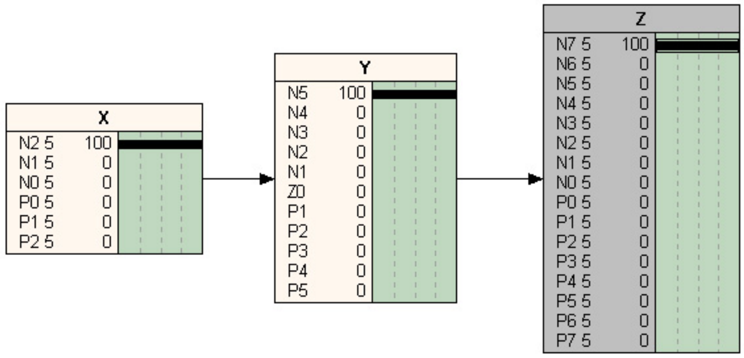

Fig. 8. The marginal distributions given evidence that $Z = -7.5$ created using the discretization in Table 1.

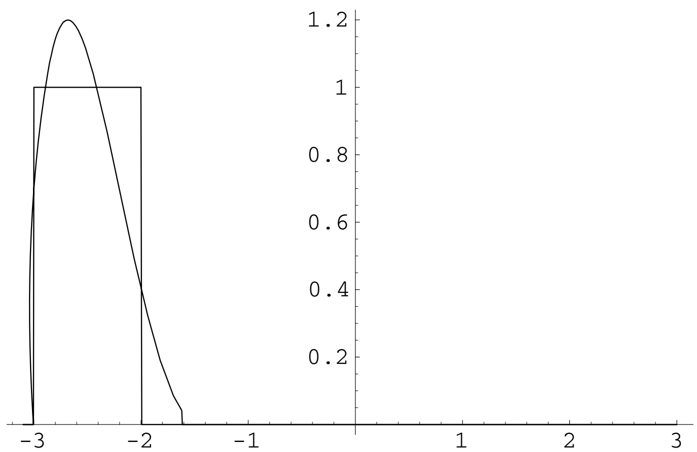

Fig. 9. The posterior marginal distribution for X given $Z = -7.5$ using MTE potentials. The uniform distribution from -3 to -2 represents the result from discretization.

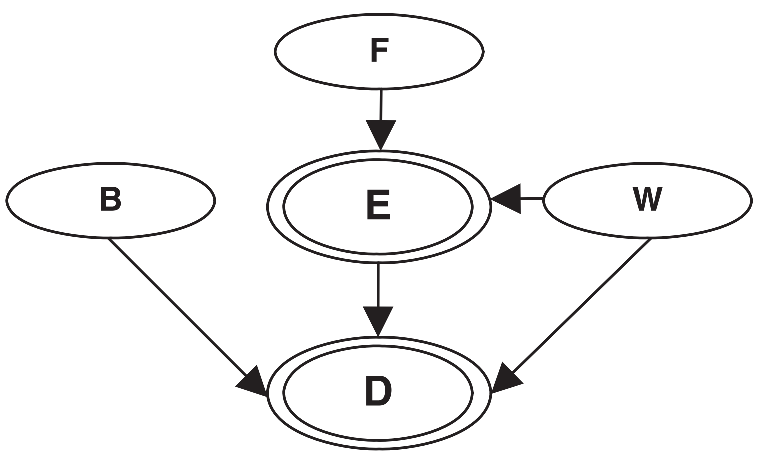

Fig. 10. The hybrid Bayesian network for the simple Waste example.

burning regime (B), and the filter efficiency (E).

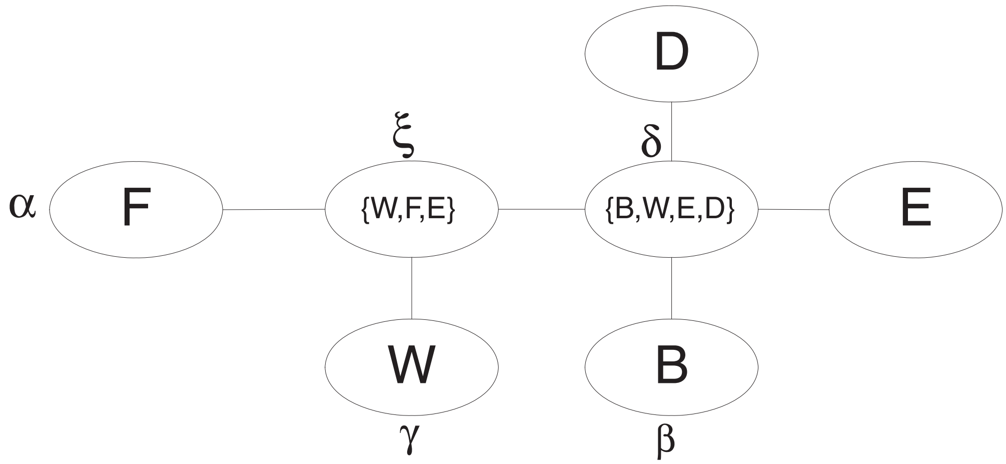

Fig. 11. The join tree for the simple Waste example.

