Appeared in: J. Vejnarova and T. Kroupa (eds.), Proceedings of the 8th Workshop on Uncertainty Processing (WUPES'09), Sept. 2009, 202--212, University of Economics, Prague.

Mixtures of Polynomials in Hybrid Bayesian Networks with Deterministic Variables

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

James C. West
School of Business
University of Kansas
cully@ku.edu

June 3, 2009

Abstract

The main goal of this paper is to describe inference in hybrid Bayesian networks (BNs) using mixtures of polynomials (MOP) approximations of probability density functions (PDFs). Hybrid BNs contain a mix of discrete, continuous, and conditionally deterministic random variables. The conditionals for continuous variables are typically described by conditional PDFs. A major hurdle in making inference in hybrid BNs is marginalization of continuous variables, which involves integrating combinations of conditional PDFs. In this paper, we suggest the use of MOP approximations of PDFs, which are similar in spirit to using mixtures of truncated exponentials (MTEs) approximations. MOP functions can be easily integrated, and are closed under combination and marginalization. This enables us to propagate MOP potentials in the extended Shenoy-Shafer architecture for inference in hybrid BNs that can include deterministic variables. MOP approximations have several advantages over MTE approximations of PDFs. They are easier to find, even for multi-dimensional conditional PDFs, and are applicable for a larger class of deterministic functions in hybrid BNs.

1 Introduction

Bayesian networks (BNs) and influence diagrams (IDs) were invented in the mid 80s (see e.g., [17], [7]) to represent and reason with large multivariate discrete probability models and decision problems, respectively. Several efficient algorithms exist to compute exact marginals of posterior distributions for discrete BNs (see e.g., [11], [23], and [9]) and to solve discrete IDs exactly (see e.g., [16], [20], [21], and [8]).

The state of the art exact algorithm for mixtures of Gaussians hybrid BNs is the Lauritzen-Jensen algorithm [10]. This requires the conditional PDFs of continuous variables to be conditional linear Gaussians, and that discrete variables do not have continuous parents. Marginals of multivariate normal distributions can be found easily without the need for integration. The disadvantages are that in the inference process, continuous variables have to be marginalized before discrete ones. In some problems, this restriction can lead to large cliques [12].

If a BN has discrete variables with continuous parents, Murphy [15] uses a variational approach to approximate the product of the potentials associated with a discrete variable and its parents with a conditional linear Gaussian. [13] uses a numerical integration technique called Gaussian quadrature to approximate non-conditional linear Gaussian distributions with conditional linear Gaussians, and this same technique can be used to approximate the product of potentials associated with a discrete variable and its continuous parents. Murphy's and Lerner's approach is then embedded in the Lauritzen-Jensen algorithm [10] to solve the resulting mixtures of Gaussians BN.

Shenoy [22] proposes approximating non-conditional linear Gaussian distributions by mixtures of Gaussians using a nonlinear optimization technique, and using arc reversals to ensure discrete variables do not have continuous parents. The resulting mixture of Gaussians BN is then solved using Lauritzen-Jensen algorithm [10].

[14] proposes approximating PDFs by mixtures of truncated exponentials (MTE), which are easy to integrate in closed form. Since the family of mixtures of truncated exponentials is closed under combination and marginalization, the Shenoy-Shafer architecture [23] can be used to solve a MTE BN. [4] proposes using a non-linear optimization technique for finding MTE approximations for several commonly used one-dimensional distributions. [2, 3] extend this approach to BNs with linear and non-linear deterministic variables. In the latter case, they approximate non-linear deterministic functions by piecewise linear ones.

In this paper, we propose using mixtures of polynomials (MOP) approximations of PDFs. Mixtures of polynomials are widely used in many domains including computer graphics, font design, approximation theory, and numerical analysis. They were first studied by Schoenberg [18]. When the MOP functions are continuous, they are referred to as polynomial splines [19]. The use of splines to approximate PDFs was initially suggested by [5]. For our purposes, continuity is not an essential requirement, and we will restrict our analysis to piecewise polynomial approximations of PDFs.

