Appeared in: Proceedings of the Tenth Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2004, pp. 429—436.

Approximating Probability Density Functions with Mixtures of Truncated Exponentials

Barry R. Cobb
Prakash P. Shenoy
University of Kansas School of Business
1300 Sunnyside Ave., Summerfield Hall
Lawrence, KS 66045–7585
{brcobb, pshenoy}@ku.edu

Rafael Rumí
Dept. of Statistics
and Applied Mathematics
University of Almeria
Almeria, Spain
rrumi@ual.es

Abstract

Mixtures of truncated exponentials (MTE) potentials are an alternative to discretization for approximating probability density functions (PDF's). This paper presents MTE potentials that approximate standard PDF's and applications of these potentials for solving inference problems in hybrid Bayesian networks.

Keywords: Hybrid Bayesian networks, MTE potentials, Shenoy-Shafer architecture.

1 Introduction

Bayesian networks model knowledge about propositions in uncertain domains using graphical and numerical representations. At the qualitative level, a Bayesian network is a directed acyclic graph where nodes represent variables and the (missing) edges represent conditional independence relations among the variables. At the numerical level, a Bayesian network consists of a factorization of a joint probability distribution into a set of conditional distributions, one for each variable in the network. Hybrid Bayesian networks contain both discrete probability mass functions (PMF's) and continuous conditional probability density functions (PDF's) as numerical inputs.

Poland [6] proposes using a finite mixture of Gaussians to fit arbitrary continuous distributions for chance variables in hybrid Bayesian networks. An alternative to using mixtures of Gaussians for approximating continuous chance variables is mixtures of truncated exponentials (MTE) potentials [4]. General formulations for standard PDF's and procedures for estimating parameters in MTE potentials for approximating arbitrary PDF's can allow implementation of hybrid Bayesian networks to a broader range of problems.

In this paper, we describe MTE approximations for three standard probability distributions. The remainder of this paper is organized as follows. Section 2 defines MTE potentials and describes a method of estimating parameters for MTE potentials. Section 3 presents MTE approximations to standard PDF's. Section 4 demonstrates inference in hybrid Bayesian networks using MTE potentials. Section 5 summarizes the paper. This paper is a brief version of a longer unpublished working paper [2].

2 Estimating Parameters for Mixtures of Truncated Exponentials (MTE) Potentials

2.1 Notation

Random variables in a hybrid Bayesian network will be denoted by capital letters, e.g. A, B, C. Sets of variables will be denoted by boldface capital letters, $\mathbf{Y}$ if all variables are discrete, $\mathbf{Z}$ if all variables are continuous, or $\mathbf{X}$ if some of the components are discrete and some are continuous. If $\mathbf{X}$ is a set of variables, $\mathbf{x}$ is a configuration of specific states of those variables. The discrete, continuous, or mixed state space of $\mathbf{X}$ is denoted by $\Omega_{\mathbf{X}}$.

MTE probability potentials and discrete probability potentials are denoted by lowercase greek letters, e.g. α, β, γ. Subscripts are used for fragments of MTE potentials or conditional probability tables when different parameters or values are required for each configuration of a variable's discrete parents, e.g. α_1, β_2, γ_3. Subscripts are also used for discrete probabilities of elements of the state space, e.g. $\delta_0 = P(D = 0)$.

In graphical representations, continuous nodes in hybrid Bayesian networks are represented by double-border ovals, whereas discrete nodes are represented by single-border ovals.

2.2 MTE Potentials

A mixture of truncated exponentials (MTE) potential has the following definition [4].

