On probability intervals and their updating

Linda van der Gaag

RUU-CS-90-22
June 1990

Utrecht University

Department of Computer Science

Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

On probability intervals and their updating

Linda van der Gaag

Technical Report RUU-CS-90-22
June 1990

Department of Computer Science
Utrecht University
P.O.Box 80.089
3508 TB Utrecht
The Netherlands

ISSN:0924-3275

On Probability Intervals and Their Updating

Linda van der Gaag
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

Abstract

The adversaries of probability theory for dealing with uncertainty in AI systems often argue that it is not expressive enough to distinguish between uncertainty and ignorance due to incompleteness of information. Probability intervals, however, have proven to be suitable for expressing incompleteness of information. In this paper, we present a new method for computing such intervals from a partial specification of a joint probability distribution. We will show that our method allows for the successive updating of probability intervals as evidence becomes available.

1 Introduction

When building knowledge-based systems it becomes evident that in many real-life domains expert knowledge is not precisely defined, but instead is of an imprecise nature. Yet, human experts typically are able to form judgements and take decisions from uncertain, incomplete and even contradictory information. In order to be useful in an environment in which only such imprecise information is available, a knowledge-based system has to capture and exploit not only the highly-specialized expert knowledge, but the uncertainties that go with the represented pieces of information as well. Researchers in artificial intelligence therefore have sought methods for representing uncertainty and have developed reasoning procedures for manipulating uncertain knowledge.

As probability theory is one of the oldest mathematical theories concerning uncertainty, it is no wonder that this formal theory was chosen as the first point of departure for the pioneering work in the field of plausible reasoning. Applying probability theory in a knowledge-based setting, however, soon proved to be problematic. One of the problems in applying probability theory in a model for handling uncertainty in a knowledge-based system, for example, is the difficulty of obtaining a joint probability distribution on the problem domain: often only a few probabilities are known or can be estimated by an expert in the field. In such a case therefore, one is confronted with the problem of having to derive mathematically sound statements concerning probabilities of interest from only a partial and often even inconsistent specification of a joint probability distribution. The early models developed in the 1970s, such as for example the well-known certainty factor model, [Shortliffe84], were able to handle this problem, although not in a mathematically sound way.

Adversaries of probability theory for dealing with uncertainty in AI systems have argued that this theory is not expressive enough to discern between uncertainty and ignorance due to incompleteness of information. In [Pearl88], however, J. Pearl has hinted at the desirability of probability intervals for expressing incompleteness of information. In this paper, we present a framework for computing such intervals for probabilities of interest from an incomplete but consistent set of probabilities; the general idea of our approach is to take the initially given probabilities as defining constraints on a yet unknown joint probability distribution. The problem of computing probability intervals has already been addressed before by several authors, see for example [Cooper86] and [Nilsson86]. In the sequel, we will comment on these papers. Here, we simply mention that our approach differs from the mentioned ones mainly by allowing for the successive updating of probability intervals as evidence becomes available.

After providing some preliminaries in Section 2, we introduce in Section 3 the notion of a partial specification of a joint probability distribution and show how linear programming techniques may be used for computing intervals for probabilities of interest from such a partial specification. Section 4 discusses an algorithm for updating a partial specification of a joint probability distribution, taking the conceptual framework from the preceding section for an invariant.

2 Preliminaries

In this subsection, we provide some preliminaries concerning probability theory, departing from an algebraic point of view.

In an expert system, knowledge concerning the problem domain usually is represented in a special knowledge-representation formalism such as for example the production-rule formalism, [Lucas90]. In this paper we do not consider these knowledge-representation schemes nor do we discuss the reasoning methods associated with these formalisms. Here, we assume that knowledge is simply represented in statistical variables. We assume that these variables can only take one of two values, thus allowing to view them as logical, propositional variables. The generalization to variables with discrete multiple values, however, is straightforward.

In the following definition the notion of a Boolean algebra of propositions is introduced.

Definition 2.1 *A* Boolean algebra $\mathcal{B}$ *is a set of elements with two binary operations* $\wedge$ *(conjunction) and* $\vee$ *(disjunction), a unary operation* $\neg$ *(negation) and two constants* $false$ *and* $true$ *which (by equality according to logical truth tables) adhere to the usual axioms. On a Boolean algebra* $\mathcal{B}$ *we define a* partial order $\preccurlyeq$ *as follows: for any* $x_1, x_2 \in \mathcal{B}$*, we say that* $x_1 \preccurlyeq x_2$ *if* $x_2 = x_1 \vee x_2$ *or (equivalently) if* $x_1 = x_1 \wedge x_2$*. A subset of elements* $\mathcal{G} = \{g_1, \ldots, g_n\}$*,* $n \geq 1$*, of a Boolean algebra* $\mathcal{B}$ *is said to be a set of* generators *for* $\mathcal{B}$ *if each element of* $\mathcal{B}$ *can be represented in terms of the elements* $g_i \in \mathcal{G}$*,* $i = 1, \ldots, n$*, and the operations* $\wedge$*,* $\vee$ *and* $\neg$*. A set of generators* $\mathcal{G}$ *for* $\mathcal{B}$ *is said to be* free *if every mapping of elements of* $\mathcal{G}$ *into an arbitrary Boolean algebra* $\mathcal{B}'$ *can be extended to a homomorphism of* $\mathcal{B}$ *into* $\mathcal{B}'$*. A Boolean algebra* $\mathcal{B}$ *is* free *if it has a finite set* $\mathcal{A} = \{a_1, \ldots, a_n\}$*,* $n \geq 1$*, of free generators; we say that* $\mathcal{B}$ *is (finitely)* generated *by* $\mathcal{A}$*. We use* $\mathcal{B}(a_1, \ldots, a_n)$ *to denote the free Boolean algebra* $\mathcal{B}$ *generated by* $\mathcal{A}$*; from now on, we will refer to* $\mathcal{A}$ *as the set of* atomic propositions *and to* $\mathcal{B}$ *as the* Boolean algebra of propositions.

We introduce the notion of a probability distribution on a Boolean algebra of propositions.

Definition 2.2 *Let $\mathcal{B}$ be a Boolean algebra of propositions as defined above. Let Pr be a function $Pr\colon \mathcal{B} \to [0,1]$ such that*

1. *Pr is* positive, *that is, for all $x \in \mathcal{B}$, we have $Pr(x) \geq 0$, and furthermore $Pr(false) = 0$,*

2. *Pr is* normed, *that is, we have $Pr(true) = 1$, and*

3. *Pr is* additive, *that is, for all $x_1, x_2 \in \mathcal{B}$, if $x_1 \wedge x_2 = false$ then $Pr(x_1 \vee x_2) = Pr(x_1) + Pr(x_2)$.*

Pr is called a probability distribution *on $\mathcal{B}$.*

It can easily be shown that the probability of an event is equivalent to the probability of the truth of the proposition asserting the occurrence of the event: we have that a probability distribution Pr on a Boolean algebra of propositions $\mathcal{B}$ has the usual properties. In the sequel, we will often take the point of view of a Boolean algebra of propositions $\mathcal{B}(a_1, \ldots, a_n)$ as a sample space being 'spanned' by a set of statistical variables A_i, $i = 1, \ldots, n$, each taking values from $\{a_i, \neg a_i\}$. Conditional probabilities are defined as customary.

The following lemma states how we may compute a revised probability distribution given a specific piece of evidence.

