Representation of Interrelationships among Binary Variables under Dempster-Shafer Theory of Belief Functions

Rajendra P. Srivastava*
Ernst & Young Distinguished Professor and Director
E&Y CARAT, School of Business, The University of Kansas
1300 Sunnyside Avenue, Lawrence, KS 66045
Phone: 785-864-7590, Fax: 785-864-5328
Email: rsrivastava@ku.edu

Lei Gao
Assistant Professor
School of Accountancy, University of Nebraska-Lincoln
P.O.Box 880488, Lincoln NE 68588-0488
Phone: 402-472-2902, Fax: 402-472-4100
Email: lgao3@unl.edu

Peter R. Gillett
Associate Professor
Accounting and Information Systems Area
Faculty of Management
Rutgers University-Newark
Newark, NJ 07102-1895
Email: gillett@everest.rutgers.edu

March 18, 2008

---

*Corresponding author

Representation of Interrelationships among Binary Variables under Dempster-Shafer Theory of Belief Functions

ABSTRACT

This paper presents an algorithm for developing models under Dempster-Shafer theory of belief functions for categorical and 'uncertain' logical relationships among binary variables. We illustrate the use of the algorithm by developing belief-function representations of the following categorical relationships: 'AND', 'OR', 'Exclusive OR (EOR)' and 'Not Exclusive OR (NEOR)', and 'AND-NEOR' and of the following uncertain relationships: 'Discounted AND', 'Conditional OR', and 'Weighted Average'. Such representations are needed to fully model and analyze a problem with a network of interrelated variables under Dempster-Shafer theory of belief functions. In addition, we compare our belief-function representation of the 'Weighted Average' relationship with the 'Weighted Average' representation developed and used by Shenoy and Shenoy[8]. We find that Shenoy and Shenoy representation of the weighted average relationship is an approximation and yields significantly different values under certain conditions.

Key words: Interrelationships, Logical Relationships, Uncertain relationship, Belief Functions, Dempster-Shafer theory of Belief Functions

Representation of Interrelationships among Binary Variables under Dempster-Shafer Theory of Belief Functions

1. INTRODUCTION

This article has three primary objectives. First objective is to describe an algorithm to develop belief-function[1] representations of interrelationships among binary variables. The second objective is to demonstrate the use of the algorithm for developing belief-function representations of simple categorical relationships such as 'AND', 'OR', 'Exclusive OR', 'Not Exclusive OR', etc. The third objective is to use the algorithm to develop belief-function representations of uncertain relationships such as 'Discounted AND', 'Conditional OR' and 'Weighted Average'. These uncertain relationships are described later in Section 4. Belief-function representations of such interrelationships are essential to model and analyze "real world" problems with multiple variables under Dempster-Shafer theory of belief functions (hereafter referred to as DS theory) using software such as Auditor Assistant developed by Shafer et al.[2] In the present article, we assume that readers are familiar with DS theory and Dempster's rule of combination. Although Shafer[1] remains the classic reference for DS theory, there are more recent publications such as Yager et al.[3] and Srivastava and Mock[4] that provide a good background on DS theory. For an illustration of Dempster's rule, one can see Sun et al.[5]

In a decision problem under DS theory involving multiple interrelated variables, one usually develops an evidential diagram and propagates uncertainties through the network of variables using Shenoy and Shafer[6] approach of local computations. Usually, such networks tend to be quite complex and propagating uncertainties through such networks is not an easy task without the use of software such as Auditor Assistant developed by Shafer et al.[2] Srivastava and Mock[7] use such an approach to analyze WebTrust assurance services. Shenoy and Shenoy[8] use

this approach to model and analyze a financial portfolio. Srivastava and his co-researchers have used evidential reasoning approach to several other business decisions such as information systems security risk assessment[5], audit risk assessment[9], fraud risk assessment[10], information quality assessment[11]. Modeling of uncertain relationships among the variables in a network for a "real world" problem has been a challenge. For example, Shenoy and Shenoy[8] while modeling financial portfolio using belief functions recognized the need of 'weighted average' relationship among the variables of interest but because of the unavailability of a belief-function representation of such a relationship, they used what we would call an 'approximate weighted average' relationship in their analysis. As we will show in Section 4, such an approximate relationship may lead to undesirable consequences. As demonstrated in this article, the algorithm presented here facilitates the development of belief-function representations of categorical and uncertain relationships among binary variables, which would eventually help build realistic models of complex "real world" problems.

The remaining part of the paper is divided into four sections. Section 2 provides the background research relevant to the paper. Section 3 describes the algorithm and demonstrates the use of the algorithm in modeling categorical relationships. Section 4 develops the belief-function representations of the following uncertain relationships: 'Discounted AND', 'Conditional OR', and 'Weighted Average'. Finally, Section 5 provides a conclusion.

2. BACKGROUND RESEARCH

Most of the interrelationships used under DS theory in evidential networks for propagating beliefs have been categorical relationships such as 'AND', and 'OR'. For example, Gillett[12], Srivastava[13], and Srivastava et al[10, 14, 15] have used simple categorical relationships. However, Gillett[12] first time introduced the 'Discounted And' in an accounting situation where

two binary variables, say X_1 and X_2, are related to a third binary variable, say Z, in such a way that belief masses propagating from X_1 and X_2 to Z follow the 'AND' relationship but when the belief masses are propagated from Z to X_1 and X_2, they are discounted (see Figure 1). Such situations are very common in the accounting domain as elaborated later (see also Gillett[12] for a detailed discussion).

----- Figure 1 here -----

Gao[16] introduced 'Conditional OR' in order to propagate frequency information through an evidential network while applying DS theory in assessing fraud risk in financial statements. As mentioned earlier, having such a relationship modeled under DS theory makes it much easier to use software like "Auditor Assistant" developed by Shafer et al.[2] to integrate prior knowledge of relationships among variables into an evidential network.

