Tractable Inference for Hybrid Bayesian Networks with NAT-Modeled Dynamic Discretization

Yang Xiang, Hanwen Zheng
School of Computer Science, University of Guelph, Canada

Abstract

Hybrid BNs (HBNs) extend Bayesian networks (BNs) to both discrete and continuous variables. Among inference methods for HBNs, we focus on dynamic discretization (DD) that converts HBN to discrete BN for inference. Complexity of BN inference is exponential on treewidth, which extends to DD for HBNs. We presents a novel framework where HBN is transformed into NAT-modeled BN (NAT: Non-impeding noisy-AND Tree) for tractable inference. A case-study under the framework is presented on sum of continuous variables. We report significant efficiency gain of approximate inference by NAT-modeled DD over alternative methods.

1 Introduction

BNs consist of discrete variables and are extended into HBNs to allow continuous variables. Several methods exist for inference with HBN (Lauritzen and Jensen 2001; Moral, Rumi, and Salmeron 2001; Shenoy and West 2011). We focus on DD (Neil, Tailor, and Marquez 2007), which converts HBNs into BNs to enable BN techniques for HBN inference without limitation of static discretization. BN inference complexity is exponential on treewidth, which extends to DD for HBNs. Our contribution includes a framework for tractable HBN inference by extending DD with local models, and identifying and solving key technical issues.

To validate and demonstrate this novel framework, we present a case-study on HBNs that involve sum of continuous variables (which we will refer to as sum of reals). That is, we have a set of continuous variables of various prior distributions. We either infer about their sum distribution or infer about their posterior distributions upon observation of their sum (and possibly some addends). Our case-study focuses on sum of reals since no tractable method is known to the best of our knowledge. “Calculating such a distribution represents a major challenge for most BN software” (Fenton and Neil 2012), due to the large treewidth involved.

In the case-study, we extend DD by NAT modeling (Xiang 2012) to convert HBN into NAT-modeled BN for tractable DD inference. NAT models are among several local models that encode BN CPTs (Conditional Probability Tables) efficiently (Pearl 1988; Henrion 1989; Diez 1993; Boutilier et al. 1996; Savicky and Vomlel 2007; Maaskant and Druzdzel 2008; Woudenberg, van der Gaag, and Rademaker 2015). Merits of NAT models include simple causal interactions (reinforcing & undermining), expressiveness (recursive mixture of interactions, multi-valued, ordinal or nominal), generality (generalizing noisy-OR, noisy-MAX, and DeMorgan), and being orthogonal to context-specific independence. Although efficient tensor decomposition for sum of integers exists (Savicky and Vomlel 2007), whether it supports tractable inference with sum of reals remains a research issue. We resolve key issues under the framework for NAT-modeled DD, and report tractable approximate inference with sum of reals.

In the remainder, Section 2 reviews background on DD, inference with sum of reals, and NAT modeling. Sections 4 analyzes limitations of alternative methods. Our case-study on NAT-modeled DD for sum inference is covered in Sections 5 to 7. Experimental evaluation is reported in Section 9.

2 Background

[**Dynamic discretization**] Inference for HBNs can be performed by converting HBN into BN using static discretization, where domain of each (continuous) variable is discretized into bins. However, the inference is inaccurate if too few bins per variable are used, and inference cost grows quickly if many bins are used. DD (Fenton and Neil 2012) overcomes the limitation by revising bins dynamically.

DD inference goes in multi-rounds, starting with a static discretization and resultant BN. After each round, approximate Kullback-Leibler (KL) distance/error between (unknown) true PDF (probability density function) and the discretized PDF is evaluated for each bin of each variable. Each bin with a large error is split into two, and adjacent bins of little probability mass are merged. The resultant BN is under a new static discretization, where bins are refined as needed, and coarsened when justified to keep the number of bins low. The new BN is used in the next round of inference.

Since a BN is used for inference in each round, and the complexity is exponential on its treewidth, the exponential complexity extends to DD inference for HBNs.

[**Inference with sum of continuous variables**] Let $u_1, ..., u_n$ be independent continuous variables (addends), and $w = \sum_{k=1}^{n} u_k$ be their sum. Let $u'_1, ..., u'_n, w'$ denote discrete variables from discretizing $u_1, ..., u_n, w$. We denote PDFs of u_k and w by $p_{u_k}()$ and $p_w()$. An HBN (segment) on $u_1, ..., u_n$ and w consists of a structure where w is the child with parents $u_1, ..., u_n$. Inference over addends and sum may compute (prior) PDF of $p_w(w)$ given $p_{u_k}(u_k)$. Inference may also compute posterior PDFs of addends, given prior PDF for each u_k, and observed values of w and some addends. PDF of $w = u + v$ can be obtained by convolution (Grinstead and Snell 2003),

$$p_w(w) = \int_{-\infty}^{\infty} p_u(w - v) p_v(v) dv, \tag{1}$$

and PDF of $w = \sum_{k=1}^{n} u_k$ can be obtained by $n-1$ pairwise convolutions. Sum PDF where addends are uniformly distributed, denoted by $u_k \sim U(a,b)$, has been studied (Killmann and von Collani 2001; Kang et al. 2010). If each $u_k \sim U(0,1)$, PDF of w is Irwin-Hall distribution (Irwin 1927; Hall 1927), with cumulative distribution function (CDF)

$$F(w;n) = \frac{1}{n!}\sum_{k=0}^{\lfloor w \rfloor}(-1)^k C(n,k)(w-k)^n. \tag{2}$$

The above methods can compute prior PDF of sum, but are not directly applicable to posterior on addends and sum. For general inference (prior and posterior) with DD, an approximate method (referred to below as *uniform mixture*) computes $P(w'|u'_1, ..., u'_n)$ based on mixture of uniform distributions (Fenton and Neil 2012).