MTE potential. Let $\mathbf{X}$ be a mixed n-dimensional random variable. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)$ and $\mathbf{Z} = (Z_1, \ldots, Z_c)$ be the discrete and continuous parts of $\mathbf{X}$, respectively, with $c + d = n$. A function $\phi : \Omega_{\mathbf{X}} \mapsto \mathbb{R}^+$ is an MTE potential if one of the next two conditions holds:

1. The potential ϕ can be written as

$$\phi(\mathbf{x}) = \phi(\mathbf{y}, \mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{ \sum_{j=1}^{d} b_i^{(j)} y_j + \sum_{k=1}^{c} b_i^{(d+k)} z_k \right\} \quad (1)$$

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

2. There is a partition $\Omega_1, \ldots, \Omega_k$ of $\Omega_{\mathbf{X}}$ verifying that the domain of continuous variables, $\Omega_{\mathbf{Z}}$, is divided into hypercubes, the domain of the discrete variables, $\Omega_{\mathbf{Y}}$, is divided into arbitrary sets, and such that ϕ is defined as

$$\phi(\mathbf{x}) = \phi_i(\mathbf{x}) \qquad \text{if } \mathbf{x} \in \Omega_i, \quad (2)$$

where each $\phi_i, i = 1, ..., k$ can be written in the form of equation (1) (i.e. each ϕ_i is an MTE potential on Ω_i).

2.3 Kullback-Leibler (KL) Divergence

When approximating a standard PDF with an MTE potential, we measure the Kullback–Leibler (KL) divergence introduced by the approximation and minimize this measure in the process of finding parameters for the MTE potential, subject to certain constraints.

The relative entropy or KL divergence [3] between PDF's $f_X(x)$ and $\tilde{f}_X(x)$ is defined as

$$D_{KL}(f_X(x) \,||\, \tilde{f}_X(x)) = \int_S f_X(x) \log \frac{f_X(x)}{\tilde{f}_X(x)}\, dx. \quad (3)$$

Define $p_{f_{Xi}}$ and $q_{\tilde{f}_{Xi}}$ as the probability masses of $f_X(x)$ and $\tilde{f}_X(x)$, respectively, in the interval $(x_{i-1}, x_i]$. A discrete approximation to the KL divergence statistic over a set of points $x_i, i = 0, ..., n$ can be calculated as follows:

$$D'_{KL}(f_X(x) \,||\, \tilde{f}_X(x)) = \sum_{i=1}^{n} p_{f_{Xi}} \log \frac{p_{f_{Xi}}}{q_{\tilde{f}_{Xi}}}. \quad (4)$$

The function $g(x) = \log\left(f_X(x)/\tilde{f}_X(x)\right)$ can be interpreted as the information contained in x for distinguishing between $f_X(x)$ and $\tilde{f}_X(x)$. Thus, KL divergence is the expectation of the information content over the domain S taken with respect to the distribution $f_X(x)$. By minimizing this expectation when determining parameters for MTE approximations to standard PDF's—subject to probability mass constraints—we ensure a small chance of distinguishing between results obtained from inference with standard PDF's and corresponding MTE approximations.

2.4 Estimation Procedure

The numerical representation of a hybrid Bayesian network requires a conditional probability potential for each variable in the network, given its parents. We first consider the problem of estimating parameters for an MTE potential approximating a marginal PDF. This technique can be extended in a straightforward way to estimate the parameters for a conditional MTE potential by using the mixed tree structure in [5].

2.4.1 Partitioning the Domain

To estimate the parameters of an MTE potential for a continuous variable X, a partition $\Omega_1, \ldots, \Omega_k$ of Ω_X must be determined. Typically, in each interval of the partition, the PDF to be approximated should show no changes in concavity/convexity or increase/decrease. To increase efficiency in the inference process, we may choose to exclude a small amount of probability density in the tails when approximating a PDF. PDF's whose basic shape does not change dramatically when the distribution parameters change, such as the normal PDF, can be fit by partitioning the domain with respect to changes in increase/decrease only.