Shenoy and Shenoy[8] and Sun et al.[5] have modeled 'Weighted Average' relationship but only in an approximate way. They use the 'discounting' method to express the 'weighted average' relationship under DS theory. However, their approach fails to represent the weighted average relationship correctly. We discuss this problem in detail in Section 4.

3. THE ALGORITHM TO MODEL RELATIONSHIPS

In this section, we describe an algorithm to model various interrelationships among binary variables for propagating beliefs in a network of variables. The current approach is based on an algorithm developed by Srivastava and Cogger[17] for converting belief masses (m-values) on individual variables from a single source of evidence to belief masses (m-values) on the joint space of the variables on which the evidence pertains. We describe the algorithm in the following form and illustrate its applications for modeling the categorical relationships such as 'AND', 'OR', 'Exclusive OR' etc. in the present section. We demonstrate its application for

4

modeling uncertain relationships such as 'Discounted AND', 'Conditional OR' and 'Weighted Average' in the next section.

3.1. Algorithm for Modeling Relationships Among Variables under DS Theory

Step 1: Create a table with the top row as the values of the states that are given, and the first column with the values of the outcome states. Fill the body of the table with the m-values for each outcome given the values of the states (See Table 1 for 'AND' relationship as an illustration).

Step 2: Select the smallest non-zero m-value from each column. These values are written inside rectangular boxes in our examples. Select the smallest m-values from this set of chosen m-values from each column. This m-value will become the value for the focal element created in Step 3 (See Table 1).

Step 3: The union of the set of elements identified in Step 2 in each column becomes the focal elements for the m-value chosen in Step 2 (See Table 1).

Step 4: Subtract the smallest m-value obtained in Step 3 from each selected m-value in Step 2.

Step 5: Repeat Steps 2 - 4 until all entries are zero.

3.2. Logical 'AND'

Assume that we have three binary variables: X_1, X_2 and Z with the corresponding lower case letters representing their values that the variables are true and the symbol '~' in front of the lower case letters representing the negation of the variables. For example, 'x_1' represents the value or state that variable X_1 is true, and '$\sim x_1$' represents the value or state that variable X_1 is not true. An 'AND' relationship between Z, and X_1 and X_2 implies that Z is true if and only if X_1 and X_2 are true (i.e., $Z = X_1 \cap X_2$). Such a relationship will allow only the following set of values for the variables: $\{zx_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}$. One can express the above relationship through the following conditional probabilities:

$P(z|x_1x_2) = 1, P(\sim z|x_1x_2) = 0,$

$P(z|x_1\sim x_2) = 0, P(\sim z|x_1\sim x_2) = 1,$

$P(z|{\sim}x_1 x_2) = 0, P({\sim}z|{\sim}x_1 x_2) = 1,$

$P(z|{\sim}x_1 {\sim}x_2) = 0, P({\sim}z|{\sim}x_1 {\sim}x_2) = 1.$

In the belief-function framework, the above relationship is represented by[14]:

$$m(\{zx_1 x_2, {\sim}zx_1 {\sim}x_2, {\sim}z{\sim}x_1 x_2, {\sim}z{\sim}x_1 {\sim}x_2\}) = 1.$$

Let us apply the algorithm described above to model the 'AND' relationship as represented above. Step 1 yields Table 1 with the first row being the values of the joint space resulting from the Cartesian product of the state spaces of X_1 and X_2 and the first column being the state space of Z. The body of the table represents the belief masses for 'z' or '~z' given that the state space listed in the top row of the corresponding column is true. Step 2 selects '1.0' in each column. Step 3 determines the belief mass, i.e., m-value for the focal element formed by the union of all the states identified in Step 2. Step 4 yields zero for all the elements in each column. This process yield the following belief -function representation of 'AND' relationship: $m(\{zx_1 x_2, {\sim}zx_1 {\sim}x_2, {\sim}z{\sim}x_1 x_2, {\sim}z{\sim}x_1 {\sim}x_2\}) = 1.0$. These results are given in Table 1.

Table 1: 'AND' Relationship

Z	$x_1 x_2$	$x_1 {\sim}x_2$	${\sim}x_1 x_2$	${\sim}x_1 {\sim}x_2$
z	1.0	0.0	0.0	0.0
~z	0.0	1.0	1.0	1.0
	$m(\{zx_1 x_2, {\sim}zx_1 {\sim}x_2, {\sim}z{\sim}x_1 x_2, {\sim}z{\sim}x_1 {\sim}x_2\}) = 1.0$			

3.3. Logical 'OR', 'Exclusive OR (EOR)' and 'Not Exclusive OR (NEOR)'

Let us use the algorithm to model an 'OR' relationship between Z, and X_1 and X_2 (i.e., $Z = X_1 \cup X_2$). In this case, we will have the following possible values: $\{zx_1 x_2, zx_1 {\sim}x_2, z{\sim}x_1 x_2, {\sim}z{\sim}x_1 {\sim}x_2\}$. This relationship implies that Z is true when either both X_1 and X_2 are true or when

6

any one of them is true but it is false when both X_1 and X_2 are false. In terms of probabilities, we can express the above relationship as:

$P(z|x_1 x_2) = 1, P(\sim z|x_1 x_2) = 0,$

$P(z|x_1 \sim x_2) = 1, P(\sim z|x_1 \sim x_2) = 0,$

$P(z|\sim x_1 x_2) = 1, P(\sim z|\sim x_1 x_2) = 0,$

$P(z|\sim x_1 \sim x_2) = 0, P(\sim z|\sim x_1 \sim x_2) = 1.$

Applying the algorithm described earlier for this case, we obtain the results as depicted in Table 2. As one can see, similar to the previous case of 'AND' relationship, Step 3 yields the following representation of 'OR' relationship: $m(\{zx_1 x_2, zx_1 \sim x_2, z\sim x_1 x_2, \sim z\sim x_1 \sim x_2\}) = 1.0$.

