Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

L.C. van der Gaag

Partially specified probability measures in expert systems

Computer Science/Department of Software Technology Report CS-R8842 October

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Partially Specified Probability Measures in Expert Systems

L.C. van der Gaag
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

During the last decade, plausible reasoning has emerged as an important issue in expert systems research and artificial intelligence research more in general. Beside other mathematical theories, Bayesian probability theory has been chosen as a foundation for the development of models for handling uncertain information in expert systems. In the problem domains in which expert systems are developed however, a fully specified probability measure usually is not available. This observation, amongst others, prevents probability theory being applied as a model for handling uncertainty in a traditional way. In this paper, we present a model in which partially and even inconsistently specified probability measures are employed to establish upper and lower bounds on probabilities of interest; we depart from George Boole's ideas on probability presented in his 'Laws of Thought' as exposed by T. Hailperin,

1980 Mathematics Subject Classification (1985): 06Exx, 60Bxx, 68Txx
1987 CR Categories: I.2.3. **[Deduction and Theorem Proving]**: uncertainty, probabilistic reasoning; I.2.5. **[Programming Languages and Software]**: expert system tools and techniques.

Key Words & Phrases: expert systems, plausible reasoning, probability theory, linear programming.

1. INTRODUCTION

Early in expert systems research it has been observed that many domain experts are able to handle imprecise, incomplete and uncertain information to solve the practical problems they encounter in their field. This observation has led to the introduction of models for handling uncertain information in expert systems and the research area of plausible reasoning. The first point of departure in this research area has been Bayesian probability theory. Many models derived from this theory have been proposed for handling uncertainty in (mainly rule-based) expert systems, for example [1,2,3]. However, up till now none of these models can be considered the ultimate solution for the problem. One of the major stumbling blocks has been the frequent absence of completely and consistently specified probability measures in the fields of concern. In response to these problems other points of departure, i.e. other mathematical foundations for the development of models for handling uncertainty have been proposed and investigated, such as the fuzzy set theory. In this paper, we depart from the Bayesian probability theory.

When developing an expert system a domain expert is requested to elucidate his knowledge concerning the domain to be modelled. His knowledge concerning the problem domain is generally represented in a so-called knowledge representation formalism. Such a knowledge representation formalism provides amongst other things a method for modelling the uncertainties that go with the represented information. For example, it may provide a means for expressing an associated probability. With a knowledge representation formalism corresponds a reasoning method that is used to derive new information from the information available to the system. Such a reasoning method also provides a means to calculate the measures of uncertainty to be associated with the newly derived information; for an introduction to the principles of expert systems see [4]. When a probability measure has been specified for the domain concerned, the probabilities of interest can be computed from this measure. However, in the domains in which expert systems are employed, such completely

Report CS-R8842
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

specified probability measures usually are not available. Often only a few probabilities are known or can be estimated by an expert in the field. In expert systems therefore we are confronted with the problem of calculating the probability of a certain event given only a partially and often inconsistently specified probability measure. The problem of determining the probability of an event given a partially specified probability measure has already been investigated as early as halfway the nineteenth century by G. Boole, [5]. However, Boole's ideas on probability theory have received little attention. In an excellent book providing a thorough exposition of Boole's work on logic and probability in terms of modern algebra, propositional logic and probability theory, Hailperin states the following, ([6], page 215):

"Never clearly understood, and considered anyhow to be wrong, Boole's ideas on probability were simply by-passed by the history of the subject, which developed along other lines."

In our opinion Boole's ideas have become topical once more in the context of plausible reasoning in artificial intelligence. In this paper, we propose a mathematically well-founded and computationally feasible model for handling partially specified probability measures in expert systems, based on Boole's ideas. We have used Hailperin's book [6] as a guide to the work of Boole.

2. Preliminaries

In our introduction we have mentioned that in expert systems knowledge concerning the problem domain is represented in a knowledge representation formalism. In this paper we do not consider such formalisms in detail nor do we discuss the reasoning methods that have been associated with these formalisms. Here, we assume that knowledge is simply represented in logical propositions.

Definition 2.1. *Let $\mathscr{A}$ denote a finite set of atomic propositions: $\mathscr{A} = \{a_1,...,a_n\}$, $n \geqslant 1$. Let $\mathscr{B}$ be the free Boolean algebra generated by $\mathscr{A}$, i.e.*

(1) *for all $x \in \mathscr{A}$, $x \in \mathscr{B}$,*
(2) *for all $x_1, x_2 \in \mathscr{B}$, $x_1 \wedge x_2 \in \mathscr{B}$ and $x_1 \vee x_2 \in \mathscr{B}$, and*
(3) *for all $x \in \mathscr{B}$, $\neg x \in \mathscr{B}$.*

$\mathscr{B}$ is called the set of Boolean combinations of atomic propositions.

On $\mathscr{B}$ we define a partial order $\preccurlyeq$: for any $x_1,x_2 \in \mathscr{B}$, we say $x_1 \preccurlyeq x_2$ if $x_2 = x_1 \vee x_2$ or (equivalently) $x_1 = x_1 \wedge x_2$.

According to the convention in logic we denote the universal lower bound in the algebra $\mathscr{B}$ by false and the universal upper bound by true.

Since the set $\mathscr{B}$ of Boolean combinations of atomic propositions is a Boolean algebra we have equality according to logical truth tables. It will be obvious that a universal lower bound and upper bound exist. Furthermore, since $\mathscr{B}$ is a free Boolean algebra we have that the atomic propositions $a_i \in \mathscr{A}$ are algebraically independent, meaning that each of the 2^n conjunctions of the form $\bigwedge_{i=1}^{n} A_i$, where for $i = 1,...,n$ either $A_i = a_i$ or $A_i = \neg a_i$, is different from *false*.

In the next definition we introduce the notion of a probability function on a Boolean algebra in general and define a probability algebra.

Definition 2.2. *Let $\mathscr{E}$ be a Boolean algebra. Let P be a function $P : \mathscr{E} \to [0,1]$ such that*

(1) *P is positive, i.e. for all $x \in \mathscr{E}$, $P(x) \geqslant 0$, and furthermore $P(false) = 0$,*
(2) *P is normed, i.e. $P(true) = 1$, and*
(3) *P is additive, i.e. for all $x_1, x_2 \in \mathscr{E}$, if $x_1 \wedge x_2 = false$ then $P(x_1 \vee x_2) = P(x_1) + P(x_2)$.*

Then, P is called a probability function on $\mathscr{E}$. The pair $(\mathscr{E},P)$ is called a probability algebra.

The statements in the following lemma can easily be proven.

LEMMA 2.3. *Let $(\mathscr{E},P)$ be a probability algebra. Then,*

(1) $P(x) + P(\neg x) = 1$, *for all* $x \in \mathscr{E}$,
(2) $P(x_1 \vee x_2) + P(x_1 \wedge x_2) = P(x_1) + P(x_2)$, *for all* $x_1, x_2 \in \mathscr{E}$, *and*
(3) *if* $x_1 \preccurlyeq x_2$ *then* $P(x_1) \leqslant P(x_2)$, *for all* $x_1,x_2 \in \mathscr{E}$.

We use the notion of a probability function on a Boolean algebra to define on the set $\mathscr{B}$ of Boolean combinations of propositions a function *Pr* that associates with each proposition in $\mathscr{B}$ a probability of the truth of the proposition.

In an expert system, the uncertainties of the truths of some of the propositions have to be made explicit to enable the system to express the uncertainties of the truths of the new propositions it has derived. In probability theory we are used to associate probabilities with sets. Since in this paper knowledge is represented in logical propositions, we like to have an isomorphism that transforms logical propositions into sets and that preserves probability, for in that case we have that the probability of an event is equivalent to the probability of the truth of the proposition asserting the occurrence of the event. Already Boole had fully understood this equivalence. The following theorem provides such an isomorphism. Although proving the theorem is rather straightforward we nevertheless provide a part of the proof since it conveys many ideas that we will return to in the next section.

THEOREM 2.4. *Let $\mathscr{B}$ be defined according to Definition 2.1 and let Pr be a probability function on $\mathscr{B}$. Then, there exists a probability space $(\Omega, \mathscr{F}, P)$ and an isomorphism $\iota: \mathscr{B} \to \mathscr{F}$ such that*

(1) $\iota(x_1 \wedge x_2) = \iota(x_1) \cap \iota(x_2)$, *for all* $x_1,x_2 \in \mathscr{B}$,
(2) $\iota(x_1 \vee x_2) = \iota(x_1) \cup \iota(x_2)$, *for all* $x_1,x_2 \in \mathscr{B}$,
(3) $\iota(\neg x) = \overline{\iota(x)}$, *for all* $x \in \mathscr{B}$,
(4) $\mathscr{F}$ *is the free Boolean algebra generated by* $\{\iota(x) \mid x \in \mathscr{B}\}$, *and*
(5) $P(\iota(x)) = Pr(x)$, *for each* $x \in \mathscr{B}$.

The probability space $(\Omega, \mathscr{F}, P)$ is unique in the sense that P is uniquely defined by Pr. We have that $(\mathscr{F},P)$ is a probability algebra. Furthermore, the probability algebras $(\mathscr{F},P)$ and $(\mathscr{B},Pr)$ are isomorphic.

