Generalized Evidence Pre-propagated Importance Sampling for Hybrid Bayesian Networks

Changhe Yuan
Department of Computer Science and Engineering
Mississippi State University
Mississippi State, MS 39762
cyuan@cse.msstate.edu

Marek J. Druzdzel
Decision Systems Laboratory
School of Information Sciences and
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA 15260
marek@sis.pitt.edu

Abstract

In this paper, we first provide a new theoretical understanding of the Evidence Pre-propagated Importance Sampling algorithm (EPIS-BN) (Yuan & Druzdzel 2003; 2006b) and show that its importance function minimizes the KL-divergence between the function itself and the exact posterior probability distribution in Polytrees. We then generalize the method to deal with inference in general hybrid Bayesian networks consisting of deterministic equations and arbitrary probability distributions. Using a novel technique called *soft arc reversal*, the new algorithm can also handle evidential reasoning with observed deterministic variables.

Introduction

Evidence Pre-propagated Importance Sampling algorithm (EPIS-BN) (Yuan & Druzdzel 2003; 2006b) computes an importance function by pre-propagating evidence in Bayesian networks. In this paper, we provide a new theoretical understanding of the algorithm and show that its importance function minimizes the KL-divergence between the function itself and the exact posterior probability distribution in polytrees.

We then generalize the method to deal with evidential reasoning in hybrid Bayesian networks. In order not to limit the modeling power of a Bayesian network-based tool, we should make the representation of hybrid Bayesian networks as general as possible. The representation should not only allow mixtures of discrete and continuous variables, but also allow arbitrary orderings among them, including discrete variables with continuous parents. Conditional relations can be either modeled as deterministic equations or as arbitrary probability distributions with parameters as functions of parent variables.

Excel offers complete modeling freedom and allows users to specify any interaction among its cells (these can be viewed as variables). Add-ons to spreadsheet programs, such as Excel, apply Monte Carlo sampling in solving spreadsheet models. Another group of modeling tools that offer a complete modeling freedom are visual spreadsheets, with Analytica (http://www.lumina.com/) being a prominent example. However, the main shortcoming of both spreadsheets and visual spreadsheets is that they only allow forward sampling and have no capability for evidential reasoning. Evidential reasoning is hard, mainly because the a-priori probability of observed variables is extremely low.

Since the general case is difficult, many existing approaches focus on special instances, such as *Conditional Linear Gaussians* (CLGs) (Lauritzen 1992; Lauritzen. & Jensen 2001; Gogate & Dechter 2005). CLG networks do not allow discrete variables with continuous parents, and all continuous variables must have conditional linear Gaussian distributions. To address the limitation, CLGs was later extended with softmax function to model discrete child with continuous parents (Lerner, Segal, & Koller 2001). Recent work approximates general hybrid Bayesian networks using *Mixture of Truncated Exponentials* (MTE) (Cobb & Shenoy 2005; Moral, Rumi, & Salmeron 2001) or CLGs (Shenoy 2006). Although inference in the approximated models can be done exactly, accuracy of the results is subject to the quality of the approximation, and striving for good approximation may bring numerical instability for these methods. Other approaches to inference in hybrid Bayesian networks include dynamic discretization (Kozlov & Koller 1997), *junction tree algorithm with sample potentials* (Koller, Lerner, & Angelov 1999) and generalized belief propagation algorithms for hybrid models (Heskes & Zoeter 2003; Sudderth *et al.* 2003; Yuan & Druzdzel 2006a).

Although extending EPIS-BN to hybrid models seems natural, two major difficulties need to be overcome in order to make the generalization work. First, although algorithms exist for performing belief propagation in hybrid models (Heskes & Zoeter 2003; Sudderth *et al.* 2003; Yuan & Druzdzel 2006a), it is not straightforward how to utilize them to compute an importance function for sampling. Second, it is extremely difficult to generate a valid sample by forward sampling methods when there are observed deterministic variables. To tackle the difficulties, we propose two techniques: *online importance function generation* and *soft arc reversal*. The resulting algorithm gives full modeling freedom, because it is general enough to deal with evidential reasoning in hybrid Bayesian networks that contain linear or nonlinear deterministic equations and arbitrary probability distributions, and it also guarantees to converge to correct posterior probability distributions.

