

Nonlinear Deterministic Relationships in Bayesian Networks

Barry R. Cobb and Prakash P. Shenoy

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

Abstract. In a Bayesian network with continuous variables containing a variable(s) that is a conditionally deterministic function of its continuous parents, the joint density function does not exist. Conditional linear Gaussian distributions can handle such cases when the deterministic function is linear and the continuous variables have a multi-variate normal distribution. In this paper, operations required for performing inference with nonlinear conditionally deterministic variables are developed. We perform inference in networks with nonlinear deterministic variables and non-Gaussian continuous variables by using piecewise linear approximations to nonlinear functions and modeling probability distributions with mixtures of truncated exponentials (MTE) potentials.

1 Introduction

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

Approximate inference in Bayesian networks with continuous variables can be performed using mixtures of truncated exponentials (MTE) potentials (Moral et al. 2001). Cobb and Shenoy (2004) define operations which allow the distributions of linear deterministic variables to be determined when the continuous variables are modeled with MTE potentials. This allows MTE potentials to be used for inference in any continuous CLG model, as well as other models that have non-Gaussian and conditionally deterministic variables. This paper extends these methods to continuous Bayesian networks with nonlinear deterministic variables.

The remainder of this paper is organized as follows. Section 2 introduces notation and definitions used throughout the paper. Section 3 describes a method for approximating a nonlinear function with a piecewise linear function. Section 4 defines operations required for inference in Bayesian networks with conditionally deterministic variables. Section 5 contains examples of determining the distributions of nonlinear conditionally deterministic variables. Section 6 summarizes and states directions for future research. This paper is based on a longer, unpublished working paper (Cobb and Shenoy 2005).

2 Notation and Definitions

This section contains notation and definitions used throughout the paper.

2.1 Notation

Random variables will be denoted by capital letters, e.g., A, B, C . Sets of variables will be denoted by boldface capital letters, e.g., \mathbf{X} . All variables are assumed to take values in continuous state spaces. If \mathbf{X} is a set of variables, \mathbf{x} is a configuration of specific states of those variables. The continuous state space of \mathbf{X} is denoted by $\Omega_{\mathbf{X}}$. In graphical representations, continuous nodes are represented by double-border ovals, whereas nodes that are deterministic functions of their parents are represented by triple-border ovals.

2.2 Mixtures of Truncated Exponentials

A mixture of truncated exponentials (MTE) (Moral et al. 2001) potential has the following definition.

MTE potential. Let $\mathbf{X} = (X_1, \dots, X_n)$ be an n -dimensional random variable. 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

$$\phi(\mathbf{x}) = a_0 + \sum_{i=1}^m a_i \exp \left\{ \sum_{j=1}^n b_i^{(j)} x_j \right\} \quad (1)$$

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

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

$$\phi(\mathbf{x}) = \phi_i(\mathbf{x}) \quad \text{if } \mathbf{x} \in \Omega_{\mathbf{X}_i}, \quad i = 1, \dots, k, \quad (2)$$

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

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. We will refer to ϕ_i as the i -th piece of the MTE potential ϕ and $\Omega_{\mathbf{X}_i}$ as the portion of the domain of \mathbf{X} approximated by ϕ_i . In this paper, all MTE potentials are equal to zero in unspecified regions.

2.3 Conditional Mass Functions (CMF)

When relationships between continuous variables are deterministic, the joint probability density function (PDF) does not exist. If Y is a deterministic relationship of variables in \mathbf{X} , i.e. $y = g(\mathbf{x})$, the conditional mass function (CMF) for $\{Y \mid \mathbf{x}\}$ is defined as

$$p_{Y \mid \mathbf{x}} = \mathbf{1}\{y = g(\mathbf{x})\} , \quad (3)$$

where $\mathbf{1}\{A\}$ is the indicator function of the event A , i.e. $\mathbf{1}\{A\}(B) = 1$ if $B = A$ and 0 otherwise.

