Research Publishing Services

t:+65-6492 1137

f:+65-6747 4355

e:enquiries@rpsonline.com.sg

Proceedings of the

2011 11th International Conference on

Intelligent Systems Design and Applications

22 – 24 November 2011

Córdoba, Spain

Sebastián Ventura, *University of Córdoba, Spain*
Ajith Abraham, *Machine Intelligence Research (MIR) Labs, USA*
Krzysztof Cios, *Virginia Commonwealth University, USA*
Cristóbal Romero, *University of Córdoba, Spain*
Francesco Marcelloni, *University of Pisa, Italy*
José Manuel Benítez, *University of Granada, Spain*
Eva Gibaja, *University of Córdoba, Spain*

IEEE Catalog Number: CFP11394-CDR
ISBN: 978-1-4577-1675-1

Sponsors

Technical Sponsors

Technical Supporter

IEEE SMC Society – Technical Committee on Soft Computing

Some Practical Issues in Inference in Hybrid Bayesian Networks with Deterministic Conditionals

Prakash P. Shenoy
School of Business
University of Kansas
Lawrence, KS 66045-7601 USA
pshenoy@ku.edu

Rafael Rumí and Antonio Salmerón
Dept. Statistics and Applied Mathematics
University of Almería
E-04120 Almería, Spain
{rrumi, antonio.salmeron}@ual.es

***Abstract*—In this paper we analyze the use of hybrid Bayesian networks in domains that include deterministic conditionals for continuous variables. We show how exact inference can become infeasible even for small networks, due to the difficulty in handling functional relationships. We compare two strategies for carrying out the inference task, using mixtures of polynomials (MOPs) and mixtures of truncated exponentials (MTEs).**

***Index Terms*—hybrid Bayesian networks; mixtures of polynomials; mixtures of truncated exponentials; deterministic conditionals;**

I. Introduction

Hybrid Bayesian networks are Bayesian networks that include discrete and continuous random variables. The first proposal of an efficient algorithm for handling this kind of Bayesian networks was proposed for the case in which the joint distribution is *mixture of Gaussians* (MoG) [1], [2]. Some limitations of the MoG model are that it is not compatible with network topologies in which discrete variables have continuous parents, and all conditionals for continuous variables have to be conditional linear Gaussians.

A more general inference procedure, based on the use of *mixtures of truncated exponentials* (MTEs), was proposed in [3]. The MTE model does not impose any structural restriction to its corresponding networks, and is compatible with any efficient algorithm for exact inference that requires only the combination and marginalization operations, as the Shenoy-Shafer [4] and variable elimination methods [5]. Furthermore, MTEs have shown a remarkable ability for fitting many commonly used univariate probability distributions [6], [7].

The most recent proposal for dealing with hybrid Bayesian networks is based on the use of *mixtures of polynomials* (MOPs) [8]. Like MTEs, MOPs have high expressive power, but the latter are superior in dealing with deterministic conditionals for continuous variables [9], [10].

In this paper we discuss some practical issues that have to be addressed in order to make inference in hybrid Bayesian networks feasible when deterministic conditionals are present. We compare the performance of MOPs and MTEs in this context, through an example consisting of a stochastic PERT network [11].

II. MTEs and MOPs

In this section we formally define the MTE and MOP models, which will be used throughout the paper. We will use uppercase letters to denote random variables, and boldfaced uppercase letters to denote random vectors, e.g. $\mathbf{X} = \{X_1, \ldots, X_n\}$, and its domain will be written as $\Omega_{\mathbf{X}}$. By lowercase letters x (or $\mathbf{x}$) we denote some element of Ω_X (or $\Omega_{\mathbf{X}}$). The MTE model [3] is defined as follows.

Definition 1: Let $\mathbf{X}$ be a mixed n-dimensional random vector. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)^\mathsf{T}$ and $\mathbf{Z} = (Z_1, \ldots, Z_c)^\mathsf{T}$ be the discrete and continuous parts of $\mathbf{X}$, respectively, with $c + d = n$. We say that a function $f : \Omega_{\mathbf{X}} \mapsto \mathbb{R}_0^+$ is a *mixture of truncated exponentials (MTE) potential* if for each fixed value $\mathbf{y} \in \Omega_{\mathbf{Y}}$ of the discrete variables $\mathbf{Y}$, the potential over the continuous variables $\mathbf{Z}$ is defined as:

$$f(\mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{\mathbf{b}_i^\mathsf{T}\mathbf{z}\right\}, \tag{1}$$

for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$, where $a_i \in \mathbb{R}$ and $\mathbf{b}_i \in \mathbb{R}^c$, $i = 1, \ldots, m$. We also say that f is an MTE potential if there is a partition $D_1, \ldots, D_k$ of $\Omega_{\mathbf{Z}}$ into hypercubes and in each one of them, f is defined as in Eq. (1). In this case, we say f is a k-piece, m-term MTE potential.

