Solving the inverse representation problem

Gabriele Kern-Isberner[1]

Abstract. In this paper, we present an approach to extract most relevant information from a (semi-)quantitative knowledge base, e.g., from a probability distribution. Relevance here is meant with respect to some appropriate inductive inference process, like maximum entropy inference (*ME-inference*) in probabilistics. So in particular, the method developed in this paper is apt to solve the *inverse maxent problem*, computing from a distribution in a non-heuristic way a set of conditionals that ME-represents that distribution. Since we only make use of one special characteristic of ME-inference, this method may as well be applied to other, similar inference processes.

1 Introduction

In many practical examples and applications, the available knowledge is neither certain nor complete. So classical deduction often seems to be inappropriate to yield useful inferences, and also the techniques of nonmonotonic reasoning may prove to be too weak. A way out of this dilemma is offered by selecting a *model* most adequately representing the available knowledge, and using this model for inferences. This policy is pursued, for instance, by reasoning at maximum entropy in a probabilistic setting [7], and by system-Z for ranking functions [2]. The way how to solve this *representation problem* is usually guided by fundamental principles of reasoning, or by good heuristics. The appropriate handling of conditional information, constituting crucial pieces of knowledge, is a peculiarly challenging task.

On the other hand, incompleteness of knowledge can also be a deliberate consequence of focusing on most relevant information and relationships. So, conversely, given some model, one may ask which propositional and conditional statements constitute central knowledge. In particular, which part of the knowledge inherent to that model is apt to generate it with respect to some given inductive representation technique?

In this paper, we will present an approach to solve this *inverse representation problem* for quantitative and semi-quantitative representation methods satisfying a conditional indifference property. The method based on maximum entropy (abbreviated by *ME*) is known to fulfill this property (see [3]), so we will exemplify our ideas in a probabilistic setting. Given some probability distribution P, we will show how to calculate a finite set $\mathcal{R}^{prob}$ of probabilistic conditionals such that P is the ME-model of $\mathcal{R}^{prob}$. Therefore, in particular, the techniques developed in this paper constitute an approach to solve the *inverse maxent problem*. They are, however, useful in a more general environment. The basic idea is to exploit numerical relationships as manifestations of interactions of underlying conditional knowledge. So the techniques to be presented may also be applied e.g. to ranking functions representing epistemic states. Although starting from numbers, our method is essentially algebraic, elaborating the *conditional structure* of worlds.

Our approach differs from usual knowledge discovery and data mining methods in that it takes explicitly inductive representation, or inference, respectively, into consideration. It is not based on observing conditional independencies, but aims at learning conditional dependencies in a non-heuristic way. As a further novelty, our method computes not single, isolated rules, but yields as a result a set of rules in taking into account highly complex interaction of rules. When applied to a probability distribution gained from statistical data, the set of conditionals $\mathcal{R}^{prob}$ obtained by the inverse maxent procedure may also be supposed to represent most informative knowledge to be discovered from the data.

This paper is organized as follows: In the next section, we summarize some basic notations and techniques for conditionals, and we sketch how to obtain an ME-representation. Section 3 introduces the notion of conditional structures, which we will use in Section 4 to define *conditional indifference*. Section 5 sketches the procedure of how to compute a set of conditionals with respect to which a distribution is indifferent, and the theoretical background is elucidated. Then in Section 6, this procedure is applied to solve the maxent problem for an example. We conclude this paper with an outlook in Section 7. All proofs are omitted, but may be found in [4].

2 Basic notations and techniques

We consider a propositional language $\mathcal{L}$ with finitely many atomic propositions. Let Ω denote the corresponding set of possible worlds, that is, Ω is a complete set of propositional interpretations of $\mathcal{L}$. The conjunction operator, $\wedge$, will usually be omitted, so AB will mean $A \wedge B$, and negation is indicated by barring, i.e. $\overline{A} = \neg A$.

Conditionals are written in the form $(B|A)$, with antecedents, A, and consequents, B, both propositional formulas in $\mathcal{L}$. Let $(\mathcal{L} \mid \mathcal{L})$ denote the set of all conditionals over $\mathcal{L}$. Each conditional $(B|A)$ can be represented as a *generalized indicator function* on worlds, setting

$$(B|A)(\omega) = \begin{cases} 1 & : \quad \omega \models AB \\ 0 & : \quad \omega \models A\overline{B} \\ u & : \quad \omega \models \overline{A} \end{cases} \qquad (1)$$

where u stands for *undefined* (cf. [1]). Two conditionals are *equivalent* iff they yield the same indicator function. *Single-elementary conditionals* are conditionals whose antecedents are conjunctions of literals, and whose consequents consist of one single literal.

We introduce the following relation $\sqsubseteq$ between conditionals:

$$(D|C) \sqsubseteq (B|A) \quad \text{iff } CD \models AB \text{ and } C\overline{D} \models A\overline{B}$$

If $(D|C) \sqsubseteq (B|A)$, then $(D|C)$ is called a *subconditional* of $(B|A)$. For any two conditionals $(B|A), (D|C) \in (\mathcal{L} \mid \mathcal{L})$ with $ABC\overline{D} \equiv$

[1] FernUniversität Hagen, Dept. of Computer Science, LG Prakt. Informatik VIII, P.O. Box 940, D-58084 Hagen, Germany, gabriele.kern-isberner@fernuni-hagen.de

$A\overline{B}CD \equiv \bot$, the supremum $(B|A) \sqcup (D|C)$ in $(\mathcal{L} \mid \mathcal{L})$ with respect to $\sqsubseteq$ exists and is given by $(B|A) \sqcup (D|C) \equiv (AB \vee CD | A \vee C)$ (cf. [4]). In particular, for two conditionals $(B|A), (B|C)$ with the same consequent, we have $(B|A) \sqcup (B|C) \equiv (B|A \vee C)$. The following lemma provides an easy characterization for the relation $\sqsubseteq$ to hold between single-elementary conditionals:

Lemma 1 *Let $(b|A)$ and $(d|C)$ be two single-elementary conditionals. Then $(d|C) \sqsubseteq (b|A)$ iff $C \models A$ and $b = d$.*

This lemma may be slightly generalized to hold for conditionals $(b|A)$ and $(d|C)$ where A and C are disjunctions of conjunctions of literals not containing b and d, respectively.

