

A Re-definition of Mixtures of Polynomials for Inference in Hybrid Bayesian Networks

Prakash P. Shenoy

University of Kansas School of Business,
1300 Sunnyside Ave., Summerfield Hall, Lawrence, KS 66045-7601 USA
pshenoy@ku.edu

Abstract. We discuss some issues in using mixtures of polynomials (MOPs) for inference in hybrid Bayesian networks. MOPs were proposed by Shenoy and West for mitigating the problem of integration in inference in hybrid Bayesian networks. In defining MOP for multi-dimensional functions, one requirement is that the pieces where the polynomials are defined are hypercubes. In this paper, we discuss relaxing this condition so that each piece is defined on regions called hyper-rhombuses. This relaxation means that MOPs are closed under transformations required for multi-dimensional linear deterministic conditionals, such as $Z = X + Y$. Also, this relaxation allows us to construct MOP approximations of the probability density functions (PDFs) of the multi-dimensional conditional linear Gaussian distributions using a MOP approximation of the PDF of the univariate standard normal distribution. We illustrate our method using conditional linear Gaussian PDFs in two and three dimensions.

1 Introduction

An hybrid Bayesian network (BN) is a BN with a mix of discrete and continuous random variables. A random variable is said to be *discrete* if the cardinality of its state space is countable, and *continuous* otherwise. Each variable in a BN is associated with conditional distributions for the variable, one for each state of its parents. A conditional for a variable is said to be *deterministic* if the conditional variances of its conditional distributions are all zeroes.

Marginalizing a continuous variable involves integration of the product of all potentials that contain the variable in their domains. Often, these potentials are not integrable in closed form. This is a major problem in making inferences in hybrid BNs. We will call this problem the *integration* problem.

One solution to the integration problem is to approximate conditional PDFs by a family of functions called mixtures of truncated exponentials (MTEs) [4]. MTE functions are piecewise functions that are defined on regions called hypercubes, and the functions themselves are exponential functions of a linear function of the variables. Such functions are easy to integrate, and the family of MTE functions are closed under multiplication, addition, and integration, three operations that are used in finding marginals using the extended Shenoy-Shafer

architecture [7]. Cobb et al. [1] describe MTE approximations of several commonly used one-dimensional PDFs. Moral et al. [5] describe a mixed-tree method for representing an MTE approximation of a 2-dimensional CLG distribution.

Another method that is similar in principle to the mixture of truncated exponentials method is the mixture of polynomials (MOP) method proposed by Shenoy and West [8]. Instead of using piecewise exponential functions, the MOP method uses piecewise polynomials. Although a detailed comparison of MTE and MOP methods has yet to be done, an advantage of the MOP method is that one can easily find MOP approximations of differentiable PDFs using the Taylor series expansion of the PDF [8], or by using Lagrange interpolating polynomials [6].

In both the MTE and the MOP methods, the multi-dimensional piecewise functions are defined on regions called hypercubes. One advantage of this restriction is that such multi-dimensional piecewise functions are easy to integrate. However, the hypercube restriction poses two limitations. It is difficult to find an MTE or a MOP approximation of a multi-dimensional conditional PDF for dimensions greater than two. The mixed-tree method proposed by Moral et al. [5] and the Taylor series method proposed by Shenoy and West [8] do not scale up to higher dimensions in practice, i.e., the approximations using these methods have too many pieces or too many terms or have too high a degree for practical use.

Another problem is the presence of deterministic conditionals for continuous variables. For example, suppose X has PDF $f_X(x)$ and suppose Y has conditional PDF $f_{Y|x}(y)$, and suppose Z has a deterministic conditional given by the linear function $Z = X + Y$. To find the marginal distribution of Z , we need to combine $f_X(x)$ and $f_{Y|x}(z-x)$ and then integrate x out of the combination. The problem is that even if $f_{Y|x}(y)$ was defined on hypercubes, $f_{Y|x}(z-x)$ is no longer defined on hypercubes. This problem applies equally to the MTE and MOP methods.