Mixtures of polynomials (MOPs) were initially proposed as modeling tools for hybrid Bayesian networks in [8]. The original definition is similar to MTEs, in the sense that they are piecewise functions defined on hypercubes. A more general definition was given in [10], where the hypercube condition is relaxed. The details are as follows.

Definition 2: Let $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ be as in Def. 1. We say that a function $f : \Omega_{\mathbf{X}} \mapsto \mathbb{R}_0^+$ is a *mixture of polynomials (MOP) potential* if for each fixed value $\mathbf{y} \in \Omega_{\mathbf{Y}}$ of the discrete variables $\mathbf{Y}$, the potential over the continuous variables $\mathbf{Z}$ is defined as:

$$f(\mathbf{z}) = P(\mathbf{z}), \tag{2}$$

for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$, where $P(\mathbf{z})$ is a multivariate polynomial in variables $\mathbf{Z} = (Z_1, \ldots, Z_c)^\mathsf{T}$. We also say that f is a MOP potential if there is a partition $D_1, \ldots, D_k$ of $\Omega_{\mathbf{Z}}$ into *hyper-rhombuses* and in each one of them, f is defined as in Eq. (2).

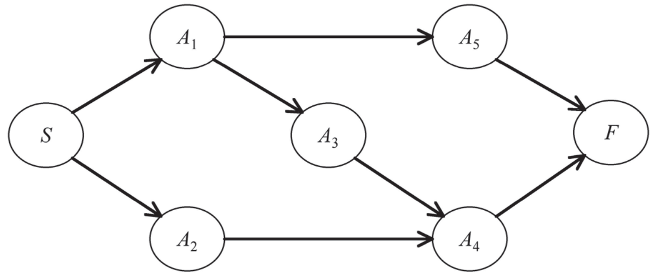

Fig. 1. A PERT network with five activities

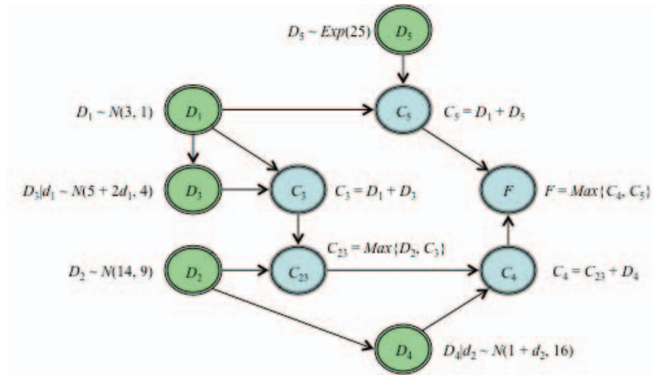

Fig. 2. A Bayesian network representing the PERT network in Fig. 1

The fact that the elements in the partition are hyper-rhombuses, means that for any order of the variables $Z_1, \ldots, Z_c$, for each D_i it holds that

$$\begin{aligned} l_{1i} \leq \quad & z_1 \quad \leq u_{1i}, \\ l_{2i}(z_1) \leq \quad & z_2 \quad \leq u_{2i}(z_1), \\ & \vdots \\ l_{ci}(z_1, \ldots, z_{c-1}) \leq \quad & z_c \quad \leq u_{ci}(z_1, \ldots, z_{c-1}), \end{aligned}$$

where l_{1i} and u_{1i} are constants, and $l_{ji}(z_1, \ldots, z_{j-1})$ and $u_{ji}(z_1, \ldots, z_{j-1})$ are linear functions of $z_1, \ldots, z_{j-1}$ for $j = 2, \ldots, c$, and $i = 1, \ldots, k$.

MTEs and MOPs are closed under multiplication, addition, and integration. However, integrating over hyper-rhombuses is in general more complex than over hypercubes. The advantage is that using hyper-rhombuses, it is easier to represent models like the *conditional linear Gaussian*, where the domain of a variable may depend on the value of its parents in the network. Unfortunately, MTEs cannot be defined on hyper-rhombuses, as the integration operation would not remain closed for that class.

III. A PERT Hybrid Bayesian Network

We will illustrate the inference process in hybrid Bayesian networks using a *stochastic PERT network* [11]. PERT stands for *Program Evaluation and Review Technique*, and is one of the commonly used project management techniques [12]. PERT networks are directed acyclic networks where the nodes represent duration of activities and the arcs represent precedence constraints in the sense that before we can start any activity, all the parent activities have to be completed. The term *stochastic* refers to the fact that the duration of activities are modeled as continuous random variables.