5.2.1 *Definition of Potentials*

The potentials for the five nodes are the probability tables and distributions shown below:

$$\begin{aligned}\alpha(0) = P(F=0) = 0.90, \alpha(1) = P(F=1) = 0.10,\\ \beta(0) = P(B=0) = 0.80, \beta(1) = P(B=1) = 0.20,\\ \gamma(0) = P(W=0) = 0.20, \gamma(1) = P(W=1) = 0.80,\end{aligned}$$

$$\begin{aligned}\pounds(E \mid F=0, W=0) \sim N(\mu_{\varepsilon_1}, \sigma^2_{\varepsilon_1}) \Rightarrow \varepsilon_1(e) \sim N(5,1),\\ \pounds(E \mid F=0, W=1) \sim N(\mu_{\varepsilon_2}, \sigma^2_{\varepsilon_2}) \Rightarrow \varepsilon_2(e) \sim N(8,1),\\ \pounds(E \mid F=1, W=0) \sim N(\mu_{\varepsilon_3}, \sigma^2_{\varepsilon_3}) \Rightarrow \varepsilon_3(e) \sim N(0,1),\\ \pounds(E \mid F=1, W=1) \sim N(\mu_{\varepsilon_4}, \sigma^2_{\varepsilon_4}) \Rightarrow \varepsilon_4(e) \sim N(1,1),\end{aligned}$$

$$\begin{aligned}\pounds(D \mid B=0, W=0, E) \sim N(\mu_{\delta_1}(e), \sigma^2_{\delta_1}) \Rightarrow \delta_1(d, e, B=0, W=0) \sim N(5+e, 4),\\ \pounds(D \mid B=0, W=1, E) \sim N(\mu_{\delta_2}(e), \sigma^2_{\delta_2}) \Rightarrow \delta_2(d, e, B=0, W=1) \sim N(6+e, 4),\\ \pounds(D \mid B=1, W=0, E) \sim N(\mu_{\delta_3}(e), \sigma^2_{\delta_3}) \Rightarrow \delta_3(d, e, B=1, W=0) \sim N(7+e, 4),\\ \pounds(D \mid B=1, W=1, E) \sim N(\mu_{\delta_4}(e), \sigma^2_{\delta_4}) \Rightarrow \delta_4(d, e, B=1, W=1) \sim N(8+e, 4).\end{aligned}$$

The MTE potential fragment δ_1 is defined as follows:

$$\delta_1(d, e, B = 0, W = 0) =$$
$$\begin{cases} -0.005296 + 98.794606 \exp\{1.1284217(d - e - 5)\} \\ -231.344255 \exp\{1.1717059(d - e - 5)\} + 132.754957 \exp\{1.2021635(d - e - 5)\} \\ \qquad\qquad\qquad \text{if } e - 1 \leq d < e + 5 \\ -0.005296 + 98.794606 \exp\{-1.284217(d - e - 5)\} \\ -231.344255 \exp\{-1.1717059(d - e - 5)\} + 132.754957 \exp\{-1.2021635(d - e - 5)\} \\ \qquad\qquad\qquad \text{if } e + 5 \leq d \leq e + 11. \end{cases}$$

The MTE potential fragments $\delta_2(d, e, B = 0, W = 1)$, $\delta_3(d, e, B = 1, W = 0)$, and $\delta_4(d, e, B = 1, W = 1)$ are defined similarly. The fragments $\delta_1, ..., \delta_4$ constitute the potential δ for $\{B, W, E, D\}$. Similarly, four MTE potential fragments $\varepsilon_1, ..., \varepsilon_4$ which are defined according the the 2-piece MTE approximation to the normal PDF in (9) constitute the potential ε for $\{W, F, E\}$.

5.2.2 Propagation

At node $\{W, F, E\}$, the potentials α, γ, and ε are combined as follows:

$$\xi_1(e, F = 0, W = 0) = \alpha(0) \cdot \gamma(0) \cdot \varepsilon_1(e),$$
$$\xi_2(e, F = 0, W = 1) = \alpha(0) \cdot \gamma(1) \cdot \varepsilon_2(e),$$
$$\xi_3(e, F = 1, W = 0) = \alpha(1) \cdot \gamma(0) \cdot \varepsilon_3(e),$$
$$\xi_4(e, F = 1, W = 1) = \alpha(1) \cdot \gamma(1) \cdot \varepsilon_4(e).$$

The fragments $\xi_1, ..., \xi_4$ constitute the potential ξ for $\{W, F, E\}$.

Discrete variable F is deleted by summation from the combination of α, γ, and ε and new MTE potential fragments are determined as follows:

$$\theta_0(e, W = 0) = \xi_1(e, F = 0, W = 0) + \xi_3(e, F = 1, W = 0)\,,$$
$$\theta_1(e, W = 1) = \xi_2(e, F = 0, W = 1) + \xi_4(e, F = 1, W = 1)\,.$$

The potential fragment θ_0 is shown graphically in Fig. 12. This message is sent from $\{W, F, E\}$ to $\{B, W, E, D\}$ in the join tree.

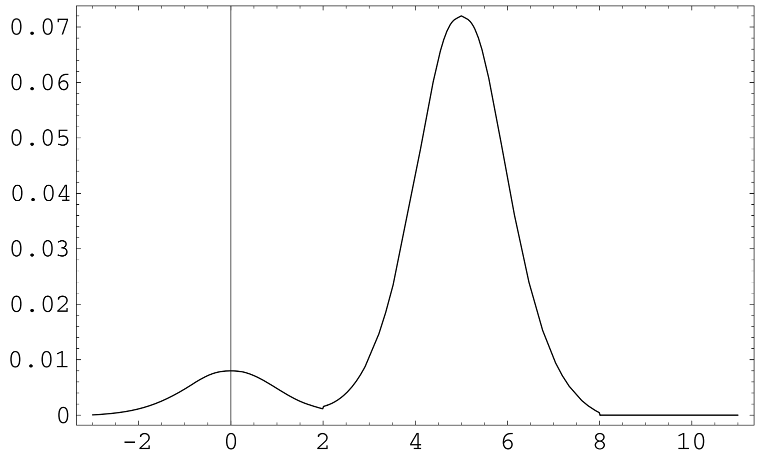

Fig. 12. The potential fragment θ_0 formed by deleting F from $(\alpha \otimes \gamma \otimes \varepsilon)$.

