In J. Vejnarova (ed.), Proceedings of 6th Workshop on Uncertainty Processing (WUPES-2003), 47--63, VSE-Oeconomica Publishers.

Inference in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials

Barry R. Cobb
School of Business
University of Kansas
brcobb@ku.edu

Prakash P. Shenoy
School of Business
University of Kansas
pshenoy@ku.edu

Abstract

Mixtures of truncated exponentials (MTE) potentials are an alternative to discretization for solving hybrid Bayesian networks. Any probability density function can be approximated with an MTE potential, which can always by 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 a 4-piece MTE potential that approximates an arbitrary normal probability density function with any mean and a positive variance. The properties of this MTE potential are presented, along with examples that demonstrate its 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.

1 Introduction

Hybrid Bayesian networks contain both discrete and continuous conditional probability distributions as numerical inputs. A commonly used type of hybrid Bayesian network is the conditional linear Gaussian (CLG) model [Lauritzen 1992, Cowell *et al.* 1999]. 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.

Discretization of continuous distributions can allow approximate inference in a hybrid Bayesian network without limitations on relationships among continuous and discrete variables. An alternative to discretization is suggested by Moral *et al.* [2001], which proposes using mixtures of truncated exponentials (MTE) potentials to approximate probability density functions 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 an example. 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 an MTE approximation for an arbitrary normal probability density function. Section 4 reviews the operations required for propagation in hybrid Bayesian networks with MTE potentials using the Shenoy-Shafer architecture. Section 5 contains two examples that demonstrate propagation of MTE potentials. Finally, section 6 summarizes and states some directions for future research. This paper is derived from a larger unpublished working paper [Cobb and Shenoy 2003].

2 Notation and Definitions

2.1 Notation

Random variables in a 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 values for those variables. The state space of $\mathbf{X}$ is denoted by $\Omega_{\mathbf{X}}$.

MTE probability distributions or discrete probability distributions are denoted by lower-case greek letters, e.g. α, β, γ. Subscripts are used for MTE potentials when different parameters are required for each configuration of a variable's discrete parents, e.g. α_1, β_2, γ_3. Subscripts are used for discrete probability potentials as an index on the probabilities of elements in the state space, e.g. $\delta_0 = P(D = 0)$.

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 [Lauritzen 1992, Cowell *et al.* 1999] in hybrid Bayesian networks have the form

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

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 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 distribution of all variables in the network is a mixture of Gaussians.

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, $P(A = a_2 \mid \mathbf{z}) = 1 - P(A = a_1 \mid \mathbf{z})$.

3 Mixtures of Truncated Exponentials

3.1 Definitions

A mixture of truncated exponentials (MTE) [Moral *et al.* 2001] 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 $\phi : \Omega_{\mathbf{X}} \mapsto \mathbb{R}^+$ 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).

3.2 MTE Approximation to the Normal PDF

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

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

The general formulation for a 4-piece MTE potential which approximates a normal distribution is

$$\phi(x) = \begin{cases} \frac{-0.017203}{\sigma} + \frac{0.930964}{\sigma} e^{1.27(\frac{x-\mu}{\sigma})} & \text{if } \mu - 3\sigma \le x < \mu - \sigma \\ \frac{0.442208}{\sigma} - \frac{0.038452}{\sigma} e^{-1.64(\frac{x-\mu}{\sigma})} & \text{if } \mu - \sigma \le x < \mu \\ \frac{0.442208}{\sigma} - \frac{0.038452}{\sigma} e^{1.64(\frac{x-\mu}{\sigma})} & \text{if } \mu \le x < \mu + \sigma \\ \frac{-0.017203}{\sigma} + \frac{0.930964}{\sigma} e^{-1.27(\frac{x-\mu}{\sigma})} & \text{if } \mu + \sigma \le x < \mu + 3\sigma \\ 0 & \text{elsewhere.} \end{cases} \tag{7}$$

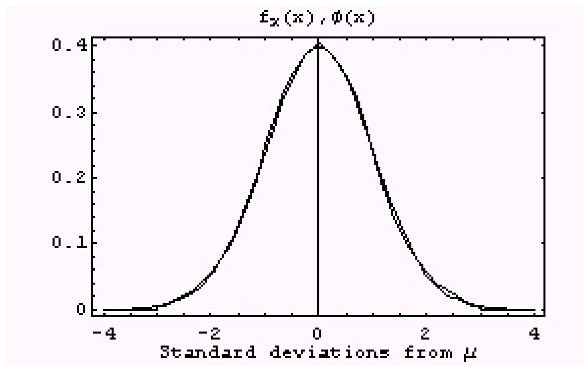

Figure 1: 4-piece MTE approximation overlayed on the standard normal distribution.

