Piecewise Linear Approximations of Nonlinear Deterministic Conditionals in Continuous Bayesian Networks

Barry R. Cobb
Virginia Military Institute
Lexington, Virginia, USA
cobbbr@vmi.edu

Prakash P. Shenoy
University of Kansas
Lawrence, Kansas, USA
pshenoy@ku.edu

Abstract

To enable inference in continuous Bayesian networks containing nonlinear deterministic conditional distributions, Cobb and Shenoy (2005) have proposed approximating nonlinear deterministic functions by piecewise linear ones. In this paper, we describe two principles and a heuristic for finding piecewise linear approximations of nonlinear functions. We illustrate our approach for some commonly used one- and two-dimensional nonlinear deterministic functions.

1 Introduction

This paper is concerned with inference in continuous Bayesian networks containing nonlinear deterministic conditional distributions for some continuous variables. In a BN, each variable is associated with a conditional probability distribution (or a conditional, in short) for each value of its parent variables. A conditional for a variable is said to be *deterministic* if the variances of the conditional are all zeroes (for all values of the variable's parents). If a continuous variable has a deterministic conditional, then the joint probability density function for all continuous variables does not exist, and this must be taken into account in a propagation algorithm for computing posterior marginals. Recently, Shenoy and West (2011a) have described an extension of the Shenoy-Shafer architecture for discrete BNs (Shenoy and Shafer, 1990), where deterministic conditionals for continuous variables are represented by Dirac delta functions (Dirac, 1927).

A major problem in inference in continuous BNs is marginalizing continuous variables, which involves integration. Often, there are no closed form solutions for the result of the integration, making representation of the intermediate functions difficult. We will refer to this as the *integration* problem.

One of the earliest non-Monte Carlo methods for inference in BNs with continuous variables was proposed by Lauritzen and Jensen (2001) for the special case where all continuous variables have conditional linear Gaussian (CLG) distributions. Since marginals of multivariate Gaussian distributions are multivariate Gaussian distributions whose parameters can be easily found from the parameters of the original distributions, this obviates the need to do integrations. However, the requirement that all continuous conditional distributions are CLG restricts the class of hybrid BNs that can be represented using this method.

Another method for dealing with the integration problem is the mixture of truncated exponentials (MTE) model proposed by Moral et al. (2001). The main idea here is to approximate conditional probability density functions (PDFs) by piecewise exponential functions, whose exponents are linear functions of the variables in the domain, and where the pieces are defined on hypercubes, i.e., intervals for each variable. Such functions are called MTEs, and this class of functions is closed under multiplication, addition, and integration, operations that are done in the propagation algorithm. Thus, the MTE method can be used for BNs that do not contain deterministic conditionals and any conditional distribution can be

used as long as they can be approximated by MTE functions.

Similar to the MTE method, Shenoy and West (2011b) have proposed another method called mixture of polynomials (MOP) to address the integration problem. The main idea is to approximate conditional PDFs by piecewise polynomials defined on hypercubes. In all other respects, the MOP method is similar in spirit to the MTE method.

Recently, Shenoy (2012) has proposed a generalization of the MOP function by allowing the pieces to be defined on regions called hyper-rhombuses, which are a generalization of hypercubes. One advantage of MOPs defined on hyper-rhombuses is that such functions are closed under transformations needed for multi-dimensional linear deterministic functions.

Cobb and Shenoy (2005) extend the applicability of MTE and MOP methods to continuous BNs containing nonlinear deterministic conditionals. The main idea is to approximate a nonlinear function by a piecewise linear (PL) function, and then apply the usual MTE or MOP method.

In this paper, we propose two principles and a heuristic for finding piecewise linear approximations of nonlinear functions, and illustrate it for an one-dimensional function $Y = X^2$, and a two-dimensional function $W = X \cdot Y$.

An outline of the remainder of the paper is as follows. In Section 2, we briefly define mixtures of polynomials functions. Also, we describe some numerical measures of goodness of an approximation of a PDF or cumulative distribution function (CDF). In Section 3, we describe two basic principles, and a heuristic, for finding piecewise linear approximations of a nonlinear functions in one and two dimensions, and we illustrate this technique for the functions $Y = X^2$ in the one-dimensional case, and $W = X \cdot Y$ for the two-dimensional case. Finally, in Section 4, we summarize our contributions and discuss some issues for further research.