Lemma 2.3 *Let $\mathcal{B}(a_1, \ldots, a_n)$, $n \geq 1$, be a Boolean algebra of propositions. Let Pr be a joint probability distribution on $\mathcal{B}(a_1, \ldots, a_n)$. Then, for a given $e \in \{a_i, \neg a_i\}$, $1 \leq i \leq n$, the conditional probabilities $Pr(x|e)$ for all $x \in \mathcal{B}(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$ define a probability distribution on $\mathcal{B}(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$.*

The probability distribution defined by the probabilities $Pr(x|e)$ as in the preceding lemma is called the *updated probability distribution* given e. We will use the phrase to *update* a probability distribution to denote the process of computing the updated probability distribution given some piece of evidence. For updating a joint probability distribution for successively obtained evidence the preceding lemma may be applied recursively.

3 Partial Specification of a Joint Probability Distribution

In this section we deal with the situation in which only a partial specification of a joint probability distribution is available for making mathematically sound statements concerning probabilities of interest. Such a partial specification of a joint probability distribution may for example have been obtained from a domain expert who assessed some probabilities of concern. In the following definition the notion of a partial specification is formally defined.

Definition 3.1 *Let $\mathcal{B}$ be a Boolean algebra of propositions. A* partial specification of a joint probability distribution *on $\mathcal{B}$ is a total function $P\colon \mathcal{C} \to [0,1]$ where $\mathcal{C} \subseteq \mathcal{B}$. A partial specification $P\colon \mathcal{C} \to [0,1]$ is* consistent *if there exists at least one joint probability distribution Pr on $\mathcal{B}$ such that $Pr\,|_{\mathcal{C}} = P$; otherwise, P is said to be* inconsistent. *Furthermore, we say that P (uniquely)* defines *Pr, or alternatively that P is a* definition *for Pr, if Pr is the only joint probability distribution on $\mathcal{B}$ such that $Pr\,|_{\mathcal{C}} = P$.*

In this section we will often use the incomplete phrase partial specification to denote a partial specification of a joint probability distribution on a given Boolean algebra of propositions as long as ambiguity cannot occur.

The problem of determining the probability of a given event from a partial specification of a joint probability distribution has already been investigated as early as halfway the nineteenth century by G. Boole, [Boole54]. Boole's work on probability theory, however, has received little attention. In our opinion Boole's ideas have become topical once more in the context of reasoning with uncertainty in knowledge-based systems. In fact, our method for deriving mathematically sound statements concerning probabilities of interest is based on Boole's work; we have used [Hailperin86] as a guide to the work of Boole.

In Section 3.1 we will present a method for computing bounds on probabilities of interest from a consistent partial specification of a joint probability distribution. For ease of exposition, we will assume that such a partial specification only comprises prior probabilities; in Section 3.2 it will be shown that the method we have developed can deal with conditional probabilities in the same way in which it handles prior ones.

3.1 Computing Bounds on Probabilities of Interest

In the following definition we introduce the notion of a basis for a joint probability distribution. This notion will play an important role in the remainder of this section.

Definition 3.2 *Let $\mathcal{B}$ be a Boolean algebra of propositions. A set $\mathcal{C} \subseteq \mathcal{B}$ is called a* basis *for a joint probability distribution on $\mathcal{B}$ if for any consistent partial specification $P\colon \mathcal{C} \to [0,1]$ defined on $\mathcal{C}$, there exists a joint probability distribution Pr on $\mathcal{B}$ such that P is a definition for Pr.*

It will be evident that in a Boolean algebra $\mathcal{B}$ we can identify several different bases. However, one basis in specific will be shown to have some convenient properties.

Definition 3.3 *Let $\mathcal{A} = \{a_1, \ldots, a_n\}$, $n \geq 1$, be a set of atomic propositions and let $\mathcal{B}(a_1, \ldots, a_n)$ be the Boolean algebra of propositions generated by $\mathcal{A}$. We define the set $\mathcal{B}_0 \subseteq \mathcal{B}(a_1, \ldots, a_n)$ such that $\mathcal{B}_0 = \{\bigwedge_{i=1}^{n} L_i \mid L_i = a_i \text{ or } L_i = \neg a_i, a_i \in \mathcal{A}\}$.*

Note that the set $\mathcal{B}_0$ has 2^n elements, essentially being the 'smallest' ones from $\mathcal{B}(a_1, \ldots, a_n)$. It can easily be shown that $\mathcal{B}_0$ indeed is a basis. Furthermore, note that by definition we have that each consistent partial specification $P\colon \mathcal{B}_0 \to [0,1]$ defined on $\mathcal{B}_0$ uniquely defines a joint probability distribution Pr on the entire Boolean algebra of propositions $\mathcal{B}(a_1, \ldots, a_n)$.

It will be evident that a basis for a joint probability distribution on a Boolean algebra of propositions with n free generators, $n \geq 1$, has at least $2^n - 1$ elements. Note that it does not follow that when less than $2^n - 1$ probabilities have been specified initially, they cannot define a joint probability distribution on a Boolean algebra of propositions $\mathcal{B}$ with n free generators uniquely: it may well be that a consistent partial specification P defined on a subset $\mathcal{C} \subseteq \mathcal{B}$ with $|\mathcal{C}| \leq 2^n - 1$ is a definition for a joint probability distribution on $\mathcal{B}$. We say that a set $\mathcal{C}$ is a minimal basis for a joint probability distribution on $\mathcal{B}$ if $\mathcal{C}$ is a basis as defined in Definition 3.2 and if in addition we have $|\mathcal{C}| = 2^n - 1$. The basis $\mathcal{B}_0 \subseteq \mathcal{B}$ contains just one element too many to be a minimal basis. For, since the Boolean algebra of propositions $\mathcal{B}$ is finite we have for each joint probability distribution Pr on

$\mathcal{B}$ that any probability $Pr(b_i)$, $b_i \in \mathcal{B}_0$, can be expressed in terms of the probabilities of all other elements from $\mathcal{B}_0$: $Pr(b_i) = 1 - \sum_{j=1,j\neq i}^{2^n} Pr(b_j)$. The deletion of an arbitrary element from $\mathcal{B}_0$ therefore yields a minimal basis.

The following three lemmas state some general properties concerning the basis $\mathcal{B}_0$.

Lemma 3.4 *Let $\mathcal{B}$ be a Boolean algebra of propositions with n free generators, $n \geq 1$. Let the basis $\mathcal{B}_0 \subseteq \mathcal{B}$ be defined according to Definition 3.3 and let its elements be enumerated as b_i, $i = 1, \ldots, 2^n$. Then, for any joint probability distribution Pr on $\mathcal{B}$ we have*

$$\sum_{i=1}^{2^n} Pr(b_i) = 1$$

The probabilities $Pr(b_i)$ for the elements $b_i \in \mathcal{B}_0$, $i = 1, \ldots, 2^n$, as mentioned in the preceding lemma will be called the *constituent probabilities* of Pr.

Lemma 3.5 *Let $\mathcal{B}$ be a Boolean algebra of propositions with n free generators, $n \geq 1$. Let the basis $\mathcal{B}_0 \subseteq \mathcal{B}$ be defined according to Definition 3.3 and let its elements be enumerated as b_i, $i = 1, \ldots, 2^n$. Then, for each $b \in \mathcal{B}$ there exists a unique set of indices $\mathcal{I}_b \subseteq \{1, \ldots, 2^n\}$ such that $b = \bigvee_{i\in\mathcal{I}_b} b_i$.*

The unique set of indices $\mathcal{I}_b$ for an element $b \in \mathcal{B}$ having the property mentioned in the previous lemma will be called the *index set* for b.