Let P be a probability distribution over the alphabet of $\mathcal{L}$. Within a probabilistic framework, conditionals can be quantified and interpreted probabilistically via conditional probabilities: $P \models (B|A)[x]$ iff $P(AB) = xP(A)$ for some $x \in [0,1]$. Suppose $\mathcal{R}^{prob} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$ is a consistent set of probabilistic conditionals. Then the *ME-representation of $\mathcal{R}^{prob}$*, *ME($\mathcal{R}^{prob}$)*, is the unique distribution Q^* that maximizes the entropy $H(Q) = -\sum_\omega Q(\omega) \log Q(\omega)$ subject to $Q \models \mathcal{R}^{prob}$ (cf. [7]). If $\mathcal{R}^{prob} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$, then $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ denotes the set of structural (i.e. unquantified) conditionals, and vice versa.

In this paper, we will only consider positive probability distributions P. Correspondingly, we will assume that the ME-representation of any set $\mathcal{R}^{prob}$ dealt with in the sequel is positive. In particular, all probabilities x_i of conditionals in $\mathcal{R}^{prob}$ have to be different from 0 and 1. This is but a technical prerequisite, to focus on the most interesting cases, and for the sake of brevity. The general case may be dealt with in a similar manner (cf. [4]).

3 Conditional structures

When we consider (finite) sets of conditionals $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\}$, we have to modify representation (1) appropriately to identify the effect of each conditional in $\mathcal{R}$ on worlds in Ω. This leads to introducing the functions $\sigma_i = \sigma_{(B_i|A_i)}$ below (see (2)) which generalize (1) by replacing the numbers 0 and 1 by abstract symbols. Moreover, we will make use of a group structure to represent the joint impact of conditionals on worlds.

To each conditional $(B_i|A_i)$ in $\mathcal{R}$ we associate two symbols $\mathbf{a}_i^+, \mathbf{a}_i^-$. Let

$$\mathcal{F}_\mathcal{R} = \langle \mathbf{a}_1^+, \mathbf{a}_1^-, \ldots, \mathbf{a}_n^+, \mathbf{a}_n^- \rangle$$

be the free abelian group with generators $\mathbf{a}_1^+, \mathbf{a}_1^-, \ldots, \mathbf{a}_n^+, \mathbf{a}_n^-$, i.e. $\mathcal{F}_\mathcal{R}$ consists of all elements of the form $(\mathbf{a}_1^+)^{r_1}(\mathbf{a}_1^-)^{s_1} \ldots (\mathbf{a}_n^+)^{r_n}(\mathbf{a}_n^-)^{s_n}$ with integers $r_i, s_i \in \mathbb{Z}$ (the ring of integers). Each element of $\mathcal{F}_\mathcal{R}$ can be identified by its exponents, so that $\mathcal{F}_\mathcal{R}$ is isomorphic to $\mathbb{Z}^{2n}$ (cf. [6]). The commutativity of $\mathcal{F}_\mathcal{R}$ corresponds to the fact that the conditionals in $\mathcal{R}$ shall be effective simultaneously, without assuming any order of application.

For each $i, 1 \leqslant i \leqslant n$, we define a function $\sigma_i : \Omega \to \mathcal{F}_\mathcal{R}$ by setting

$$\sigma_i(\omega) = \begin{cases} \mathbf{a}_i^+ & \text{if} \quad (B_i|A_i)(\omega) = 1 \\ \mathbf{a}_i^- & \text{if} \quad (B_i|A_i)(\omega) = 0 \\ 1 & \text{if} \quad (B_i|A_i)(\omega) = u \end{cases} \tag{2}$$

$\sigma_i(\omega)$ represents the manner in which the conditional $(B_i|A_i)$ applies to the possible world ω. The neutral element 1 of $\mathcal{F}_\mathcal{R}$ corresponds to the non-applicability of $(B_i|A_i)$ in case that the antecedent A_i is not satisfied. The function

$$\sigma_\mathcal{R}(\omega) = \prod_{1 \leqslant i \leqslant n} \sigma_i(\omega) = \prod_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i B_i}} \mathbf{a}_i^+ \prod_{\substack{1 \leqslant i \leqslant n \\ \omega \models A_i \overline{B}_i}} \mathbf{a}_i^- \tag{3}$$

describes the all-over effect of $\mathcal{R}$ on ω. $\sigma_\mathcal{R}(\omega)$ is called *(a representation of) the conditional structure of ω with respect to $\mathcal{R}$*.

Example 2 Consider the set $\mathcal{R} = \{(d|a), (d|b), (d|c)\}$ of conditionals using the atoms a, b, c, d. Let $\mathbf{a}_1^\pm, \mathbf{a}_2^\pm, \mathbf{a}_3^\pm$ be the group generators associated with $(d|a), (d|b), (d|c)$, respectively. The conditional structure of $\omega = ab\overline{c}d$, e.g., is $\sigma_\mathcal{R}(ab\overline{c}d) = \mathbf{a}_1^+\mathbf{a}_2^+$, since $ab\overline{c}d$ confirms the first two conditionals, and the third conditional is not applicable to it. Moreover, by conditional structures, it is possible to compare worlds, or sets of worlds, as to their behavior with respect to the conditionals in $\mathcal{R}$. For instance, we have

$$\begin{aligned} \sigma_\mathcal{R}(ab\overline{c}d)\sigma_\mathcal{R}(a\overline{b}cd)\sigma_\mathcal{R}(\overline{a}bcd) &= (\mathbf{a}_1^+\mathbf{a}_2^+)(\mathbf{a}_1^+\mathbf{a}_3^+)(\mathbf{a}_2^+\mathbf{a}_3^+) \\ = (\mathbf{a}_1^+)^2(\mathbf{a}_2^+)^2(\mathbf{a}_3^+)^2 &= (\mathbf{a}_1^+\mathbf{a}_2^+\mathbf{a}_3^+)^2 = \sigma_\mathcal{R}(abcd)^2. \end{aligned}$$

That is to say, that the elements of the set $\{ab\overline{c}d, a\overline{b}cd, \overline{a}bcd\}$ show collectively the same conditional behavior as two copies of $abcd$. ■

To compare worlds conveniently with respect to their conditional structures, we impose a multiplication on the set of worlds Ω by considering the worlds ω as formal symbols. That means, we introduce the free abelian group $\widehat{\Omega} := \langle \omega \mid \omega \in \Omega \rangle$, generated by all $\omega \in \Omega$, and consisting of all words $\widehat{\omega} = {\omega_1}^{r_1} \ldots {\omega_m}^{r_m}$ with $\omega_1, \ldots, \omega_m \in \Omega$ and integers $r_1, \ldots r_m$[2]. Now $\sigma_\mathcal{R}$ may be extended to $\widehat{\Omega}$ in a straightforward manner by setting $\sigma_\mathcal{R}(\widehat{\omega}) = \sigma_\mathcal{R}(\omega_1)^{r_1} \ldots \sigma_\mathcal{R}(\omega_m)^{r_m}$, yielding a *homomorphism of groups* $\sigma_\mathcal{R} : \widehat{\Omega} \to \mathcal{F}_\mathcal{R}$.