2 Mixtures of Polynomials

In this section, we briefly define mixture of polynomials functions. For the remainder of the paper, all functions are assumed to equal zero in undefined regions.

The definition of mixture of polynomials given here is taken from (Shenoy, 2012).

An m-dimensional function $f : \mathbb{R}^m \to \mathbb{R}$ is said to be a MOP function if

$$f(x_1, x_2, \ldots, x_m) = \begin{cases} P_i(x_1, x_2, \ldots, x_m) \\ \qquad \text{for } (x_1, x_2, \ldots, x_m) \in A_i, i = 1, \ldots, k. \end{cases}$$

where $P_i(x_1, x_2, \ldots, x_m)$ are multivariate polynomials in m variables for all i, and the regions A_i are disjoint and as follows. Suppose π is a permutation of $\{1, ..., m\}$. Then each A_i is of the form:

$$\begin{aligned} l_{1i} \le x_{\pi(1)} \le u_{1i}, \\ l_{2i}(x_{\pi(1)}) \le x_{\pi(2)} \le u_{2i}(x_{\pi(1)}), \\ \vdots \\ l_{mi}(x_{\pi(1)}, \ldots, x_{\pi(m-1)}) \le x_{\pi(m)} \\ \le u_{mi}(x_{\pi(1)}, \ldots, x_{\pi(m-1)}) \end{aligned} \quad (2.1)$$

where l_{1i} and u_{1i} are constants, and $l_{ji}(x_{\pi(1)}, \ldots, x_{\pi(j-1)})$ and $u_{ji}(x_{\pi(1)}, \ldots, x_{\pi(j-1)})$ are linear functions of $x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(j-1)}$ for $j = 2, \ldots, m$, and $i = 1, \ldots, k$. We will refer to the nature of the region described in Equation (2.1) as a *hyper-rhombus*.

A *hypercube* is a special case of a hyper-rhombus where $l_{1i}, \ldots, l_{mi}, u_{1i}, \ldots, u_{mi}$ are all constants.

Example 1. An example of a 2-piece, 3-degree MOP approximation $g_1(\cdot)$ of the standard normal PDF in 1-dimension is as follows:

$$g_1(x) = \begin{cases} 0.424 + 0.128x - 0.085x^2 - 0.028x^3 \\ \qquad \text{if } -3 < x < 0, \\ 0.424 - 0.128x - 0.085x^2 + 0.028x^3 \\ \qquad \text{if } 0 \le x < 3 \end{cases} \quad (2.2)$$

$g_1(\cdot)$ was found using Lagrange interpolating polynomial with Chebyshev points (Shenoy, 2012).

The family of MOP functions is closed under multiplication, addition and integration, the operations that are done during propagation of potentials in the extended Shenoy-Shafer architecture for BNs. They are also closed under transformations needed for linear deterministic functions.

In this paper, we focus on the use of PL approximations in conjunction with MOP functions to facilitate inference in BNs. In many cases, the PL approximations can also be used with MTE functions. This is demonstrated in (Cobb and Shenoy, 2012).

2.1 Quality of MOP Approximations

In this section, we discuss some quantitative ways to measure the accuracy of a MOP approximation of PDFs.

We will measure the accuracy of a PDF with respect to another defined on the same domain by four different measures, the Kullback-Leibler (KL) divergence, maximum absolute deviation, absolute error in the mean, and absolute error in the variance.

If f is a PDF on the interval (a, b), and g is a PDF that is an approximation of f such that $g(x) > 0$ for $x \in (a, b)$, then the *KL divergence* between f and g, denoted by $KL(f, g)$, is defined as

$$KL(f, g) = \int_a^b \ln\left(\frac{f(x)}{g(x)}\right) f(x)\, dx.$$

$KL(f, g) \geq 0$, and $KL(f, g) = 0$ if and only if $g(x) = f(x)$ for all $x \in (a, b)$. We do not know the semantics associated with the statistic $KL(f, g)$.