[**NAT models**] A NAT model (Xiang 2012; Xiang and Jiang 2018) is over an effect e and a set of n causes $C = \{c_1, ..., c_n\}$, where $e \in D_e = \{e^0, ..., e^\eta\}$ $(\eta \geq 1)$ and $c_i \in \{c_i^0, ..., c_i^{m_i}\}$ $(i = 1, ..., n, m_i \geq 1)$. C and e form a family (a child and its parents) in BN, whose dependence is quantified by a CPT by default. Values e^0 and c_i^0 are *inactive*. Other values (may be written as e^+ or c_i^+) are *active*.

A causal event is a *success* or *failure* depending on if e is active up to a given value, is *single-* or *multi-causal* depending on the number of active causes, and is *simple* or *congregate* depending on value range of e. For instance, $P(e^k \leftarrow c_i^j) = P(e^k|c_i^j, c_z^0 : \forall z \neq i)$ $(j > 0)$ is probability of a *simple single-causal success*, and

$$P(e \geq e^k \leftarrow c_1^{j_1}, ..., c_q^{j_q}) = P(e \geq e^k|c_1^{j_1}, ..., c_q^{j_q}, c_z^0 : c_z \in C \setminus X)$$

is probability of a *congregate multi-causal success*, where $j_1, ..., j_q > 0$, $X = \{c_1, ..., c_q\}$ $(q > 1)$. The latter may be denoted as $P(e \geq e^k \leftarrow \underline{x}^+)$. Interactions among causes may be reinforcing or undermining as defined below.

Definition 1 *Let e^k be an active effect value, $R = \{W_1, ..., W_m\}$ $(m \geq 2)$ be a given partition of a set $X \subseteq C$ of causes, $S \subset R$, and $Y = \cup_{W_i \in S} W_i$. Sets of causes in R reinforce each other relative to e^k, iff $\forall S$ $P(e \geq e^k \leftarrow \underline{y}^+) \leq P(e \geq e^k \leftarrow \underline{x}^+)$. They undermine each other iff $\forall S$ $P(e \geq e^k \leftarrow \underline{y}^+) > P(e \geq e^k \leftarrow \underline{x}^+)$.*

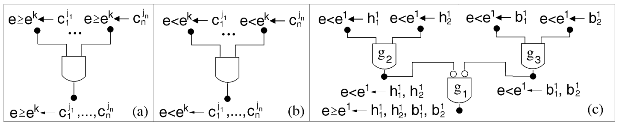

Figure 1: Direct gate (a), dual gate (b), and NAT (c).

A NAT has multiple NIN-AND gates. A *direct* gate involves disjoint sets of causes $W_1, ..., W_m$. Each input event is a success $e \geq e^k \leftarrow \underline{w}_i^+$ $(i = 1, ..., m)$, e.g., Fig. 1 (a) where W_i is a singleton. Output event $e \geq e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+$ has probability $\prod_{i=1}^m P(e \geq e^k \leftarrow \underline{w}_i^+)$. Direct gates encode undermining causal interactions.

Each input of *dual* gate is a failure $e < e^k \leftarrow \underline{w}_i^+$, e.g., Fig. 1 (b). Output event $e < e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+$, has probability $\prod_{i=1}^m P(e < e^k \leftarrow \underline{w}_i^+)$ and satisfies relation $P(e < e^k \leftarrow ...) = 1 - P(e \geq e^k \leftarrow ...)$. Dual gates encode reinforcement causal interactions.

Fig. 1 (c) shows a NAT, where causes h_1 and h_2 reinforce each other, and so do b_1 and b_2. However, the two groups undermine each other. That is, for gate g_1, each W_i (as in Def. 1) is a general set. See (Xiang 2012) for a formal definition of NAT. From the NAT and probabilities of its input events, in the general form $P(e^k \leftarrow c_i^j)$ $(j, k > 0)$, called *single-causals*, $P(e \geq e^1 \leftarrow h_1^1, h_2^1, b_1^1, b_2^1)$ can be obtained. From the single-causals and all derivable NATs, CPT $P(e|h_1, h_2, b_1, b_2)$ is uniquely specified. A NAT model is specified by the topology and single-causals with space linear in n.

The *leaky* cause for e represents all causes of e not explicitly named. If a leaky cause is always active, it is *persistent* (Henrion 1989). Special issues arise when NAT-models have persistent leaky causes (Xiang and Jiang 2018).

A BN where CPT of some family is NAT model is a *NAT-modeled BN*. A BN has $O(N\ \kappa^n)$ space, where N is number of variables, κ bounds domain size of variables, and $n+1$ bounds family size. A NAT-modeled BN where every family of size > 2 is NAT model has $O(N\ \kappa\ n)$ space. A CPT can be approximated into a NAT model by *compression* (Xiang and Jiang 2018). By compressing each CPT, a BN is approximated by a NAT-modeled BN. Common inference methods for BNs can be used for NAT-modeled BNs by converting them into BNs, e.g., through trans-causalization (Xiang and Loker 2020). The inference is tractable if NAT-modeled BNs have high treewidth and low density [1].