3 Piecewise Linear Approximations to Nonlinear Functions

3.1 Dividing the Domain

Suppose that a random variable Y is a deterministic function of a single variable X , $Y = g(X)$. The function $Y = g(X)$ can be approximated by a piecewise linear function. Define a set of ordered points $x = (x_0, \dots, x_n)$ in the domain of X , with x_0 and x_n defined as the endpoints of the domain. A corresponding set of points $y = (y_0, \dots, y_n)$ is determined by calculating the value of the function $y = g(x)$ at each point x_i , $i = 0, \dots, n$. The piecewise linear function (with n pieces) approximating $Y = g(X)$ is the function $Y^{(n)} = g^{(n)}(X)$ defined as follows:

$$g^{(n)}(x) = \begin{cases} \left(y_0 - \frac{y_1 - y_0}{x_1 - x_0} \cdot x_0 \right) + \frac{y_1 - y_0}{x_1 - x_0} \cdot x & \text{if } x_0 \leq x < x_1 \\ \left(y_1 - \frac{y_2 - y_1}{x_2 - x_1} \cdot x_1 \right) + \frac{y_2 - y_1}{x_2 - x_1} \cdot x & \text{if } x_1 \leq x < x_2 \\ \vdots & \vdots \\ \left(y_{n-2} - \frac{y_{n-1} - y_{n-2}}{x_{n-1} - x_{n-2}} \cdot x_{n-2} \right) + \frac{y_{n-1} - y_{n-2}}{x_{n-1} - x_{n-2}} \cdot x & \text{if } x_{n-2} \leq x < x_{n-1} \\ \left(y_{n-1} - \frac{y_n - y_{n-1}}{x_n - x_{n-1}} \cdot x_{n-1} \right) + \frac{y_n - y_{n-1}}{x_n - x_{n-1}} \cdot x & \text{if } x_{n-1} \leq x \leq x_n . \end{cases} \quad (4)$$

Let $g_i^{(n)}(x)$ denote the i -th piece of the piecewise linear function in (4). We refer to $g^{(n)}$ as an n -point (piecewise linear) approximation of g . In this paper, all piecewise linear functions equal zero in unspecified regions. If a variable is a deterministic function of multiple variables, the definition in (4) can be extended by dividing the domain of the parent variables into hypercubes and creating an approximation of each function in each hypercube.

3.2 Algorithm for Splitting Regions

An initial piecewise approximation is defined (minimally) by splitting the domain of X at extreme points and points of change in concavity and convexity in the function $y = g(x)$, and at endpoints of pieces of the MTE potential for X . This initial set of bounds on the pieces of the approximation is defined as $x = (x_0^S, \dots, x_\ell^S)$. The absolute value of the difference between the approximation and the function will increase, then eventually decrease within each region of the approximation. This is due to the fact that the approximation in (4) always lies “inside” the actual function.

Additional pieces may be added to improve the fit between the nonlinear function and the piecewise approximation. Define an allowable error bound, ϵ , for the distance between the function $g(x)$ and its piecewise linear approximation. Define an interval η used to select the next point at which to test the distance between $g(x)$ and the piecewise approximation. The piecewise linear approximation in (4) is completely defined by the sets of points $x = (x_0, \dots, x_n)$ and $y = (y_0, \dots, y_n)$. The following procedure in pseudo-code determines the sets of points x and y which define the piecewise linear approximation when a deterministic variable has one parent.

```

INPUT :=  $x_0^S, \dots, x_\ell^S, g(x), \epsilon, \eta$ 
OUTPUT :  $x = (x_0, \dots, x_n), y = (y_0, \dots, y_n)$ 
INITIALIZATION
 $x \leftarrow \{(x_0^S, \dots, x_\ell^S)\}$  /* Endpoints, extrema, and inflection points in  $\Omega_{\mathbf{X}}$  */
 $y \leftarrow \{g(x_0^S), \dots, g(x_\ell^S)\}$ 
 $i = 0$  /* Index for the intervals in the domain of  $X$  */

DO WHILE  $i < |x|$  /* Continue until all intervals are refined */
 $j = 1$  /* Index for number of test points in an interval */
 $a = 0$  /* Previous distance between  $g(x)$  and approximation */
 $b = 0$  /* Current distance between  $g(x)$  and approximation */
  FOR  $j = 1 : (x_{i+1} - x_i) / \eta$ 
     $b = g(x_i + (j - 1) \cdot \eta) -$ 
       $\left( \left( y_i - \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \cdot x_i \right) + \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \cdot (x_i + (j - 1) \cdot \eta) \right)$ 
    IF  $|b| \geq a$  /* Compare current and previous distance */
       $a = |b|$  /* Distance increased; test next point */
    ELSE
      BREAK /* Distance did not increase; break loop */
    END IF
  END FOR
  IF  $a > \epsilon$  /* Test max. distance versus allowable error bound */
     $x \leftarrow \text{Rank}(x \cup \{x_i + (j - 2) \cdot \eta\})$  /* Update  $x$  and re-order */
     $y \leftarrow \text{Rank}(y \cup \{g(x_i + (j - 2) \cdot \eta)\})$  /* Update  $y$  and re-order */
  END IF
   $i = i + 1$ 
END DO