Fig. 1 shows a PERT network with 5 activities ($A_1, \ldots, A_5$). Nodes S and F represent the start and finish times of the project. The links among activities mean that an activity cannot be started until after all its predecessors have been completed. Assume we are informed that the durations of A_1 and A_3 are positively correlated, and the same is true with A_2 and A_4. Then, this PERT network can be transformed into a Bayesian network as follows.

Let D_i and C_i denote the duration and the completion time of the activity i, respectively. The activity nodes in the PERT network are replaced with activity completion times in the BN. Next, activity durations are added with a link from D_i to C_i, so that each activity will be represented by two nodes, its duration D_i and its completion time C_i. Notice that the completion times of the activities which do not have any predecessors will be the same as their durations. Hence, activities A_1 and A_2 will be represented just by their durations, D_1 and D_2. As A_3 and A_1 have positively correlated durations, a link will connect D_1 and D_3 in the Bayesian network. For the same reason, another link will connect D_2 and D_4. The completion time of A_3 is $C_3 = D_1 + D_3$. Let $C_{23} = \max\{D_2, C_3\}$ denote the completion time of activities A_2 and A_3. The completion time of activity A_5 is $C_5 = D_1 + D_5$, and for activity A_4, it is $C_4 = C_{23} + D_4$.

We assume that the project start time is zero and each activity is started as soon as all the preceding activities are completed. Accordingly, F represents the completion time of the project, which is the maximum of C_5 and C_4. The resulting PERT Bayesian network is given in Fig. 2.

Notice that the conditionals for the variables C_3, C_{23}, C_4, C_5 and F are deterministic, in the sense that their conditional distributions given their parents have zero variances. On the other hand, variables $D_1, \ldots, D_5$ are continuous random variables, and their corresponding conditional distributions are depicted next to their corresponding nodes in Fig. 2. The parameters μ and σ^2 of the Normal distribution are in units of $days$ and $days^2$, respectively. The parameter μ of the exponential distribution is in units of $days$.

The problem of inference in hybrid BNs with deterministic conditionals has been studied in [9]. The first problem we find when attempting to carry out inference in the network in Fig. 2 is concerned with variables C_{23} and F, as their

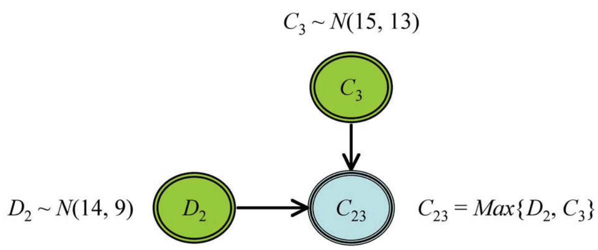

Fig. 3. A max conditional

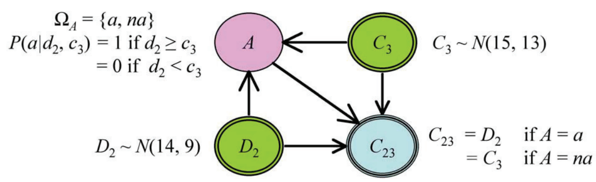

Fig. 4. Transformation of a max conditional

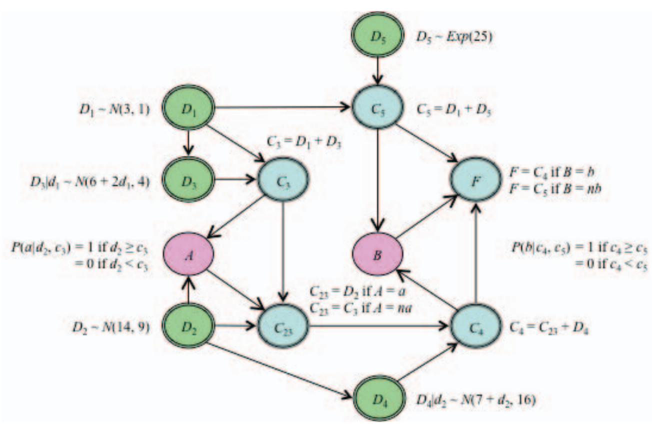

Fig. 5. A hybrid Bayesian network representing the PERT network in Fig. 1

deterministic conditional is the $\max$ function, which is not easy to use directly in probabilistic inference. However, we can convert the $\max$ deterministic function to a linear function [9]. Consider the situation in Fig. 3. We can remove the $\max$ by introducing a discrete indicator variable A with two states a and na, which denote whether $D_2 \geq C_3$ or $D_2 < C_3$, respectively, obtaining the equivalent representation displayed in Fig. 4. After applying this transformation to C_{23} and F in Fig. 2, we obtain the hybrid Bayesian network shown in Fig. 5. Notice that it is hybrid because now there are both discrete and continuous variables in the same network.

