ELSEVIER

Decision Support Systems 16 (1996) 39–53

Decision Support Systems

A linear programming framework for logics of uncertainty [1]

K.A. Andersen [a,*], J.N. Hooker [b]

[a] *Matematisk Institut, Aarhus Universitet DK-8000 Århus C, Denmark*
[b] *Graduate School of Industrial Administration Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Abstract

Several logics for reasoning under uncertainty distribute "probability mass" over sets in some sense. These include probabilistic logic, Dempster-Shafer theory, other logics based on belief functions, and second-order probabilistic logic. We show that these logics are instances of a certain type of linear programming model, typically with exponentially many variables. We also show how a single linear programming package can implement these logics computationally if one "plugs in" a different column generation subroutine for each logic, although the practicality of this approach has been demonstrated so far only for probabilistic logic.

Keywords: Linear programming; Logic; Uncertainty

1. Introduction

Several logics for reasoning under uncertainty are variations on a theme. Numbers, perhaps probabilities, are assigned to propositions to indicate degrees of confidence. The object is to determine the degree of confidence one can have in a conclusion inferred from the propositions. Dependencies among the propositions require that some of the "probability mass" assigned to one proposition be distributed to others. Solution of this distribution problem yields a range of confidence levels for the conclusion.

The oldest uncertainty logic of this kind is Boole's probabilistic logic [2,3], which was revived a few years ago by Hailperin [12,13], rediscovered by Nilsson [18], and recently discussed by a number of others [5,8–11,15–17,19]. But Dempster-Shafer theory has a similar structure [20], as does a logic based on Shafer's belief functions suggested by Dubois, Prade and others [7,16,17]. A number of other logics can be devised along similar lines.

It seems to be widely recognized that several uncertainty logics can be viewed as probability mass distribution problems in some sense. Here we not only make this sense precise but show the following; (a) Inference in all these logics can all be formulated as a linear programming problem of a certain type, typically with exponentially many variables in the worst case. (b) At least in the logics discussed here, the exponential number of variables can be dealt with computationally by using column generation schemes, a well known device for such situations. The practicality of this approach has already been demonstrated for

* Corresponding author

[1] The second author is partially supported by ONR grant N00014-92-J-1028 and AFOSR grant 91-0287.

SSDI 0167-9236(94)00055-7

probabilistic logic [15]. This suggests that a single linear programming code can implement several uncertainty logics, if one plugs in a different column generation subroutine for each logic. Computational testing, however, has not been carried out on logics other than probabilistic logic.

We will show how several logics fit into this framework and will describe the column generation subproblem in each case. In probabilistic logic, column generation is a pseudo-boolean optimization problem, as is already well known. In Dempster-Shafer theory it is an integer programming problem. We will introduce a second-order probability logic in which it is a mixed integer/linear programming problem. In the logic of belief functions mentioned above, there is no exponential explosion of columns, and a column generation technique is likely to be unnecessary.

For some applications one may wish to add nonlinear constraints, although we do not pursue this possibility here. For instance, Dempster's combination rule, which is a key ingredient of Dempster-Shafer theory, combines a renormalization device with an independence assumption. The rule can be used perfectly well, and in many cases more appropriately, without the independence assumption, and we do so here. But the independence assumption can be imposed by adding nonlinear constraints to the otherwise linear model. Probabilistic logic can also be augmented with independence assumptions, such as those depicted by a Bayesian network, by adding nonlinear constraints. In [1] we show when and how this can be done without an exponential growth in the number of nonlinear constraints. It is unclear at this point whether column generation techniques may be successfully extended to nonlinear problems.

We begin below with a statement of the general linear programming model. After discussing briefly how a column generation approach is implemented computationally, we show how several logics can be placed in this framework. These include probabilistic logic, a version of probabilistic logic with unreliable sources of information, Dempster-Shafer theory, second-order probabilistic logic (which allows for unreliable sources in a different way), and a simple logic of belief functions. The expositions is clarified by using some small examples.

2. The general model

We are given propositions $F_1, \dots, F_h$ and some information about the level of confidence we may have in them. The confidence level for F_i is indicated by its "mass," which is a number in the interval [0,1]. The interpretation of mass varies from one logic to another; in probabilistic logic, for instance, it is probability mass in the classical sense. Since the precise mass of F_i may be unknown, we will suppose that an interval $[L_i, U_i]$ is given, within which the mass lies. If nothing is known about the mass of F_i, we set $L_i = 0$, $U_i = 1$.

Let S_i be the set of possible *outcomes* that make proposition F_i true. In probabilistic logic, S_i is the set of "possible worlds" in which F_i is true. In Dempster-Shafer theory, it is a subset of the "frame of discernment." We let $\mu(S_i)$ denote the mass of F_i, which we also refer to as the mass of S_i. The sets $S_1, \dots, S_h$ need not all be distinct.

We are interested in knowing how much confidence we can place in a proposition F_t whose mass is not given. Its mass may be constrained by the fact that S_t intersects some of the sets for which masses are given. In other words, F_t may be logically related to $F_1, \dots, F_h$. A fundamental problem, therefore, is to find out how much of the mass assigned a set S_i can or must be associated with other sets.

2.1. Example 1

Suppose we assign masses 0.8 and 0.7 respectively to propositions F_1 and F_2 and wish to infer something about the mass of a third proposition F_3. The logical relations among the propositions are indicated by the relations among the sets S_1, S_2, S_3 depicted in Fig. 1. For instance, the figure implies that F_1 and F_2 can both be true (since S_1 and S_2 intersect), and if they are both true, F_3 must also be true (since all outcomes in $S_1 \cap S_2$ are in S_3).

The mass 0.8 assigned S_1 can be regarded as lying in S_1's circle. Some of this mass may lie in S_3's circle, and similarly for the mass of 0.7 as-

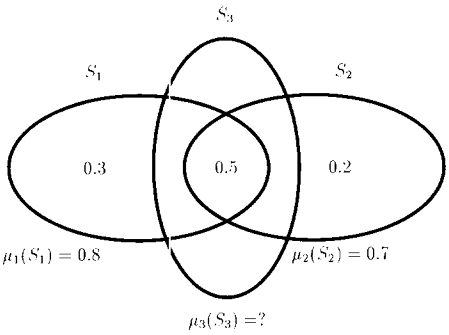


Fig. 1. Relations among the sets S_1, S_2 and S_3.

signed S_2. The inference problem is to determine the minimum and maximum mass that may lie in the S_3 circle. It is therefore a mass distribution problem.

For reasons that will become evident as we examine particular logics, we will formulate this problem as one of distributing each set's mass over its intersections with other sets. For convenience let us write the intersection $\cap_{i \in J} S_i$ as S_J. Then we wish to distribute each S_i's mass over S_i's intersections S_J with other sets. Or more precisely, we will distribute S_i's mass over the *index sets* J that indicate which sets are intersected. The two are not the same, because possibly $S_J = S_{J'}$ when $J \neq J'$. In Fig. 1, for instance, $S_{\{1,2\}} = S_{\{1,2,3\}}$. The notion of distributing mass over index sets may seem odd at this point, but it is one of the keys to unifying logics of uncertainty.

The precise distribution problem varies from logic to logic because the structure of the logic dictates which index sets J may receive mass. We will show how the distribution problem of Fig. 1 would be formulated in probabilistic logic, and then how it would be formulated in Dempster-Shafer theory. We will explain the motivation for these formulations in the ensuing sections of the paper. Our sole purpose here is to illustrate that the distribution problem can be formulated in different ways.

