Efficient Computation of Belief Theoretic Conditionals

Lalintha G. Polpitiya, Kamal Premaratne, Manohar N. Murthi and Dilip Sarkar

University of Miami
Coral Gables, Florida (USA)

10th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA ’17)
Lugano, Switzerland, 10-14 July 2017

Acknowledgement: This work is based on research supported by the U.S. Office of Naval Research (ONR) via grant #N00014-10-1-0140, and the U.S. National Science Foundation (NSF) via grant #1343430.

12 July 2017
Outline

1. Challenges and Contributions
2. Theoretical Foundation
3. DS-Conditional-One Computational Model
4. Computation of Conditionals
5. Experiments and Concluding Remarks
Motivation and Challenges

- **Dempster-Shafer (DS)** theory offers greater expressiveness and flexibility in evidential reasoning.

- **Conditional operation**: plays a pivotal role in DST strategies for evidence updating and fusion.

- **Implementations**: restricted to smaller frames of discernment (FoDs) because of the prohibitive computational burden that larger FoDs impose on existing methods.

- **Several approximation methods**: compromise the quality of the generated results for computational efficiency.

- **Exact computation of conditionals is of paramount importance**: quality of results generated from DS theoretic (DST) strategies depend directly on the precision of the conditional.
Contributions
to fill the void between what DS theory can offer and its practical implementation

- The main contribution: a generalized computational model (DS-Conditional-One) for computing DST conditionals.

- This model can be employed to compute both the Fagin-Halpern (FH) and Dempster’s conditional beliefs of an arbitrary proposition.

- This is exactly the challenge that Shafer refers to in [Shafer, 1990, p.348], viz., “It remains to be seen how useful the fast Möbius transform will be in practice. It is clear, however, that it is not enough to make arbitrary belief function computations feasible.”
Contributions
to fill the void between what DS theory can offer and its practical implementation

- This new model can also be utilized as a visualization tool for conditional computations and in analyzing characteristics of conditioning operations.

- An outcome of this research is a conditional computation library which is available online.
  (https://profuselab.github.io/Conditional-Computation-Library/)
## Basic notions of belief theory

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta$</td>
<td><em>Frame of discernment (FoD)</em>, i.e., the set of all possible mutually exclusive and exhaustive propositions.</td>
</tr>
<tr>
<td>$\theta_i$</td>
<td><em>Singletons</em>, i.e., the lowest level of discernible information, i.e., $\Theta = {\theta_0, \ldots, \theta_{n-1}}$, here $n =</td>
</tr>
<tr>
<td>$\overline{A}$</td>
<td><em>Complement</em> of the proposition $A \subseteq \Theta$, i.e., those singletons that are not in $A$.</td>
</tr>
<tr>
<td>$m(\cdot)$</td>
<td><em>Basic belief assignment (BBA)</em> or <em>mass assignment</em> $m : 2^\Theta \rightarrow [0, 1]$ where $\sum_{A \subseteq \Theta} m(A) = 1$ and $m(\emptyset) = 0$.</td>
</tr>
</tbody>
</table>

**Focal element**

Singleton or composite (i.e., non-singleton) proposition that receives a non-zero mass.

**$\mathcal{F}$**

*Core*, the set of focal elements.

**$\mathcal{E}$**

*Body of evidence (BoE)* represented via the triplet $\{\Theta, \mathcal{F}, m\}$.

**Subset propositions of $A$**

Subsets of $A$ and $A$ itself, here $m = |A|$.
The following notation is useful for our work:

\[ S(A; B) = \sum_{\emptyset \neq C \subseteq A; \emptyset \neq D \subseteq B} m(C \cup D). \]

\( S(A; B) \) denotes the sum of all masses of propositions that ‘straddle’ both \( A \subseteq \Theta \) and \( B \subseteq \Theta \).
The following result is of critical importance for our work:

**Proposition 1**

Consider the body of evidence (BoE) \( \mathcal{E} = \{ \Theta, \emptyset, m(\cdot) \} \) and \( A \subseteq \Theta \). For \( B \subseteq \Theta \), consider the mappings \( \Gamma_A : 2^\Theta \mapsto [0, 1] \) and \( \Pi_A : 2^\Theta \mapsto [0, 1] \), where

\[
\Gamma_A(B) = \sum_{\emptyset \neq X \subseteq \overline{A}} m((A \cap B) \cup X); \quad \Pi_A(B) = \sum_{Y \subseteq (A \cap B)} \Gamma_A(Y).
\]

Then the following are true:

(i) \( \Gamma_A(A \cap B) = \Gamma_A(B) \) and \( \Pi_A(A \cap B) = \Pi_A(B) \). So, w.l.o.g., we assume that \( B \subseteq A \).

(ii) \( \Gamma_A(\emptyset) = B \cap \overline{A} \).
Fagin-Halpern (FH) conditional

Most natural generalization of the probabilistic conditional notion - close connection with the inner and outer conditional probability measures.

**Definition 2 (Fagin-Halpern (FH) conditional)**

[Fagin and Halpern, 1990]

Consider the BoE $\mathcal{E} = \{\emptyset, \mathcal{F}, m(\cdot)\}$ and $A \in \mathcal{F}$. The conditional belief $Bl(B|A)$ of $B$ given the conditioning event $A$ is

$$Bl(B|A) = \frac{Bl(A \cap B)}{Bl(A \cap B) + Pl(A \cap \overline{B})}.$$

**Proposition 3**

Consider the BoE $\mathcal{E} = \{\emptyset, \mathcal{F}, m(\cdot)\}$ and $A \in \mathcal{F}$. Then, we may express $Bl(B|A)$ as

$$Bl(B|A) = \frac{Bl(A \cap B)}{1 - Bl(\overline{A}) - S(\overline{A}; A \cap B)}, \quad B \subseteq \emptyset.$$
Dempster’s conditional

Perhaps the most widely employed DST conditional notion.

Definition 4 (Dempster’s conditional) [Shafer, 1976]

Consider the BoE $\mathcal{E} = \{\Theta, \emptyset, m(\cdot)\}$ and $A \subseteq \Theta$ s.t. $Bl(\overline{A}) \neq 1$, or equivalently, $Pl(A) \neq 0$. The conditional belief $Bl(B|A)$ of $B$ given the conditioning event $A$ is

$$Bl(B|A) = \frac{Bl(\overline{A} \cup B) - Bl(\overline{A})}{1 - Bl(\overline{A})}.$$ 

Proposition 5

Consider the BoE $\mathcal{E} = \{\Theta, \emptyset, m(\cdot)\}$ and $A \subseteq \Theta$ s.t. $Bl(\overline{A}) \neq 1$. Then, $Bl(B|A)$ can be expressed as

$$Bl(B|A) = \frac{Bl(A \cap B) + S(\overline{A}; A \cap B)}{1 - Bl(\overline{A})}, \quad B \subseteq \Theta.$$
Consider the FoD $\Theta = \{\theta_0, \theta_1, \ldots, \theta_{n-1}\}$.

Suppose we desire to determine the belief potential associated with $A = \{\theta_{k_1}, \theta_{k_2}, \ldots, \theta_{k_\ell}\} \subseteq \Theta$.

The REGAP property allows us to recursively generate the propositions that are relevant for this computation: Start with $\{\emptyset\}$. 