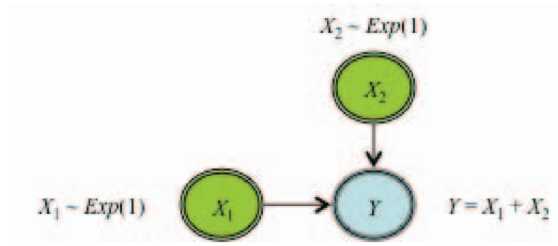

Fig. 6. A sum conditional

A. MTE representation of conditionals

MTEs can be used to accurately approximate several univariate distributions, including the ones in the PERT hybrid BN [6]. The approximation of conditional densities using MTEs is more difficult, as it was recently shown [13] that the value of any conditional MTE density has to be constant with respect to the value of the parent variables. Therefore, it means that a conditional density is approximated by MTEs by partitioning the domain of the parent variables into hypercubes and then fitting a univariate MTE in each hypercube. The resulting conditional MTE density is called a *mixed tree* [14].

An additional problem is found when attempting to deal with deterministic conditionals. MTEs are not closed with respect to the *convolution* operation required by the sum conditional. Consider the situation depicted in Fig. 6. The marginal PDF of Y is $Gamma[r = 2, \mu = 1]$, which is not an MTE function. The reason is that such a marginal is obtained through the so-called *convolution* operation as follows:

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X_1}(x_1)\, f_{X_2}(y - x_1)\, \mathrm{d}x_1 \\
&= \int_0^y e^{-x_1}\, e^{-(y-x_1)}\, \mathrm{d}x_1 \qquad (3) \\
&= y\, e^{-y} \qquad \text{if } y > 0,
\end{aligned}
$$

which is not an MTE as y appears outside the exponent of e, thus violating Def. 1. This is a consequence of the fact that even though $f_{X_2}(x_2)$ is defined on hypercubes, $f_{X_2}(y - x_1)$ is no longer defined on hypercubes, and therefore, the limits of integration in Eq. 3 are not all constants.

One solution to this problem is to approximate the function $f_{X_2}(y - x_1)$ on hypercubes using a *mixed tree* approach as described in [14]. If $Z \sim N(0, 1)$, $f(\cdot)$ is an MTE approximation of the PDF of Z, and $Y = \sigma Z + \mu$, where σ and μ are real constants, then $Y \sim N(\mu, \sigma^2)$, and an MTE approximation of the PDF of Y is given by $g(y) = \frac{1}{|\sigma|} f(\frac{y-\mu}{\sigma})$. For example, suppose $X_1 \sim N(3, 1)$, $X_2|x_1 \sim N(6 + 2x_1, 4)$ and $Y = X_1 + X_2$. Suppose $f(\cdot)$ is an MTE approximation of the PDF of $N(0, 1)$ on the domain $(-3, 3)$ (see [6], [7]). Then, $g_1(x_1) = f(x_1 - 3)$ is an MTE approximation of the PDF of $X_1 \sim N(3, 1)$ on the domain $(0, 6)$. Finally, $g_2(x_1, x_2)$, as described in Eq. (4), is an MTE approximation of the conditional PDF of $X_2 \mid x_1$:

$$g_2(x_1, x_2) = \begin{cases} f(\frac{x_2-8}{2})/2 & \text{if } 0 \le x_1 < 2, \\ f(\frac{x_2-12}{2})/2 & \text{if } 2 \le x_1 < 4, \\ f(\frac{x_2-16}{2})/2 & \text{if } 4 \le x_1 \le 6. \end{cases} \tag{4}$$

Notice that this mixed-tree method can also be used with MOPs.

B. MOP representation of conditionals

The problem of fitting univariate density functions using MOPs was studied in [8]. Also, a procedure for approximating conditionals was given in that same work. More recently, the ability of MOPs to fit conditional distributions was significantly improved by means of allowing the functions to be defined into hyper-rhombuses rather than into hypercubes [10].

To illustrate this improvement, consider three random variables D_1, D_3, and C_3, with $D_1 \sim N(3,1)$, $D_3|d_1 \sim N(6+2d_1, 4)$, and $C_3 = D_1 + D_3$. Suppose $f(\cdot)$ is a MOP approximation of the PDF of $N(0,1)$ on the domain $(-3,3)$. Then, $g_1(d_1) = f(d_1 - 3)$ is a MOP approximation of the PDF of $D_1 \sim N(3,1)$ on the domain $(0,6)$. Finally, $g_2'(d_1, d_3) = f(\frac{d_3-6-2d_1}{2})/2$ is a MOP approximation of the conditional PDF of D_3 given d_1. Notice that $g_2'(d_1, d_3)$ is defined on hyper-rhombuses, i.e. $0 \le \frac{d_3-6-2d_1}{2} < 3$, etc.