2.2. Example 1; Probabilistic interpretation

In probabilistic logic we view the 0.8 mass assigned S_1 as distributed over the three regions of the S_1 circle in Fig. 1, and similarly for S_2. One possible distribution is shown in the figure. Once a distribution of this sort is specified, the mass inside S_3 is determined.

To view this as distribution of mass over intersections, we add propositions $F_4, \ldots, F_7$ to the original three, where F_4, F_5, F_6 are respectively $\neg F_1$, $\neg F_2$, $\neg F_3$, and F_7 is a tautology. S_4, S_5, S_6 are therefore the complementary sets $\bar{S}_1$, $\bar{S}_2$, $\bar{S}_3$, respectively. We also assign mass 1 to the tautology.

The distribution problem just described is equivalent to distributing each S_i's mass over the sets J for which S_J is one of the regions into which S_i is divided in Fig. 1. Thus the 0.8 mass of S_1 is distributed over the index sets {1,2,3}, {1,5,3}, {1,5,6}, which correspond to the three regions of the S_1 circle, namely $S_{\{1,2,3\}} = S_1 \cap S_2 \cap S_3$, $S_{\{1,5,3\}} = S_1 \cap \bar{S}_2 \cap S_3$, and $S_{\{1,5,6\}} = S_1 \cap \bar{S}_2 \cap \bar{S}_3$.

Let q_J be the mass distributed to J. Then in general we have,

$$\mu(S_i) = \sum_{J \in I(i)} q_J, \tag{1}$$

where $I(i) \subset I$ is the family of index sets over which the mass $\mu(S_i)$ of S_i is to be distributed. Since $\mu(S_i)$ must lie in $[L_i, U_i]$, we have the constraints,

$$L_i \leq \mu(S_i) \leq U_i, \; i = 1, \ldots, m. \tag{2}$$

$$q_J \geq 0, \text{ all } J \in I, \tag{3}$$

where $\mu(S_i)$ is given by (1).

We can place bounds on the mass of S_t by solving the two optimization problems,

$$\min/\max \frac{\mu^*(S_t)}{1 - \mu^*(\emptyset)} \quad \text{s.t. } (2), (3). \tag{4}$$

The notation $\mu^*(S_t)$ is used to indicate that mass may be distributed differently in the objective function than in the constraints. Thus we have,

$$\mu^*(S_t) = \sum_{J \in I^*(t)} q_J, \tag{5}$$

where the index set $I^*(t)$ depends on the logic.

2.3. Example 1; Probabilistic interpretation, continued

In the probabilistic logic interpretation, the mass $\mu(S_1) = 0.8$ is distributed over the index sets in $I(1) = \{\{1,2,3\},\{1,5,3\},\{1,5,6\}\}$. So,

$$\mu(S_1) = q_{\{1,2,3\}} + q_{\{1,5,3\}} + q_{\{1,5,6\}},$$

and similarly for $\mu(S_2)$. The unknown mass $\mu^*(S_3)$ is distributed in the same way. No mass is assigned to the empty set, so that $\mu^*(\emptyset) = 0$. So the optimization problems (4) become,

$$\begin{aligned}
&\min/\max\ q_{\{1,2,3\}} + q_{\{1,5,3\}} + q_{\{4,2,3\}} + q_{\{4,5,3\}}\\
&\text{s.t. } 0.8 \le q_{\{1,2,3\}} + q_{\{1,5,3\}} + q_{\{1,5,6\}} \le 0.8\\
&0.7 \le q_{\{1,2,3\}} + q_{\{4,2,3\}} + q_{\{4,2,6\}} \le 0.7\\
&1 \le q_{\{1,2,3\}} + q_{\{4,2,3\}} + q_{\{1,5,3\}} + q_{\{4,5,3\}}\\
&\quad + q_{\{4,2,6\}} + q_{\{1,5,6\}} + q_{\{4,5,6\}} \le 1\\
&q_J \ge 0
\end{aligned}$$

It is easy to see that the distribution in Fig. 1, in which $\mu(S_3) = 0.5$, minimizes $\mu(S_3)$. The maximum value of $\mu(S_3)$ is 1, obtained by setting $(q_{\{1,2,3\}}, q_{\{1,5,3\}}, q_{\{4,2,3\}}) = (0.5, 0.3, 0.2)$.

2.4. Example 1; Dempster-Shafer interpretation

In Dempster-Shafer theory, the mass 0.8 assigned S_1 is divided into a) mass that lies specifically in $S_1 \cap S_2 \cap S_3$ b) additional mass that could lie anywhere in $S_1 \cap S_2$ c) additional mass that could lie anywhere in $S_1 \cap S_3$ and d) additional mass that could lie anywhere in S_1. (In this example, it happens that $S_1 \cap S_2 = S_1 \cap S_2 \cap S_3$.) So $\mu(S_1) = 0.8$ is distributed among $q_{\{1\}}$, $q_{\{1,2\}}$, $q_{\{1,3\}}$, $q_{\{1,2,3\}}$. Also we add a tautology to the original three propositions and assign it mass 1.

Unlike probabilistic logic, Dempster-Shafer theory distributes the inferred mass $\mu^*(S_t)$ differently than the given masses $\mu(S_i)$. Here $\mu^*(S_3)$ is regarded as the sum of masses distributed to J's for which $S_J \subset S_3$. Since S_3, $S_1 \cap S_2$, $S_1 \cap S_3$, $S_2 \cap S_3$ and $S_1 \cap S_2 \cap S_3$ are all subsets of S_3, $\mu^*(S_3)$ is distributed to the index sets in $I^*(3) = \{\{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$. Here $\mu^*(\emptyset) = 0$ because none of the intersections S_J is a subset of the empty set.

The optimization problems (4) become,

$$\begin{aligned}
&\min/\max \frac{q_{\{3\}} + q_{\{1,2\}} + q_{\{1,3\}} + q_{\{2,3\}} + q_{\{1,2,3\}}}{1-0}\\
&\text{s.t. } 0.8 \le q_{\{1\}} + q_{\{1,2\}} + q_{\{1,3\}} + q_{\{1,2,3\}} \le 0.8\\
&0.7 \le q_{\{2\}} + q_{\{1,2\}} + q_{\{2,3\}} + q_{\{1,2,3\}} \le 0.7\\
&1 \le q_{\emptyset} + q_{\{1\}} + q_{\{2\}} + q_{\{3\}} + q_{\{1,2\}} + q_{\{1,3\}} + q_{\{2,3\}}\\
&\quad + q_{\{1,2,3\}} \le 1\\
&q_J \ge 0
\end{aligned}$$

(The variable $q_{\emptyset}$ represents mass that is assigned to no particular set.) It happens that $\mu^*(S_3)$ is restricted to the same interval as in the probabilistic interpretation, namely [0.5,1]. It is 0.5 when $(q_{\{1\}}, q_{\{2\}}, q_{\{1,2\}}) = (0.3, 0.2, 0.5)$. It is 1 when $(q_{\{3\}}, q_{\{1,2\}}, q_{\{1,3\}}) = (0.2, 0.7, 0.1)$.

Note that the objective function in (4) is normalized by dividing by the mass that is not assigned to the empty set. Among the logics we discuss, this plays a role only in Dempster-Shafer theory. (In the above example, $\mu^*(\emptyset)$ happened to be zero.) In the other logics, no mass is assigned to the empty set, so that $\mu^*(\emptyset) = 0$ and (4) is a linear programming problem. When $\mu^*(\emptyset) \neq 0$, (4) becomes a fractional programming problem that is readily transformed to a linear programming problem using well-known methods [4].

