Basic Independence Results for Maximum Entropy Reasoning Based on Relational Conditionals

Marco Wilhelm[1], Gabriele Kern-Isberner[1], and Andreas Ecke[2]

[1] Technical University of Dortmund, Dortmund, Germany
[2] Technical University of Dresden, Dresden, Germany
✉ marco.wilhelm@tu-dortmund.de

Abstract

Maximum entropy reasoning (ME-reasoning) based on relational conditionals combines both the capability of ME-distributions to express uncertain knowledge in a way that excellently fits to commonsense, and the great expressivity of an underlying first-order logic. The drawbacks of this approach are its high complexity which is generally paired with a costly domain size dependency, and its non-transparency due to the non-existent a priori independence assumptions as against in Bayesian networks. In this paper, we present some independence results for ME-reasoning based on the aggregating semantics for relational conditionals that help to disentangle the composition of ME-distributions, and therefore, lead to a problem reduction and provide structural insights into ME-reasoning.

1 Introduction

In recent years, *relational probabilistic logics* [3, 7, 8, 12] became the focus of interest due to their expressive power when modeling uncertain knowledge about interactions between individual objects. The *principle of maximum entropy* (ME-principle) [5] then again constitutes a most appropriate form of commonsense probabilistic reasoning [9]. As it fulfills the paradigm of informational economy [2], it provides a probability distribution which satisfies given probabilistic knowledge and adds as little information as possible. Hence, ME-reasoning based on relational conditionals combines both the capability of maximum entropy distributions to express uncertain knowledge in a way that excellently fits to commonsense, and the great expressivity of the underlying first-order logic. The drawbacks of the maximum entropy approach are, on the one hand, its high complexity and therefore the need of elaborate strategies to deal with large numbers of objects. And on the other hand, its non-transparency due to the non-existent a priori independence assumptions in contrast to Bayesian networks [11]. Due to this, the ME-principle is often regarded as a black box methodology.

In this paper, we present some independence results for ME-reasoning based on relational conditionals that help to disentangle the composition of ME-distributions, and therefore, provide structural insights into ME-reasoning. Formally, we consider a reasoner's knowledge base to consist of two distinct sets: A finite set of factual knowledge and a finite set of conditional beliefs. While facts are represented by grounded first-order formulas and are treated as being definitely true, conditional beliefs represent defeasible rules: A conditional $(B|A)[p]$ is a formal representation of the statement "if A holds, then B follows with probability p", where A and B are not necessarily grounded first-order formulas. As a common ground, both facts and beliefs constitute constraints for probability distributions to serve as the reasoner's epistemic state. As an intuition, these distributions assign the likelihood of representing the real world to possible worlds, based on the assumptions made in the reasoner's knowledge base. While facts just force the distributions to assign the probability zero to certain worlds, conditional beliefs need a more comprehensive semantics. Here, we rely on the so-called aggregating semantics

[8] which is inspired by statistical approaches, but sums up probabilities instead of just counting instances. More precisely, it stipulates that the probabilities of all possible worlds, each probability weighted with the number of verified ground instances of the conditional (i.e., both the premise and the conclusion of the ground instance are true within the respective possible world), sums up to the probabilities of all possible worlds, each weighted with the number of applicable ground instances (i.e., the premise is true) and multiplied with the probability of the conditional. Therefore, the aggregating semantics combines the statistical and the subjective view on probabilities proposed by Halpern [4].

The ME-distribution then is the unique probability distribution which maximizes entropy while satisfying the constraints given by the facts and beliefs according to the aggregating semantics. In reference to its property of *system independence* [13], the ME-distribution factorizes if these constraints only make an impact on the single factors, i.e., if the constraints themselves are independent. Here, we show that the latter happens in two different cases: If the knowledge base splits into syntactically independent parts (= "independence based on syntax splitting"), and if the ground instances of the conditionals show a certain isomorphism property (= "independence based on isomorphic operands"). Furthermore, these independence results carry over to ME-reasoning: Under some additional constraints, drawing inferences and revising epistemic states can be performed on the individual independent parts of the ME-distribution.

The rest of this paper is organized as follows: First, we recall essentials of ME-reasoning for some deeper syntactical and structural analysis. Afterwards, we devote a separate section to both independence studies (independence based on syntax splitting and on isomorphic operands). Finally, we conclude. Most of the proofs are omitted due to space restrictions.