The *maximum absolute deviation* between f and g, denoted by $MAD(f, g)$, is given by:

$$MAD(f, g) = sup\{|f(x) - g(x)| : a < x < b\}$$

One semantics associated with $MAD(f, g)$ is as follows. If we compute the probability of some interval $(c, d) \subseteq (a, b)$ by computing $\int_c^d g(x)\, \mathrm{d}x$, then the error in this probability is bounded by $(d - c) \cdot MAD(f, g)$.

The maximum absolute deviation can also be applied to CDFs. Thus, if $F(\cdot)$ and $G(\cdot)$ are the CDFs corresponding to $f(\cdot)$, and $g(\cdot)$, respectively, then the maximum absolute deviation between F and G, denoted by $MAD(F, G)$, is

$$MAD(F, G) = sup\{|F(x) - G(x)| : a < x < b\}$$

The *absolute error of the mean*, denoted by $AEM(f, g)$, and the *absolute error of the variance*, denoted by $AEV(f, g)$, are given by:

$$AEM(f, g) = |E(f) - E(g)| \tag{2.3}$$
$$AEV(f, g) = |V(f) - V(g)| \tag{2.4}$$

where $E(\cdot)$ and $V(\cdot)$ denote the expected value and the variance of a PDF, respectively.

3 Finding PL Approximations of Nonlinear Functions

When we have nonlinear deterministic conditionals, Cobb and Shenoy (2005) propose approximating these nonlinear functions by piecewise linear (PL) ones. The family of MOP functions is closed under operation needed for linear deterministic functions.

There are many ways in which one can approximate a nonlinear function by a PL function. In this section, we describe two basic principles and a heuristic for minimizing the errors in the marginal distribution of the variable with the deterministic conditional represented by the PL approximation.

3.1 One-Dimensional Function $Y = X^2$

To illustrate PL approximations, consider a simple BN as follows: $X \sim N(0, 1)$, $Y = X^2$. The exact marginal distribution of Y is chi-square with 1 degree of freedom. We will use the 2-piece, 3-degree MOP $g_1(\cdot)$ defined in Equation (2.2) on the domain $(-3, 3)$, for the MOP approximation of the PDF of $N(0, 1)$.

3.1.1 Two Basic Principles

In constructing PL approximations, we will adhere to two basic principles. First, the domain of the marginal PDF of the variable with the deterministic conditional should remain unchanged. Thus, in the chi-square example, since the PDF of X is defined on the domain $(-3, 3)$,

and $Y = X^2$, the domain of Y is $(0, 9)$, and we need to ensure that any PL approximation of the function $Y = X^2$ results in the marginal PDF of Y on the domain $(0, 9)$.

Second, if the PDF of X is symmetric about some point, and the deterministic function is also symmetric about the same point, then we need to ensure that the PL approximation retains the symmetry. In the chi-square example, the PDF of X is symmetric about the point $X = 0$, and $Y = X^2$ is also symmetric about the point $X = 0$ on the domain $(-3, 3)$. Therefore, we need to ensure that the PL approximation is also symmetric about the point $X = 0$.

3.1.2 AIC-like Heuristic

In the statistics literature, Akaike's information criterion (AIC) (Akaike, 1974) is a heuristic for building statistical models from data. For example, in a multiple regression setting, if we have a data set with p explanatory variables and a response variable, we could always decrease the sum of squared errors in the model by using more explanatory variables. However, this could lead to over-fitting and lead to poor predictive performance. Thus, we need a measure that has a penalty factor for including more explanatory variables than is necessary. If we have a model with p explanatory variables, and $\widehat{\sigma}^2$ is an estimate of σ^2 in the regression model, the AIC heuristic is to minimize $n \times \ln(\widehat{\sigma}^2) + 2p$, where the $2p$ term acts like a penalty factor for using more explanatory variables than are necessary.