Suppose the domain of the continuous variable has been divided into K intervals denoted $D_1, ..., D_K$. To estimate parameters for an MTE potential which approximates a standard PDF within a given interval D_k, we choose a set of points $x = (x_0, ..., x_n)$ by evenly dividing the portion of the domain of the PDF represented by D_k. A set of points $y = (y_0, ..., y_n)$ is determined by calculating the value of the PDF at each point x_i, $i = 0, ..., n$.

2.4.2 Approximation by Nonlinear Optimization

Defining an MTE approximation to a PDF $f_X(x; \mathbf{\Theta}_m)$ (abbreviated $f_X(x)$) requires estimating constants a_{0k}, a_{ik} and $b_{ik}^{(j)}$ in (1) for each interval D_k. We assume $\mathbf{\Theta}_m$ is an arbitrary vector of parameters of a standard PDF and the MTE approximation will be fitted for potential parameter vectors, $m = 1, ..., M$. The formulation in (1) allows an independent term (a_{0k}) and an unlimited number of exponential terms in each interval D_k; however, we will restrict the MTE potential to three exponential terms in each interval to increase efficiency during the inference process. Additionally, we assume one MTE potential will be defined for each configuration of a variable's discrete parents, so in each exponential term, parameters $b_{ik}^{(j)}$ are only defined for $j = 1, ..., d$ in (1). Thus, the parameters to be estimated are a_{0k}, a_{1k}, a_{2k},a_{3k}, $b_{1k}^{(j)}$,$b_{2k}^{(j)}$ and $b_{3k}^{(j)}$.

Define $\hat{\phi}^{(k)}(x; \theta_{mk})$ (abbreviated $\hat{\phi}^{(k)}(x)$) as the initial MTE approximation for PDF $f_X(x)$ in interval D_k. To estimate the parameters $\theta_{mk} = \{a_{0mk}, a_{1mk}, a_{2mk}, a_{3mk}, b_{1mk}^{(j)}, b_{2mk}^{(j)}, b_{3mk}^{(j)}\}$ in (1), the discrete approximation to KL divergence between the standard PDF and the MTE approximation is minimized subject to continuity, probability mass and non-negativity constraints for each selected paremeter vector $\mathbf{\Theta}_m$, $m = 1, ...M$.

To create the approximation to the gamma, beta and lognormal PDF's in Section 3, the following optimization problem is solved,

$$
\begin{aligned}
\underset{\theta_{mk}}{\text{argmin}} \quad & \sum_{i=1}^{n} p_{f_{Xi}} \log \frac{p_{f_{Xi}}}{q_{\hat{\phi}_{Xi}^{(k)}}} \\
\text{subject to} \quad & f_X(x_0) = \hat{\phi}^{(k)}(x_0) \\
& f_X(x_n) = \hat{\phi}^{(k)}(x_n) \\
& \int_{x_0}^{x_n} \left(f_X(x) - \hat{\phi}^{(k)}(x) \right) dx = 0 \\
& \hat{\phi}^{(k)}(x_i) \geq 0, \; i = 0, ..., n,
\end{aligned}
$$

where $p_{f_{Xi}}$ and $q_{\hat{\phi}_{Xi}^{(k)}}$ are the probability masses between x_i and x_{i-1} for $f_X(x)$ and $\hat{\phi}^{(k)}(x)$, respectively. The solution to the above optimization problem is defined as $\hat{\theta}_{mk}$. The first and second constraints ensure that the end points in adjacent regions of the MTE potential are equal. The first constraint can be relaxed in region D_1 and the second constraint can be relaxed in region D_K.

3 MTE Approximations to Standard PDF's

An MTE potential can be used to approximate any PDF. In this section, we present MTE approximations to three standard PDF's. Two MTE approximations to the normal PDF are presented in [1].