2 Background: Knowledge Representation and Reasoning Formalisms with Respect to Maximum Entropy

We consider a function-free first-order language FOL over the signature $\Sigma = (\mathsf{Pred}, \mathsf{Const})$ consisting of a finite set of predicates Pred with a fixed arity and a finite set of constants Const. An atom $P(t_1, \ldots, t_k)$ is a predicate P of arity k followed by terms $t_1, \ldots, t_k$, where a term is either a variable or a constant. We typically use the symbols x, y, z for variables and a, b, c for constants. A *literal* is an atom or its negation. Formulas in FOL are inductively built from atoms using negation $\neg$, binary connectives $\wedge$ and $\vee$, as well as quantification $\exists x \in \mathsf{C}.\phi$ and $\forall x \in \mathsf{C}.\phi$ with $\mathsf{C} \subseteq \mathsf{Const}$. A variable is called *free* if it is not bounded by a quantifier, and a formula without variables is *ground*. The set of all *ground atoms* based on the signature Σ is denoted with $\mathcal{G}(\Sigma)$. Every formula $F \in \mathsf{FOL}$ can be grounded by substituting every free variable in F with a constant, and by carrying out all the quantifications. We denote the set of all *ground instances* of F built this way with $\mathsf{Grnd}(F)$. Further, $\mathsf{Lit}(\mathcal{G})$ denotes the set of all *ground literals* built on the ground atoms in $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$. In order to shorten mathematical expressions, we abbreviate $A \wedge B$ with AB, $\neg A$ with $\overline{A}$, and $A \vee \overline{A}$ with $\top$ for formulas $A, B \in \mathsf{FOL}$.

2.1 Probabilistic Relational Conditional Knowledge Bases

In order to represent uncertain knowledge, we use *(probabilistic) conditionals* of the form $(B|A)[p]$ with $A, B \in \mathsf{FOL}$ and $p \in [0, 1]$.[1] Such a conditional serves as a formalization of

[1] Principally, it would be possible to formulate conditionals with interval probabilities. However, reasoning with interval probabilities quickly leads to non-informative results (cf. [6]). In this paper, we therefore identify probabilities with best expectation values.

the statement "if A holds, then B follows with probability p", albeit we have to clarify its meaning when A or B contains free variables, which is explicitly allowed in our framework. In addition, we allow to add a constraint set $\mathcal{C}$ to a conditional $(B|A)[p]$, which restricts the domain of the free variables in A and B. The constraints in $\mathcal{C}$ have to be of the form $x \in \mathsf{C}$ or $x \neq y$ where x and y are free variables in A or B and C is a subset of Const. A *ground instance* of such a constrained conditional $(B|A)\langle\mathcal{C}\rangle[p]$ is obtained by grounding A and B (as well as the constraints in $\mathcal{C}$) such that free variables occurring in both A and B are substituted with the same constant in A and B, and all the (grounded) constraints are satisfied. If $\mathcal{C}$ is empty or contains valid formulas only, we omit $\langle\mathcal{C}\rangle$ in the notation of conditionals. The set of all ground instances of $(B|A)\langle\mathcal{C}\rangle[p]$ is denoted with $\mathsf{Grnd}((B|A)\langle\mathcal{C}\rangle[p])$.

Example 1.

1. *Let* $r_1 = (R(x,y)|Q(x))[p]$ *be a conditional, and let* $a, b \in \mathsf{Const}$. *Then, both* $(R(a,b)|Q(a))[p]$ *and* $(R(a,a)|Q(a)[p]$ *are proper ground instances of* r_1, *but the conditional* $(R(a,b)|Q(b))[p]$ *is not.*
2. *Let* $r_2 = (R(x,y)|\top)\langle\{x \neq y\}\rangle[p]$ *be a conditional, and let* $\mathsf{Const} = \{a, b\}$. *Then, the set of all ground instances of* r_2 *is* $\mathsf{Grnd}(r_2) = \{(R(a,b)|\top)[p], (R(b,a)|\top)[p]\}$.

A *knowledge base* $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ consists of two components: $\mathcal{F}$ is a finite set of ground formulas in FOL representing *factual* knowledge which is incontrovertible true, and $\mathcal{B}$ is a finite set of conditionals with non-trivial probabilities (i.e., $p \notin \{0,1\}$) representing what is *believed* to be true by an agent. When dropping the restriction $p \notin \{0,1\}$, the factual knowledge could principally be integrated into $\mathcal{B}$ (just add the conditionals $(F|\top)[1]$ for every $F \in \mathcal{F}$ to $\mathcal{B}$). However, we want to maintain the differentiation between facts and beliefs in order to highlight their different semantics.