The sets $I(i)$ and $I^*(t)$ give instructions for generating the columns of the coefficient matrix in (4). Each column corresponds to a set $J \in I$. Assuming $\mu^*(\emptyset) = 0$, it has the form $(y_0, y_1, \ldots, y_m)$, where y_i $(i \ge 1)$ is the coefficient of q_J in constraint i and y_0 its coefficient in the objective function. The column is given by,

$$y_i = \begin{cases} 1 & \text{if } J \in I(i) \\ 0 & \text{otherwise,} \end{cases} \quad i = 1, \ldots, m,$$

$$y_0 = \begin{cases} 1 & \text{if } J \in I^*(t) \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

A similar column definition can be given for the linearized version of (4) when $\mu^*(\emptyset) \neq 0$.

Table 1
Linear programming structure of four uncertainty logics

	Probabilistic logic	Dempster-Shafer theory	2nd-order probabilistic logic	Belief functions
Interpretation of set S_i	Set of possible worlds in which proposition F_i is true	Subset of frame of discernment associated with some evidence source k	Half-space in probability space defined by $Pr(F_i) \le \pi$	Subset of frame of discernment
Constraints, where $\mu(S_i) = \Sigma_{J \in I(i)} q_J$	Bounds on prior probabilities [1]: $L \le \mu(S_i) \le U$	Value of basic probability function: $\mu(S_i) = m_k(S_i)$	Bounds on 2nd-order probability that $Pr(F_i) \le \pi$: $L \le \mu(S_i) \le U$	Bounds on value of belief function $Bel(S_i)$: $L \le \mu(S_i) \le U$
Objective function, where $\mu^*(S_i) = \Sigma_{J \in I^*(i)} q_J$	Posterior probability: $\dfrac{\mu^*(S_i \cap S_{c(i)})}{\mu^*(S_{c(i)})}$	Normalized probability: $\dfrac{\mu^*(S_t)}{1-\mu^*(\emptyset)}$	2nd-order probability that $Pr(F_i) \le \pi$: $\mu^*(S_t)$	Value of belief function $Bel(S_t)$: $\mu^*(S_t)$
Intersections $S_J = \cap_{i \in J} S_i$	All minimal [2] nonempty intersections of sets S_i	All intersections of one S_i associated with each evidence source k	All minimal nonempty intersections of half-spaces S_i	The sets S_i
I(i) contains all J for which:	$S_J \subset S_i$	$i \in J$	$S_J \subset S_i$	$S_J \subset S_i$
I*(i) contains all J for which:	$S_J \subset S_i$	$S_J \subset S_i$	$S_J \subset S_i$	$S_J \subset S_i$
Practical generation of columns q_J	Pseudo-boolean optimization	Integer programming	Mixed integer programming	None required

[1] Can also place bounds on conditional probabilities $L_i \le Pr(F_i \mid F_{c(i)}) \le U_i$ by using the constraints $L_i \mu(S_{c(i)}) \le \mu(S_i \cap S_{c(i)})$ and $U_i \mu(S_{c(i)}) \le \mu(S_i \cap S_{c(i)})$.
[2] A minimal intersection S_J is one containing no other nonempty intersection of S_i's.

Logics that use conditional probabilities require a model in which the masses in (4) are replaced with "conditional masses." A conditional mass $\mu(S|T)$ is defined to be equal to $\mu(S \cap T)/\mu(T)$. Intervals $[L_i, U_i]$ are given for conditional masses $\mu(S_i|S_{c(i)})$, where $S_{c(i)}$ is one of the sets $S_1, \ldots, S_h$. Thus the constraints (2) become,

$$L_i \leq \frac{\mu(S_i \cap S_{c(i)})}{\mu(S_{c(i)})} \leq U_i.$$

These constraints can be written in linear form,

$$0 \leq \mu(S_i \cap S_{c(i)}) - L_i \mu(S_{c(i)})$$
$$\mu(S_i \cap S_{c(i)}) - U_i \mu(S_{c(i)}) \leq 0. \tag{7}$$

Assuming $\mu^*(\emptyset) = 0$, the linear programming problem (4) becomes a fractional programming problem,

$$\min/\max \frac{\mu^*(S_t \cap S_{c(t)})}{\mu^*(S_{c(t)})}$$
$$\text{s.t. } (7), (3). \tag{8}$$

This is again convertible to a linear programming problem.

When the objective function of (8) is an unconditional mass $\mu^*(S_t)$, column J of (8) has the form $(y_0, y_1, z_1, \ldots, y_m, z_m)$, where

$$y_i = \begin{cases} 1 - L_i & \text{if } J \in I(i) \cap I(c(i)) \\ -L_i & \text{if } J \in I(c(i)) \setminus I(i) \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \ldots, m,$$

$$z_i = \begin{cases} 1 - U_i & \text{if } J \in I(i) \cap I(c(i)) \\ -U_i & \text{if } J \in I(c(i)) \setminus I(i), \\ 0 & \text{otherwise}, \end{cases} \quad i = 1, \ldots, m, \tag{9}$$

$$y_o = \begin{cases} 1 & \text{if } J \in I^*(t) \\ 0 & \text{otherwise} \end{cases}$$

When the objective function is conditional, a similar column definition can be given for the linearized form of (8).

Table 1 shows how the four uncertainty logics considered here fit into this pattern.

3. The column generation subproblem

We have the model

$$\min/\max \mu^*(S_t) = \sum_{J \in I^*(t)} q_J \tag{10}$$
$$\text{s.t. } \mu(S_i) = \sum_{J \in I(i)} q_J \leq U_i, \ i = 1, \ldots, m.$$
$$\mu(S_i) = \sum_{J \in I(i)} q_J \geq L_i, \ i = 1, \ldots, m.$$
$$q_J \geq 0, \ J \in I$$

In general, depending on the logic, there can be an exponential number of columns in the above program. Therefore it might be a good idea to use a column generation procedure. For an introduction to column generation procedures, especially Dantzig-Wolfe, see [6]. Assume that we have only generated some columns

$$\min/\max \mu^* S_t = \sum_{J \in \bar{I}^*(t)} q_J \tag{11}$$
$$\text{s.t. } \mu(S_i) = \sum_{J \in \bar{I}(i)} q_J \leq U_i, \ i = 1, \ldots, m.$$
$$\mu(S_i) = \sum_{J \in \bar{I}(i)} q_J \geq L_i, \ i = 1, \ldots, m.$$
$$q_J \geq 0, \ J \in \bar{I}$$

Here the index sets $\bar{I}^*(t)$, $\bar{I}(i)$ and $\bar{I}$ denote the set of columns generated so far. What we need now is to determine if the above columns are sufficient for solving the program and if not how a new (improving) column may be generated. This is usually done by constructing a subproblem. By maximizing or minimizing a certain objective function over some set, it is possible to decide if an improving column does exist. If one exists it is added to the program which is then resolved. If no improving column exists the optimal solution to the program has been found. The procedure can be started with any set of known columns, possibly none, in which case the program only contains slack- and surplus variables in the constraints.