Evidence Pre-propagated Importance Sampling

Importance sampling-based algorithms, such as AIS-BN (Cheng & Druzdzel 2000), *Dynamic* IS (Moral & Salmeron 2003), and EPIS-BN (Yuan & Druzdzel 2003; 2006b), have been proven effective in solving discrete Bayesian networks. These algorithms essentially differ only in the methods that they use to compute importance functions. The role taken by the importance function in importance sampling is similar to that of a heuristic function in search algorithms; the more precise the importance function is, the better importance sampling performs. In this section, we develop a new theoretical understanding for the EPIS-BN algorithm (Yuan & Druzdzel 2003; 2006b).

The EPIS-BN Algorithm: A Review

Suppose $P(\mathbf{X})$ models the joint probability distribution over a set of variables $\{X_1, X_2, ..., X_n\}$ in a Bayesian network. Let the evidence set be $\mathbf{E}$. The importance function of EPIS-BN is defined as:

$$\rho(\mathbf{X}\backslash\mathbf{E}) = \prod_{i=1}^{n} P(X_i|\text{PA}(X_i), \mathbf{E}) \; , \tag{1}$$

where each $P(X_i|\text{PA}(X_i), \mathbf{E})$ is an *importance conditional probability table* (ICPT) (Cheng & Druzdzel 2000). The following theorem shows that we can calculate the ICPTs in polytrees exactly (Yuan & Druzdzel 2003; 2006b).

Theorem 1 *Let X_i be a variable in a polytree, and $\mathbf{E}$ be the set of evidence variables. The exact* ICPT *$P(X_i|PA(X_i), \mathbf{E})$ for X_i is*

$$\alpha(PA(X_i))P(X_i|PA(X_i))\lambda(X_i) \; , \tag{2}$$

where $\alpha(PA(X_i))$ is a normalizing constant dependent on $PA(X_i)$, and $\lambda(X_i)$ is the message to X_i sent from its descendants, as defined in (Pearl 1988).

Based on the observation that LBP provides surprisingly good results for many networks with loops (Murphy, Weiss, & Jordan 1999), Yuan and Druzdzel (2003; 2006b) propose to use a small number of LBP steps to estimate the importance functions for loopy networks in the EPIS-BN algorithm. Experimental results show that subsequent importance sampling converges to the posterior probability distribution, even in those cases in which LBP does not converge to the right posterior. The resulting algorithm, EPIS-BN, achieves a considerable improvement over the previous state-of-the-art algorithm, AIS-BN (Cheng & Druzdzel 2000).

A New Theoretical Justification

We now develop a new theoretical understanding of the success of the EPIS-BN algorithm. Given a particular representation of importance function, we want to parameterize it such that it is as close as possible to the posterior probability distribution. The posterior distribution of the network hence has the following form: $P(X_1, ..., X_n|\mathbf{E})$. Let $Q(X_1, ..., X_n)$ be the importance function. It has been shown that when diagnostic evidence is introduced, the optimal importance function no longer factorizes in the same way as the original network (Yuan & Druzdzel 2005). Nevertheless, most existing importance sampling algorithms still use the prior network structure as the representation for their importance function. We also use the representation in this work, i.e.,

$$Q(\mathbf{X}) = \prod_{i=1}^{n} Q(X_i|\text{PA}(X_i)) \; . \tag{3}$$

First, we have the following theorem.

Theorem 2 *The importance function that takes the original network structure and minimizes the KL-divergence from the posterior probability distribution has the following form*

$$Q(\mathbf{X}) \propto \prod_{i=1}^{n} P(X_i, PA(X_i)|\mathbf{E}) \; . \tag{4}$$

Proof: *The KL-divergence between the importance function and the posterior probability distribution is*

$$\begin{aligned} &KL(P(\mathbf{X}|\mathbf{E}), Q(\mathbf{X})) \\ &= \sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{E}) \ln \frac{P(\mathbf{X}|\mathbf{E})}{Q(\mathbf{X})} \\ &= \sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{E}) \ln P(\mathbf{X}|\mathbf{E}) - \sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{E}) \ln Q(\mathbf{X}) \\ &= -H(P) + H(P, Q), \end{aligned}$$

where

$$\begin{aligned} &H(P, Q) \\ &= -\sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{E}) \ln \prod_{i=1}^{n} Q(X_i|PA(X_i)) \\ &= -\sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{E}) \sum_{i=1}^{n} \ln Q(X_i|PA(X_i)) \\ &= -\sum_{i=1}^{n} \sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{E}) \ln Q(X_i|PA(X_i)) \\ &= -\sum_{i=1}^{n} \sum_{X_i, PA(X_i)} P(X_i, PA(X_i)|\mathbf{E}) \ln Q(X_i|PA(X_i)). \end{aligned}$$

