Combining Propositional Logic with Maximum Entropy Reasoning on Probability Models

Manfred Schramm[1] and **Stephan Schulz**[2]

Abstract. We present a system for non-monotonic reasoning based on the probability calculus. This calculus incorporates this type of reasoning in two ways: Non-monotonic decisions (which can be treated as decisions under incomplete knowledge as well) can be the result of reasoning in a single probability model (via conditionalization) or in a set of probability models (via additional principles of rational decisions). But probability theory is too fine-grained to model common sense reasoning in general (think about "paradoxes" due to the unexpected existence of certain P-Models ([2, 15])). The remaining degrees of freedom have to be filled (of course without introducing subjective biases). We therefore use additional (context-sensitive) constraints (resp. principles), which are able to support rational decisions based on incomplete knowledge. These principles have to be global (context-dependent on all assumptions) to avoid loosing the sensitivity of the language to the assumptions. The central principle of rational decisions used by our system is the method of Maximum Entropy (MaxEnt), which is a well founded extension of probability theory with global properties ([12, 10, 8, 16]).

1 Introduction

Propositional logic is a well-researched area of science. It allows the specification of many kinds of exact knowledge and has applications in as diverse fields as machine learning, integrated circuit design and mathematical group theory. The interest in these areas has led to the development of efficient decision procedures for propositional logic and mature implementations. However, propositional logic lacks the features necessary for modeling most real world situations and common sense reasoning. Consequently, such attempts are relegated to introductory textbooks.

There are two particular features missing from propositional logic for modeling common sense reasoning: The ability to describe *uncertainty* and the ability to support plausible reasoning under *incomplete knowledge*. In this paper we will deal with this issues by mating propositional logic with reasoning on probability models.

To achieve this goal we allow the propositional assumptions to be accompanied by probabilities (or probability intervals) to specify a range of certainty. We also use classical conditional probability statements to model the common sense conjunctive *if-then*. We call such a (conditional or unconditional) probability statement a *constraint*, and a specification in our approach consists of a set of these constraints.

For any set of constraints we can compute a unique *probability model* (or *P-Model*) over the interpretations of the propositional variables (or, as they are also called, *possible worlds*), that is, we can compute a probability for each possible assignment of truth values to the propositional variables. We achieve this by searching for the P-Model with maximum entropy that is consistent with the set of constraints.

Given a complete P-Model, we can compute the probability of an arbitrary propositional sentence by summing the probabilities of the interpretations which evaluate the sentence to true. Similarly, we can calculate the probabilities of conditional queries.

The search for the P-Model with maximum entropy (the *MaxEnt-Model*) is expensive, as the number of interpretations grows exponentially with the number of propositional variables. However, we can use known properties of the MaxEnt-Model to reduce the search space. In particular, we can eliminate interpretations that are inconsistent at a propositional level, and we can apply the principle of *indifference* to represent sets of equivalent interpretations by a single element.

We have implemented these ideas in the PIT system for reasoning with uncertain and incomplete knowledge.

This paper is organized as follows: After this overview, we will give a more formal introduction to probability models on propositional interpretations. We will then describe and justify the maximum entropy principle and touch its connections to the principle of indifference and independence. The next section deals with the efficient calculation of the MaxEnt-Model for a given set of constraints. After that, we will describe our implemented system, PIT. We will conclude with an example of common sense reasoning with the system. This example also illustrates how default reasoning can be implemented in our approach.

2 Propositional Logic and P-Models

In this section we will give a short introduction to the concepts of probability models on propositional interpretations. For a more formal description we refer to [19] or any introduction to reasoning with probabilities.

Let $V = \{x_1 \ldots x_n\}$ be a finite set of propositional variables (*atoms*), and let $Form(V)$ be the set of propositional formulas in these variables (using e.g. the operators $\{\neg, \vee, \wedge, \rightarrow\}$ for *not*, *or*, *and* and the *material implication*). A literal is either an atom or a negated atom. An *interpretation* of the variables

[1] Institut für Informatik, Technische Universität München, Arcisstraße 20, D-80290 München, Germany, +49/89/521096, schramma@informatik.tu-muenchen.de

[2] ditto, schulz@informatik.tu-muenchen.de

is a mapping $I : V \to \{true, false\}$. An interpretation can be extended to a function $I : Form(V) \to \{true, false\}$ by the usual rules for the application of the operators. We also call an interpretation a *possible world* and denote it by the unique full conjunction[3] of literals that evaluates to *true* under it. A propositional specification S is a set of formulas, and an interpretation that evaluates all formulas from S to true is called a *propositional model* of S.