In this paper, we suggest replacing the hypercube condition with a more general condition called hyper-rhombus. For one-dimensional functions, the two conditions coincide. However, for dimensions two or greater, the hyper-rhombus condition is a generalization of the hypercube condition. The hyper-rhombus condition has several important advantages. First, it allows us to define MOP approximations of high-dimensional CLG distributions using a MOP approximation of the one-dimensional standard normal PDF. Second, MOP functions defined on hyper-rhombuses are closed under operations required for multi-dimensional linear deterministic functions. This is not true for MTE functions, i.e., if the definition of MTE functions were generalized so that the hypercube condition was replaced by the hyper-rhombus condition, then MTE functions would not be closed under operations required for multi-dimensional linear deterministic functions. For example, the sum of two independent variables with exponential PDFs has a gamma PDF, which is not a MTE function. Third, MOP functions that are defined on hyper-rhombuses are closed under integration. Fourth, the computational penalty incurred by having hyper-rhombus condition for MOPs (compared to the hypercube condition) appears to be small.

An outline of the remainder of the paper is as follows. In Section 2, we provide a re-definition of high-dimensional MOP functions that are defined on regions

called hyper-rhombuses. In Section 3, we describe how we can use the MOP approximation of the standard normal PDF to find MOP approximations of the PDFs of two- and three-dimensional CLG distributions. Finally in Section 4, we summarize our findings and discuss some issues for further research.

2 Mixture of Polynomials Functions

In this section, we define MOP functions. The definition we provide here is slightly more general than the definition provided in Shenoy and West [8] for the case of multi-dimensional functions.

2.1 MOP Functions

A one-dimensional function $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be a *mixture of polynomials* (MOP) function if it is a piecewise function of the form:

$$f(x) = \begin{cases} a_{0i} + a_{1i}x + \cdots + a_{ni}x^n & \text{for } x \in A_i, i = 1, \dots, k, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where A_1, \dots, A_k are disjoint intervals in \mathbb{R} that do not depend on x , and a_{0i}, \dots, a_{ni} are constants for all i . We will say that f is a k -piece (ignoring the 0 piece), and n -degree (assuming $a_{ni} \neq 0$ for some i) MOP function.

The definition given in Equation (1) is exactly the same as in Shenoy and West [8]. The main motivation for defining MOP functions is that such functions are easy to integrate in closed form, and that they are closed under multiplication, integration, and addition, the main operations in making inferences in hybrid Bayesian networks. The requirement that each piece is defined on an interval A_i is also designed to ease the burden of integrating MOP functions.

A multivariate polynomial is a polynomial in several variables. For example, a polynomial in two variables is as follows:

$$P(x_1, x_2) = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 + a_{20}x_1^2 + a_{02}x_2^2 + a_{21}x_1^2x_2 + a_{12}x_1x_2^2 + a_{22}x_1^2x_2^2 \quad (2)$$

The degree of the polynomial in Equation (2) is 4 assuming a_{22} is a non-zero constant. In general, the degree of a multivariate polynomial is the largest sum of the exponents of the variables in the terms of the polynomial.

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

$$f(x_1, x_2, \dots, x_m) = \begin{cases} P_i(x_1, x_2, \dots, x_m) & \text{for } (x_1, x_2, \dots, x_m) \in A_i, i = 1, \dots, k, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

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

$$\begin{aligned}
l_{1i} &\leq x_{\pi(1)} \leq u_{1i}, \\
l_{2i}(x_{\pi(1)}) &\leq x_{\pi(2)} \leq u_{2i}(x_{\pi(1)}), \\
&\vdots \\
l_{mi}(x_{\pi(1)}, \dots, x_{\pi(m-1)}) &\leq x_{\pi(m)} \leq u_{mi}(x_{\pi(1)}, \dots, x_{\pi(m-1)})
\end{aligned} \tag{4}$$

where l_{1i} and u_{1i} are constants, and $l_{ji}(x_{\pi(1)}, \dots, x_{\pi(j-1)})$ and $u_{ji}(x_{\pi(1)}, \dots, x_{\pi(j-1)})$ are linear functions of $x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(j-1)}$ for $j = 2, \dots, m$, and $i = 1, \dots, k$. We will refer to the nature of the region described in Equation (4) as a *hyper-rhombus*. Although we have defined the hyper-rhombus as a closed region in Equation (4), each of the $2m$ inequalities can be either strictly $<$ or \leq .