The generators $\mathbf{a}_i^+$ are mere symbols, representing the effects of the corresponding conditional on worlds. As can easily be seen, however, the *kernel* of such a representation homomorphism

$$\textit{ker}\ \sigma_\mathcal{R} := \{\widehat{\omega} \in \widehat{\Omega} \mid \sigma_\mathcal{R}(\widehat{\omega}) = 1\}$$

does not depend on the particular symbols chosen. Therefore, it is an invariant of $\mathcal{R}$. *ker* $\sigma_\mathcal{R}$ contains exactly all group elements $\widehat{\omega} \in \widehat{\Omega}$ with a balanced conditional structure, that means, where all effects of conditionals in $\mathcal{R}$ on worlds occurring in $\widehat{\omega}$ are completely cancelled.

Having the same conditional structure defines an equivalence relation $\equiv_\mathcal{R}$ on $\widehat{\Omega}$: $\widehat{\omega}_1 \equiv_\mathcal{R} \widehat{\omega}_2$ iff $\sigma_\mathcal{R}(\widehat{\omega}_1) = \sigma_\mathcal{R}(\widehat{\omega}_2)$, i.e. iff $\widehat{\omega}_1\widehat{\omega}_2^{-1} \in$ *ker* $\sigma_\mathcal{R}$. Thus the kernel of $\sigma_\mathcal{R}$ plays an important part in identifying the conditional structure of elements $\widehat{\omega} \in \widehat{\Omega}$, in particular of worlds ω, with respect to $\mathcal{R}$. No nontrivial relations hold between different group generators $\mathbf{a}_1^+, \mathbf{a}_1^-, \ldots, \mathbf{a}_n^+, \mathbf{a}_n^-$ of $\mathcal{F}_\mathcal{R}$, so we have $\sigma_\mathcal{R}(\widehat{\omega}) = 1$ iff $\sigma_i(\widehat{\omega}) = 1$ for all i, $1 \leqslant i \leqslant n$, and this means *ker* $\sigma_\mathcal{R} = \bigcap_{i=1}^n$ *ker* σ_i. In this way, each conditional in $\mathcal{R}$ contributes to *ker* $\sigma_\mathcal{R}$. Besides the explicit representation of knowledge by $\mathcal{R}$, often implicit normalizing constraints (such as $P(\top|\top) = 1$ for probability functions or $\kappa(\top|\top) = 0$ for ordinal conditional functions) have to be taken into account. It is easy to check that *ker* $\sigma_{(\top|\top)} = \widehat{\Omega}_0 := \{\widehat{\omega} = {\omega_1}^{r_1} \cdot \ldots \cdot {\omega_m}^{r_m} \in \widehat{\Omega} \mid \sum_{j=1}^m r_j = 0\}$. Two elements $\widehat{\omega}_1 = \omega_1^{r_1} \ldots \omega_m^{r_m}$, $\widehat{\omega}_2 = \nu_1^{s_1} \ldots \nu_p^{s_p} \in \widehat{\Omega}$ are equivalent modulo $\widehat{\Omega}_0$, $\widehat{\omega}_1 \equiv_\top \widehat{\omega}_2$, iff $\widehat{\omega}_1\widehat{\Omega}_0 = \widehat{\omega}_2\widehat{\Omega}_0$, i.e. iff $\sum_{1 \leqslant j \leqslant m} r_j = \sum_{1 \leqslant k \leqslant p} s_k$. This means that $\widehat{\omega}_1$ and $\widehat{\omega}_2$ are equivalent modulo $\widehat{\Omega}_0$ iff they both are a (cancelled) product of the same

[2] We will often use fractional representations for the elements of $\widehat{\Omega}$, that is, for instance, we will write $\dfrac{\omega_1}{\omega_2}$ instead of $\omega_1\omega_2^{-1}$.

number of generators, each generator being counted with its corresponding exponent. Set

$$ker_0\ \sigma_{\mathcal{R}} := ker\ \sigma_{\mathcal{R}} \cap \widehat{\Omega}_0 = ker\ \sigma_{\mathcal{R}\cup\{(\top|\top)\}}.$$

4 Indifferent representations of conditional knowledge

In this section, we will study conditional interactions in positive probability functions P. Each such function may be extended to a homomorphism $P : \widehat{\Omega} \to (\mathbb{R}^+, \cdot)$ from $\widehat{\Omega}$ into the positive real numbers by setting $P({\omega_1}^{r_1} \ldots {\omega_m}^{r_m}) = P(\omega_1)^{r_1} \cdot \ldots \cdot P(\omega_m)^{r_m}$. This allows us to analyze numerical relationships in order to elaborate the conditionals whose structures P follows, that means, to determine sets of conditionals $\mathcal{R}$ with respect to which P is *indifferent*:

Definition 3 Suppose P is a positive probability distribution, and let $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ be a set of conditionals.

P is *indifferent with respect to* $\mathcal{R}$ iff $P(\widehat{\omega}_1) = P(\widehat{\omega}_2)$ whenever $\sigma_{\mathcal{R}}(\widehat{\omega}_1) = \sigma_{\mathcal{R}}(\widehat{\omega}_2)$, for all $\widehat{\omega}_1, \widehat{\omega}_2 \in \widehat{\Omega}$ with $\widehat{\omega}_1 \equiv_\top \widehat{\omega}_2$.

If P is indifferent with respect to $\mathcal{R}$, then it does not distinguish between elements $\widehat{\omega}_1 \equiv_\top \widehat{\omega}_2$ with the same conditional structure with respect to $\mathcal{R}$. Conversely, any deviation $P(\widehat{\omega}) \neq 1$ can be explained by the conditionals in $\mathcal{R}$ acting on $\widehat{\omega}$ in a non-balanced way. Note that the notion of indifference only aims at observing conditional structures, without making use of any probabilities associated with the conditionals.