Now let $\Omega = \{\omega_1, \dots, \omega_m\}$ be a finite set of disjoint *elementary events* (i.e. only one of the ω_i can occur at any one time or situation). A *probability measure* on Ω is a function $P : \Omega \to [0, 1]$ with

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

The probability measure is continued to a function $P : 2^\Omega \to [0, 1]$ by

$$P(A) := \sum_{\omega \in A} P(\omega) \text{ for } A \subseteq \Omega$$

A *linear probabilistic constraint* (or simply *constraint* from here on) is a statement of the form $P(A) \in J$ or $P(A|B) \in J$, where $A, B \subseteq \Omega$ and J is a subinterval of $[0, 1]$. [4]

A *probability model (P-Model)* for a set of constraints DB is a probability measure which fulfills the constraints from DB. We now chose the interpretations for a set of propositional variables as the elementary events of a probability measure. We can then identify a propositional formula with the set of its (propositional) models. Thus, probabilistic constraints can be written as (conditional) probabilities on propositional formulas. Probabilistic constraints with the probability 0 or 1 still act like purely propositional statements. We call them *logical constraints.*

If we have a unique P-Model for a set of constraints, we can easily compute the probability of a proposition by summing the probabilities of its models. We can also calculate a conditional probability $P(a|b)$ by comparing the total probability of the models of $a \wedge b$ to the probability of models of b $(a, b \in Form(V))$. Unfortunately, in most cases the constraints of a specification do not, without further principles, determine a unique P-Model, but rather a (infinite) set of P-Models. To select a single, representative model from this set we apply the principle of *maximum entropy.*

3 The Maximum Entropy Principle

The maximum entropy principle is a well known principle to choose a certain (preferred) probability model from an (infinite) set of possible P-Models. ([10, 13, 16, 5]). It solves this task without introducing arbitrary and unnecessary information or personal bias by the choice. Formally, the principle of maximum entropy (*MaxEnt*) selects the P-Model with maximal entropy. Following [20], the entropy for a given P-Model P over Ω is defined as

$$H(P) = -\sum_{\omega \in \Omega} P(\omega) ln(P(\omega))$$

[3] A *full conjunction* is a conjunction of literal where each atom from V appears either in atomic form or in negated form.

[4] We write $P(\dots) = p$ for the special case of $J = [p, p]$.

If the set of P-Models is generated by linear inequalities on $P(A)$ with $A \subseteq \Omega$, as in our case, this choice is known to lead to an unique P-Model P^* (e.g. [13])

In addition to explaining the connection between $H(P)$ and the amount of information in the distribution P, MaxEnt can be justified via stating a set of plausible axioms of inductive reasoning and deriving MaxEnt as unique consequence of this axioms ([16, 21, 22]), or with more quantitative arguments like the Concentration Theorem or the Wallis Derivation of entropy (see [11, 12, 9] for detailed descriptions). These quantitative arguments observe that for discrete models of the probabilities (e.g. urns with a given number N of balls), where the probability mass of a distribution can be calculated by classical discrete (combinatorial) methods, the mass of the distributions concentrates around the MaxEnt distribution P^* if we refine the model (by increasing N).

We now make use of MaxEnt in two different forms, whose relation can again be obtained from Fig. 1. The weaker form (less conclusions are supported) we call *worst case analysis.* Here we choose the values in the intervals first and apply MaxEnt afterwards. A conclusion is only supported if MaxEnt accepts the conclusion for any possible choice of the values in the intervals. The stronger form (more conclusions are supported) is also the simpler form: just take the P-Model with maximal entropy, where the constraints may range in the specified intervals.

4 MaxEnt, Indifference and Independence

Solving the MaxEnt-Problem (i.e. calculating the MaxEnt-Model for an under-constrained probabilistic problem) is not a new idea (see e.g. [7]). It is also well known that the MaxEnt principle has connections to some basic form of the principle of indifference, and that it also provides connections to the notion of independence.