Therefore, the minimum is achieved when

$$\forall i \; Q(X_i|PA(X_i)) \propto P(X_i, PA(X_i)|\mathbf{E}) \; . \tag{5}$$

□

Given the above theorem, we immediately have the following corollary for polytrees.

Corollary 1 *In a polytree, the importance function that takes the original network structure and minimizes the KL-divergence from the posterior probability distribution has the following form*

$$Q(\mathbf{X}) \propto \prod_{i=1}^{n} \lambda(X_i)P(X_i|PA(X_i)) \prod_{U_j \in PA(X_i)} \pi_{X_i}(U_j) \; , \tag{6}$$

where $\lambda(X_i)$ is the message sent from children of X_i and $\pi_{X_i}(U_j)$ is the message sent from parent U_j, as defined in (Pearl 1988).

Proof: *In a polytree, the evidence set can be divided into two sets, i.e., $\mathbf{E} = \mathbf{E}^+_{X_i} \cup \mathbf{E}^-_{X_i}$, where $\mathbf{E}^+_{X_i}$ is the evidence set connected to parents of X_i, and $\mathbf{E}^-_{X_i}$ is the evidence set connected to children of X_i. Therefore,*

$$
\begin{aligned}
& Q(X_i|PA(X_i)) \\
& \propto P(X_i, PA(X_i)|\mathbf{E}) \\
& \propto \frac{P(X_i, PA(X_i), \mathbf{E}^+_{X_i}, \mathbf{E}^-_{X_i})}{P(\mathbf{E})} \\
& \propto P(\mathbf{E}^-_{X_i}|X_i)P(X_i|PA(X_i))P(PA(X_i)|\mathbf{E}^+_{X_i}) \\
& \propto P(\mathbf{E}^-_{X_i}|X_i)P(X_i|PA(X_i)) \prod_{U_i \in PA(X_i)} P(U_i|\mathbf{E}^+_{U_i}). \\
& \propto \lambda(X_i)P(X_i|PA(X_i)) \prod_{U_i \in PA(X_i)} \pi_{X_i}(U_j) \ .
\end{aligned}
$$

□

Although the above result is apparently different from Theorem 1, note that if we sample a Bayesian network in its topological order, the parents of a variable X_i are already instantiated before we sample X_i. Hence, the π messages from its parents are all constants. Therefore, Corollary 1 reduces to Theorem 1. Another important implication of this theoretical result is that belief propagation not only calculates the exact marginal posterior probability distributions, but also calculates a joint posterior probability distribution over unobserved variables that minimizes the KL-divergence from the exact posterior probability distribution.

Importance Sampling in Hybrid Bayesian Networks

In this section, we generalize the EPIS-BN algorithm to deal with evidential reasoning in general hybrid Bayesian networks.

The Approach

Monte Carlo sampling methods put minimal restriction on the representation of the models as long as we know how to sample from the probability distributions being used. This makes sampling a natural choice for inference in general hybrid Bayesian networks. However, not much work has been done in this direction due to two major difficulties. First, obtaining a good importance function is difficult in such general models. Second, no methods exist for evidential reasoning with observed deterministic variables. Such case is difficult because forward sampling methods will have difficulty in obtaining valid samples; almost all the samples obtained will have zero probabilities.

We dedicate the next two sections to addressing the two difficulties. In the next section, we discuss how to use the *Hybrid Loopy Belief Propagation* (HLBP) algorithm (Yuan & Druzdzel 2006a) to compute an importance function and generalize EPIS-BN to general hybrid Bayesian networks. In the section after that, we propose a technique called *soft arc reversal* to deal with evidential reasoning when observed deterministic variables are present.