3 Tractable Inference by Extended DD

Complexity of BN inference is exponential on its treewidth, which extends to DD for HBNs. Since DD inference goes in multiple rounds, this exponential cost is amplified in DD. To enable tractable inference with HBNs, we propose a novel framework by extending DD with local modeling:

We apply local modeling to discretized continuous variables in each round of DD, to create a BN with tractable inference. Since the BN in each round has specific static discretization (revised bin settings), local modeling must be revised. Hence, our framework requires extra tasks: The 1st creates initial local models for continuous variables (at initial DD round). The 2nd adapts local models for continuous variables newly discretized (at each DD round).

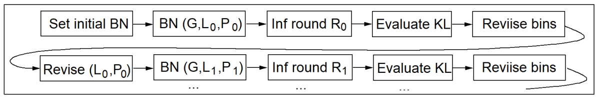

Figure 2: The extended DD framework for HBN inference.

Fig. 2 illustrates the framework (one row per DD round). G denotes HBN structure (directed acyclic graph), L_i $(i = 0, 1, ...)$ denotes local models of ith round, and P_i denotes probability parameters. The 1st task occurs in 1st box of 1st row. The 2nd task occurs in 1st box of each other row.

The two tasks must be efficient. Otherwise, their cost can outweigh savings from tractable inference in each round (3rd

[1] Low density does not imply tractability: Tree BNs (low density) of large n values (large treewidth) are exponential on n.

box in each row). For example, if achieved by *compression*, they are super-exponential on family size $n + 1$ (Xiang and Jiang 2018). Below, we exemplify the framework through a case-study on sum of reals, and present efficient solutions on the two tasks, when local modeling are NATs.

4 Alternative Methods for Sum of Reals

Consider discrete convolution (DCov) that replaces integration in Eqn. (1) by summing over even bins. For 2 PDFs discretized into k and $m \geq k$ bins, DCov takes $k + m$ rounds, each of $O(k)$ multiplications, with $O(k(k+m))$ time. Prior inference for sum needs $n - 1$ rounds of DCov. With q bins per addend, 1st round has $k = m = q$, and 2nd round has $k = q$ and $m = 2q$. Hence, prior for sum has $O(n^2\ q^2)$ time.

To compute posteriors over addends and sum, $P(w'|u'_1, ..., u'_n)$ is needed: q^n CPDs (conditional probability distributions). Since addend bins are uneven during DD, each of the q bins needs to be divided up to x even sub-bins in order to perform DCov. Cost for each CPD is $O(n^2\ x^2)$ and that for $P(w'|u'_1, ..., u'_n)$ is $O(n^2\ x^2\ q^n)$: intractable for large q and n.

In summary, DCov for prior of sum is efficient and *discretely exact* (inaccuracy due to discretization only), but DCov for posterior is exponential. Hence, we use DCov as golden standard for accuracy on priors only.

Uniform mixture (Fenton and Neil 2012) (D.3.3) can get approximate sum CPT $P(w'|u'_1, ..., u'_n)$ (not equivalent to convolution, as counter-example can be constructed), with uneven bins. For each $(u'_1, ..., u'_n)$, get bounds l and h, that defines distribution $U(l, h)$. For each bin of w', set $P(w'|u'_1, ..., u'_n)$ percentage of w' overlapping with $[l, h]$. With q bins per addend and $O(q^n)$ CPT values, cost to get $P(w'|u'_1, ..., u'_n)$ is $O(2n\ q^n)$: exponential time.

Other relevant work includes tensor rank-one decomposition for sum of integers (Savicky and Vomlel 2007). Whether it supports tractable inference with HBNs requires further research. The same holds for arithmetic circuits (Darwiche 2003). Parent divorcing (Olesen et al. 1989) is integrated into NAT-modeled DD through trans-causalization (Xiang and Loker 2020) (see Section 7).

5 NAT-modeled DD with Even Bins

We resolve 1st task in Section 3 on creating initial local models. We do so with even bins, as they can be adjusted by subsequent DD.

Without losing generality, assume $u_k \geq 0$ for $k = 1, ..., n$ and hence $w \geq 0$. First, consider even bin width $L = 1$. Let each bin be $[i, i+1)$, denoted by b_i, where $i = 0, 1, 2, ...$. We use $u'_k = b_i$ to denote $u_k \sim U(i, i+1)$ and hence $i = \lfloor u_k \rfloor$. If $u'_k = b_0$ for each k, then $w' \in \{b_0, ..., b_{n-1}\}$. By Eqn. (2), $P(w' = b_i|u'_1 = b_0, ..., u'_n = b_0) = F(i+1; n) - F(i; n)$ $(i = 0, ..., n-1)$. For example, define bins $b_0 = [0, 1)$, $b_1 = [1, 2)$, etc. Consider $P(w'|u'_1 = b_0, u'_2 = b_0, u'_3 = b_0) = P(w'|b_0, b_0, b_0)$, where $u'_k = b_0$ denotes $u_k \sim U(0, 1)$, and $w' \in \{b_0, b_1, b_2\}$. The result is $P(w' = b_i|b_0, b_0, b_0) = 1/6$ $(i = 0, 2)$ and $P(w' = b_1|b_0, b_0, b_0) = 2/3$.