Basically, aspects of these properties can already be recognized in some axiomatic derivations of MaxEnt. Informally, we can state that MaxEnt maximizes indifference and independence. By defining different logics on P-Models by adding different axioms (Logic on P-Models with indifference, Logic on P-Models with independence, Logic on P-Models with indifference and independence, Logic with worst case analysis of MaxEnt-Models and Logic with MaxEnt, see Fig. 1 for an example of these logics and their logical relations; see [19] for a more extended description), these connections can be clarified and used for reducing the search space of the optimization process.

Both principles, indifference and independence, can be used for this purpose. The following lines will shortly state our use of the principle of indifference. The strong connections between MaxEnt and qualitative models of independence (as delivered by undirected graphs ([17])) are not currently implemented in our system, because our main application (in a medical domain, see section 8) contains highly connected concepts. Therefore, the use of indifference is more likely to lead to significant simplifications than the use of independence.

The Principle of Indifference

The history of this famous principle goes back to Laplace and Keynes. Let us quote ([10]) for a short and informal definition

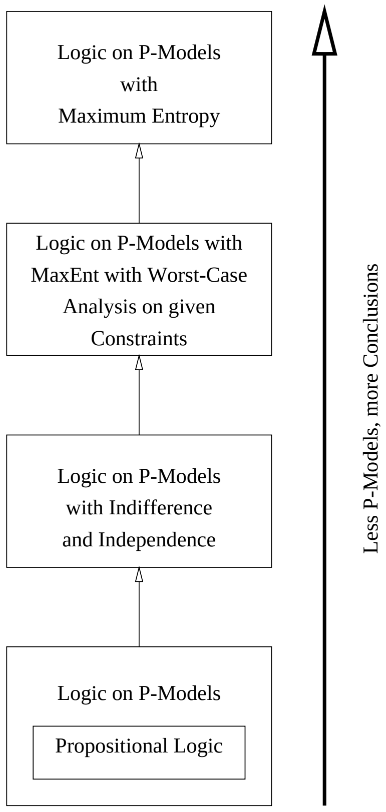

Figure 1. Logics on P-Models

of this principle:

> "If the available evidence gives us no reason to consider proposition a_1 either more or less likely as a_2, then the only honest way we can describe that state of knowledge is to assign them equal probabilities: $P(a_1) = P(a_2)$."

Certainly, this qualitative description leaves open many questions of its precise meaning, but it describes a more extended principle than the definition of indifference as *assigning equal degrees of belief to all basic "situations" consistent with the knowledge base* (compare e.g. [1]), which often is simplified to demanding that the equation $P(\omega) = 1/|\Omega|$ holds if no constraint is present.
The following definition will sketch a method on how to decide whether the available evidence (i.e. a set of constraints) gives us reason to consider a proposition more or less likely than an other (see [19] for a more extended description).
We define a possible world ω_1 to to be indifferent from ω_2 if a renaming (permutation) of all possible worlds exists, which does not change the set of P-Models, and if ω_1 and ω_2 are within the same cycle of this permutation.
As consequence of detecting two possible worlds ω_1, ω_2 as indifferent, we demand $P(\omega_1) = P(\omega_2)$ as an additional constraint. It can be shown that this additional constraint does not change the resulting MaxEnt-Model.

5 Efficient Calculation of the MaxEnt-Model

We search for the MaxEnt-Model by applying a numerical optimization algorithm. It can be shown that the optimization problem for the entropy function with linear constraints is very simple. The optimization searches in a convex space, and the second order derivations of the optimization problem are defined, and can be calculated efficiently. Therefore, efficient algorithms (developed by mathematicians) for the task already exist. We decided to use an implementation of the SQP (Sequential Quadratic Programming, see [23]) algorithm and an implementation of the KOMPLEX-Algorithm for the (more difficult) worst case analysis[5]([3]).
Corresponding to the theoretical connections of the different logics in 1, we know explain how propositional constraints and the principle of indifference does help for an efficient calculation of the MaxEnt-Model by reducing the search space for the optimization process.