```

The algorithm refines the piecewise approximation to the function $y = g(x)$ until the maximum distance between the function and the piecewise approximation is no larger than the specified error bound. A smaller error bound, ϵ , produces more pieces in the linear approximation and a closer fit in the theoretical and approximate density functions for the deterministic variable (see, e.g., Section 5.1 of (Cobb and Shenoy 2005)). A closer approximation using more pieces, however, requires greater computational expense in the inference process.

4 Operations with Linear Deterministic Variables

Consider a random variable Y which is a monotonic function, $Y = g(X)$, of a random variable X . The joint cumulative distribution function (CDF) for $\{X, Y\}$ is given by $F_{X,Y}(x, y) = F_X(g^{-1}(y))$ if $g(X)$ is monotonically increasing and $F_{X,Y}(x, y) = F_X(x) - F_X(g^{-1}(y))$ if $g(X)$ is monotonically decreasing. The CDF of Y is determined as $F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y)$. Thus, $F_Y(y) = F_X(g^{-1}(y))$ if $g(X)$ is monotonically increasing and $F_Y(y) = 1 - F_X(g^{-1}(y))$ if $g(X)$ is monotonically decreasing. By differentiating the CDF of Y , the PDF of Y is obtained as

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} (g^{-1}(y)) \right|, \quad (5)$$

when $Y = g(X)$ is monotonic. If Y is a conditionally deterministic linear function of X , i.e. $Y = g(x) = ax + b$, $a \neq 0$, the following operation can be used to determine the marginal PDF for Y :

$$f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right). \quad (6)$$

The following definition extends the operation defined in (6) to accommodate piecewise linear functions. Suppose Y is a conditionally deterministic piecewise linear function of X , $Y = g(X)$, where $g_i(x) = a_i x + b_i$, with each $a_i \neq 0$, $i = 1, \dots, n$. Assume the PDF for X is an MTE potential ϕ with k pieces, where the j -th piece is denoted ϕ_j for $j = 1, \dots, k$. Let n_j denote the number of linear segments of g that intersect with the domain of ϕ_j and notice that $n = n_1 + \dots + n_j + \dots + n_k$. The CMF $p_{Y|x}$ represents the conditionally deterministic relationship of Y on X . The following definition will be used to determine the marginal PDF for Y (denoted $\chi = (\phi \otimes p_{Y|x})^{\downarrow Y}$):

$$\chi(y) = (\phi \otimes p_{Y|x})^{\downarrow Y}(y) \triangleq \begin{cases} 1/a_1 \cdot \phi_1((y-b_1)/a_1) & \text{if } y_0 \leq y < y_1 \\ 1/a_2 \cdot \phi_1((y-b_2)/a_2) & \text{if } y_1 \leq y < y_2 \\ \vdots & \vdots \\ 1/a_{n_1} \cdot \phi_1((y-b_{n_1})/a_{n_1}) & \text{if } y_{n_1-1} \leq y < y_{n_1} \\ \vdots & \vdots \\ 1/a_n \cdot \phi_k((y-b_n)/a_n) & \text{if } y_{n-1} \leq y < y_n \end{cases}, \quad (7)$$

with ϕ_j being the piece of ϕ whose domain is a superset of the domain of g_i . The normalization constants for each piece of the resulting MTE potential ensures that the CDF of the resulting MTE potential matches the CDF of the theoretical MTE potential at the endpoints of the domain of the resulting PDF. From Theorem 3 in (Cobb and Shenoy 2004), it follows that the class of MTE potentials is closed under the operation in (7); thus, the operation can be used for inference in Bayesian networks with deterministic variables. Note that the class of MTE potentials is not closed under the operation in (5), which is why we approximate nonlinear functions with piecewise linear functions.

5 Examples

The following examples illustrate determination of the distributions of random variables which are nonlinear deterministic functions of their parents, as well as inference in a simple Bayesian network with a nonlinear deterministic variable.