Using MOP is similar in spirit to using MTEs. MOP functions can be easily integrated, and they are closed under combination and marginalization. Thus, the extended Shenoy-Shafer architecture [25] can be used to make inferences in BN with deterministic variables. However, there are several advantages of MOP functions over MTEs.

First, we can find MOP approximations of differentiable PDFs easily by using the Taylor series approximations. Finding MTE approximations as suggested by [4] necessitates solving non-linear optimization problems, which is not as easy a task as it involves navigating among local optimal solutions.

Second, for the case of conditional PDFs with several parents, finding a good MTE approximation can be extremely difficult as it involves solving a non-linear optimization problem in high-dimensional space for each piece. The Taylor series expansion can also be used for finding MOP approximations of conditional PDFs. In [24], we describe a MOP approximation for a 2-dimensional CLG distribution.

Third, if a hybrid BN contains deterministic functions, then the MTE approach can be used directly only for linear deterministic functions. By directly, we mean without approximating a non-linear deterministic function by a piecewise linear one. This is because the MTE functions are not closed under transformations needed for non-linear deterministic functions. MOP functions are closed under a larger family of deterministic functions including linear functions and quotients [24]. This enables propagation in a bigger family of hybrid BNs than is possible using MTEs.

An outline of the remainder of the paper is as follows. In Section 2, we define MOP functions and describe how one can find MOP approximations with illustration for the univariate normal distribution. In Section 3, we solve a small example designed to demonstrate the feasibility of using MOP approximations with a non-differentiable deterministic function. Finally, in Section 4, we end with a summary and discussion of some of the challenges associated with MOP approximations.

2 Mixtures of Polynomials Approximations

In this section, we describe MOP functions and some methods for finding MOP approximations of PDFs. We illustrate our method for the normal distribution. In [24], we also describe MOP approximations of the PDFs of the chi-square distribution, and the conditional linear Gaussian distribution in two dimensions.

2.1 MOP Functions

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

$$f(x) = \begin{cases} a_{0i} + a_{1i}x + a_{2i}x^2 + \cdots + a_{ni}x^n & \text{for } x \in A_i, i = 1, \ldots, k, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

where $A_1, \ldots, A_k$ are disjoint intervals in $\mathcal{R}$ that do not depend on x, and $a_{0i}, \ldots, a_{ni}$ 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 main motivation for defining MOP functions is that such functions are easy to integrate in closed form, and that they are closed under multiplication and integration. They are also closed under differentiation and addition.

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

$$f(x_1, \ldots, x_m) = f_1(x_1) \cdot f_2(x_2) \cdots f_m(x_m) \tag{2.2}$$

where each $f_i(x_i)$ is a one-dimensional MOP function as defined in Equation (2.1). If $f_i(x_i)$ is a k_i-piece, n_i-degree MOP function, then f is a $(k_1 \cdots k_m)$-piece, $(n_1 + \ldots + n_m)$-degree MOP function. Therefore it is important to keep the number of pieces and degrees to a minimum.

2.2 Finding MOP Approximations of PDFs

Consider the univariate standard normal PDF $\phi(z) = (1/\sqrt{2\pi})e^{-z^2/2}$. A 1-piece, 28-degree, MOP approximation $\phi_{1p}(z)$ of $\phi(z)$ in the interval $(-3, 3)$ is as follows:

$$\phi_{1p}(z) = \begin{cases} c^{-1}(1 - z^2/2 + z^4/8 - \ldots + z^{28}/1428329123020800) & \text{if } -3 < z < 3, \\ 0 & \text{otherwise} \end{cases}$$

where $c^{-1} \approx 0.4$. This MOP approximation was found using the Taylor series expansion of $e^{-z^2/2}$, at $z = 0$, to degree 28, restricting it to the region $(-3, 3)$, verifying that $\phi_{1p}(z) \geq 0$ in the region $(-3, 3)$, and normalizing it with constant c so that $\int \phi_{1p}(z)dz = 1$ (whenever the limits of integration are not specified, the entire range $(-\infty, \infty)$ is to be understood). We will denote these operations by writing:

$$\phi_{1p}(z) = \begin{cases} TSeries[e^{-z^2/2}, z = 0, d = 28] & \text{if } -3 < z < 3 \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