When $u'_k = b_i \geq b_0$, define $\psi = \sum_{k=1}^{n} \lfloor u_k \rfloor$. It follows that $w \in [\psi, \psi+n)$ and $w' \in \{b_\psi, ..., b_{\psi+n-1}\}$. By Eqn. (2),

$$P(w' = b_{\psi+i}|u'_1, ..., u'_n) = F(i+1; n) - F(i; n) \quad (3)$$

holds for $i = 0, ..., n-1$. Eqn. (3) specifies sum CPT with unit bin width $L = 1$. Table 1 shows an example where domains of u_1, u_2, u_3 are $[0, 2]$, $[0, 3]$, $[0, 2]$, respectively.

$w' = b_0$	b_1	b_2	b_3	b_4	b_5	b_6	u'_1	u'_2	u'_3
1/6	2/3	1/6	0.0	0.0	0.0	0.0	b_0	b_0	b_0
0.0	1/6	2/3	1/6	0.0	0.0	0.0	b_0	b_1	b_0
0.0	0.0	1/6	2/3	1/6	0.0	0.0	b_0	b_2	b_0

Table 1: $P(w'|u'_1, u'_2, u'_3)$ (only 3 rows out of 12 are shown).

Next, consider even bin width $L \neq 1$. Let bin $[i\ L, (i+1)L)$ be denoted by b_i, and $u'_k = b_i$ denote $u_k \sim U(i\ L, (i+1)L)$. Hence, $i = u_k/L$ by integer division. For example, if $L = 1.5$ and $u_k = 3.3$, then $i = 2$. If $u'_k = b_0$ for each k, then $w' \in \{b_0, ..., b_{n-1}\}$. It is easy to see that

$$P(w' = b_i|u'_1 = b_0, ..., u'_n = b_0) = F(i+1; n) - F(i; n)$$

holds for $(i = 0, ..., n-1)$, identically to Eqn. (3) when $L = 1$. It follows that inference for $L \neq 1$ can be performed by scaling variable domains with $1/L$ and using unit bins. Hence, we assume unit bins when even bins are involved.

Once a sum CPT is specified, it must be compressed into a NAT model (Section 2) for efficient inference. Compression determines both NAT and single-causals. It is infeasible to be entirely offline (infinitely many potential sum CPTs). It is costly to be entirely online, since it involves searching through NAT topologies exponential in n.

We propose semi-offline compression: Compress offline a range of sum CPTs to identify the suitable NAT. For a particular sum CPT, compress online given the NAT to determine single-causals (much more efficient due to fixed NAT).

Our offline experiment (details omitted for space) found that the best NAT is dual NIN-AND gate (Fig. 1 (b)) with a persistent leaky cause. Table 2 shows a NAT model with persistent leaky cause $c_0 \in \{c_0^0, c_0^1\}$. The sum CPT has 72 parameters, while the NAT model has 30. Euclidean distance (ED) between NAT CPT and sum CPT is 0.159.

w'	b_1	b_2	b_3	b_4	b_5	b_6
$P(w' \leftarrow c_0^1)$	.534	.219	.109	.01	.01	.01
$P(w' \leftarrow u'_1 = b_1)$	.202	.292	.192	.193	.01	.01
$P(w' \leftarrow u'_2 = b_1)$	.195	.508	.166	.01	.01	.01
$P(w' \leftarrow u'_2 = b_2)$	.109	.170	.232	.160	.179	.039
$P(w' \leftarrow u'_3 = b_1)$	.147	.120	.150	.273	.059	.01

Table 2: NAT single causals for sum CPT in Table 1.

6 Model Sum CPT with Bin Merge & Split

Next, we consider 2nd task in Section 3 on how to adapt local model at each DD round. This task is driven by bin revisions at either addends or sum, and revised bins require revised NAT model for sum CPT. As bin revision results in uneven bins, the above method for 1st task does not apply.

We resolve NAT-modeling on uneven bins as follows: Given initial NAT model for sum CPT with even bins, modify the NAT model directly according to revised bins. The new NAT model is on uneven bins. Bin merge or split may occur to addend or sum, forming 4 operations analyzed below. For accuracy, we require that NAT CPT after bin revision be *consistent* with the NAT CPT before revision:

Definition 2 *Let M be NAT model on $V = \{u'_1, ..., u'_n, w'\}$ (addends and sum) and N be NAT model after bin revision on $v \in V$. If sum CPT from N differs only on terms that involve modified bins, the difference is probabilistically sound, and other terms are invariant, then N is consistent with M.*

[**Merge sum bins**] Let w' bins b_i, b_{i+1} be merged into $b_{i,i+1}$. We have single-causals $P(b_i \leftarrow u'_k)$, $P(b_{i+1} \leftarrow u'_k)$, and NAT CPT terms $P(b_i|U')$, $P(b_{i+1}|U')$ before merging, where $U' = \{u'_1, ..., u'_n\}$. After merging, we set new NAT model as follows without change to other single-causals:

$$P(b_{i,i+1} \leftarrow u'_k) = P(b_i \leftarrow u'_k) + P(b_{i+1} \leftarrow u'_k), \quad (k = 1, ..., n). \quad (4)$$

Theorem 1 holds, whose proof is omitted due to space.

