

The CPT Structure of Variable Elimination in Discrete Bayesian Networks

C.J. Butz, W. Yan, P. Lingras, Y.Y. Yao

Abstract We show that a *conditional probability table* (CPT) is obtained after every multiplication and every marginalization step when eliminating variables from a discrete Bayesian network. The main advantage of our work is an improvement in presentation. The probability distributions constructed during variable elimination in Bayesian networks have always been denoted as potentials. Since CPTs are a special case of potential, our description is more precise and readable.

1 Introduction

A discrete *Bayesian network* [2, 3, 4, 5, 9, 10, 15] consists of a *directed acyclic graph* (DAG) and a corresponding set of *conditional probability tables* (CPTs). Bayesian networks serve as a clear and concise semantic modeling tool for managing uncertainty in complex domains by representing variables in the problem domain as vertices in the DAG, qualifying direct relationships between variables with directed arcs in the DAG, and quantifying these relationships with CPTs. The *probabilistic conditional independencies* [14] encoded in the DAG indicate that the product of the CPTs is a *joint probability distribution* (JPD). While a JPD over m binary variables would comprise 2^m probabilities, a Bayesian network models a JPD compactly and indirectly as a product of CPTs.

The task of *variable elimination* (VE) is central to reasoning with Bayesian networks. A variable v is eliminated using a simple two-step procedure. First, the probability distributions involving v are multiplied together. Second, v is marginalized

C.J. Butz, W. Yan and Y.Y. Yao

Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada, S4S 0A2,
e-mail: {butz, yanwe111, yyao}@cs.uregina.ca

P.Lingras

Department of Mathematics and Computing Science, Saint Mary's University, Halifax, Nova Scotia, Canada, B3H 3C3, e-mail: Pawan.Lingras@stmarys.ca

out of the product obtained in the first step. Other variables can be eliminated in a recursive manner. The probabilistic reasoning literature has always denoted the probability distributions constructed during VE as *potentials*. This description is not as precise as it should be.

In this chapter, we show that every multiplication operation and every marginalization operation involved in eliminating variables from a discrete Bayesian network yields a CPT. The concept of *expanded form* is introduced to define each potential constructed by the VE algorithm in terms of a sequence of multiplication and marginalization operators on the given Bayesian network CPTs. We then establish that each expanded form can be equivalently rewritten in *normal form*, that is, as the marginalization of a product of Bayesian network CPTs. By applying our key observation, it is established that every distribution constructed by VE is indeed a CPT.

The significance of this result resides in the description of the VE algorithm. The distributions constructed by VE have always been denoted as potentials. Potentials do not have clear physical interpretation [2], as they are unnormalized probability distributions [15]. In contrast, CPTs have clear semantic meaning [2], since the probabilities in the distribution must necessarily obey a specific pattern. Thereby, establishing that the distributions constructed by VE are, in fact, CPTs rather than potentials yield a description that is more precise and readable.

The chapter is organized as follows. Section 2 reviews Bayesian networks and the VE algorithm. That each multiplication and addition during VE yields a CPT is shown in Section 3. We then show the advantages of our work in Section 4. The conclusion is presented in Section 5.

2 Background Knowledge

Let $U = \{v_1, v_2, \dots, v_n\}$ be a finite set of variables. Each variable $v_i \in U$ has a finite domain, denoted $dom(v_i)$, representing the values v_i can assume. For a subset $X \subseteq U$, we write $dom(X)$ for the Cartesian product of the domains of the individual variables in X . Each element $x \in dom(X)$ is called a *configuration* of X . If c is a configuration on X and $Y \subseteq X$, then $c.Y$ denotes the restriction of c onto Y . As done in relational databases [7], we assume that there is a value λ such that $c.\emptyset = \lambda$ for any configuration c .

Definition 1. [4] A potential on $dom(X)$ is a function ϕ on $dom(X)$ such that $\phi(x) \geq 0$, for each configuration $x \in dom(X)$, and $\sum_{x \in dom(X)} \phi(x) > 0$.

For brevity, we refer to a potential as a distribution on X rather than $dom(X)$, and we call X , not $dom(X)$, its domain [13]. Also, for simplified notation, we use XY to denote $X \cup Y$, and may write $\{v_i\}$ as v_i in this chapter.