Table 2: 'OR' Relationship

Z	$x_1 x_2$	$x_1 \sim x_2$	$\sim x_1 x_2$	$\sim x_1 \sim x_2$
z	1.0	1.0	1.0	0.0
~z	0.0	0.0	0.0	1.0
	$m(\{zx_1 x_2, zx_1 \sim x_2, z\sim x_1 x_2, \sim z\sim x_1 \sim x_2\}) = 1.0$			

Similarly, we can determine the belief-function representation of 'Exclusive OR (EOR)' or its negation (NEOR). An 'Exclusive OR (EOR)' relationship between Z, and X_1 and X_2 implies that Z is true only when either X_1 is true or X_2 is true but it is false when both X_1 and X_2 are either false or true. Such a relationship will allow only the following set of values on the joint space: $\{\sim zx_1 x_2, zx_1 \sim x_2, z\sim x_1 x_2, \sim z\sim x_1 \sim x_2\}$. Extending the reasoning of the previous models for 'AND' and 'OR', we can express the belief-function representation of 'EOR' as follows:

$$m(\{\sim zx_1 x_2, zx_1 \sim x_2, z\sim x_1 x_2, \sim z\sim x_1 \sim x_2\}) = 1.0.$$

Similarly one can express the belief-function representation of 'Not Exclusive OR (NEOR)' for the three variables, X_1, X_2, and Z, where Z is true only when both X_1 and X_2 are either true or false, and Z is false when either X_1 is true and X_2 is false or X_1 is false and X_2 is true as: $m(\{zx_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1x_2, z\sim x_1\sim x_2\}) = 1.0$. This relationship allows only the following possible set of values on the joint space: $\{zx_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1x_2, z\sim x_1\sim x_2\}$.

3.4. Algebraic Relationship 'AND-NEOR'

Srivastava and Lu[18] define this new logical relationship 'AND-NEOR' for the situation where three variables X_1, X_2, and Z are related through an algebraic relationship such as: $Z = X_1 \pm X_2$. In this case, Z is true when both X_1 and X_2 are true, and it is false when any one of them (X_1 or X_2) is false but the other one is true. However, when both X_1 and X_2 are false, we do not know whether Z is true or false because there could be off-setting errors that might make Z true. This relationship is symmetric in all the three variables, as it should be, because they have the same relationship: $X_1 = X_2 \pm Z$, or $X_2 = X_1 \pm Z$. Moreover, simply knowing that Z is true, we cannot say anything about the state of X_1 or X_2 because there could be infinite different wrong values of X_1 and X_2 that can make Z true. Such a relationship will only allow the following set of values: $\{zx_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1x_2, z\sim x_1\sim x_2, \sim z\sim x_1\sim x_2\}$. As one can see, these values are symmetric in all the three variables. Also, one can see that the above set of values are the union of the values for 'AND' and 'NEOR' relationships. The belief function representation of the above relationship can be obtained by applying the algorithm to the above case which yields the following representation:

$$m(\{zx_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1x_2, z\sim x_1\sim x_2, \sim z\sim x_1\sim x_2\}) = 1.0.$$

The above relationship allows us to propagate beliefs in a network of variables with linear algebraic relationships as done by Srivastava and Lu[18]. This type of situation is quite

common in accounting. For example, the accounts receivable balance at the end of a fiscal year is equal to the beginning balance of the account plus the sales on credit minus the cash receipts during the fiscal year. The auditor combines the evidence gathered for sales on credit and for cash receipts during the period with the direct evidence for the ending balance of the accounts receivable from confirmations. 'AND-NEOR' will help us combine these items of evidence in an evidential network containing accounts receivable as one variable, and cash receipts for the period and sales on credit for the period being the other two variables.

All the above cases are examples of the situation where the entire frame of the joint space is assigned a belief mass of 1.0. Shenoy and Shenoy[8] call such a relationship a *deterministic* relationship. Obtaining belief-function representations for such relationships is an easy task; just assign a belief mass of 1.0 to the entire frame of the joint space. However, it is a challenge to determine a belief-function representation for uncertain relationships. In the next section, we will show how one can obtain such representations using the above algorithm.

4. MODELING UNCERTAIN RELATIONSHIPS

4.1. 'Discounted AND' Relationship

Gillett[12] has argued for the use of 'Discounted AND' in place of the logical 'AND' relationship as used in the auditing literature between transaction streams and account balances. For example, the auditing literature assumes that the Accounts Receivables balance is related to Sales and Cash Receipts for the period through the 'AND' relationship[19], implying that Accounts Receivable balance (AR) is fairly stated if and only if Sales (S) and Cash Receipts (CR) for the period are fairly stated. This relationship implies that if AR is fairly stated then both S and CR for the period may be misstated but because of the off-setting errors AR would be fairly stated. Thus, the assumed 'AND' relationship among these accounts is not appropriate. Gillett[12] defines

a "Discounting AND" relationship between such accounts as AR, S, and CR, which means that when the information in terms of belief masses, i.e., m-values, is being propagated from S and CR to AR then it will follow 'AND' relationship but if the information in terms of m-values is being propagated from AR to S and CR it will be discounted.

Let us consider a 'Discounted AND' relationship between the following three binary variables X_1, X_2 and Z such that Z is true if X_1 is true and X_2 is true. Also, Z is true 30% percent of the times if both X_1 and X_2 are not true because of the off-setting errors. However, Z is not true when either X_1 is not true but X_2 is true or X_1 is true but X_2 is not true. Also, Z is not true 70% of the times if both X_1 and X_2 are not true. Applying the algorithm described above, we get Table 3 and the corresponding m-value for the first focal element after performing Steps 1-3.

Table 3: 'Discounted AND' Relationship between Z, and X_1 and X_2 where Z is true 30 percent of the times even if both X_1 and X_2 are not true.