Lemma 3.6 *Let $\mathcal{B}$ be a Boolean algebra of propositions with n free generators, $n \geq 1$. Let the basis $\mathcal{B}_0 \subseteq \mathcal{B}$ be defined as in the foregoing and let its elements be enumerated as $b_i, i = 1, \ldots, 2^n$. Furthermore, let $b \in \mathcal{B}$ and let $\mathcal{I}_b$ be the index set for b as in Lemma 3.5. Then, for each joint probability distribution Pr on $\mathcal{B}$ we have*

$$Pr(b) = \sum_{i\in\mathcal{I}_b} Pr(b_i)$$

We will exploit the set $\mathcal{B}_0$ and its properties for computing intervals for probabilities of interest from an arbitrary partial specification. Before we do so, we compare our approach with similar approaches described in [Cooper86] and [Nilsson86], respectively. In [Cooper86], G.F. Cooper uses a set of propositions similar to our set $\mathcal{B}_0$ for computing probability intervals; he, however, does not introduce the notion of a basis. In his work on probabilistic logic, N.J. Nilsson defines the notion of possible worlds, a notion that at first sight is closely related to our notion of a basis, [Nilsson86]; in general, however, a set of possible worlds will not be discriminating enough for the purpose of computing arbitrary probability intervals.

Now, suppose that we are given probabilities for a number of arbitrary Boolean combinations of atomic propositions, that is, we consider the case in which we are given a consistent partial specification P of a joint probability distribution on $\mathcal{B}$, which is defined on an arbitrary subset $\mathcal{C} \subseteq \mathcal{B}$. The problem of finding a joint probability distribution on $\mathcal{B}$ which is an extension of P will now be transformed into an equivalent problem in linear algebra. The general idea is to take the initially given probabilities as defining constraints on a yet unknown joint probability distribution.

Let $\mathcal{B}$ once more be a Boolean algebra of propositions with n free generators, $n \geq 1$. Let $\mathcal{B}_0 \subseteq \mathcal{B}$ be the basis defined in Definition 3.3 and let its elements be enumerated as

b_i, $i = 1, \ldots, 2^n$. Let $\mathcal{C} = \{c_1, \ldots, c_m\}$, $m \geq 1$, be a subset of $\mathcal{B}$ and let P: $\mathcal{C} \rightarrow [0, 1]$ be a consistent partial specification of a joint probability distribution on $\mathcal{B}$. We now consider an arbitrary (yet unknown) joint probability distribution Pr on $\mathcal{B}$ with $Pr\,|_{\mathcal{C}} = P$. Let the constituent probabilities $Pr(b_i)$, $b_i \in \mathcal{B}_0$, of Pr be denoted by x_i, $i = 1, \ldots, 2^n$. Let the initially specified probabilities $P(c_i) = Pr(c_i), c_i \in \mathcal{C}$, $i = 1, \ldots, m$, be denoted by p_i. From Lemma 3.4 and Lemma 3.6 we obtain the following inhomogeneous system of linear equations:

$$\begin{array}{ccccccc} d_{1,1}x_1 & + & \ldots & + & d_{1,2^n}x_{2^n} & = & p_1 \\ \vdots & & & & \vdots & & \vdots \\ d_{m,1}x_1 & + & \ldots & + & d_{m,2^n}x_{2^n} & = & p_m \\ x_1 & + & \ldots & + & x_{2^n} & = & 1 \end{array}$$

where $d_{i,j} = \begin{cases} 0 & \text{if } j \notin \mathcal{I}_{c_i} \\ 1 & \text{if } j \in \mathcal{I}_{c_i} \end{cases}$, $i = 1, \ldots, m$, $j = 1, \ldots, 2^n$, in which $\mathcal{I}_{c_i}$ is the index set for $c_i \in \mathcal{C}$. This system of linear equations has the 2^n unknowns $x_1, \ldots, x_{2^n}$. Now, let $\boldsymbol{p}$ denote the column vector of right-hand sides of this system of linear equations and let $\boldsymbol{x}$ denote the column vector of unknowns. Furthermore, let $\boldsymbol{D}$ denote the coefficient matrix of the system. From now on, we will use the matrix equation $\boldsymbol{Dx} = \boldsymbol{p}$ to denote the system of linear equations obtained from a partial specification P as described above.

The following lemma states the relation between extensions of a consistent partial specification of a joint probability distribution and solutions to the matrix equation obtained from it.

Lemma 3.7 *Let $\mathcal{B}$ be a Boolean algebra of propositions with n free generators, $n \geq 1$. Let the basis $\mathcal{B}_0 \subseteq \mathcal{B}$ be defined according to Definition 3.3 and let its elements be enumerated as b_i, $i = 1, \ldots, 2^n$. Let $\mathcal{C} \subseteq \mathcal{B}$ and let P: $\mathcal{C} \rightarrow [0, 1]$ be a consistent partial specification of a joint probability distribution on $\mathcal{B}$. Let $\boldsymbol{Dx} = \boldsymbol{p}$ be the matrix equation obtained from P as described in the foregoing. Then, the following properties hold:*

- *For any joint probability distribution Pr on $\mathcal{B}$ such that $Pr\,|_{\mathcal{C}} = P$, we have that the vector $\boldsymbol{x}$ of constituent probabilities $x_i = Pr(b_i)$, $b_i \in \mathcal{B}_0$, $i = 1, \ldots, 2^n$, is a solution to the matrix equation $\boldsymbol{Dx} = \boldsymbol{p}$.*

- *For any nonnegative solution vector $\boldsymbol{x}$ with components x_i, $i = 1, \ldots, 2^n$, to the matrix equation $\boldsymbol{Dx} = \boldsymbol{p}$, we have that $Pr(b_i) = x_i$, $b_i \in \mathcal{B}_0$, defines a joint probability distribution Pr on $\mathcal{B}$ such that $Pr\,|_{\mathcal{C}} = P$.*

Note that although every joint probability distribution Pr which is an extension of a consistent partial specification P corresponds uniquely with a solution to the matrix equation $\boldsymbol{Dx} = \boldsymbol{p}$ obtained from P, not every solution to $\boldsymbol{Dx} = \boldsymbol{p}$ corresponds with a 'probabilistic' extension of P: $\boldsymbol{Dx} = \boldsymbol{p}$ may have solutions in which at least one of the x_i's is less than zero.

From Lemma 3.7 we derive a necessary and sufficient condition for a consistent partial specification to be a definition of a joint probability distribution.

Corollary 3.8 *Let $\mathcal{B}$ be a Boolean algebra of propositions. Let P be a consistent partial specification of a joint probability distribution on $\mathcal{B}$. Let $\boldsymbol{Dx} = \boldsymbol{p}$ be the matrix equation obtained from P as described in the foregoing. P uniquely defines a joint probability distribution on $\mathcal{B}$ if and only if $\boldsymbol{Dx} = \boldsymbol{p}$ has a unique nonnegative solution.*

Now consider the case in which we are given a consistent partial specification P which can be extended in more than one way to an actual joint probability distribution. For making statements concerning probabilities of interest, we can simply select a single 'probabilistic' extension of P and use the selected joint probability distribution for computing the probabilities we are interested in. Since P can be extended in more that one way to a joint probability distribution on $\mathcal{B}$ we have that the matrix equation $\boldsymbol{D}\boldsymbol{x} = \boldsymbol{p}$ obtained from P has infinitely many solutions. For the rank r of the coefficient matrix $\boldsymbol{D}$ we have that $r < 2^n$. So, in $\boldsymbol{D}\boldsymbol{x} = \boldsymbol{p}$ we have r basic variables and $2^n - r$ free variables. To obtain a particular solution to the matrix equation, we choose the values of the free variables, that is, some of the constituent probabilities, more or less freely, although subject to the constraints from the matrix equation and $x_i \geq 0$, $i = 1, \ldots, 2^n$; from these values the values of the basic variables can then be computed uniquely.