PROOF. We take $\Omega = \{\bigwedge_{i=1}^{n} L_i \mid L_i = a_i \text{ or } L_i = \neg a_i,\ a_i \in \mathscr{A},\ i = 1,\ldots,n\}$, i.e. the elements of Ω are conjunctions of length n in which for each $i \leqslant n$ either a_i or $\neg a_i$ occurs. We have that Ω has 2^n elements. In the rest of this proof the elements of Ω are enumerated as $\omega_1,\ldots,\omega_{2^n}$.

Using De Morgan's laws and the distributive laws, any element of $\mathscr{B}$ can be represented as a disjunction of elements of Ω: for each $x \in \mathscr{B}$ there exists a unique set of indices $\mathscr{I} \subseteq \{1,\ldots,2^n\}$ such that $x = \bigvee_{i \in \mathscr{I}} \omega_i$ where $\omega_i \in \Omega$. For $x = \bigvee_{i \in \mathscr{I}} \omega_i$ we say that x is in *disjunctive normal form.*

We define a mapping ι as follows: for $x = \bigvee_{i \in \mathscr{I}_x} \omega_i$ where $\mathscr{I}_x \subseteq \{1,\ldots,2^n\}$, we take $\iota(x) = \{\omega_i \mid i \in \mathscr{I}_x\}$. Notice that ι is well-defined since for a given x the set $\mathscr{I}_x$ is unique.

It is obvious that we have $\iota(\mathit{false}) = \varnothing$ and $\iota(\mathit{true}) = \Omega$. Furthermore, we have the following properties of this mapping ι:

(1) $\iota(x_1 \wedge x_2) = \iota(x_1) \cap \iota(x_2)$, for all $x_1, x_2 \in \mathscr{B}$.

Suppose we have $x_1 = \bigvee_{i_1 \in \mathscr{I}_1} \omega_{i_1}$ and $x_2 = \bigvee_{i_2 \in \mathscr{I}_2} \omega_{i_2}$, where $\mathscr{I}_1, \mathscr{I}_2 \subseteq \{1,\ldots,2^n\}$. So, $x_1 \wedge x_2 = (\bigvee_{i_1 \in \mathscr{I}_1} \omega_{i_1}) \wedge (\bigvee_{i_2 \in \mathscr{I}_2} \omega_{i_2})$. Using the distributive laws, $x_1 \wedge x_2$ can be written as $\bigvee_{i_1 \in \mathscr{I}_1, i_2 \in \mathscr{I}_2} (\omega_{i_1} \wedge \omega_{i_2})$.

From our definition of Ω, we have that $\omega_{i_1} \wedge \omega_{i_2} = false$ for $i_1 \neq i_2$. It follows that $x_1 \wedge x_2 = \bigvee_{i \in \mathcal{I}_1 \cap \mathcal{I}_2} \omega_i$, and consequently $\iota(x_1 \wedge x_2) = \{\omega_i \mid i \in \mathcal{I}_1 \cap \mathcal{I}_2\} = \{\omega_{i_1} \mid i_1 \in \mathcal{I}_1\} \cap \{\omega_{i_2} \mid i_2 \in \mathcal{I}_2\} = \iota(x_1) \cap \iota(x_2)$.

(2) $\iota(x_1 \vee x_2) = \iota(x_1) \cup \iota(x_2)$, for all $x_1, x_2 \in \mathcal{B}$.

(3) $\iota(\neg x) = \overline{\iota(x)}$, for all $x \in \mathcal{B}$.

It is evident that the mapping ι is an isomorphism, i.e. has an inverse mapping ι^{-1}.

The atoms of the algebra $\mathcal{F}$ are the singletons of 2^Ω. The algebra $\mathcal{F}$ obviously equals $\{\iota(x) \mid x \in \mathcal{B}\}$. We have that $\mathcal{F}$ is a free Boolean algebra.

From the properties of the probability function Pr on $\mathcal{B}$ we have that P $(= Pr \circ \iota^{-1})$ is additive and [0,1]-valued on $\mathcal{F}$ and therefore is a probability function on $\mathcal{F}$. So, we have that $(\mathcal{F},P)$ is a probability algebra. Furthermore, it can easily be shown that $(\mathcal{B},Pr)$ and $(\mathcal{F},P)$ are isomorphic. ■

In the sequel, we will proceed from the point of view given by the probability algebra $(\mathcal{B},Pr)$.

3. Partially Specified Probability Functions

In the foregoing section we have introduced the notion of a probability algebra $(\mathcal{B},Pr)$ in which Pr is a probability function on $\mathcal{B}$ that assigns a number in the interval [0,1] to every Boolean combination of atomic propositions, and we have shown that a (unique) probability space can be constructed from such a probability algebra. In our introduction we have argued that in the fields in which expert systems are employed, such a total probability function Pr on the Boolean algebra of propositions of interest is generally not known. Only a number of specific probabilities are known or can be estimated by an expert, i.e. Pr is only specified on a limited number of Boolean combinations of atomic propositions. We have mentioned before that an expert system derives new information from the information available to the system by employing a reasoning process; the probability of the truth of this new information has to be computed from the probabilities that are known to the system, i.e. from the partially specified probability function Pr.

We discern several types of partial specifications:

(1) The 'partially specified' probability function is defined uniquely by the values that have been specified initially. So, in this case all probabilities of interest can be determined uniquely from the given values. In Section 3.1 we deal with such uniquely defined probability functions.

(2) The partially specified probability function can be extended in more than one way to an actual probability function: there do not exist unique values for the probabilities we are interested in. This case will be the topic of the Sections 3.2 and 3.3.

(3) There is no 'underlying' probability function that coincides with the initially given values of the partially specified 'probability function'. In this case the values that have been specified do not represent actual probabilities. We address this problem in Section 3.4.

The following definition provides some terminology for the notions introduced above.

Definition 3.1. *Let $\mathcal{B}$ be the Boolean algebra as defined in Definition 2.1 and let $\mathcal{C} \subseteq \mathcal{B}$. Let $P: \mathcal{C} \to [0,1]$ and $P': \mathcal{B} \to [0,1]$ be total functions. We use $P' \mid_{\mathcal{C}} = P$ to denote that P' restricted to $\mathcal{C}$ equals P, i.e. that P' is an extension of P to $\mathcal{B}$.*

P is consistently specified (or consistent, for short) if there is at least one probability function Pr on $\mathcal{B}$ such that $Pr \mid_{\mathcal{C}} = P$; otherwise, P is said to be inconsistently specified (or inconsistent). Furthermore, we say that P (uniquely) defines Pr or alternatively that P is a definition for Pr, if Pr is the only probability function on $\mathcal{B}$ such that $Pr \mid_{\mathcal{C}} = P$.

We remark that D.V. Lindley, A. Tversky and R.V. Brown use the terms *coherent* and *incoherent* instead of *consistent* and *inconsistent*, [7].

3.1. Uniquely Defined Probability Functions

In this section, we characterize necessary and sufficient conditions for a probability function Pr to be uniquely defined by the probabilities that have been specified initially.

We introduce the notion of a basis for a probability function.

DEFINITION 3.2. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1 and let $\mathscr{C} \subseteq \mathscr{B}$. $\mathscr{C}$ is called a basis for a probability function on $\mathscr{B}$ if for any consistent function P: $\mathscr{C} \to [0,1]$, there exists a unique probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$.*

DEFINITION 3.3. *Let $\mathscr{A}$ and $\mathscr{B}$ be defined according to Definition 2.1. We define the set $\mathscr{B}_0 \subseteq \mathscr{B}$ such that $\mathscr{B}_0 = \{\bigwedge_{i=1}^{n} L_i \,|\, L_i = a_i$ or $L_i = \neg a_i,\ a_i \in \mathscr{A},\ i = 1,...,n\}$.*

Notice that the set $\mathscr{B}_0$ has been introduced before in the proof of Theorem 2.4. The following result is rather straightforward.

LEMMA 3.4. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1 and let $\mathscr{B}_0$ be defined as above. Then, $\mathscr{B}_0$ is a basis for a probability function on $\mathscr{B}$.*

This basis $\mathscr{B}_0$ will be used frequently throughout the remainder of this paper. Notice that by definition we have that each consistent function P: $\mathscr{B}_0 \to [0,1]$ defines a probability function Pr on $\mathscr{B}$. The following lemma states another two sets that can easily be shown to be bases for probability functions on $\mathscr{B}$.

LEMMA 3.5. *Let $\mathscr{B}$ be defined according to Definition 2.1. Then, the set $\{\bigwedge_{i \in I} a_i \,|\, I \subseteq \{1,...,n\},\ a_i \in \mathscr{A}\}$ is a basis for a probability function on $\mathscr{B}$, and so is the set $\{\bigvee_{i \in I} a_i \,|\, I \subseteq \{1,...,n\},\ a_i \in \mathscr{A}\}$.*

PROOF. We only prove the lemma for $\{\bigwedge_{i \in I} a_i \,|\, I \subseteq \{1,...,n\},\ a_i \in \mathscr{A}\}$. The case for $\{\bigvee_{i \in I} a_i \,|\, I \subseteq \{1,...,n\},\ a_i \in \mathscr{A}\}$ follows by symmetry.