Example 1. Let a, b, c, d be four binary variables. Two potentials $\phi(a, b)$ and $\phi(c, d)$ are shown in Table 1.

Table 1 Two potentials $\phi(a,b)$ and $\phi(c,d)$.

a	b	$\phi(a,b)$	c	d	$\phi(c,d)$
0	0	0.2	0	0	1.6
0	1	0.8	0	1	0.0
1	0	0.0	1	0	0.5
1	1	1.0	1	1	0.3

Definition 2. The *unity-potential* $1(v_i)$ for a single variable v_i is a function 1 mapping every element of $\text{dom}(v_i)$ to one. More generally, the unity-potential $1(X)$ for a set $X = \{v_1, v_2, \dots, v_k\}$ of variables is defined as follows: $1(X) = 1(v_1) \cdot 1(v_2) \cdot \dots \cdot 1(v_k)$. That is, $1(X)$ is table on X , where the probability value is one for each row.

Note that $\phi(Y) = \phi(Y) \cdot 1(X)$, if $X \subseteq Y$.

Definition 3. A conditional probability table (CPT) on a set X of variables given a disjoint set Y of variables, denoted $p(X|Y)$, is a potential on the union of X and Y such that for each configuration $y \in \text{dom}(Y)$, $\sum_{x \in \text{dom}(X)} p(X = x|Y = y) = 1$.

Example 2. Let $a, b, c, d, e, f, g, h, i$ be binary variables. Table 2 shows CPTs $p(a)$, $p(b)$, $p(c|a)$, $p(d|a,b)$, $p(e|b)$, $p(f|c,d)$, $p(g|e,f)$, $p(h|e,g)$ and $p(i|f)$.

Table 2 CPTs $p(a)$, $p(b)$, $p(c|a)$, $p(d|a,b)$, $p(e|b)$, $p(f|c,d)$, $p(g|e,f)$, $p(h|e,g)$ and $p(i|f)$.

a	$p(a)$	b	$p(b)$	a	c	$p(c a)$	a	b	d	$p(d a,b)$	b	e	$p(e b)$	
0	0.496	0	0.423	0	0	0.123	0	0	0	0.408	0	0	0.437	
1	0.504	1	0.577	0	1	0.877	0	0	1	0.592	0	1	0.563	
				1	0	0.057	0	1	0	0.101	1	0	0.421	
				1	1	0.943	0	1	1	0.899	1	1	0.579	
							1	0	0	0.123				
							1	0	1	0.877				
							1	1	0	0.027				
							1	1	1	0.973				
c	d	f	$p(f c,d)$	e	f	g	$p(g e,f)$	e	g	h	$p(h e,g)$	f	i	$p(i f)$
0	0	0	1.0	0	0	0	0.739	0	0	0	0.562	0	0	0.739
0	0	1	0.0	0	0	1	0.261	0	0	1	0.438	0	1	0.261
0	1	0	1.0	0	1	0	0.278	0	1	0	0.406	1	0	0.498
0	1	1	0.0	0	1	1	0.722	0	1	1	0.594	1	1	0.502
1	0	0	1.0	1	0	0	0.567	1	0	0	0.421			
1	0	1	0.0	1	0	1	0.433	1	0	1	0.579			
1	1	0	0.0	1	1	0	0.303	1	1	0	0.353			
1	1	1	1.0	1	1	1	0.697	1	1	1	0.647			

Whenever $p(X|Y)$ is written with X and Y not disjoint, we mean $p(X|Y - X)$ to satisfy the disjointness condition of CPTs. In [8], three special cases of CPTs are

denoted as $p(X|\emptyset)$, $p(\emptyset|Y)$ and $p(\emptyset|\emptyset)$. However, in BN literature, they are more commonly written as $p(X)$, $1(Y)$ and 1, respectively.