To describe the column generation procedure suppose we are maximizing the programs (10) and (11). The procedure is similar when minimizing. Let λ_i, $i = 1,\ldots,m$, denote the dual variables to the constraints $\Sigma_{J \in \bar{I}(i)} q_J \leq U_i$, and let γ_i, $i = 1,\ldots,m$, denote the dual variables to the constraints $\Sigma_{J \in \bar{I}(i)} q_J \geq L_i$. Then $\lambda_i \geq 0$ and $\gamma_i \leq 0$, $i = 1,\ldots,m$. Suppose we construct a set, say P, such that the extreme points of P are exactly the possible columns q_J, $J \in I$. Then the subproblem becomes:

$$\min(\lambda - \gamma)y - y_0$$

s.t.

$$y = (y_1,\ldots,y_m) \in P \tag{12}$$

where $\lambda = (\lambda_1,\ldots,\lambda_m)$ and $\gamma = (\gamma_1,\ldots,\gamma_m)$.

If the optimal solution to this problem is strictly less than 0, then an improving column q_J has been found. The index J is added to the index sets $\bar{I}^*(t)$, $\bar{I}(i)$ and $\bar{I}$, meaning that column q_J is added to the program (11), which is then resolved. If the optimal solution to the contrary is at least 0, then an optimal solution to (10) has been determined. We notice that the problem is in a sense to state the set P in a reasonable way. For the logics described in this paper the sets P are described in Sections 4–7.

4. Probabilistic logic

In probabilistic logic, conditional probabilities $Pr(F_i|F_{c(i)})$ are constrained to lie in intervals $[L_i, U_i]$, where the F_i's are formulas of propositional logic. (An unconditioned probability $Pr(F_i)$ can be given by letting $F_{c(i)}$ be a tautologous proposition.) The object is to compute bounds on a probability mass $Pr(F_t|F_{c(t)})$ that are consistent with the given probabilistic information.

The formulas F_i contain atomic propositions $x_1,\ldots,x_n$. A *possible world* is an assignment v: $\{x_1,\ldots,x_n\} \to \{0,1\}^n$ of truth values to the atomic propositions. F_i is true in a possible world v when the assignment v makes it true, which we indicate by writing $v(F_i) = 1$.

Let S_i be the set of possible worlds in which F_i is true. Thus $Pr(F_i)$ is the probability that the actual world lies in S_i. We therefore interpret $\mu(S_i)$ to be the probability $Pr(F_i)$ and $\mu(S_i|S_{c(i)})$ to be the conditional probability $Pr(F_i|F_{c(i)})$.

The given intervals $[L_i,U_i]$ impose the constraints (7). Since the probability of all possible worlds must sum to one, we use one of the constraints (7) to assign a mass of one to a tautologous proposition.

By the law of total probability, $\mu(S_i)$ is the sum of the probabilities of the possible worlds in S_i. The probability mass $\mu(S_i)$ must therefore be distributed over these worlds. If we let q_v denote the probability of world v, we have

$$\mu(S_i) = \mu^*(S_i) = \sum_{v(S_i)=1} q_v. \tag{13}$$

The interference problem is to place bounds on a conditional probability $Pr(F_t|F_{c(t)})$. Thus we solve (8), where μ and μ^* are given by (13).

4.1. Example 2

We now give a small example. Suppose we have the following information:

$$Pr(x_1) \in [L_1,U_1]$$
$$Pr(x_1 \to x_2) \in [L_2,U_2]$$
$$Pr(x_2 \to x_3) \in [L_3,U_3]$$
$$Pr(x_3|x_1 \wedge x_2) \in [L_4,U_4]$$

where x_1, x_2 and x_3 are atomic propositions. We are interested in computing the bounds on $Pr(x_3)$, given the above information.

In this case:

$$F_1 = \{x_1\}, F_2 = \{x_1 \to x_2\}, F_3 = \{x_2 \to x_3\},$$
$$F_4 = \{x_3\}, F_{c(4)} = \{x_1 \wedge x_2\}, F_t = \{x_3\}$$

Table 2
Truth table for Example 2

(x_1,x_2,x_3)	Probability	$x_1 \to x_2$	$x_2 \to x_3$	$x_1 \wedge x_2$
(0,0,0)	p_1	1	1	0
(0,0,1)	p_2	1	1	0
(0,1,0)	p_3	1	0	0
(0,1,1)	p_4	1	1	0
(1,0,0)	p_5	0	1	0
(1,0,1)	p_6	0	1	0
(1,1,0)	p_7	1	0	1
(1,1,1)	p_8	1	1	1

Let $(p_1,\ldots,p_8)$ denote the probabilities of the eight possible worlds (x_1, x_2, x_3) as shown in Table 2.

Now let us write the usual LP for determining the bounds on the probability of the atomic proposition x_3.

$$\min/\max\ p_2+p_4+p_6+p_8$$

s.t.

$$\begin{pmatrix} 0&0&0&0&1&1&1&1\\ 1&1&1&1&0&0&1&1\\ 1&1&0&1&1&1&0&1\\ 0&0&0&0&0&0&-U_4&(1-U_4)\\ 0&0&0&0&1&1&1&1\\ 1&1&1&1&0&0&1&1\\ 1&1&0&1&1&1&0&1\\ 0&0&0&0&0&0&-L_4&(1-L_4)\\ 1&1&1&1&1&1&1&1 \end{pmatrix} \begin{pmatrix} p_1\\ p_2\\ p_3\\ p_4\\ p_5\\ p_6\\ p_7\\ p_8 \end{pmatrix} \begin{matrix} \le\\ \le\\ \le\\ \le\\ \ge\\ \ge\\ \ge\\ \ge\\ = \end{matrix} \begin{pmatrix} U_1\\ U_2\\ U_3\\ 0\\ L_1\\ L_2\\ L_3\\ 0\\ 1 \end{pmatrix}$$

$$p_i \ge 0,\ i=1,\ldots,8$$

Note that to solve this program it is not necessary to distinguish between the probabilities p_2 and p_4. The two columns they represent are identical, so that in this case it is sufficient to use seven probabilities.

We now show that probabilistic logic fits the general model of the previous section. Whenever an interval, $[L_i,U_i]$ is given for $Pr(F_i)$, the interval $[1-U_i,1-L_i]$ is implicity given for $Pr(\neg F_i)$. We therefore assume without harm that $\neg F_i$ belongs to the list $F_1,\ldots, F_h$ whenever F_i does; that is, the complement $\bar{S}_i$ of S_i belongs to the list $S_1,\ldots,S_h$ whenever S_i does. Since the probability mass attributed to a set is spread over the possible worlds in the set, empty sets cannot receive probability mass.

The intersections S_J, $J \in I$, are all minimal nonempty intersections of $S_1,\ldots,S_h$. A *minimal* intersection S_J is one that properly contains no nonempty intersection. More precisely,

$$I=\{J\subset\{1,\ldots,h\}|S_J\neq\emptyset \text{ and } S_{J'}\not\subset S_J \text{ for all } J'\subset\{1,\ldots,h\}\}$$

Due to the fact that $\bar{S}_i$ is among $S_i,\ldots,S_h$ whenever S_i is, the distinct intersections S_J for $J \in I$ partition the set of all possible worlds. Note that S_i may contain S_J when $i \notin J$. To distribute the probabilities $Pr(S_i)$ over the q_J's as in (1), we let

$$I(i)=\{J\in I|S_J\subset S_i\}. \tag{14}$$

We must now show that distributing probability over the variables q_J as in (1) results in the same problem (8) as distributing it over possible world probabilities q_v as in (13). We can do this by showing that (8) has the same columns in either case. Note first that for any $J \in I$, q_J occurs in a constraint or objective function of (8) with a given coefficient if and only all variables $q_{J'}$ with $S_{J'}=S_J$ occur with that coefficient. These variables can therefore be collapsed into one, say q_J, which represents the set S_J in the partition of possible worlds. But the column corresponding to q_J is identical to that for q_v, where v is any possible world in S_J. Thus the possible worlds v generate the same column as the sets J.