Let $\mathscr{C} = \{\bigwedge_{i \in I} a_i \,|\, I \subseteq \{1,...,n\},\ a_i \in \mathscr{A}\}$. Let P: $\mathscr{C} \to [0,1]$ be a consistent function on $\mathscr{C}$. We consider a probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. By definition we have $Pr(c) = P(c)$ for all $c \in \mathscr{C}$, i.e. the probabilities $Pr(c)$ coincide with the initially specified function values $P(c)$. First, the probabilities of conjunctions comprising negated elements of $\mathscr{A}$ are determined by recursively applying the rule

$$Pr(c \wedge \neg a) = Pr(c) - Pr(c \wedge a),\ c \in \mathscr{C},\ a \in \mathscr{A}.$$

Notice that this rule is derived from Lemma 2.3(2). The function values $Pr(x)$ for all other elements $x \in \mathscr{B} \setminus \mathscr{C}$ are now uniquely determined by recursively using the following properties from Lemma 2.3:

(1) $Pr(x_1 \vee x_2) + Pr(x_1 \wedge x_2) = Pr(x_1) + Pr(x_2)$, for all $x_1, x_2 \in \mathscr{B}$, and
(2) $Pr(x) + Pr(\neg x) = 1$, for all $x \in \mathscr{B}$.

Therefore, Pr is a unique extension of P. Since P is an arbitrary consistent function on $\mathscr{C}$, we have that $\mathscr{C}$ is a basis for a probability function on $\mathscr{B}$. ■

We now consider the more general case where we are only given probabilities for a number of arbitrary boolean combinations of atomic propositions. We first state some general results.

LEMMA 3.6. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1 and let the basis $\mathscr{B}_0$ be defined according to Definition 3.3. Let the elements of $\mathscr{B}_0$ be enumerated as b_i, $i = 1,...,2^n$. Then, for any probability function Pr on $\mathscr{B}$ we have*

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

The probabilities $Pr(b_i)$ will be called constituent probabilities.

PROOF. From our definition of $\mathscr{B}_0$ we have that for any i,j, $i \neq j$, $b_i \wedge b_j = false$, and furthermore $\bigvee_{i=1}^{2^n} b_i = true$. The result now follows from $Pr(true) = 1$ and the additivity of the probability function Pr. ■

LEMMA 3.7. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1. Let $\mathscr{B}_0$ be defined according to Definition 3.3 and let its elements be enumerated as b_i, $i = 1,...,2^n$. Then, for each $b \in \mathscr{B}$ there exists a unique set of indices $\mathscr{I}_b \subseteq \{1,...,2^n\}$, called the index set for b, such that $b = \bigvee_{i \in \mathscr{I}_b} b_i$.*

PROOF. Each element $b \in \mathscr{B}$ can be written in disjunctive normal form, i.e. can be represented uniquely as a disjunction of elements of $\mathscr{B}_0$ using De Morgan's laws and the distributive laws. So, there is a unique set of indices $\mathscr{I}_b \subseteq \{1,...,2^n\}$ such that $b = \bigvee_{i \in \mathscr{I}_b} b_i$. ■

LEMMA 3.8. *Let $\mathscr{B}$ be defined according to Definition 2.1. Let $\mathscr{B}_0$ with elements b_i, $i = 1,...,2^n$, be defined as in the foregoing. Furthermore, let $b \in \mathscr{B}$ and let $\mathscr{I}_b$ be the index set for b. Then for each probability function Pr on $\mathscr{B}$, $Pr(b) = \sum_{i \in \mathscr{I}_b} Pr(b_i)$.*

PROOF. From Lemma 3.7 we have $b = \bigvee_{i \in \mathscr{I}_b} b_i$. From the additivity of Pr and $b_i \wedge b_j = false$ for any i,j, $i \neq j$, we obtain the result stated above. ■

Now let $\mathscr{C} \subseteq \mathscr{B}$ and let $P\colon \mathscr{C} \to [0,1]$ be a consistent function on $\mathscr{C}$. Let $\mathscr{B}_0$ be the basis defined according to Definition 3.3. We consider a probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. Let the constituent probabilities $Pr(b_i)$, $b_i \in \mathscr{B}_0$, be denoted by x_i. Let the initially specified probabilities $P(c_i) = Pr(c_i)$, $c_i \in \mathscr{C}$, be denoted by p_i. From Lemma 3.8 and Lemma 3.6 we have obtained the following inhomogeneous *system of* $|\mathscr{C}| + 1$ *linear equations*:

$$\begin{array}{ccccccc}
d_{1,1}x_1 & + \ldots + & d_{1,2^n}x_{2^n} & = & p_1 \\
\cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot \\
\cdot & & \cdot & & \cdot \\
d_{k,1}x_1 & + \ldots + & d_{k,2^n}x_{2^n} & = & p_k \\
x_1 & + \ldots + & x_{2^n} & = & 1
\end{array}$$

where $k = |\mathscr{C}|$ and $d_{i,j} = \begin{cases} 0 & \text{if } j \notin \mathscr{I}_{c_i} \\ 1 & \text{if } j \in \mathscr{I}_{c_i} \end{cases}$. This system of linear equations has the 2^n unknowns $x_1,...,x_{2^n}$. We now use some of the notions from linear algebra. Further details on these notions can be found in any introductory book on linear algebra, for instance [8,9].

Let $\boldsymbol{p}$ denote the column vector of right-hand sides of the system of linear equations and $\boldsymbol{x}$ the column vector of unknowns. Furthermore, let $\boldsymbol{D}$ denote the *coefficient matrix*

$$\boldsymbol{D} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,2^n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ d_{k,1} & \cdots & d_{k,2^n} \\ 1 & \cdots & 1 \end{bmatrix}$$

and $(\boldsymbol{D}; \boldsymbol{p})$ the *augmented matrix*

$$(\boldsymbol{D}; \boldsymbol{p}) = \begin{bmatrix} d_{1,1} & \cdots & d_{1,2^n} & p_1 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ d_{k,1} & \cdots & d_{k,2^n} & p_k \\ 1 & \cdots & 1 & 1 \end{bmatrix}.$$

Then, the system shown above is equivalent to the following *matrix equation*:

$$\boldsymbol{Dx} = \boldsymbol{p}$$

In the sequel, we use the symbol $\mathscr{B}_0^{(\boldsymbol{D};\,\boldsymbol{p})}$ to denote the matrix equation obtained from the function P.

LEMMA 3.9. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1. Let $\mathscr{C} \subseteq \mathscr{B}$ and let P: $\mathscr{C} \rightarrow [0,1]$ be a consistent function on $\mathscr{C}$. Let $\mathscr{B}_0^{(\boldsymbol{D};\,\boldsymbol{p})}$ be the matrix equation obtained from P as described in the foregoing. For any probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$ we have that the vector of constituent probabilities $Pr(b_i)$, $b_i \in \mathscr{B}_0$, $i = 1,\ldots,2^n$, is a solution to the matrix equation $\mathscr{B}_0^{(\boldsymbol{D};\,\boldsymbol{p})}$.*

PROOF. The result immediately follows from the Lemmas 3.6 and 3.8. ■

We now assume that we are given a function P: $\mathscr{C} \rightarrow [0,1]$ which has been specified consistently, i.e. there exists at least one probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. From the previous lemma we have that the constituent probabilities $Pr(b_i)$ $(= x_i)$ of any such extension Pr of P constitute a solution vector to the matrix equation obtained from P. Our assumption therefore guarantees that $\mathscr{B}_0^{(\boldsymbol{D};\,\boldsymbol{p})}$ has at least one nonnegative solution, i.e. a solution in which for all i, $x_i \geqslant 0$. When solving $\mathscr{B}_0^{(\boldsymbol{D};\,\boldsymbol{p})}$ we have two possibilities:

(1) The matrix equation has a unique solution.
(2) The matrix equation has more than one solution.

Notice that if the matrix equation has more than one solution it may have nonnegative solutions as well as solutions in which at least one of the x_i's is less than zero. Case (2) will be addressed in the Sections 3.2 and 3.3. The following proposition concerns case (1).

PROPOSITION 3.10. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1. Let $\mathscr{C} \subseteq \mathscr{B}$ and let $P\colon \mathscr{C} \to [0,1]$ be a consistent function on $\mathscr{C}$. Let $\mathscr{B}_0^{(D;\,p)}$ be the matrix equation obtained from P as described in the foregoing. P uniquely defines a probability function on $\mathscr{B}$ if and only if $\mathscr{B}_0^{(D;\,p)}$ has a unique solution.*

PROOF.