Our context here is slightly different from statistics. In statistics, we have data, and the true model is unknown. In our context, there is no data and the true model is known (the true model could be a nonlinear model estimated from data). However, there are some similarities. We could always decrease the error in the fit between the nonlinear function and the PL approximation by using more parameters (pieces), but doing so does not always guarantee that the error in the marginal distribution of the deterministic variable with the nonlinear function will be minimized. Making inferences with MOPs that have many pieces can be intractable (Shenoy et al., 2011). For this reason, we need to keep the number of pieces as small as possible.

Suppose $f_X(x)$ denotes the PDF of X and suppose we approximate a non-linear deterministic function $Y = r(X)$ by a PL function, say $Y = r_1(X)$. The MSE of the PL approximation r_1, denoted by $MSE(r_1)$, is given as follows:

$$MSE(r_1) = \int_{-\infty}^{\infty} f_X(x) \left(r(x) - r_1(x)\right)^2 \mathrm{d}x. \tag{3.1}$$

The AIC-like heuristic finds a PL approximation $Y = r_1(X)$ with p free parameters such that the $AIC(r_1) = \ln(MSE(r_1)) + p$ is minimized subject to the domain and symmetry principles.

3.1.3 Example

For the chi-square BN, the domain and symmetry principles require use of $(-3, 9)$, $(0, 0)$, and $(3, 9)$ knots. The knots are the points that are connected by straight lines to form the PL approximation. Suppose we wish to find a 4-piece PL approximation. Let (x_1, y_1) and $(-x_1, y_1)$ denote the two additional knots where $-3 < x_1 < 0$, and $0 < y_1 < 9$. Such a PL approximation would consist of 2 free parameters (where the parameters are x_1 and y_1). Solving for the minimum $MSE(r_1)$ with MOP $g_1(x)$ as the PDF of X results in the solution: $x_2 = -1.28$, $y_2 = 1.16$, minimum $MSE(r_1) = 0.043$, and $AIC(r_1) = -1.141$. The PL approximation $Y = r_1(X)$ is as follows (see Figure 1):

$$Y = \begin{cases} -4.66 - 4.55X & \text{if } -3 < X < -1.28 \\ -0.91X & \text{if } -1.28 \leq X < 0 \\ 0.91X & \text{if } 0 \leq X < 1.28 \\ -4.66 + 4.55X & \text{if } 1.28 \leq X < 3 \end{cases} \tag{3.2}$$

If we approximate $Y = X^2$ by a PL approximation $Y = r_2(X)$ with, say 6 pieces (4 free parameters), then the value of $MSE(r_2)$ is 0.0060, and the value of $AIC(r_2)$ is -1.124, which is higher than $AIC(r_1)$. Similarly, if we use an 8-piece approximation (6 free parameters), then the value of $MSE(r_3)$ is 0.002, and

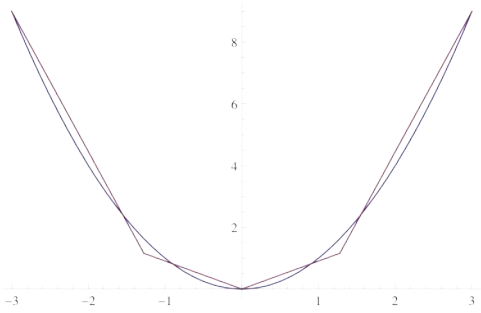

Figure 1: $Y = r_1(X)$ overlaid on $Y = X^2$

the value of $AIC(r_3)$ is -0.421, which is higher than $AIC(r_1)$ and $AIC(r_2)$. Thus, the AIC heuristic suggests a 2-piece PL approximation $Y = r_1(X)$. The accuracies of the marginal PDF of Y computed using MOP $g_1(x)$ for the PDF of X, and the three PL approximations r_1, r_2, and r_3 are shown in the table below (best values are shown in boldface). The model used as the exact PDF to calculate the goodness of fit statistics is the marginal PDF of Y found using $g_1(\cdot)$ and $Y = X^2$.