Unlike MTEs, MOPs are closed under the operations required for sum conditionals. Thus, after the elimination of D_3, $g_4(d_1, c_3) = g_2'(d_1, c_3 - d_1)$ is a MOP, and after the elimination of D_1,

$$g_5(c_3) = \int_{-\infty}^{\infty} g_1(d_1) g_4(d_1, c_3)\, \mathrm{d}d_1$$

is also a MOP.

In order to compare the hyper-rhombus and the hypercube approaches, we will model the same problem using MOPs defined on hypercubes. Thus, assume $f(z)$ is a 2-piece, 3-degree MOP approximation of the PDF of $N(0,1)$ [8]. Then as before, $g_1(d_1) = f(d_1 - 3)$ is a 2-piece, 3-degree MOP approximation of the PDF of $N(3,1)$ on $(0,6)$, and $g_2(d_1, d_3)$ as defined in Equation (4) is a 6-piece 3-degree MOP approximation of the conditional PDF of D_3 given D_1.

The combination of $g_1(d_1)$ and $g_2(d_1, d_3)$ results in a 8-piece, 6-degree joint PDF of D_1, D_3. Computing $g_6(d_3) = \int_{-3}^{3} g_1(d_1)\, g_2(d_1, d_3)\, \mathrm{d}d_1$ takes about 1.8 seconds in *Mathematica*® v. 8.0.1, and results in a 8-piece, 3-degree MOP approximation of the PDF of D_3 on the interval $(0, 24)$. The computation of such density following the hyper-rhombus approach takes 4 seconds in the same platform. Therefore, computations using hypercubes are faster, as the integration operation is easier. However, the hypercube approximation (using mixed trees [14]) of the conditional PDF of D_3 has more pieces (6) than the hyper-rhombus approximation (2). Also, the hypercube approximation of the joint PDF of D_1 and D_3 has more pieces (8) than the hyper-rhombus approximation (4), and both have the same degree. The hypercube approximation of the marginal PDF of D_3 has more pieces (8) than the hyper-rhombus approximation (4), but smaller degree (3 vs. 7). In both cases, after integration, the number of pieces remains unchanged. For hypercubes, after integration, the degree goes down from 6 to 3, whereas for hyper-rhombuses, the degree goes up from 6 to 7.

IV. Solving the PERT hybrid Bayesian network

A. Solution using MOPs

We start with a 2-piece, 3-degree MOP $f_1(\cdot)$ as an approximation of the PDF of the standard normal. Using $f_1(\cdot)$, we define a 2-piece, 3-degree MOP approximation of the PDFs of D_1 and D_2. Using mixed trees, we define a 6-piece, 3-degree MOP of the conditional PDFs of D_3 and D_4. We used a 2-piece, 3-degree MOP approximation of the $Exp(25)$ density for D_5.

The initial potentials in the PERT hybrid BN in Fig. 5 are denoted as follows. The conditionals for $D_1, \ldots, D_5$ are $f_{D_1}, \ldots, f_{D_5}$ respectively. The deterministic conditional for C_3 is $f_{C_3}(d_1, d_3, c_3) = \delta(c_3 - d_1 - d_3)$, where δ stands for the Dirac's delta function. The deterministic conditional for C_{23} is $f_{C_{23a}}(d_2, c_3, c_{23}) = \delta(c_{23} - d_2)$ if $A = a$, and $f_{C_{23na}}(d_2, c_3, c_{23}) = \delta(c_{23} - c_3)$ if $A = na$. For C_4, we have $f_{C_4}(c_{23}, c_4, d_4) = \delta(c_4 - c_{23} - d_4)$. For the discrete variable A, we have $p_{Aa}(d_2, c_3)$ denoting $P(A = a|d_2, c_3)$ and $p_{Ana}(d_2, c_3)$ for $P(A = na|d_2, c_3)$. The conditional for C_5 is denoted as $f_{C_5}(d_1, d_5, c_5) = \delta(c_5 - d_1 - d_5)$. For the discrete variable B, we have $p_{Bb}(c_4, c_5)$ meaning $P(B = b|c_4, c_5)$ and $p_{Bnb}(c_4, c_5)$ meaning $P(B = nb|c_4, c_5)$. Finally, the conditional for F is denoted as $f_{Fb}(c_4, c_5, f) = \delta(f - c_4)$ if $B = b$ and as $f_{Fnb}(c_4, c_5, f) = \delta(f - c_5)$ if $B = nb$.

The goal is to compute the marginal density for F. To that end, we choose an elimination order of the remaining variables in the network, namely $D_5, D_3, D_1, D_4, D_2, C_{23}, A, C_3, C_4, C_5$, and B.