Z	$x_1 x_2$	$x_1 {\sim} x_2$	${\sim} x_1 x_2$	${\sim} x_1 {\sim} x_2$
z	1.0	0.0	0.0	0.3
~z	0.0	1.0	1.0	0.7
$m(\{zx_1x_2, {\sim}zx_1{\sim}x_2, {\sim}z{\sim}x_1x_2, z{\sim}x_1{\sim}x_2\}) = 0.3$				

Next, according to Step 4, we subtract the smallest m-value, 0.3 in the present case, from the chosen m-values in each of the other columns. The new values are given in Table 4 below. We repeat Steps 2 and 3 and obtain the next focal element and the corresponding m-value as given in Table 4.

Table 4: Step 4 for 'Discounted AND' Relationship

Z	x_1x_2	$x_1\sim x_2$	$\sim x_1x_2$	$\sim x_1\sim x_2$
z	0.7	0.0	0.0	0.0
~z	0.0	0.7	0.7	0.7

$$m(\{zx_1x_2,\ \sim zx_1\sim x_2,\ \sim z\sim x_1x_2,\ \sim z\sim x_1\sim x_2\}) = 0.7$$

Thus, the belief-function representation of 'Discounted AND' relationship is given by the following set of m-values:

$$m(\{zx_1x_2,\ \sim zx_1\sim x_2,\ \sim z\sim x_1x_2,\ \sim z\sim x_1\sim x_2\}) = 0.7,$$
$$m(\{zx_1x_2,\ \sim zx_1\sim x_2,\ \sim z\sim x_1x_2,\ z\sim x_1\sim x_2\}) = 0.3. \tag{1}$$

These m-values yield the desired result. For example, if we assume that Z is true, i.e., $m(z) = 1.0$, propagating this information through the above relationship yields the following m-values for variables X_1 and X_2: $m(x_1) = 0.7$, $m(\sim x_1) = 0$, and $m(\{x_1, \sim x_1\}) = 0.3$, and $m(x_2) = 0.7$, $m(\sim x_2) = 0$, and $m(\{x_2, \sim x_2\}) = 0.3$. On the other hand, the fact that Z is not true, i.e., $m(\sim z) = 1.0$, when this information is propagated through the 'Discounted AND' relationship to X_1 and X_2, it yields the following m-values at the two variables: $m(\{x_1, \sim x_1\}) = 1.0$, and $m(\{x_2, \sim x_2\}) = 1.0$). This result is similar to the result obtained in an 'AND' relationship.

4.2. 'Conditional OR (CR)' Relationship

Gao[16] has used the 'Conditional OR (CR)' for the interrelationship between presence or absence of fraud in the financial statements of a company and the accounts in which fraud may have been committed. We show in this article how one can model such a relationship under DS theory for a general situation. Let us first consider a simple case to illustrate the process. Assume that Z represents the variable 'Fraud exists in the financial Statements' with its two values 'z' and

'~z', respectively, representing the states that fraud is present, and fraud is not present. Suppose there are two possible accounts, X_1, and X_2, through which management can perpetrate fraud. We assume X_1 and X_2 to be binary variables with values x_1, $\sim x_1$, x_2, and $\sim x_2$, respectively, representing the state that fraud is present or absence in X_1 and in X_2. The 'Conditional OR' between Z, and X_1 and X_2 implies that if fraud is suspected in the financial statement then there is a_1 percent possibility that fraud is in X_1 and a_2 percent possibility that fraud is in X_2, where a_1 + a_2 = 1. However, if there is no fraud suspected in the financial statements, i.e., '~z' is true then there is no fraud in X_1 or in X_2, i.e., $\sim x_1$ and $\sim x_2$ are true. At the same time, the 'Conditional OR' implies that if X_1 has fraud then irrespective of whether X_2 has fraud or not the financial statements will have fraud, i.e., z will be true. Also, ~z will be true if and only if both $\sim x_1$ and $\sim x_2$ are true. In other words, the financial statements will not have fraud only when both accounts X_1 and X_2 are not fraudulent. The 'CR' relationship is particularly useful for auditors to evaluate how fraud can be perpetrated through various accounts and schemes.

Let us apply the algorithm described in Section 3.1. Step 1 yields the top row and first column of Table 5 and the body of the table. Note that the given states are the values of Z. The 'Conditional OR' defined above dictates that if z is true then x_1 is true a_1 percent of the times and x_2 is true a_2 percent of the times. In addition, if z is true then there is zero chance that '$\sim x_1$' or '$\sim x_2$' would be true. These values are represented in the second and fourth columns of Table 5. However, when '~z' is true, both '$\sim x_1$' and '$\sim x_2$' are true 100 percent of the times. Columns 6 and 8 represent these conditions. Step 2 of the algorithm yields the values in the rectangular boxes in Table 5 with the assumption that a_1 is less than a_2. Step 3 yields the following m-value for the first focal element: $m(\{zx_1x_2, zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = a_1$.

Table 5: 'Conditional OR' Relationship between Z, and X_1 and X_2. If z is true then x_1 is true a_1 percent of the times and x_2 is true a_2 percent of the time, assuming $a_1 < a_2$.

Variable Z							
z is true				~z is true			
x_1	a_1	x_2	a_2	x_1	0	x_2	0
~x_1	0	~x_2	0	~x_1	1.0	~x_2	1.0
$\{x_1, ~x_1\}$	a_2	$\{x_2, ~x_2\}$	a_1	$\{x_1, ~x_1\}$	0	$\{x_2, ~x_2\}$	0
$m(\{zx_1x_2, zx_1~x_2, ~z~x_1~x_2\}) = a_1$							

Next, according to Step 4, we subtract the smallest m-value, a_1 in the present case, from the chosen m-values in each of the other columns. We then repeat Steps 2 and 3 and obtain the next focal element with following m-value: $m(\{zx_1x_2, z~x_1x_2, ~z~x_1~x_2\}) = a_2$, as given in Table 6.