The definition of a m -dimensional MOP function stated in Equation (3) is more general than the corresponding definition stated in Shenoy and West [8], which is as follows:

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

$$f(x_1, \dots, x_m) = f_1(x_1) \cdot f_2(x_2) \cdots f_m(x_m) \tag{5}$$

where each $f_i(x_i)$ is a one-dimensional MOP function as defined in Equation (1).

It is easy to see that an m -dimensional function satisfying the condition in Equation (5) will also satisfy the condition in Equation (3), but the converse is not true. Thus, a function as follows:

$$f(x_1, x_2) = \begin{cases} x_1x_2^2 + x_1^2x_2 & \text{for } -3 \leq x_1 \leq 3 \text{ and } x_1 - 3 \leq x_2 \leq x_1 + 3 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

satisfies Equation (3) but not Equation (5) for two reasons. First, $x_1x_2^2 + x_1^2x_2$ cannot be obtained by a product of two one-dimensional polynomials. Second, the first piece is defined on the region $-3 \leq x_1 \leq 3$, $x_1 - 3 \leq x_2 \leq x_1 + 3$, which is not a hypercube, but is a hyper-rhombus.

Finally, high-dimensional MOP function defined on hyper-rhombuses remain MOP functions after integration. Thus, the family of MOP functions are closed under multiplication, addition, and integration. A disadvantage of the new definition is that it takes longer to integrate such functions compared to hypercubes. An advantage is that we can more easily construct high dimensional conditional PDFs such as the conditional linear Gaussian distributions. This is described in Section 3.

3 Fitting MOPs to Two- and Three-Dimensional CLG PDFs

In this section, we will find MOP approximations of the PDFs of 2- and 3-dimensional conditional linear Gaussian (CLG) distributions based on a MOP approximation of the 1-dimensional standard normal PDF. Our revised definition of multi-dimensional MOP functions in Equation (3) facilitates the task of finding MOP approximations of the PDFs of CLG conditional distributions.

3.1 Measuring Goodness of Fit of Approximations

There are several ways of measuring the goodness of fit of an approximation. First, we can use the Kullback-Liebler (KL) divergence [2] as a measure of the goodness of fit. If f is a PDF on the range (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

$$KL(f, g) = \int_a^b \ln \left(\frac{f(x)}{g(x)} \right) f(x) dx. \quad (7)$$

$KL(f, g) \geq 0$, and $KL(f, g) = 0$ if and only if $g(x) = f(x)$ for all $x \in (a, b)$. Typically, $KL(f, g) \leq .001$ is considered a good approximation.

Another measure of goodness of a fit is the maximum absolute deviation. Thus, if f is a PDF on the range (a, b) , and g is a PDF that is an approximation of f , then the maximum absolute deviation between f and g , denoted by $MAD(f, g)$, is given by:

$$MAD(f, g) = \sup\{|f(x) - g(x)| : a < x < b\} \quad (8)$$

Finally, other measures of goodness of fit are the absolute errors in the means and variances. Thus 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:

$$AEM(f, g) = |E(f) - E(g)| \quad (9)$$

$$AEV(f, g) = |V(f) - V(g)| \quad (10)$$

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

3.2 One-Dimensional CLG Distributions

We start with a 4-piece, 3-degree MOP approximation g_1 of the truncated standard normal PDF φ described as follows:

$$g_1(z) = \begin{cases} 0.627801 + 0.503039z + 0.128574z^2 + 0.0100516z^3 & \text{if } -3 < z \leq -1, \\ 0.401108 - 0.00675769z - 0.240746z^2 - 0.0761647z^3 & \text{if } -1 < z \leq 0, \\ 0.401108 + 0.00675769z - 0.240746z^2 + 0.0761647z^3 & \text{if } 0 < z \leq 1, \\ 0.627801 - 0.503039z + 0.128574z^2 - 0.0100516z^3 & \text{if } 1 < z < 3 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