We can verify that $\phi_{1p}(z) \geq 0$ as follows. First, we plot the unnormalized MOP approximation, denoted by, say, $\phi_u(z)$. From the graph, we identify approximately the regions where $\phi_u(z)$ could possibly be negative. Then starting from a point in each these regions, we compute the local minimum of $\phi_u(z)$ using, e.g., gradient descent. Since MOP functions are easily differentiable, the gradients can be easily found. If $\phi_u(z) \geq 0$ at all the local minimums, then we have verified that $\phi_{1p}(z) \geq 0$. If $\phi_u(z) < 0$ at a local minimum, then we need to either increase the degree of the polynomial approximation, or increase the number of pieces, or both.

We have some very small coefficients in the MOP approximation. Rounding these off to a certain number of decimal places could cause numerical instability. Therefore, it is important to keep the coefficients in their rational form.

A graph of the MOP approximation $\phi_{1p}(z)$ overlaid on the actual PDF $\phi(z)$ is shown in Figure 1 and it shows that there are not many differences between the two functions in the interval $(-3, 3)$. The main difference is that ϕ_{1p} is restricted to $(-3, 3)$, whereas ϕ is not. The mean of ϕ_{1p} is 0, and its variance ≈ 0.976. Most of the error in the variance is due to the restriction of the distribution to the interval $(-3, 3)$. If we restrict the standard normal density

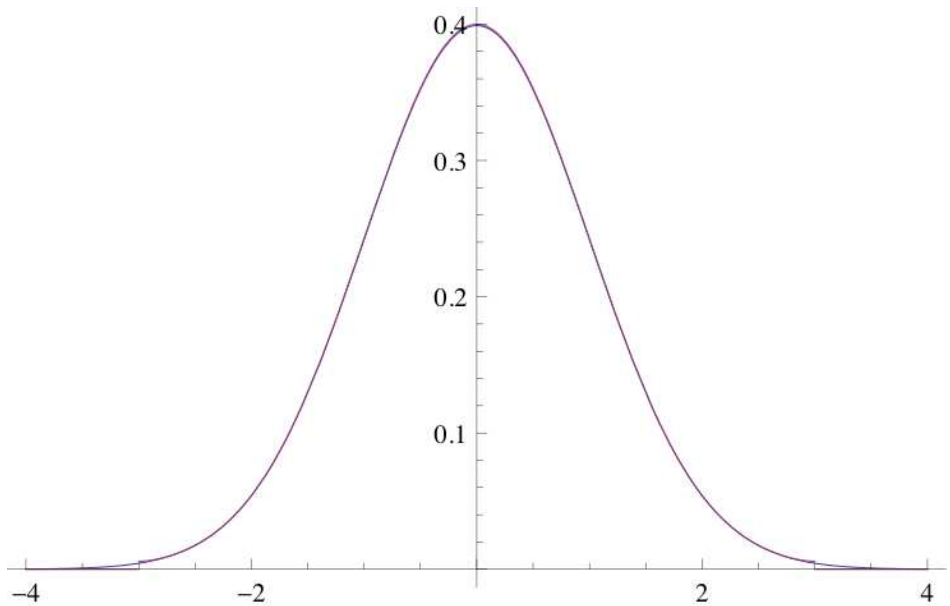

Figure 1: A graph of $\phi_{1p}(z)$ overlaid on $\phi(z)$

ϕ function to the interval $(-3, 3)$, renormalize it so that it is a PDF, then its variance ≈ 0.973.