5.1 Example One

Suppose X is normally distributed with a mean of 0 and a standard deviation of 1, i.e. $X \sim N(0, 1^2)$, and Y is a conditionally deterministic function of X , $y = g(x) = x^3$. The distribution of X is modeled with an two-piece, three-term MTE potential as defined in (Cobb et al. 2003). The MTE potential is denoted by ϕ and its two pieces are denoted ϕ_1 and ϕ_2 , with $\Omega_{X_1} = \{x : -3 \leq x < 0\}$ and $\Omega_{X_2} = \{x : 0 \leq x \leq 3\}$.

Piecewise Approximation. Over the region $[-3, 3]$, the function $y = g(x) = x^3$ has an inflection point at $x = 0$, which is also an endpoint of a piece of the MTE approximation to the PDF of X . To initialize the algorithm in Sect. 3.2, we define $x = (x_0^S, x_1^S, x_2^S) = (-3, 0, 3)$ and $y = (y_0^S, y_1^S, y_2^S) = (-27, 0, 27)$. For this example, define $\epsilon = 1$ and $\eta = 0.06$ (which divides the domain of X into 100 equal intervals).

The procedure in Sect. 3.2 terminates after finding sets of points $x = (x_0, \dots, x_8)$ and $y = (y_0, \dots, y_8)$ as follows:

$$\begin{aligned} x &= (-3.00, -2.40, -1.74, -1.02, 0.00, 1.02, 1.74, 2.40, 3.00) , \\ y &= (-27.000, -13.824, -5.268, -1.061, 0.000, 1.061, 5.268, 13.824, 27.000) . \end{aligned}$$

The function representing the eight-point linear approximation is defined as

$$g^{(8)}(x) = \begin{cases} 21.960x + 38.880 & \text{if } -3.00 \leq x < -2.40 \\ 12.964x + 17.289 & \text{if } -2.40 \leq x < -1.74 \\ 5.843x + 4.898 & \text{if } -1.74 \leq x < -1.02 \\ 1.040x & \text{if } -1.02 \leq x < 0 \\ 1.040x & \text{if } 0 \leq x < 1.02 \\ 5.843x - 4.898 & \text{if } 1.02 \leq x < 1.74 \\ 12.964x - 17.289 & \text{if } 1.74 \leq x < 2.40 \\ 21.960x - 38.880 & \text{if } 2.40 \leq x \leq 3.00 . \end{cases} \quad (8)$$

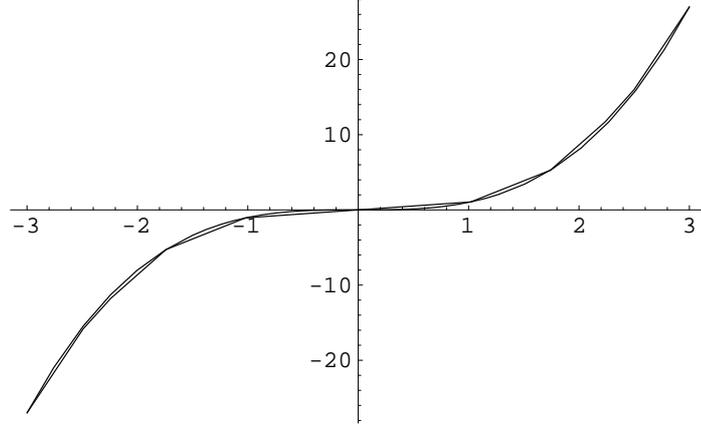


Fig. 1. The piecewise linear approximation $g^{(8)}(x)$ overlaid on the function $y = g(x)$

The piecewise linear approximation $g^{(8)}(x)$ is shown in Fig. 1, overlaid on the function $y = g(x)$.