The deletion of D_5 is carried out by computing

$$f_2(d_1, c_5) = \int f_{D_5}(d_5)\, f_{C_5}(d_1, d_5, c_5)\, \mathrm{d}d_5 = f_{D_5}(c_5 - d_1).$$

Next we remove D_3 and D_1 by computing

$$\begin{aligned} f_3(d_1, c_3) &= \int f_{D_3}(d_1, d_3)\, f_{C_3}(d_1, d_3, c_3)\, \mathrm{d}d_3 \\ &= f_{D_3}(d_1, c_3 - d_1), \\ f_4(c_3, c_5) &= \int f_{D_1}(d_1)\, f_2(d_1, c_5)\, \mathrm{d}d_1, \end{aligned}$$

where f_4 is computed as a 24-piece, 6-degree MOP f_4, and takes 6.6 seconds to be calculated. Next, we re-approximate f_4 by a 8-piece, 3-degree MOP using Lagrange interpolating polynomials [15], in order to reduce the complexity (number of pieces and degree).

The next step is to delete D_4:

$$f_5(c_{23}, d_2, c_4) = \int f_{D_4}(d_2, d_4)\, f_{C_4}(c_{23}, c_4, d_4)\, \mathrm{d}d_4 = f_{D_4}(d_2, c_4 - c_{23}).$$

Next, the deletion of D_2 yields the following functions:

$$f_{6a}(c_3, c_{23}, c_4) = \int f_{D_2}(d_2)\, f_{C_{23a}}(d_2, c_3, c_{23})\, f_5(c_{23}, d_2, c_4) p_{Aa}(d_2, c_3)\, \mathrm{d}d_2 = f_{D_2}(c_{23})\, f_5(c_{23}, c_{23}, c_4)\, p_{Aa}(c_{23}, c_3)$$

$$f_{6na}(c_3, c_{23}, c_4) = f_{C_{23na}}(d_2, c_3, c_{23}) \int f_{D_2}(d_2)\, f_5(c_{23}, d_2, c_4)\, p_{Ana}(d_2, c_3)\, \mathrm{d}d_2 = f_{C_{23na}}(d_2, c_3, c_{23}) \int_{-\infty}^{c_3} f_{D_2}(d_2)\, f_5(c_{23}, d_2, c_4)\, \mathrm{d}d_2.$$

Next we marginalize out C_{23} as follows:

$$f_{7a}(c_3, c_4) = \int f_{6a}(c_3, c_{23}, c_4)\, \mathrm{d}c_{23} = \int f_{D_2}(c_{23})\, f_5(c_{23}, c_{23}, c_4)\, p_{Aa}(c_{23}, c_3)\, \mathrm{d}c_{23} = \int_{c_3}^{\infty} f_{D_2}(c_{23})\, f_5(c_{23}, c_{23}, c_4)\, \mathrm{d}c_{23}.$$

$$f_{7na}(c_3, c_4) = \int f_{6na}(c_3, c_{23}, c_4)\, \mathrm{d}c_{23} = \int_{-\infty}^{c_3} f_{D_2}(d_2)\, f_5(c_3, d_2, c_4)\, \mathrm{d}d_2.$$

For deleting A, we compute $f_7(c_3, c_4) = f_{7a}(c_3, c_4) + f_{7na}(c_3, c_4)$. f_7 is computed as a 30-piece, 6-degree MOP and it requires 54 seconds to be computed. Given the large number of pieces, we re-approximate f_7 by a 8-piece, 5-degree MOP using Lagrange interpolating polynomials [15].

After further elimination of C_3, f_8 is computed as a 45-piece, 8-degree MOP in 16 seconds: $f_8(c_4, c_5) = \int f_4(c_3, c_5)\, f_7(c_3, c_4)\, \mathrm{d}c_3$.

The elimination of C_4 takes 24.7 seconds, and consists of computing

$$f_{9b}(c_5, f) = \int p_{Bb}(c_4, c_5)\, f_{Fb}(c_4, c_5, f)\, f_8(c_4, c_5)\, \mathrm{d}c_4 = p_{Bb}(f, c_5)\, f_8(f, c_5)$$

and

$$f_{9nb}(c_5, f) = f_{Fnb}(c_4, c_5, f) \int p_{Bnb}(c_4, c_5)\, f_8(c_4, c_5)\, \mathrm{d}c_4 = f_{Fnb}(c_4, c_5, f) \int_{-\infty}^{c_5} f_8(c_4, c_5)\, \mathrm{d}c_4.$$

The deletion of C_5 takes 57.8 seconds, required to calculate $f_{10b}(f) = \int f_{9b}(c_5, f)\, \mathrm{d}c_5$ and $f_{10nb}(f) = \int f_{9nb}(c_5, f)\, \mathrm{d}c_5$.

Finally, by deleting B, we obtain $f_{11}(f) = f_{10b}(f) + f_{10nb}(f)$, which is a 14-piece, 9-degree MOP representation of the PDF of F. Using f_{11}, we compute the expectation and standard deviation of F: $E(F) = 51.98$ days and $\sigma_F = 22.7$ days, respectively. A plot of f_{11} is shown in Fig. 7. Evaluation of the entire Mathematica notebook (all computations including re-approximation) takes about 180 seconds.