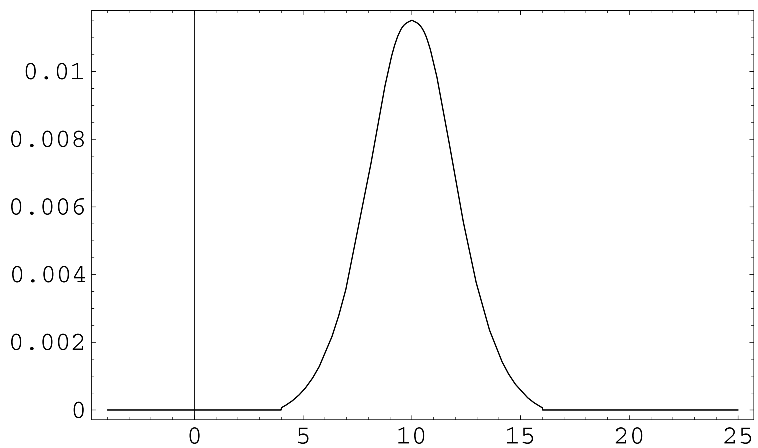

Fig. 13. The potential fragment η_1 formed by combining $\beta(0)$, θ_0 and δ_1 restricted to $E = 5$.

At $\{B, W, E, D\}$ potential fragments $\eta_1, ..., \eta_4$ are determined as follows:

$$\eta_1(d, e, B = 0, W = 0) = \beta(0) \cdot \theta_0(e, W = 0) \cdot \delta_1(d, e, B = 0, W = 0)\,,$$
$$\eta_2(d, e, B = 0, W = 1) = \beta(0) \cdot \theta_1(e, W = 1) \cdot \delta_2(d, e, B = 0, W = 1)\,,$$
$$\eta_3(d, e, B = 1, W = 0) = \beta(1) \cdot \theta_0(e, W = 0) \cdot \delta_3(d, e, B = 1, W = 0)\,,$$
$$\eta_4(d, e, B = 1, W = 1) = \beta(1) \cdot \theta_1(e, W = 1) \cdot \delta_4(d, e, B = 1, W = 1)\,.$$

The potential fragment η_1 is shown graphically in Fig. 13 for a value of $E = 5$. The potential fragments $\eta_1, ..., \eta_4$ constitute the potential η for $\{B, W, E, D\}$.

5.2.3 Posterior Marginals

Prior to calculating the marginal distributions for D and E, integration limits are defined using the parameters from the original potentials in the problem. Although the integration limits can always be set to $-\infty$ and ∞, defining the limits as real numbers facilitates easier

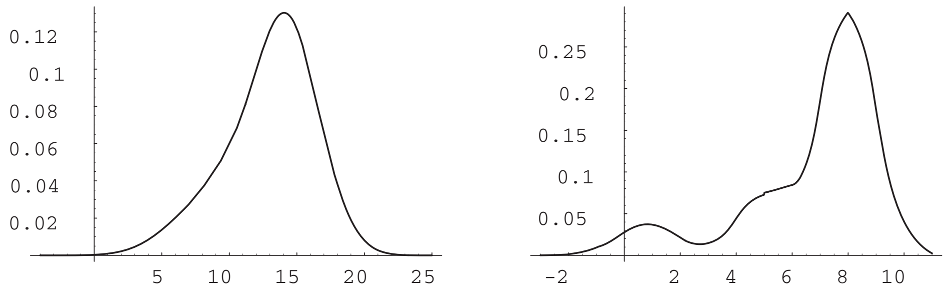

Fig. 14. The marginal potentials for D (left) and E (right) in the simple Waste example.

calculations.

The lower and upper limits of integration for E are denoted by λ_E and κ_E, respectively, and are calculated as follows:

$$\lambda_E = \text{Min}\left\{\mu_{\varepsilon_1} - 3\sigma^2_{\varepsilon_1}, \mu_{\varepsilon_2} - 3\sigma^2_{\varepsilon_2}, \mu_{\varepsilon_3} - 3\sigma^2_{\varepsilon_3}, \mu_{\varepsilon_4} - 3\sigma^2_{\varepsilon_4}\right\} = -3\,,$$
$$\kappa_E = \text{Max}\left\{\mu_{\varepsilon_1} + 3\sigma^2_{\varepsilon_1}, \mu_{\varepsilon_2} + 3\sigma^2_{\varepsilon_2}, \mu_{\varepsilon_3} + 3\sigma^2_{\varepsilon_3}, \mu_{\varepsilon_4} + 3\sigma^2_{\varepsilon_4}\right\} = 11\,,.$$

The lower and upper limits of integration for D are a function of the lower and upper limits of integration for E, since the mean of each potential for D is a linear function of E. The upper and lower limits of integration for D are denoted by λ_D and κ_D, respectively, and are calculated as follows:

$$\lambda_D = \text{Min}\left\{\mu_{\delta_1}(\lambda_E) - 3\sigma^2_{\delta_1}, \mu_{\delta_2}(\lambda_E) - 3\sigma^2_{\delta_2}, \mu_{\delta_3}(\lambda_E) - 3\sigma^2_{\delta_3}, \mu_{\delta_4}(\lambda_E) - 4\sigma^2_{\delta_4}\right\} = -3\,,$$
$$\kappa_D = \text{Max}\left\{\mu_{\delta_1}(\kappa_E) + 3\sigma^2_{\delta_1}, \mu_{\delta_2}(\kappa_E) + 3\sigma^2_{\delta_2}, \mu_{\delta_3}(\kappa_E) + 3\sigma^2_{\delta_3}, \mu_{\delta_4}(\kappa_E) + 3\sigma^2_{\delta_4}\right\} = 25\,.$$

The limits λ_D and κ_D are the lower and upper limits for the entire integration operation. The limits of integration for a given piece of the MTE potential may include linear functions of the remaining variables and require a replacement of a resulting linear term, as explained in Section 4.5.