In this formulation, the mean, μ, of X may be represented by a linear function of its continuous parents, as in (1).

The MTE potential constructed from the formulation 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. $\int_{\mu-3\sigma}^{\mu+3\sigma} x\phi(x)dx = \mu$
5. $\int_{\mu-3\sigma}^{\mu+3\sigma} (x-\mu)^2\phi(x)dx = 0.989532\sigma^2$.

Areas in the extreme tails of the normal distribution are re-assigned to the four regions of the MTE in proportion to the areas in the related regions of the normal distribution. The general MTE distribution is shown in Figure 1, overlayed on a graph of the standard normal distribution over the same region.

4 Propagation in MTE Networks

This section will describe the operations required to carry out propagation of MTE potentials in a hybrid Bayesian network. Most of this material is described in Moral *et al.* [2001].

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 $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$. Assume a set of variables $\mathbf{X}' = (\mathbf{Y}', \mathbf{Z}') \subseteq \mathbf{X}$, 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:

$$\phi^{R(\mathbf{X}'=\mathbf{x}')}(\mathbf{w}) = \phi^{R(\mathbf{Y}'=\mathbf{y}', \mathbf{Z}'=\mathbf{z}')}(\mathbf{w}) = \phi(\mathbf{x}) \tag{8}$$

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 $\mathbf{x}'$.

4.2 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 $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$. The marginal of ϕ for a set of variables $\mathbf{X}' = (\mathbf{Y}', \mathbf{Z}') \subseteq \mathbf{X}$ is an MTE potential computed as

$$\phi^{\downarrow \mathbf{X}'}(\mathbf{y}', \mathbf{z}') = \sum_{\mathbf{y} \in \Omega_{\mathbf{Y} \setminus \mathbf{Y}'}} \Big(\int_{\Omega_{\mathbf{Z} \setminus \mathbf{Z}'}} \phi(\mathbf{y}, \mathbf{z}) \, d\mathbf{z}'' \Big) \tag{9}$$

where $\mathbf{z}'' = \mathbf{z} \setminus \mathbf{z}'$.

4.3 Combination

Combination of MTE potentials is pointwise multiplication. Let ϕ_1 and ϕ_2 be MTE potentials for $\mathbf{X}_1 = (\mathbf{Y}_1, \mathbf{Z}_1)$ and $\mathbf{X}_2 = (\mathbf{Y}_2, \mathbf{Z}_2)$. 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

$$\phi(\mathbf{x}) = \phi(\mathbf{x}^{\downarrow \Omega_{\mathbf{x}_1}}) \cdot \phi(\mathbf{x}^{\downarrow \Omega_{\mathbf{x}_2}}) \text{ for all } \mathbf{x} \in \Omega_{\mathbf{X}}. \tag{10}$$

4.4 Normalization

Let $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$ 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

$$K = \sum_{\mathbf{y} \in \Omega_{\mathbf{Y}}} \Big(\int_{\Omega_{\mathbf{Z}}} \phi'(\mathbf{y}, \mathbf{z}) \, d\mathbf{z} \Big). \tag{11}$$

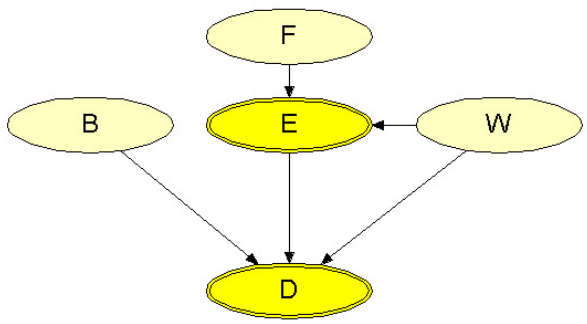

Figure 2: The hybrid Bayesian network for the simple Waste example.