The following proposition rephrases conditional indifference by establishing a relationship between the kernels of $\sigma_{\mathcal{R}}$ and P:

Proposition 4 *Suppose P is a positive probability distribution, and let $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ be a set of conditionals. P is indifferent with respect to $\mathcal{R}$ iff $ker_0\ \sigma_{\mathcal{R}} \subseteq ker_0\ P$.*

If $ker_0\ \sigma_{\mathcal{R}} = ker_0\ P$, then $P(\widehat{\omega}_1) = P(\widehat{\omega}_2)$ iff $\sigma_{\mathcal{R}}(\widehat{\omega}_1) = \sigma_{\mathcal{R}}(\widehat{\omega}_2)$ and $\widehat{\omega}_1 \equiv_\top \widehat{\omega}_2$. In this case, P completely follows the conditional structures imposed by $\mathcal{R}$ – it observes $\mathcal{R}$ *faithfully*.

The next theorem characterizes indifferent probability functions:

Theorem 5 *A (positive) probability function P is indifferent with respect to a set $\mathcal{R} = \{(B_1|A_1), \ldots, (B_n|A_n)\}$ iff there are positive real numbers $\alpha_0, \alpha_1^+, \alpha_1^-, \ldots, \alpha_n^+, \alpha_n^- \in \mathbb{R}^+$, such that*

$$P(\omega) = \alpha_0 \prod_{\substack{1\leqslant i\leqslant n\\ \omega\models A_iB_i}} \alpha_i^+ \prod_{\substack{1\leqslant i\leqslant n\\ \omega\models A_i\overline{B_i}}} \alpha_i^-,\ \omega \in \Omega. \qquad (4)$$

Conditional indifference is the crucial ingredient to realize the *principle of conditional preservation* in belief revision theory which may govern the revision of an epistemic state by a set of conditionals (cf. [5]). This principle may be reformulated for the inductive representation of conditional probabilistic knowledge, as follows:

Principle of conditional preservation for representations:
A probability distribution P representing a set $\mathcal{R}^{prob}$ of conditionals satisfies the principle of conditional preservation (with respect to $\mathcal{R}^{prob}$) iff P is indifferent with respect to R.

Representations observing this principle handle even complex interdependencies between the conditionals involved in a very accurate way. Therefore, they are especially well-designed for conditional inferences (see [5]).

In particular, each ME-distribution $ME(\mathcal{R}^{prob})$ is indifferent with respect to its generating set of conditionals, which may be seen directly from Theorem 5 (cf. [3]). As an example in a qualitative framework, each system-Z^* representation satisfies the principle of conditional preservation, too (cf. [5]).

5 Discovering conditional structures

In this section, as the main result of this paper, we will present an approach to computing sets $\mathcal{R}$, or $\mathcal{R}^{prob}$, respectively, of conditionals that may be apt to generate some given (positive) probability function P via an appropriate inductive inference method. Appropriate here means obeying the principle of conditional preservation, as e.g. ME-inference (see Section 4 above). Our method addresses quite new aspects in knowledge discovery:

- It is based on numbers but not on probabilities close to 1; actually it aims at discovering structures of conditional knowledge.
- The method is able to disentangle highly complex interactions between conditionals.
- We are going to discover not single, isolated rules but a set of rules, thus taking into regard the collective effects of several conditionals.

The method to be presented is guided by the following idea: If P is the result of an inductive inference procedure using a set $\mathcal{R}^{prob}$ of conditionals as knowledge base and observing the principle of conditional preservation (e.g. $P = ME(\mathcal{R}^{prob})$), then P is necessarily indifferent with respect to $\mathcal{R}$, i.e. $ker_0\ \sigma_{\mathcal{R}} \subseteq ker_0\ P$ by Proposition 4. Ideally, we would have P to represent $\mathcal{R}$ faithfully, that is,

$$P \models \mathcal{R} \quad \text{and} \quad ker_0\ P = ker_0\ \sigma_{\mathcal{R}}. \qquad (5)$$

Assuming faithfulness means presupposing that no equation $P(\widehat{\omega}) = 1$ is fulfilled accidentally, but that any of these equations is induced by $\mathcal{R}$. Thus the structures of the conditionals in $\mathcal{R}$ become manifest in the elements of $ker_0\ P$, that is, in elements $\widehat{\omega} \in \widehat{\Omega}$ with $P(\widehat{\omega}) = 1$. As a further prerequisite, we will assume that this knowledge inherent to P is representable by a set of single-elementary conditionals. This restriction should not be considered as a heavy drawback, bearing in mind the expressibility of single-elementary conditionals.

So assume $\mathcal{R}^{prob} = \{(b_1|A_1)[x_1], \ldots, (b_n|A_n)[x_n]\}$ is an existing, but hidden set of single-elementary conditionals, such that (5) holds. Let us further suppose that $ker_0\ P$ is known from exploiting numerical relationships. Without loss of generality, only to simplify notation, we assume all consequents b_i to be positive literals. Since conditional indifference is a structural notion, we omit the quantifications x_i of the conditionals in what follows. Let $\sigma_{\mathcal{R}} : \widehat{\Omega} \to \mathcal{F}_{\mathcal{R}} = \langle \mathbf{a}_1^+, \mathbf{a}_1^-, \ldots, \mathbf{a}_n^+, \mathbf{a}_n^- \rangle$ denote a conditional structure homomorphism with respect to $\mathcal{R}$.