The marginal potential for D is calculated by removing B and W from η (references to B and W in the domains are omitted) and integrating over E as follows:

$$\varphi(d) = \int_{\lambda_E}^{\kappa_E} \left(\eta_1(d,e) + \eta_2(d,e) + \eta_3(d,e) + \eta_4(d,e)\right)\, de\ \ .$$

This potential is shown graphically in Fig. 14, plotted over the region from λ_D to κ_D.

The following integrals are calculated to verify that $\varphi(d)$ is a PDF for D and to calculate the expected value and variance of the distribution:

$$\int_{\lambda_D}^{\kappa_D} \varphi(d)\; dd = 1.0000\,,$$
$$E(D) = \int_{\lambda_D}^{\kappa_D} d \cdot \varphi(d)\; dd = 12.9402\,,$$
$$Var(D) = \int_{\lambda_D}^{\kappa_D} (d - E(D))^2 \cdot \varphi(d)\; dd = 11.8380\,.$$

These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of 12.9400 and 11.9284, respectively.

The marginal potential for E is calculated by removing B and W from η (references to B and W in the domains are omitted) and integrating over D as follows:

$$\psi(e) = \int_{\lambda_D}^{\kappa_D} \left(\eta_1(d,e) + \eta_2(d,e) + \eta_3(d,e) + \eta_4(d,e)\right)\; dd\,.$$

This potential is shown graphically in Fig. 14, plotted over the region from λ_E to κ_E.

The following integrals are calculated to verify that $\psi(e)$ is a PDF for E and to calculate the expected value and variance of the distribution:

$$\int_{\lambda_E}^{\kappa_E} \psi(e)\; de = 1.0000\,,$$
$$E(E) = \int_{\lambda_E}^{\kappa_E} e \cdot \psi(e)\; de = 6.7403\,,$$
$$Var(E) = \int_{\lambda_E}^{\kappa_E} (e - E(E))^2 \cdot \psi(e)\; de = 6.2143\,.$$

These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of 6.7400 and 6.2324, respectively.

5.2.4 Entering Evidence

Suppose evidence is obtained that $D = 10$. The existing potential fragments for $\{B, W, E, D\}$ are restricted to $\eta_1(10, e, B = 0, W = 0), ..., \eta_4(10, e, B = 1, W = 1)$. These potential fragments are summed (to remove B and W) and then integrated over the domain of E to obtain a normalization constant,

$$K = \int_{\lambda_E}^{\kappa_E} (\eta_1(10,e) + \eta_2(10,e) + \eta_3(10,e) + \eta_4(10,e)) \; de = 0.0594\,.$$

The normalized marginal distribution for E is

$$\vartheta(e) = K^{-1} \cdot (\eta_1(10,e) + \eta_2(10,e) + \eta_3(10,e) + \eta_4(10,e)) \quad .$$

The following integrals are calculated to verify that $\vartheta(e)$ is a PDF for E and to calculate the expected value and variance of the distribution:

$$\int_{\lambda_E}^{\kappa_E} \vartheta(e) \; de = 1.0000\,,$$

$$E(E) = \int_{\lambda_E}^{\kappa_E} e \cdot \vartheta(e) \; de = 5.3078\,,$$

$$Var(E) = \int_{\lambda_E}^{\kappa_E} (e - E(E))^2 \cdot \vartheta(e) \; de = 4.1803\,.$$

These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of 5.2935 and 4.1552, respectively.

To calculate revised marginal probabilities for discrete nodes W and F, the evidence that $D = 10$ is used to restrict the potential δ as follows:

$$\chi_1(e, B=0, W=0) = \delta_1(10, e, B=0, W=0)\,,$$
$$\chi_2(e, B=0, W=1) = \delta_2(10, e, B=0, W=1)\,,$$
$$\chi_3(e, B=1, W=0) = \delta_3(10, e, B=1, W=0)\,,$$
$$\chi_4(e, B=1, W=1) = \delta_4(10, e, B=1, W=1)\,.$$

These revised potential fragments are then sent to node $\{W, F, E\}$ in the join tree after marginalizing B in the following message:

$$\tau_0(e, W=0) = \beta(0) \cdot \chi_1(e, B=0, W=0) + \beta(1) \cdot \chi_3(e, B=1, W=0)\,,$$
$$\tau_1(e, W=1) = \beta(0) \cdot \chi_2(e, B=0, W=1) + \beta(1) \cdot \chi_4(e, B=1, W=1)\,.$$

At node $\{W, F, E\}$ the message is combined with the existing potential as follows:

$$\zeta_1(e, F=0, W=0) = \xi_1(e, F=0, W=0) \cdot \tau_0(e, W=0)\,,$$
$$\zeta_2(e, F=0, W=1) = \xi_2(e, F=0, W=1) \cdot \tau_1(e, W=1)\,,$$
$$\zeta_3(e, F=1, W=0) = \xi_3(e, F=1, W=0) \cdot \tau_0(e, W=0)\,,$$
$$\zeta_4(e, F=1, W=1) = \xi_4(e, F=1, W=1) \cdot \tau_1(e, W=1)\,.$$

To calculate revised probabilities for discrete node F, E must be removed by integration and W and B must be removed by summation. The normalization constant K is still valid. The revised probabilities $P(F=0)$ and $P(F=1)$ are calculated as follows:

$$\varrho(0) = P(F=0) = K^{-1}\left(\int_{\lambda_E}^{\kappa_E} (\zeta_1(e, F=0, W=0) + \zeta_2(e, F=0, W=1))\; de\right) = 0.8700\,,$$
$$\varrho(1) = P(F=1) = K^{-1}\left(\int_{\lambda_E}^{\kappa_E} (\zeta_3(e, F=1, W=0) + \zeta_4(e, F=1, W=1))\; de\right) = 0.1300\,.$$

These answers are comparable with exact results obtained using Hugin software, which calculates revised probabilities $P(F=0) = 0.8691$ and $P(F=1) = 0.1309$.