The conditional distribution for Y is represented by a CMF as follows:

$$\psi^{(8)}(x, y) = p_{Y|x}(y) = \mathbf{1}\{y = g^{(8)}(x)\} .$$

Determining the Distribution of Y . The marginal distribution for Y is determined by calculating $\chi^{(8)} = (\phi \otimes \psi^{(8)})^{1Y}$. The MTE potential for Y is

$$\chi^{(8)}(y) = \begin{cases} (1/21.960) \cdot \phi^{(1)}(0.0455y - 1.7705) & \text{if } -27.000 \leq y < -13.824 \\ (1/12.964) \cdot \phi_1(0.0771y - 1.3336) & \text{if } -13.824 \leq y < -5.268 \\ (1/5.843) \cdot \phi_1(0.1712y - 0.8384) & \text{if } -5.268 \leq y < -1.061 \\ (1/1.040) \cdot \phi_1(0.9612y) & \text{if } -1.061 \leq y \leq 0.000 \\ (1/1.040) \cdot \phi_2(0.9612y) & \text{if } 0.000 \leq y < 1.061 \\ (1/5.843) \cdot \phi_2(0.1712y + 0.8384) & \text{if } 1.061 \leq y < 5.628 \\ (1/12.964) \cdot \phi_2(0.0771y + 1.3336) & \text{if } 5.628 \leq y < 13.824 \\ (1/21.960) \cdot \phi_2(0.0455y + 1.7705) & \text{if } 13.824 \leq y \leq 27.000 . \end{cases}$$

The CDF associated with the eight-piece MTE approximation is shown in Fig. 2, overlaid on the CDF associated with the PDF from the transformation

$$f_Y(y) = f_X(g_1^{-1}(y)) \frac{d}{dy} (g_1^{-1}(y)) . \quad (9)$$

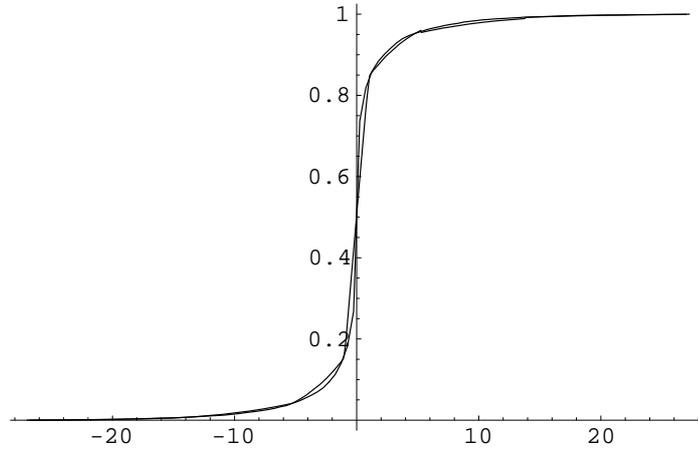


Fig. 2. CDF for the eight-piece MTE approximation to the distribution for Y overlaid on the CDF created using the transformation in (9)

5.2 Example Two

The Bayesian network in this example (see Fig. 3) contains one variable (X) with a non-Gaussian potential, one variable (Z) with a Gaussian potential, and one variable (Y) which is a deterministic linear function of its parent. The probability distribution for X is a beta distribution, i.e. $\mathcal{L}(X) \sim \text{Beta}(\alpha = 2.7, \beta = 1.3)$. The PDF for X is approximated (using the methods described in (Cobb et al. 2003))



Fig. 3. The Bayesian network for Example Two

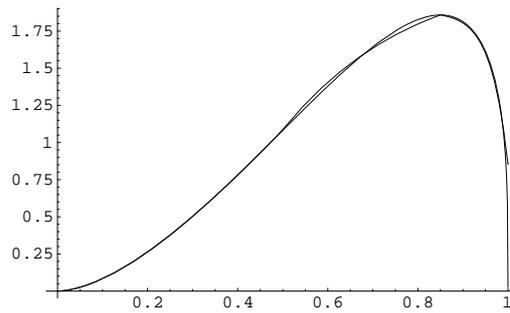


Fig. 4. The MTE potential for X overlaid on the actual $\text{Beta}(2.7, 1.3)$ distribution

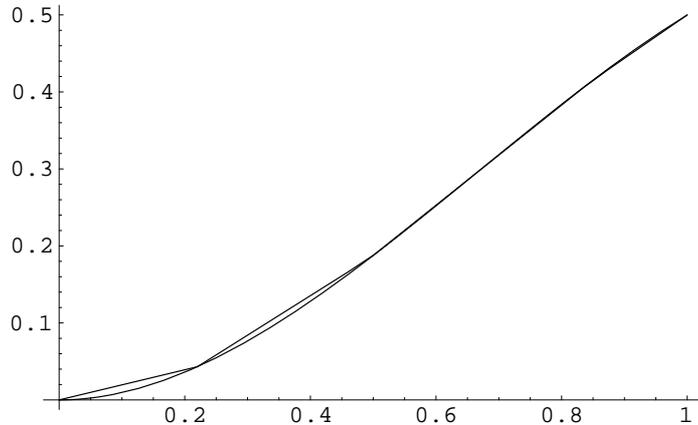


Fig. 5. The piecewise linear approximation $g^{(5)}(x)$ overlaid on the function $g(x)$ in Example Two

by a three-piece, two-term MTE potential. The MTE potential ϕ for X is shown graphically in Figure 4, overlaid on the actual $Beta(2.7, 1.3)$ distribution.