In some examples, working with a 28-degree polynomial may not be tractable. In this case, we can include more pieces to reduce the degree of the polynomial. For example, a 6-piece, 3-degree MOP approximation of $\phi(z)$ is as follows:

$$\phi_{6p}(z) = \begin{cases} TSeries[e^{-z^2/2}, z = -5/2, d = 3] & \text{if } -3 < z < -2, \\ TSeries[e^{-z^2/2}, z = -3/2, d = 3] & \text{if } -2 \leq z < -1, \\ TSeries[e^{-z^2/2}, z = -1/2, d = 3] & \text{if } -1 \leq z < 0, \\ TSeries[e^{-z^2/2}, z = 1/2, d = 3] & \text{if } 0 \leq z < 1, \\ TSeries[e^{-z^2/2}, z = 3/2, d = 3] & \text{if } 1 \leq z < 2, \\ TSeries[e^{-z^2/2}, z = 5/2, d = 3] & \text{if } 2 \leq z < 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

Notice that ϕ_{6p} is discontinuous at the end points of the intervals. Also, $E(\phi_{6p}) = 0$, and $V(\phi_{6p}) \approx 0.974$. The variance of ϕ_{6p} is closer to the variance of the truncated normal (≈ 0.973) than ϕ_{1p}.

In some examples, for reasons of precision, we may wish to work with a larger interval than $(-3, 3)$ for the standard normal. For example, an 8-piece, 4-degree MOP approximation of ϕ in the interval $(-4,4)$ is as follows:

$$\phi_{8p}(z) = \begin{cases} TSeries[e^{-z^2/2}, z=-7/2, d=4] & \text{if } -4 < z < -3, \\ TSeries[e^{-z^2/2}, z=-5/2, d=3] & \text{if } -3 \le z < -2, \\ TSeries[e^{-z^2/2}, z=-3/2, d=3] & \text{if } -2 \le z < -1, \\ TSeries[e^{-z^2/2}, z=-1/2, d=3] & \text{if } -1 \le z < 0, \\ TSeries[e^{-z^2/2}, z=1/2, d=3] & \text{if } 0 \le z < 1, \\ TSeries[e^{-z^2/2}, z=3/2, d=3] & \text{if } 1 \le z < 2, \\ TSeries[e^{-z^2/2}, z=5/2, d=3] & \text{if } 2 \le z < 3, \\ TSeries[e^{-z^2/2}, z=7/2, d=4] & \text{if } 3 \le z < 4, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Notice that the degrees of the first and the eighth pieces are 4 to avoid $\phi_{8p}(z) < 0$. $E(\phi_{8p}(z)) = 0$, and $V(\phi_{8p}(z)) \approx 0.99985$. Due to the larger interval, the variance is closer to 1 than the variance for ϕ_{6p}. If we truncate the PDF of the standard normal to the region $(-4,4)$ and renormalize it, then its variance is ≈ 0.99893.

To find a MOP approximation of the PDF of the $N(\mu, \sigma^2)$ distribution, where μ and $\sigma > 0$ are constants, we exploit the fact that MOP functions are invariant under linear transformations. Thus, if $f(x)$ is a MOP function, then $f(ax+b)$ is also a MOP function. If $Z \sim N(0,1)$, its PDF is approximated by a MOP function $\phi_p(z)$, and $X = \sigma Z + \mu$, then $X \sim N(\mu, \sigma^2)$, and a MOP approximation of the PDF of X is given by $\xi(x) = (1/\sigma)\phi_p((x-\mu)/\sigma)$.

3 An Example

In this section, we illustrate the use of MOP functions for solving a small hybrid Bayesian network (BN) with a deterministic variable. We use the extended Shenoy-Shafer architecture described in [25]. In [24], we solve more hybrid BNs with deterministic variables including the quotient and the product deterministic functions.

Consider a BN as shown in Figure 2. X and Y are continuous variables and W is deterministic with a non-differentiable function of X and Y, $W = \max\{X, Y\}$.