4.2. Example 2, continued

In this case the sets S_i's are as follows:

$$S_1=\{(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}$$

$$S_2=\{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,1,0),(1,1,1)\}$$

$$S_3=\{(0,0,0),(0,0,1),(0,1,1),(1,0,0),(1,0,1),(1,1,1)\}$$

$$S_4=\{(0,0,1),(0,1,1),(1,0,1),(1,1,1)\}$$

$$S_{c(4)}=\{(1,1,0),(1,1,1)\}$$

$$S_t=S_4$$

The minimal intersections S_J of the sets S_1, S_2, S_3, S_4, $S_{c(4)}$, $\bar{S}_1$, $\bar{S}_2$, $\bar{S}_3$, $\bar{S}_4$ and $\bar{S}_{c(4)}$ are given by:

$$S_{J_1}=\bar{S}_1\cap S_3\cap\bar{S}_4=\{(0,0,0)\}$$

$$S_{J_2}=\bar{S}_1\cap S_4=\{(0,0,1),(0,1,1)\}$$

$$S_{J_3}=\bar{S}_3\cap S_{c_{(4)}}=\{(0,1,0)\}$$

$$S_{J_4}=\bar{S}_2\cap\bar{S}_4=\{(1,0,0)\}$$

$$S_{J_5}=\bar{S}_2\cap S_4=\{(1,0,1)\}$$

$S_{J_6} = \bar{S}_3 \cap \bar{S}_{c(4)} = \{(1,1,0)\}$

$S_{J_7} = S_3 \cap S_{c(4)} = \{(1,1,1)\}$

We notice that the set I is given by:

$I = \{J_1, J_2, J_3, J_4, J_5, J_6, J_7\}$

Furthermore we have

$I(1) = \{J \in I | S_J \subset S_1\} = \{J_4, J_5, J_6, J_7\}$

$I(2) = \{J \in I | S_J \subset S_2\} = \{J_1, J_2, J_3, J_6, J_7\}$

$I(3) = \{J \in I | S_J \subset S_3\} = \{J_1, J_2, J_4, J_5, J_7\}$

$I(4) = \{J \in I | S_J \subset S_4\} = \{J_2, J_5, J_7\}$

$I(c(4)) = \{J \in I | S_J \subset S_{c(4)}\} = \{J_6, J_7\}$

$I(t) = \{J \in I | S_J \subset S_t\} = \{J_2, J_5, J_7\}$

If we formulate the program (8) using the descriptions of the columns (9), the problem below is obtained:

min/max $q_{J_2} + q_{J_5} + q_{J_7}$

s.t.

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -U_4 & (1-U_4) \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -L_4 & (1-L_4) \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} q_{J_1} \\ q_{J_2} \\ q_{J_3} \\ q_{J_4} \\ q_{J_5} \\ q_{J_6} \\ q_{J_7} \end{pmatrix} \begin{matrix} \le \\ \le \\ \le \\ \le \\ \ge \\ \ge \\ \ge \\ \ge \\ = \end{matrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ 0 \\ L_1 \\ L_2 \\ L_3 \\ 0 \\ 1 \end{pmatrix}$$

$q_{J_i} \ge 0, i = 1, \ldots, 7$

The above problem is almost the same as the one we found earlier using the possible world probabilities. The only difference is the notation and that two identical columns corresponding to the possible world probabilities p_2 and p_4 have been collapsed together.

The column description (14) is not computationally practical. Determining whether $J \in I$ involves solving a satisfiability problem to check whether S_J is empty. Thus we generate columns corresponding to possible worlds v rather than sets J. In the nonconditional problem (4) this yields columns $(y_0, y_1, \ldots, y_m)$ with $y_i = v(F_i)$. If (for instance) F_i is a logical clause $\bigvee_{j \in P} x_j \vee \bigvee_{j \in S} \neg x_j$, then y_i can be written as a pseudo-boolean function $y_i = 1 - \prod_{j \in P}(1 - x_j)\prod_{j \in N} x_j$. Thus the column generation subproblem, which is to minimize (12), becomes a pseudo-boolean optimization problem. The situation is similar for the conditional problem (8).

Whether generated by J's or possible worlds, some of the columns of (4) are identical. But two identical columns will never appear in the basis of the solution. Thus when several possible worlds generate the same column, only one of the worlds will in practice absorb all the probability attributed to the set S_J containing those worlds.

5. Probabilistic logic with unreliable sources

In probabilistic logic, the probabilities $Pr(F_i)$ of logical formulas F_i are delivered from a single evidence source. However, it might be useful to extend the model, such that it is possible to allow for more than one evidence source to supply probabilities (or estimates of these) to the logical formulas in question. This can be done in a straightforward manner.

Suppose there are K evidence sources, denoted by ES_k, $k = 1, \ldots, K$. If $K = 1$, the ordinary model for probabilistic logic is obtained. If $K \ge 2$, we instead get conditional probabilities $Pr(F_i | ES_k)$, for all i, all k. The interpretation of these probabilities is: the probability of F_i given that evidence source k is reliable. For each of the logical formulas F_i, there are K probabilities, namely the ones obtained from the K evidence sources.

Let S_i be the set of possible worlds in which F_i is true, and let R_k be the set of possible worlds, in which evidence source k is reliable (with certainty). Notice that here is a slight difference from probabilistic logic. The model not only has possible worlds in which some logical formulas are true but also possible worlds in which evidence sources are reliable. Suppose that evidence source k informs us that the probability of S_i is in the interval $[L_i^k, U_i^k]$. This gives rise to the set of linear constraints:

$$L_i^k \le Pr(F_i | ES_k) \le U_i^k,\ i, \ldots, m,\ k = 1, \ldots, K.$$

These constraints can be rewritten as

$$0 \leq Pr(S_i \cap R_k) - L_i^k Pr(R_k) \quad (15)$$

$$Pr(S_i \cap R_k) - U_i^k Pr(R_k) \leq 0 \quad (16)$$

It is of course also possible for each evidence source to specify conditional probabilities. Suppose that evidence source k informs us that the conditional probability of F_i given $F_{c(i)}$ belongs to some interval $[L_{ic(i)}^k, U_{ic(i)}^k]$. The following set of constraints is then obtained:

$$L_{ic(i)}^k \leq Pr\left(F_i | F_{c(i)}, ES_k\right) \leq U_{ic(i)}^k,$$
$$i = 1,\dots,m, k = 1,\dots,K.$$

These constraints can be rewritten as follows:

$$0 \leq Pr\left(S_i \cap S_{c(i)} \cap R_k\right) - L_{ic(i)}^k Pr\left(S_{c(i)} \cap R_k\right). \quad (17)$$

$$Pr\left(S_i \cap S_{c(i)} \cap R_k\right) - U_{ic(i)}^k Pr\left(S_{c(i)} \cap R_k\right) \leq 0. \quad (18)$$

In ordinary probabilistic logic $K = 1$, it is implicitly assumed that the evidence source is reliable. This need of course not be the case. Therefore, in addition to the above mentioned conditional probabilities, it is possible to specify probability intervals $[L^k, U^k]$, $k = 1,\dots,K$, indicating to which degree the different evidence sources are reliable. This gives rise to the set of constraints:

$$L^k \leq Pr(ES_k) \leq U^k, \ k = 1,\dots,K. \quad (19)$$

If one believes in some of the evidence sources with certainty, the corresponding probability intervals should overlap, since otherwise the model is inconsistent. We cannot with certainty believe, for instance, that a probability is in the interval [0.2, 0.3] and at the same time with certainty believe that it is in the interval [0.5, 0.6].