Convention 1. *Let* $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ *be a knowledge base. Throughout the rest of the paper, we refer to the i-th conditional in* $\mathcal{B}$ *with* $r_i = (B_i|A_i)\langle\mathcal{C}_i\rangle[p_i]$, *i.e.,* A_i *(resp.* B_i, $\mathcal{C}_i$, *and* p_i*) refers to the premise (resp. consequence, constraint set, and probability) of the i-th conditional in* $\mathcal{B}$. *Further,* $n = |\mathcal{B}|$ *denotes the number of conditionals in* $\mathcal{B}$.

To be able to separate ground instances of conditionals as well as whole knowledge bases on a syntactical level, we introduce the notion of the *support* of a conditional respectively a knowledge base. Informally, the support is the set of all ground atoms that may occur in any ground instance of the conditional respectively of any conditional in the knowledge base.

Definition 1 (Support)**.** *The* support $\mathsf{Supp}(F)$ of a ground formula $F \in \mathsf{FOL}$ *is the set of all ground atoms that occur in* F. *We further inductively define the support of*

1. *an arbitrary formula* $F \in \mathsf{FOL}$ *as* $\mathsf{Supp}(F) = \bigcup_{F' \in \mathsf{Grnd}(F)} \mathsf{Supp}(F')$,
2. *a conditional* r *as* $\mathsf{Supp}(r) = \bigcup_{(B|A)[p] \in \mathsf{Grnd}(r)} \big(\mathsf{Supp}(A) \cup \mathsf{Supp}(B)\big)$,
3. *a knowledge base* $\mathcal{KB}$ *as* $\mathsf{Supp}(\mathcal{KB}) = \bigcup_{F \in \mathcal{F}} \mathsf{Supp}(F) \cup \bigcup_{r \in \mathcal{B}} \mathsf{Supp}(r)$.

Example 2. *The support of conditional* r_2 *from Example 1 is* $\mathsf{Supp}(r_2) = \{R(a,b), R(b,a)\}$.

The formal semantics of knowledge bases is given by probability distributions over possible worlds. Let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms. A *possible world* ω *over* $\mathcal{G}$ is a complete conjunction of the ground literals in $\mathsf{Lit}(\mathcal{G})$. We denote the set of all possible worlds over $\mathcal{G}$ with $\Omega(\mathcal{G})$. A ground literal $L \in \mathsf{Lit}(\mathcal{G}(\Sigma))$ is entailed by a possible world ω, written $\omega \models L$, iff L occurs in ω. This entailment relation shall be extended to arbitrary ground formulas in the usual way. Further, let $\mathcal{F}$ be a set of ground formulas. We write $\omega \models \mathcal{F}$ iff $\omega \models F$ for all $F \in \mathcal{F}$, and $\omega \not\models \mathcal{F}$ iff $\neg(\omega \models \mathcal{F})$ holds. Provided that $\mathcal{G}$ is a proper subset of $\mathcal{G}(\Sigma)$, a possible world $\omega \in \Omega(\mathcal{G})$ more precisely is a *partial* possible world, as it does not describe the (possible) state of the *whole* world, but refers to the disjunction $\omega = \bigvee_{\omega' \in \Omega(\mathcal{G}(\Sigma)):\ \omega' \models \omega} \omega'$ of truly possible worlds. We also have the following definition.

Definition 2 (Projection of Possible Worlds)**.** *Let $\mathcal{G}, \mathcal{G}' \subseteq \mathcal{G}(\Sigma)$ be sets of ground atoms with $\mathcal{G} \subseteq \mathcal{G}'$, and let $\omega \in \Omega(\mathcal{G}')$ be a possible world over $\mathcal{G}'$. The* projection of ω on $\mathcal{G}$ *is the possible world $\omega|_{\mathcal{G}} \in \Omega(\mathcal{G})$ defined by*

$$\omega|_{\mathcal{G}} = \bigwedge_{\substack{l \in \mathsf{Lit}(\mathcal{G}) \\ \omega \models l}} l.$$

Example 3. *Let $\mathcal{G}(\Sigma) = \{A(a), A(b), B(a), B(b)\}$ and $\mathcal{G} = \{A(a), B(a)\}$. The projection of the possible world $\omega = \overline{A(a)}\ A(b)\ B(a)\ \overline{B(b)} \in \Omega(\mathcal{G}(\Sigma))$ on $\mathcal{G}$ is $\omega|_{\mathcal{G}} = \overline{A(a)}\ B(a)$.*

We make use of this fine-grained notion of (partial) possible worlds and projections of possible worlds when obtaining marginal distributions.