4.5 Shenoy-Shafer Architecture

Moral *et al.* [2001] shows that the class of MTE potentials is closed under restriction, marginalization, and combination. Thus, MTE potentials can be propagated using the Shenoy-Shafer architecture [Shenoy and Shafer 1990], since only 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 Waste Example

This example is derived from Cowell *et al.* [1999] 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 binary join tree [Shenoy 1997] for the problem are shown in Figures 2 and 3, respectively. Expected value and variance calculations in this example were verified with Hugin software.

5.1.1 Definition of Potentials

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

$$\alpha_0 = P(F = 0) = 0.90, \alpha_1 = P(F = 1) = 0.10,$$
$$\beta_0 = P(B = 0) = 0.80, \beta_1 = P(B = 1) = 0.20,$$

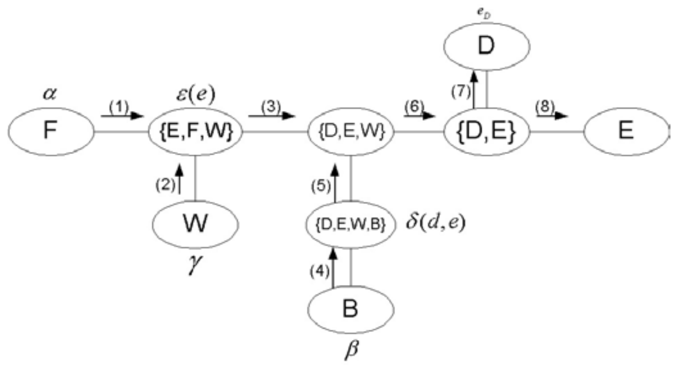

Figure 3: The binary join tree for the simple Waste example.

$$\gamma_0 = P(W=0) = 0.20, \gamma_1 = P(W=1) = 0.80,$$

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

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

Each of the potentials are translated to MTE distributions, the potentials. For instance, ε_1 and δ_1 are defined as follows

$$\varepsilon_1(e) = \begin{cases} -0.017203 + 0.930964e^{1.27(e-5)} & \text{if } 2 \le e < 4\\ 0.442208 - 0.038452e^{-1.64(e-5)} & \text{if } 4 \le e < 5\\ 0.442208 - 0.038452e^{1.64(e-5)} & \text{if } 5 \le e < 6\\ -0.017203 + 0.930964e^{-1.27(e-5)} & \text{if } 6 \le e < 7, \end{cases}$$

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

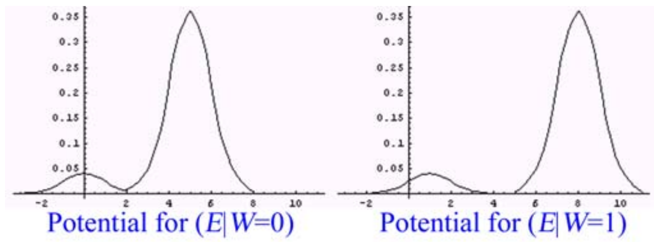

Figure 4: The potentials formed by deleting F from $(\varepsilon \otimes \gamma \otimes \alpha)$.

5.1.2 Computing Messages

The relevant messages in the binary join tree are described below.

1. `Deleting` F `from` $(\varepsilon \otimes \gamma \otimes \alpha)$
   This message is sent from $\{E, F, W\}$ to $\{D, W, E\}$ in the binary join tree and is denoted by (3) in Figure 3. Discrete variable F is deleted by summation from the combination of α and ε and new potentials are determined as follows

$$\xi_0(e) = \alpha_0\varepsilon_1(e) + \alpha_1\varepsilon_3(e)$$

$$\xi_1(e) = \alpha_0\varepsilon_2(e) + \alpha_1\varepsilon_4(e).$$

   These potentials are shown graphically in Figure 4.