3.1 Gamma PDF

3.1.1 Function Characteristics

Suppose we have a Poisson process with constant rate λ per unit of time. Let the random variable X denote the waiting time for the rth event. The variable X has the *gamma distribution with parameters r and λ* where

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} \exp\{-\lambda x\}\ ,\ \ x > 0.$$

for any $r > 0$ and $\lambda > 0$, where $\Gamma(r)$ is the *gamma function.*

For $r > 1$, the gamma PDF has an absolute maximum where its first derivative equals zero, or $m = (r-1)/\lambda$. For $r > 2$, the gamma PDF has inflection points (changes in concavity) where its second derivative equals zero, or $x = (r-1)/\lambda \pm \sqrt{(r-1)}/\lambda$.

Define $d = \sqrt{(r-1)}/\lambda$ so that the inflection points are defined as $x = m \pm d$. The gamma PDF has two inflection points and one critical point (which is always a maximum) when $r \geq 3$. When $1 < r < 3$, the gamma PDF is a concave down function to the left of the critical point. When $r = 1$, the gamma pdf is a special case of the exponential PDF and is a monotonically decreasing, concave up function. For $0 < r \leq 1$, we approximate the gamma PDF with the exponential PDF.

3.1.2 MTE Approximation

The gamma PDF has four regions where no changes in increase/decrease or concavity/convexity occur: $(0, m-d)$, $[m-d, m)$, $[m, m+d)$, and $[m+d, \infty)$. The procedure in Section 2.4 is used to fit a 4-piece MTE approximation to the gamma PDF.

The MTE approximations to the gamma PDF with parameters $r = 6, 8$ and 11 and $\lambda = 1$ are displayed graphically in Figure 1. These MTE approximations have KL divergence statistics of 0.002095, 0.000856, and 0.000283, respectively.

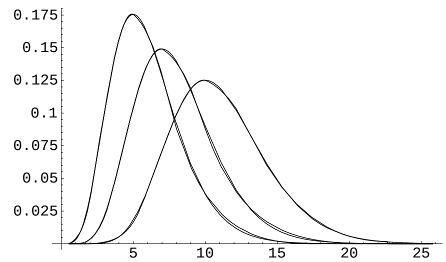

Figure 1: The MTE approximations to gamma PDF's with parameters $r = 6, 8$ and 11 and $\lambda = 1$ overlayed on the graph of the gamma PDF's.

3.2 Beta PDF

3.2.1 Function Characteristics

A distribution of a random proportion, such as the proportion of defective items in a shipment, can be represented with the beta PDF. The beta PDF for a random variable X which represents a random proportion is

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\ ,\ \ 0 < x < 1.$$

for any $\alpha > 0$ and $\beta > 0$ where $B(\alpha,\beta)$ is the *beta function.*

For most parameters α and β, the beta PDF has a critical point (either an absolute maximum or minimum) where its first derivative equals zero, or $m = (1-\alpha)/(2-\alpha-\beta)$. For some parameters α and β, the beta PDF has inflection points (changes in concavity), $d^{\pm}$, where its second derivative equals zero, or $d^{\pm} = \frac{(\alpha-1)(\alpha+\beta-3)\pm\sqrt{(\beta-1)(\alpha-1)(\alpha+\beta-3)}}{(\alpha+\beta-3)(\alpha+\beta-2)}$.

The behavior of the beta PDF is summarized in Figure 2.

3.2.2 MTE Approximation

For each of the regions defined by the parameters in Figure 2, we could define an MTE approximation, but the symmetry of the beta PDF allows us to reduce the parameter space. If $\pounds(X) \sim Beta(\alpha,\beta)$, then $f_X(x) = f_Y(1-x)$

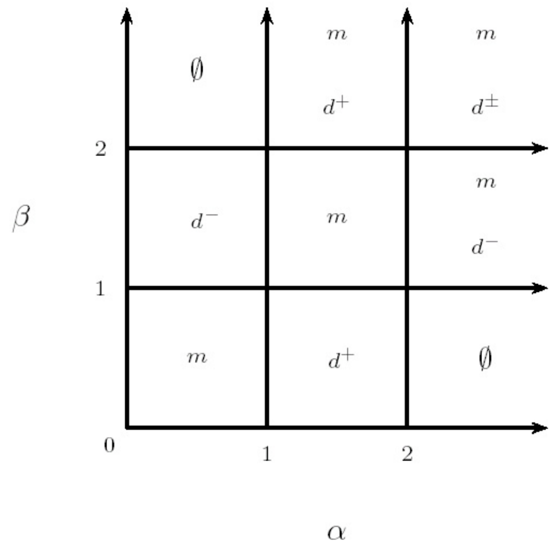

Figure 2: Critical and inflection points for the different parameters of the beta distribution.