Our method is a *bottom-up approach* generalizing conditionals in accordance with the conditional structure revealed by $ker_0\ P$. We start with considering *basic single-elementary conditionals*, which are single-elementary conditionals with antecedents of maximal length. For each atom $v \in \mathcal{L}$, choose an arbitrary, but fixed numbering of the remaining atoms $w \neq v$, $(w_0, w_1, \ldots, w_{\#(atoms)-1})$. Then *basic single-elementary conditionals* are conditionals of the form

$$\psi_{v,l} = (v \mid \bigwedge_j w_j^{\epsilon_j}) \qquad (6)$$

with $\epsilon_j \in \{0, 1\}$, $w_j^1 := w_j$, $w_j^0 := \overline{w_j}$, $0 \leqslant j \leqslant \#(atoms) - 1$ and $l = \sum_j \epsilon_j 2^j$. We will abbreviate the antecedent of $\psi_{v,l}$ by $C_{v,l}$. Let

$$\mathcal{B} = \{\psi_{v,l} \mid v \text{ atom in } \mathcal{L}, 0 \leqslant l \leqslant 2^{\#(atoms)-1} - 1\}$$

denote the set of all basic single-elementary conditionals in $(\mathcal{L} \mid \mathcal{L})$, and let $\mathcal{F}_{\mathcal{B}} = \langle \mathbf{b}_{v,l}^+, \mathbf{b}_{v,l}^- \mid v \text{ atom in } \mathcal{L}, 0 \leqslant l \leqslant 2^{\#(atoms)-1} -$

$1\rangle$ be the free abelian group corresponding to $\mathcal{B}$ with conditional structure homomorphism $\sigma_{\mathcal{B}} : \widehat{\Omega} \to \mathcal{F}_{\mathcal{B}}$,

$$\sigma_{\mathcal{B}} = \prod_{v,l} \sigma_{v,l}, \ \ \sigma_{v,l}(\omega) = \begin{cases} \mathbf{b}^{+}_{v,l}, & \text{if } \omega = C_{v,l} v \\ \mathbf{b}^{-}_{v,l}, & \text{if } \omega = C_{v,l} \overline{v} \\ 1, & \text{else} \end{cases}$$

Lemma 6 *$\sigma_{\mathcal{B}}$ is injective, i.e. ker$_0$ $\sigma_{\mathcal{B}} = \{1\}$.*

So $\sigma_{\mathcal{B}}$ provides the most finely grained conditional structure on $\widehat{\Omega}$: No different elements $\widehat{\omega}_1 \neq \widehat{\omega}_2$ are equivalent with respect to $\mathcal{B}$.

Next, we define a homomorphism $g : \mathcal{F}_{\mathcal{B}} \to \mathcal{F}_{\mathcal{R}}$ via

$$g(\mathbf{b}^{\pm}_{v,l}) = \prod_{\substack{1 \leqslant i \leqslant n \\ \psi_{v,l} \sqsubseteq (b_i | A_i)}} \mathbf{a}^{\pm}_i = \prod_{\substack{1 \leqslant i \leqslant n \\ b_i = v, C_{v,l} \models A_i}} \mathbf{a}^{\pm}_i, \tag{7}$$

where the equalities hold according to Lemma 1. Note that g – as $\mathcal{R}$ – is not known but only assumed to exist.

It is important to note that for different atoms v and v', only different $\mathbf{a}^{+}_i$ occur in $g(\mathbf{b}^{+}_{v,l})$ and $g(\mathbf{b}^{+}_{v',l'})$, respectively, by Lemma 1 (analogically for $\mathbf{a}^{-}_i$ and $g(\mathbf{b}^{-}_{v,l})$ and $g(\mathbf{b}^{-}_{v',l'})$). Moreover, each $\mathbf{a}^{+}_i$ and $\mathbf{a}^{-}_i$ occurs at most once in each $g(\mathbf{b}^{+}_{v,l})$ and $g(\mathbf{b}^{-}_{v,l})$, respectively. This will be used several times in the sequel. g establishes a connection between the conditional structures with respect to $\mathcal{B}$ and to – the unknown, but existing – $\mathcal{R}$:

Theorem 7 *Let $g : \mathcal{F}_{\mathcal{B}} \to \mathcal{F}_{\mathcal{R}}$ be as in (7). Then $\sigma_{\mathcal{R}} = g \circ \sigma_{\mathcal{B}}$.*

Theorem 7 provides immediately a method for determining *ker* g by considering $\sigma_{\mathcal{B}}$ and *ker*$_0$ $\sigma_{\mathcal{R}}$ = *ker*$_0$ P (cf. (5)).

Corollary 8 *$\widehat{\omega} \in$ ker$_0$ $\sigma_{\mathcal{R}}$ iff $\widehat{\omega} \in \widehat{\Omega}_0$ and $\sigma_{\mathcal{B}}(\widehat{\omega}) \in$ ker g.*

Proposition 9 *Let $\widehat{\omega} = \omega_1^{r_1} \cdot \ldots \cdot \omega_m^{r_m} \in \widehat{\Omega}_0$.*
Then $\sigma_{\mathcal{B}}(\omega_1^{r_1} \cdot \ldots \cdot \omega_m^{r_m}) \in$ ker g iff for all atoms v in $\mathcal{L}$,

$$\prod_{\substack{1 \leqslant k \leqslant m \\ \omega_k = C_{v,l} v}} (\mathbf{b}^{+}_{v,l})^{r_k}, \quad \prod_{\substack{1 \leqslant k \leqslant m \\ \omega_k = C_{v,l} \overline{v}}} (\mathbf{b}^{-}_{v,l})^{r_k} \in \textit{ker } g. \tag{8}$$

So each (generating) element of *ker*$_0$ $\sigma_{\mathcal{R}}$ gives rise to an equation modulo *ker* g for the generators $\mathbf{b}^{+}_{v,l}, \mathbf{b}^{-}_{v,l}$ of $\mathcal{F}_{\mathcal{B}}$.

Corollary 10 *Let v be an atom of the language $\mathcal{L}$. Set $\mathbf{b}_{v,l} = \dfrac{\mathbf{b}^{+}_{v,l}}{\mathbf{b}^{-}_{v,l}}$. $\prod_{1 \leqslant k \leqslant m} (\mathbf{b}^{+}_{v,l_k})^{r_k} \in$ ker g iff $\prod_{1 \leqslant k \leqslant m} (\mathbf{b}^{-}_{v,l_k})^{r_k} \in$ ker g iff $\prod_{1 \leqslant k \leqslant m} (\mathbf{b}_{v,l_k})^{r_k} \in$ ker g.*

The idea of the procedure to be described in the sequel is to exploit the relations mod *ker* g holding between the group elements $\mathbf{b}_{v,l} \in \mathcal{F}_{\mathcal{B}}$ with the aim to define a finite sequence of sets $\mathcal{S}^{(0)}, \mathcal{S}^{(1)}, \ldots$ of conditionals approximating $\mathcal{R}$:

$$\textit{ker}_0\ \sigma_{\mathcal{S}^{(0)}} \subseteq \textit{ker}_0\ \sigma_{\mathcal{S}^{(1)}} \subseteq \ldots \subseteq \textit{ker}_0\ \sigma_{\mathcal{R}} \tag{9}$$

We will first present the fundamental techniques and state the necessary theoretical results. In the next section, the procedure will be explained by an example and applied to ME-reasoning.