Online Importance Function Generation

We first discuss how to apply HLBP to calculate importance function in hybrid Bayesian networks. The main idea of HLBP is that, since no closed-form solutions exist for the LBP messages in general hybrid models, HLBP formulates the message calculation as Monte Carlo integration problem and approximates the messages using mixture of Gaussians (MoGs) (Titterington, Smith, & Makov 1985).

Just as in EPIS-BN, we are interested in making use of the λ messages estimated by HLBP to calculate the ICPTs. In discrete Bayesian networks, we are able to multiply the λ messages with CPTs to obtain the ICPTs before sampling. It is much more complex in hybrid Bayesian networks, and we have to distinguish the following several scenarios. First, the λ messages are useless in the following case.

Deterministic continuous variable with arbitrary parents: As soon as the parents of a deterministic node are instantiated, we simply evaluate the deterministic relation to get its value. Its λ messages can be simply discarded.

In the following two cases, a node has all discrete parents. ICPTs can also be computed before the sampling step.

Discrete variable with only discrete parents: This scenario reduces to the same as in discrete Bayesian networks.

Stochastic continuous variable with only discrete parents: All the conditional relations in this case are continuous probability distributions with fixed parameters. We can approximate the conditional distributions with MoGs (Titterington, Smith, & Makov 1985) and multiply with the λ message to produce new MoGs as the ICPTs using either Gibbs sampling (Sudderth *et al.* 2003) or importance sampling (Yuan & Druzdzel 2006a).

Finally, if a node has continuous parents, its conditional relations may depend on the values of these parents, such as a softmax function (Lerner, Segal, & Koller 2001) or a probability distribution with parameters as functions of the parents. Pre-computing ICPT is typically infeasible. We propose a technique called *online importance function generation* to deal with the problem. The main idea is to delay combining the λ message with the conditional relation until the parents are instantiated. The technique is applied during sampling and adopted in the following two cases.

Discrete variable with continuous parents: Once parents of a discrete variable are instantiated, the conditional relation becomes a discrete distribution which can be multiplied by the discrete λ message to produce the importance function.

Stochastic continuous variable with continuous parents: Once the parents are instantiated, the parameters of the conditional distributions become constants. We can approximate the distributions by MoGs and multiply with the λ message to get the importance function.

Evidential Reasoning with Deterministic Variables

Unlikely evidence is known to pose problems for importance sampling. It is even more the case in hybrid Bayesian networks. In particular, observed deterministic variables make sampling difficult. Note that importance sampling algorithms are typically forward sampling methods, i.e., they sample a network in its topological order. However, samples thus obtained may conflict with evidence observed on a deterministic node and hence have zero probabilities. Cobb and Shenoy (2005) approximate nonlinear deterministic equations using piecewise linear equations; their method is not applicable here because it may produce invalid samples. To address the problem, we propose a technique named *soft arc reversal*: we draw samples for all the parents except one and solve the remaining parent using the deterministic equation. Of course we have to assign appropriate weights to samples in order to correct bias thus introduced. This technique is similar to performing a physical arc reversal on the network.

$P(C\|A,B)$
$C = A + 2 * B$

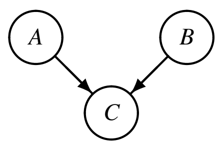


Figure 1: A simple hybrid Bayesian network.

We explain soft arc reversal using a concrete example. Suppose we have a small hybrid Bayesian network with three continuous variables A, B, and C as in Figure 1. C is a deterministic node dependent on A and B. We also assume that C is observed to be 4.0. A plain forward sampling algorithm would sample A and B first and accept the evidence on C as the sample. However, such sample would almost surely have zero probability, because the chance of sampling a value pair for A and B that satisfy the equation $P(C|A, B)$ is almost zero. Note that once we get the value for either A or B, the value of the other variable is already determined. For example, suppose we get value 2.0 for A. Then B can only take value 1.0. Therefore, instead of sampling B, we should solve for it and take 1.0 as the sample.

This problem is not really unique to hybrid models. Suppose there is a discrete Bayesian network with the same structure as in Figure 1, and the CPT of node C models the relation $C = \text{OR}(A, B)$ among three binary variables. Let C be observed to be 0. If we sample A and B first, it is also hard to hit a valid sample, because only one scenario $(A = 0, B = 0)$ is compatible with the evidence, and many samples will be wasted. However, since there are only four scenarios in total, we can still hit valid samples. The issue was not addressed in any previous algorithms to our best knowledge.