This MOP approximation was found using Lagrange interpolating polynomial with Chebyshev points [6]. A graph of $g_1(z)$ overlaid on the graph of $\varphi(z)$ is shown in Figure 1. The goodness of fit measures are as follows:

$$\begin{aligned} KL(\varphi, g_1) &\approx 0.00005, & MAD(\varphi, g_1) &\approx 0.0024, \\ AEM(\varphi, g_1) &= 0, & AEV(\varphi, g_1) &\approx 0.0003. \end{aligned}$$

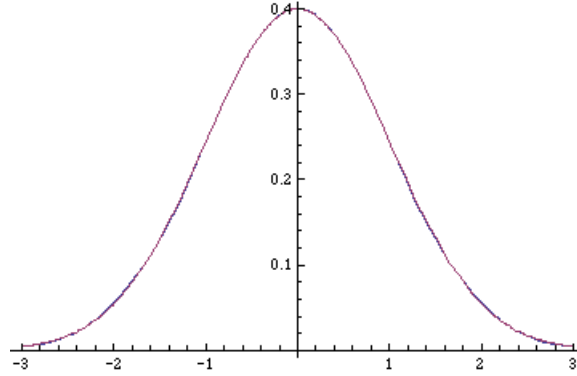


Fig. 1. A graph of $g_1(z)$ versus z (in red) overlaid on the graph of $\varphi(z)$ (in blue)

To find a MOP approximation of the PDF of the general $N(\mu, \sigma^2)$ distribution, where μ and σ are real constants such that $\sigma \neq 0$, we exploit the fact that MOP functions are closed under linear transformations. Thus, if $f(x)$ is a MOP function, then given any real constants a and b , $f(ax+b)$ is also a MOP function. If $Z \sim N(0, 1)$, its PDF is approximated by a MOP function $g(z)$, and $X = \sigma Z + \mu$, then $X \sim N(\mu, \sigma^2)$, and a MOP approximation of the PDF of X is given by $(1/|\sigma|)g((x - \mu)/\sigma)$. Notice that $(1/|\sigma|)g((x - \mu)/\sigma)$ remains a MOP even if μ is a variable (and not a constant) as long as σ is a non-zero constant.

3.3 Two-Dimensional CLG Distributions

Consider the CLG conditional distribution $Y|z \sim N(z, 1)$ (where $Z \sim N(0, 1)$). We will find a MOP approximation of the conditional PDF of $Y|z$ on the two-dimensional region $z - 3 < y < z + 3$. In Shenoy and West [8], a 12-piece, 14-degree MOP approximation is found by covering the two-dimensional region $-3 < z < 3, z - 3 < y < z + 3$ by 12 squares (hypercubes in two dimensions), and then by using two-dimensional Taylor series approximation at the mid-point of each square. Here, we can use the one-dimensional 4-piece, 3-degree MOP approximation $g_1(z)$ of the standard normal distribution as follows. Let $h_1(z, y)$ denote a MOP approximation of the conditional PDF of $Y|z$. Then,

$$h_1(z, y) = g_1(y - z) \quad (12)$$

It follows from the remark at the end of Subsection 3.2, that $h_1(z, y)$ as defined in Equation (12) represents a MOP approximation of the PDF of $N(z, 1)$. Since $g_1(z)$ is a PDF, it follows that $h_1(z, y)$ is a PDF, i.e., $h_1(z, y) \geq 0$, and $\int h_1(z, y)dy = 1$ for all z . Notice that the four pieces of $h_1(z, y)$ are not defined on hypercubes, but rather on hyper-rhombuses (since we now have regions such as $-3 < y - z \leq -1$, etc). A 3-dimensional plot of $h_1(z, y)$ is shown in Figure 2.