Theorem 1 *Bin merging for sum w' according to Eqn. (4) is consistent, satisfies below, and is invariant otherwise:*

$$P(b_{i,i+1}|U') = P(b_i|U') + P(b_{i+1}|U'). \quad (5)$$

[**Split sum bins**] Consider splitting bin b_i of sum w' into b_{ia} and b_{ib}, where $|b_{ia}| = |b_{ib}| = |b_i|/2$. Since it is inverse of merging that is consistent, we guide splitting by Eqns. (5) and (4). As we only have left-hand of Eqn. (5), its right-hand cannot be derived, nor can right-hand of Eqn. (4). Hence, we split single-causal on w' bin b_i to b_{ia} and b_{ib} to enable

$$P(b_i \leftarrow u'_k) = P(b_{ia} \leftarrow u'_k) + P(b_{ia} \leftarrow u'_k), \quad (6)$$

and set right-hand terms by single-causal density in bins adjacent to b_i through 4 cases (Fig. 3), where density $f_i = P(b_i \leftarrow u'_k)/|b_i|$. Case 1: $f_{i-1} \leq f_i \leq f_{i+1}$ (increasing). Case 2: $f_{i-1} \geq f_i \geq f_{i+1}$ (decreasing). Case 3: $f_i > max(f_{i-1}, f_{i+1})$ (convex). Case 4: $f_i < min(f_{i-1}, f_{i+1})$ (concave).

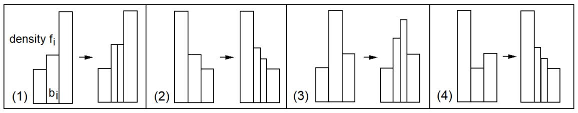

Figure 3: Cases of sum bin split.

Case 1: Estimate density over b_{ia} and b_{ib} as $f_{ia} = f_{ib} = (f_{i-1} + f_{i+1})/2$. Set their probability mass tentatively to $t_{ia} = t_{ib} = f_{ia}\ |b_i|/2$. Note that f_i is not referenced. The mass distribution may not sum to one, to be handled below.

Case 2: We estimate density over b_{ia} and b_{ib} as $f_{ia} = (f_i + f_{i-1})/2$ and $f_{ib} = (f_i + f_{i+1})/2$, respectively.

Case 3: Estimate density over b_{ia} and b_{ib} so that new bin adjacent to the bin with density $max(f_{i-1}, f_{i+1})$ has higher density. In particular, we estimate density over b_{ia} and b_{ib} as $f_{ia} = (f_{i-1} + 3f_i)/4$ and $f_{ib} = (3f_i + f_{i+1})/4$.

Case 4: We estimate density over b_{ia} and b_{ib} as $f_{ia} = (2f_{i-1} + f_{i+1})/3$ and $f_{ib} = (f_{i-1} + 2f_{i+1})/3$. It aims to avoid erroneous valley in sum distribution by removing the valley in single-causal distribution. As the result, $f_{ia}, f_{ib} > min(f_{i-1}, f_{i+1})$ where $(f_{i-1}, f_{ia}, f_{ib}, f_{i+1})$ is monotonic.

Since mass distribution from above (for $P(w' \leftarrow u'_k)$ where $w' \in \{b_0, ..., b_{i-1}, b_{ia}, b_{ib}, b_{i+1}, ...\}$) may not sum to one, we scale single-causal mass in a window W of bins to renders new $P(w' \leftarrow u'_k)$ summing to one. We omit details due to space. Note that the 4 case rules are not compelled, and are approximately consistent due to scaling.

[**Merge addend bins**] Let bins b_i and b_{i+1} of addend u'_k be merged into $b_{i,i+1}$. Before merging, we have single-causals $P(w' \leftarrow u'_k = b_i)$, $P(w' \leftarrow u'_k = b_{i+1})$, and NAT CPT items $P(w'|u'_k = b_i, U'_-)$ and $P(w'|u'_k = b_{i+1}, U'_-)$, where $U'_- = U' \setminus \{u'_k\}$. After merging, set NAT model as follows without change to other single-causals:

$$\begin{aligned} & P(w' \leftarrow u'_k = b_{i,i+1}) \\ = \quad & \rho_i P(w' \leftarrow u'_k = b_i) + \rho_{i+1} P(w' \leftarrow u'_k = b_{i+1}), \end{aligned} \quad (7)$$

where $\rho_j = P(b_j)/(P(b_i) + P(b_{i+1}))$. Theorem 2 holds.

Theorem 2 *Bin merging for addend u'_k according to Eqn. (7) is consistent and satisfies*

$$\begin{aligned} & P(w'|b_{i,i+1}, U'_-) \\ = \; & \tfrac{P(w'|b_i,U'_-)P(b_i)}{P(b_i)+P(b_{i+1})} + \tfrac{P(w'|b_{i+1},U'_-)P(b_{i+1})}{P(b_i)+P(b_{i+1})}. \end{aligned} \quad (8)$$

[**Splitting addend bins**] We split bin b_i of addend u'_k into b_{ia} and b_{ib}, where $|b_{ia}| = |b_{ib}| = |b_i|/2$. As it is inverse of merging which is consistent, we guide splitting by Eqn. (8). Before splitting, we know only left-hand of Eqn. (8) and right-hand denominator. Remaining terms cannot be determined. We set $P(w' \leftarrow u'_k = b_{ia})$ and $P(w' \leftarrow u'_k = b_{ib})$ in new NAT model as follows without changing other single-causals. Suppose we have

$$\begin{aligned} P(u'_k \in b_{ia})/P(u'_k \in b_i) &= |b_{ia}|/|b_i| \\ P(u'_k \in b_{ib})/P(u'_k \in b_i) &= |b_{ib}|/|b_i|. \end{aligned} \quad (9)$$