The conditional associated with W is represented by the Dirac potential $\omega(x,y,w) = \delta(w - \max\{x,y\})$, where δ is a Dirac delta function [6]. To compute the marginal PDF of W, we need to evaluate the integral

$$f_W(w) = \int f_X(x)(\int f_Y(y)\delta(w - \max\{x,y\})dy)dx \tag{3.1}$$

where $f_W(w)$, $f_X(x)$, and $f_Y(y)$ are the marginal PDFs of W, X, and Y, respectively. Since the deterministic function is not differentiable, the integrals in Equation (3.1) cannot be evaluated as written.

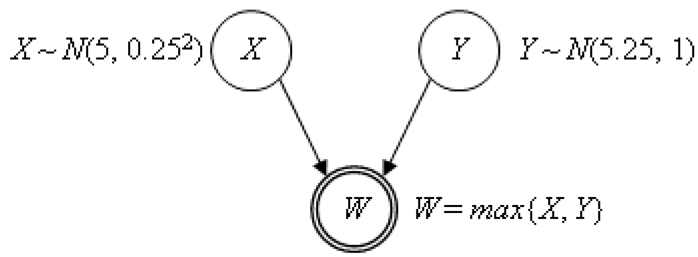

Figure 2: A BN with a max deterministic function

One solution to finding the marginal PDF of W is to use theory of order statistics. Let $F_W(w)$, $F_X(x)$, and $F_Y(y)$ denote the marginal cumulative distribution functions (CDFs) of W, X, and Y, respectively. Then:

$$F_W(w) = P(W \leq w) = P(X \leq w, Y \leq w) = F_X(w)F_Y(w). \tag{3.2}$$

Differentiating both sides of Equation (3.2) with respect to w, we have:

$$f_W(w) = f_X(w)F_Y(w) + F_X(w)f_Y(w). \tag{3.3}$$

In our example, X and Y have normal PDFs, which does not have a closed form CDF. However, using MOP approximations of the normal PDF, we can easily compute a closed form expression for the CDFs, which will remain MOP functions. Then, using Equation (3.3), we will have a closed-form MOP approximation for the PDF of W. Assuming we start with the 8-piece, 4-degree MOP approximation ϕ_{8p} of $N(0,1)$ on the interval $(-4,4)$ as described in Equation (2.5), we can find MOP approximations of the PDFs of $N(5, 0.25^2)$ and $N(5.25, 1)$ as discussed in Section 2 as follows:

$$\xi(x) = 4\phi_{8p}(4(x-5)), \tag{3.4}$$
$$\psi(y) = \phi_{8p}(y-5.25). \tag{3.5}$$

Next we find the MOP approximations of the CDFs of X and Y, and then the MOP approximation of the PDF of W using Equation (3.3). A graph of the MOP approximation of $f_W(w)$ is shown in Figure 3.

The mean and variance of the MOP approximation of f_W are computed as 5.5484 and 0.4574. [1] provides formulae for exact computation of the mean and variance of the max of two normals as follows:

$$E(W) = E(X)F_Z(b) + E(Y)F_Z(-b) + af_Z(b), \tag{3.6}$$
$$E(W^2) = (E(X)^2 + V(X))F_Z(b) + (E(Y)^2 + V(Y))F_Z(-b) + (E(X)+E(Y))af_z(b), \tag{3.7}$$

where $a^2 = V(X) + V(Y) - 2C(X,Y)$, $b = (E(X) - E(Y))/a$, and f_Z and F_Z are the PDF and CDF of $N(0,1)$, respectively.