Table 6: Step 4 for 'Conditional OR' Relationship

Variable Z							
Z is true (z)				Z is not rue (~z)			
x_1	0	x_2	a_2	x_1	0	x_2	0
~x_1	0	~x_2	0	~x_1	$1-a_1 = a_2$	~x_2	$1- a_1 = a_2$
$\{x_1, ~x_1\}$	a_2	$\{x_2, ~x_2\}$	0	$\{x_1, ~x_1\}$	0	$\{x_2, ~x_2\}$	0
$m(\{zx_1x_2, z~x_1x_2, ~z~x_1~x_2\}) = a_2$							

Thus, we obtain the following m-values to represent the conditional OR relationship. CR, as described earlier under DS theory:

$$m(\{zx_1x_2, zx_1{\sim}x_2, {\sim}z{\sim}x_1{\sim}x_2\}) = a_1$$

$$m(\{zx_1x_2, z{\sim}x_1x_2, {\sim}z{\sim}x_1{\sim}x_2\}) = a_2 \qquad (2)$$

One can easily see that the CR relationship as defined in (2) has the right properties. For example, if we assume that fraud is committed in the financial statements, i.e., $m(z) = 1$, then after propagating this information through the CR relationship (See Figure 1), we obtain $m(x_1) = a_1$, $m({\sim}x_1) = 0$, and $m(x_2) = a_2$, $m({\sim}x_2) = 0$. If we assume that there is no fraud in the financial statements, i.e., $m({\sim}z) = 1$, then we obtain the following m-values after propagating this information through the CR relationship to X_1 and X_2: $m(x_1) = 0$, $m({\sim}x_1) = 1$, $m(x_2) = 0$, and $m({\sim}x_2) = 1$. This is a logical result; if the overall financial statements are not fraudulent then there is no fraud in individual accounts. Also, one can show that if there is fraud in one of the accounts, say X_1, i.e., $m(x_1) = 1$, then irrespective of the value of X_2, we obtain $m(z) = 1$ when this information is propagated through CR to Z from X_1 and X_2 (see Figure 1). This is again a logical result; if fraud is detected in one of the accounts then the overall financial statements contain fraud.

We can extend the above relationship to a situation where Z is related to n binary variables through the CR relationships. Such a relationship becomes important when one is developing fraud models with several accounts being part of the financial statements, not just two. Here is the result obtained through a direct extension of (2):

$$m(\{zx_i \bigcap_{\substack{j \\ j \neq i}}^{n} \Theta_j, {\sim}z{\sim}x_i \bigcap_{\substack{j \\ j \neq i}}^{n} {\sim} x_j \}) = a_i, \text{ where } i = 1, 2, \ldots n. \qquad (3)$$

where Θ_j represents the frame of X_j, i.e., $\Theta_j = \{x_j, {\sim}x_j\}$ and a_i represents the weight associated with variable X_i. These weights add to one, i.e., $a_1 + a_2 + a_3 + \ldots + a_n = 1$.

4.3. 'Weighted Average' Relationship

Weighted average, by assigning weights to variables to reflect their relative importance, is an important relationship that is frequently used in evidential reasoning networks. For example, Sun et al.[5] have used 'Weighted Average' relationship to develop information systems risk assessment model under DS theory. Shenoy and Shenoy[8] have used this relationship to model financial portfolios under DS theory. However, as discussed later in this section, their belief-function model of weighted average relationship does not represent the generally accepted notion of weighted average relationship.

In this section, we develop a belief-function model for the 'Weighted Average' relationship between a binary variable Z and n binary variables, X_1, X_2, …. X_n. First, we derive the model with n = 2 and then extend it to the general situation. For the case where n = 2, we expect the following results when belief masses, i.e., m-values, at the two variables X_1 and X_2 are propagated through the 'Weighted Average' relationship to Z (See Figure 1):

$$m(z) = w_1 m_1(x_1) + w_2 m_2(x_2),$$

$$m(\sim z) = w_1 m_1(\sim x_1) + w_2 m_2(\sim x_2),$$

$$m(\{z, \sim z\}) = w_1 m_1(\{x_1, \sim x_1\}) + w_2 m_2(\{x_2, \sim x_2\}), \tag{4}$$

where symbols m_1 and m_2, respectively, stand for the belief masses defined at X_1 and X_2, w_1 and w_2, respectively, determine the weight put on the variables X_1 and X_2 such that $w_1 + w_2 = 1$. For the current example, the 'Weighted Average' relationship also implies that if we know that 'z' is true, i.e., $m(z) = 1$, then we know that x_1 is true with weight w_1 and x_2 is true with weight w_2.

Let us apply the algorithm described in Section 3.1 to 'Weighted Average' relationship to determine a belief-function representation of the relationship. Step1 yields Table 7 and its body. Note that, like modeling CR in the previous section, we have values of the variable Z in the first

row as given. The elements of X_1 and X_2 are listed as columns. We know that if 'z' is true then 'x$_1$' is true with support (i.e., weight) w_1 and 'x$_2$' is true with support (i.e., weight) w_2. These requirements are reflected in Table 7 through columns 2, 4, and 8. Performing Steps 2 and 3 yields the first focal point and its m-value as: $m(\{zx_1x_2, zx_1\sim x_2, \sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1$.

Applying Step 4 and repeating Steps 2 and 3 yields the second focal element and its m-value as: $m(\{zx_1x_2, z\sim x_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = w_2$.

Table 7: 'Weighted Average' Relationship between Z, and X_1 and X_2, assuming $w_1 < w_2$

Variable Z							
Z is true (z)				Z is not true (~z)			
x_1	$\boxed{w_1}$	x_2	w_2	x_1	0	x_2	0
$\sim x_1$	0	$\sim x_2$	0	$\sim x_1$	$\boxed{w_1}$	$\sim x_2$	w_2
$\{x_1, \sim x_1\}$	w_2	$\{x_2, \sim x_2\}$	$\boxed{w_1}$	$\{x_1, \sim x_1\}$	w_2	$\{x_2, \sim x_2\}$	$\boxed{w_1}$
$m(\{zx_1x_2, zx_1\sim x_2, \sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1$							
$m(\{zx_1x_2, z\sim x_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = w_2$							

Thus, applying the algorithm in Section 3.1 to 'Weighted Average' relationship, we obtained the following set of m-values defining this relationship between Z and the two variables X_1 and X_2:

$$m(\{zx_1x_2, zx_1\sim x_2, \sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1,$$
$$m(\{zx_1x_2, z\sim x_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = w_2. \tag{5}$$

Analysis of the 'Weighted Average' Relationship

Here we analyze the performance of the weighted average relationship between Z and the two variables X_1 and X_2, under the following two situations: 1) Propagate belief masses from the two variables, X_1 and X_2, to Z (see Figure 1), and 2) Propagate belief masses from Z to X_1 and X_2. Let us consider the following set of m-values defined at X_1 and X_2: $m_1(x_1)$, $m_1(\sim x_1)$, and $m_1(\{x_1, \sim x_1\})$; and $m_2(x_2)$, $m_2(\sim x_2)$, and $m_2(\{x_2, \sim x_2\})$.