Since $|b_{ia}| = |b_{ib}|$ and $|b_{ia}| + |b_{ib}| = |b_i|$, we have $\frac{|b_{ia}|}{|b_i|} = \frac{|b_{ib}|}{|b_i|} = 1/2$. We set new NAT model with

$$P(w' \leftarrow u'_k = b_{ia}) = P(w' \leftarrow u'_k = b_{ib}) = P(w' \leftarrow u'_k = b_i). \quad (10)$$

Since this setting satisfied Eqn. (7) with $\rho_i = \rho_{i+1} = 1/2$, it follows that Eqn. (8) holds, assuming condition in Eqn. (9).

At the time of bin split, we have $P(u'_k \in b_i)$ from previous round of DD inference. We do not have $P(u'_k \in b_{ia})$ and $P(u'_k \in b_{ib})$, nor do we guarantee Eqn. (9). Hence, addend bin split by Eqn. (10) is approximately consistent.

7 Bin Merge & Split for Trans-Causalization

Since NAT-modeled BNs are converted to BNs for inference, e.g., by trans-causalization, a new trans-causalization is needed at each DD round, after sum or addend bin revision. To enhance efficiency, we save trans-causalizing cost by applying bin revision directly to trans-causalized BN:

From Section 5, best NAT model of sum CPT is dual NIN-AND gate with persistent leaky cause. For n addends, there are $n + 1$ causes. They form $n + 1$ root nodes (Fig. 4 (a)) in trans-causalized structure, each with 1 probabilistic child (z-node). Each z-node has 1 deterministic child (y-node). Each y-node has 2 parents: at least one z-node and at most one y-node. Each y-node has at most one child (y-node), and y_n stands for sum. Domains of z and y-nodes are the same as sum w'. Each y-node has a (deterministic) MAX CPT. CPT of each z-node captures single-causals of its parent cause.

Fig. 4 (b)) shows a BN for sum $sum0$ of addends $ad1, ..., ad4$, and marginal distributions at end of DD. Addend PDFs are Gaussian, Beta, Triangular, and Uniform, respectively. The persistent leaky cause is $plc0$. It has $n = 4$, 5 z-nodes $z6, ..., z10$, and 4 y-nodes $y11, y12, y13, sum0$.

After bin revision during DD, we directly modify domains and CPTs for necessary z-nodes and y-nodes only. For instance, if only two addends revise bins, only CPTs of two z-nodes are revised. Revision of z-node CPT is per Section 6 and revision of y-node CPT is functional.

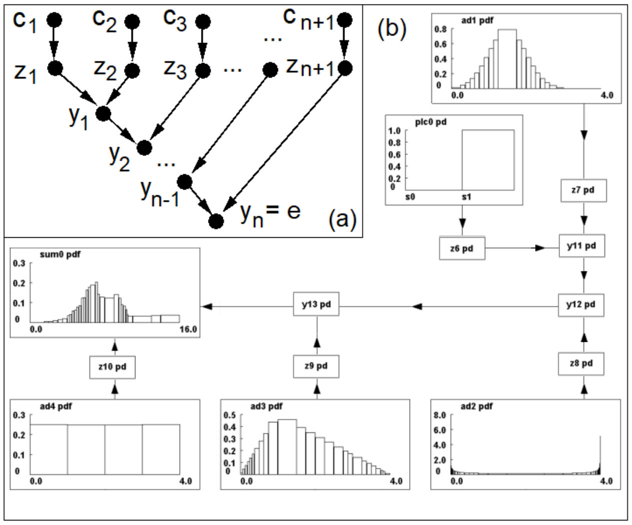

Figure 4: (a) Trans-causalized structure. (b) Example structure for sum of 4 addends.

8 NAT-Modeled DD and Complexity

Algorithm 1 specifies NAT-Modeled DD, integrating above techniques. Inference at each round uses junction tree (JT) message passing (Jensen, Lauritzen, and Olesen 1990). Lines 1 through 5 are performed offline once for all. Remaining lines are online for each new set of observations.

Algorithm 1

```
Input: HBN $B_0$ specifying addend PDFs and observation for some variables;
1  convert $B_0$ to BN $B_1$ with even bins and Irwin-Hall distributions for sum CPT;
2  compress sum CPT to dual-gate NAT model $M$ with persistent leaky cause;
3  convert $B_1$ to NAT-modeled BN $B_2$ using $M$;
4  trans-causalize $B_2$ into BN $B_3$;
5  compile $B_3$ into JT $T$;
6  for $i = 1$ to $MaxRound$, do
7      enter observation into $T$;
8      message passing in $T$ for posterior marginals;
9      for each discretized variable $u'$, do
10         compute approximate KL error on $u'$;
11         if KL error on $u' > ErrorBound$,
12             revise bins for $u'$;
13             revise $B_3$ and $T$ locally accordingly;
14     if no $u'$ is found whose KL error $> ErrorBound$, break;
15 return posterior marginals for all variables in $B_1$ from $T$;
```