Definition 4. Let $\phi_1(X)$ and $\phi_2(Y)$ be two potentials. The product of $\phi_1(X)$ and $\phi_2(Y)$, denoted $\phi_1 \cdot \phi_2$, is defined as: for each configuration c of XY ,

$$(\phi_1 \cdot \phi_2)(c) = \phi_1(c.X) \cdot \phi_2(c.Y).$$

Example 3. Table 3 (left) depicts the product $\phi_1(e, f, g, h)$ of the two CPTs $p(g|e, f)$ and $p(h|e, g)$ in Table 2, namely,

$$\phi_1(e, f, g, h) = p(g|e, f) \cdot p(h|e, g). \quad (1)$$

Table 3 (left) The product $\phi_1(e, f, g, h)$ of $p(g|e, f)$ and $p(h|e, g)$ in Table 2. (right) The marginalization of $\phi_1(e, f, g, h)$ onto $\{e, f, h\}$ yields $\phi_1(e, f, h)$.

e	f	g	h	$\phi_1(e, f, g, h)$	e	f	h	$\phi_1(e, f, h)$
0	0	0	0	0.415	0	0	0	0.521
0	0	0	1	0.324	0	0	1	0.479
0	0	1	0	0.106	0	1	0	0.449
0	0	1	1	0.155	0	1	1	0.551
0	1	0	0	0.156	1	0	0	0.392
0	1	0	1	0.122	1	0	1	0.608
0	1	1	0	0.293	1	1	0	0.374
0	1	1	1	0.429	1	1	1	0.626
1	0	0	0	0.239				
1	0	0	1	0.328				
1	0	1	0	0.153				
1	0	1	1	0.280				
1	1	0	0	0.128				
1	1	0	1	0.175				
1	1	1	0	0.246				
1	1	1	1	0.451				

The key concept of a Bayesian network can now be defined.

Definition 5. [10] A Bayesian network on $U = \{v_1, v_2, \dots, v_n\}$ is a pair (D, C) . D is a directed acyclic graph (DAG) on U . C is a set of CPTs $\{p(v_1|P_1), p(v_2|P_2), \dots, p(v_n|P_n)\}$ such that for each variable $v_i \in D$, there is a CPT $p(v_i|P_i)$ for v_i given its parents P_i in D .

One salient feature of Bayesian networks is that the product of the given CPTs is a joint distribution on the set U of variables.

Example 4. The *coronary heart disease* (CHD) [4, 12] Bayesian network, inspired by a research project in the field of epidemiology, is the DAG in Figure 1 on the set U of variables, along with the CPTs in Table 2. It follows that:

$$\begin{aligned} p(U) \\ = p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \cdot p(i|f). \end{aligned} \quad (2)$$

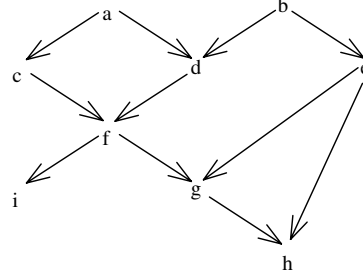


Fig. 1 The CHD Bayesian network consists of this DAG together with the CPTs in Table 2.

It is assumed in the literature that it is feasible to store the Bayesian network CPTs in computer memory, but not the joint distribution $p(U)$. The marginalization operator, denoted \sum , is used to further manipulate the stored CPTs.

Definition 6. Given a potential $\phi(Z)$, let $X \subseteq Z$ and $Y = Z - X$. The marginal of $\phi(Z)$ onto X , denoted $\phi(X)$, is defined as: for each configuration x of X ,

$$\phi(x) = \sum_{y \in \text{dom}(Y)} \phi(x, y),$$

where x, y is the configuration of Z obtained by combining x with the configuration y of Y .

Example 5. The marginalization of potential $\phi_1(e, f, g, h)$ in Table 3 (left) onto $\{e, f, h\}$ yields the potential $\phi_1(e, f, h)$ in Table 3 (right).

Lemmas 1 and 2 state two important properties that are used in Bayesian network reasoning.

Lemma 1. [13] *Let ϕ be a potential on Z , and $X \subseteq Y \subseteq Z$. Marginalizing ϕ onto Y and subsequently onto X is the same as marginalizing ϕ directly onto X .*

Lemma 1 states that variables can be eliminated in any order. While the order can affect the amount of computation performed [13], it has no bearing on the results of this chapter. The minimum deficiency search (MDS) [1, 6] algorithm, a technique for finding a good elimination ordering, is used in this chapter.