# pieces	4	6	8
p	**2**	4	6
MSE	0.043	0.006	**0.002**
AIC	**−1.141**	−1.124	−0.421
KL	0.250	0.153	**0.110**
MAD of PDF	35.332	35.092	**34.713**
MAD of CDF	0.164	0.133	**0.101**
AEM	**0.059**	0.189	0.168
AEV	**0.065**	0.106	0.186
CPU	**4.774***s*	13.151s	22.932s

The MAD of the PDFs occurs near zero where the values of the actual PDF and the approximations are very large. The CPU row in the table above represents the combined time to obtain the PL approximation and calculate the goodness of fit statistics. The latter is an indication of the relative computing time that would be required to perform inference in a BN using these PL approximation with MOPs. All computations were made in Mathematica 8.0 on a computer with Intel Core 2 Duo processor (2.93 GHz) with 16 GB of memory.

The minimum AIC heuristic uses the information in the PDF of X as well as the nature of the deterministic function $Y = g(X)$ to find a PL approximation. Its main disadvantage is that determining the knots of a minimum AIC PL approximation involves solving a nonlinear optimization problem. The more pieces there are in a PL approximation, the more variables there are in the nonlinear optimization problem, and more complex it is to solve the nonlinear optimization problem. However, we can often exploit special features of the PDF and the nonlinear function (such as the domain and symmetry principles) to minimize the number of variables in a optimization problem to make its solution tractable. Since the AIC-like heuristic includes a penalty for adding pieces to the PL approximation, the reduction in MSE associated with doing so may not be justified. In fact in this example, the AEM and AEV statistics for the resulting marginal distribution of interest are better when fewer pieces are used in the PL approximation. As expected, computation time increases with the number of pieces in the PL approximation.

3.2 Multi-Dimensional Function $W = X \cdot Y$

For multi-dimensional nonlinear functions, we can use the same two basic principles and the minimum AIC-like heuristic as for the one-dimensional case.

Consider this example: $X \sim N(5, 0.5^2)$, $Y \sim N(15, 4^2)$, and $W = r(X, Y) = X \cdot Y$. We construct a 2-piece, 3-degree MOP $g_X(x) = g_1(\frac{x-5}{0.5})/0.5$ of the PDF of X on the domain $(3.5, 6.5)$, and a 2-piece, 3-degree MOP $g_Y(y) = g_1(\frac{y-15}{4})/4$ of the PDF of Y on the domain $(3, 27)$ (here $g_1(\cdot)$ is the 2-piece, 3-degree MOP approximation of the standard normal PDF on the domain $(-3, 3)$ as described in Equation 2.2).

Using these two MOP approximations of the PDFs of X and Y, we can find an "exact" marginal PDF of W as follows:

$$g_W(w) = \int_{-\infty}^{\infty} g_X(x) \left(\int_{-\infty}^{\infty} g_Y(y)\, \delta(w - x \cdot y)\, \mathrm{d}y \right) \mathrm{d}x, \tag{3.3}$$

where δ is the Dirac delta function (Shenoy

and West, 2011a). $\delta(w - x \cdot y)$ represents the conditional distribution of W given $X = x$ and $Y = y$. $g_W(\cdot)$ is not a MOP, but we do have a representation of it, can graph it, and can compute its mean ($E(g_W) = 75$) and variance ($V(g_W) = 458.96$). Unfortunately, we cannot compute the CDF corresponding to $g_W(\cdot)$. So we do not report any MAD for the CDFs statistics.

Suppose we wish to find a 2-piece PL approximation of $W = X \cdot Y$. The domain of the joint distribution of X and Y is a rectangle $(3.5 < X < 6.5) \times (3 < Y < 27)$. The exact domain of W is $(10.5, 175.5)$. Because g_X is symmetric about the axis $X = 5$, and g_Y is symmetric about the axis $Y = 15$, there are several ways one can find a two-piece region using the symmetry principle. The PDF of X is symmetric about the line $X = 5$, while the PDF of Y is symmetric about the line $Y = 15$. The slope and intercept of the line connecting the points $(3.5, 3)$ and $(6.5, 27)$ are 8 and -25, respectively, so the joint PDF of (X, Y) is symmetric about the line $Y = 8X - 25$. The joint PDF of (X, Y) is similarly symmetric about the line $Y = -8X+55$. There is no symmetry in the function $W = X \cdot Y$ about any axis. Thus, we can divide the domain vertically using the hyperplane $X = 5$ or horizontally using $Y = 15$ or diagonally using $Y = 8X-25$ or $Y = -8X+55$. The best approximation (lowest AIC score) obtained was by dividing the domain of X and Y vertically using the hyperplane $X = 5$. We conjecture that this works better than dividing the rectangle vertically since X has a smaller variance than Y.