We use Shenoy and Shafer[6] approach to propagate the belief masses from X_1 and X_2 to Z through the 'Weighted Average' relationship as defined in (5). First, we vacuously (see vacuous extension examples in Srivastava[13]) extend the belief masses at X_1 and X_2 to the frame of the 'Weighted Average' relationship. Next, we combine the three sets of belief masses at the relationship node using Dempster's rule of combination and then marginalize (see marginalization examples in Srivastava[13]) them to the variable Z. The frame of the relationship is defined as $\Theta = (\{zx_1x_2, zx_1\sim x_2, z\sim x_1x_2, \sim z\sim x_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\})$ which is obtained by finding the union of the two focal elements in (5).

The following are the belief masses as a result of combining the belief masses at X_1 and the m-values at the 'Weighted Average' relationship:

$$m(\{zx_1x_2, zx_1\sim x_2\}) = w_1 m_1(x_1),$$

$$m(\{zx_1x_2, \sim zx_1\sim x_2\}) = w_2 m_1(x_1),$$

$$m(\{\sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1 m_1(\sim x_1),$$

$$m(\{z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_2 m_1(\sim x_1),$$

$$m(\{zx_1x_2, zx_1\sim x_2, \sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1 m_1(\{x_1, \sim x_1\}),$$

$$m(\{zx_1x_2, z\sim x_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = w_2 m_1(\{x_1, \sim x_1\}). \tag{6}$$

Next, we combine the belief masses at X_2 with the above belief masses given in (6) and then marginalize them to variable Z. This process yields exactly the same belief masses as desired in (4):

$$m(z) = w_1 m_1(x_1) + w_2 m_2(x_2),$$

$$m(\sim z) = w_1 m_1(\sim x_1) + w_2 m_2(\sim x_2),$$

$$m(\{z, \sim z\}) = w_1 m_1(\{x_1, \sim x_1\}) + w_2 m_2(\{x_2, \sim x_2\}). \tag{7}$$

Let us consider a situation where belief masses are being propagated from Z to the variables X_1 and X_2. Again, we use Shenoy and Shafer[6] approach of local computations to achieve this goal. First, we vacuously extend the belief masses at Z to the 'Weighted Average' relationship and combine the belief masses at the relationship using Dempster's rule. Next, we marginalize the belief masses obtained in the previous step. Suppose the belief masses at Z are represented by the following set of m-values: $m_Z(z)$, $m_Z(\sim z)$, and $m_Z(\{z, \sim z\})$. These belief masses when combined with the belief masses in (5) yield the following set of belief masses:

$$m(\{zx_1x_2, zx_1\sim x_2\}) = w_1 m_Z(z),$$

$$m(\{zx_1x_2, z\sim x_1x_2\}) = w_2 m_Z(z),$$

$$m(\{\sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1 m_Z(\sim z),$$

$$m(\{\sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = w_2 m_Z(\sim z),$$

$$m(\{zx_1x_2, zx_1\sim x_2, \sim z\sim x_1x_2, \sim z\sim x_1\sim x_2\}) = w_1 m_Z(\{z, \sim z\}),$$

$$m(\{zx_1x_2, z\sim x_1x_2, \sim zx_1\sim x_2, \sim z\sim x_1\sim x_2\}) = w_2 m_Z(\{z, \sim z\}). \tag{8}$$

After marginalizing the above m-values at X_1 and X_2, we get the following set of m-values:

$$m(x_1) = w_1 m_Z(z),$$

$$m(\sim x_1) = w_1 m_Z(\sim z),$$

$$m(\{x_1, \sim x_1\}) = w_1 m_Z(\{z, \sim z\}) + w_2, \tag{9}$$

and

$$m(x_2) = w_2 m_Z(z),$$

$$m(\sim x_2) = w_2 m_Z(\sim z),$$

$$m(\{x_2, \sim x_2\}) = w_2 m_Z(\{z, \sim z\}) + w_1. \tag{10}$$

The results in (9) and (10) are logical. If we know that z is true with $m(z) = 1$, then we know that x_1 is true with a belief mass of weight w_1 and x_2 is true with a belief mass of w_2, and there is no evidence in support of '$\sim x_1$', and '$\sim x_2$', i.e., $m(\sim x_1) = 0$, and $m(\sim x_2) = 0$. Similar logical results are obtained if we set $m(\sim z) = 1$, i.e., we get $m(x_1) = 0$, $m(\sim x_1) = w_1$, and $m(x_2) = 0$, $m(\sim x_2) = w_2$.

'Weighted Average' Relationship for n binary variables

Here we generalize the 'Weighted Average' relationship to a situation where Z is related to n binary variables through this relationship. Such a relationship becomes important when one is developing models that contain assigned weights to variables in order to reflect their relative importance as used in Shenoy and Shenoy[8] and Sun et al.[5]. Here is the result obtained through a direct extension of (5):

$$m(\{z x_i \bigcap_{\substack{j \\ j \neq i}}^{n} \Theta_j, \sim z \sim x_i \bigcap_{\substack{j \\ j \neq i}}^{n} \Theta_j\}) = w_i, \text{ where } i = 1, 2, \dots n. \tag{11}$$

where Θ_j represents the frame of X_j, i.e., $\Theta_j = \{x_j, \sim x_j\}$ and w_i represents the weight associated with variable X_i. These weights add to one, i.e., $w_1 + w_2 + w_3 + \dots + w_n = 1$.