Lemma 2. [13] *If ϕ_1 is a potential on X and ϕ_2 is a potential on Y , then the marginalization of $\phi_1 \cdot \phi_2$ onto X is the same as ϕ_1 multiplied with the marginalization of ϕ_2 onto $X \cap Y$.*

Lemma 2 means that only the potentials involving the variable v being eliminated need be multiplied together before v is marginalized away. *Variable elimination* (VE), a central component in Bayesian network reasoning, is now defined.

Algorithm 1. [13] VE(C, X)

Input: C - a set of CPTs defining a Bayesian network on a set of variables U ,

X - a non-empty, proper subset of variables in U to be eliminated.

Output: F - the factorization obtained by marginalizing X from C .

begin

Let $F = C$.

while X is not empty

Let $X = X - \{v\}$.

Let F' be the set of potentials in F involving v .

Let $\phi(Y)$ be the product of the potentials in F' .

Let $\phi(Y - v)$ be obtained by marginalizing v out of $\phi(Y)$.

Let $F = (F - F') \cup \{\phi(Y - v)\}$.

end while

return F .

end

Example 6. Consider VE(C, X), where C are the CPTs in Table 2 defining the CHD Bayesian network and $X = \{a, c, d, e, g\}$ are the variables to be eliminated. The MDS algorithm suggests to use the elimination ordering g, c, a, d, e . Let F be a copy of C and let v be variable g . Those potentials involving g , namely, $F' = \{p(g|e, f), p(h|e, g)\}$ are collected. The product $\phi_1(e, f, g, h)$, shown in Table 3 (left), of the potentials in F' is obtained by Equation (1). Then $\phi_1(e, f, h)$, shown in Table 3 (right), is obtained by $\sum_g \phi_1(e, f, g, h)$. Next, the set F of potentials is updated as $F = (F - \{p(g|e, f), p(h|e, g)\}) \cup \{\phi_1(e, f, h)\}$. The remainder of the example follows in a similar manner.

Regarding Example 6, let us emphasize the distributions multiplied and marginalized in terms of equations:

$$\sum_{a,c,d,e,g} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \cdot p(i|f) \quad (3)$$

$$= \sum_{a,c,d,e} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(i|f) \cdot \sum_g p(g|e,f) \cdot p(h|e,g) \quad (4)$$

$$= \sum_{a,c,d,e} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(i|f) \cdot \sum_g \phi_1(e, f, g, h) \quad (5)$$

$$= \sum_{a,c,d,e} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(i|f) \cdot \phi_1(e,f,h) \quad (6)$$

$$= \sum_{a,d,e} p(a) \cdot p(b) \cdot p(d|a,b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \sum_c p(c|a) \cdot p(f|c,d) \quad (7)$$

$$= \sum_{a,d,e} p(a) \cdot p(b) \cdot p(d|a,b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \sum_c \phi_2(a,c,d,f) \quad (8)$$

$$= \sum_{a,d,e} p(a) \cdot p(b) \cdot p(d|a,b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \phi_2(a,d,f) \quad (9)$$

$$= \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \sum_a p(a) \cdot p(d|a,b) \cdot \phi_2(a,d,f) \quad (10)$$

$$= \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \sum_a \phi_3(a,b,d) \cdot \phi_2(a,d,f) \quad (11)$$

$$= \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \sum_a \phi_4(a,b,d,f) \quad (12)$$

$$= \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \phi_4(b,d,f) \quad (13)$$

$$= \sum_e p(b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \sum_d \phi_4(b,d,f) \quad (14)$$

$$= \sum_e p(b) \cdot p(e|b) \cdot p(i|f) \cdot \phi_1(e,f,h) \cdot \phi_4(b,f) \quad (15)$$

$$= p(b) \cdot p(i|f) \cdot \phi_4(b,f) \cdot \sum_e p(e|b) \cdot \phi_1(e,f,h) \quad (16)$$

$$= p(b) \cdot p(i|f) \cdot \phi_4(b,f) \cdot \sum_e \phi_5(b,e,f,h) \quad (17)$$

$$= p(b) \cdot p(i|f) \cdot \phi_4(b,f) \cdot \phi_5(b,f,h). \quad (18)$$

3 The CPT Structure of Variable Elimination

Here we prove that a CPT is obtained after every multiplication operation and every marginalization operation when applying the VE algorithm on a discrete Bayesian network. A key observation is that the product of any non-empty subset of Bayesian network CPTs is itself a CPT. Our claim is then shown by rewriting the factorization to exploit our key observation.