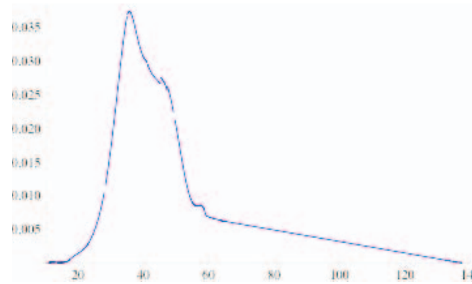

Fig. 7. MOP approximation of the marginal for F

Re-approximation of f_4 and f_7 is crucial to solving the PERT hybrid BN. Without re-approximation of these two functions, computing f_8 results in "lack of memory" warning message even though the computer is equipped with 8GB of RAM.

B. Solution using MTEs

We start with a 1-piece, 6-terms MTE, $f_1(\cdot)$, as an approximation of the PDF of the standard normal density [7]. Using $f_1(\cdot)$, we define a 1-piece, 6-terms MTE approximation of the PDF of D_1 and also of the PDF of D_2. Using mixed trees, we define a 2-piece, 6-terms MTE approximation of the conditional PDFs of D_3 and D_4. Since the PDF of the exponential distribution is already a 1-piece, 1-term MTE function, no approximation is needed for the PDF of D_5.

In order to compute the marginal for F, we use the same elimination order as for the case of MOPs. After the elimination of D_5, D_3, and D_1, we compute a 4-piece, 6-terms MTE $f_4(c_3, c_5)$ in 8.02 seconds. After further elimination of D_4, D_2, C_{23} and A, we compute a 16-piece, 72-terms MTE $f_7(c_3, c_5)$ in 65 seconds. We did not re-approximate either f_4 or f_7, as Lagrange interpolating polynomials cannot be applied in this case. We are currently looking for a method to do similar re-approximations with MTEs as we did with MOPs.

After further elimination of C_3, we obtain a 20-piece, 6-terms MTE $f_8(c_4, c_5)$ in 1,460 seconds. Then we delete C_4 (need 47 seconds) and C_5 (need 35 seconds), and B, obtaining a 8-piece, 17-terms MTE approximation of the marginal PDF of F. A graph of it is displayed in Fig. 8. We compute the expectation and standard deviation: $E(F) = 44.17$ days and $\sigma_F = 16.32$ days, respectively. The evaluation of the entire Mathematica notebook (all computations) takes about 1,600 seconds (about 27 minutes).

C. Solution using simulation

In order to have a clearer idea of the true marginal for F, we compute an estimation of it by simulating a sample of size $1,000,000$ of $D_1, \ldots, D_5$ and then computing the corresponding values of C_3, C_{23}, C_4, C_5 and F for each record in the sample, according to their definition. Then, we fitted a

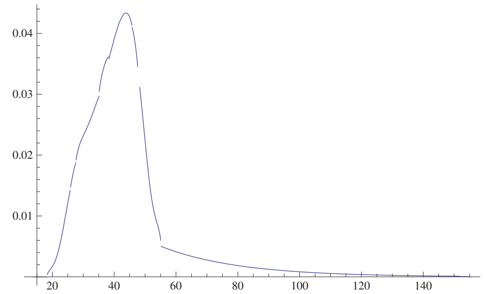

Fig. 8. MTE approximation of the marginal for F

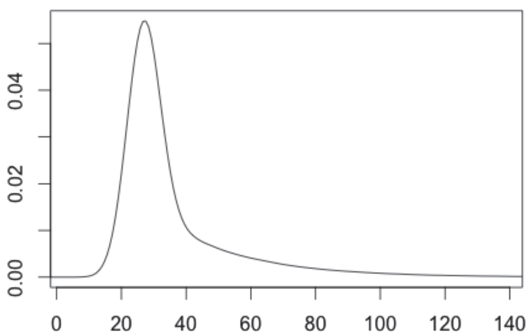

Fig. 9. Kernel approximation of the marginal PDF of F

Gaussian kernel density to the values obtained for F. The result is displayed in Fig. 9. In this case, point estimates of the expectation and standard deviation of F are $\hat{E}(F) = 36.19$ and $\hat{\sigma}_F = 20.28$ days, respectively.

V. Summary and Conclusions

We have described some practical issues in solving hybrid BNs that include deterministic conditionals using MTEs and MOPs, and have solved a PERT hybrid BN consisting of 2 discrete and 10 continuous variables, 5 of which have linear deterministic conditionals.