As in ordinary probabilistic logic, we have:

$$Pr(S_i \cap R_k) = \sum_{v(S_i \cap R_k) = 1} q_v$$

where q_v denotes the probability of world v.

The inference problem is to place bounds on a conditional probability $Pr(F_t | F_{c(t)})$. As in ordinary probabilistic logic, we solve a fractional linear program similar to,

$$\min/\max \frac{Pr\left(S_t \cap S_{c(t)}\right)}{Pr\left(S_{c(t)}\right)}$$

$s.t.$

(15),(16),(17),(18),(19).

The sets $S_1,\dots,S_h$ are those defined above: the sets of possible worlds in which the given formulas $F_i \wedge ES_k, F_i \wedge F_{c(i)} \wedge ES_k, ES_k, i = 1,\dots,m$, $k = 1,\dots,K$, respectively, are true.

The probabilities of these formulas can be expressed in terms of possible worlds as explained in Section 4.

We see that the only difference from probabilistic logic is that we now have possible worlds in which some formulas are true and possible worlds in which an evidence source is reliable (with certainty). Instead of just having probabilities of formulas, we have probabilities conditioned on evidence sources. Furthermore, it is possible to state the probability of the reliability of some evidence source. If any of the probabilities are unknown, they are simply left unspecified.

So, everything in this section has been formulated in the same way as was done in the section on probabilistic logic. In particular, the model falls into the general framework. The column generation procedure is as in probabilistic logic.

6. Dempster-Shafer theory

In Dempster-Shafer theory there are several evidence sources, indexed by $k = 1,\dots,K$. Each evidence source k distributes a probability mass of one over a family $\{S_i | i \in H_k\}$ of distinct sets of possible outcomes. The index sets H_k are disjoint, but a set S_i in one family may be equal to a set S_i in another family. The union of the S_i' s is the *frame of discernment*, denoted Θ.

For each i $\in H_k$, we interpret the mass $\mu(S_i)$ to be the value $m_k(S_i)$ of the *basic probability function* m_k, which indicates the strength of k's evidence that the actual outcome lies in S_i. Thus we have constraints (2) with $L_i = U_i = m_{k(i)}(S_i)$, where i $\in H_{k(i)}$. Although $\Sigma_{i \in H_k} m_k(S_i) = 1$ for each evidence source k, some of its mass can be assigned to a set representing the universe Θ, which contains all possible situations. This mass represents evidence that supports no particular proposition.

When there are K evidence sources, the intersections S_J are associated with cells of a K-dimensional cube. Each cell corresponds to an index set J that contains for each $k \in \{1,\ldots,K\}$ an index $i \in H_k$ representing the "coordinate" of the cell along dimension k. Since S_J may be the same for different J's, the same S_J may be associated with several cells. The mass $\mu(S_i)$ for $i \in H_k$ is distributed over the masses q_J of all cells J whose k-th coordinate is i; i.e., all cells J with $i \in J$. Thus we have (1), with $I(i) = \{J \in I | i \in J\}$.

Classical Dempster-Shafer theory uses the particular distribution dictated by Dempster's combination rule:

$$q_J = \prod_{i \in J} m_{k(i)}(S_i), \tag{20}$$

which assumes that the evidence sources are in some sense independent. But we will allow any distribution observing (1), since independence assumptions may be unjustified. The classical theory can be obtained by adding the constraints (20) to the model.

This distribution scheme has the curious result that an empty set S_J may receive probability mass. But it also implies that the constraints (2) and (3) always have a feasible solution. This can be seen by noting that the probabilities dictated by Dempster's combination rule satisfy them.

Whenever $S_J \subset S_t$, evidence that S_J contains the actual outcome adds to the evidence that S_t does. So we interpret $\mu^*(S_t)$ as the sum of the masses of all $S_J \subset S_t$. Thus we have (5) with $I^*(t) = \{J | S_J \subset S_t\}$.

Since empty S_J's receive mass, this mass is ignored and the rest renormalized so that it sums to one. Thus the inference problem is to obtain bounds on

$$Bel(S_t) = \frac{\mu^*(S_t)}{1 - \mu^*(\varnothing)}, \tag{21}$$

where "*Bel*" is Shafer's notation. We therefore obtain the fractional programming problem (4).

6.1. Example 3

Suppose that a detective is investigating a burglary of a shop. From one source he gets evidence supporting the beliefs that the thief is left-handed and that he is not an insider. From another source he gets evidence supporting the beliefs that the theft was an inside job and that the thief is right-handed. One of the clerks in the shop is left-handed, and the detective must decide with what certainty he can accuse the clerk.

For solving this problem we define two atomic propositions:

x_1: The thief is left – handed.

x_2: The thief is an insider.

A possible outcome is indicated by a pair of truth values (x_1, x_2), and the frame of discernment consists of the four possible pairs. We are given the following six quantities:

$m_1(S_1) = m_1(\{(1,0),(1,1)\})$

(source 1 evidence for x_1)

$m_1(S_2) = m_1(\{(0,0),(1,0)\})$

(source 1 evidence for $\neg x_2$)

$m_1(S_3) = m_1(\Theta)\ (= 1 - m_1(S_1) - m_1(S_2))$

$m_2(S_4) = m_2(\{(0,1),(1,1)\})$

(source 2 evidence for x_2)

$m_2(S_5) = m_2(\{(0,0),(0,1)\})$

(source 2 evidence for $\neg x_1$)

$m_2(S_6) = m_2(\Theta)(= 1 - m_2(S_4) - m_2(S_5))$

We construct the following table of intersections:

Table 3
Intersections

$S_6 = \Theta$	{(1,0),(1,1)}	{(0,0),(1,0)}	Θ
$S_5 = \{(0,0),(0,1)\}$	$\varnothing$	{(0,0)}	{(0,0),(0,1)}
$S_4 = \{(0,1),(1,1)\}$	{(1,1)}	$\varnothing$	{(0,1),(1,1)}
	$S_1 = \{(1,0),(1,1)\}$	$S_2 = \{(0,0),(1,0)\}$	$S_3 = \Theta$

Notice that the empty set $\emptyset$ may occur in the table, and that the same set may occur several times. In this particular case the empty set $\emptyset$ occurs twice.

Now let us associate a mass q_J with each particular cell:

Table 4
q_J's

$m_2(\Theta)$	q_{J_1}	q_{J_4}	q_{J_7}
$m_2(S_5)$	q_{J_2}	q_{J_5}	q_{J_8}
$m_2(S_4)$	q_{J_3}	q_{J_6}	q_{J_9}
	$m_1(S_1)$	$m_1(S_2)$	$m_1(\Theta)$

Each $m_k(S_i)$ is distributed over the corresponding row or column. For instance, $m_1(S_1) = q_{J1} + q_{J2} + q_{J3}$.