The reversal parents are selected in a preprocessing step. We prefer to choose stochastic nodes as reversal parents, because once a stochastic node is solved, we can update the weight of the current sample using the conditional probability distribution of the node. However, if all parents are deterministic, we have no choice but to pick a deterministic parent to reverse; the choice can be arbitrary in such case. Then, we need to apply the selecting procedure recursively to the reversal parent until we find a stochastic node or a root node.

We need an equation solver in order to solve deterministic equations. In our implementation, we use the *Newton's method for solving nonlinear set of equations* (Kelley 2003). Not all equations are solvable by this equation solver or any equation solver for that matter. It is desirable to choose a parent that is easier to solve, which can be tested using a preprocessing step. In more difficult cases, we have to resort to modeler's help and ask for specifying which parent to solve or even specify the inverse functions manually. When there are multiple choices, one helpful heuristic is to choose the parent with the largest variance.

The HEPIS-BN Algorithm

Now, since we know how to calculate all the ICPTs and how to deal with observed deterministic variables, we are ready to present the *hybrid evidence pre-propagated importance sampling* algorithm (HEPIS-BN). The complete algorithm is outlined in Figure 2. It essentially boils down to drawing a single importance sample, as in Steps 5 and 6, from the importance function and repeating the process until we get enough samples.

Experimental Results

We tested our algorithm using two benchmark models and a modified benchmark model. We compute the exact solutions by a massive amount of computation (likelihood weighting (LW) (Fung & Chang 1989; Shachter & Peot 1989) with 100M samples). To evaluate how well HEPIS-BN performs, we discretized the ranges of continuous variables to 50 intervals and then calculated the Hellinger's distance between the results of HEPIS-BN and the exact solutions. We use likelihood weighting as our baseline because of its applicability to general hybrid models. We tried comparing our algorithm against the Gibbs sampling implemented in BUGS but encountered convergence difficulties similar to those reported in (Lerner, Segal, & Koller 2001). All results are the average of 50 runs of the experiments.

Results on the Emission Networks

In the first experiment, we tested the HEPIS-BN algorithm on two benchmark models, emission network (Lauritzen 1992) and augmented emission network (Lerner, Segal, & Koller 2001).

Several existing approaches only produce the first two moments as the solution. However, we would like to note that mean and variance alone provide only limited information about the actual posterior probability distributions. Figure 4 shows the posterior probability distribution of node DustEmission when observing CO2Emission, Penetrability and WasteType to be $-1.6, 0.5, and 1$ respectively. We also plot in the same figure the corresponding normal approximation with mean 3.77 and variance 1.74. We see that the normal approximation does not reflect the true posterior. While

Algorithm: HEPIS-BN
Input: Hybrid Bayesian network B, a set of evidence variables **E**, and a set of non-evidence variables **X**;
Output: The marginal distributions of non-evidence variables.

1. Order the nodes in the topological order of the network.
2. Select reversal ancestors for evidence nodes.
3. Perform hybrid loopy belief propagation (HLBP) for several steps.
4. Precompute the ICPTs for all the nodes with only discrete parents using messages by HLBP.
5. Initialize sample weight w to be 1.0.
6. **for** each node X in the ordering
    - **if** X is a reversal parent
        - $\backslash *$ Do nothing! $* \backslash$
    - **else if** X is stochastic with no evidence
        - Calculate ICPT $I(x|u)$ if not already computed and draw a sample for x.
        - Set $w = w * P(x|u)/I(x|u)$.
    - **else if** X is deterministic with no evidence
        - Evaluate $P(x|u)$ and keep the weight unchanged.
    - **else if** X is stochastic with evidence e
        - Take e as the sample and set $w = w * P(e|u)$.
    - **else** $\backslash *$ X must be deterministic with evidence $* \backslash$
        - Use soft arc reversal to solve reversal parents.
        - Assign appropriate weight for the current sample.
    - **end if**

    end for
7. Repeat steps 5–6 and draw a desirable number of weighted samples.
8. Estimate the posterior marginal probability distributions for all unobserved variables.

Figure 2: The HEPIS-BN algorithm.

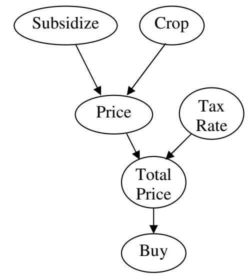


Figure 3: Augmented crop network.