2. `Deleting` B `from` $(\delta \otimes \beta)$
   This message is sent from $\{D, B, W, E\}$ to $\{D, W, E\}$ in the binary join tree and is denoted by (5) in Figure 3. Discrete variable B is deleted by summation from the combination of β and δ and new potentials are determined as follows

$$\eta_0(d, e) = \beta_0\delta_1(d) + \beta_1\delta_3(d)$$

$$\eta_1(d, e) = \beta_0\delta_2(d) + \beta_1\delta_4(d).$$

3. `Deleting` W `from` $(\xi \otimes \eta)$
   This message is sent from $\{D, W, E\}$ to $\{D, E\}$ in the binary join tree and is denoted by (6) in Figure 3. The potentials ξ and η calculated previously are combined, then the variable W is removed by summation from the resulting potential. This operation is performed as follows

$$\theta(d, e)) = \gamma_0\ \xi_1(e)\ \eta_0(d, e) + \gamma_1\ \xi_2(e)\ \eta_1(d, e).$$

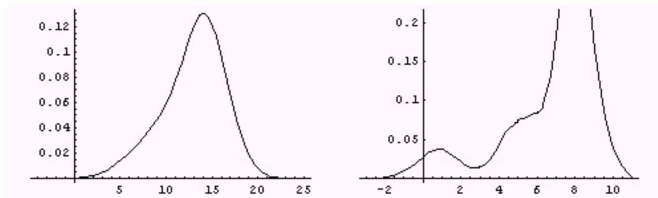

Figure 5: The marginal potentials for D (left) and E (right) in the simple Waste example.

5.1.3 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 calculations.

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

$$\lambda_E = \text{Min}(\mu_{\varepsilon_1} - 3\sigma^2_{\varepsilon_1}, \mu_{\varepsilon_2} - 3\sigma^2_{\varepsilon_2}, \mu_{\varepsilon_3} - 3\sigma^2_{\varepsilon_3}, \mu_{\varepsilon_4} - 3\sigma^2_{\varepsilon_4}) = -3$$

$$\kappa_E = \text{Max}(\mu_{\varepsilon_1} + 3\sigma^2_{\varepsilon_1}, \mu_{\varepsilon_2} + 3\sigma^2_{\varepsilon_2}, \mu_{\varepsilon_3} + 3\sigma^2_{\varepsilon_3}, \mu_{\varepsilon_4} + 3\sigma^2_{\varepsilon_4}) = 11.$$

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 marginal potential for D (shown graphically in Figure 5) is calculated from θ as follows

$$\varphi(d) = \int_{\lambda_E}^{\kappa_E} \theta(d, e)\, de.$$

The expected value and variance of D are calculated as

$$E(D) = \int_{\lambda_D}^{\kappa_D} d\ \varphi(d)\, dd = 12.9401,$$

$$Var(D) = \int_{\lambda_D}^{\kappa_D} (d - E(D))^2 \; \varphi(d) \; dd = 11.876.$$

The marginal potential for E (shown graphically in Figure 5) is calculated from θ as follows

$$\psi(e) = \int_{\lambda_D}^{\kappa_D} \theta(d, e) \; dd.$$

The expected value and variance of E are calculated as

$$E(E) = \int_{\lambda_E}^{\kappa_E} e \; \psi(e) \; de = 6.74014,$$

$$Var(E) = \int_{\lambda_E}^{\kappa_E} (e - E(E))^2 \; \psi(e) \; de = 6.22196.$$

5.1.4 Entering Evidence

Suppose evidence is obtained that $D = 10$. This evidence is sent from node D to node $\{D, E\}$ in the binary join tree, then the existing potential for $\{D, E\}$ is restricted to $\theta(10, e)$. This potential is then integrated over the domain of E to obtain a normalization constant,

$$K = \int_{\lambda_E}^{\kappa_E} \theta(10, e) \; de = 0.0592091,$$

which represents the probability of the observed evidence. The normalized marginal distribution for E is $\vartheta(e) = K^{-1} \; \theta(10, e)$.

The following integrals are calculated to determine the expected value and variance of $\vartheta(e)$

$$E(E) = \int_{\lambda_E}^{\kappa_E} e \; \vartheta(e) \; de = 5.30754,$$

$$Var(E) = \int_{\lambda_E}^{\kappa_E} (e - E(E))^2 \; \vartheta(e) \; de = 4.19925.$$

To calculate revised marginal probabilities for discrete nodes F and W, the evidence that $D = 10$ is sent to node $\{D, E, W\}$ in the join tree, where the existing potentials are restricted, becoming $\eta_0(10, e)$ and $\eta_1(10, e)$. These restricted potentials are sent to node $\{E, F, W\}$ in the join tree and combined with the existing potentials $\varepsilon_1(e), \ldots, \varepsilon_4(e)$.