The variable Y is a conditionally deterministic function of X , $y = g(x) = -0.5x^3 + x^2$. The five-point linear approximation is characterized by points $x = (x_0, \dots, x_5) = (0, 0.220, 0.493, 0.667, 0.850, 1)$ and $y = (y_0, \dots, y_5) = (0, 0.043, 0.183, 0.296, 0.415, 0.500)$. The points x_0, x_2, x_3 , and x_5 are defined according to the endpoints of the pieces of ϕ . The point x_4 is an inflection point in the function $g(x)$ and the point $x_1 = 0.220$ is found by the algorithm in Sect. 3.2 with $\epsilon = 0.015$ and $\eta = 0.01$. The function representing the five-piece linear approximation (denoted as $g^{(5)}$) is shown graphically in Fig. 5 overlaid on $g(x)$.

The conditional distribution for Y given X is represented by a CMF as follows:

$$\psi^{(5)}(x, y) = p_{Y|x}(y) = \mathbf{1}\{y = g^{(5)}(x)\} .$$

The probability distribution for Z is defined as $\mathcal{L}(Z | y) \sim N(2y + 1, 1)$ and is approximated by χ , which is a two-piece, three-term MTE approximation to the normal distribution (Cobb et al. 2003).

5.3 Computing Messages

The join tree for the example problem is shown in Fig. 6.

The messages required to calculate posterior marginals for each variable in the network without evidence are as follows:

- 1) ϕ from $\{X\}$ to $\{X, Y\}$
- 2) $(\phi \otimes \psi^{(5)})^{\downarrow Y}$ from $\{X, Y\}$ to $\{Y\}$ and $\{Y\}$ to $\{Y, Z\}$
- 3) $((\phi \otimes \psi^{(5)})^{\downarrow Y} \otimes \chi)^{\downarrow Z}$ from $\{Y, Z\}$ to $\{Z\}$



Fig. 6. The join tree for the example problem

5.4 Posterior Marginals

The posterior marginal distribution for Y is the message sent from $\{X, Y\}$ to $\{Y\}$ and is calculated using the operation in (7). The expected value and variance of this distribution are calculated as 0.3042 and 0.0159, respectively. The posterior marginal distribution for Z is the message sent from $\{Y, Z\}$ to $\{Z\}$ and is calculated by point-wise multiplication of MTE functions, followed by marginalization (see operations defined in (Moral et al. 2001)). The expected value and variance of this distribution are calculated as 1.6084 and 1.0455, respectively.

5.5 Entering Evidence

Suppose we observe evidence that $Z = 0$ and let e_Z denote this evidence. Define $\varphi = (\phi \otimes \psi^{(5)}) \downarrow^Y$ and $\psi'^{(5)}(x, y) = \mathbf{1}\{x = g^{(5)-1}(y)\}$ as the potentials resulting from the reversal of the arc between X and Y (Cobb and Shenoy 2004). The evidence e_Z is passed from $\{Z\}$ to $\{Y, Z\}$ in the join tree, where the existing potential is restricted to $\chi(y, 0)$. This likelihood potential is passed from $\{Y, Z\}$ to $\{Y\}$ in the join tree.

Denote the unnormalized posterior marginal distribution for B as $\xi'(y) = \varphi(y) \cdot \chi(y, 0)$. The normalization constant is calculated as $K = \int_y (\varphi(y) \cdot \chi(y, 0)) dy = 0.0670$. Thus, the normalized marginal distribution for Y is found as $\xi(y) =$