We start with setting $\mathcal{S}^{(0)} = \mathcal{B}$. Lemma 6 states *ker*$_0$ $\sigma_{\mathcal{S}^{(0)}} = 1$, so (9) trivially holds. Let $\equiv_g$ denote the equivalence relation mod *ker* g on $\mathcal{F}_{\mathcal{B}}$, i.e. $\mathbf{b}_1 \equiv_g \mathbf{b}_2$ iff $g(\mathbf{b}_1) = g(\mathbf{b}_2)$ for any two elements $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{F}_{\mathcal{B}}$. For each (generating) element $\widehat{\omega} = \omega_1^{r_1} \cdot \ldots \cdot \omega_m^{r_m}$ of *ker*$_0$ P = *ker*$_0$ $\sigma_{\mathcal{R}}$, set up an equation modulo *ker* g:

$$\sigma_{\mathcal{B}}(\widehat{\omega}) \equiv_g 1,$$

and split up these equations according to Proposition 9 and Corollary 10. Set $g^{(0)} := g$. The basic idea of the method is to eliminate, or to join conditionals by $\sqcup$, respectively, in accordance with the equations modulo *ker* g. We first summarize the technical prerequisites for each step:

Prerequisites: $\mathcal{S}^{(t)}$ is a set of conditionals $\varphi^{(t)}_{v,j}$ with a single atom v in the conclusion, and the antecedent $D^{(t)}_{v,j}$ of $\varphi^{(t)}_{v,j}$ is a disjunction of elementary conjunctions not containing v. Let $\mathcal{F}_{\mathcal{S}^{(t)}} = \langle \mathbf{s}^{(t)\,+}_{v,j}, \mathbf{s}^{(t)\,-}_{v,j} \rangle_{v,j}$ be the free abelian group associated with $\mathcal{S}^{(t)}$, and let $g^{(t)} : \mathcal{F}_{\mathcal{S}^{(t)}} \to \mathcal{F}_{\mathcal{R}}$ be the homomorphism defined by

$$g^{(t)}(\mathbf{s}^{(t)}_{v,j}) = \prod_{\substack{1 \leqslant i \leqslant n \\ v = b_i, D^{(t)}_{v,j} \models A_i}} \mathbf{a}_i$$

such that $g^{(t)} \circ \sigma_{\mathcal{S}^{(t)}} = \sigma_{\mathcal{R}}$. Let $\equiv_{g^{(t)}}$ mean $\equiv$ modulo *ker* $g^{(t)}$.

We show how to exploit equations of the form

$$\mathbf{s}^{(t)}_{v,j_0} \equiv_{g^{(t)}} \mathbf{s}^{(t)}_{v,j_1} \ldots \mathbf{s}^{(t)}_{v,j_m} \tag{10}$$

to modify $\mathcal{S}^{(t)}$ appropriately. To obtain this modified set $\mathcal{S}^{(t+1)}$,

1. eliminate $\varphi^{(t)}_{v,j_0}$ from $\mathcal{S}^{(t)}$;
2. replace each $\varphi^{(t)}_{v,j_k}$ by

$$\varphi^{(t+1)}_{v,j_k} = \varphi^{(t)}_{v,j_0} \sqcup \varphi^{(t)}_{v,j_k} = (v \mid D^{(t)}_{v,j_0} \vee D^{(t)}_{v,j_k}),$$

   for $1 \leqslant k \leqslant m$. Set $D^{(t+1)}_{v,j_k} = D^{(t)}_{v,j_0} \vee D^{(t)}_{v,j_k}$, $1 \leqslant k \leqslant m$;
3. retain all other $\varphi^{(t)}_{w,l}$, i.e.

$$\varphi^{(t+1)}_{w,l} = \varphi^{(t)}_{w,l} \quad \text{for} \quad w \neq v \text{ or } l \notin \{j_0, j_1, \ldots, j_m\}.$$

This also includes the case $m = 0$, i.e. $\varphi^{(t)}_{v,j_0} \equiv_{g^{(t)}} 1$; in this case, Step 2 is vacuous and therefore is left out. Define homomorphisms $h^{(t+1)} : \mathcal{F}_{\mathcal{S}^{(t)}} \to \mathcal{F}_{\mathcal{S}^{(t+1)}}$ and $g^{(t+1)} : \mathcal{F}_{\mathcal{S}^{(t+1)}} \to \mathcal{F}_{\mathcal{R}}$ by

$$h^{(t+1)}(\mathbf{s}^{(t)}_{w,l}) = \begin{cases} \prod_{1 \leqslant k \leqslant m} \mathbf{s}^{(t+1)}_{v,j_k} & \text{if } w = v, l = j_0 \\ \mathbf{s}^{(t+1)}_{v,j_k} & \text{if } w = v, l = j_k, 1 \leqslant k \leqslant m \\ \mathbf{s}^{(t+1)}_{w,l} & \text{else} \end{cases}$$

and

$$g^{(t+1)}(\mathbf{s}^{(t+1)}_{w,l}) = \prod_{\substack{1 \leqslant i \leqslant n \\ w = b_i, D^{(t+1)}_{w,l} \models A_i}} \mathbf{a}_i.$$

Lemma 11 *Let $\mathcal{S}^{(t+1)}, h^{(t+1)}, g^{(t+1)}$ be defined as above. Then the following relationships hold:*

(i) $g^{(t+1)} \circ h^{(t+1)} = g^{(t)}$;
(ii) $h^{(t+1)} \circ \sigma_{\mathcal{S}^{(t)}} = \sigma_{\mathcal{S}^{(t+1)}}$;
(iii) $g^{(t+1)} \circ \sigma_{\mathcal{S}^{(t+1)}} = \sigma_{\mathcal{R}}$.