1. The first step consists of reducing the set of propositional interpretations by checking for logical constraints. For the representation of this set we use *Binary Decision Diagrams* ([4]), which allow an efficient representation for domains of the size we are interested in (less than 64 Boolean variables). If the set of assumptions is already inconsistent at the propositional level, the process stops here and gives appropriate feedback. Also, the percentage of possible worlds in which some propositional expressions are true can be obtained at this stage.
2. To further reduce the number of variables for the optimization process, we generate a table of the linear propositional constraints in some normal form, perform consistency checks (especially in the use of the intervals), and apply the principle of indifference.
   To check for indifferent events algorithmically, we use sufficient conditions to detect efficiently some special permutations in the table of constraints that do not change the space of P-Models. If we find some possible worlds to be indifferent, these are collected into an equivalence class by joining the corresponding columns of the table of constraints. By that operation, some rows (constraints) of the table become redundant (this does not indicate that the constraints can be omitted in the specification), and can be omitted (which also leads to a more efficient optimization). See section 7 for an example of the effect.
3. At the next step, the simplified optimization problem is passed to our subtools SQP and KOMPLEX. They continue in simplifying the problem (e.g. they detect linear dependencies between constraints) and finally they calculate the desired P-Model. To be more specific, the worst case analysis is implemented as a second optimization process by using KOMPLEX, which tries to minimize the probability of a certain question by calling the MaxEnt optimization (with SQP) with varying values from the intervals.

6 Description of the PIT System

Fig 2 shows the central components of our system PIT (for Probability Induction Tool):

[5] Contrary to the MaxEnt problem, this is usually not a convex problem, and thus cannot be solved by SQP.

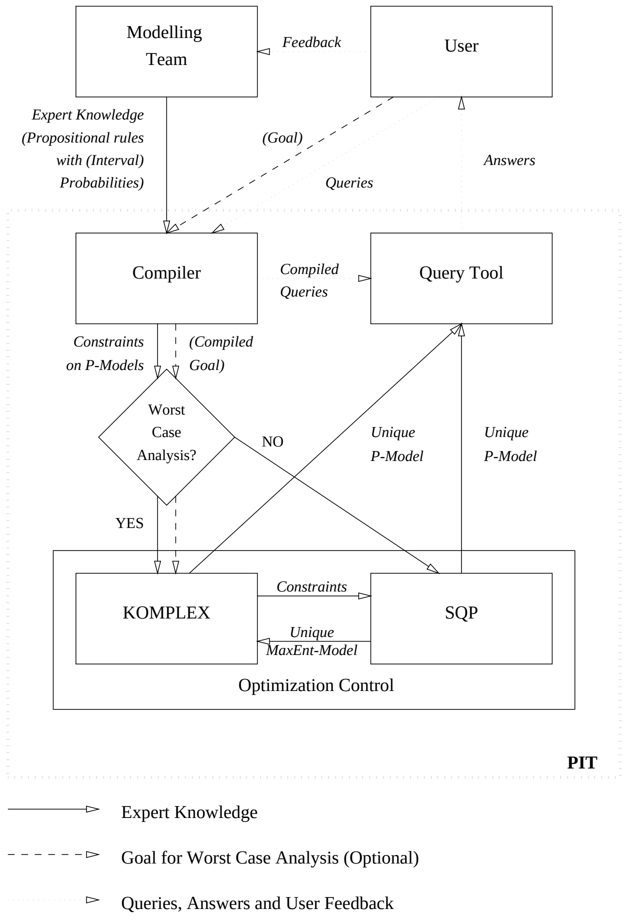

Figure 2. The PIT System and its Environment

- The task of the modeling team is to specify the knowledge of an expert as conditional probabilities ranging on intervals. Probabilities can be specified for arbitrary propositional expressions. The common-sense implication (*if-then*) is modeled as a conditional probability (and not as the material implication).
- The task of the compiler is to
  - make consistency checks and prove propositional problems. If the problem is a purely propositional one (all probabilities are 0 or 1), the compiler already calculates the answers to the queries.
  - translate the symbolic input into tables of linear constraints (for KOMPLEX and SQP) to be used as input for the optimization tool.
  - simplify the table of linear constraints by detecting indifferent constraints and indifferent events. Our experiments show that indifference is indeed a very powerful principle to reduce the complexity of the representation.
- The optimization control component governs the use of the subroutines KOMPLEX and SQP:
  - If no worst-case analysis is desired, SQP generates the unique P-Model with maximum entropy that is consistent with the constraints. This P-Model is then used to calculate answers for the user (compare the top entry in Fig. 1).
  - Otherwise, the constraints are handed to KOMPLEX, and KOMPLEX searches for the unique MaxEnt P-Model that minimizes the probability of a single goal and is still consistent with the given constraints. In this case KOMPLEX uses some (user-defined) constraints and generates new point constraints from them. It then uses SQP to calculate the MaxEnt-Model for these point constraints and the remaining, fixed constraints. The model finally resulting from this search is then used to calculate answers as above. This realizes the logic of the second entry *(Logic on P-Models with Worst-Case Analysis on given constraints)* of Fig. 1.
- The task of the Query Tool is to save the correlation between the events and variables (for the numerical optimization) and to use this information for calculating the answer to the queries after the optimization has found a model.
- Finally, the user specifies questions, receives the answers and gives feed-back to the modeling team. Of course, during the modeling phase the team also assumes the role of the user and can use the feedback directly.