---

**Figure:** REGAP: REcursive Generation of and Access to Propositions, Start with $\emptyset$
**REcursive Generation of and Access to Propositions**

REGAP: Inserting singleton \(\{\theta_{k3}\}\)

---

**Figure:** REGAP: REcursive Generation of and Access to Propositions, inserting \(\{\theta_{k3}\}\)

- Inserting another singleton \(\{\theta_{k3}\} \in A\) brings the new propositions
  \(
  \emptyset \cup \{\theta_{k3}\} = \{\theta_{k3}\}, \quad \{\theta_{k1}\} \cup \{\theta_{k3}\} = \{\theta_{k1}, \theta_{k3}\}, \quad \{\theta_{k2}\} \cup \{\theta_{k3}\} = \{\theta_{k2}, \theta_{k3}\},
  \)
  and
  \(
  \{\theta_{k1}, \theta_{k2}\} \cup \{\theta_{k3}\} = \{\theta_{k1}, \theta_{k2}, \theta_{k3}\}.
  \)

- In essence, when a new singleton is added, new propositions associated with it can be recursively generated by adding the new singleton to each existing proposition.
REcursive Generation of and Access to Propositions

REGAP: Generalized representation

REGAP, stands for REcursive Generation of and Access to Propositions.

- All propositions of interest within the FoD $\Theta$ can be generated when $A = \Theta$.
- These recursively generated propositions can be formulated as a vector, a matrix or a tree, and utilized to represent a dynamic BoE.
DS-Conditional-One computational model

- **DS-Conditional-One** is a computational model that enables one to compute the FH and Dempster’s conditional beliefs of an *arbitrary* proposition.

- We denote the conditioning proposition $A$, its complement $\overline{A}$, and the conditioned proposition $B$ as $\{a_0, a_1, \ldots, a_{|A|-1}\}$, $\{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $\{b_0, b_1, \ldots, b_{|B|-1}\}$, respectively. Here, $\Theta = \{\theta_0, \theta_1, \ldots, \theta_{n-1}\}$ denotes the FoD and $a_i, \alpha_j, b_k \in \Theta$.

- We represent singletons of the conditioning event $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ as *column singletons* and singletons of the complement of conditioning event $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$ as *row singletons* in a DS-Matrix.
Figure: **DS-Conditional-One model.** Quantities related to $Bl(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$. 
The proposed DS-Conditional-One computational model allows direct identification of \( \text{REGAP}(A) \), \( \text{REGAP}(\overline{A}) \), \( \text{REGAP}(A \cap B) \), \( (\text{REGAP}(\overline{A}) \times \text{REGAP}(A \cap B)) \), \( (\text{REGAP}(A) \times \text{REGAP}(\overline{A})) \), and \( \Gamma_A(C), \forall C \subseteq B \).

We use a lookup table named \textit{power} to enhance the computational efficiency. It contains 2 to the power of singleton indexes in increasing order.

\( \text{index}[] \) is a dynamic array which keeps the indexes of subset propositions of \( A \cap B \).
Compute $BL(A \cap B)$ (with time complexity $O(2^{|A \cap B|})$)

Figure: **DS-Conditional-One model.** Quantities related to $BL(B | A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A.$
Compute $BL(A \cap B)$: Algorithm 1

```plaintext
1: procedure $BLB$(Singletons $A$, Singletons $B$, DS-Matrix $BBA$)
2:     believe $\leftarrow 0$
3:     count $\leftarrow 0$
4:     for each $a_i$ in $A \cap B$ do
5:         index[count] $\leftarrow$ power[$i$]
6:         temp $\leftarrow$ count
7:         count $\leftarrow$ count + 1
8:     for $j \leftarrow 0$, temp $-$ 1 do
9:         index[count] $\leftarrow$ index[$j$] + power[$i$]
10:        count $\leftarrow$ count + 1
11:     end for
12: end for
13: for $i \leftarrow 0$, power[|$A \cap B$|] $-$ 2 do
14:    belief $\leftarrow$ belief + $BBA[0][index[i]]$
15: end for
16: Return $belief$
17: end procedure
```
Compute $Bl(\overline{A})$ (with time complexity $O(2^{\overline{|A|}})$)

Figure: DS-Conditional-One model. Quantities related to $Bl(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$. 
Compute $B\ell(\bar{A})$: Algorithm 2