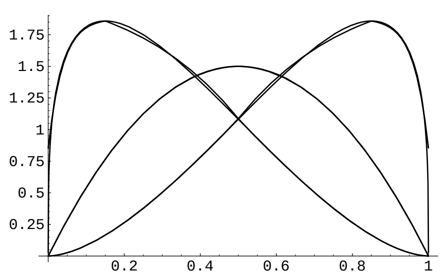

Figure 3: The MTE approximations to beta PDF's with parameters $(\alpha, \beta) = (2,2), (2.7,1.3)$ and $(1.3,2.7)$ overlayed on the graph of the beta PDF's.

where $\pounds(Y) \sim Beta(\beta, \alpha)$. This property allows us to define an MTE approximation for parameters (α, β) fulfilling the property $\alpha \geq \beta$. The MTE approximation will have a different number of pieces, depending on the critical point and the existence of inflection points.

The MTE approximations to the beta PDF with parameters $(\alpha, \beta) = (2,2), (2.7,1.3)$ and $(1.3,2.7)$ are displayed graphically in Figure 3. The MTE parameters for $Beta(1.3, 2.7)$ are obtained from $Beta(2.7, 1.3)$. The KL divergence statistics for these MTE approximations are $2.62118E-06$, 0.000330, and 0.000330, respectively.

3.3 Lognormal PDF

3.3.1 Function Characteristics

A random variable X is is lognormal, i.e. $\pounds(X) \sim LN(\mu, \sigma^2)$, if and only if $\pounds(\ln X) \sim N(\mu, \sigma^2)$. A lognormal random variable has the PDF

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \frac{-(\ln x - \mu)^2}{2\sigma^2} \ , \ \ x > 0.$$

for any $\sigma^2 > 0$.

The lognormal PDF has an absolute maximum where its first derivative equals zero, or $m = \exp\{\mu - \sigma^2\}$.

The inflection points, $d^{\pm}$, are defined where the second derivative of the lognormal PDF equals zero, or $d^{\pm} = \exp\left\{\frac{1}{2}(2\mu - 3\sigma^2 \pm \sigma\sqrt{4+\sigma^2})\right\}$.

3.3.2 MTE Approximation

To define upper and lower bounds for the MTE approximation to the lognormal PDF, we use the normal PDF as a benchmark and construct a potential containing the same probability mass in the lognormal PDF as contained in the normal PDF over the interval $[\mu - 3\sigma, \mu + 3\sigma]$. This probability mass—which equals 0.9973—is contained in the interval $[\ \exp\{\mu - 3\sigma\},\ \exp\{\mu + 3\sigma\}]$ of the lognormal PDF.

The MTE approximations to the lognormal PDF with parameters $\mu = 0$ and $\sigma^2 = 0.25, 0.50$ and 1 are displayed graphically in Figure 4. The KL divergence statistics for these MTE approximations are 0.000330, 0.000099, and 0.006467, respectively.

4 Applications

This section presents two applications of MTE potentials to inference problems in hybrid Bayesian networks. The operations required to perform inference with MTE potentials are described in detail in [1, 4].