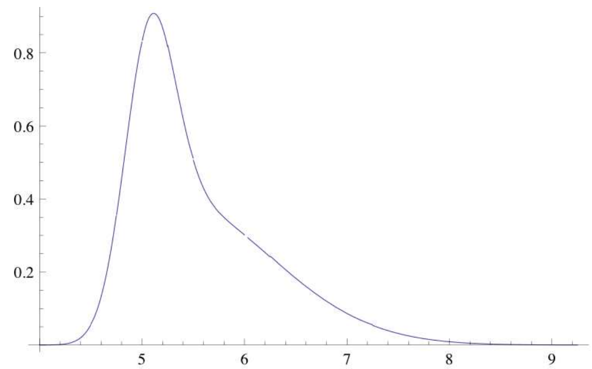

Figure 3: A graph of the MOP approximation of the PDF of W

In our example, $E(X) = 5$, $E(Y) = 5.25$, $V(X) = 0.25^2$, $V(Y) = 1$, $C(X, Y) = 0$. Thus, $E(W) \approx 5.5483$, and $V(W) \approx 0.4576$. The mean and variance of the MOP approximation of W are accurate to three decimal places. Unfortunately, the reasoning behind this computation of the marginal of W is not included in inference in BNs.

To obtain the marginal of W using BN inference, we convert the max function to a differentiable function as follows: $\max\{X, Y\} = X$ if $X \geq Y$, and $= Y$ if $X < Y$. We include a discrete variable A with two states, a and na, where a indicates that $X \geq Y$, and make it a parent of W. The revised BN is shown in Figure 4.

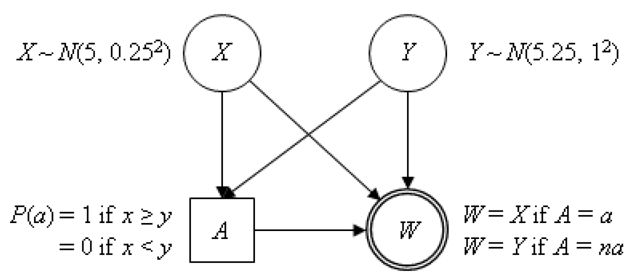

Figure 4: The revised BN for the max deterministic function

Starting with the BN in Figure 4, the marginal of W can be computed using the extended Shenoy-Shafer architecture [25]. We start with mixed potentials as follows:

$$\mu_X(x) = (1, \xi(x)); \tag{3.8}$$
$$\mu_y(y) = (1, \psi(y)); \tag{3.9}$$
$$\mu_A(a, x, y) = (H(x - y), 1), \mu_A(na, x, y) = (1 - H(x - y), 1); \tag{3.10}$$
$$\mu_W(a, x, y, w) = (1, \delta(w - x)), \mu_W(na, x, y, w) = (1, \delta(w - y)). \tag{3.11}$$

In Equation (3.10), $H(.)$ is the *Heaviside* function such that $H(x) = 1$ if $x \geq 0$, and $= 0$ otherwise. The Heaviside function is a MOP function.

To find the marginal of W, we sequentially delete X, Y, and A. To delete X, first we combine μ_X, μ_A, and μ_W, and then marginalize X from the combination:

$$
\begin{aligned}
(\mu_X \otimes \mu_A \otimes \mu_W)(a, x, y, w) &= (H(x-y), \xi(x)\delta(w-x)), && (3.12)\\
(\mu_X \otimes \mu_A \otimes \mu_W)(na, x, y, w) &= (1 - H(x-y), \xi(x)\delta(w-y)); && (3.13)\\
(\mu_X \otimes \mu_A \otimes \mu_W)^{-X}(a, y, w) &= (1, \int H(x-y)\xi(x)\delta(w-x))dx)\\
&= (1, H(w-y)\xi(w)), && (3.14)\\
(\mu_X \otimes \mu_A \otimes \mu_W)^{-X}(na, y, w) &= (1, \delta(w-y)\int(1 - H(x-y))\xi(x)dx)\\
&= (1, \delta(w-y)\theta(y)); && (3.15)
\end{aligned}
$$

where $\theta(y) = \int(1 - H(x-y))\xi(x)dx$.