One key observation is that in the process of solving the PERT hybrid BN, some of the intermediate potentials have a large number of pieces, some of which are defined on lower dimensions and which have no useful information. One solution to this is to re-approximate these potentials with a smaller number of pieces and fewer degrees/terms. In the case of MOPs, this can be done using Lagrange interpolating polynomials. In the case of MTEs, a method for re-approximation needs to be devised.

We plan to solve the PERT hybrid BN using MOPs defined on hyper-rhombuses to keep the number of pieces to a minimum, and compare the running time and accuracy with the corresponding results using hypercubes. Shenoy [15] describes the use of Lagrange interpolating polynomials with Chebyshev points for approximating univariate functions by MOPs. However, more work needs to be done in re-approximating high-dimensional joint and conditional functions by MOP using Lagrange interpolating polynomials.

In large networks with several deterministic conditionals, the inference process can be too costly even using re-approximation. We plan to explore approaches that keep the complexity (degree and # pieces) below some specified threshold. One way can be to consider fixed-size potentials and re-approximation after each marginalization operation. Another approach is to use a simulation method that is adapted for presence of deterministic conditionals.

Acknowledgment

This work has been supported by the Spanish Ministry of Science and Innovation, through projects TIN2007-67418-C03-02 and TIN2010-20900-C04-02, and by ERDF (FEDER) funds, and by a sabbatical partially funded by the University of Kansas, Lawrence, KS, to one of the co-authors. Thanks to Serafín Moral for suggesting the re-approximation strategy.

References

[1] S. Lauritzen, "Propagation of probabilities, means and variances in mixed graphical association models," *Journal of the American Statistical Association*, vol. 87, pp. 1098–1108, 1992.
[2] S. L. Lauritzen and F. Jensen, "Stable local computation with conditional Gaussian distributions," *Statistics and Computing*, vol. 11, pp. 191–203, 2001.
[3] S. Moral, R. Rumí, and A. Salmerón, "Mixtures of truncated exponentials in hybrid Bayesian networks," in *ECSQARU'01. Lecture Notes in Artificial Intelligence*, vol. 2143, 2001, pp. 135–143.
[4] P. P. Shenoy and G. Shafer, "Axioms for probability and belief function propagation," in *Uncertainty in Artificial Intelligence 4*, R. Shachter, T. Levitt, J. Lemmer, and L. Kanal, Eds. North Holland, Amsterdam, 1990, pp. 169–198.
[5] N. Zhang and D. Poole, "Exploiting causal independence in Bayesian network inference," *Journal of Artificial Intelligence Research*, vol. 5, pp. 301–328, 1996.
[6] B. Cobb, P. P. Shenoy, and R. Rumí, "Approximating probability density functions with mixtures of truncated exponentials," *Statistics and Computing*, vol. 16, pp. 293–308, 2006.
[7] H. Langseth, T. Nielsen, R. Rumí, and A. Salmerón, "Parameter estimation and model selection for mixtures of truncated exponentials," *International Journal of Approximate Reasoning*, vol. 51, pp. 485–498, 2010.
[8] P. P. Shenoy and J. West, "Inference in hybrid Bayesian networks using mixtures of polynomials," *International Journal of Approximate Reasoning*, vol. 52, pp. 641–657, 2011.
[9] ——, "Extended Shenoy-Shafer architecture for inference in hybrid Bayesian networks with deterministic conditionals," *International Journal of Approximate Reasoning*, vol. 52, pp. 805–818, 2011.
[10] P. P. Shenoy, "A re-definition of mixtures of polynomials for inference in hybrid Bayesian networks," *ECSQARU'11. Lecture Notes in Artificial Intelligence*, vol. 6717, pp. 98–109, 2011.
[11] E. Cinicioglu and P. P. Shenoy, "Using mixtures of truncated exponentials for solving stochastic pert networks," in *Proceedings of the Eighth Workshop on Uncertainty Processing (WUPES-09)*, J. Vejnarova and T. Kroupa, Eds., Prague, 2009, pp. 269–283.
[12] D. G. Malcolm, J. H. Roseboom, C. E. Clark, and W. Fazar, "Application of a technique for research and development program evaluation," *Operations Research*, vol. 7, pp. 646–669, 1959.
[13] H. Langseth, T. Nielsen, R. Rumí, and A. Salmerón, "Maximum likelihood learning of conditional MTE distributions," *ECSQARU 2009. Lecture Notes in Computer Science*, vol. 5590, pp. 240–251, 2009.
[14] S. Moral, R. Rumí, and A. Salmerón, "Approximating conditional MTE distributions by means of mixed trees," in *ECSQARU'03. Lecture Notes in Artificial Intelligence*, vol. 2711, 2003, pp. 173–183.
[15] P. P. Shenoy, "Some issues in using mixtures of polynomials for inference in hybrid Bayesian networks," University of Kansas Business School, Lawrence, KS, Working Paper No. 323, 2010.