Next, we analyze complexity for revising trans-causalized BN at each DD round (lines 12, 13). When sum bins are revised, each y-node revises its domain and CPT. At start of DD (even bins), domain size of each addend is q, and domain size of sum is $n\,q$. Each MAX CPT has size $n^3\,q^3$. Each z-node revise its domain and CPT (size $n\,q^2$), using updated single-causals. Complexity of revision is $S_1 = O(n(n^3\,q^3 + n\,q^2)) = O(n^4\,q^3)$. When bins of an addend are revised, its child z-node revises CPT, using updated single-causals. The CPT has size $n\,q^2$. No change to y-nodes is needed. In case all addends revise their bins, complexity of revision is $S_2 = O(n^2\,q^2)$. From S_1, S_2, complexity of bin revision at both sum and addends is $S_3 = O(n^4\,q^3)$.

Finally, we consider inference complexity in a DD round (line 8). The largest JT cluster has size 3 (see Fig. 4), and space size $n^3\,q^3$. Hence, complexity of inference is $S_4 = O(n^4\,q^3)$. From S_3, S_4, one round of NAT-modeled DD for sum inference has complexity of $S_5 = O(n^4\,q^3)$, reducing those (exponential) in Section 4 to polynomial.

9 Experimental Study

We implemented DCov, Irwin-Hall based static discretization (IHSD), uniform mixture based DD (UMDD), and NAT-modeled DD (NATDD), with JT based inference. MacBook Pro of 2.5GHz CPU and 16 GB memory was used.

[**Priors for sum**] For inference with prior sum probability distribution (PD), we set number of addends $n = 3, ..., 10$ with 30 HBNs for each n (240 HBNs in total), and each HBN contains a n-addend family. For $n = 3, 4$, domain size of each addend is selected from $[0, 3]$, $[0, 4]$, $[0, 5]$, and $[0, 6]$. For $n = 5, 6$, it is from $[0, 3]$ and $[0, 4]$. For $n = 7, ..., 10$, it is set to $[0, 3]$. All HBNs start with even bin size 1. At end of DD, typical minimum bin size is 0.125.

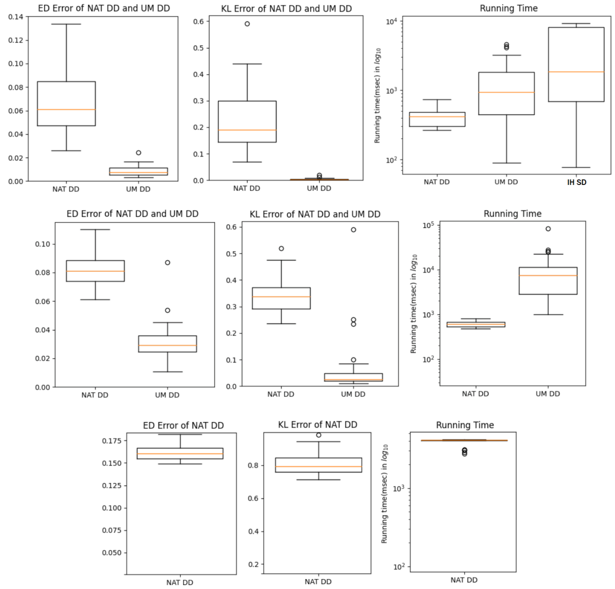

Figure 5: Prior inference for $n = 4$ (top), 6 (middle), and 10.

For $n = 3, ..., 7$, we run 3 methods per HBN and, for $n = 8, 9, 10$, we run 2 methods: 630 inference runs. We report ED and KL errors of sum PD by NATDD and UMDD from golden standard IHSD or DCov for $n = 4, 6, 10$, due to space. We also report runtime of NATDD, UMDD and IHSD (DCov does not support posterior inference and is used as golden standard for accuracy in prior inference).

For $n = 4$, NATDD, UMDD, and IHSD are applied, with IHSD as accuracy standard, and NATDD is compared with

UMDD on accuracy and efficiency. For $n = 6$, NATDD, UMDD, and DCov are applied, where DCov replaces IHSD (runs out of memory) as standard. For $n = 10$, only NATDD and DCov are run (UMDD runs out of memory).

In Fig. 5, subtitle at top-left denotes ED errors of NATDD and UMDD, relative to standard. When n increases from 4 to 10, average ED error by NATDD increases from 0.06 to 0.16: not large, but larger than UMDD. Runtime of NATDD increases from 0.4 sec ($n = 4$) to 4 sec ($n = 10$). At $n = 7$, NATDD is two orders of magnitude faster than UMDD (not shown in Fig. 5). At $n > 7$, UMDD cannot complete.

[**Posteriors on addends and sum**] For inference with posterior sum PD, we set $n = 3, 4, 5$, so that IHSD can be run with NATDD and UMDD as standard. HBNs are the same as above with $n = 3, 4, 5$. For each HBN, we observe sum and one addend, with random observation for other addends. Average ED and KL errors of posterior marginals are measured for NATDD and UMDD. Errors and runtime are summarized in Fig. 6 (note that Fig. 5 differs in $n = 4, 10$).