1: **procedure** $B\ell_{\text{Comp}}$(Singletons $\bar{A}$, DS-Matrix $BBA$)
2: \hspace{1em} belief $\leftarrow$ 0
3: \hspace{1em} for $i \leftarrow 1$, power[$|\bar{A}|$] $- 1$ do
4: \hspace{2em} belief $\leftarrow$ belief + $BBA[i][0]$
5: \hspace{1em} end for
6: \hspace{1em} Return belief
7: **end procedure**
Compute $S(\overline{A}; A \cap B)$ (with time complexity $O(2^{\left|\overline{A}\right| + \left|A \cap B\right|})$)

Figure: DS-Conditional-One model. Quantities related to $Bl(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$. 
Compute $S(\bar{A}; A \cap B)$: Algorithm 3

1: procedure $\text{Strad}(\text{Singletons } \bar{A}, \text{Singletons } A, \text{Singletons } B, \text{DS-Matrix } BBA)$

2: $belief \leftarrow 0$

3: $count \leftarrow 0$

4: for each $a_i$ in $A \cap B$ do

5: $index[count] \leftarrow power[i]$

6: $temp \leftarrow count$

7: $count \leftarrow count + 1$

8: for $j \leftarrow 0, temp - 1$ do

9: $index[count] \leftarrow index[j] + power[i]$

10: $count \leftarrow count + 1$

11: end for

12: end for

13: for $i \leftarrow 1, power[|\bar{A}|] - 1$ do

14: for $j \leftarrow 0, power[|A \cap B|] - 2$ do

15: $belief \leftarrow belief + BBA[i][index[j]]$

16: end for

17: end for

18: Return $belief$

19: end procedure
FH conditional belief of an **arbitrary** proposition

Use the expression in Proposition 3, where $Bl(A \cap B)$, $Bl(\overline{A})$ and $S(\overline{A}; A \cap B)$ are obtained via Algorithms 1, 2, and 3, respectively.

**Proposition 3**

Consider the BoE $\mathcal{E} = \{\Theta, \mathcal{F}, m(\cdot)\}$ and $A \in \mathcal{F}$. Then, we may express $Bl(B|A)$ as

$$Bl(B|A) = \frac{Bl(A \cap B)}{1 - Bl(\overline{A}) - S(\overline{A}; A \cap B)}, \quad B \subseteq \Theta.$$  

The computational complexity is $O(2^{|\overline{A}|} + |A \cap B|)$.

**$Bl(B|A)$ when $B = \{a_0, a_2\}$**

$$Bl(B|A) = \frac{Bl(A \cap B)}{1 - \Gamma_A(\emptyset) - \Gamma_A(\{a_0\}) - \Gamma_A(\{a_2\}) - \Gamma_A(\{a_0, a_2\}).$$
Dempster’s conditional belief of an **arbitrary** proposition

Use the expression in Proposition 5, where $BL(A \cap B)$, $BL(\bar{A})$ and $S(\bar{A}; A \cap B)$ are obtained via Algorithms 1, 2, and 3, respectively.

**Proposition 5**

Consider the BoE $E = \{\Theta, \emptyset, m(\cdot)\}$ and $A \subseteq \Theta$ s.t. $BL(\bar{A}) \neq 1$. Then, $BL(B\mid A)$ can be expressed as

$$BL(B\mid A) = \frac{BL(A \cap B) + S(\bar{A}; A \cap B)}{1 - BL(\bar{A})}, \ B \subseteq \Theta.$$

The computational complexity is $O(2^{|\bar{A}|} + |A \cap B|)$.

**$BL(B\mid A)$ when $B = \{a_0, a_2\}$**

$$BL(B\mid A) = \frac{BL(A \cap B) + \Gamma_A(\{a_0\}) + \Gamma_A(\{a_2\}) + \Gamma_A(\{a_0, a_2\})}{1 - \Gamma_A(\{\emptyset\})}.$$
Dempster’s conditional mass using specialization matrix
[Smets, 2002]

- It employs a $2^{|\Theta|} \times 2^{|\Theta|}$-sized stochastic matrix $\mathcal{S}_A$ (with each entry ‘0’ or ‘1’) referred to as the **conditioning specialization matrix** and a $2^{|\Theta|} \times 1$-sized vector $m(\cdot)$ containing the focal elements.