So the new set $\mathcal{S}^{(t+1)}$ is apt to continue the set chain (9):

Corollary 12 *With the same notation as in Lemma 11, it holds that*

$$\textit{ker}_0\ \sigma_{\mathcal{S}^{(t)}} \subseteq \textit{ker}_0\ \sigma_{\mathcal{S}^{(t+1)}} \subseteq \textit{ker}_0\ \sigma_{\mathcal{R}} \tag{11}$$

By replacing each group element $\mathbf{s}^{(t)}_{v,l}$ by $h^{(t+1)}(\mathbf{s}^{(t)}_{v,l})$, equations holding modulo *ker* $g^{(t)}$ are transformed into equations modulo *ker* $g^{(t+1)}$:

$$g^{(t)}(\prod_k (\mathbf{s}^{(t)}_{v,l_k})^{r_k}) = 1 \quad \text{iff} \quad g^{(t+1)}(\prod_k h^{(t+1)}(\mathbf{s}^{(t)}_{v,l_k})^{r_k}) = 1,$$

due to Lemma 11(i). Note that while neither $\mathcal{R}$ nor g are known, the homomorphisms $h^{(t)}$ will approximate $\mathcal{R}$ in a constructive way.

If all equations modulo *ker* g can be solved successfully, then, finally, no non-trivial equations modulo *ker* $g^{(t')}$ are left for some t'. That is, for any $\widehat{\omega} \in$ *ker*$_0$ $\sigma_{\mathcal{R}}$, $1 = \sigma_{\mathcal{R}}(\widehat{\omega}) = g^{(t')} \circ \sigma_{\mathcal{S}^{(t')}}(\widehat{\omega})$ holds trivially, i.e. due to $\sigma_{\mathcal{S}^{(t')}}(\widehat{\omega}) = 1$. But this means *ker*$_0$ $\sigma_{\mathcal{R}} \subseteq$ *ker*$_0$ $\sigma_{\mathcal{S}^{(t')}} \subseteq$ *ker*$_0$ $\sigma_{\mathcal{R}}$, so *ker*$_0$ $\sigma_{\mathcal{S}^{(t')}} =$ *ker*$_0$ $\sigma_{\mathcal{R}}$. In this case, repeating the procedure described above yields a suitable set $\mathcal{S}^{(t')}$ of conditionals which is faithfully represented by P. The appertaining probabilities may be calculated directly from P.

6 Example – solving the inverse maxent problem

We will now illustrate the method described in the previous section by an example. Given some positive probability distribution P, we will show how to calculate efficiently a set $\mathcal{S}^{prob}$ of (probabilistic) conditionals such that $P = ME(\mathcal{S}^{prob})$. P is indifferent with respect to each such set $\mathcal{S}^{prob}$, so we have *ker*$_0$ $\mathcal{S} \subseteq$ *ker*$_0$ $P =: K$.

We consider formulas involving the three atomic propositions a - *being a student*, b - *being young*, and c - *being single (i.e. unmarried)*. The distribution P over a, b, c is given as follows:

ω	$P(\omega)$	ω	$P(\omega)$	ω	$P(\omega)$	ω	$P(\omega)$
abc	0.1950	$ab\overline{c}$	0.1758	$a\overline{b}c$	0.0408	$a\overline{b}\overline{c}$	0.0519
$\overline{a}bc$	0.1528	$\overline{a}b\overline{c}$	0.1378	$\overline{a}\overline{b}c$	0.1081	$\overline{a}\overline{b}\overline{c}$	0.1378

Here important relationships between probabilities are revealed by $P(\overline{a}b\overline{c}) = P(\overline{a}\overline{b}\overline{c})$, $P(\frac{abc}{ab\overline{c}}) = P(\frac{\overline{a}bc}{\overline{a}b\overline{c}})$, $P(\frac{a\overline{b}c}{a\overline{b}\overline{c}}) = P(\frac{\overline{a}\overline{b}c}{\overline{a}\overline{b}\overline{c}})$, determining the kernel of P as $K = \left\langle \frac{\overline{a}b\overline{c}}{\overline{a}\overline{b}\overline{c}}, \frac{abc \cdot \overline{a}b\overline{c}}{ab\overline{c} \cdot \overline{a}bc}, \frac{a\overline{b}c \cdot \overline{a}\overline{b}\overline{c}}{a\overline{b}\overline{c} \cdot \overline{a}\overline{b}c} \right\rangle$.

We list the twelve basic single-elementary conditionals $\psi_{v,l}$ of $\mathcal{B}$:

$$\begin{array}{lll} \psi_{a,0} = (a \mid \overline{b}\overline{c}) & \psi_{b,0} = (b \mid \overline{a}\,\overline{c}) & \psi_{c,0} = (c \mid \overline{a}\overline{b}) \\ \psi_{a,1} = (a \mid \overline{b}c) & \psi_{b,1} = (b \mid \overline{a}c) & \psi_{c,1} = (c \mid \overline{a}b) \\ \psi_{a,2} = (a \mid b\overline{c}) & \psi_{b,2} = (b \mid a\overline{c}) & \psi_{c,2} = (c \mid a\overline{b}) \\ \psi_{a,3} = (a \mid bc) & \psi_{b,3} = (b \mid ac) & \psi_{c,3} = (c \mid ab) \end{array}$$

with corresponding generators $\mathbf{b}^+_{v,l}, \mathbf{b}^-_{v,l}$ of $\mathcal{F}_{\mathcal{B}}$. The generators of K yield the following equations modulo *ker* g, due to Corollary 8:

$$1 \equiv_g \sigma_{\mathcal{B}}\left(\frac{\overline{a}b\overline{c}}{\overline{a}\overline{b}\overline{c}}\right) = \frac{\mathbf{b}^-_{a,2}\mathbf{b}^+_{b,0}\mathbf{b}^-_{c,1}}{\mathbf{b}^-_{a,0}\mathbf{b}^-_{b,0}\mathbf{b}^-_{c,0}}$$

$$1 \equiv_g \sigma_{\mathcal{B}}\left(\frac{abc \cdot \overline{a}b\overline{c}}{ab\overline{c} \cdot \overline{a}bc}\right) = \frac{\mathbf{b}^+_{a,3}\mathbf{b}^+_{b,3}\mathbf{b}^+_{c,3} \cdot \mathbf{b}^-_{a,2}\mathbf{b}^+_{b,0}\mathbf{b}^-_{c,1}}{\mathbf{b}^+_{a,2}\mathbf{b}^+_{b,2}\mathbf{b}^-_{c,3} \cdot \mathbf{b}^-_{a,3}\mathbf{b}^+_{b,1}\mathbf{b}^+_{c,1}}$$

$$1 \equiv_g \sigma_{\mathcal{B}}\left(\frac{a\overline{b}c \cdot \overline{a}\overline{b}\overline{c}}{a\overline{b}\overline{c} \cdot \overline{a}\overline{b}c}\right) = \frac{\mathbf{b}^+_{a,1}\mathbf{b}^-_{b,3}\mathbf{b}^+_{c,2} \cdot \mathbf{b}^-_{a,0}\mathbf{b}^-_{b,0}\mathbf{b}^-_{c,0}}{\mathbf{b}^+_{a,0}\mathbf{b}^-_{b,2}\mathbf{b}^-_{c,2} \cdot \mathbf{b}^-_{a,1}\mathbf{b}^-_{b,1}\mathbf{b}^+_{c,0}}$$