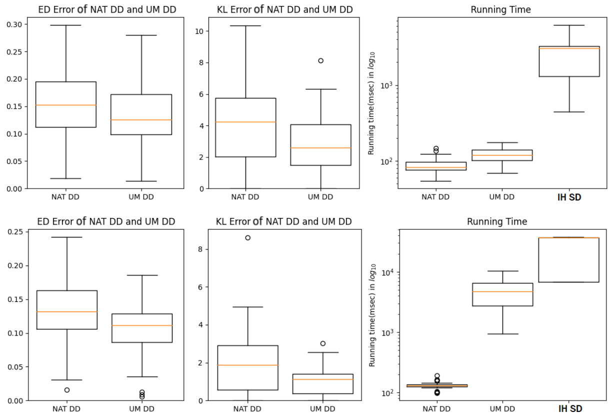

Figure 6: Posterior inference for $n = 3$ (top) and 5 (bottom).

ED errors for NATDD are slightly higher than UMDD, and difference is less pronounced than in prior inference. This is consistent with (Xiang and Baird 2018): Posterior errors are generally smaller than CPT compression errors. On the other hand, at $n = 5$, NATDD is about 30 times faster than UMDD.

10 Conclusion

We contribute a novel framework of NAT-modeled DD for inference with HBNs and case study with sum of reals. Key tasks are creation of initial local models, and their efficient adaptation during DD. We developed techniques to accomplish them efficiently. Experiment showed significant efficiency gain over alternative methods, while incurring low inference errors. Much of the techniques are generalizable to NAT-modeled DD with other HBNs, which forms a further research direction.

References

Boutilier, C.; Friedman, N.; Goldszmidt, M.; and Koller, D. 1996. Context-specific independence in Bayesian networks. In *Proc. 12th Conf. on Uncertainty in Artificial Intelligence*, 115–123.

Darwiche, A. 2003. A differential approach to inference in Bayesian networks. *J. ACM* 50(3):280–305.

Diez, F. 1993. Parameter adjustment in Bayes networks: The generalized noisy OR-gate. In Heckerman, D., and Mamdani, A., eds., *Proc. 9th Conf. on Uncertainty in Artificial Intelligence*, 99–105. Morgan Kaufmann.

Fenton, N., and Neil, M. 2012. *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press.

Grinstead, C., and Snell, J. 2003. *Introduction to Probability*. American Mathematical Society.

Hall, P. 1927. The distribution of means for samples of size n drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika* 19(3/4):240–245.

Henrion, M. 1989. Some practical issues in constructing belief networks. In Kanal, L.; Levitt, T.; and Lemmer, J., eds., *Uncertainty in Artificial Intelligence 3*. Elsevier Science Publishers. 161–173.

Irwin, J. 1927. On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to pearson's type ii. *Biometrika* 19(3/4):225–239.

Jensen, F.; Lauritzen, S.; and Olesen, K. 1990. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly* (4):269–282.

Kang, J.; Kim, S.; Kim, Y.; and Jang, Y. 2010. Generalized convolution of uniform distributions. *J. Appl. Math. & Informatics* 28(5-6):1573–1581.

Killmann, F., and von Collani, E. 2001. A note on the convolution of the uniform and related distributions and their use in quality control. *Economic Quality Control* 16(1):17–41.

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

Maaskant, P., and Druzdzel, M. 2008. An independence of causal interactions model for opposing influences. In Jaeger, M., and Nielsen, T., eds., *Proc. 4th European Workshop on Probabilistic Graphical Models*, 185–192.

Moral, S.; Rumi, R.; and Salmeron, A. 2001. Mixtures of truncated exponentials in hybrid Bayesian networks. In *Lecture Notes in Artificial Intelligence*, volume 2143. Springer-Verlag. 135–143.

Neil, M.; Tailor, M.; and Marquez, D. 2007. Inference in hybrid Bayesian networks using dynamic discretization. *Statistics and Computing* 17(3):219–233.

Olesen, K.; Kjaerulff, U.; Jensen, F.; Jensen, F.; Falck, B.; Andreassen, S.; and Andersen, S. 1989. A munin network for the median nerve-a case study on loops. *Applied Artificial Intelligence* 3(2-3):385–403.

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

Savicky, P., and Vomlel, J. 2007. Exploiting tensor rank-one decomposition in probabilistic inference. *Kybernetika* 43(5):747–764.

Shenoy, P., and West, J. 2011. Inference in hybrid Bayesian networks using mixtures of polynomials. *Inter. J. of Approximate Reasoning* 52(5):641–657.

Woudenberg, S.; van der Gaag, L.; and Rademaker, C. 2015. An intercausal cancellation model for Bayesian-network engineering. *Inter. J. Approximate Reasoning* 63:32–47.

Xiang, Y., and Baird, B. 2018. Compressing Bayesian networks: Swarm-based descent, efficiency, and posterior accuracy. In Bagheri, E., and Cheung, J., eds., *Canadian AI 2018, LNAI 10832*. Springer. 3–16.

Xiang, Y., and Jiang, Q. 2018. NAT model based compression of Bayesian network CPTs over multi-valued variables. *Computational Intelligence* 34(1):219–240.

Xiang, Y., and Loker, D. 2020. Trans-causalizing NAT-modeled Bayesian networks. *IEEE Trans. Cybernetics, DOI: 10.1109/TCYB.2020.3009929.*

Xiang, Y. 2012. Non-impeding noisy-AND tree causal models over multi-valued variables. *International J. Approximate Reasoning* 53(7):988–1002.