To calculate revised probabilities for discrete node F, E must be removed by integration and W must be removed by summation. Normalization constant K is still valid at node F, since it represents the probability of the observed evidence. The revised probabilities $P(F = 0)$ and $P(F = 1)$ are calculated as follows

$$\varrho_0 = P(F=0) = K^{-1}\alpha_0\left(\int_{\lambda_E}^{\kappa_E} (\gamma_0\ \varepsilon_1(e)\ \eta_0(10, e) + \gamma_1\ \varepsilon_2(e)\ \eta_1(10, e))\ de\right) = 0.870007,$$

$$\varrho_1 = P(F=1) = K^{-1}\alpha_1\left(\int_{\lambda_E}^{\kappa_E} (\gamma_0\ \varepsilon_3(e)\ \eta_0(10, e) + \gamma_1\ \varepsilon_4(e)\ \eta_1(10, e))\ de\right) = 0.129973.$$

To calculate revised probabilities for discrete node W, E must be removed by integration and F must be removed by summation. Normalization constant K is still valid at node W, since it represents the probability of the observed evidence. The revised probabilities $P(W = 0)$ and $P(W = 1)$ are calculated as follows

$$\nu_0 = P(W=0) = K^{-1}\gamma_0\left(\int_{\lambda_E}^{\kappa_E} (\alpha_0\ \varepsilon_1(e)\ \eta_0(10, e) + \alpha_1\ \varepsilon_3(e)\ \eta_1(10, e))\ de\right) = 0.517355,$$

$$\nu_1 = P(W=1) = K^{-1}\gamma_1\left(\int_{\lambda_E}^{\kappa_E} (\alpha_0\ \varepsilon_2(e)\ \eta_0(10, e) + \alpha_1\ \varepsilon_4(e)\ \eta_1(10, e))\ de\right) = 0.482585.$$

5.2 Crop Network Example

This example is from Murphy [1999] and Binder *et al.* [1997] 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 Bayesian network appears in Figure 6. 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. The probability that $B = 0$ given price (P) equals

$$\beta_0(p) = P(B = 0|P = p) = \frac{1}{1 + e^{wp+g}}. \tag{12}$$

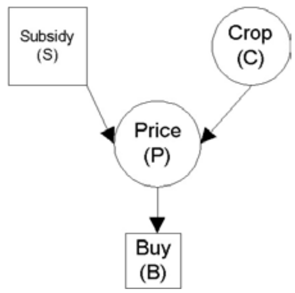

Figure 6: The hybrid Bayesian network for the Crop example.

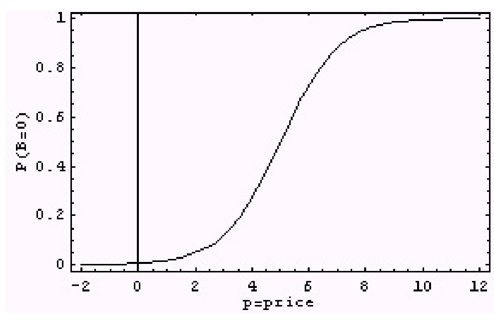

Figure 7: The logistic function representing $P(B = 0|P = p)$ in the Crop network.

For the Crop example, $P(B = 0|P = p)$ is given by parameters $w = -1$ and $g = 5$. Since B is binary, $\beta_1(p) = P(B = 1|P = p) = 1 - P(B = 0|P = p)$. The logistic function for $P(B = 0|P = p)$ is shown graphically in Figure 7.

5.2.1 Binary Join Tree Initialization

When all continuous distributions are approximated with MTE potentials, 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 Figure 8. The potential for $(B \mid p)$ was stated previously. The potentials 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) has a normal distribution, $\pounds(C) \sim N(5, 1)$, which is described by the following MTE potential:

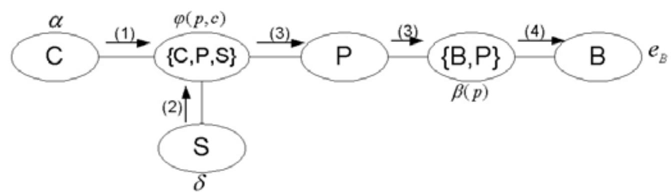

Figure 8: The join tree for the Crop example.