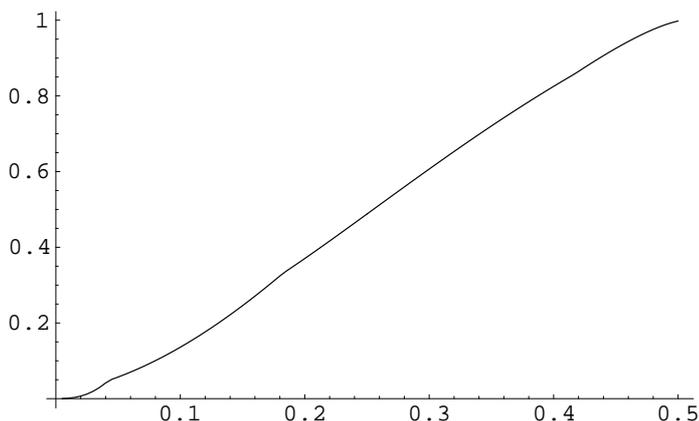


Fig. 7. The posterior marginal CDF for Y considering the evidence $Z = 0$

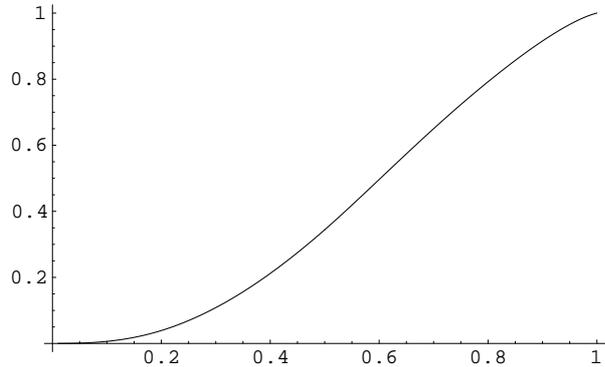


Fig. 8. The posterior marginal CDF for X considering the evidence ($Z = 0$)

$K^{-1} \cdot \xi'(y)$. The expected value and variance of this distribution (whose CDF is displayed in Fig. 7) are calculated as 0.2560 and 0.0167, respectively.

Using the operation in (7), we determine the posterior marginal distribution for X as $\vartheta = (\xi \otimes \psi^{(5)}) \downarrow^X$. The expected value and variance of this distribution are calculated as 0.5942 and 0.0480, respectively. The posterior marginal CDF for X considering the evidence is shown graphically in Figure 8.

6 Summary and Conclusions

This paper has described operations required for inference in Bayesian networks containing variables that are nonlinear deterministic functions of their continuous parents. Since the joint PDF for a network with deterministic variables does not exist, the operations required are based on the method of convolutions from probability theory. By estimating nonlinear functions with piecewise linear approximations, we ensure the class of MTE potentials are closed under these operations. Bayesian networks in this paper contain only continuous variables. In future work, we plan to design a general inference algorithm for Bayesian networks that contain a mixture of discrete and continuous variables, with some continuous variables defined as deterministic functions of their continuous parents.

References

- Cobb, B.R. and P.P. Shenoy: Inference in hybrid Bayesian networks with deterministic variables. In P. Lucas (ed.): *Proceedings of the Second European Workshop on Probabilistic Graphical Models (PGM-04)* (2004) 57–64, Leiden, Netherlands.
- Cobb, B.R. and P.P. Shenoy: Modeling nonlinear deterministic relationships in Bayesian networks. School of Business Working Paper No. 310, University of Kansas, Lawrence, Kansas (2005). Available for download at: <http://www.people.ku.edu/~brcobb/WP310.pdf>

- Cobb, B.R., Shenoy, P.P. and R. Rumí: Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials. Working Paper No. 303, School of Business, University of Kansas, Lawrence, Kansas (2003). Available for download at: <http://www.people.ku.edu/~brcobb/WP303.pdf>
- Kullback, S. and R.A. Leibler: On information and sufficiency. *Annals of Mathematical Statistics* **22** (1951) 79–86.
- Larsen, R.J. and M.L. Marx: *An Introduction to Mathematical Statistics and its Applications* (2001) Prentice Hall, Upper Saddle River, N.J.
- S.L. Lauritzen and F. Jensen: Stable local computation with conditional Gaussian distributions. *Statistics and Computing* **11** (2001) 191–203.
- Moral, S., Rumí, R. and A. Salmerón: Mixtures of truncated exponentials in hybrid Bayesian networks. In P. Besnard and S. Benferhart (eds.): *Symbolic and Quantitative Approaches to Reasoning under Uncertainty*, Lecture Notes in Artificial Intelligence **2143** (2001) 156–167, Springer-Verlag, Heidelberg.