There are, however, other joint probability distributions on $\mathcal{B}$ respecting the initially given probabilities which are not equal to the one defined by the chosen solution vector: every other nonnegative vector differing from the selected one by a vector in the nullspace of $\boldsymbol{D}$ defines another joint probability distribution on $\mathcal{B}$ which is also an extension of P. It will be evident that the more free variables occur in the matrix equation, the more arbitrary the selected probability distribution will be. The results from using one solution vector for computing probabilities of interest can therefore differ considerably from the results from using another solution vector. Selecting a single, not unique extension of a partial specification of a joint probability distribution to serve as the basis for further computations as sketched in the foregoing, therefore does not render a reliable result.

We abandon the idea of selecting a single extension of a partial specification of a joint probability distribution for further computation: we introduce a method for finding best possible upper and lower bounds on probabilities of interest. The idea of finding bounds on probabilities from a partial specification of a joint probability distribution originated with G. Boole, as well as the idea of obtaining the 'narrowest limits' ([Hailperin86], page 338).

We define the notions of the least upper bound and greatest lower bound functions relative to a partial specification of a joint probability distribution.

Definition 3.9 *Let $\mathcal{B}$ be a Boolean algebra of propositions. Let $\mathcal{C} \subseteq \mathcal{B}$ and let $P\colon \mathcal{C} \to [0,1]$ be a consistent partial specification of a joint probability distribution on $\mathcal{B}$. The function $lub_P\colon \mathcal{B} \to [0,1]$ defined by $lub_P(b) = sup\{Pr(b) | Pr$ is a joint probability distribution on $\mathcal{B}$ such that $Pr|_{\mathcal{C}} = P\}$ for all $b \in \mathcal{B}$, is called the* least upper bound function *relative to P. The* greatest lower bound function *relative to P, denoted by glb_P, is defined symmetrically.*

Note that the least upper bound function relative to a partial specification P in general is not a joint probability distribution; of course, the same remark can be made concerning the greatest lower bound function. For a given $b \in \mathcal{B}$, the length of the interval $[glb_P(b), lub_P(b)]$ expresses the lack of knowledge concerning the probability of the truth of the proposition b. The two types of bounds are interrelated by $lub_P(b) = 1 - glb_P(\neg b)$, for each $b \in \mathcal{B}$.

Let P be a partial specification of a joint probability distribution on a Boolean algebra of propositions $\mathcal{B}$. The following lemma now states that we can find for each $b \in \mathcal{B}$ a joint probability distribution Pr on $\mathcal{B}$ being an extension of P such that $Pr(b) = lub_P(b)$; again, a similar observation can be made concerning glb_P.

Lemma 3.10 *Let $\mathcal{B}$ be a Boolean algebra of propositions. Let $\mathcal{C} \subseteq \mathcal{B}$ and let P: $\mathcal{C} \to [0,1]$ be a consistent partial specification of a joint probability distribution on $\mathcal{B}$. Furthermore, let the functions lub_P and glb_P be defined according to Definition 3.9. Then, for each $b \in \mathcal{B}$ we have $lub_P(b) = max\{Pr(b) | Pr$ is a joint probability distribution on $\mathcal{B}$ such that $Pr|_\mathcal{C} = P\}$. A similar property holds for $glb_P(b)$.*

Proof. The property stated in the lemma will readily be seen using the observation that the Boolean algebra of propositions $\mathcal{B}$ is finite. The lemma has been proven formally by Th. Hailperin, [Hailperin65]. ■

On the basis of the properties stated in Lemma 3.10, it can be shown that the problems of finding for a given $b \in \mathcal{B}$ the least upper bound $lub_P(b)$ and the greatest lower bound $glb_P(b)$ relative to a partial specification P of a joint probability distribution on $\mathcal{B}$, are equivalent to the following linear programming problems, respectively:

1. maximize $Pr(b)$ subject to $\boldsymbol{Dx} = \boldsymbol{p}$ and $\boldsymbol{x} \geq \boldsymbol{0}$; and
2. minimize $Pr(b)$ subject to $\boldsymbol{Dx} = \boldsymbol{p}$ and $\boldsymbol{x} \geq \boldsymbol{0}$,

where $\boldsymbol{Dx} = \boldsymbol{p}$ is the matrix equation obtained from P. The equivalence will be stated formally in Proposition 3.11. First, we consider case (1) in some detail in order to obtain a more traditional representation of the linear programming problem.

Let $\mathcal{B}$ be a Boolean algebra of propositions with n free generators, $n \geq 1$. Let $\mathcal{B}_0 \subseteq \mathcal{B}$ be the basis defined according to Definition 3.3 and let its elements be enumerated as b_i, $i = 1, \ldots, 2^n$. Now let $b \in \mathcal{B}$ and let $\mathcal{I}_b$ be its index set. For each joint probability distribution Pr on $\mathcal{B}$, we have that

$$Pr(b) = \sum_{i \in \mathcal{I}_b} Pr(b_i) = \sum_{i \in \mathcal{I}_b} x_i$$

Now, let for b constants c_i, $i = 1, \ldots, 2^n$, be defined such that $c_i = \begin{cases} 0 & \text{if } i \notin \mathcal{I}_b \\ 1 & \text{if } i \in \mathcal{I}_b \end{cases}$. Then, we have that

$$Pr(b) = \sum_{i=1}^{2^n} c_i x_i$$

So, our aim is to find the best upper bound for this function $\sum_{i=1}^{2^n} c_i x_i$.

We recall that in the matrix equation $\boldsymbol{Dx} = \boldsymbol{p}$ obtained from a partial specification P: $\mathcal{C} \to [0,1]$, $\mathcal{C} \subseteq \mathcal{B}$, $\boldsymbol{D}$ denotes a $(|\mathcal{C}| + 1) \times 2^n$ matrix, $\boldsymbol{x}$ is the 2^n column vector of constituent probabilities $Pr(b_i)$ and $\boldsymbol{p}$ is the $|\mathcal{C}| + 1$ column vector of initially given probabilities. The partial problem (1) can therefore be reformulated in the following more traditional representation of a linear programming problem:

maximize $\quad \sum_{i=1}^{2^n} c_i x_i$

subject to (i) $\quad \sum_{j=1}^{2^n} d_{i,j} x_j = p_i$, for $i = 1, \ldots, |\mathcal{C}| + 1$, and

(ii) $\quad x_j \geq 0$, for $j = 1, \ldots, 2^n$

where the constants $d_{i,j}$ constitute the matrix $\boldsymbol{D}$. Note that we have added nonnegativity constraints to $\boldsymbol{Dx} = \boldsymbol{p}$ explicitly to allow for nonnegative solutions only. The linear programming problem (2) can be treated analogously by taking for the objective function $-\sum_{i=1}^{2^n} c_i x_i$.