First, one salient feature of Bayesian networks is shown, namely, that the product of the given CPTs is a joint probability distribution. For example, in the CHD Bayesian network, we have:

$$\begin{aligned} & p(a,b,c,d,e,f,g,h,i) \\ &= p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \cdot p(i|f). \end{aligned} \quad (19)$$

Equation (19) can be verified by showing that the following equation equals one (1):

$$\sum_U p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \cdot p(i|f).$$

Let \prec be a *topological ordering* [11] of the variables $\{a,b,c,d,e,f,g,h,i\}$ in the CHD DAG, say $b \prec a \prec c \prec d \prec e \prec f \prec g \prec h \prec i$. By marginalizing the variables in reverse order of \prec , the variable v_i being marginalized only appears in one CPT, $p(v_i|P_i)$. By the definition of CPT, $\sum_{v_i} p(v_i|P_i) = 1(P_i)$. The claim follows. For example, the above equation can be rewritten as

$$\begin{aligned} & \sum_{U-i} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \\ & \cdot \sum_i p(i|f) \\ = & \sum_{U-i} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \\ & \cdot 1(f) \\ = & \sum_{U-i} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g). \end{aligned}$$

The remaining variables can be removed similarly, thereby establishing that the product of all CPTs in Table 2 is a joint probability distribution. This well-known proof is a special case of a more general result.

Lemma 3. Consider a Bayesian network on variables $U = \{v_1, v_2, \dots, v_n\}$ with DAG D and CPTs $C = \{p(v_1|P_1), p(v_2|P_2), \dots, p(v_n|P_n)\}$. Let $C' = \{p(v_i|P_i), p(v_j|P_j), \dots, p(v_l|P_l), p(v_m|P_m)\}$ be any non-empty subset of C . The product of the CPTs in C' is a CPT of the variables X given Y , where $X = \{v_i, v_j, \dots, v_l, v_m\}$ and $Y = (P_i P_j \dots P_l P_m) - X$.

Proof. Let $C' = \{p(v_i|P_i), p(v_j|P_j), \dots, p(v_l|P_l), p(v_m|P_m)\}$. Similar to the proof that the product of all Bayesian network CPTs is a joint probability distribution, we show that the product of the Bayesian network CPTs in C' is a CPT by establishing

$$\sum_X p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot p(v_m|P_m) = 1(Y). \quad (20)$$

Let \prec denote a topological ordering of the variables in D . Without loss of generality, let $v_i \prec v_j \prec \dots \prec v_l \prec v_m$. This \prec and the fact that D is a DAG mean that v_m can only appear in one CPT of C' , namely, $p(v_m|P_m)$. Thereby, we have

$$\begin{aligned} & \sum_X p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot p(v_m|P_m) \\ = & \sum_{X-v_m} p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot \sum_{v_m} p(v_m|P_m) \\ = & \sum_{X-v_m} p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot 1(P_m) \\ = & \sum_{X-v_m} p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot 1(P_m - X) \cdot 1(P_m \cap X) \end{aligned}$$

$$= 1(P_m - X) \cdot \sum_{X - v_m} p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot 1(P_m \cap X). \quad (21)$$

Since $v_m \notin P_m$, by definition, $(P_m \cap X) \subseteq X - v_m$. That is, $(P_m \cap X) \subseteq \{v_i, v_j, \dots, v_l\}$. Thereby, $1(P_m \cap X)$ can be factorized into unity-potentials $1(v)$ on its singleton variables v , and each $1(v)$ can be multiplied with the CPT $p(v|P_v)$. As $p(v|P_v) \cdot 1(v)$ gives $p(v|P_v)$, Equation (21) can be rewritten as

$$1(P_m - X) \cdot \sum_{X - v_m} p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l).$$

By a similar argument for variables v_l, \dots, v_j, v_i , we obtain our desired result:

$$\begin{aligned} & \sum_X p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot p(v_m|P_m) \\ &= 1(P_i - X) \cdot 1(P_j - X) \cdot \dots \cdot 1(P_l - X) \cdot 1(P_m - X) \\ &= 1(Y). \quad \square \end{aligned}$$

Lemma 3 establishes that the product of a subset of CPTs from a Bayesian network is a CPT. While it is not guaranteed that the product is $p(X|Y)$, namely, a CPT defined with respect to the joint distribution p , we make this assumption in this chapter.