$\Rightarrow$ Let P define the probability function Pr on $\mathscr{B}$. Then, by definition Pr is the only probability function on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. So, the constituent probabilities $Pr(b_i)$ for each $b_i \in \mathscr{B}_0$ are determined uniquely by the function values $P(c) = Pr(c)$, for all $c \in \mathscr{C}$, and can be computed using the properties of Pr mentioned in Definition 2.2 and Lemma 2.3. From Lemma 3.9 we have that the vector of constituent probabilities $Pr(b_i)$ is a solution to the matrix equation $\mathscr{B}_0^{(D;\,p)}$. Now suppose that this solution is not a unique one. Then, the matrix equation has infinitely many solutions and at least one other nonnegative one. This means that there is another vector of constituent probabilities that can be calculated from $P(c)$; but then Pr is not a unique extension of P, and therefore P does not define Pr. So, the vector of values $Pr(b_i)$ is a unique solution to the matrix equation.

$\Leftarrow$ Let $\mathscr{B}_0^{(D;\,p)}$ have a unique solution. Since we assumed P to be a consistent function we are guaranteed that this solution is a nonnegative one. This nonnegative solution is a vector of values $Pr(b_i)$ for all $b_i \in \mathscr{B}_0$, $i = 1,\ldots,2^n$. We take these values to be constituent probabilities of a probability function Pr on $\mathscr{B}$ and then use them to extend Pr to all other elements of $\mathscr{B}$. Since $\mathscr{B}_0$ is a basis, we have that all function values of Pr can be calculated uniquely from these constituent probabilities $Pr(b_i)$. Furthermore, since the constituent probabilities $Pr(b_i)$ have been obtained from the function values $P(c) = Pr(c)$, for all $c \in \mathscr{C}$, it follows that all function values of Pr can be calculated from these $P(c)$. From $\mathscr{B}_0^{(D;\,p)}$ having a unique solution, we have that Pr is the only probability function on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. So, P defines Pr.

■

COROLLARY 3.11. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1. A basis for a probability function on $\mathscr{B}$ has at least $2^n - 1$ elements.*

PROOF. Let $\mathscr{C}$ be a subset of $\mathscr{B}$ such that $\mathscr{C}$ is a basis for a probability function on $\mathscr{B}$. We assume that $|\mathscr{C}| < 2^n - 1$. Consider the matrix equation $\mathscr{B}_0^{(D;\,p)}$ obtained from a consistent function $P\colon \mathscr{C} \to [0,1]$. We recall that this matrix equation is of the form $\boldsymbol{Dx} = \boldsymbol{p}$. The matrix equation $\boldsymbol{Dx} = \boldsymbol{p}$ has a unique solution if $rank(\boldsymbol{D}) = |\mathscr{B}_0| = 2^n$. In general, we have $rank(\boldsymbol{D}) \leqslant |\mathscr{C}| + 1$. Since $|\mathscr{C}| < 2^n - 1$, we have $rank(\boldsymbol{D}) < 2^n$. But then, the matrix equation does not have a unique solution and from Proposition 3.10 it follows that P does not uniquely define a probability function on $\mathscr{B}$; from Definition 3.2 we have that $\mathscr{C}$ is not a basis. So, $\mathscr{C}$ has at least $2^n - 1$ elements. ■

Notice that from Corollary 3.11 we only have a necessary condition for a consistent function $P\colon \mathscr{C} \to [0,1]$ where $\mathscr{C} \subseteq \mathscr{B}$, to be a definition of a probability function on $\mathscr{B}$ and not a sufficient one, whereas Proposition 3.10 states a necessary and sufficient condition.

DEFINITION 3.12. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1 and let $\mathscr{C} \subseteq \mathscr{B}$. $\mathscr{C}$ is called a minimal basis if $\mathscr{C}$ is a basis and $|\mathscr{C}| = 2^n - 1$.*

Let $P\colon \mathscr{C} \to [0,1]$ be a consistent function on $\mathscr{C}$. P is called a minimal definition of a probability function Pr on $\mathscr{B}$ if P defines Pr and $\mathscr{C}$ is a minimal basis.

COROLLARY 3.13. *Let $\mathscr{B}$ be the Boolean algebra as defined in Definition 2.1 and let $\mathscr{B}_0$ be defined according to Definition 3.3. Then, $\mathscr{B}_0$ is not a minimal basis.*

Notice that $\mathscr{B}_0$ contains just one element too many to be a basis. For, since $\mathscr{B}_0$ is finite we have that each constituent probability $Pr(b_i)$, $b_i \in \mathscr{B}_0$, can be expressed in terms of all other constituent probabilities: $Pr(b_i) = 1 - \sum_{j=1,\, j \neq i}^{2^n} Pr(b_j)$. The deletion of an arbitrary element from $\mathscr{B}_0$ therefore renders a minimal basis.

3.2. Partially Defined Probability Functions

In the foregoing section we have transformed the problem of determining a total probability function on $\mathscr{B}$ from a given partially and consistently specified probability function, into the problem in linear algebra of finding a nonnegative solution to an inhomogeneous system of linear equations: from a set $\mathscr{C} \subseteq \mathscr{B}$ and a consistent function $P: \mathscr{C} \rightarrow [0,1]$ on $\mathscr{C}$, we have constructed a matrix equation and we have shown that the function P uniquely defines a probability function Pr on $\mathscr{B}$ if this matrix equation has a unique solution. From this, it followed that a basis for a probability function on $\mathscr{B}$ has at least $2^n - 1$ elements. This means that when less than $2^n - 1$ probabilities have been specified initially, it is not possible to determine a unique probability function from these values, in general: there is more than one way to extend the function P to a probability function Pr on $\mathscr{B}$. In this and the following section we address the case where the matrix equation has more than one (nonnegative) solution, i.e. the case in which we are given a consistent function P that is not a definition of a probability function on $\mathscr{B}$ and therefore does not provide enough information to derive from its values a unique probability function.

In Proposition 3.10 we have shown that a consistent function $P: \mathscr{C} \rightarrow [0,1]$ uniquely defines a probability function on $\mathscr{B}$ if and only if the matrix equation $\mathscr{B}_0^{(D;\,p)}$ obtained from P has a unique solution. Notice that our consistency assumption guaranteed that the solution was a nonnegative one. In fact, a similar result can easily be proven for the case where P can be extended in more than one way to a probability function on $\mathscr{B}$.

PROPOSITION 3.14. *Let $\mathscr{B}$ be defined according to Definition 2.1. Let $\mathscr{C} \subseteq \mathscr{B}$ and let $P: \mathscr{C} \rightarrow [0,1]$ be a consistent function on $\mathscr{C}$. Let $\mathscr{B}_0^{(D;\,p)}$ be the matrix equation obtained from P. There exists more than one probability function Pr on $\mathscr{B}$ such that $Pr \mid_{\mathscr{C}} = P$ if and only if $\mathscr{B}_0^{(D;\,p)}$ has infinitely many solutions.*

Notice that although every probability function Pr corresponds uniquely with a solution to the matrix equation $\mathscr{B}_0^{(D;\,p)}$, not every solution to $\mathscr{B}_0^{(D;\,p)}$ corresponds with a probability function: $\mathscr{B}_0^{(D;\,p)}$ may have solutions in which at least one of the x_i's is less than zero.

From Proposition 3.14 we have that the problem of finding an extension of a partially and consistently specified probability function is equivalent to the problem of finding a particular nonnegative solution to the matrix equation obtained from the initially specified probabilities. In this subsection, we show how we may obtain a single extension of a partially specified probability function that can be extended in more than one way. In Section 3.3 we abandon the idea of determining a single extension of a partially specified probability function and we show how such a function can be used to calculate upper and lower bounds on those probabilities that are of interest. We will argue that the latter approach is more realistic within the context of expert systems.

Let $\mathscr{C}$ be a subset of $\mathscr{B}$ and let $P: \mathscr{C} \rightarrow [0,1]$ be a consistent function on $\mathscr{C}$ which is not a definition of a probability function on $\mathscr{B}$. Notice that $\mathscr{C}$ therefore is not a basis for a probability function on $\mathscr{B}$. Let $\mathscr{B}_0^{(D;\,p)}$ be the matrix equation obtained from P in the manner described in Section 3.1. Since P is not a definition of a probability function on $\mathscr{B}$ we have that $\mathscr{B}_0^{(D;\,p)}$ has infinitely many solutions; for the rank r of the coefficient matrix $\boldsymbol{D}$ we have that $r < 2^n$. In $\mathscr{B}_0^{(D;\,p)}$ we therefore have r *basic variables* and $2^n - r$ *free variables*. In general, to obtain a *particular solution* to the matrix equation the values of the free variables can be chosen arbitrarily, from which the values of the basic variables can then be computed uniquely. Every solution vector differs from the particular solution by a vector in the nullspace of $\boldsymbol{D}$.

Since we are dealing with probabilities however, in our case the values of the free variables cannot be chosen freely for we are only interested in solutions in which $0 \leqslant x_i \leqslant 1$ for all x_i. Again we are guaranteed that at least one nonnegative solution exists since we have assumed that P has been specified consistently. We confine ourselves to giving an example.