Comparison with Shenoy and Shenoy Model of Weighted Average Relationship

In this section, we compare and contrast the 'Weighted Average' model proposed and used by Shenoy and Shenoy[8] with the model developed here. For the comparison purpose, we consider a simple case where a binary variable Z is related to two other binary variables X_1 and X_2 through the 'Weighted Average' relationship as defined by Shenoy and Shenoy. From here onward, we will call Shenoy and Shenoy 'Weighted Average' model as the 'Approximate Weighted Average' model and the model developed in this paper as the 'Weighted Average' model. In the 'Approximate Weighted Average' relationship, Shenoy and Shenoy discount the belief masses at the variables X_1 and X_2 by (1 - the respective weights) to propagate to variable Z and then combine them using Dempster's rule. According to Shenoy and Shenoy, we discount belief masses at X_1 by $(1- w_1)$, and at X_2 by $(1- w_2)$, where w_1 and w_2, respectively, represent the relative weights or importance of X_1 and X_2 and $w_1 + w_2 = 1$. Thus, the belief masses propagated to the variable Z from X_1 and X_2 as a result of discounting will yield the following values:

Belief masses at Z from X_1:

$$m_{1Z}(z) = w_1 m_1(x_1), \; m_{1Z}(\sim z) = w_1 m_1(\sim x_1), \text{ and } m_{1Z}(\Theta_Z) = w_1 m_1(\Theta_1) + w_2.$$

Belief masses at Z from X_2:

$$m_{2Z}(z) = w_2 m_2(x_2), \; m_{2Z}(\sim z) = w_2 m_2(\sim x_2), \text{ and } m_{2Z}(\Theta_Z) = w_2 m_2(\Theta_2) + w_1.$$

The above two sets of belief masses are combined using Dempster's rule. This combination yields the following belief masses at variable Z (labeled with a subscript S for Shenoy and Shenoy) as a result of the propagation of belief masses from X_1 and X_2 through the 'Approximate Weighted Average' relationship as defined by Shenoy and Shenoy.

$$m_S(z) = [w_1 m_1(x_1)+w_2 m_2(x_2)-w_1 w_2[m_1(x_1)m_2(x_2)+m_1(x_1)m_2(\sim x_2)+m_1(\sim x_1)m_2(x_2)]]/K, \quad (12)$$

$$m_S(\sim z) = [w_1 m_1(\sim x_1) + w_2 m_2(\sim x_2) - w_1 w_2 [m_1(\sim x_1) m_2(\sim x_2) + m_1(x_1) m_2(\sim x_2) + m_1(\sim x_1) m_2(x_2)]]/K, \quad (13)$$

$$m_S(\Theta_Z) = [w_1 m_1(\Theta_1) + w_2 m_2(\Theta_2) + w_1 w_2 (1 - m_1(\Theta_1))(1 - m_2(\Theta_2))]/K, \quad (14)$$

where K is the renormalization constant in Dempster's rule as given below:

$$K = 1 - w_1 w_2 [m_1(x_1) m_2(\sim x_2) + m_1(\sim x_1) m_2(x_2)]. \quad (15)$$

As one can see from (12) - (14), the belief masses at Z are not equal to the exact weighted average belief masses in (7). For $w_1 = 0.5$, $m_1(x_1) = 1.0$ and $m_2(x_2) = 1.0$, the difference between m(z) in (7) and $m_S(z)$ in (12) is 0.25, the maximum. As one can see, for these values, (7) yields m(z) = 1.0, and (12) yields $m_S(z) = 0.75$, a big difference. Figure 2 plots the difference between m(z) from (7) and $m_S(z)$ from (12), and the difference between m(~z) from (7) and $m_S(\sim z)$ from (13) as a function of the weight w_1 for the following m-values for X_1 and X_2: $m_1(x_1) = 0.6$, $m_1(\sim x_1) = 0.2$, and $m_2(x_2) = 0.8$, $m_2(\sim x_2) = 0.1$. As one can see in Figure 2, both differences peak at $w_1 = 0.5$. Thus, using the Shenoy and Shenoy 'Weighted Average' relationship would yield an undesirable result when propagating belief in a network containing such a relationship.

----- Figure 2 here -----

5. SUMMARY AND CONCLUSION

This paper has presented an algorithm for developing models under DS theory for categorical and 'uncertain' logical relationships among binary variables. We have illustrated the use of the algorithm in modeling belief-function representations of the following categorical relationships: 'AND', 'OR', 'Exclusive OR (EOR)' and 'Not Exclusive OR (NEOR)', and 'AND-NEOR' and of the following uncertain relationships: 'Discounted AND', 'Conditional OR', and 'Weighted Average'. Gillett[12] has shown the importance of 'Discounted AND' relationship in the accounting context. Gao[16] has demonstrated the use of 'Conditional OR' in analyzing and