Example 7. Consider the CPTs $\{p(c|a), p(e|b), p(f|c, d)\}$, which are a subset of the CHD Bayesian network of Figure 1. By Lemma 3,

$$p(c, e, f|a, b, d) = p(c|a) \cdot p(e|b) \cdot p(f|c, d). \quad (22)$$

The notion of *expanded form* is introduced to express potentials built by VE equivalently in terms of multiplication and marginalization operators on a subset of Bayesian network CPTs.

Definition 7. Let ϕ be any potential constructed in a given instance of $VE(C, X)$. The expanded form of ϕ is the unique expression defining how VE built ϕ using the multiplication and marginalization operators on the Bayesian network CPTs in C .

Example 8. The expanded form of potential $\phi_4(b, f)$ in Equation (15) is:

$$\phi_4(b, f) = \sum_d \left(\sum_a \left((p(a) \cdot p(d|a, b)) \cdot \left(\sum_c (p(c|a) \cdot p(f|c, d)) \right) \right) \right),$$

which is determined recursively as follows:

$$\begin{aligned} \phi_4(b, f) &= \sum_d \phi_4(b, d, f) \\ &= \sum_d \left(\sum_a \phi_4(a, b, d, f) \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_d \left(\sum_a (\phi_3(a, b, d) \cdot \phi_2(a, d, f)) \right) \\
&= \sum_d \left(\sum_a ((p(a) \cdot p(d|a, b)) \cdot \phi_2(a, d, f)) \right) \\
&= \sum_d \left(\sum_a \left((p(a) \cdot p(d|a, b)) \cdot \left(\sum_c \phi_2(a, c, d, f) \right) \right) \right) \\
&= \sum_d \left(\sum_a \left((p(a) \cdot p(d|a, b)) \cdot \left(\sum_c (p(c|a) \cdot p(f|c, d)) \right) \right) \right).
\end{aligned}$$

Definition 8. The expanded form of a potential ϕ constructed in a given instance of $VE(C, X)$ is said to be in normal form, if all marginalizations take place on the product of all CPTs used to build ϕ .

Example 9. The expanded form of potential $\phi_4(b, f)$ in Example 8 is not in normal form, since, for instance, the marginalization of variable c takes place on a product not involving the CPTs $p(a)$ and $p(d|a, b)$.

The next result is critical to applying our key observation.

Lemma 4. *The expanded form of any potential ϕ constructed in a given instance of $VE(C, X)$ can always be equivalently rewritten in normal form.*

Proof. Let \sum_{v_i} be any marginalization in the expanded form. There are two cases to consider. First, consider the case when there is another marginalization \sum_{v_j} to the immediate left of \sum_{v_i} . By Lemma 1, we can equivalently rewrite $\sum_{v_j} \sum_{v_i}$ as $\sum_{v_i} \sum_{v_j}$. Otherwise, consider the case when a multiplication operator appears to the immediate left of \sum_{v_i} , say $\phi_1 \cdot \sum_{v_i} \phi_2$. By construction of VE, all distributions involving v_i are multiplied together as ϕ_2 before v_i is marginalized out. This means that v_i does not appear in ϕ_1 . By Lemma 2, $\phi_1 \cdot \sum_{v_i} \phi_2$ can be equivalently rewritten as $\sum_{v_i} \phi_1 \cdot \phi_2$. By repeated argument, \sum_{v_i} can be pulled to the left of all multiplication operators in the expanded form. This argument holds for all other marginalization signs. By definition, the expanded form can be equivalently rewritten into normal form. \square