Proposition 3.11 *Let $\mathcal{B}$ be a Boolean algebra of propositions. Let $\mathcal{C} \subseteq \mathcal{B}$ and let $P\colon \mathcal{C} \to [0,1]$ be a consistent partial specification of a joint probability distribution on $\mathcal{B}$. Let $\boldsymbol{Dx} = \boldsymbol{p}$ be the matrix equation obtained from P. Furthermore, let the functions lub_P and glb_P be defined according to Definition 3.9. Then, for any $b \in \mathcal{B}$ we have that $lub_P(b)$ is equal to the solution of the linear programming problem*

$$\begin{array}{lll} \textit{maximize} & & Pr(b) \\ & & \\ \textit{subject to} & \text{(i)} & \boldsymbol{Dx} = \boldsymbol{p}, \textit{ and} \\ & \text{(ii)} & \boldsymbol{x} \geq \boldsymbol{0}. \end{array}$$

A similar statement can be made concerning $glb_P(b)$.

Now consider application of the linear programming approach in a model for handling uncertainty in a knowledge-based system. In short, a domain expert is requested to assess several probabilities. The assessed probabilities are used in the manner described in this section to generate a system of linear constraints. From this system of constraints upper and lower bounds on the probabilities that are of interest to the user of the system are computed. The following example illustrates the idea.

Example 3.12 Let $\mathcal{A} = \{a_1, a_2, a_3\}$ and let $\mathcal{B}(a_1, a_2, a_3)$ be the free Boolean algebra generated by $\mathcal{A}$. Let $\mathcal{C} = \{a_1 \wedge a_2, \neg a_1 \vee a_3, a_2, a_2 \wedge \neg a_3\}$. Note that $\mathcal{C}$ cannot be a basis for a joint probability distribution on $\mathcal{B}(a_1, a_2, a_3)$ since it only contains four elements. Let P be a consistent partial specification defined on $\mathcal{C}$ which can be extended in more than one way to a joint probability distribution on $\mathcal{B}(a_1, a_2, a_3)$. We consider such a 'probabilistic' extension Pr. Suppose that we have the following function values of Pr coinciding with the corresponding initially given function values of P:

$$\begin{array}{lcl} Pr(a_1 \wedge a_2) & = & 0.23 \\ Pr(\neg a_1 \vee a_3) & = & 0.62 \\ Pr(a_2) & = & 0.43 \\ Pr(a_2 \wedge \neg a_3) & = & 0.18 \end{array}$$

Now let the elements of the basis $\mathcal{B}_0 \subseteq \mathcal{B}(a_1, a_2, a_3)$ be enumerated as follows:

$$\begin{array}{lcllllll} b_1 & = & a_1 & \wedge & a_2 & \wedge & a_3 \\ b_2 & = & \neg a_1 & \wedge & a_2 & \wedge & a_3 \\ b_3 & = & a_1 & \wedge & \neg a_2 & \wedge & a_3 \\ b_4 & = & a_1 & \wedge & a_2 & \wedge & \neg a_3 \\ b_5 & = & \neg a_1 & \wedge & \neg a_2 & \wedge & a_3 \\ b_6 & = & \neg a_1 & \wedge & a_2 & \wedge & \neg a_3 \\ b_7 & = & a_1 & \wedge & \neg a_2 & \wedge & \neg a_3 \\ b_8 & = & \neg a_1 & \wedge & \neg a_2 & \wedge & \neg a_3 \end{array}$$

Furthermore, let the constituent probabilities $Pr(b_i)$ be denoted by x_i, $i = 1, \ldots, 8$. From P we obtain the following system of linear equations:

$$\begin{array}{rcl}
x_1 + x_4 & = & 0.23 \\
x_1 + x_2 + x_3 + x_5 + x_6 + x_8 & = & 0.62 \\
x_1 + x_2 + x_4 + x_6 & = & 0.43 \\
x_4 + x_6 & = & 0.18 \\
x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 & = & 1
\end{array}$$

We add the constraints

$$x_i \geq 0, i = 1, \ldots, 8$$

explicitly. Now, suppose that we are interested in bounds on the probability of the truth of the atomic proposition a_3. From Proposition 3.11 we have that the problem of determining the best upper bound for $Pr(a_3)$ is equal to maximizing the objective function

$$x_1 + x_2 + x_3 + x_5$$

subject to the constraints shown above. Applying the simplex method for this purpose we obtain $lub_P(a_3) = 0.62$ and $glb_P(a_3) = 0.25$. ∎

It is well-known that an LP-problem can be solved in polynomial time, that is, polynomial in the size of the problem. The size of an LP-problem is dependent, among other factors, upon the number of variables it comprises. The specific type of problem discussed in the foregoing has exponentially many variables, that is, exponential in the number of statistical variables discerned in the problem domain. Therefore, these problems cannot be solved in polynomial time; computing bounds on probabilities of interest requires an exponential number of steps.

3.2 Dealing with Conditional Probabilities

In the previous subsection we have presented a linear programming method for computing bounds on probabilities of interest from a partial specification of a joint probability distribution. This method has been developed for partial specifications comprising prior probabilities only. In the domains in which knowledge-based systems are employed, however, it often is easier to assess or otherwise obtain conditional probabilities than it is to obtain prior ones. Moreover, the user of the system will often be interested in conditional probabilities. We will show that conditional probabilities can be introduced into the linear programming method without requiring much effort.

We first examine the case in which we are initially given some conditional probabilities. Let $\mathcal{B}$ once more be a Boolean algebra of propositions with n free generators, $n \geq 1$. Furthermore, let $\mathcal{B}_0 \subseteq \mathcal{B}$ be the basis as defined in Definition 3.3 and let its elements be enumerated as b_i, $i = 1, \ldots, 2^n$. Let P be a consistent partial specification of a joint probability distribution on $\mathcal{B}$; we consider a joint probability distribution Pr on $\mathcal{B}$ which is an extension of P. Now suppose that an expert has assessed the value $P(c_1|c_2) = Pr(c_1|c_2) = p_0$, where $c_1, c_2 \in \mathcal{B}$, to be taken as a conditional probability. Note that it follows implicitly that $Pr(c_2) \neq 0$. By definition, we have $Pr(c_1|c_2) = \frac{Pr(c_1 \wedge c_2)}{Pr(c_2)}$. From Lemma 3.5 and Lemma 3.6 we have that there exist an index set $\mathcal{I}_{c_1 \wedge c_2}$ for $c_1 \wedge c_2$ such that

$$Pr(c_1 \wedge c_2) = \sum_{i \in \mathcal{I}_{c_1 \wedge c_2}} Pr(b_i)$$

and an index set $\mathcal{I}_{c_2}$ for c_2 such that

$$Pr(c_2) = \sum_{i \in \mathcal{I}_{c_2}} Pr(b_i)$$

where $Pr(b_i)$ are the constituent probabilities of Pr. We therefore have that

$$Pr(c_1|c_2) = \frac{\sum_{i \in \mathcal{I}_{c_1 \wedge c_2}} Pr(b_i)}{\sum_{i \in \mathcal{I}_{c_2}} Pr(b_i)} = p_0$$

It follows that

$$\sum_{i \in \mathcal{I}_{c_1 \wedge c_2}} Pr(b_i) = p_0 \cdot \sum_{i \in \mathcal{I}_{c_2}} Pr(b_i)$$

We now obtain the constraint

$$\sum_{i \in \mathcal{I}_{c_1 \wedge c_2}} Pr(b_i) - p_0 \cdot \sum_{i \in \mathcal{I}_{c_2}} Pr(b_i) = 0$$

which is similar in concept to the ones we have encountered in our linear programming problems; it can therefore be treated likewise. (Note that we have to guarantee that $\sum_{i \in \mathcal{I}_{c_2}} Pr(b_i) = 0$ is not a solution to the obtained system of linear constraints.)