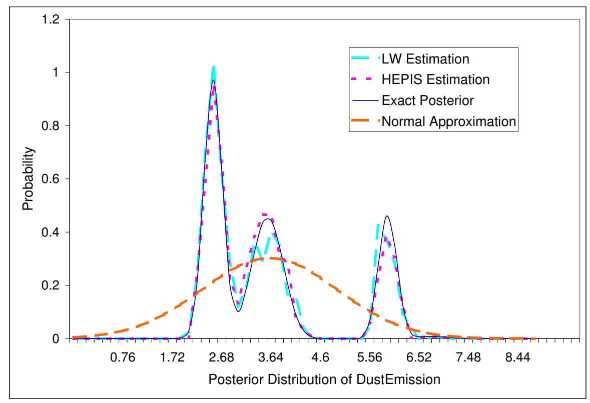


Figure 4: The posterior probability distributions estimated by LW, HEPIS-BN, normal approximation of DustEmission when observing CO2Emission, Penetrability and WasteType to be $-1.6, 0.5, 1$ respectively in emission network.

the actual posterior distribution has a multimodal shape, the normal approximation does not tell us where the mass really is. On the contrary, LW and HEPIS-BN with 4K samples were both able to correctly estimate the shape of the posterior probability distribution, with HEPIS-BN demonstrating better performance than LW.

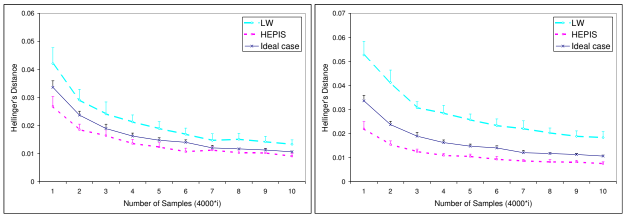
We compared the convergence rates and plotted the error curves of both algorithms in Figure 5(a). We also plotted the results of LW on emission network with no evidence (the ideal case for importance sampling). Although results thus obtained are not strict lower bounds, they can serve as an indication of the limiting precision of importance sampling on the network. We observe that HEPIS-BN had a precision even slightly better than the ideal case, maybe due to fewer nodes to estimate. However, HEPIS-BN requires more time to generate a sample than LW, roughly twice as much as LW on the emission network.

We also used a more unlikely evidence with CO2Emission, Penetrability and WasteType being $-0.1, 0.5, and 1$ respectively to test the robustness of LW and HEPIS-BN. It is more unlikely because when we set Penetrability to 0.5 and WasteType to 1, the posterior probability distribution of CO2Emission has a mean of -1.55. The error curves are shown in Figure 5(b). We can see that LW clearly performed worse, but HEPIS-BN seemed quite robust to low likelihood of the evidence. To develop a better understanding, we gradually changed the observed value of CO2Emission to make the evidence less likely and ran both algorithms on these cases. The results are plotted in Figure 6. We observe that LW kept deteriorating in face of unlikely evidence, while HEPIS-BN's performance was fairly stable.

We also plotted the convergence results on the augmented emission (Lerner, Segal, & Koller 2001) network with evidence of CO2Sensor and DustSensor being true in Figure 7(a). We used the same evidence as in (Lerner, Segal, & Koller 2001). Again, we observe that HEPIS-BN achieves a precision better than LW and comparable to the ideal case.

Evidential Reasoning with Deterministic Variables

We propose soft arc reversal to deal with observed deterministic nodes. To verify that the proposed technique works



(a) **(b)**

Figure 5: Convergence curve comparison of LW and HEPIS-BN on emission network with evidence (a) CO2Emission $= -1.6$, Penetrability $= 0.5$ and WasteType $= 1$ (b) CO2Emission $= -0.1$, Penetrability $= 0.5$ and WasteType $= 1$, together with Ideal case (LW on emission with no evidence).