Next, we need to find a PL approximation for $r(X, Y) = X \cdot Y$ in each of the two rectangles that satisfies the domain principle. The smallest value of $W = X \cdot Y$ is 10.5 at the point $(X, Y) = (3.5, 3)$, and the largest value of W is 175.5 at the point $(X, Y) = (6.5, 27)$. For the first rectangle, $3.5 < X < 5, 3 < Y < 27$, consider a PL approximation $r_{11}(X, Y) = a_1X + b_1Y + c_1$, where a_1, b_1, c_1 are constants. One way to satisfy the lower bound domain constraint is by selecting the PL approximation r_1 such that $r_{11}(3.5, 3) = 10.5$. Thus, we can eliminate one of the three parameters, e.g., $c_1 = 10.5 - 3.5a_1 - 3b_1$. To find values of the remaining two parameters, we solve an optimization problem as follows:

$$\begin{array}{l} \text{Find } a_1, b_1, c_1 \text{ so as to} \\ Minimize \int_{3.5}^{5} g_X(x) \left(\int_3^{27} (r(x,y) \right. \\ \qquad \left. -r_{11}(x,y))^2 g_Y(y)\, \mathrm{d}y \right) \mathrm{d}x \\ \text{subject to : } c_1 = 10.5 - 3.5a_1 - 3b_1 \end{array} \tag{3.4}$$

For the second rectangle $5 \leq X < 6.5, 3 < Y < 27$, consider a PL approximation $r_{12}(X, Y) = a_2X + b_2Y + c_2$. To satisfy the upper domain constraint, we impose the constraint $r_{12}(6.5, 27) = 175.5$, i.e., $c_2 = 175.5 - 6.5a_2 - 27b_2$. To find values of the remaining two parameters, we solve the optimization problem:

$$\begin{array}{l} \text{Find } a_2, b_2, c_2 \text{ so as to} \\ Minimize \int_{5}^{6.5} g_X(x) \left(\int_3^{27} (r(x,y) \right. \\ \qquad \left. -r_{12}(x,y))^2 g_Y(y)\, \mathrm{d}y \right) \mathrm{d}x \\ \text{subject to : } c_2 = 175.5 - 6.5a_2 - 3b_2, \text{and} \\ \qquad 5a_2 + 3b_2 + c_2 \geq 10.5 \end{array} \tag{3.5}$$

The logic behind the second constraint in (3.5) is that assuming $a_2 \geq 0$ and $b_2 \geq 0$ (which are implicitly satisfied in minimizing MSE), the smallest value of $W = r_{12}(X, Y)$ is at the point $(X, Y) = (5, 3)$. Thus, in order to satisfy the domain principle, we need to ensure that $r_{12}(5, 3) \geq 10.5$. This constraint is binding at the optimal solution. Solving the optimization problems in (3.4) and (3.5), we obtain a PL approximation r_1 as follows:

$$\begin{array}{l} r_1(X, Y) = \\ \begin{cases} 8.15X + 4.18Y - 30.55 & \text{if } X < 5 \\ 25.51X + 5.28Y - 132.89 & \text{if } X \geq 5 \end{cases} \end{array} \tag{3.6}$$

The total MSE for $r_1(X, Y)$ when compared to $r(X, Y)$ using PDFs $g_X(x)$ and $g_Y(y)$ is 14.98. Since we have 4 free parameters (a_1, b_1, a_2, b_2) in the PL approximation $r_1(X, Y)$, the AIC value is $AIC(r_1) = 6.71$.