EXAMPLE 3.15. Let $\mathscr{A} = \{a_1, a_2, a_3\}$ and let $\mathscr{B}$ be the free Boolean algebra generated by $\mathscr{A}$. Let $\mathscr{C} = \{a_1 \wedge a_2, \neg a_1 \vee a_3, a_2, a_2 \wedge \neg a_3\}$ and let $P: \mathscr{C} \rightarrow [0,1]$ be a consistent function on $\mathscr{C}$. Notice that $\mathscr{C}$ cannot be a basis since it only contains four elements. We consider a probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. Suppose we have the following function values of Pr coinciding with the corresponding, initially given function values of P:

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

Now, let the elements of the basis $\mathscr{B}_0$ be enumerated as follows:

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

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

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

and the following corresponding matrix equation:

$$
\begin{pmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix}
=
\begin{pmatrix} 0.23 \\ 0.62 \\ 0.43 \\ 0.18 \\ 1 \end{pmatrix}
$$

We bring the augmented coefficient matrix $(D; p)$ in echelon form, thus obtaining:

$$\left[\begin{array}{cccccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0.23 \\ 0 & 1 & 1 & -1 & 1 & 1 & 0 & 1 & 0.39 \\ 0 & 0 & -1 & 1 & -1 & 0 & 0 & -1 & -0.19 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0.18 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0.20 \end{array}\right]$$

Since $rank((D; p)) = 5 < |\mathscr{B}_0| = 8$, we have $8 - rank((D; p)) = 3$ free variables and 5 basic variables. The constituent probabilities x_5, x_7 and x_8 are the free variables. Now, the values of these variables cannot be chosen freely, but must be chosen subject to $0 \leqslant x_i \leqslant 1$ and a number of restrictions that can be derived from the augmented matrix in echelon form.

For instance, from the last row of the echelon matrix we have that x_7 has to be chosen subject to $0.20 \leqslant x_7 \leqslant 1$. Suppose we choose $x_7 = 0.35$. Then, from the last row of the echelon matrix we obtain $x_6 = 0.15$. Using the fourth row, we have that $x_4 = 0.03$. From the first row, it follows that $x_1 = 0.2$. Now, from the third row it follows that we have to choose the values of the free variables x_5 and x_8 such that $0 \leqslant x_5 + x_8 \leqslant 0.19 + x_4 = 0.22$. We choose $x_5 = 0.1$ and $x_8 = 0.1$. The value of x_3 can now be calculated from the third row: $x_3 = 0.02$. From the second row, we obtain $x_2 = 0.05$. So, the vector

$$(0.2\ 0.05\ 0.02\ 0.03\ 0.1\ 0.15\ 0.35\ 0.1)^{\mathrm{T}}$$

is a particular solution to the matrix equation obtained from P. We recall that this solution vector provides a set of constituent probabilities. From these constituent probabilities we have a uniquely determined probability measure Pr on $\mathscr{B}$ that respects the probabilities that initially have been given. ■

In the foregoing example we have demonstrated in an informal manner how a single probability function Pr on $\mathscr{B}$ can be determined from a consistently specified function P that does not define Pr uniquely. Since the values of the free variables, i.e. some of the constituent probabilities, have been chosen more or less freely, there are other probability functions on $\mathscr{B}$ respecting the initially given probabilities, not equal to the function Pr defined by the solution vector mentioned above: every other nonnegative solution vector, differing from the particular solution shown in the example by a vector in the nullspace of D, defines another probability function on $\mathscr{B}$ which is an extension of P. It will be obvious from Example 3.15 that the more free variables occur in the matrix equation, the more arbitrary the selected probability function will be.

When we consider an application in expert systems, the results from using one solution vector can considerably differ from the results from using another solution vector. The method sketched in the previous example therefore does not render a reliable result.

3.3. Finding Bounds on Probabilities of Interest

In the foregoing subsection our aim has been to select a single probability function which is an extension of a partially and consistently specified probability function. Of the selected probability function the constituent probabilities were fixed, obtaining a representation of the selected function by which it is described uniquely. The probability function thus specified can then be employed by the expert system to compute all probabilities of interest. In this section we abandon the idea of selecting a single probability function that has to serve as the basis for further computations. We introduce a method for finding best possible upper and lower bounds on the probabilities we are interested in. The idea of finding bounds on probabilities originated with Boole, as well as the idea of obtaining the 'narrowest limits' ([6], page 338). In defining the notions of (best) upper bound and (best) lower bound we closely follow Hailperin, [6,10].

DEFINITION 3.16. *Let $\mathcal{B}$ be defined according to Definition 2.1. Let $\mathcal{C} \subseteq \mathcal{B}$ and let P: $\mathcal{C} \rightarrow [0,1]$ be a consistent function on $\mathcal{C}$. The function bub_P: $\mathcal{B} \rightarrow [0,1]$ is the best upper bound function relative to P if the following properties hold:*

(1) *for each probability function Pr on $\mathcal{B}$ such that $Pr|_{\mathcal{C}} = P$, we have $Pr(b) \leqslant bub_P(b)$ for each $b \in \mathcal{B}$, and*

(2) *bub_P is the minimal function satisfying* (1).

The best lower bound function blb_P is defined symmetrically.

Notice that the function bub_P relative to P in general is not a probability function; of course, the same remark can be made concerning blb_P. For a given $b \in \mathcal{B}$, the length of the interval $[blb_P(b), bub_P(b)]$ expresses the lack of knowledge concerning the probability of the truth of b.

LEMMA 3.17. *Let $\mathcal{B}$ be defined according to Definition 2.1. Let $\mathcal{C} \subseteq \mathcal{B}$ and let P: $\mathcal{C} \rightarrow [0,1]$ be a consistent function on $\mathcal{C}$. Furthermore, let bub_P and blb_P be as defined above. For each $b \in \mathcal{B}$ there exists a probability function Pr on $\mathcal{B}$ with $Pr|_{\mathcal{C}} = P$ such that $Pr(b) = bub_P(b)$. A similar result holds for blb_P.*

The two types of bounds are interrelated as stated in the following lemma.

LEMMA 3.18. *Let $\mathcal{B}$ be defined according to Definition 2.1. Let $\mathcal{C} \subseteq \mathcal{B}$ and let P: $\mathcal{C} \rightarrow [0,1]$ be a consistent function on $\mathcal{C}$. Let the functions bub_P and blb_P be defined according to Definition 3.16. Then for each $b \in \mathcal{B}$, $bub_P(b) = 1 - blb_P(\neg b)$.*

PROOF. Let $\mathcal{B}_0$ be the basis as defined in Definition 3.3 and let its elements be enumerated b_i, $i = 1,...,2^n$. Let $b \in \mathcal{B}$ and let $\mathcal{I}_b \subseteq \{1,...,2^n\}$ be the index set for b. From Lemma 3.8 we have that for each probability function Pr on $\mathcal{B}$ with $Pr|_{\mathcal{C}} = P$ the following holds:

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

Using the proof of Theorem 2.4, it can easily be shown that for each Pr

$$Pr(\neg b) = \sum_{i \in \overline{\mathcal{I}_b}} Pr(b_i) = 1 - Pr(b).$$

Since bub_P is an upper bound function and blb_P is a lower bound function relative to P, we have by definition for each Pr and each $b \in \mathcal{B}$

$$Pr(b) \leqslant bub_P(b), \text{ and}$$
$$blb_P(\neg b) \leqslant Pr(\neg b),$$

From $Pr(\neg b) = 1 - Pr(b)$, we have $blb_P(\neg b) \leqslant 1 - Pr(b)$, thus obtaining

$$Pr(b) \leqslant 1 - blb_P(\neg b),$$

for each $b \in \mathcal{B}$. From Definition 3.16(2) we have

$$bub_P(b) \leqslant 1 - blb_P(\neg b).$$

Reversing the argument, we can show

$$1 - blb_P(\neg b) \leqslant bub_P(b),$$

from which we obtain the desired result. ■

Let $\mathcal{C} \subseteq \mathcal{B}$ and let P: $\mathcal{C} \rightarrow [0,1]$ be a consistent function. Hailperin has shown that the problems of finding the best upper bound $bub_P(b)$ and the best lower bound $blb_P(b)$ relative to P for a given

$b \in \mathcal{B}$ are equivalent to the following *linear programming problems*:

(1) maximize $Pr(b)$ subject to $\mathcal{B}_0^{(D;\,p)}$ and $x_i \geqslant 0$ for all $i = 1,...,2^n$; and
(2) minimize $Pr(b)$ subject to $\mathcal{B}_0^{(D;\,p)}$ and $x_i \geqslant 0$ for all $i = 1,...,2^n$,

where $\mathcal{B}_0^{(D;\,p)}$ is the matrix equation obtained from P.