Since we are using the one-dimensional MOP approximation $g_1(z)$, the goodness of fit of $h_1(z, y)$ is same as that of $g_1(z)$. One question is how much of a

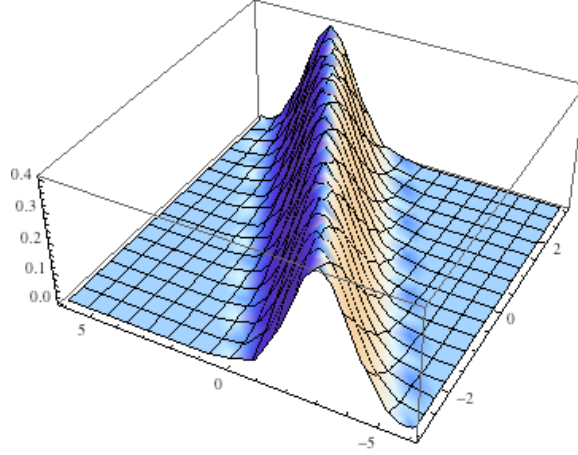


Fig. 2. A 3-dimensional plot of $h_1(z, y)$

penalty do we have to pay for using hyper-rhombuses instead of hypercubes. To answer this, we do two simple experiments.

First, we compute the marginal PDF of Y as follows. $g_1(z)h_1(z, y)$ represents a MOP approximation of the joint PDF of (Z, Y) . To find the marginal PDF of Y , we integrate out Z . Thus, a MOP approximation of the marginal PDF of Y is given by:

$$h_2(y) = \int_{-\infty}^{\infty} g_1(z)h_1(z, y)dz \quad (13)$$

It takes Mathematica[®] ≈ 14 seconds (on a laptop personal computer) to do the multiplication and integration in Equation (13), and $h_2(y)$ is computed as a 7-degree MOP function on the domain $(-6, 6)$. The exact joint distribution of (Z, Y) is bivariate normal with parameters $\mu_Z = \mu_Y = 0$, $\sigma_Z^2 = 1$, $\sigma_Y^2 = 2$, and $\sigma_{ZY} = 1$. Therefore, the exact marginal distribution of Y is $N(0, 2)$. Let $f_Y(y)$ denote the exact PDF of $N(0, 2)$ truncated to $(-6, 6)$. A plot of $h_2(y)$ overlaid on the plot of $f_Y(y)$ is shown in Figure 3. The goodness of fit between $f_Y(y)$ and $h_2(y)$ are as follows:

$$\begin{aligned} KL(f_Y, h_2) &\approx 0.0005, & MAD(f_Y, h_2) &\approx 0.0015 \\ AEM(f_Y, h_2) &\approx 0.0000, & AEV(f_Y, h_2) &\approx 0.0532. \end{aligned}$$

Second, consider the Bayesian network as shown in Figure 4 that includes W with a deterministic conditional, $W = Z + Y$. Suppose we use $g_1(z)$ as a MOP approximation of $N(0, 1)$, and $h_1(z, y)$ as a MOP approximation of $N(z, 1)$. The marginal distribution of W is then given by the convolution formula:

$$h_3(w) = \int_{-\infty}^{\infty} g_1(z)h_1(z, w - z)dz \quad (14)$$

It takes Mathematica[®] ≈ 15 seconds to do the multiplication and integration in Equation (14). h_3 is computed as a 7-degree MOP function in the region $(-9, 9)$.

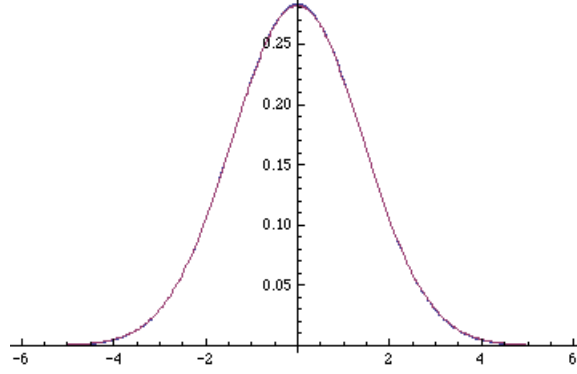


Fig. 3. A graph of $h_2(y)$ (in red) overlaid on the graph of $f_Y(y)$ (in blue)