7 An Example for Reasoning with PIT

We will explain reasoning with our system for an example from Lifschitz's collection of non-monotonic benchmarks (The example can be found as B4 in [14]). Consider the following colloquial specification:

(a1) Quakers are normally pacifists.
(a2) Republicans are normally hawks.
(a3) Pacifists are normally politically active.
(a4) Hawks are normally politically active.
(a5) Pacifists are not hawks.

We will show that the following common sense conclusions hold under our system:

(c1) Quakers who are not republicans are [normally] pacifists.
(c2) Republicans who are not Quakers are [normally] not pacifists.
(c3) Quakers, Republicans, Pacifists and Hawks are [normally] politically active.

We will first translate the assumptions into probabilistic constraints. Please note that the assumptions can be read as default statements. Statement (a1), for example, can be interpreted as *In the absence of other evidence, it is reasonable to assume that a Quaker is a pacifist.* It can, of course, be formalized in many ways. However, at least to our understanding, it requires that for an arbitrary Quaker the probability that he is a pacifist is greater than 0.5. We will therefore chose the *weakest possible formalization* for the term *normally* and only require that the corresponding probability ranges in the interval $(0.5, 1]$.[6]

[6] Regardless of the exact subinterval of $(0.5, 1]$ we chose, we can show that the conclusions hold within our approach.
One (mathematical) argument for using this interval (especially the lower boarder of 0.5) is that this is the only possibility to

So we express defaults by very weak conditional probability constraints using the largest plausible intervals. This representation is strong enough to solve many benchmarks of non-monotonic reasoning (e.g. benchmarks from [14]) even if we perform a worst-case analysis on the intervals.

We chose conditional probabilities to model the conditional statements in both assumptions and conclusions (Note that statement (a1) is equivalent to the conditional statement *If somebody is a Quaker, then he normally is a pacifist*). This is a well-known and widely used convention (compare e.g. [6]). To represent the five concepts *Quaker, Pacifist, Republican, Hawk* and *Politically active* we chose the set of variables $V = \{Qu, Pa, Re, Ha, Po\}$. With these decisions, we can restate the problem in our modeling language:

(a1) $P(Pa|Qu) \in (0.5; 1]$[7]
(a2) $P(Ha|Re) \in (0.5; 1]$
(a3) $P(Po|Pa) \in (0.5; 1]$
(a4) $P(Po|Ha) \in (0.5; 1]$
(a5) $P(\neg Ha|Pa) = 1$

To show the desired conclusions, we have to demonstrate that the following constraints hold in the resulting P-Model:

(c1) $P(Pa|Qu \wedge \neg Re) \in (0.5; 1]$
(c2) $P(\neg Pa|\neg Qu \wedge Re) \in (0.5; 1]$
(c3) $P(Po|Qu \vee Re \vee Pa \vee Ha) \in (0.5; 1]$[8]

As we have 5 propositional variables, the set of elementary events Ω contains $2^5 = 32$ possible worlds. The logical constraint (a5) eliminates 8 of these interpretations as impossible (i.e. their probability is 0). By applying the principle of indifference (compare section 5), we can restrict the number of equivalence classes to 11. (Roughly stated, the principle of indifference shows the symmetry between $(Qu \wedge \neg Re)$ and $(\neg Qu \wedge Re)$ as well as the symmetry between $(Pa \wedge \neg Ha)$ and $(\neg Pa \wedge Ha)$) .