We consider case (1) in some detail in order to introduce some notions from linear programming. For further information on linear programming the reader is referred to [9,11,12]. Let $\mathcal{B}_0$ be the basis as defined in Definition 3.3 and let its elements be enumerated b_i, $i = 1,...,2^n$. Let $b \in \mathcal{B}$. From Lemma 3.8 we have that b has a unique index set $\mathcal{I}_b$. For this index set $\mathcal{I}_b$ we have

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

for each probability function Pr with $Pr\,|_{\mathscr{C}} = P$. Now, let constants c_i for b be defined by

$$c_i = \begin{cases} 0 & \text{if } i \notin \mathcal{I}_b \\ 1 & \text{if } i \in \mathcal{I}_b \end{cases}$$

Then,

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

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 $\mathcal{B}_0^{(D;\,p)}$ is a matrix equation of the form $\boldsymbol{Dx} = \boldsymbol{p}$ where $\boldsymbol{D}$ denotes a $(|\mathscr{C}| + 1) \times 2^n$ matrix, $\boldsymbol{x}$ is the 2^n column vector of constituent probabilities $Pr(b_i)$ and $\boldsymbol{p}$ is a $|\mathscr{C}| + 1$ 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 $\sum_{i=1}^{2^n} c_i x_i$

subject to

(i) $\sum_{j=1}^{2^n} d_{i,j} x_j = p_i$ for $i = 1,...,|\mathscr{C}| + 1$, and

(ii) $x_j \geqslant 0$ for $j = 1,...,2^n$.

where the constants $d_{i,j}$ constitute the matrix $\boldsymbol{D}$. The function $\sum c_i x_i$ is called the *objective function* of the linear programming problem. The conditions under which the objective function has to be maximized are called the *constraints*. Notice that we have added the constraints $x_j \geqslant 0$ for $j = 1,...,2^n$ explicitly to the set of constraints to allow only nonnegative solutions. A vector $\boldsymbol{x}$ of length 2^n satisfying the constraints is called a *feasible solution*. It can easily be shown, that the set of feasible solutions is a convex set, [11]. In the sequel the convex set of feasible solutions will be denoted by S. Notice that if the problem has more than one solution it has in fact an infinite number of solutions. A feasible solution that maximizes the objective function is called an *optimal solution*.

A linear programming problem usually is interpreted geometrically, in our case in 2^n-dimensional space. The constraints of type (i) each span a hyperplane and the constraints of type (ii) together denote the positive orthant in the 2^n-dimensional space. The *solution space* of the problem is the convex polyhedron in the positive orthant of the 2^n-dimensional space obtained from intersecting the $|\mathscr{C}| + 1$ hyperplanes spanned by the constraints (i). This solution space is the earlier mentioned convex set of feasible solutions S. By definition a convex polyhedron has a finite number of extreme points. Furthermore, every feasible solution in S can be expressed as a convex combination of the extreme points of S. From linear programming we have the following theorem, [11]: the objective function assumes its maximum at an extreme point of the solution space S. If it assumes its maximum at more than one extreme point, then it takes on the same value for every convex combination of those points. From this theorem it follows that we only have to consider the extreme points of the convex polyhedral solution space to determine an optimal solution. This idea is the basis of a well-known computation procedure for solving linear programming problems known as the *simplex method* which has been developed by G.B. Dantzig. Once any extreme-point feasible solution has been found, this method determines an optimal solution in a finite number of steps. In each step a neighboring extreme point is selected whose corresponding value of the objective function exceeds the value of the objective function for the preceding solution. This process is continued until an optimal solution has been reached. We do not consider this method in detail. Further details can be found in [11].

We conclude our discussion of linear programming by presenting another view to finding an optimal solution. We introduce one more notion: a *supporting hyperplane* to a convex polyhedron T is a hyperplane containing at least one point in T and having all of T on one side of the hyperplane. Now consider the family of parallel hyperplanes $\sum c_i x_i = k$ where $k \geqslant 0$. If for any arbitrary value k, there is a point X on the hyperplane $\sum c_i x_i = k$ that is in S, then the hyperplane contains at least one feasible solution. Now consider the largest value k^* such that $\sum c_i x_i = k^*$ contains at least one feasible solution but does not contain any interior points of S. Then, $\sum c_i x_i = k^*$ only contains boundary points of S and therefore is a supporting hyperplane of S. From the theorem mentioned above we have that each extreme point in S in the hyperplane $\sum c_i x_i = k^*$ represents an optimal solution. In fact, it will be evident that all points which $\sum c_i x_i = k^*$ and S have in common are optimal solutions to the linear programming problem. Since each point in the solution space represents a vector of constituent probabilities, it should intuitively be clear that this linear programming problem yields the value $bub_P(b)$ we are looking for. Hailperin has provided a formal proof for this statement in [6,10], using the argument we have stated in Lemma 3.17. Our linear programming problem (2) can be treated uniformly by taking for the objective function $-\sum c_i x_i$.

PROPOSITION 3.19. *Let $\mathscr{B}$ be defined according to Definition 2.1. Let $\mathscr{C} \subseteq \mathscr{B}$ and let P: $\mathscr{C} \rightarrow [0,1]$ be a consistent function on $\mathscr{C}$. Let $Dx = p$ be the matrix equation obtained from P. Furthermore, let the functions bub_P and blb_P be defined according to Definition 3.16. Then for any $b \in \mathscr{B}$, $bub_P(b)$ is equal to the solution of the linear programming problem*

Maximize $Pr(b)$
subject to

(*i*) $Dx = p$, *and*
(*ii*) $x \geqslant \mathbf{0}$.

A similar result holds for $blb_P(b)$.

Notice that since determining an upper bound on a given probability does not 'cut' feasible solutions from the solution space, we have that several different objective functions can be maximized independently.

We propose using the linear programming approach in a model for handling uncertainty in an expert system. To summarize, an expert is requested to assess certain probabilities. The assessed probabilities subsequently are used in the manner described in this section to compute upper and lower bounds on the probabilities that are of interest to the user in a given consultation. The following example illustrates the idea.

EXAMPLE 3.20. In this example we reconsider the situation described in Example 3.15 once more. Let $\mathscr{A} = \{a_1, a_2, a_3\}$ and let $\mathscr{B}$ be the free Boolean algebra generated by $\mathscr{A}$. Let $\mathscr{C} = \{a_1 \wedge a_2, \neg a_1 \vee a_3, a_2, a_2 \wedge \neg a_3\}$ and let P: $\mathscr{C} \rightarrow [0,1]$ be a consistent function on $\mathscr{C}$. We consider a probability function Pr on $\mathscr{B}$ such that $Pr\,|_{\mathscr{C}} = P$. From Example 3.15 we have

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

Let the elements of the basis $\mathscr{B}_0$ be enumerated as b_i, $i = 1,...,2^n$, as shown in Example 3.15. From P we obtained 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 now add the constraints

$$x_i \geqslant 0,\ i = 1,...,2^n$$

explicitly. Suppose we are interested in bounds on the probability of the truth of a_3. From Proposition 3.19 we have that the problem of determining the best upper bound of $Pr(a_3)$ 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 we obtain $bub_P(a_3) = 0.62$ and $blb_P(a_3) = 0.25$. Notice that in Example 3.15 we found $Pr(a_3) = 0.37$. ■

Although the linear programming approach is the heart of our method, several issues more or less specific to expert systems remain to be addressed. In Section 3.4 we discuss the problem that arises when the expert's assessed probabilities are inconsistent. Section 4 addresses some miscellaneous issues.

3.4. Inconsistently Specified Probability Functions

In the foregoing subsections we have dealt with partially specified probability functions that could be extended in at least one way to an actual probability function. Here we address the case where we are given a set of 'probabilities' which is inconsistent in the sense that when the given values are looked upon as values of a partially specified 'probability function', it is not possible to extend it to an actual total probability function.

EXAMPLE 3.21. Let $\mathscr{B}$ be the free Boolean algebra generated by $\mathscr{A} = \{a_1, a_2\}$. Let $\mathscr{C} = \{a_1 \wedge a_2, a_1\}$. Now consider the function P: $\mathscr{C} \rightarrow [0,1]$ of initially given values, defined by

$$P(a_1 \wedge a_2) = 0.34$$
$$P(a_1) = 0.28$$

It is evident that this function P cannot be extended to a probability function Pr on $\mathscr{B}$, since in every probability function the property if $x_1 \preccurlyeq x_2$ then $Pr(x_1) \leqslant Pr(x_2)$ holds for any $x_1, x_2 \in \mathscr{B}$. ■

In the foregoing example the inconsistency is easy to see. This type of evident inconsistency generally does not occur in a set of initially specified probabilities. The more probabilities are specified initially however, the harder it is to detect inconsistency. Furthermore, it will be obvious that the more probabilities are initially given, the more likely the 'probability function' is to be inconsistent.

We have discussed that the problem of finding an extension of a partially and consistently specified probability function P is equivalent to the problem of finding a nonnegative vector x that is a solution to the matrix equation $Dx = p$ obtained from P. We recall that a nonnegative solution vector x to the above mentioned equation denotes a vector of constituent probabilities. Since in the foregoing we assumed that the initially given probabilities were specified consistently we were guaranteed that the matrix equation $Dx = p$ had at least one nonnegative solution vector x. Now we have to reckon with the possibility of an inconsistently specified probability function, and we therefore are not guaranteed that this matrix equation has nonnegative solutions; in fact, the matrix equation can have no solution at all or can have only solutions x where at least one of the components x_i is less than zero. We therefore have to add the constraint $x \geqslant 0$ explicitly. In the sequel we use the phrase *system of linear constraints* to denote the matrix equation and the constraint $x \geqslant 0$. If we interpret the system of linear constraints geometrically in 2^n-dimensional space as we have done in Section 3.3, it will be evident that inconsistency corresponds with an empty solution space.