Example 10. In Example 8, the expanded form of potential $\phi_4(b, f)$ can be equivalently rewritten in normal form as follows:

$$\begin{aligned}
&\sum_d \left(\sum_a \left((p(a) \cdot p(d|a, b)) \cdot \left(\sum_c (p(c|a) \cdot p(f|c, d)) \right) \right) \right) \\
&= \sum_d \left(\sum_a \left(\sum_c ((p(a) \cdot p(d|a, b)) \cdot (p(c|a) \cdot p(f|c, d))) \right) \right) \\
&= \sum_{d, a, c} p(a) \cdot p(d|a, b) \cdot p(c|a) \cdot p(f|c, d).
\end{aligned}$$

The main result of this chapter is given next.

Theorem 1. *Every multiplication step and every marginalization step in $VE(C,X)$ yields a CPT.*

Proof. Let ϕ be any potential built during an instance of $VE(C,X)$. By Definition 7, the expanded form of ϕ is the unique expression defining how VE built ϕ using the multiplication and marginalization operators on the Bayesian network CPTs in C . By Lemma 4, the expanded form can be equivalently rewritten in normal form, say:

$$\sum_{X'} p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot p(v_m|P_m).$$

Lemma 3 establishes that

$$p(v_i, v_j, \dots, v_l, v_m | P_i P_j \dots P_l P_m) = p(v_i|P_i) \cdot p(v_j|P_j) \cdot \dots \cdot p(v_l|P_l) \cdot p(v_m|P_m).$$

It follows that X' is a subset of $\{v_i, v_j, \dots, v_l, v_m\}$, since the VE algorithm requires that all distributions involving the variable being eliminated be multiplied together before the variable is marginalized away. Thus, by definition of CPT,

$$\sum_{X'} p(v_i, v_j, \dots, v_l, v_m | P_i P_j \dots P_l P_m)$$

yields a CPT. Therefore, any potential built by the VE algorithm is a CPT. \square

Example 11. Recall the elimination of variables a, c, d, e and g from the Bayesian network in Example 4. By Theorem 1, we now have:

$$\begin{aligned} & \sum_{a,c,d,e,g} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(g|e,f) \cdot p(h|e,g) \\ & \cdot p(i|f) \\ = & \sum_{a,c,d,e} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(i|f) \cdot \sum_g p(g|e,f) \\ & \cdot p(h|e,g) \\ = & \sum_{a,c,d,e} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(i|f) \cdot \sum_g p(g,h|e,f) \\ = & \sum_{a,c,d,e} p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|a,b) \cdot p(e|b) \cdot p(f|c,d) \cdot p(i|f) \cdot p(h|e,f) \\ = & \sum_{a,d,e} p(a) \cdot p(b) \cdot p(d|a,b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot \sum_c p(c|a) \cdot p(f|c,d) \\ = & \sum_{a,d,e} p(a) \cdot p(b) \cdot p(d|a,b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot \sum_c p(c,f|a,d) \\ = & \sum_{a,d,e} p(a) \cdot p(b) \cdot p(d|a,b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot p(f|a,d) \\ = & \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot \sum_a p(a) \cdot p(d|a,b) \cdot p(f|a,d) \\ = & \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot \sum_a p(a,d|b) \cdot p(f|a,d) \end{aligned}$$

$$\begin{aligned}
&= \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot \sum_a p(a,d,f|b) \\
&= \sum_{d,e} p(b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot p(d,f|b) \\
&= \sum_e p(b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot \sum_d p(d,f|b) \\
&= \sum_e p(b) \cdot p(e|b) \cdot p(i|f) \cdot p(h|e,f) \cdot p(f|b) \\
&= p(b) \cdot p(i|f) \cdot p(f|b) \cdot \sum_e p(e|b) \cdot p(h|e,f) \\
&= p(b) \cdot p(i|f) \cdot p(f|b) \cdot \sum_e p(e,h|b,f) \\
&= p(b) \cdot p(i|f) \cdot p(f|b) \cdot p(h|b,f).
\end{aligned}$$

The significance of Theorem 1 is that the potentials $\phi_1(e, f, g, h)$, $\phi_1(e, f, h)$, $\phi_2(a, c, d, f)$, $\phi_2(a, d, f)$, $\phi_3(a, b, d)$, $\phi_4(a, b, d, f)$, $\phi_4(b, d, f)$, $\phi_5(b, e, f, h)$ and $\phi_5(b, f, h)$ constructed in Equations (3) - (18) of Example 6 are actually the CPTs $p(g, h|e, f)$, $p(h|e, f)$, $p(c, f|a, d)$, $p(f|a, d)$, $p(a, d|b)$, $p(a, d, f|b)$, $p(d, f|b)$, $p(e, h|b, f)$ and $p(h|b, f)$, respectively.