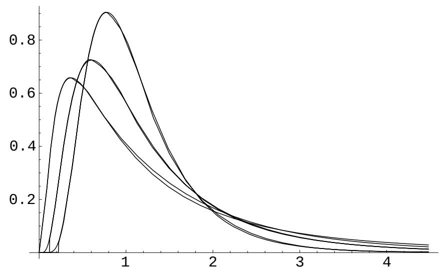

Figure 4: The MTE approximations to lognormal PDF's with parameters $\mu = 0$ and $\sigma^2 = 0.25, 0.5$ and 1.0 overlayed on the graph of the lognormal PDF's.

4.1 Bank Example

A small town has 50 residents and one bank. The number of daily arrivals to the bank (R) follows a Poisson distribution with rate $\lambda = 0.24$, i.e. $\pounds(R) \sim Poisson$ (12). Let ρ denote the discrete probability potential for R.

The service rate of customers (S) is normally distributed with a mean of 3.0 per hour and a standard deviation of 0.25, i.e. $\pounds(S) \sim N(3.0, 0.0625)$. Let φ be the potential for S, which is a 2-piece MTE approximation to the normal PDF [1]. The time to serve all customers arriving in one day (T) has a gamma distribution that is conditional on random variables R and S, i.e. $\pounds(T \mid S = s, R = r) \sim \Gamma(R, S)$, which is represented by the MTE approximation ϑ. Hire (H) is a binary, discrete random variable representing whether or not the bank manager hires an additional teller, and is modeled with a binary sigmoid function. This sigmoid function is approximated using a general MTE formulation [1].

The MTE potential fragments ψ_0 and ψ_1 constitute the MTE potential ψ for $\{H, T\}$. The MTE potential fragment for $\{H = 1, T\}$ is shown graphically in Figure 5.

Solving the problem of calculating marginal distributions for each variable in the network will require the estimate of the parameter vectors $\hat{\theta}_{mk} = \hat{\theta}_{rk}$ for $r = 2, ..., 20$, where $\hat{\theta}_{2k}$ represents parameters needed to approximate the gamma PDF with $r = 2$, etc.

The hybrid Bayesian network for the Bank example is depicted in Figure 6. A join tree

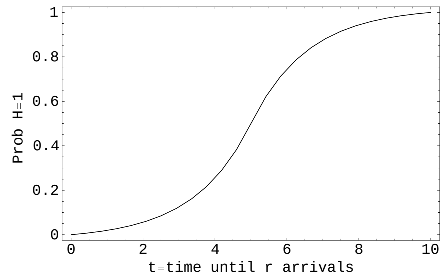

Figure 5: The MTE approximation to the sigmoid function representing $P(H = 1 \mid T = t)$ in the Bank network.

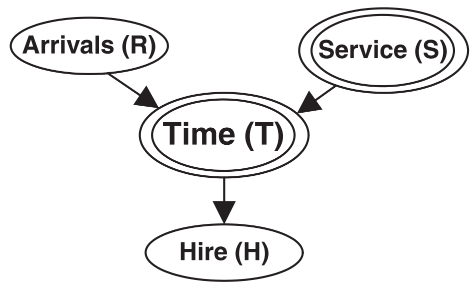

Figure 6: Hybrid Bayesian network for the Bank example.

for the Bank example is shown in Figure 7.

4.1.1 Posterior Marginals

The message sent from $\{R, S, T\}$ to $\{T\}$ is the marginal distribution for T and is calculated as follows

$$\tau(t) = \int_s \left(\varphi(s) \left(\sum_{r=0}^{20} \rho(r) \cdot \vartheta(s,t) \right) \right) \, ds.$$

The expected value and variance of the marginal distribution for T are 4.0782 and 3.2859, respectively. The posterior marginal distribution for T is shown graphically in Figure 8.

To calculate the posterior marginal probabil-

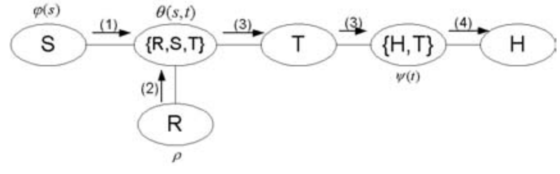

Figure 7: The join tree for the Bank example.