We address the question for which vectors q the system of linear constraints $Dx = q$ and $x \geqslant 0$ does have a solution. Let D_j denote the jth $|\mathscr{C}| + 1$ column vector of D. Then, the system of linear constraints may be expressed as

$$\sum_{j=1}^{2^n} x_j D_j = q, \text{ and}$$

$$x \geqslant 0$$

The system of linear constraints is consistent if and only if q lies in the convex polyhedral cone spanned by D_j, $j = 1,...,2^n$. In linear programming this convex cone is often called the *requirement space*. Already Boole has addressed the question of inconsistency: he has called the conditions on q for which consistency holds the *conditions of possible experience*, [6].

When we have a system of linear constraints consisting of only equalities, the well-known *Gaussian elimination* procedure can be employed to decide inconsistency or to derive the conditions of possible experience. Since our system of linear constraints also comprises inequalities we cannot use Gaussian elimination for this purpose. For systems comprising inequalities however there exists an equivalent computational procedure: the *Fourier-Motzkin elimination* method. This method can be employed to derive conditions of possible experience given a matrix D. We merely mention the existence of this elimination method; further details can be found in [6,12].

The linear programming approach discussed in Section 3.3 cannot be applied directly when we are given an inconsistently specified 'probability function': for a correct application of the method we have to have a consistently specified function. In Section 4 we introduce a method for obtaining a consistent function from the initially given inconsistent function. We emphasize that the method we discuss can only be viewed as an approximation technique and therefore is ad hoc; the best solution to the problem of an inconsistently specified 'probability function' is to reassess the function values.

4. Application in Expert Systems

In the foregoing section we have presented a method for computing upper and lower bounds on probabilities of interest from a given partially specified probability function. We have suggested before that this method can be employed in a model for plausible reasoning in expert systems. To summarize, an expert is requested to elucidate his knowledge concerning the domain to be modelled in the expert system. He furthermore is asked to assess certain probabilities to be associated with the represented pieces of information. During an actual consultation the system derives new information from the available information and determines which probabilities are of interest to the user. For these probabilities upper and lower bounds are established using the linear programming approach discussed in Section 3.3. In this section we address some remaining issues mostly typical for handling uncertainty in expert systems.

4.1. Reasoning With Production Rules

Many of the models that have been proposed during the last decade for handling uncertainty are devised to be incorporated in a rule-based expert system. In a rule-based expert system the domain knowledge is represented using the *production rule formalism*. The domain knowledge an expert has is formulated in statements having the following form: if certain conditions are fulfilled then a certain conclusion may be drawn. The domain expert is asked to estimate the probability of the truth of each of these statements. During an actual consultation of the system the production rules are used to derive new information. The production rules that actually succeed during such a consultation constitute an *inference network*, [13]. The internal nodes of an inference network represent intermediary conclusions. The models for handling uncertainty to be incorporated in a rule-based system aim at computing probabilities not only for the external nodes of interest to the user but also for all internal nodes, i.e. for all intermediary conclusions as well, [1,2,13]. None of these models although departing from Bayesian probability theory are fully correct, see for instance [14]. In particular since the production rules do not all represent causal or logical relationships, the developers of models for handling uncertainty in rule-based systems were faced with insuperable problems. We feel that it is not necessary to compute probabilities for all intermediary conclusions since only some conclusions are of interest to the user. We therefore propose that an inference network is employed in a heuristical manner to select those conclusions that are of interest and that only for the selected conclusions bounds on the probability are computed using the linear programming approach.

4.2. Conditional Probabilities

In the domains in which expert systems are employed it often is easier to estimate or otherwise obtain conditional probabilities than it is to obtain a priori probabilities. Moreover, the user of the system usually is interested in conditional probabilities. In the foregoing sections we have only considered a priori probabilities. In this subsection we show that conditional probabilities can be introduced in our method without any extra difficulties.

We first examine the case where we are initially given some conditional probabilities. Let $\mathscr{B}$ again be the free algebra of Boolean combinations of propositions as defined in Definition 2.1 and let $c_1, c_2 \in \mathscr{B}$. Let $\mathscr{B}_0$ be the basis as defined in Definition 3.3 and let its elements be enumerated b_i, $i = 1,\ldots,2^n$. We consider a probability function Pr on $\mathscr{B}$ which is an extension of a given partially specified function. Now suppose the expert has assessed the value $Pr(c_1 | c_2) = p_0$ to be taken as a conditional probability. By definition we have $Pr(c_1 | c_2) = \dfrac{Pr(c_1 \wedge c_2)}{Pr(c_2)}$. From Lemmas 3.7 and 3.8 we have that there exist an index set $\mathscr{I}_{c_1 \wedge c_2}$ for $c_1 \wedge c_2$ such that $Pr(c_1 \wedge c_2) = \sum_{i \in \mathscr{I}_{c_1 \wedge c_2}} Pr(b_i)$ and an index set $\mathscr{I}_{c_2}$ for c_2 such that $Pr(c_2) = \sum_{i \in \mathscr{I}_{c_2}} Pr(b_i)$ where $Pr(b_i)$ are constituent probabilities of Pr. We therefore have

$$\frac{\sum_{i \in \mathscr{I}_{c_1 \wedge c_2}} Pr(g_i)}{\sum_{i \in \mathscr{I}_{c_2}} Pr(g_i)} = p_0.$$

It follows that $\sum_{i \in \mathscr{I}_{c_1 \wedge c_2}} Pr(g_i) = p_0 \cdot \sum_{i \in \mathscr{I}_{c_2}} Pr(g_i)$. So, we have obtained the constraint

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

This constraint is similar in concept to the ones we have encountered in our linear programming problems and can therefore be treated likewise.

In the case we are interested in lower and upper bounds on a conditional probability, we have a fractional objective function in our linear programming problem. A fractional linear programming problem can be reduced to solving a related ordinary linear programming problem with one more variable. The following theorem formulated in [6] but originally due to Charnes, states this result.

THEOREM 4.1. *The linear fractional problem*

Maximize $\frac{\mathbf{cx}}{\mathbf{gx}}$ *subject to* $\mathbf{Dx} = \mathbf{p}$, *and* $\mathbf{x} \geqslant \mathbf{0}$.

is equivalent to the linear programming problem

Maximize $\mathbf{cy}$ *subject to* $\mathbf{Dy} = t\mathbf{p}$, $\mathbf{gy} = 1$, $\mathbf{y} \geqslant \mathbf{0}$, *and* $t \geqslant 0$.

4.3. *Extension to Inequalities*

In Section 3.3 we have shown that computing upper and lower bounds on a probability of interest from a partially specified probability function is equivalent to a linear programming problem in which all constraints except $\mathbf{x} \geqslant \mathbf{0}$ are equalities. In general, a linear programming problem having 2^n variables takes the following form:

Maximize $\sum_{j=1}^{2^n} c_j x_j$

subject to

(*i*) $\sum_{j=1}^{2^n} d_{i,j} x_j \geqslant p_i$ for $i = 1,\ldots,k$, $k \geqslant 0$,

(*ii*) $\sum_{j=1}^{2^n} d_{i,j} x_j = p_i$ for $i = 1,\ldots,l$, $l \geqslant 0$,

(*iii*) $\sum_{j=1}^{2^n} d_{i,j} x_j \leqslant p_i$ for $i = 1,\ldots,m$, $m \geqslant 0$,

(*iv*) $x_i \geqslant 0$ for $i = 1,\ldots,2^n$, $n \geqslant 1$.

It is evident that the linear programming problem we have obtained in the foregoing is a special case of the one shown above. In our case we have $k = m = 0$; so, we only have constraints of the types (*ii*) and (*iv*). It will be obvious that our method can be extended to include constraints of the other types as well: an expert is then provided with the possibility of expressing bounds on probabilities instead of point estimates.

4.4. *Dealing With Inconsistently Specified Probability Functions*

In an expert system context it is likely that the 'probabilities' which have been assessed by the expert are inconsistent. Such an inconsistently specified 'probability function' cannot be employed in establishing upper and lower bounds on the probabilities of interest using the linear programming approach proposed in the foregoing. In this subsection we propose a method for obtaining a consistent set of probabilities from the initially inconsistently specified 'probability function'.

The following proposition applies to an inconsistently specified 'probability function' in which the initially specified values are in correct proportion, i.e. in which these values behave additively. In such a case the inconsistency is due to the fact that the sum of the 'constituent probabilities' is not equal to 1. Lindley, Tversky and Brown state that such inconsistencies are quite common when the number of events exceeds two or three, [7].