modeling management fraud in the financial statements of a company. Sun et al[5] and Shenoy and Shenoy[8] have identified the need for the belief-function representation of 'Weighted Average' relationship, but in absence of such a model they used an 'Approximate Weighted Average' model as proposed by Shenoy and Shenoy[8]. However, the 'Approximate Weighted Average' model of Shenoy and Shenoy fails to represent the correct relationship as demonstrated in this paper. Having belief-function representations of such relationships would make it much more convenient to develop belief-function models of practical problems in a variety of domains such as accounting, finance, and information systems.
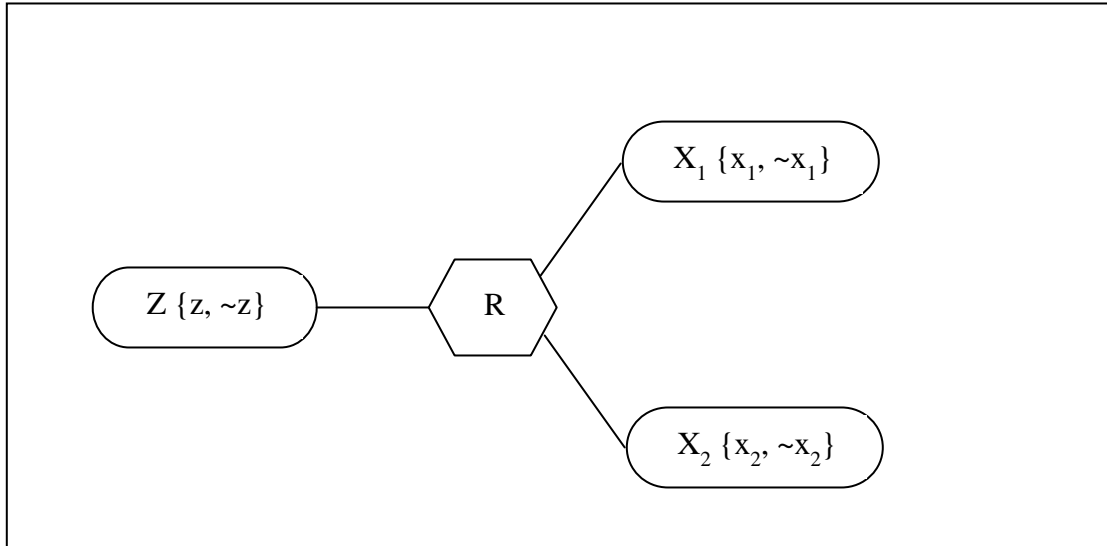
REFERENCE

1. Shafer G. *A Mathematical Theory of Evidence*. Princeton University Press;1976.

2. Shafer G, Shenoy PP, Srivastava RP. AUDITOR'S ASSISTANT: A Knowledge Engineering Tool For Audit Decisions. *Proceedings of the 1988 Touche Ross University of Kansas Symposium on Auditing Problems*, May, 1988. pp 61-79.

3. Yager RR, Kacprzyk J, Fedrizzi M. *Advances in the Dempster-Shafer Theory of Evidence*. New York, NY: John Wiley and Sons, 1994.

4. Srivastava RP, Mock TJ. *Belief Functions in Business Decisions.* Physica-Verlag, Heidelberg, Springer-Verlag Company; 2002.

5. Sun L, Srivastava RP, Mock TJ. An Information Systems Security Risk Assessment Model under Dempster-Shafer Theory of Belief Functions. *Journal of Management Information Systems,* Vol. 22, No. 4: 2006, pp 109-142.

6. Shenoy PP, Shafer P. Propagating Belief Functions with Local Propagation. *IEEE Expert*, Vol. 1; 1986. pp 43-52.

7. Srivastava RP, Mock TJ. Evidential Reasoning for WebTrust Assurance Services. *Journal of Management Information Systems*, Vol. 16, No. 3; Winter 2000. pp 11-32.

8. Shenoy C, Shenoy P. Modeling Financial Portfolios using Belief Functions. In *Belief Functions in Business Decisions,* edited by R. P. Srivastava and T. Mock, Physica-Verlag, Heidelberg, Springer-Verlag Company: 2002. pp 316-332.

9. Srivastava RP, Shafer G. Belief-Function Formulas for Audit Risk. *The Accounting Review*, April: 1992. pp 249-283.

10. Srivastava RP, Mock TJ, Turner J. Analytical Formulas for Risk Assessment for a Class of Problems where Risk Depends on Three Interrelated Variables. *International Journal of Approximate Reasoning,* Vol. 45; 2007. pp 123–151.

11. Bovee M, Srivastava RP, Mak B. A Conceptual Framework and Belief-Function Approach to Assessing Overall Information Quality. *International Journal of Intelligent Systems*, Volume 18, No. 1, January: 2003. pp 51-74.

12. Gillett PR. A Comparative Study of Audit Evidence and Audit Planning Models Using Uncertain Reasoning. Doctoral Dissertation, School of Business, University of Kansas; 1996.

13. Srivastava RP. The Belief-Function Approach to Aggregating Audit Evidence. *International Journal of Intelligent Systems,* Vol. 10, No. 3, March: 1995. pp 329-356.

14. Srivastava RP, Shenoy PP, Shafer G. Propagating Beliefs in an 'AND' Tree." *International Journal of Intelligent Systems,* Vol. 10: 1995. pp 647-664.

15. Srivastava RP, Dutta SK, Johns R. An Expert System Approach to Audit Planning and Evaluation in the Belief-Function Framework. *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 5, No. 3; 1996. pp 165-183.

16. Gao L. Investigation of the Perpetration and Concealment Process of Management Fraud: An Empirical Analysis of Fraud Schemes. Doctoral Dissertation, School of Business, University of Kansas; 2005.

17. Srivastava RP. Audit Decisions Using Belief Functions: A Review. *Control and Cybernetics*. Vol. 26, No.2; 1997. pp 135-160.

18. Srivastava R P, Lu H. Structural Analysis of Audit Evidence using Belief Functions. *Fuzzy Sets and Systems*, Vol. 131, Issues No. 1, October; 2002. pp 107-120.

19. Leslie DA, Aldersley SJ, Cockburn DJ, Reiter CJ. An assertion-based approach to auditing (discussant's remarks). *Proceedings of the 1986 Touche Ross/University of Kansas Symposium on Auditing Problems*. Lawrence, KS: School of Business, University of Kansas; 1986. pp 31-67.

Figure 2: The Difference of two m-values at variable Z; one set based on the exact 'Weighted Average' model and the other set based on the 'Approximate Weighted Average' model of Shenoy and Shenoy[8] as a function of the Weight w_1. The following m-values are assumed for the graph:
$m_1(x_1) = 0.6$, $m_1(\sim x_1) = 0.2$, and $m_2(x_2) = 0.8$, $m_2(\sim x_2) = 0.1$.