To calculate revised probabilities for discrete node W, E must be removed by integration and F and B must be removed by summation. The normalization constant K is still valid. The revised probabilities $P(W=0)$ and $P(W=1)$ are calculated as follows:

$$\nu(0) = P(W=0) = K^{-1}\left(\int_{\lambda_E}^{\kappa_E} (\zeta_1(e, F=0, W=0) + \zeta_3(e, F=1, W=0))\; de\right) = 0.5164\,,$$
$$\nu(1) = P(W=1) = K^{-1}\left(\int_{\lambda_E}^{\kappa_E} (\zeta_2(e, F=0, W=1) + \zeta_4(e, F=1, W=1))\; de\right) = 0.4836\,.$$

These answers are comparable with exact results obtained using Hugin software, which calculates revised probabilities $P(W=0) = 0.5156$ and $P(W=1) = 0.4844$.

To calculate revised probabilities for discrete node B, the restricted potential χ is combined with the potential θ sent to $\{B, W, E, D\}$, then E is removed by integration and W is removed by summation from the result. The normalization constant K is still valid. The revised probabilities $P(B=0)$ and $P(B=1)$ (references to B and W in the domains

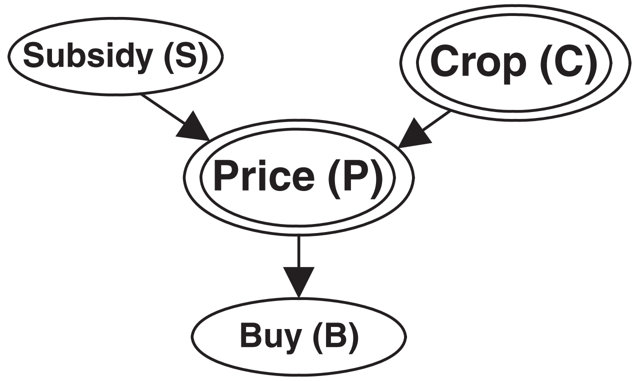

Fig. 15. The hybrid Bayesian network for the Crop example.

omitted) are calculated as follows:

$$\rho_0 = P(B=0) = K^{-1} \cdot \beta(0) \cdot \left(\int_{\lambda_E}^{\kappa_E} (\chi_1(e) \cdot \theta_0(e) + \chi_2(e) \cdot \theta_1(e)) \; de \right) = 0.8682\,,$$

$$\rho_1 = P(B=1) = K^{-1} \cdot \beta(1) \cdot \left(\int_{\lambda_E}^{\kappa_E} (\chi_3(e) \cdot \theta_0(e) + \chi_4(e) \cdot \theta_1(e)) \; de \right) = 0.1318\,.$$

These answers are comparable with exact results obtained using Hugin software, which calculates revised probabilities $P(B=0) = 0.8669$ and $P(B=1) = 0.1331$.

5.3 Crop Network Example

This example is used by Binder *et al.* [1] and Murphy [21] and will provide a simple example of inference using MTE potentials in a hybrid Bayesian network with a discrete child of a continuous parent. A diagram of the hybrid Bayesian network appears in Fig. 15. In this model, the price (P) of a crop is assumed to decrease linearly with the amount of crop (C) produced. If the government subsidizes prices ($S=1$), the price will be raised by a fixed amount. The consumer is likely to buy ($B=1$) if the price drops below a certain amount.

The discrete variable (B) is modeled by a *softmax* or *logistic* distribution, which in the case of a binary discrete variable, reduces to a sigmoid function. For the Crop example, $P(B=0|P=p)$ is given by sigmoid function parameters $w=-1$ and $g=5$ and is approximated by an MTE potential of the form presented in (10). The MTE potential fragment for $B=0$ given price (P) is

$$\beta_0(B=0,p) = P(B=0 \mid P=p) = \begin{cases} 0 & \text{if } p < 0 \\ -0.021704 + 0.021804 \exp\{0.635p\} & \text{if } 0 \le p < 5 \\ 1.021704 - 12.4827 \exp\{-0.635p\} & \text{if } 5 \le p \le 10 \\ 1 & \text{if } p > 10. \end{cases} \quad (16)$$

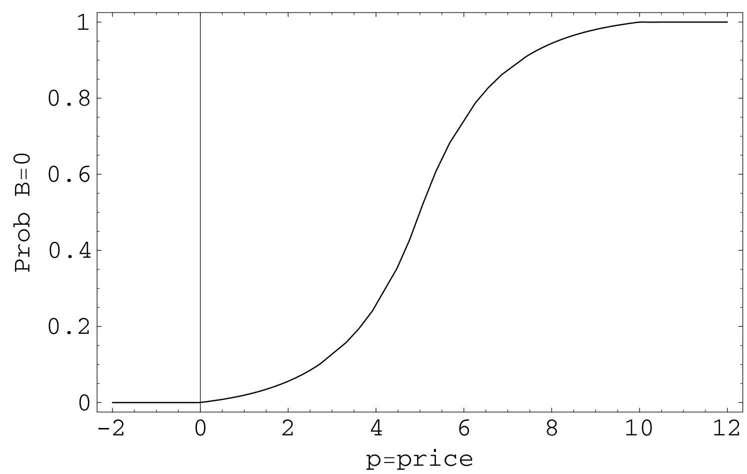

Fig. 16. The MTE approximation to the sigmoid function representing $P(B = 0 \mid P = p)$ in the Crop network.