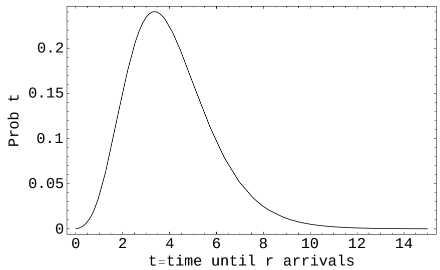

Figure 8: The posterior marginal distribution for T in the Bank example.

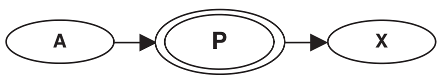

Figure 9: The hybrid Bayesian network for the Binomial example.

ities for H, the marginal distribution for T is combined with the conditional MTE potential fragments ψ_0 and ψ_1 as follows

$$\begin{aligned} \varrho_1(t) &= \psi_1(t) \cdot \tau(t), \\ \varrho_0(t) &= \psi_0(t) \cdot \tau(t). \end{aligned}$$

The marginal probability of the bank manager hiring an additional teller ($H = 1$) is 33.6%, which is found by removing T from ϱ_1.

4.2 Binomial Example

In a quality control process, a random sample of output is taken and evaluated on whether or not each unit meets a pre–determined standard. Suppose the prior distribution for the success parameter P of the binomial distribution (where $0 < p < 1$) characterizing the sample output has a beta distribution which depends on the state of the system (A) with $\Omega_A = \{0 = poor,\ 1 = average,\ 2 = good\}$. A discrete random variable X represents the number of successes in 5 trials, i.e. $\pounds(X) \sim Binomial\,(5, P)$, which is denoted as ψ. The Bayesian network for this example is shown in Figure 9.

4.2.1 MTE Potentials

Assume the following discrete distribution for A: $\varphi_0 = P(A = 0) = 0.05$, $\varphi_1 = P(A = 1) = 0.15$, $\varphi_2 = P(A = 2) = 0.80$.

Figure 10: The join tree for the Binomial example.

The potential fragment for $\{P, A = 2\}$ is an MTE approximation to the beta PDF with parameters $\alpha = 1.3$ and $\beta = 2.7$:

$$\rho_2(p, A = 2) = P(P \mid A = 2) =$$

$$\begin{cases} -5.951669 + 5.573316 \exp\{0.461388p\} \\ -0.378353 \exp\{-6.459391p\} \\ \qquad \text{if } 0 < p < 0.492929 \\ 0.473654 - 6.358483 \exp\{-2.639474p\} \\ +2.729395 \exp\{-0.331472p\} \\ \qquad \text{if } 0.492929 \le p < 0.85 \\ 1.823067 - (5.26E - 12) \exp\{26.000041p\} \\ +0.035775 \exp\{0.529991p\} \\ \qquad \text{if } 0.85 \le p < 1 \\ 0 \qquad \text{elsewhere.} \end{cases}$$

If the system is in state $A = 0$, P has a beta distribution with $\beta = 1.3$ and $\alpha = 2.7$. Due to the symmetry of the beta PDF, the potential fragment for $\{P \mid A = 0\}$ is approximated as $\rho_0(p, A = 0)$=$P(P \mid A = 0)$=$\rho_2(1-p, A = 2)$.

The potential fragment for $\{P, A = 1\}$ is an MTE approximation to the beta PDF with parameters $\alpha = 2$ and $\beta = 2$. The potential fragments constituting the potential ρ for $\{P, A\}$ are shown graphically in Figure 3 in Section 3.

The binary join tree for the Binomial example is shown in Figure 10.