Let $g_{W_1}(\cdot)$ denote the marginal PDF of W computed using $g_X(x)$, $g_Y(y)$, and $\delta(w - r_1(x, y))$. $g_{W_1}(\cdot)$ is computed as a 13-piece, 7-degree MOP on the domain $(10.5, 175.5)$. A

graph of $g_{W_1}(\cdot)$ overlaid on the graph of $g_W(\cdot)$ is shown in Figure 2.

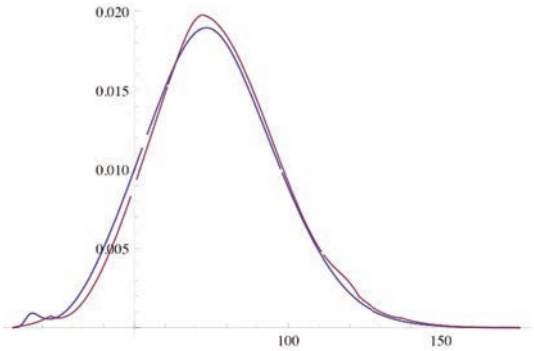

Figure 2: $g_{W_1}(\cdot)$ overlaid on $g_W(\cdot)$

The goodness of fit statistics of g_{W_1} compared to g_W are as follows:

Errors	*Values*
$KL(g_W, g_{W_1})$	0.0069
$MAD(g_W, g_{W_1})$	0.0012
$AEM(g_W, g_{W_1})$	1.8074
$AEV(g_W, g_{W_1})$	12.7308

One way to reduce the AIC value for the PL approximation is to reduce its number of parameters. In solving the optimization problems (3.4) and (3.5), if we add the constraints $c_1 = -a_1 \cdot b_1$ and $c_2 = -a_2 \cdot b_2$, we obtain a PL approximation r_2 as follows:

$$r_2(X,Y) = \begin{cases} 3.00X + 4.60Y - 13.81 & \text{if } X < 5 \\ 27.00X + 5.19Y - 140.06 & \text{if } X \geq 5 \end{cases} \tag{3.7}$$

The approximation $W = r_2(X,Y)$ has a MSE of 18.10, compared to MSE of 14.98 for $W = r_1(X,Y)$. Notice that the approximation $W = r_2(X,Y)$ has only 2 free parameters (compared to 4 for $W = r_1(X,Y)$). The corresponding values of the AIC heuristic is $AIC(r_2) = 4.90$. Let $g_{W_2}(\cdot)$ denote the marginal PDF of W computed using $g_X(x)$, $g_Y(y)$, and $\delta(w - r_2(x,y))$. $g_{W_2}(\cdot)$ is computed as a 13-piece, 7-degree MOP on the domain $(10.5, 175.5)$. A graph of $g_{W_2}(\cdot)$ overlaid on the graph of $g_W(\cdot)$ is shown in Figure 3.

The goodness of fit statistics of g_{W_2} compared to g_W are as follows:

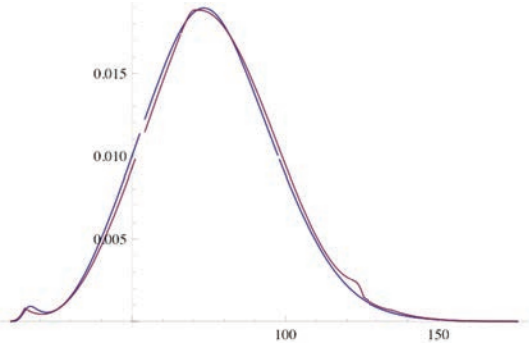

Figure 3: $g_{W_2}(\cdot)$ overlaid on $g_W(\cdot)$

Errors	*Values*
$KL(g_W, g_{W_2})$	0.0023
$MAD(g_W, g_{W_2})$	0.0008
$AEM(g_W, g_{W_2})$	1.2620
$AEV(g_W, g_{W_2})$	10.5493

Comparing these statistics with those obtained without the constraints $c = -a \cdot b$, we see that even though the MSE of r_2 is higher, all goodness of fit statistics for g_{W2} (computed using r_2) are better than the corresponding ones for g_{W1} (computed using r_1). This is probably because the AIC value of r_2 (4.90) is lower than the AIC value of r_1 (6.71). The AIC value of r_2 is lower than the AIC value of r_1 since r_2 has 2 less parameters than r_1.