Now consider the case in which we are interested in lower and upper bounds on a conditional probability. From the foregoing discussion, it will be evident that we have a fractional objective function for our problem. Such a problem, called a *fractional* linear programming problem, however, can be reduced to a related 'ordinary' linear programming problem with one additional variable. The following proposition formulated in [Hailperin86] but originally due to A. Charnes, states this result.

Proposition 3.13 *The fractional linear programming problem*

maximize $\frac{\boldsymbol{cx}}{\boldsymbol{gx}}$

subject to (i) $\boldsymbol{Dx} = \boldsymbol{p}$
(ii) $\boldsymbol{x} \geq \boldsymbol{0}$

is equivalent to the linear programming problem

maximize $\boldsymbol{cy}$

subject to (i) $\boldsymbol{Dy} = t\boldsymbol{p}$
(ii) $\boldsymbol{gy} = 1$
(iii) $\boldsymbol{y} \geq \boldsymbol{0}$
(iv) $t \geq 0$

The linear programming problems we obtained in Section 3.1 were in standard form: all constraints except the nonnegativity constraints are equalities. It will be evident that our method is able to deal with LP-problems in general form as well. Allowing inequalities in our method provides a domain expert with a flexible means for expressing probabilistic information: besides prior and conditional probabilities, he may specify bounds on probabilities instead of point estimates and he may give certain probabilities relative to other ones.

4 Processing Evidence

In this section we address the problem of updating a partial specification of a joint probabiliy distribution P as evidence becomes available, that is, we address the problem of processing evidence in the conceptual framework we have presented in the preceding section taking this framework for an invariant. To this end, we discern two types of evidence:

- evidence concerning the partial specification itself, called *case-independent* evidence, and
- evidence observed for a specific case, called *case-dependent* evidence.

Case-independent evidence is merely *new* knowledge concerning the partial specification P rendering it 'more specified': it is information we did not have before. This type of evidence is dealt with just by adding another constraint representing the piece of evidence to the system of linear constraints obtained from P. The bounds obtained after processing this type of evidence are modified monotonically: new evidence merely leads to the same or narrower probability intervals. Note that this property allows for stepwise filling-in a probability distribution.

The second type of evidence concerns information that for the specific case we are looking at we have observed that a certain statistical variable has a certain value. From the way an updated probability distribution in the fully specified case is computed, it will be evident that we cannot simply add this evidence as a new constraint to the system of constraints obtained from the partial specification P; the addition of such a constraint will probably even render the system infeasible. Adhering to the basic idea of our framework, our aim now is to arrive at a method for 'updating' the system of constraints obtained from P, yielding a new system of linear constraints such that this new system defines the possible extensions of the partial specification P after the evidence has been processed; in that case, after processing a piece of evidence we can compute bounds on probabilities of interest just like before the evidence was processed.

Let $P\colon \mathcal{C} \to [0,1]$ be a partial specification of a joint probability distribution on the Boolean algebra of propositions $\mathcal{B}(a_1,\ldots,a_n)$, $n \geq 1$. From P we obtain the system of linear constraints $\boldsymbol{Dx} = \boldsymbol{p}, \boldsymbol{x} \geq \boldsymbol{0}$, where $\boldsymbol{x}$ is a 2^n-column vector of constituent probabilities, as described in the preceding section. Let F denote the feasible set of this system of constraints. Recall that each vector $\boldsymbol{x} \in F$ defines a joint probability distribution Pr on $\mathcal{B}(a_1,\ldots,a_n)$ which is an extension of P. Now, suppose that we obtain the case-dependent evidence that the statistical variable $A \in \{A_1,\ldots,A_n\}$ has the value *true* (the case in which we have observed that A has the value *false* is dealt with analogously).

We consider the joint probability distribution Pr defined by a specific vector $\boldsymbol{x} \in F$ and investigate the updating of Pr. We observe that 2^{n-1} constituent probabilities of Pr

specify a, that is, $A = true$, and that the remaining 2^{n-1} ones specify $\neg a$. It will be evident that updating the joint probability distribution Pr amounts to setting all constituent probabilities specifying $\neg a$ equal to zero, and then normalizing the remaining constituent probabilities in order to render the result again a joint probability distribution. Now consider the defining vector x. Without loss of generality, we take the components of x to be ordered in such a way that the first 2^{n-1} components correspond to the constituent probabilities of Pr specifying a and that the remaining components correspond to the constituent probabilities specifying $\neg a$. The following definition introduces an update mapping such that when applied to the vector x the updated vector defines the updated joint probability distribution (for ease of exposition, we take the updated vector to be of the same dimension as the original one).

Definition 4.1 *The* update mapping U: $\mathbf{R}^{2^n} \rightarrow \mathbf{R}^{2^n}$, $n \geq 1$, *is the partial mapping defined by*

1. $U((x_1, \ldots, x_{2^n})) = \left(\frac{x_1}{\sum_{i=1}^{2^{n-1}} x_i}, \ldots, \frac{x_{2^{n-1}}}{\sum_{i=1}^{2^{n-1}} x_i}, 0, \ldots, 0 \right)$ *if* $\sum_{i=1}^{2^{n-1}} x_i \neq 0$, *and*
2. $U(x) = undefined$, *otherwise.*

The case in which we apply the update mapping U to some vector $x \in \mathbf{R}^{2^n}$ of constituent probabilities for which we have $\sum_{i=1}^{2^{n-1}} x_i = 0$ deserves some special attention. For the joint probability distribution Pr defined by such a vector x, we evidently have $Pr(a) = 0$. Evidence that $Pr(a) = 1$ contradicts this prior information. For this case, we take $U(x) = undefined$; this is an arbitrary choice. We return to this observation shortly.

Since we are primarily interested in applying mappings to vectors representing joint probability distributions, we will frequently restrict the discussion to unit simplices.

Definition 4.2 *The* unit simplex *in* $\mathbf{R}^n$, $n \geq 1$, *denoted by* S_n, *is the convex set in the positive orthant of* $\mathbf{R}^n$ *such that for each* $x \in S_n$ *we have* $\sum_{i=1}^{n} x_i = 1$.

The following lemma states the evident property that when applied to a vector representing a joint probability distribution the update mapping U yields a vector which again represents a joint probability distribution, provided of course that the result is defined.

Lemma 4.3 *Let the mapping* U: $\mathbf{R}^{2^n} \rightarrow \mathbf{R}^{2^n}$, $n \geq 1$, *be defined as above. For each vector* $x \in S_{2^n}$, *we have that either* $U(x) \in S_{2^n}$ *or* $U(x) = undefined$.

Until now we have only looked at the updating of a single vector. Recall that such a vector is an element of a convex polytope F of vectors, each defining a joint probability distribution which is an extension of the initially given partial specification P. For processing the evidence that the statistical variable A has adopted the value *true*, we apply the update mapping U to each vector $x \in F$. We therefore are interested in the image $U(F)$ of F. Since the mapping U is non-linear, the question arises whether the image of a convex polytope under U again is a convex polytope. We will show that this question may be answered in the affirmative.