4.2.2 Posterior Marginals

The marginal distribution for P and is calculated as $\varrho(p)$= $0.05 \cdot \rho_0(p, A = 0)$ $+0.15 \cdot \rho_1(p, A = 1)$+ $0.80 \cdot \rho_2(p, A = 2)$. The expected value and variance of this distribution are 0.6308 and 0.0536, respectively.

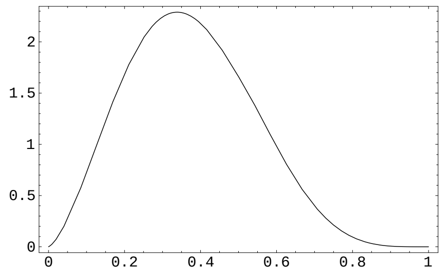

Figure 11: The revised marginal distribution for (P) incorporating the evidence $X = 1$.

Marginal probabilities for X are calculated by evaluating $((\varphi \otimes \rho)^{\downarrow P} \otimes \psi)^{\downarrow X}$. For $X = 0, \ldots, X = 5$, these probabilities are 0.0563, 0.1081, 0.1604, 0.2068, 0.2374, and 0.2310, respectively.

4.2.3 Entering Evidence

Suppose a sample of output from the system is taken and only one unit meets the quality standard. This can be expressed by evidence potential $e_X = 1$ and passed as a message in the join tree of Figure 10 from $\{X\}$ to $\{X, P\}$. New potential fragments for $\{P, A\}$ are determined as

$$\begin{aligned} \pi_0(p, A = 0) &= \psi(1, p) \cdot \rho_0(p, A = 0) \\ \pi_1(p, A = 1) &= \psi(1, p) \cdot \rho_1(p, A = 1) \\ \pi_2(p, A = 2) &= \psi(1, p) \cdot \rho_2(p, A = 2). \end{aligned}$$

The revised probabilities for A given the evidence are determined by integrating these potential fragments over P and normalizing. These probabilities are: $P(A = 0) = 0.1193$, $P(A = 1) = 0.2477$, and $P(A = 2) = 0.6330$.

The revised marginal distribution for P is shown in Figure 11 and has a mean and expected value of 0.3735 and 0.0260, respectively.

5 Summary

We have described a method of estimating the required parameters for MTE potentials which approximate standard PDF's and presented MTE approximations to three standard PDF's. Two examples of inference in hybrid Bayesian networks which use these MTE approximations were presented. Propagation in these examples is exact, so the only error in the solution is introduced in the approximation of the standard PDF's.

References

[1] Cobb, B.R. and P.P. Shenoy (2003), "Inference in hybrid Bayesian networks with mixtures of truncated exponentials," School of Business Working Paper No. 294, University of Kansas. Available for download at: www.people.ku.edu/∼brcobb/WP294.pdf

[2] Cobb, B.R., Shenoy, P.P. and R. Rumí (2004), "Approximating probability density functions with mixtures of truncated exponentials," School of Business Working Paper No. 303, University of Kansas. Available for download at: www.people.ku.edu/∼brcobb/WP303.pdf

[3] Kullback, S. and R.A. Leibler (1951), "On information and sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

[4] Moral, S., Rumí, R. and A. Salmerón (2001), "Mixtures of truncated exponentials in hybrid Bayesian networks," in P. Besnard and S. Benferhart (eds.), *Symbolic and Quantitative Approaches to Reasoning under Uncertainty*, Lecture Notes in Artificial Intelligence, 2143, 156–167.

[5] Moral, S., Rumí, R. and A. Salmerón (2003), "Approximating conditional MTE distributions by means of mixed trees," in T.D. Nielsen and N.L. Zhang (eds.), *Symbolic and Quantitative Approaches to Reasoning under Uncertainty*, Lecture Notes in Artificial Intelligence, 2711, 173–183.

[6] Poland, W.B. (1994), "Mixtures of Gaussians and minimum relative entropy techniques for modeling continuous distributions," Ph.D. Thesis, Department of Engineering–Economic Systems, Stanford University, Stanford, CA.