PROPOSITION 4.2. *Let* $\mathscr{B}$ *be defined according to Definition 2.1. Let* $\mathscr{C} \subseteq \mathscr{B}$ *and let* $P: \mathscr{C} \to [0,1]$ *be an inconsistent function on* $\mathscr{C}$. *Let* $\mathbf{Dx} = \mathbf{p}$ *and* $\mathbf{x} \geqslant \mathbf{0}$ *constitute the system of linear constraints obtained* *from P, where $\boldsymbol{D}$ is a $(|\mathscr{C}|+1) \times 2^n$ matrix, $\boldsymbol{x}$ is a 2^n column vector and $\boldsymbol{p}$ is a $|\mathscr{C}|+1$ column vector. Furthermore, let $\boldsymbol{D}^-$ denote the $|\mathscr{C}| \times 2^n$ matrix obtained from $\boldsymbol{D}$ by omitting its last row; let $\boldsymbol{p}^-$ equally denote the $|\mathscr{C}|$ column vector obtained from $\boldsymbol{p}$ by omitting its last component. If the system of linear constraints $\boldsymbol{D}^-\boldsymbol{x} = \boldsymbol{p}^-$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ has a solution and $\boldsymbol{p}^- \neq \boldsymbol{0}$, then there exists a scalar $k > 0$ such that the system of linear constraints $\boldsymbol{D}\boldsymbol{x} = \begin{bmatrix} k\boldsymbol{p}^- \\ 1 \end{bmatrix}$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ has a solution.*

PROOF. Let the system of linear constraints $\boldsymbol{D}^-\boldsymbol{x} = \boldsymbol{p}^-$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ have at least one solution. We consider such a solution vector $\boldsymbol{x}'$ with components x_j', $j = 1,...,2^n$. For each component of $\boldsymbol{p}^-$, p_i, $i = 1,...,|\mathscr{C}|$, we have

$$\sum_{j=1}^{2^n} d_{i,j}x_j' = p_i,$$

where $d_{i,j}$ constitutes $\boldsymbol{D}^-$. From the system of linear constraints $\boldsymbol{D}\boldsymbol{x} = \boldsymbol{p}$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ not having a solution, we have that $\sum x_j' < 1$ or $\sum x_j' > 1$. Furthermore, from $\boldsymbol{p}^- \neq \boldsymbol{0}$ it follows that $\sum x_j' > 0$. So, we have

$$\frac{\sum_{j=1}^{2^n} d_{i,j}x_j'}{\sum_{j=1}^{2^n} x_j'} = \frac{p_i}{\sum_{j=1}^{2^n} x_j'}$$

for $i = 1,...,|\mathscr{C}|$. Now, let y_j' denote $\dfrac{x_j'}{\sum_{j=1}^{2^n} x_j'}$. Then we have

$$\sum_{j=1}^{2^n} d_{i,j}y_j' = \frac{p_i}{\sum_{j=1}^{2^n} x_j'}$$

for $i = 1,...,|\mathscr{C}|$. Since $\boldsymbol{x}'$ is a solution to $\boldsymbol{D}^-\boldsymbol{x} = \boldsymbol{p}^-$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ we have that $\boldsymbol{y}'$ is a solution to $\boldsymbol{D}^-\boldsymbol{x} = k\boldsymbol{p}^-$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ where $k = \dfrac{1}{\sum_{j=1}^{2^n} x_j'}$. Furthermore, we have $\sum y_j' = 1$. So, $\boldsymbol{y}'$ is a solution to $\boldsymbol{D}\boldsymbol{x} = \begin{bmatrix} k\boldsymbol{p}^- \\ 1 \end{bmatrix}$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$. ■

The basic idea of Proposition 4.2 is that the solution space of the system of linear constraints $\boldsymbol{D}^-\boldsymbol{x} = \boldsymbol{p}^-$ and $\boldsymbol{x} \geqslant \boldsymbol{0}$ is moved along the $\boldsymbol{p}^-$ vector towards the origin or just away from the origin dependent upon whether $\sum x_i > 1$ or $\sum x_i < 1$, so that the intersection of the shifted solution space and the hyperplane $\sum x_i = 1$ is not empty. It will be obvious that there exist many scalars having the property mentioned in the proposition, obtained from different points in the solution space of the original system of linear constraints.

The method for obtaining a consistent set of probabilities from an inconsistently specified 'probability function' which does not behave additively, is based on the idea of allowing an expert to make a certain mistake in his assessments. For each probability the expert has estimated we add two inequalities to the system of linear constraints instead of one equality. Let m be a constant value far less than 1 representing the margin we allow the expert to be mistaken in his assessments. When an expert has given the value $Pr(c) = \sum_{j=1}^{2^n} d_{i,j}x_j = p_0$, we add the following inequalities to the system of constraints:

$$(1)\quad \sum_{j=1}^{2^n} d_{i,j}x_j \leqslant p_0^+ \text{ where } p_0^+ = \begin{cases} p_0 + m & \text{if } p_0 + m < 1 \\ 1 & \text{otherwise} \end{cases}$$

$$(2)\quad \sum_{j=1}^{2^n} d_{i,j}x_j \geqslant p_0^- \text{ where } p_0^- = \begin{cases} p_0 - m & \text{if } p_0 - m > 0 \\ 0 & \text{otherwise} \end{cases}$$

Notice that instead of a hyperplane we have specified a 'band' in 2^n-dimensional space. We consider the system of linear constraints that in this way has been obtained from the expert's assessments. When this system has an empty solution space, the assessments cannot be used and the expert has to reassess the requested probabilities. When the system of these constraints however has at least one feasible solution, the equation $\sum x_i = 1$ is added (after applying Proposition 4.2, if necessary). We proceed with the resulting system of linear constraints. A problem with this approach is that the constraints in the original matrix equation are treated as being equally trustworthy. If the expert however is more certain of some of his assessments than of the other ones, we can attach for each constraint a weighting factor to the margin m thus obtaining a constraint-specific margin determining the width of the specified band.

In [7], several other methods for repairing inconsistently assessed 'probability functions' are proposed. We feel that the method we have presented is much more to the point in an expert system context.

4.5. Computational Complexity of the Linear Programming Approach

An important issue concerning the applicability of our method is the computational complexity of solving a linear programming problem. In this section we state a few results concerning the complexity of some computational methods. These results have all been formulated in considerable detail by A. Schrijver in [12].

In Section 3.3 we have mentioned the well-known simplex method for solving linear programming problems. The worst-case behaviour of this method is exponential in the size of the problem, i.e. the number of constituent probabilities. On the average however the simplex method can be shown to be polynomial. In practice, the simplex method has proved to be very efficient: experience suggests that the number of steps necessary to solve a given linear programming problem is about linear in the size of the problem. The simplex method however is not a polynomial-time method. Beside this simplex method other computational methods for solving linear programming problems have been developed, some of which have been shown to be polynomial in time, such as Khachiyan's ellipsoid method and a recently presented method by Karmarkar.

5. Conclusion

Bayesian probability theory seems a natural point of departure for the development of a model for dealing with uncertain information in expert systems. Probability theory however cannot be applied as a model for handling uncertainty in a traditional way since in the problem domains for which expert systems are being developed usually only partially and inconsistently specified probability functions are known. In this paper we have presented a method for determining upper and lower bounds of the probability of a certain event given such a partially and inconsistently specified function. Hereto we have employed in the context of expert systems ideas that originated with G. Boole and that have been extended by T. Hailperin.

The method we propose is firmly rooted in linear algebra, a mathematically well-founded theory. Since the heart of our method consists of employing linear programming techniques, our method is well-founded and can be shown to be computationally feasible. For embedding our linear programming approach in an expert system context, we have proposed solutions to some of the notorious problems one encounters when handling uncertainty in such a knowledge-based system. Only our proposals for obtaining a consistent probability function from an inconsistently specified set of 'probabilities' can be considered to be ad hoc.

In the model we have presented however, we are only able to deal with a 'naive' representation of a probability function; for instance, it is not evident how we may represent and exploit (statistical) independence of propositions. Future research will address this problem.

REFERENCES

[1] 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.

[2] R.O. DUDA, P.E. HART, N.J. NILSSON (1976). *Subjective Bayesian Methods for Rule-Based Inference Systems*, Technical Note 124, Artificial Intelligence Center, SRI International, Menlo Park.

[3] D.J. SPIEGELHALTER (1986). Probabilistic reasoning in predictive expert systems, in: L.N. KANAL & J.F. LEMMER (eds.), *Uncertainty in Artificial Intelligence*, vol. 1, North-Holland, Amsterdam.

[4] P.J.F. LUCAS & L.C. VAN DER GAAG (1988). *Principles of Expert Systems*, Academic Service (in Dutch), to appear in English.

[5] 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).

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

[7] D.V. LINDLEY, A. TVERSKY & R.V. BROWN (1979). On the reconciliation of probability assessments, *Journal of the Royal Statistical Society, Series A*, vol. 142, part 2, pp. 146 - 180.

[8] S. LIPSCHUTZ (1968). *Linear Algebra*, Schaum's Outline Series, McGraw-Hill, New York.

[9] G. STRANG (1976). *Linear Algebra and Its Applications*, Academic Press, New York.

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

[11] S.I. GASS (1975). *Linear Programming. Methods and Applications*, McGraw-Hill Book Company, New York.

[12] A. SCHRIJVER (1986). *Theory of Linear and Integer Programming*, John Wiley & Sons, Chichester.

[13] L.C. VAN DER GAAG (1987). *A Network Approach to the Certainty Factor Model*, Report CS-R8757, CWI, Amsterdam, to appear in: *International Journal of Approximate Reasoning*.

[14] L.C. VAN DER GAAG (1988). *The Certainty Factor Model and Its Basis in Probability Theory*, Report CS-R8816, CWI, Amsterdam.