- The computational and space complexity of the specialization matrix multiplication is $O(2^{|\Theta|} \times 2^{|\Theta|})$, a prohibitive burden even for modest FoD sizes.
Size of the DS-Conditional-One matrix vs Specialization matrix
When FoD size is 4

Figure: Size of the DS-Conditional-One matrix vs Specialization matrix: When FoD size is 4
Size of the DS-Conditional-One matrix vs Specialization matrix

When FoD size is 10

Figure: Size of the DS-Conditional-One matrix vs Specialization matrix: When FoD size is 10
Experiments

- For a given FoD size, we selected a random set of focal elements, with randomly selected mass values, and conducted 10,000 conditional computations for randomly chosen propositions $A$ and $B \subseteq A$.

- With the DS-Conditional-One model (which applies to both FH and Dempster’s conditionals), we use a ‘brute force’ approach to compute all the conditional beliefs. We then use the fast Möbius transform (FMT) [Thoma, 1989] to get the conditional masses for all the propositions.

- All conditional computations for an arbitrary proposition were done on an iMac running Mac OS X 10.12.3 (with 2.9GHz Intel Core i5 processor and 8GB of 1600MHz DDR3 RAM). Conditional computations for all propositions were done on the same iMac for smaller FoDs and on a supercomputer (https://ccs.miami.edu/pegasus) for larger FoDs (underlined in Table 1).
Experiments

<table>
<thead>
<tr>
<th>Method → Conditional</th>
<th>DS-Conditional-One Model FH or Dempster’s Specializa. Dempster’s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Bl(B</td>
</tr>
<tr>
<td>FoD</td>
<td>$Bl(B</td>
</tr>
<tr>
<td>Θ</td>
<td>Max.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
</tr>
<tr>
<td>8</td>
<td>255</td>
</tr>
<tr>
<td>10</td>
<td>1,023</td>
</tr>
<tr>
<td>12</td>
<td>4,095</td>
</tr>
<tr>
<td>14</td>
<td>16,383</td>
</tr>
<tr>
<td>16</td>
<td>65,535</td>
</tr>
<tr>
<td>18</td>
<td>262,143</td>
</tr>
<tr>
<td>20</td>
<td>1,048,575</td>
</tr>
</tbody>
</table>

Table: DS-Conditional-One model versus specialization matrix based method. Average computational times (ms). (*** denotes computations not completed within a feasible time).
Concluding Remarks

- The significant speed advantage offered by the proposed computational model over the specialization matrix based approach is evident from experiments.

- For larger FoDs, the computational burden associated with the specialization matrix based approach becomes prohibitive because of its space complexity of $\mathcal{O}(2^{2|\Theta|} \times 2^{2|\Theta|})$. For example, an FoD of size 20 would need 128 ($= 2^{20} \times 2^{20}/8$) GB of memory to represent the specialization matrix, if each matrix entry occupies only 1 bit.

- Our experiment results demonstrate that the average computational time taken to compute the conditional belief of an arbitrary proposition by the proposed approach is less than $2 \ (\mu s)$ for a FoD of size 10 and $0.7 \ (ms)$ for a FoD of size 20 ($\sim 1$ million focal elements).

- It also appears possible to further enhance the algorithms that we have developed via parallel computing optimizations because of the underlying matrix structure.
References

A new approach to updating beliefs.
In *Proc. 6th Conf. Uncertainty in Artificial Intelligence (UAI)*, pages 347–374, Cambridge, MA.

A framework for efficient computation of belief theoretic operations.
In *Proc. 19th Int. Conf. Information Fusion (FUSION)*, pages 1570–1577, Heidelberg, Germany.

*A Mathematical Theory of Evidence*.
Princeton Univ. Press, Princeton, NJ.

Perspectives on the theory and practice of belief functions.

The application of the matrix calculus to belief functions.

*Factorization of belief functions*.
Thank you very much!