It will be evident from the preceding informal discussion that the update mapping U is composed of a multiplication and a projective mapping. The multiplication mapping is defined in the following definition.

Definition 4.4 *The* multiplication mapping M: $\mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *is the partial mapping defined by*

1. $M((x_1, \ldots, x_{2^n})) = \left(\frac{x_1}{\sum_{i=1}^{2^{n-1}} x_i}, \ldots, \frac{x_{2^n}}{\sum_{i=1}^{2^{n-1}} x_i} \right)$ *if* $\sum_{i=1}^{2^{n-1}} x_i \neq 0$, *and*
2. $M(x) = undefined$, *otherwise.*

The geometrical idea of applying the multiplication mapping M to a vector from the unit simplex S_{2^n} is sketched in Figure 3.1.

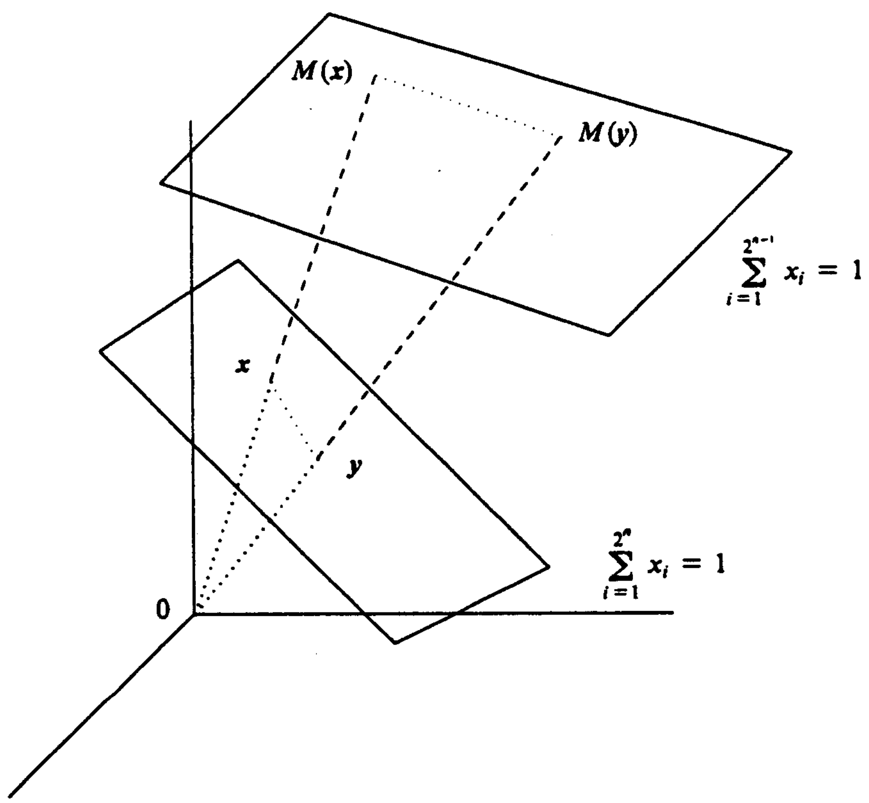

Figure 3.1

We consider this mapping M in some detail.

Lemma 4.5 *Let the multiplication mapping* M: $\mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *be defined according to Definition 4.4. Furthermore, let* $F \subseteq S_{2^n}$ *be a convex polytope such that for each* $x \in F$ *we have that* $M(x) \neq undefined$. *Then, the image* $M(F)$ *of* F *is a convex polytope.*

Proof. We prove the lemma by showing that by applying M each line segment in F is mapped into a line segment.

Let $\boldsymbol{x}, \boldsymbol{y} \in F$; let $M(\boldsymbol{x})_i$ denote the i-th component of the vector $M(\boldsymbol{x})$. Then, $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}$, $0 \leq \lambda \leq 1$, represents the line segment between the two points $\boldsymbol{x}$ and $\boldsymbol{y}$. Now, consider $M(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})$; note that the conditions of the lemma guarantee that $M(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \neq undefined$. We now have to show that there exists a scalar μ, $0 \leq \mu \leq 1$, such that the property $M(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) = \mu M(\boldsymbol{x}) + (1-\mu)M(\boldsymbol{y})$ holds. We have

$$M(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})_i = \frac{\lambda x_i + (1-\lambda)y_i}{\sum_{j=1}^{2^{n-1}} (\lambda x_j + (1-\lambda)y_j)} = \frac{\lambda x_i + (1-\lambda)y_i}{\lambda \sum_{j=1}^{2^{n-1}} x_j + (1-\lambda)\sum_{j=1}^{2^{n-1}} y_j}$$

for $i = 1, \ldots, 2^n$. Now let $\alpha = \sum_{j=1}^{2^{n-1}} x_j$ and $\beta = \sum_{j=1}^{2^{n-1}} y_j$. Then,

$$M(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})_i = \frac{\lambda x_i + (1-\lambda)y_i}{\lambda\alpha + (1-\lambda)\beta} = \frac{\lambda}{\lambda\alpha + (1-\lambda)\beta} x_i + \frac{(1-\lambda)}{\lambda\alpha + (1-\lambda)\beta} y_i$$

By definition, we have $M(\boldsymbol{x})_i = \frac{x_i}{\alpha}$ and $M(\boldsymbol{y})_i = \frac{y_i}{\beta}$. So,

$$\begin{aligned} M(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})_i &= \frac{\lambda\alpha}{\lambda\alpha + (1-\lambda)\beta} M(\boldsymbol{x})_i + \frac{(1-\lambda)\beta}{\lambda\alpha + (1-\lambda)\beta} M(\boldsymbol{y})_i = \\ &= \mu M(\boldsymbol{x})_i + (1-\mu)M(\boldsymbol{y})_i \end{aligned}$$

for $i = 1, \ldots, 2^n$, where $\mu = \frac{\lambda\alpha}{\lambda\alpha + (1-\lambda)\beta}$. Note that $0 \leq \mu \leq 1$. Furthermore, note that $\lambda = 0$ corresponds with $\mu = 0$ and that $\lambda = 1$ corresponds with $\mu = 1$. We have that $\mu M(\boldsymbol{x}) + (1-\mu)M(\boldsymbol{y})$ is the line segment between the points $M(\boldsymbol{x})$ and $M(\boldsymbol{y})$. It follows that $M(F)$ is convex. ■

Lemma 4.6 *Let the multiplication mapping* $M: \mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *be defined according to Definition 4.4. Furthermore, let* $F \subseteq S_{2^n}$ *be a convex polytope such that for each* $\boldsymbol{x} \in F$ *we have that* $M(\boldsymbol{x}) \neq undefined$. *Then,* $\boldsymbol{x}$ *is a vertex of* F *if and only if* $M(\boldsymbol{x})$ *is a vertex of* $M(F)$.

Proof. The lemma follows immediately from the proof of the previous lemma and the observation that for each $\boldsymbol{x}, \boldsymbol{y} \in F$ such that $\boldsymbol{x} \neq \boldsymbol{y}$, we have $M(\boldsymbol{x}) \neq M(\boldsymbol{y})$. ■

Recall that the update mapping U from Definition 4.1 is composed of the multiplication mapping M and a projective mapping. We now turn our attention to this projective mapping.