4 Summary and Conclusions

This paper is concerned with inference in BNs containing nonlinear deterministic conditionals using MOPs. The family of MOP functions is not closed under operations needed for inference with nonlinear deterministic conditionals, but is closed for inference with liner deterministic conditionals. Cobb and Shenoy (2005) suggest approximating nonlinear deterministic conditionals by piecewise linear ones. However, there are many ways of finding such approximations, and a very naïve heuristic was used in (Cobb and Shenoy, 2005), which examined only 1-dimensional nonlinear functions.

In this paper, we describe a principled approach to finding PL approximations of nonlinear functions. The domain principle ensures that the domain of the PL approximation should be exactly the same as in the nonlinear

case, and the symmetry principle states that the approximation should retain symmetry of the nonlinear function (if any), and the symmetry of the PDFs of the parent variables (if any). An AIC-like heuristic for finding PL approximation is described.

Using these two principles and the AIC-like heuristic, PL approximations of some commonly used nonlinear functions are described. For the nonlinear functions $Y = X^2$ and $W = X \cdot Y$, we find the marginal of the variable with the nonlinear deterministic conditional using PL approximations, and compare it with the marginal found using the exact nonlinear function, and estimate the errors in the marginals using MOP approximations of PDFs.

Cobb and Shenoy (2012) use the principles and heuristic in this paper to find PL approximations of $Y = e^X$ and $W = 3X/Y$, and also solve two small hybrid Bayesian networks that contain nonlinear deterministic conditionals. The use of MTE functions in combination with PL approximations to nonlinear deterministic conditionals is also demonstrated.

Acknowledgments

This research has been funded in part by the Spanish Ministry of Economy and Competitiveness through project TIN2010-20900-C04-01, and by the European Regional Development Funds (FEDER).

References

H. Akaike. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

B. R. Cobb and P. P. Shenoy. 2005. Nonlinear deterministic relationships in Bayesian networks. In L. Godo, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Lecture Notes in Artificial Intelligence 3571, pages 27–38. Springer, Berlin.

B. R. Cobb and P. P. Shenoy. 2012. Modeling nonlinear deterministic conditional distributions in hybrid Bayesian networks. Working Paper 328, University of Kansas, School of Business, Lawrence, KS.

P. A. M. Dirac. 1927. The physical interpretation of the quantum dynamics. *Proceedings of the Royal Society of London, Series A*, 113(765):621–641.

S. L. Lauritzen and F. Jensen. 2001. Stable local computation with conditional Gaussian distributions. *Statistics & Computing*, 11(2):191–203.

S. Moral, R. Rumí, and A. Salmerón. 2001. Mixtures of truncated exponentials in hybrid Bayesian networks. In S. Benferhat and P. Besnard, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Lecture Notes in Artificial Intelligence 2143, pages 156–167. Springer, Berlin.

P. P. Shenoy and G. Shafer. 1990. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, pages 169–198. North-Holland.

P. P. Shenoy and J. C. West. 2011a. Extended Shenoy-Shafer architecture for inference in hybrid Bayesian networks with deterministic conditionals. *International Journal of Approximate Reasoning*, 52(6):805–818.

P. P. Shenoy and J. C. West. 2011b. Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52(5):641–657.

P. P. Shenoy, R. Rumí, and A. Salmerón. 2011. Some practical issues in inference in hybrid Bayesian networks with deterministic conditionals. In S. Ventura, A. Abraham, K. Cios, C. Romero, F. Marcelloni, J. M. Benitez, and E. Gibaja, editors, *Proceedings of the 2011 Eleventh International Conference on Intelligent Systems Design and Applications*, pages 605–610. IEEE Research Publishing Services, Piscataway, NJ.

P. P. Shenoy. 2012. Two issues in using mixtures of polynomials for inference in hybrid Bayesian networks. *International Journal of Approximate Reasoning*, 53(5):847–866.