Considering these equations for each atom a, b, c separately and omitting the $\{+, -\}$-signs (see Proposition 9 and Corollary 10), we obtain

$$\mathbf{b}_{a,0} \equiv_g \mathbf{b}_{a,2}, \mathbf{b}_{c,0} \equiv_g \mathbf{b}_{c,1}, \mathbf{b}_{b,0} \equiv_g 1,$$
$$\mathbf{b}_{a,2} \equiv_g \mathbf{b}_{a,3}, \mathbf{b}_{c,1} \equiv_g \mathbf{b}_{c,3}, \mathbf{b}_{b,0}\mathbf{b}_{b,3} \equiv_g \mathbf{b}_{b,1}\mathbf{b}_{b,2},$$
$$\mathbf{b}_{a,0} \equiv_g \mathbf{b}_{a,1}, \mathbf{b}_{c,0} \equiv_g \mathbf{b}_{c,2}.$$

This yields
$$\mathbf{b}_{a,0} \equiv_g \mathbf{b}_{a,1} \equiv_g \mathbf{b}_{a,2} \equiv_g \mathbf{b}_{a,3},$$
$$\mathbf{b}_{c,0} \equiv_g \mathbf{b}_{c,1} \equiv_g \mathbf{b}_{c,2} \equiv_g \mathbf{b}_{c,3},$$
$$\mathbf{b}_{b,0} \equiv_g 1, \mathbf{b}_{b,3} \equiv_g \mathbf{b}_{b,1}\mathbf{b}_{b,2}.$$

Eliminating $\psi_{b,0}$ and joining conditionals according to these equations, as described by the algorithm in Section 5, results in the following conditionals:

$$\psi_{a,0} \sqcup \psi_{a,1} \sqcup \psi_{a,2} \sqcup \psi_{a,3} \equiv (a|\top);$$
$$\psi_{c,0} \sqcup \psi_{c,1} \sqcup \psi_{c,2} \sqcup \psi_{c,3} \equiv (c|\top);$$
$$\psi_{b,3} \sqcup \psi_{b,1} \equiv (b|c); \quad \psi_{b,3} \sqcup \psi_{b,2} \equiv (b|a).$$

Associating the proper probabilities with these structural conditionals, we obtain $\mathcal{S}^{prob} = \{(a|\top)[0.4635], (c|\top)[0.4967], (b|a)[0.8], (b|c)[0.7]\}$ as an ME-generating set for P, i.e. $P = ME(\mathcal{S}^{prob})$.

7 Outlook

In general, the techniques described in Section 5 will not suffice to eliminate all equations modulo *ker* g, and we will be left with more complex equations modulo *ker* $g^{(t)}$ of the form

$$\prod_k (\mathbf{s}^{(t)}_{v,j_k})^{r_k} \equiv_{g^{(t)}} \prod_l (\mathbf{s}^{(t)}_{v,j_l})^{s_l}, \tag{12}$$

all $r_k, s_l > 0$. The great variety of relationships possibly holding between the conditionals involved makes it difficult, if not impossible in general, to construct a new appropriate set $\mathcal{S}^{(t+1)}$ of conditionals in a straightforward way.

Nevertheless, the method developed so far already illustrates the central idea of how to find the conditionals whose structures some probability function P (or some ordinal conditional function κ) follows: By investigating relationships between the numerical values of P, the effects of conditionals are analyzed and isolated, and conditionals are joined suitably so as to fit the conditional structures inherent to P. The operations on conditionals are based on equations between group elements representing these conditionals.

The applicability of the method presented in this paper neither depends on the presupposition of P being a faithful respresentation nor on having a complete description of *ker*$_0$ P available: Each numerical relationship found amongst the values of P corresponds to an element of *ker*$_0$ P and may be used to set up equations for the group elements in $\mathcal{F}_{\mathcal{B}}$ modulo *ker* g. The generators of *ker*$_0$ P are particularly appropriate for this task, in that they yield basic equations, but any other element will do, too. If P fails to be a faithful representation of some suitable set of conditionals, then too many equations modulo *ker* g will have to be solved trivially. In this case, backtracking will be necessary, undoing the last joining of conditionals.

Though at the present state, the method is not guaranteed to terminate successfully, we will find that in many cases, it will yield a useful approximation of the hidden set $\mathcal{R}$ of conditionals. Treating equations of form (12) is a topic of our ongoing research, and results will be published in a further paper.

REFERENCES

[1] B. DeFinetti, *Theory of Probability*, volume 1,2, John Wiley and Sons, New York, 1974.

[2] M. Goldszmidt and J. Pearl, ‘Qualitative probabilities for default reasoning, belief revision, and causal modeling’, *Artificial Intelligence*, **84**, 57–112, (1996).

[3] G. Kern-Isberner, ‘Characterizing the principle of minimum cross-entropy within a conditional-logical framework’, *Artificial Intelligence*, **98**, 169–208, (1998).

[4] G. Kern-Isberner. A unifying framework for symbolic and numerical approaches to nonmonotonic reasoning and belief revision, 1999. (manuscript).

[5] G. Kern-Isberner, ‘Conditional preservation and conditional indifference’, in *Proceedings 8th International Workshop on Nonmonotonic Reasoning (NMR’2000)*, Breckenridge, Colorado, (2000).

[6] R.C. Lyndon and P.E. Schupp, *Combinatorial group theory*, Springer, Berlin Heidelberg New York, 1977.

[7] J.B. Paris and A. Vencovská, ‘In defence of the maximum entropy inference process’, *International Journal of Approximate Reasoning*, **17**, 77–103, (1997).