Definition 4.7 *The* projective mapping R: $\mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *is the mapping defined by*

$$R((x_1, \ldots, x_{2^n})) = (x_1, \ldots, x_{2^{n-1}}, 0, \ldots, 0)$$

Note that if we take $U = R \circ M$, we formally have to deal with the case where $M(\boldsymbol{x})$ is undefined for some vector $\boldsymbol{x}$. For ease of exposition we disregard such cases for the moment. The following two lemmas should be evident.

Lemma 4.8 *Let the projective mapping* $R\colon \mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *be defined as above. Let* $F \subseteq S_{2^n}$ *be a convex polytope. Then, the image* $R(F)$ *of* F *is a convex polytope.*

Lemma 4.9 *Let the projective mapping* $R\colon \mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *be defined as above. Let* $F \subseteq S_{2^n}$ *be a convex polytope. Then,* $R(x)$ *is a vertex of* $R(F)$ *only if* x *is a vertex of* F.

Note that the reverse property of Lemma 4.9 does not hold, that is, not every vertex x of F corresponds with a vertex of $R(F)$.

Now, recall that we are interested in the result of applying the update mapping U to the set F of vectors each defining a joint probability distribution which is an extension of the initially given partial specification P. We combine the Lemmas 4.5, 4.6, 4.8 and 4.9 to yield the following lemma concerning U.

Lemma 4.10 *Let the update mapping* $U : \mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *be defined according to Definition 4.1. Let* $F \subseteq S_{2^n}$ *be a convex poytope such that for each* $x \in F$ *we have that* $U(x) \neq undefined$. *Then,*

1. *the image* $U(F)$ *of* F *is a convex polytope, and*

2. $U(x)$ *is a vertex of* $U(F)$ *only if* x *is a vertex of* F.

Furthermore, let $vert(F)$ *be the set of vertices of* F *and let* $hull(U(vert(F)))$ *be the convex hull of* $U(vert(F))$. *Then,* $U(F) = hull(U(vert(F)))$.

Consider the last statement of the preceding lemma once more. It will be evident that for a given polytope $F \subseteq S_{2^n}$, $n \geq 1$, having the mentioned property, the set $U(vert(F))$ is not the minimal spanning set of $U(F)$ since it may contain some interior points from $U(F)$ as well.

The last lemma now provides us with a theoretical means for updating the system of constraints obtained from the partial specification P. In the following algorithm we assume that the feasible set of this system does not comprise any vectors x for which $U(x) = undefined$.

Algorithm 4.11 *Let the update mapping* $U : \mathbf{R}^{2^n} \to \mathbf{R}^{2^n}$, $n \geq 1$, *be defined according to Definition 4.1. Let* $F \subseteq S_{2^n}$ *be a convex polytope such that for each* $x \in F$ *we have that* $U(x) \neq undefined$. *Then, the following algorithm yields a system of constraints having* $U(F)$ *for its feasible set:*

1. *Compute the set* $vert(F)$ *of all vertices of* F.

2. *Apply the operator* U *to each element* $x \in vert(F)$, *thus obtaining the set* $U(vert(F))$.

3. *Use* $U(vert(F))$ *to span the convex hull* $U(F) = hull(U(vert(F)))$.

4. *Construct the supporting hyperplanes of* $U(F)$ *and generate the appropriate system of constraints.*

This algorithm for processing case-dependent evidence of course is rather inefficient; step (1) in itself already takes exponential time. It shows however that updating the system of constraints can actually be achieved.

We return to the case in which the feasible set F of the system of constraints obtained from P contains at least one vector for which the image under U equals *undefined*. Note that in this case F has at least one vertex $\boldsymbol{x}$ such that $U(\boldsymbol{x}) = \mathit{undefined}$. We distinguish two cases: if $vert(F)$ only consists of vectors $\boldsymbol{x}$ for which $U(\boldsymbol{x}) = \mathit{undefined}$, then the observed evidence we are trying to process evidently is inconsistent with the prior information; the evidence cannot be processed and the system should report the detected inconsistency. If, on the other hand, $vert(F)$ also comprises some vertices for which the image under U is defined, then the observed evidence can be processed. In computing $U(F)$, however, we have to exclude all vectors defining marginal distributions the piece of evidence is inconsistent with. Easy geometrical observations suffice to show that the above-mentioned algorithm yields the correct result after just ignoring those vertices of F for which the image under U is undefined; further details are provided in [Gaag90a].

5 Conclusion

In this paper we have introduced the notion of a partial specification of a joint probability distribution. We have exploited this notion in devising a method for computing intervals for probabilities of interest from an incomplete set of initially assessed probabilities and for successively updating them as evidence becomes available. By means of this method we have proven the claim that probability theory is not expressive enough to deal with incomplete information to be a misconception. Furthermore, we have endorsed Pearl's viewpoint concerning the suitability of probability intervals for expressing incompleteness of information as mentioned in our introduction. Although our method serves the purpose of supporting these observations, it can only be considered a theoretical result. Since representing an arbitrary joint probability distribution requires an exponential number of probabilities, any algorithm based on the ideas presented in this paper will take exponential time.

In case many independency relationships hold between the statistical variables discerned, however, far less probabilities suffice for uniquely representing a joint probability distribution, see for example [Lauritzen88]. So, it will be advantageous to exploit such independency relationships in a linear programming approach. In [Cooper86], G.F. Cooper already briefly addressed the incorporation of independency relationships in his method. In [Gaag90b], we will show how independency relationships can be exploited in our framework, yielding an efficient algorithm for computing intervals, having under certain conditions a polynomial time complexity.

References

[Boole54] G. Boole (1854). *An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities*, Walton and Maberley, London (reprinted in 1951 by Dover Publications, Inc., New York).

[Cooper86] G.F. Cooper (1986). A diagnostic method that uses causal knowledge and linear programming in the application of Bayes' formula, *Computer Methods and Programs in Biomedicine*, vol. 22, pp. 223 - 237.

[Gaag90a] L.C. van der Gaag (1990). *Probability-Based Models for Plausible Reasoning*, Ph.D. thesis, University of Amsterdam, to appear.

[Gaag90b] L.C. van der Gaag (1990). Computing probability intervals under independency constraints, to be presented at the *Sixth Conference on Uncertainty in AI.*

[Hailperin65] T. Hailperin (1965). Best possible inequalities for the probability of a logical function of events, *American Mathematical Monthly*, vol. 72, pp. 343 - 359.

[Hailperin86] T. Hailperin (1986). *Boole's Logic and Probability*, 2nd Edition, North-Holland, Amsterdam.

[Lauritzen88] S.L. Lauritzen, D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion), *Journal of the Royal Statistical Society, Series B*, vol. 50, no. 2, pp. 157- 224.

[Lucas90] P.J.F. Lucas, L.C. van der Gaag (1990). *Principles of Expert Systems*, Addison-Wesley, Wokingham, to appear.

[Nilsson86] N.J. Nilsson (1986). Probabilistic logic, *Artificial Intelligence*, vol. 28, pp. 71 - 87.

[Pearl88] J. Pearl (1988). On probability intervals, *International Journal of Approximate Reasoning*, vol. 2, pp. 211 - 216.

[Shortliffe84] E.H. Shortliffe, B.G. Buchanan (1984). A model of inexact reasoning in medicine, in: B.G. Buchanan, E.H. Shortliffe (eds.). *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts, pp. 233 - 262.