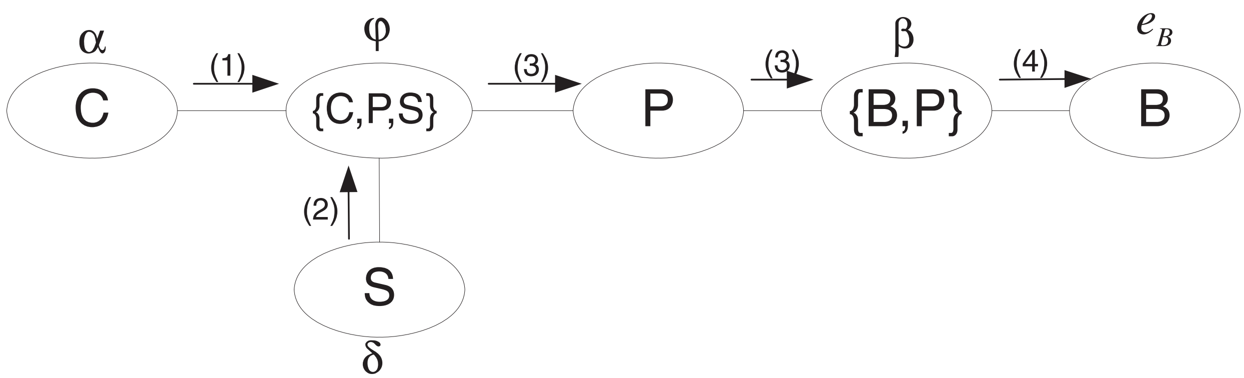

Fig. 17. The join tree for the Crop example.

Since B is binary, $\beta_1(B = 1, p) = P(B = 1 \mid P = p) = 1 - P(B = 0 \mid P = p)$. The MTE potential fragments β_0 and β_1 constitute the potential β for $\{B, P\}$. The MTE potential fragment $\beta_0(B = 0, p)$ is shown graphically in Fig. 16.

5.3.1 Binary Join Tree Initialization

Other algorithms developed for inference in hybrid networks with discrete children of continuous parents place special restrictions on the process of initializing the network. For instance, Lerner *et al.* [16] requires a preprocessing phase where all potentials except those for the discrete children of continuous parents are inserted. The algorithm suggested by Murphy [21] requires any logistic or softmax functions to be converted to Gaussian potentials by using a variational lower bound.

When all probability potentials in the hybrid Bayesian network are approximated with MTE potentials (note that discrete potentials are a special case of the MTE potential), no restrictions are placed on join tree initialization. The join tree is initialized as usual and propagation takes place according to the Shenoy–Shafer architecture.

A join tree for the Crop example is shown in Fig. 17. The potential for $\{B, P\}$ was stated previously. The potential for the Subsidy variable (S) is the following binary discrete distribution:

$$\delta(0) = P(S = 0) = 0.30,\ \delta(1) = P(S = 1) = 0.70.$$

The crop variable (C) follows a normal PDF, $\pounds(C) \sim N(5, 1)$, which is approximated by the following MTE potential:

$$\alpha(c) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27(c-5)\} & \text{if } 2 \le c < 4 \\ 0.442208 - 0.038452 \exp\{-1.64(c-5)\} & \text{if } 4 \le c < 5 \\ 0.442208 - 0.038452 \exp\{1.64(c-5)\} & \text{if } 5 \le c < 6 \\ -0.017203 + 0.930964 \exp\{-1.27(c-5)\} & \text{if } 6 \le c < 8. \end{cases}$$

The price variable (P) decreases linearly with the amount of crop (C) produced and is increased by a fixed amount if the government subsidizes prices ($S = 1$). Thus, $\pounds(P|S = 0, C) \sim N(10 - c, 1)$ and $\pounds(P|S = 1, C) \sim N(20 - c, 1)$, which are represented by the following MTE potential fragments:

$$\varphi_0(p, c, S = 0) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27(p+c-10)\} & \text{if } 7 - c \le p < 9 - c \\ 0.442208 - 0.038452 \exp\{-1.64(p+c-10)\} & \text{if } 9 - c \le p < 10 - c \\ 0.442208 - 0.038452 \exp\{1.64(p+c-10)\} & \text{if } 10 - c \le p < 11 - c \\ -0.017203 + 0.930964 \exp\{-1.27(p+c-10)\} & \text{if } 11 - c \le p < 13 - c, \end{cases}$$

$$\varphi_1(p, c, S = 1) = \begin{cases} -0.017203 + 0.930964 \exp\{1.27(p+c-20)\} & \text{if } 17 - c \le p < 19 - c \\ 0.442208 - 0.038452 \exp\{-1.64(p+c-20)\} & \text{if } 19 - c \le p < 20 - c \\ 0.442208 - 0.038452 \exp\{1.64(p+c-20)\} & \text{if } 20 - c \le p < 21 - c \\ -0.017203 + 0.930964 \exp\{-1.27(p+c-20)\} & \text{if } 21 - c \le p < 23 - c. \end{cases}$$

5.3.2 Computing Messages

The following messages are required to compute the prior marginal distributions for P and B in the Crop example

1) α from $\{C\}$ to $\{C, P, S\}$
2) δ from $\{S\}$ to $\{C, P, S\}$
3) $(\alpha \otimes \delta \otimes \varphi)^{\downarrow P}$ from $\{C, P, S\}$ to $\{P\}$
4) $((\alpha \otimes \delta \otimes \varphi)^{\downarrow P} \otimes \beta)^{\downarrow B}$ from $\{B, P\}$ to $\{B\}$.

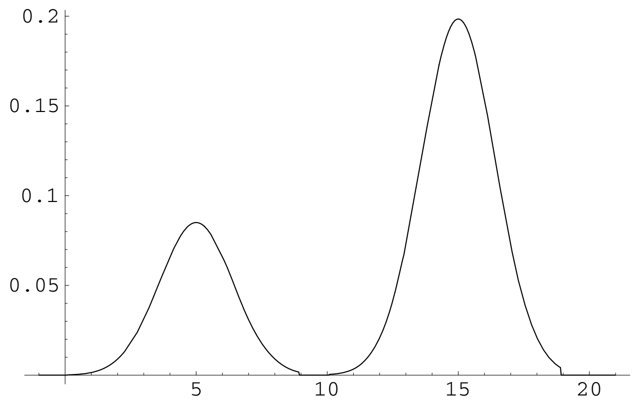

Fig. 18. The prior marginal distribution for P in the Crop example.