Next, we delete Y. To do so, we combine $(\mu_X \otimes \mu_A \otimes \mu_W)^{-X}$ and μ_Y, and then marginalize Y:

$$
\begin{aligned}
((\mu_X \otimes \mu_A \otimes \mu_W)^{-X} \otimes \mu_Y)(a, y, w) &= (1, H(w-y)\xi(w)\psi(y)), && (3.16)\\
((\mu_X \otimes \mu_A \otimes \mu_W)^{-X} \otimes \mu_Y)(na, y, w) &= (1, \delta(w-y)\theta(y)\psi(y)); && (3.17)\\
((\mu_X \otimes \mu_A \otimes \mu_W)^{-X} \otimes \mu_Y)^{-Y}(a, w) &= (1, \xi(w)\int H(w-y)\psi(y)dy)\\
&= (1, \xi(w)\rho(w)), && (3.18)\\
((\mu_X \otimes \mu_A \otimes \mu_W)^{-X} \otimes \mu_Y)^{-Y}(na, w) &= (1, \theta(w)\psi(w)); && (3.19)
\end{aligned}
$$

where $\rho(w) = \int H(w-y)\psi(y)dy$.

Finally, we delete A by marginalizing A from $((\mu_X \otimes \mu_A \otimes \mu_W)^{-X} \otimes \mu_Y)^{-Y}$:

$$
\begin{aligned}
(((\mu_X \otimes \mu_A \otimes \mu_W)^{-X} \otimes \mu_Y)^{-Y})^{-A}(w) &= (1, \xi(w)\rho(w) + \theta(w)\psi(w))\\
&= (1, \omega(w)); && (3.20)
\end{aligned}
$$

where $\omega(w) = \xi(w)\rho(w) + \theta(w)\psi(w)$. $\omega(w)$ is a MOP approximation of $f_W(w)$. Notice that

$$
\begin{aligned}
\rho(w) &= \int H(w-y)\psi(y)dy = F_Y(w), \text{and} && (3.21)\\
\theta(w) &= \int(1 - H(x-y))\xi(x)dx = 1 - P(X > w) = F_X(w), && (3.22)
\end{aligned}
$$

and therefore, $\omega(w) = \xi(w)\rho(w)+\theta(w)\psi(w)$ is a MOP approximation of $f_X(w)F_Y(w)+F_X(w)f_Y(w)$. We get exactly the same results as those obtained by using theory of order statistics but using BN inference.

4 Summary and Discussion

The biggest problem associated with inference in hybrid BNs is the integration involved in marginalization of continuous variables. As a remedy, we have proposed MOP approximations for PDFs in the same spirit as MTE approximations [14]. Like MTE functions, MOP functions are easy to integrate, and are closed under combination and marginalization. This allows propagation of MOP potentials using the extended Shenoy-Shafer architecture [25].

MOP approximations have several advantages over MTE approximations of PDFs. First, they are easy to find using the Taylor series expansion of differentiable functions. Second, finding MOP approximations of multi-dimensional conditional PDFs is also relatively straightforward using the multi-dimensional Taylor series expansion. Third, MOP approximations are closed for a larger family of deterministic functions including the quotient functions. Beyond these observations, a formal empirical comparison of MOP vs. MTE approximations is an issue that needs further study.

Some issues associated with MOP approximations that need to be investigated are as follows. There is a tradeoff between the number of pieces and the degree of the polynomial. More pieces mean smaller intervals and consequently smaller degrees. Assuming the goal is to find marginals most efficiently, what is the optimal number of pieces/degrees?

Another challenge is to describe the effect of pieces/terms on the errors in the moments of marginals. It appears that most of the errors in the moments are caused by truncating the domain of variables to some finite intervals. Thus, it may be possible to decide on what intervals should be used if we wish to compute marginals within some prescribed error bounds for the moments of the marginal of variable of interest.

High degree MOP approximations lead to very small coefficients that need to be kept in rational form. This may decrease the efficiency of computation, and may limit the size of BN models that can be solved. One solution here is to use more pieces, which lowers the degrees of the MOP approximations.

MOP approximations are not closed for many classes of deterministic functions such as products and exponentiation. If we can expand the class of MOP functions to include positive and negative rational exponents and maintain the properties of MOP functions—easily integrable, closed under combination and marginalization—then we can solve hybrid BNs with a larger class of deterministic functions.

References

[1] C. E. Clark. The greatest of a finite set of random variables. *Operations Research*, 9(2):145–162, 1961.