We wish to determine the mass $\mu^*(\{1,1)\})$ that can be associated with the proposition $x_1 \wedge x_2$ that the thief is a lefthanded insider. The normalized objective function of (4) is given by $q_{J_3}/(1 - q_{J_2} - q_{J_6})$ (in this particular case, $\mu^*(\emptyset) = q_{J2} + q_{J6}$). The resulting program (4) is:

$$\min/\max \frac{q_{J_3}}{1 - q_{J_2} - q_{J_6}}$$

s.t.

$$q_{J_1} + q_{J_2} + q_{J_3} = m_1(S_1)$$
$$q_{J_4} + q_{J_5} + q_{J_6} = m_1(S_2)$$
$$q_{J_3} + q_{J_6} + q_{J_9} = m_2(S_4)$$
$$q_{J_2} + q_{J_5} + q_{J_8} = m_2(S_5)$$
$$\Sigma_{i=1}^{9} q_{J_i} = 1$$
$$q_{J_i} \geq 0,\ i = 1,\ldots,9$$

The classical Dempster-Shafer theory associates the mass

$$m_1(S_1)m_2(S_4)$$
$$/[1 - m_1(S_1)m_2(S_5) - m_1(S_2)m_2(S_4)]$$

with the proposition $x_1 \wedge x_2$.

Assuming for simplicity that all intersections S_J are nonempty so that $\mu^*(\emptyset) = 0$, the columns $(y_0, y_1, \ldots, y_m)$ of (4) satisfy

$$\sum_{i \in H_k} y_i = 1,\ k = 1,\ldots,K. \tag{22}$$

The objective function coefficient y_o is 1 if and only if $\bigcap_{y_i=1} S_i \subset S_t$, which is to say $\wedge_{yi=1} F_i \supset F_t$. If the F_i's are formulas of propositional logic, additional constraints and 0-1 variables representing atomic propositions can be used to define y_0 in terms of $y_1, \ldots, y_m$, using well-known methods [14]. The column generation subproblem becomes an integer programming problem: minimize (12) subject to (22) and the additional constraints. When $\mu^*(\emptyset) \neq 0$, the linearized version of (4) can be similarly treated.

7. Second-Order probabilistic logic

Second-order probabilistic logic assigns probability distributions to $Pr(F_j|F_{c(j)})$, rather than specifying $Pr(F_j|F_{c(j)})$ directly. These second-order distributions are approximated by specifying the probability that $Pr(F_j|F_{c(j)})$ lies in each of several intervals $[0,\pi_i]$. Probabilistic information is therefore given in the form,

$$L_i \leq Pr\left(\frac{Pr(F_{j_i} \wedge F_{c(j_i)})}{Pr(F_{c(j_i)})} \leq \pi_i\right) \leq U_i,\ i = 1,\ldots,m. \tag{23}$$

The propositions F_j again belong to propositional logic and contain atomic propositions $x_1,\ldots,x_n$. In general, probabilities from a less reliable source will have a more dispersed second order distribution.

The first-order probability space consists of all probability distributions $(\pi_1,\ldots,\pi_{2^n})$ over the possible worlds v. Each constraint in (23) places limits on the probability mass assigned to a region S_i of this space. Specifically, S_i is the half-space in which $Pr(F_{j_i} \wedge F_{c(j_i)}) \leq \pi_i Pr(F_{c(j_i)})$, or

$$\sum_{\substack{v(F_{j_i})=1 \\ v(F_{c(j_i)})=1}} p_v \leq \pi_i \sum_{v(F_{c(j_i)})=1} p_v. \tag{24}$$

We interpret $\mu(S_i)$ as the second-order probability that $Pr(F_{j_i}|F_{c(j_i)}) \leq \pi_i$. We can again suppose that $\bar{S}_i$ belongs to the list $S_1,\ldots,S_h$ whenever S_i does. Thus the set of all minimal nonempty intersections S_J of halfspaces partitions probability space into polyhedral regions. The probability mass $Pr(S_i)$ of a halfspace is the sum of the masses q_J of the polyhedra S_J in it. Thus we have (1) with $I(i)=\{J|S_J \subset S_i\}$.

The inference problem is to find bounds on $\mu^*(S_t)=\mu(S_t)$ by solving (4), which is a linear programming problem because $\mu^*(\emptyset)=0$.

It remains to find a computationally practical way to implement the column generation scheme (6) when $I(i)=\{J|S_J \subset S_i\}$. This can be done via mixed integer programming. From (24), each S_i consists of the vectors p in probability space satisfying the i-th constraint of

$$-y_i < \sum_{\substack{v(F_{j_i})=1 \\ v(F_{c(j_i)})=1}} p_v - \pi_i \sum_{v(F_c(j_i))=1} p_v \le 1-y_i,$$

$$i=1,\dots,m, \qquad (25)$$

when $y_i=1$, and its complement consists of those satisfying this same constraint when $y_i=0$. Thus the sets S_J for $J \in I$ are precisely the nonempty solution sets of (25) over all 0-1 vectors $y=(y_1,\dots,y_m)$. The column generation subproblem is therefore to minimize (12) subject to (25) and

$$\sum_v p_v = 1, p_v \ge 0, \text{ all } v,$$

where $y_0 = y_t$.

7.1. Example 4

Suppose we have the following set of second-order probabilities:

$$Pr(Pr(x_1) \le \pi_1) \in [L_1,U_1]$$

$$Pr(Pr(x_1 \to x_2) \le \pi_2) \in [L_2,U_2]$$

$$Pr(Pr(x_1 \to x_2) \le \pi_3) \in [L_3,U_3]$$

$$Pr(Pr(x_1 \to x_2) \le \pi_4) \in [L_4,U_4]$$

$$Pr(Pr(x_3|x_1 \wedge x_2) \le \pi_5) \in [L_5,U_5]$$

Note that the distribution of $Pr(x_1 \to x_2)$ is somewhat more accurately specified than the others, because three cumulative second-order probabilities are given. We want to determine the bounds on the probability $Pr(Pr(x_3) \le \pi_5)$. Notice that all the π's are given. The sets, S_i, $1 \le i \le 4$, are given as follows:

$$S_1:\{p \in \Re^8_+ | \Sigma^8_{i=1} p_i = 1,\ p_5+p_6+p_7+p_8 \le \pi_1\}$$

$$S_2:\{p \in \Re^8_+ | \Sigma^8_{i=1} p_i = 1, p_1+p_2+p_3+p_4+p_7+p_8 \le \pi_2\}$$

$$S_3:\{p \in \Re^8_+ | \Sigma^8_{i=1} p_i = 1, p_1+p_2+p_3+p_4+p_7+p_8 \le \pi_3\}$$

$$S_4:\{p \in \Re^8_+ | \Sigma^8_{i=1} p_i = 1, p_1+p_2+p_3+p_4+p_7+p_8 \le \pi_4\}$$

$$S_5:\{p \in \Re^8_+ | \Sigma^8_{i=1} p_i = 1,\quad p_8 \le \pi_5(p_7+p_8)\},$$

where $\Re^8_+$ is the nonnegative orthant of R^8. The set S_t is given by:

$$S_t:\{p \in \Re^8_+ | \Sigma^8_{i=1} p_i = 1,\ p_2+p_4+p_6+p_8 \le \pi_5\}$$

It is not particularly easy to see what the nonempty intersections S_J of the halfspaces S_1, S_2, S_3, S_4 and S_t are. However, the columns of (4) can be found using the following system of linear constraints:

$$-y_1 < p_5+p_6+p_7+p_8-\pi_1 \le 1-y_1$$

$$-y_2 < p_1+p_2+p_3+p_4+p_7+p_8-\pi_2 \le 1-y_2$$

$$-y_3 < p_1+p_2+p_3+p_4+p_7+p_8-\pi_3 \le 1-y_3$$

$$-y_4 < p_1+p_2+p_3+p_4+p_7+p_8-\pi_4 \le 1-y_4$$

$$-y_5 < p_8+\pi_5(p_7+p_8) \le 1-y_5$$

$$-y_0 < p_2+p_4+p_6+p_8-\pi_5 \le 1-y_0$$

$$\Sigma^8_{i=1} p_i = 1$$

$$p_i \ge 0, i=1,\dots 8, y_i \in \{0,1\}, i=0,\dots,5$$

The columns of (4) are the binary vectors $(y_0, y_1, y_2, y_3, y_4, y_5)$ that are feasible for the above constraint set. An approximate second-order distribution for $Pr(x_3)$ can be found by solving this problem for several values of π_5.