The optimization process on the resulting 11 variables yields the following approximation of the MaxEnt-Model:

- $P(\omega) = 0$ for $\omega \in \{Qu \wedge Re \wedge Pa \wedge Ha \wedge Po, Qu \wedge Re \wedge Pa \wedge Ha \wedge \neg Po, Qu \wedge \neg Re \wedge Pa \wedge Ha \wedge Po, Qu \wedge \neg Re \wedge Pa \wedge Ha \wedge \neg Po, \neg Qu \wedge Re \wedge Pa \wedge Ha \wedge Po, \neg Qu \wedge Re \wedge Pa \wedge Ha \wedge \neg Po, \neg Qu \wedge \neg Re \wedge Pa \wedge Ha \wedge Po, \neg Qu \wedge \neg Re \wedge Pa \wedge Ha \wedge \neg Po\}$ (**Remark:** These are the possible worlds eliminated by the logical constraint (a5), i.e. all worlds that allow a Pacifist to also be a Hawk. They have already been detected at the compiler stage, of course with are more efficient method than enumerating them).
- $P(\omega) = 0.045086$ for $\omega \in \{\neg Qu \wedge \neg Re \wedge \neg Pa \wedge \neg Ha \wedge Po, \neg Qu \wedge \neg Re \wedge \neg Pa \wedge \neg Ha \wedge \neg Po\}$
- $P(\omega) = 0.045087$ for $\omega \in \{\neg Qu \wedge \neg Re \wedge Pa \wedge \neg Ha \wedge \neg Po, \neg Qu \wedge \neg Re \wedge \neg Pa \wedge Ha \wedge \neg Po\}$
- $P(\omega) = 0.045081$ for $\omega \in \{\neg Qu \wedge \neg Re \wedge Pa \wedge \neg Ha \wedge Po, \neg Qu \wedge \neg Re \wedge \neg Pa \wedge Ha \wedge Po\}$
- $P(\omega) = 0.027859$ for $\omega \in \{Qu \wedge \neg Re \wedge \neg Pa \wedge \neg Ha \wedge Po, Qu \wedge \neg Re \wedge \neg Pa \wedge \neg Ha \wedge \neg Po, \neg Qu \wedge Re \wedge \neg Pa \wedge \neg Ha \wedge Po, \neg Qu \wedge Re \wedge \neg Pa \wedge \neg Ha \wedge \neg Po\}$
- $P(\omega) = 0.072951$ for $\omega \in \{Qu \wedge \neg Re \wedge Pa \wedge \neg Ha \wedge \neg Po, \neg Qu \wedge Re \wedge \neg Pa \wedge Ha \wedge \neg Po\}$
- $P(\omega) = 0.072943$ for $\omega \in \{Qu \wedge \neg Re \wedge Pa \wedge \neg Ha \wedge Po, \neg Qu \wedge Re \wedge \neg Pa \wedge Ha \wedge Po\}$
- $P(\omega) = 0.027866$ for $\omega \in \{Qu \wedge \neg Re \wedge \neg Pa \wedge Ha \wedge \neg Po, \neg Qu \wedge Re \wedge Pa \wedge \neg Ha \wedge \neg Po\}$
- $P(\omega) = 0.027876$ for $\omega \in \{Qu \wedge \neg Re \wedge \neg Pa \wedge Ha \wedge Po, \neg Qu \wedge Re \wedge Pa \wedge \neg Ha \wedge Po\}$
- $P(\omega) = 0.017216$ for $\omega \in \{Qu \wedge Re \wedge \neg Pa \wedge \neg Ha \wedge Po, Qu \wedge Re \wedge \neg Pa \wedge \neg Ha \wedge \neg Po\}$
- $P(\omega) = 0.045085$ for $\omega \in \{Qu \wedge Re \wedge Pa \wedge \neg Ha \wedge \neg Po, Qu \wedge Re \wedge \neg Pa \wedge Ha \wedge \neg Po\}$
- $P(\omega) = 0.045090$ for $\omega \in \{Qu \wedge Re \wedge Pa \wedge \neg Ha \wedge Po, Qu \wedge Re \wedge \neg Pa \wedge Ha \wedge Po\}$

From this MaxEnt-Model we gain the following data about the assumptions.

(a1) $P(Pa|Qu) = 0.500001 (= 0.5 + \epsilon)$. The probability of assumption (a1) in the MaxEnt-Model is the lowest one compatible with the original constraint.
(a2) $P(Ha|Re) = 0.500001 (= 0.5 + \epsilon)$.
(a3) $P(Po|Pa) = 0.500001 (= 0.5 + \epsilon)$.
(a4) $P(Po|Ha) = 0.500001 (= 0.5 + \epsilon)$.
(a5) $P(\neg Ha|Pa) = 1$. Remember that (a5) is a logical constraint and there is no freedom to choose another probability here.