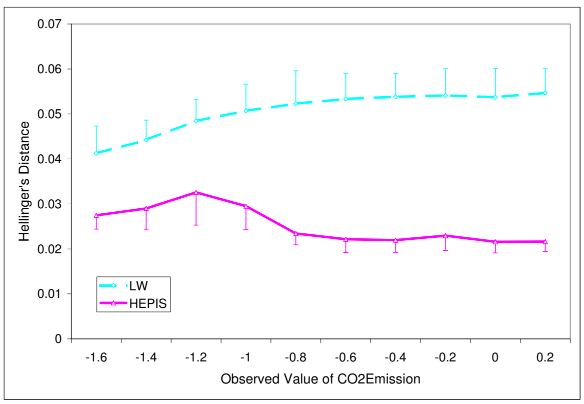
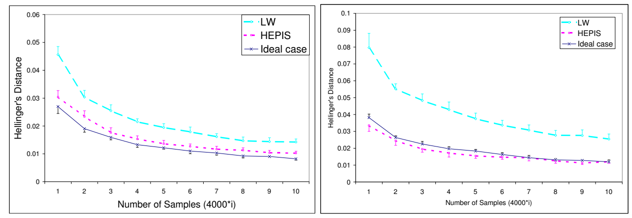


Figure 6: The influence of the observed value of CO2Emission on the precision of LW and HEPIS-BN after observing Penetrability to 0.5 and WasteType to 1 in emission network.

properly, we revised the crop network (Lerner, Segal, & Koller 2001) by adding a nonlinear deterministic node TotalPrice ($TotalPrice = Price * (1 + TaxRate)$) (as in Figure 3) and observing the value of TotalPrice to be 18.0. Although the model is quite simple, most existing approaches cannot handle it. The classic LW does not work in such case either. We had to enhance it for the purpose of our experiments with the *soft arc reversal* technique. We ran the same experiment as in the previous section, and plotted out the error curves in Figure 7(b). We again observed that HEPIS-BN performed better than LW on this network and its precision was comparable to the ideal case.

Results on Larger Networks

Given the stability demonstrated by the HEPIS-BN algorithm, we anticipate that the advantage of HEPIS-BN will be more evident in larger models. To test the hypothesis, we transformed the emission network into a dynamic model with three slices and a total of 27 variables, and observed CO2Emission at -0.1 and Penetrability at 0.5 in the first and third slice of the model. The results of the algorithms are shown in Figure 8. While LW showed a much worse



(a) **(b)**

Figure 7: Error curves LW and HEPIS-BN (a) augmented emission network when CO2Sensor and DustSensor are observed to be true (b) crop network when totalprice is observed to be 18.0.

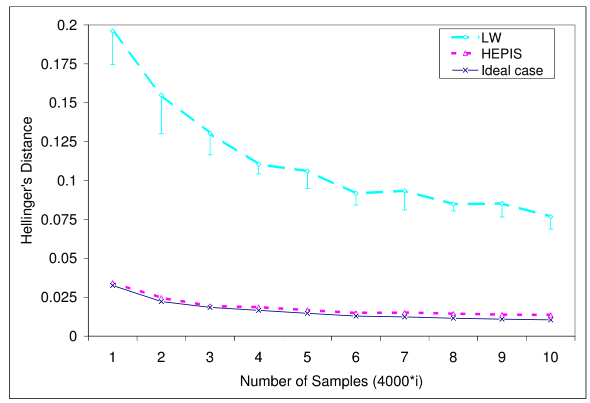


Figure 8: Error curves of LW and HEPIS-BN on the dynamic emission network.

precision, the precision of HEPIS-BN remained close the that of the ideal case. Therefore, although HEPIS-BN typically requires more running time per sample, it is expected to outperform LW on complex models.

Conclusion

In this paper, we provide a new theoretical understanding of the EPIS-BN algorithm and show that its importance function is motivated by minimizing the KL-divergence from posterior probability distribution. We then propose the HEPIS-BN algorithm that deals with evidential reasoning in general hybrid Bayesian networks using two novel techniques: online importance function generation and soft arc reversal. We tested the algorithm on several benchmark hybrid models, and observed that HEPIS-BN not only yields a much better accuracy than LW given the same number of samples, but also is more stable in face of unlikely evidence. More importantly, its performance is consistently close to the ideal case with no evidence. The new algorithm provides full modeling freedom and is applicable to a broad class of problems. HEPIS-BN is slower than LW due to its complexity, but its efficiency can be improved using more efficient density estimation techniques. We believe that the advantage of HEPIS-BN will become more evident in large models with unlikely observations.