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

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 potentials:

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

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

5.2.2 Computing Messages

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

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

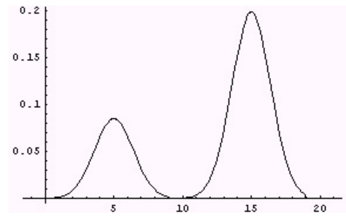

Figure 9: The posterior marginal distribution for P in the Crop example.

5.2.3 Posterior Marginals

1. `Posterior Marginal for` P

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

   $$\psi(p) = \int_c (\alpha(c)(\delta_0 \varphi_0(p,c) + \delta_1 \varphi_1(p,c)))\ dc.$$

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

   $$E(P) = \int_p p\ \psi(p)\ dp = 11.9902,$$

   $$Var(P) = \int_p (p - E(P))^2\ \psi(p)\ dp = 22.9373.$$

   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 posterior marginal distribution for P is shown graphically in Figure 9.

2. `Posterior Marginal for` B
   To calculate the posterior marginal probabilities for B, the marginal distribution for P is combined with the conditional probabilities $\beta_0(p)$ and $\beta_1(p)$. First, the joint distribution of $\{B = 0, P\}$ is calculated as follows

   $$\Psi_0(p) = \beta_0\ \psi(p) = (1 + e^{5-p})^{-1}\ \psi(p).$$

   Next, the joint distribution of $\{B = 1, P\}$ is calculated as

$$\Psi_1(p) = \beta_1\ \psi(p) = (1 - (1 + e^{5-p})^{-1})\ \psi(p).$$

The posterior probabilities for B are found by removing P as

$$P(B = 0) = \int_p \Psi_0(p)\ dp = 0.849586,$$

$$P(B = 1) = \int_p \Psi_1(p)\ dp = 0.150414.$$

6 Summary and Conclusions

We have described the details of a 4-piece MTE potential approximation to a normal probability and defined its properties. Inference in two hybrid Bayesian networks using MTE potentials was demonstrated using the Shenoy-Shafer architecture for calculating marginals.

Extensive future research on MTE potentials and their applications is needed. General formulations for other continuous probability density functions can allow implementation to a broader range of problems. These particularly include distributions that are formed when discrete variables have continuous parents. An adaption 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.

Acknowledgments

The first author was supported by a research assistantship funded through a grant to the University of Kansas Center for Research, Inc. from Sprint and Nortel Networks. The research described in this paper was partially supported by a contract from Sparta, Inc. to the second author.

References

[1] Binder, J., Koller, D., Russell, S.J. and K. Kanazawa (1997), "Adaptive probabilistic networks with hidden variables," *Machine Learning Journal*, 29(2-3), 213–244.

[2] Cobb, B.R. and P.P. Shenoy (2003), "Inference in hybrid Bayesian networks with mixtures of truncated exponentials," School of Business Working Paper No. 294, University of Kansas, Lawrence, Kansas. Available for download at: `http://www.business.ku.edu/home/pshenoy/WP294.pdf`.

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

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

[5] Lerner, U., Segal, E. and D. Koller (2001), "Exact inference in networks with discrete children of continuous parents," in J. Breese and D. Koller (eds.), *Uncertainty in Artificial Intelligence*, 17, 319–328.

[6] Moral, S., Rumí, R. and A. Salmerón (2001), "Mixtures of truncated exponentials in hybrid Bayesian networks," in P. Besnard and S. Benferhart (eds.), *Proceedings of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty (ECSQARU 2001)*, 156–167.

[7] Murphy, K. (1999), "A variational approximation for Bayesian networks with discrete and continuous latent variables," in K.B. Laskey and H. Prade (eds.), *Uncertainty in Artificial Intelligence*, 15, 467–475.

[8] Shenoy, P.P. (1997), "Binary Join Trees for Computing Marginals in the Shenoy-Shafer Architecture," *International Journal of Approximate Reasoning*, 17(2,3), 239–263.

[9] Shenoy, P.P. and Shafer, G. (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.