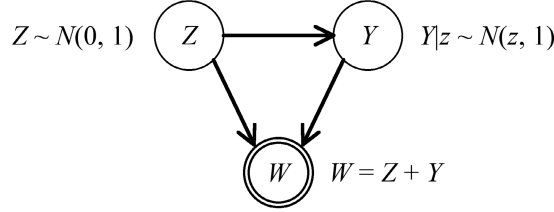


Fig. 4. A Bayesian network with a sum deterministic function

Since the exact joint distribution of (Z, Y) is bivariate normal with parameters $\mu_Z = \mu_Y = 0, \sigma_Z^2 = 1, \sigma_Y^2 = 2, \sigma_{ZY} = 1$, the exact marginal distribution of W is $N(0, 5)$. Let $f_W(w)$ denote the exact PDF of $N(0, 5)$ truncated to $(-9, 9)$. A plot of $h_3(w)$ overlaid on the the plot of $f_W(w)$ is shown in Figure 5. The goodness of fit between $f_W(w)$ and $h_3(w)$ are as follows:

$$\begin{aligned}
 KL(f_W, h_3) &\approx 0.0009, & MAD(f_W, h_3) &\approx 0.0009 \\
 AEM(f_W, h_3) &\approx 0.0000, & AEV(f_W, h_3) &\approx 0.1301.
 \end{aligned}$$

3.4 Three-Dimensional CLG Distributions

Suppose $Z \sim N(0, 1), Y|z \sim N(z, 1)$, and $X|(z, y) \sim N(z + y, 1)$. Notice that the conditional PDF of X is in three dimensions. As in the 2-dimensional case, we find a MOP approximation $h_4(z, y, x)$ of the PDF of $N(z + y, 1)$ in the 3-dimensional region $z + y - 3 < x < z + y + 3$ by using the 4-piece, 3-degree MOP approximation $g_1(z)$ for $N(0, 1)$ as follows:

$$h_4(z, y, x) = g_1(x - (z + y)) \tag{15}$$

Notice that the 4 pieces of h_4 are defined on regions $-3 < x - (z + y) \leq -1$, etc. Therefore, h_4 is a MOP by our definition in Equation (3).

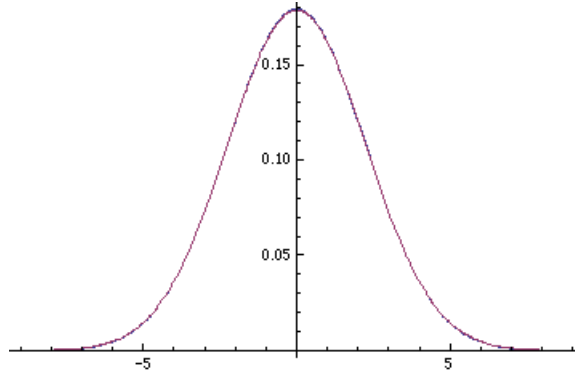


Fig. 5. A graph of $h_3(w)$ (in red) overlaid on the graph of $f_W(w)$ (in blue)

As in the two-dimensional case, we will investigate how much of a time penalty one has to pay for using hyper-rhombus condition. First, we will compute the marginal PDF of X as follows. $g_1(z)$ denotes a MOP approximation of the marginal PDF of Z , $h_1(z, y)$ denotes a MOP approximation of the conditional PDF of $Y|z$, and $h_4(z, y, x)$ denotes a MOP approximation of the conditional PDF of $X|(y, z)$. Thus, $g_1(z)h_1(z, y)h_4(z, y, x)$ denotes a MOP approximation of the joint PDF of (Z, Y, X) . Thus, a MOP approximation of the marginal PDF of X is given by:

$$\begin{aligned} h_6(x) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(z)h_1(z, y)h_4(z, y, x)dydz & (16) \\ &= \int_{-\infty}^{\infty} g_1(z) \left(\int_{-\infty}^{\infty} h_1(z, y)h_4(z, y, x)dy \right) dz \end{aligned}$$