Acknowledgements

This research was partially done when the first author was at the University of Pittsburgh. We would like to acknowledge generous support for our work from the Air Force Office of Scientific Research grant FA9550–06–1–0243 and from Intel Research. All experimental data have been obtained using SMILE, a Bayesian inference engine developed at the Decision Systems Laboratory and available at http://genie.sis.pitt.edu.

References

Cheng, J., and Druzdzel, M. J. 2000. BN-AIS: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. *Journal of Artificial Intelligence Research* 13:155–188.

Cobb, B. R., and Shenoy, P. P. 2005. Nonlinear deterministic relationships in Bayesian networks. In *Proceedings of Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-05)*, 27–38. Berlin: Springer-Verlag.

Fung, R., and Chang, K.-C. 1989. Weighing and integrating evidence for stochastic simulation in Bayesian networks. In Henrion, M.; Shachter, R.; Kanal, L.; and Lemmer, J., eds., *Uncertainty in Artificial Intelligence 5*, 209–219. New York, N. Y.: Elsevier Science Publishing Company, Inc.

Gogate, V., and Dechter, R. 2005. Approximate inference algorithms for hybrid Bayesian networks with discrete constraints. In *Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI–05)*.

Heskes, T., and Zoeter, O. 2003. Generalized belief propagation for approximate inference in hybrid Bayesian networks. In *Proceedings of AI&STAT-03*.

Kelley, C. T. 2003. *Solving Nonlinear Equations with Newton's Method*. Society for Industrial and Applied Mathematic.

Koller, D.; Lerner, U.; and Angelov, D. 1999. A general algorithm for approximate inference and its application to hybrid Bayes nets. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, 324–333. Morgan Kaufmann Publishers Inc.

Kozlov, A., and Koller, D. 1997. Nonuniform dynamic discretization in hybrid networks. In *Proceedings of the 13th Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, 314–325. San Francisco, CA: Morgan Kaufmann.

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

Lauritzen, S. 1992. Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association* 87:1098–1108.

Lerner, U.; Segal, E.; and Koller, D. 2001. Exact inference in networks with discrete children of continuous parents. In *Proceedings of the Seventeenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)*, 319–328. Morgan Kaufmann Publishers Inc.

Moral, S., and Salmeron, A. 2003. Dynamic importance sampling computation in Bayesian networks. In *Proceedings of Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-03)*, 137–148.

Moral, S.; Rumi, R.; and Salmeron, A. 2001. Mixtures of truncated exponentials in hybrid Bayesian networks. In *Proceedings of Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-01)*, 156–167.

Murphy, K.; Weiss, Y.; and Jordan, M. 1999. Loopy belief propagation for approximate inference: An empirical study. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–99)*, 467–475. San Francisco, CA: Morgan Kaufmann Publishers.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Shachter, R. D., and Peot, M. A. 1989. Simulation approaches to general probabilistic inference on belief networks. In Henrion, M.; Shachter, R.; Kanal, L.; and Lemmer, J., eds., *Uncertainty in Artificial Intelligence 5*, 221–231. New York, N. Y.: Elsevier Science Publishing Company, Inc.

Shenoy, P. 2006. Inference in hybrid Bayesian networks using mixtures of gaussians. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, 428–436. Corvallis, OR: AUAI Press.

Sudderth, E.; Ihler, A.; Freeman, W.; and Willsky, A. 2003. Nonparametric belief propagation. In *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR-03)*.

Titterington, D.; Smith, A.; and Makov, U. 1985. *Statistical Analysis of Finite Mixture Distribution*. NY: John Wiley and Sons.

Yuan, C., and Druzdzel, M. J. 2003. An importance sampling algorithm based on evidence pre-propagation. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI–03)*, 624–631.

Yuan, C., and Druzdzel, M. J. 2005. Importance sampling in Bayesian networks: An influence-based approximation strategy for importance functions. In *Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI–05)*, 650–657.

Yuan, C., and Druzdzel, M. J. 2006a. Hybrid loopy belief propagation. In *Proceedings of the Third European Workshop on Probabilistic Graphical Models (PGM'06)*, 317–324.

Yuan, C., and Druzdzel, M. J. 2006b. Importance sampling algorithms for Bayesian networks: Principles and performance. *Mathematical and Computer Modelling* 43:1189–1207.