As one can see, the MaxEnt-Model tries to be as uncommitted about the assumptions as the specification allows (i.e. it chooses a probability for the assumptions that provides the minimum amount of information). In this case, as is to be expected, the model is found at the (lower) border of the intervals. Nevertheless, all the desired conclusions are supported in it:

(c1) $P(Pa|Qu \wedge \neg Re) = 0.566900$.
(c2) $P(\neg Pa|\neg Qu \wedge Re) = 0.783404$.
(c3) $P(Po|Qu \vee Re \vee Pa \vee Ha) = 0.500001$.

The probability for all three conclusions falls into the interval $(0.5, 1]$, ergo all three conclusions hold under our modeling assumptions. As we already noted, similar results are obtained for other choices of the intervals and even the worst case analysis (choose the values in the intervals first and apply MaxEnt afterwards) does not change the result, because the MaxEnt-Model already represents the worst case for this example, that is, it chooses exactly the probabilities of the assumptions that minimize the probabilities of the three conclusions.

8 Conclusion and future work

We have shortly described our concept and implementation for reasoning with incomplete and uncertain (here: probabilistic) knowledge. Based on propositional logic, the specification

express assumptions and conclusions within the same interval. In most examples of this type of inductive reasoning, the conclusions show less certainty than the premises, so increasing the lower boarder of the interval of the assumptions does not increase the lower boarder of the conclusion by the same size.

[7] In the implemented system a half-open interval $(a, b]$ is realized as the closed interval $[a + \epsilon, b]$, where ϵ is of the same order of magnitude as the desired overall accuracy.

[8] Note that the sentence (c3) is equivalent to the conditional *If somebody is a Quaker or a Republican or a Pacifist or a Hawk, then he (normally) is politically active.* Thus the formalization uses the disjunction $\vee$ to represent the colloquial *and* of the original sentence.

can contain logical constraints, constraints with fixed probabilities or constraints with probabilities ranging in intervals. This very general features can also be used for qualitative non-monotonic reasoning. Most non-monotonic benchmarks (e.g. those from [14] using unary predicates) can be modeled within the very weak formalization we described above. [9]

For better judgments we still look for more complex examples, that are still well understood in their desired conclusions (contrary to statistical reasoning on complex domains). In our actual work we also trying to specify an application for a medical purpose within the ESPRIT Basic Research Project 9119 MIX , based on 24 propositional variables and about 150 probabilistic constraints.

Our future work will also include improvements of the subtools.

- We will try to make use of the principle of independence in the compiler to reduce the complexity of the optimization problem.
- We will try to replace the SQP optimization tool with a more specialized version that makes use of the special properties of our problem and thus is better equipped to deal with our expected standard applications.
- Finally, we will replace the KOMPLEX algorithm for the worst-case analysis with an improved algorithm that is under development in an ongoing Ph.D. thesis.

REFERENCES

[1] F. Bacchus, A.J. Grove, J.Y. Halpern, and D. Koller, 'From Statistics to Beliefs', in *Proc. of the 10^{th} National Conference on Artificial Intelligence*. AAAI, AAAI Press/MIT Press, (1992).

[2] C.R. Blyth, 'On Simpson's Paradox and the Sure-Thing Principle', *Journal of the American Statistical Association*, **67**(338), 363–381, (June 1972).

[3] M.J. Box, 'A New Method of Constrained Optimization and a Comparison with Other Methods', *The Computer Journal*, **8**, 42–52, (1965).

[4] R.Y. Bryant, 'Symbolic Boolean Manipulation with Ordered Binary Decision Diagrams', Technical Report CMU-CS-92-160, School of Computer Science, Carnegie Mellon University, (1992). Available at http://www.cs.cmu.edu/Web/Reports/1992.html.

[5] P. Cheeseman, 'A Method of Generalize Bayesian Probability Values for Expert Systems', in *Proc. of the 8^{th} International Joint Conference on Artificial Intelligence (IJCAI-83), Karlsruhe*, volume 1, pp. 198–202, San Mateo, CA, (1983). Morgan Kaufman.