The integration in Equation (16) was done in two stages in Mathematica[®]. The inner integral (with respect to y) required approximately 79 seconds (≈ 1.3 minutes), and resulted in a 2-dimensional, 7-degree, MOP. The outer integral (with respect to z) required 118 seconds (≈ 2.0 minutes), and resulted in a 1-dimensional, 11-degree, MOP on the interval $(-12, 12)$. Thus, the two multiplications and the two integrations in Equation (16) require a total of approximately 197 seconds (or about 3.3 minutes) using Mathematica[®] on a laptop computer. The exact distribution of X can be shown to be $N(0, 6)$. Let $f_X(x)$ denote the PDF of $N(0, 6)$ truncated to the region $(-12, 12)$. A graph of $h_6(x)$ overlaid on the graph of $f_X(x)$ is shown in Figure 6. The goodness of fit between $f_X(x)$ and $h_6(x)$ are as follows:

$$\begin{aligned} KL(f_X, h_6) &\approx 0.0005 & MAD(f_X, h_6) &\approx 0.0010 \\ AEM(f_X, h_6) &\approx 0.0000 & AEV(f_X, h_6) &\approx 0.1618 \end{aligned}$$

Second, consider the Bayesian network as shown in Figure 7 that includes V with a deterministic conditional, $V = Z + Y + X$. Suppose we use $g_1(z)$ as a MOP approximation of $N(0, 1)$, $h_1(z, y)$ as a MOP approximation of $N(z, 1)$, and $h_4(z, y, x)$ as a MOP approximation of $N(z+y, 1)$. The marginal distribution of V is then given by the convolution formula:

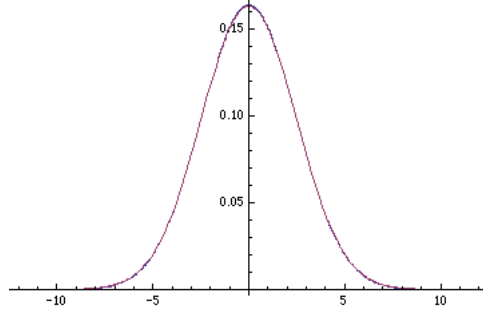


Fig. 6. A graph of $h_6(x)$ (in red) overlaid on the graph of $f_X(x)$ (in blue)

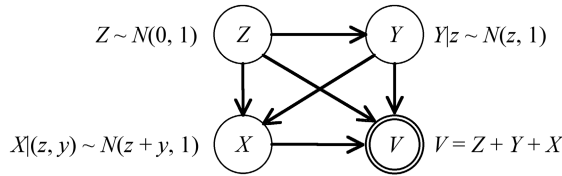


Fig. 7. A Bayesian network with a 3-dimensional conditional

$$\begin{aligned}
 h_8(v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(z)h_1(z, y)h_4(z, y, v - z - y)dydz & (17) \\
 &= \int_{-\infty}^{\infty} g_1(z) \left(\int_{-\infty}^{\infty} h_1(z, y)h_4(z, y, v - z - y)dy \right) dz
 \end{aligned}$$

The integration in Equation (17) was done in two stages in Mathematica[®]. The inner integral (with respect to y) required approximately 93 seconds (≈ 1.6 minutes), and resulted in a 2-dimensional, 7-degree, MOP. The outer integral (with respect to z) required 176 seconds (≈ 2.9 minutes), and resulted in a 1-dimensional, 11-degree, MOP on the interval $(-21, 21)$. Thus, the two multiplications and the two integrations in Equation (17) require a total of approximately 269 seconds (or ≈ 4.5 minutes) using Mathematica[®] on a laptop computer. The exact marginal distribution of V is $N(0, 21)$. Let $f_V(v)$ denote the exact PDF of $N(0, 21)$ truncated to $(-21, 21)$. A plot of $h_8(w)$ overlaid on the the plot of $f_V(v)$ is shown in Figure 8. The goodness of fit between $f_V(v)$ and $h_8(v)$ are as follows:

$$\begin{aligned}
 KL(f_V, h_8) &\approx 0.0008, & MAD(f_V, h_8) &\approx 0.0005 \\
 AEM(f_V, h_8) &\approx 0.0000, & AEV(f_V, h_8) &\approx 0.5648
 \end{aligned}$$

In summary, the hyper-rhombus condition enables us to easily represent CLG conditionals in high dimensions. The computational cost of integrating a high-dimensional MOP function with a hyper-rhombus condition does not seem high for 2 or 3 dimensional CLG distributions, and there is no loss of precision compared to one-dimensional conditionals. Shenoy [6] discusses the tradeoffs between the hyper-rhombus and hypercube conditions in greater detail.