8. Belief functions

Dempster-Shafer theory derives the plausibility $Bel(S_t)$ of S_t by combining evidence from several sources. Each source k provides support for several sets S_i that is measured by a basic probability function m_k.

A variation on this approach is to suppose that the evidence for the sets S_i is combined in advance, outside the mechanism of the theory. This provides an estimate $Bel(S_i)$ of the plausibility of each S_i. The goal is to find what values of $Bel(S_t)$ are consistent with these estimates.

Thus we interpret $\mu(S_i) = \mu^*(S_i)$ to be $Bel(S_i)$. Again any evidence for a subset of S_t is evidence for S_t. We therefore postulate an underlying basic probability function $m(S_i)$ that measures evidence in favour of the particular set S_i, with $Bel(S_i) = \Sigma_{S_j \subset S_i} t_t \; m(S_j)$. This means that we distribute the mass $\mu(S_i) = Bel(S_i)$ over the variables $q_J = q_{\{j\}} = m(S_j)$ for all subsets S_j of S_i. So the "intersections" S_J are simply the sets S_j. The distribution formula is (1) with $I(i) = \{\{j\} | S_j \subset S_i\}$.

If the precise value of $Bel(S_i)$ is given for all S_i, the underlying basic probability function is determined by the inclusion-exclusion formula,

$$m(S_i) = \sum_{S_j \subset S_i} (-1)^{|S_i \setminus S_j|} Bel(S_j).$$

But if $Bel(S_i)$ is only partially specified, several basic probability functions are possible, and $Bel(S_t)$ may be restricted to a range but not precisely determined.

If L_i, U_i are the bounds placed on $Bel(S_i) = \mu(S_i)$, we have the constraints (2). To obtain bounds on $Bel(S_t) = \mu^*(S_t)$ we solve (4), which is linear because $\mu^*(\emptyset) = 0$. Since this problem contains only one column for each S_i, column generation should not in general be necessary. Also it should normally be easy to determine whether $S_j \subset S_i$ (i.e., whether F_j implies F_i), since the given propositions F_i should normally be simple.

9. Conclusion

In this paper we have demonstrated how several logics for reasoning under uncertainty fit into the same framework. Each admits a linear programming model of essentially the same structure, except that the different logics are implemented with different column generation procedures. The column generation procedures include pseudo-boolean optimization, integer programming and mixed integer programming. The logics that we have been able to show fit into this general framework include probabilistic logic, probabilistic logic with unreliable sources of information, Dempster-Shafer theory, second order probabilistic logic and a simple logic of belief functions.

An important question is how well the column generation scheme will work in practice. It has been demonstrated to work very well in the case of probabilistic logic [15], and therefore also for the special version of probabilistic logic with unreliable sources of information. Further computational experience is needed to determine how well it will work in Dempster-Shafer theory and second-order probabilistic logic.

References

[1] Andersen, K.A., and J.N. Hooker, Bayesian logic, Decision Support Systems 11 (1994) 191–210.

[2] Boole, G., An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities. Dover Publications (New York, 1951). Original work published 1854.

[3] Boole, G., Studies in Logic and Probability, ed. by R. Rhees, Watts and Co (London) and Open Court Publishing Company (La Salle, Illinois, 1952).

[4] Charnes, A., and W.W. Cooper, Programming with linear fractionals, Naval Research Logistics Quarterly 9 (1962) 181–186.

[5] Chen, S.S., Some extensions of probabilistic logic, in J.F. Lemmer and L.N. Kanal, eds., Uncertainty in Artificial Intelligence 2, North-Holland (1988).

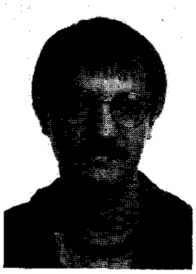
[6] Dirickx, I.M.I., and L.P. Jennergren, Systems Analysis by Multilevel Methods: With Applications to Economics and Management. Wiley, Chichester (1979).

[7] Dubois, D., and H. Prade, The principle of minimum specificity as a basis for evidential reasoning, Uncertainty in Knowledge-Based Systems, Lecture Notes in Computer Science 286 (1986) 75–84

[8] Dubois, D., and H. Prade, A tentative comparison of numerical approximate reasoning methodologies, International Journal Man-Machine Studies 27 (1987) 149–183.

[9] Georgakopolous, G., D. Kavvadias and C.H. Papadimitriou, Probabilistic satisfiability, Journal of Complexity 4 (1988) 1–11.

[10] Grosof, B.N., An inequality paradigm for probabilistic reasoning, in J.F. Lemmer and L.N. Kanal, eds., Uncertainty in Artificial Intelligence 1, North-Holland (1986).
[11] Grosof, B.N., Non-monotonicity in probabilistic knowledge, in J.F. Lemmer and L.N. Kanal, eds., Uncertainty in Artificial Intelligence 2, North-Holland (1986).
[12] Hailperin, T., Boole's Logic and Probability, Studies in Logic and the Foundations of Mathematics v. 85, North-Holland (1976).
[13] Hailperin, T., Probability logic, Notre Dame Journal of Formal Logic 25 (1984) 198–212.
[14] Hooker, J.N., A quantitative approach to logical inference, Decision Support Systems 4 (1988) 45–69.
[15] Jaumard, B., P. Hansen and M.P. Aragaö, Column generation methods for probabilistic logic, ORSA Journal on Computing 3 (1991) 135–148.
[16] McLeish, M., Probabilistic logic: some comments and possible use for nonmonotonic reasoning, in J.F. Lemmer and L.N. Kanal, eds., Uncertainty in Artificial Intelligence 2, North-Holland (1986).
[17] McLeish, M., Nilsson's probabilistic entailment extended to Dempster-Shafer theory, in Uncertainty in Artificial Intelligence 3 (1989) 23–34.
[18] Nilsson, N.J., Probabilistic logic, Artificial Intelligence 28 (1986) 71–87.
[19] Paaß, G., Probabilistic logic, in Non-standard Logics for Automated Reasoning, ed. P. Smets em et al., Academic Press (New York, 1988) 213–251.
[20] Shafer, G., A Mathematical Theory of Evidence, Princeton University Press (1976).



Kim Allan Andersen is associate professor at the Mathematical Institute at Aarhus University. He received a masters degree in mathematics and economics from Aarhus University in 1984 and a Ph.D. in operations research from Aarhus University in 1990. His research interests lie in the application of mathematical programming to logical problems, as well as integer programming.

John Hooker is Professor in the Graduate School of Industrial Administration at Carnegie Mellon University. He is interested in new modelling paradigms for operations research, including logico-mathematical models as well as empirical models, and in the mathematical structure of propositional, probabilistic and other logics.