[6] E.W. Adams, 'Probability and the Logic of Conditionals', in *Aspects of Inductive Logic*, eds., J. Hintikka and P.Suppes, North Holland, Amsterdam, (1966).

[7] S.A. Goldman and R.L. Rivest, 'A Non-Iterative Maximum Entropie Algorithm', in *Uncertainty in Artificial Intelligence*, eds., J.F. Lemmer and L.N. Kanal, volume 2, 133–148, Elsevier Science, (1988).

[8] M. Goldszmidt, P. Morris, and J. Pearl, 'A Maximum Entropy Approach to Non-Monotonic Reasoning', in *Proc. of the 8^{th} National Conference on Artificial Intelligence, Boston*, pp. 646–652. AAAI, AAAI Press/MIT Press, (1990).

[9] A.J. Grove, J.Y. Halpern, and D. Koller, 'Random Worlds and Maximum Entropy', in *Proc. of the 7^{th} Annual IEEE Symposium on Logic in Computer Science (LICS)*. IEEE, (1992).

[10] E.T. Jaynes, 'Where do we stand on Maximum Entropy?', in *Papers on Probability, Statistics and Statistical Physics*, ed., R.D. Rosenkrantz, 210–314, Kluwer Academic Publishers, (1978).

[11] E.T. Jaynes, 'On the Rationale of Maximum Entropy Methods', *Proc. of the IEEE*, **70**(9), 939–952, (1982).

[12] E.T. Jaynes, 'Probability Theory: The Logic of Science'. Continually updated, fragmentary edition of approximately 700 pages available from `ftp://bayes.wustl.edu//pub/jaynes`, 1995.

[13] J.N. Kapur and H.K. Kesavan, *Entropy Optimization Prinziples with Applications*, Academic Press, 1992.

[14] V. Lifschitz, 'Benchmark Problems for formal Non-Monotonic Reasoning', in *Non-Monotonic Reasoning: 2nd International Workshop*, ed., Reinfrank et al, volume 346 of *LNAI*, pp. 202–219. Springer, (1989).

[15] E. Neufeld and J.D. Horton, 'Conditioning on Disjunctive Knowledge: Simpson's Paradox in Default Logic', in *Uncertainty in Artificial Intelligence*, eds., M. Henrion, R.D. Shachter, L.N. Kanal, and J.F. Lemmer, volume 5, 117–125, Elsevier Science, (1990).

[16] J.B. Paris and A. Vencovska, 'A Note on the Inevitability of Maximum Entropy', *International Journal of Approximate Reasoning*, **3**, 183–223, (1990).

[17] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA, 1988.

[18] D. Poole, 'What the Lottery Paradox Tells us about Default Reasoning', in *Proc. of the 1^{st} International Conference on Principles of Knowledge Representation and Reasoning*, eds., R.J. Brachman, H.J. Levesque, and R. Reiter, pp. 333–340, San Mateo, CA, (1989). Morgan Kaufmann.

[19] M. Schramm and M. Greiner, 'Foundations: Indifference, Independence & Maxent', in *Maximum Entropy and Bayesian Methods in Science and Engeneering (Proc. of the MaxEnt'94)*, ed., J. Skilling. Kluwer Academic Publishers, (1995).

[20] C.E. Shannon and W. Weaver, *A Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

[21] J.E. Shore and R.W. Johnson, 'Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy', *IEEE Transactions on Information Theory*, **IT-26**(1), 26–37, (1980).

[22] J. Skilling, 'The Axioms of Maximum Entropy', in *Maximum Entropy and Bayesian Methods in Science and Engineereing*, eds., G.J. Erickson and C.R. Smith, volume 1 – Foundations, Kluwer Academic Publishers, (1988).

[23] S. Wagenpfeil, *Implementierung eines SQP-Verfahrens mit dem Algorithmus von Ritter und Best*, Diplomarbeit, Institut für Angewandte Mathematik und Statistik, TU München, 1991. (German Language).

[9] An exception of this is the well known proposal of the *And-rule*, but it is known that this rule causes trouble (consider e.g. the *Lottery Paradox* as described in [18]). Furthermore, as MaxEnt is based on combinatorial calculations, the result of MaxEnt has some meaning independent from proposed rules of non-monotonic reasoning. If it contradicts a proposed rule, it also calls into question this rule, or to be more concrete, the rule will not be useful for reasoning on proportions.