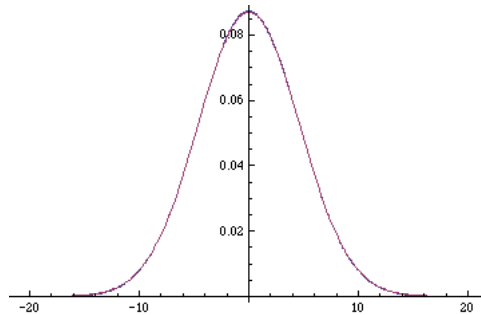


Fig. 8. A graph of $h_8(v)$ (in red) overlaid on the graph of $f_V(v)$ (in blue)

4 Summary and Discussion

A major contribution of this paper is a re-definition of multi-dimensional mixture of polynomials so that the regions where the polynomials are defined are hyper-rhombuses instead of hypercubes. This re-definition allows us to use the MOP approximation of a one-dimensional standard normal PDF to define MOP approximations of high-dimensional CLG PDFs. Also, the re-definition ensures that MOP functions are closed under transformations required for multi-dimensional linear deterministic functions, which was not true when MOP functions were defined on regions that are hypercubes.

Shenoy [6] compares the practical implications of the hyper-rhombus condition with the hypercube condition. He compares the time required for computation of marginals for a couple of simple Bayesian networks, and also the accuracy of the computed marginals.

The hyper-rhombus condition is of great use for constructing MOP approximations of multi-dimensional CLG distributions. However, it does not seem to be of much help in constructing MOP approximations of other multi-dimensional distributions such as the log-normal distribution. Constructing MOP approximations of the multi-dimensional log-normal distributions is of great interest in the finance literature where log-normal distributions are used to model stock price behavior [3]. This is a topic that needs further investigation.

Acknowledgements. The paper has benefitted from comments by three ECSQARU-11 reviewers.

References

1. Cobb, B.R., Shenoy, P.P., Rumí, R.: Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials. *Statistics & Computing* 16(3), 293–308 (2006)
2. Kullback, S., Leibler, R.A.: On information and sufficiency. *Annals of Mathematical Statistics* 22, 76–86 (1951)

3. Li, Y., Shenoy, P.P.: Solving hybrid influence diagrams with deterministic variables. In: Grünwald, P., Spirtes, P. (eds.) *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pp. 322–331. AUAI Press, Corvallis (2010)
4. Moral, S., Rumí, R., Salmerón, A.: Mixtures of truncated exponentials in hybrid bayesian networks. In: Benferhat, S., Besnard, P. (eds.) *ECSQARU 2001. LNCS (LNAI)*, vol. 2143, pp. 156–167. Springer, Heidelberg (2001)
5. Moral, S., Rumí, R., Salmerón, A.: Approximating conditional MTE distributions by means of mixed trees. In: Nielsen, T.D., Zhang, N.L. (eds.) *ECSQARU 2003. LNCS (LNAI)*, vol. 2711, pp. 173–183. Springer, Heidelberg (2003)
6. Shenoy, P.P.: Some issues in using mixtures of polynomials for inference in hybrid Bayesian networks. Working Paper 323, School of Business University of Kansas, Lawrence, KS (October 2010)
7. Shenoy, P.P., West, J.C.: Extended Shenoy-Shafer architecture for inference in hybrid Bayesian networks with deterministic conditionals. *International Journal of Approximate Reasoning* (2011) (in press)
8. Shenoy, P.P., West, J.C.: Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning* (2011) (in press)