5.3.3 Prior Marginals

(1) `Prior Marginal for` P

The message sent from $\{C, P, S\}$ to $\{P\}$ is the prior marginal distribution for P and is calculated as follows:

$$\psi(p) = \int_{\Omega_C} (\alpha(c) \cdot (\delta(0) \cdot \varphi_0(p, c, S = 0) + \delta(1) \cdot \varphi_1(p, c, S = 1)))\, dc.$$

The expected value and variance of the prior marginal distribution for P are calculated as follows:

$$E(P) = \int_{\Omega_P} p \cdot \psi(p)\, dp = 11.9902\,,$$

$$Var(P) = \int_{\Omega_P} (p - E(P))^2 \cdot \psi(p)\, dp = 22.9373\,.$$

These calculations can be verified by using Hugin software to construct a network with variables S, C, and P only. Hugin gives an expected value and variance for P of 12 and 23, respectively. The prior marginal distribution for P is shown graphically in Fig. 18.

(2) `Prior Marginal for` B

To calculate the prior marginal probabilities for B, the prior marginal distribution for P is combined with the MTE potential fragments $\beta_0(B = 0, p)$ and $\beta_1(B = 1, p)$. First, an MTE potential for $\{B = 0, P\}$ is calculated as follows:

$$\Psi_0(B = 0, p) = \beta_0(B = 0, p) \cdot \psi(p).$$

Next, an MTE potential for $\{B = 1, P\}$ is calculated as follows:

$$\Psi_1(B = 1, p) = \beta_1(B = 1, p) \cdot \psi(p).$$

Table 2
Probabilities for the discrete approximation to the prior marginal distribution for P.

Min. Value	Max. Value	p=Mid-point	$P(P = p)$	$P(B = 0 \mid P = p)$	$P(B = 1 \mid P = p)$
-1	1	0	0.0006	0.0067	0.9933
1	3	2	0.0229	0.0474	0.9526
3	5	4	0.1269	0.2689	0.7311
5	7	6	0.1269	0.7311	0.2689
7	9	8	0.0228	0.9526	0.0474
9	11	10	0.0013	0.9933	0.0067
11	13	12	0.0535	0.9991	0.0009
13	15	14	0.2960	0.9999	0.0001
15	17	16	0.2960	0.1000	0.0000
17	19	18	0.0532	1.0000	0.0000
19	21	20	0.0000	1.0000	0.0000

The prior marginal probabilities for B are found by removing P as follows:

$$P(B = 0) = \int_{\Omega_P} \Psi_0(B = 0, p)\, dp = 0.8497\,,$$

$$P(B = 1) = \int_{\Omega_P} \Psi_1(B = 1, p)\, dp = 0.1503\,.$$

To compare these answers with a discrete approximation, an 11-bin discretization was created using the prior marginal distribution for P, $\psi(p)$. This distribution was integrated over 11 regions, with the probabilities obtained listed in Table 2, along with the conditional probabilities for $\{B = 0 \mid p\}$ and $\{B = 1 \mid p\}$ calculated by using the logistic function specified in (16).

By using Bayes rule with the probabilities in Table 2, we can verify that $P(B = 0) \approx 0.85$ and $P(B = 1) \approx 0.15$, consistent with the answers obtained by using MTE potentials.

6 Summary and Conclusions

We have described the details of 2-piece and 4-piece MTE potential approximations to a normal PDF and defined their properties. Inference in three hybrid Bayesian networks using MTE potentials was demonstrated using the Shenoy–Shafer architecture for calculating marginals.

MTE potentials have tremendous promise for use as an inference tool in hybrid Bayesian networks where other methods lead to poor accuracy or high computational expense. MTE potentials can approximate any continuous PDF, so they can be used for inference without limitations on the positioning of continuous and discrete variables within the network. Additionally, other methods proposed for inference in hybrid Bayesian networks with discrete children of continuous parents only calculate the moments of the marginal distributions, whereas the approach presented in this paper can be used to calculate the complete density functions.

Extensive future research on MTE potentials and their applications is needed. General formulations for other continuous PDF's can allow implementation to a broader range of problems. These particularly include distributions that are formed when discrete variables have continuous parents. Such MTE potentials will allow application to specific problems with threshold parameters, such as the business investment project mentioned in Section 1. An adaptation of the EM algorithm for approximating continuous distributions with mixtures of Gaussian distributions may be useful in allowing mixtures of exponentials to be used for approximating probability distributions from data. Another challenge is to automate the algorithm for computing marginals in MTE networks.

Another important class of problems are those that include conditionally deterministic variables (variables that are a deterministic function of their parents). In this case, if some of the parents are continuous, then the joint density function does not exist. CLG distributions can handle such cases when the deterministic function is linear and continuous variables have Gaussian distributions. MTE methods can be used when continuous variables have non-Gaussian distributions [5] and the deterministic functions are non-linear [4].

An alternative to using MTE potentials is using mixtures of Gaussians to represent non-Gaussian distributions. However, the limitations of linear deterministic relationships and no continuous parents of discrete variables pose many problems in this endeavor. Further research is needed to make this approach viable.

Acknowledgments

This research was completed while the first author was a doctoral student at the University of Kansas School of Business and was supported by a research assistantship funded through a grant to the University of Kansas Center for Research, Inc. from Sprint and Nortel Networks. We thank Antonio Salmerón, Rafael Rumí, participants of the WUPES 2003 conference, and three anonymous referees for valuable comments, suggestions, and discussions.