[2] B. R. Cobb and P. P. Shenoy. Hybrid Bayesian networks with linear deterministic variables. In F. Bacchus and T. Jaakkola, editors, *Uncertainty in*

Artificial Intelligence: Proceedings of the 21st Conference, pages 136–144, Corvallis, OR, 2005. AUAI Press.

[3] B. R. Cobb and P. P. Shenoy. Nonlinear deterministic relationships in Bayesian networks. In L. Godo, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 8th European Conference, ECSQARU 2005*, Lecture Notes in Artificial Intelligence 3571, pages 27–38, Berlin, 2005. Springer.

[4] B. R. Cobb, P. P. Shenoy, and R. Rumi. Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials. *Statistics & Computing*, 16(3):293–308, 2006.

[5] R. M. Curds. *Propagation techniques in probabilistic expert systems.* PhD thesis, Department of Statistical Science, University College London, London, UK, 1997.

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

[7] R. A. Howard and J. E. Matheson. Influence diagrams. In R. A. Howard and J. E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 719–762. Strategic Decisions Group, 1984.

[8] F. Jensen, F. V. Jensen, and S. L. Dittmer. From influence diagrams to junction trees. In R. L. de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 10th Conference*, pages 367–373, San Francisco, CA, 1994. Morgan Kaufmann.

[9] F. V. Jensen, S. L. Lauritzen, and K. G. Olesen. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, 4:269–282, 1990.

[10] S. L. Lauritzen and F. Jensen. Stable local computation with conditional Gaussian distributions. *Statitistics and Computing*, 11:191–203, 2001.

[11] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50(2):157–224, 1988.

[12] U. Lerner and R. Parr. Inference in hybrid networks: Theoretical limits and practical algorithms. In J. Breese and D. Koller, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 17th Conference*, pages 310–318, San Francisco, CA, 2001. Morgan Kaufmann.

[13] U. Lerner, E. Segal, and D. Koller. Exact inference in networks with discrete children of continuous parents. In J. Breese and D. Koller, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 17th Conference*, pages 319–328, San Francisco, CA, 2001. Morgan Kaufmann.

[14] S. Moral, R. Rumí, and A. Salmerón. Mixtures of truncated exponentials in hybrid Bayesian networks. In S. Benferhat and P. Besnard, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 6th European Conference, ECSQARU-2001*, Lecture Notes in Artificial Intelligence 2143, pages 156–167, Berlin, 2001. Springer.

[15] K. P. Murphy. A variational approximation for Bayesian networks with discrete and continuous latent variables. In K. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 15th Conference*, pages 457–466, San Francisco, CA, 1999. Morgan Kaufmann.

[16] S. M. Olmsted. *On representing and solving decision problems*. PhD thesis, Department of Engineering-Economic Systems, Stanford University, Stanford, CA., 1983.

[17] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[18] I. J. Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics*, 4(45-99):112–141, 1946.

[19] L. L. Schumaker. *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge, UK, third edition, 2007.

[20] R. D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, 1986.

[21] P. P. Shenoy. Valuation-based systems for Bayesian decision analysis. *Operations Research*, 40(3):463–484, 1992.

[22] P. P. Shenoy. Inference in hybrid Bayesian networks using mixtures of Gaussians. In R. Dechter and T. Richardson, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 22nd Conference*, pages 428–436, Corvallis, OR, 2006. AUAI Press.

[23] P. P. Shenoy and G. Shafer. 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, 1990.

[24] P. P. Shenoy and J. C. West. Inference in hybrid Bayesian networks using mixtures of polynomials. Working Paper 321, University of Kansas School of Business, Lawrence, KS, May 2009.

[25] P. P. Shenoy and J. C. West. Inference in hybrid Bayesian networks with deterministic variables. In C. Sossai and G. Chemello, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty–10th ECSQARU*, Lecture Notes in Artificial Intelligence 5590, pages 46–58, Berlin, 2009. Springer-Verlag.