4 Advantages

By definition, a CPT is a special case of potential. For instance, all of the CPTs in Table 2 are potentials. On the contrary, not all potentials are CPTs. For example, using $\phi(a, b)$ in Table 1, it can be verified that

$$\sum_b \phi(a, b) = 1(a). \quad (23)$$

Therefore, the potential $\phi(a, b)$ is, in fact, a CPT $p(b|a)$. In contrast, the potential $\phi(c, d)$ in Table 1 cannot satisfy the definition of CPT. Nevertheless, it is more precise to label the distributions constructed by VE as CPTs rather than as potentials.

Labeling distributions as CPTs bring more clarity to the VE algorithm. Potentials do not have clear physical interpretation [2] as they are unnormalized probability distributions [15]. In contrast, CPTs have clear semantic meaning [2], since the probabilities must satisfy a rigorous pattern.

5 Conclusion

In this chapter, we considered the problem of eliminating a variable from a Bayesian network in the context of no evidence. Our main result is that each multiplication and addition in the VE algorithm generates a CPT, not a potential. Our key observa-

tion is that the product of any subset of Bayesian network CPTs is itself a CPT. The main advantage of our work is an improvement in presentation. Potentials do not have clear physical interpretation [2] as they are unnormalized probability distributions [15]. In contrast, CPTs have clear semantic meaning [2], since the probabilities must satisfy a rigorous pattern. While the distributions constructed during VE have traditionally been called potentials, we have shown that they are, in fact, CPTs.

References

1. U. Bertelè and F. Brioschi, *Nonserial Dynamic Programming*, Mathematics in Science and Engineering. Academic Press, (91) (1972).
2. E. Castillo, J. Gutiérrez, A. Hadi, *Expert Systems and Probabilistic Network Models*, Springer, New York, 1997.
3. R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Springer, New York, 1999.
4. P. Hájek, T. Havránek, R. Jiroušek, *Uncertain Information Processing in Expert Systems*, CRC Press, Ann Arbor, 1992.
5. F.V. Jensen, *An Introduction to Bayesian Networks*, UCL Press, London, 1996.
6. U. Kjaerulff. *Triangulation of Graphs - Algorithms Giving Small Total State Space*. Technical Report R 90-09, Department of Mathematics and Computer Science, Strandvejen, DK 9000 Aalborg, Denmark, 1990.
7. D. Maier, *The Theory of Relational Databases*, Computer Science Press, 1983.
8. R.E. Neapolitan, *Learning Bayesian Networks*, Prentice Hall, 2003.
9. R.E. Neapolitan, *Probabilistic Reasoning in Expert Systems*, John Wiley & Sons, Toronto, 1990.
10. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, 1988.
11. F. Preparata, R.T. Yeh, *Introduction to Discrete Structures*, Addison-Wesley, Don Mills, 1973.
12. Z. Reiniš, J. Pokorný, V. Basika, J. Tiserord, K. Goničan, D. Horáková, T. Havránek, F. Hrabovský, and E. Stuchlíková, Prognostic significance of the risk profile in the prevention of coronary heart disease, *Bratisl. Lék. Listy*, 76, 137, 1981.
13. G. Shafer, *Probabilistic Expert Systems*, SIAM, Philadelphia, 1996.
14. S.K.M. Wong, C.J. Butz, D. Wu, On the implication problem for probabilistic conditional independency, *IEEE Trans. Syst. Man Cybern. A* 30 (6) (2000) 785-805.
15. Y. Xiang, *Probabilistic Reasoning in Multiagent Systems: A graphical models approach*, Cambridge University Press, New York, 2002.