References

[1] J. Binder, D. Koller, S.J. Russell, K. Kanazawa, Adaptive probabilistic networks with hidden variables, Mach. Learn. J. 29 (2-3) (1997) 213–244.

[2] K.C. Chang, Z. Tian, Efficient inference for mixed Bayesian networks, Proceedings of the 5th ISIF/IEEE International Conference on Information Fusion, 2002, pp. 512–519.

[3] B.R. Cobb, R. Rumí, P.P. Shenoy, Approximating probability density functions with mixtures of truncated exponentials, Proceedings of the Tenth Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-04), 2004, pp. 429–436. Available for download at: `http://lark.cc.ku.edu/~pshenoy/Papers/IPMU04.pdf`

[4] B.R. Cobb, P.P. Shenoy, Nonlinear deterministic relationships in Bayesian networks, in: L. Godo (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Artificial Intelligence, Vol. 3571, Springer-Verlag, Heidelberg, 2005, pp. 27–38.

[5] B.R. Cobb, P.P. Shenoy, Hybrid Bayesian networks with linear deterministic variables, in: F. Bacchus, T. Jaakkola (Eds.), Uncertainty in Artificial Intelligence, Vol. 21, 2005, in press.

[6] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer, New York, 1999.

[7] S. Davies, A. Moore, Interpolating conditional density trees, in: A. Darwiche, N. Friedman (Eds.), Uncertainty in Artificial Intelligence, Vol. 18, Morgan Kaufmann, San Francisco, CA, 2002, pp. 119–127.

[8] R. Demirer, P.P. Shenoy, Sequential valuation networks and asymmetric decision problems, Eur. J. Oper. Res. (2005) in press.

[9] M.I. Jordan, R.A. Jacobs, Hierarchical mixtures of experts and the EM algorithm, Neural Comput. 6 (1994) 181–214.

[10] D. Koller, U. Lerner, D. Anguelov, A general algorithm for approximate inference and its application to hybrid Bayes nets, in: K.B. Laskey, H. Prade (Eds.), Uncertainty in Artificial Intelligence, Vol. 15, Morgan Kaufmann, San Francisco, CA, 1999, pp. 324–333.

[11] A.V. Kozlov, D. Koller, Nonuniform dynamic discretization in hybrid networks, in: D. Geiger, P.P. Shenoy (Eds.), Uncertainty in Artificial Intelligence, Vol. 13, Morgan Kaufmann, San Francisco, CA, 1997, pp. 314–325.

[12] S.L. Lauritzen, Propagation of probabilities, means and variances in mixed graphical association models, J. Amer. Stat. Assoc. 87 (1992) 1098–1108.

[13] S.L. Lauritzen, F. Jensen, Stable local computation with conditional Gaussian distributions, Stat. Comput. 11 (2001) 191–203.

[14] S.L. Lauritzen, D.J. Spiegelhalter, Local computations with propabilities on graphical structures and their application to expert systems, J. Roy. Stat. Soc. B. 50 (2) (1988) 157–224.

[15] U. Lerner, R. Parr, Inference in hybrid networks: Theoretical limits and practical algorithms, in: J. Breese, D. Koller (Eds.), Uncertainty in Artificial Intelligence, Vol. 17, Morgan Kaufmann, San Francisco, CA, 2001, pp. 310–318.

[16] U. Lerner, E. Segal, D. Koller, Exact inference in networks with discrete children of continuous parents, in: J. Breese, D. Koller (Eds.), Uncertainty in Artificial Intelligence, Vol. 17, Morgan Kaufmann, San Francisco, CA, 2001, pp. 319–328.

[17] P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman and Hall, London, 1983.

[18] S. Moral, R. Rumí, A. Salmerón, Mixtures of truncated exponentials in hybrid Bayesian networks, in: P. Besnard, S. Benferhart (Eds.), Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Lecture Notes in Artificial Intelligence, Vol. 2143, Springer-Verlag, Heidelberg, 2001, pp. 156–167.

[19] S. Moral, R. Rumí, A. Salmerón, Estimating mixtures of truncated exponentials from data, in: J.A. Gamez, A. Salmerón (Eds.), Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM–02), Cuenca, Spain, 2002, pp. 135–143.

[20] S. Moral, R. Rumí, A. Salmerón, Approximating conditional MTE distributions by means of mixed trees, in: T.D. Nielsen, N.L. Zhang (Eds.), Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Lecture Notes in Artificial Intelligence, Vol. 2711, Springer-Verlag, Heidelberg, 2003, pp. 173–183.

[21] K. Murphy, A variational approximation for Bayesian networks with discrete and continuous latent variables, in: K.B. Laskey, H. Prade (Eds.), Uncertainty in Artificial Intelligence, Vol. 15, Morgan Kaufmann, San Francicso, CA, 1999, pp. 467–475.

[22] V. Romero, R. Rumí, A. Salmerón, Structural learning of Bayesian networks with mixtures of truncated exponentials, in: P. Lucas (Ed.), Proceedings of the Second European Workshop on Probabilistic Graphical Models (PGM–04), Leiden, Netherlands, 2004, pp. 177–184.

[23] R. Rumí, Modelos De Redes Bayesianas Con Variables Discretas Y Continuas, Doctoral Thesis, Universidad de Almería, Departamento de Estadística Y Matemática Aplicada, Almería, Spain, 2003.

[24] P.P. Shenoy, Binary join trees for computing marginals in the Shenoy-Shafer architecture, Int. J. Approx. Reason. 17 (2,3) (1997) 239–263.

[25] P.P. Shenoy, G. Shafer, Axioms for probability and belief function propagation, in: R.D. Shachter, T.S. Levitt, J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence, Vol. 4, North-Holland, Amsterdam, 1990, pp. 169–198.

[26] D.J. Spiegelhalter, A.P. Dawid, S.L Lauritzen, R.G. Cowell, Bayesian analysis in expert systems, Stat. Sci. 8 (3) (1993) 219–283.