While the probabilistic evaluation of facts $F \in \mathcal{F}$ and conditionals $(B|A)[p] \in \mathcal{B}$ *without* free variables is very straightforward (possible worlds that contradict any $F \in \mathcal{F}$ should have probability zero, and a conditional $(B|A)[p]$ without free variables constrains a probability distribution $\mathcal{P}$ to satisfy the conditional probability $\mathcal{P}(AB) \cdot \mathcal{P}(A)^{-1} = p$ where $\mathcal{P}(A) = \sum_{\omega \models A} \mathcal{P}(\omega)$), conditionals with free variables leave some room for interpretation. In this point, we rely on the *aggregating semantics* [14], which is inspired by statistical approaches, but sums up probabilities instead of just counting instances. Therefore it combines the statistical and the subjective view on probabilities proposed by Halpern [4].

Definition 3 (Aggregating Semantics)**.** *Let $r = (B|A)\langle\mathcal{C}\rangle[p]$ be a conditional, let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms with $\mathsf{Supp}(r) \subseteq \mathcal{G}$, and let $\mathcal{P}(\mathcal{G}) : \Omega(\mathcal{G}) \to [0,1]$ be a probability distribution. $\mathcal{P}(\mathcal{G})$ is a $\mathcal{G}$-*model *of r, denoted $\mathcal{P}(\mathcal{G}) \models r$, iff*

$$\frac{\sum_{(B'|A')[p] \in \mathsf{Grnd}(r)} \mathcal{P}(\mathcal{G})(A'B')}{\sum_{(B'|A')[p] \in \mathsf{Grnd}(r)} \mathcal{P}(\mathcal{G})(A')} = p.$$

Note that the aggregating semantics reduces to the standard conditional probability fulfillment $\mathcal{P}(\mathcal{G})(AB) \cdot \mathcal{P}(\mathcal{G})(A)^{-1} = p$ if r does not contain free variables. Further, if $\mathcal{P}(\mathcal{G})$ is a uniform distribution, we end up with a purely statistical interpretaion of the conditional.

Altogether, we now can define when a probability distribution models a (whole) knowledge base $\mathcal{KB}$. If so, the probability distribution can be understood as the *epistemic state* of an agent whose knowledge is $\mathcal{KB}$.

Definition 4 ($\mathcal{G}$-Model of a Knowledge Base)**.** *Let $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ be a knowledge base, let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms with $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}$, and let $\mathcal{P}(\mathcal{G}) : \Omega(\mathcal{G}) \to [0,1]$ be a probability distribution. $\mathcal{P}(\mathcal{G})$ is a $\mathcal{G}$-*model *of*

1. *$\mathcal{F}$, written $\mathcal{P}(\mathcal{G}) \models \mathcal{F}$, iff* *$\mathcal{P}(\mathcal{G})(\omega) = 0$ for all $\omega \in \Omega(\mathcal{G})$ with $\omega \not\models \mathcal{F}$.*
2. *$\mathcal{B}$, written $\mathcal{P}(\mathcal{G}) \models \mathcal{B}$, iff* *$\mathcal{P}(\mathcal{G}) \models r$ for all $r \in \mathcal{B}$.*
3. *$\mathcal{KB}$, written $\mathcal{P}(\mathcal{G}) \models \mathcal{KB}$, iff* *$\mathcal{P}(\mathcal{G})$ is a $\mathcal{G}$-model of both $\mathcal{F}$ and $\mathcal{B}$.*

If $\mathcal{P}(\mathcal{G})$ is a $\mathcal{G}$-model of $\mathcal{KB}$ and $\mathcal{G} = \mathcal{G}(\Sigma)$, we call $\mathcal{P}(\mathcal{G})$ a model *of $\mathcal{KB}$ for short.*

Obviously, the following characterization of being a model of $\mathcal{KB}$ according to the aggregating semantics holds:

Proposition 1. *Let $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ be a knowledge base, let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms with $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}$, and let*

$$\Omega(\mathcal{G})|_{\mathcal{F}} = \{\omega \in \Omega(\mathcal{G}) \mid \omega \models \mathcal{F}\}$$

be the set of possible worlds over $\mathcal{G}$ that do not contradict the facts in $\mathcal{F}$. Further, let $\mathcal{P}'(\mathcal{G}) : \Omega(\mathcal{G})|_{\mathcal{F}} \to [0,1]$ be a probability distribution with

$$\frac{\sum_{(B'|A')[p] \in \mathsf{Grnd}(r)} \mathcal{P}'(\mathcal{G})(A'B')}{\sum_{(B'|A')[p] \in \mathsf{Grnd}(r)} \mathcal{P}'(\mathcal{G})(A')} = p \qquad \text{for all} \quad r = (B|A)\langle\mathcal{C}\rangle[p] \in \mathcal{B}.$$

Then,

$$\mathcal{P}(\mathcal{G})(\omega) = \begin{cases} \mathcal{P}'(\mathcal{G})(\omega) & \text{iff} \quad \omega \in \Omega(\mathcal{G})|_{\mathcal{F}} \\ 0 & \text{iff} \quad \omega \in \Omega(\mathcal{G}) \setminus \Omega(\mathcal{G})|_{\mathcal{F}} \end{cases}$$

is a $\mathcal{G}$-model of $\mathcal{KB}$. Every $\mathcal{G}$-model of $\mathcal{KB}$ can be built this way.

Proposition 1 allows us to investigate the facts and the beliefs in a knowledge base consecutively. Once the set of possible worlds $\Omega(\mathcal{G})|_{\mathcal{F}}$ is determined, the conditional beliefs in $\mathcal{B}$ can be evaluated with respect to $\Omega(\mathcal{G})|_{\mathcal{F}}$, and without taking the original set $\mathcal{F}$ into account.

A knowledge base $\mathcal{KB}$ is called *$\mathcal{G}$-consistent* if it has at least one $\mathcal{G}$-model. Note that if $\mathcal{KB}$ is $\mathcal{G}$-consistent for some $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$, it is $\mathcal{G}'$-consistent for all sets of ground atoms $\mathcal{G}'$ with $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}'$. Thus, we can say that $\mathcal{KB}$ is *consistent* iff it is $\mathcal{G}(\Sigma)$-consistent, i.e., iff it has a model (with respect to $\mathcal{G}(\Sigma)$). However, as we want to formulate independence results with respect to concrete sets $\mathcal{G}$ later on, we will maintain the notions of $\mathcal{G}$-consistency and $\mathcal{G}$-models.

2.2 Maximum Entropy Reasoning

For every $\mathcal{G}$-consistent knowledge base, there is a distinct $\mathcal{G}$-model based on the *principle of maximum entropy* (ME-principle) [13, 9]. This *maximum entropy distribution* (ME-distribution) $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$ is the unique probability distribution that maximizes entropy among all $\mathcal{G}$-models of $\mathcal{KB}$. According to [10], the ME-distribution is the one distribution which fits best to commonsense. It is defined as follows.

Definition 5 (Maximum Entropy Distribution)**.** *Let $\mathcal{KB}$ be a $\mathcal{G}$-consistent knowledge base. The maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$ relative to $\mathcal{G}$ and $\mathcal{KB}$ is defined by*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB}) = \arg \max_{\mathcal{P}(\mathcal{G}) \models \mathcal{KB}} - \sum_{\substack{\omega \in \Omega(\mathcal{G}) \\ \mathcal{P}(\mathcal{G})(\omega) \neq 0}} \mathcal{P}(\mathcal{G})(\omega) \cdot \log \mathcal{P}(\mathcal{G})(\omega).$$

According to [6], there is a product representation of the maximum entropy distribution based on the solution of a nonlinear equation system. Before we recall this representation, we have to introduce some further notations. For this, let $r = (B|A)\langle\mathcal{C}\rangle[p]$ be a conditional, let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms, and let $\omega \in \Omega(\mathcal{G})$. Then,

$$\mathsf{app}(\mathcal{G}, r)(\omega) = |\{(B'|A')[p] \in \mathsf{Grnd}(r) \mid \mathsf{Supp}((B'|A')[p]) \subseteq \mathcal{G} \quad \text{and} \quad \omega \models A'\}|,$$
$$\mathsf{ver}(\mathcal{G}, r)(\omega) = |\{(B'|A')[p] \in \mathsf{Grnd}(r) \mid \mathsf{Supp}((B'|A')[p]) \subseteq \mathcal{G} \quad \text{and} \quad \omega \models A'B'\}|,$$

$$\phi(\mathcal{G},r)(\omega) = \mathsf{ver}(\mathcal{G},r)(\omega) - p \cdot \mathsf{app}(\mathcal{G},r)(\omega).$$

The functions $\mathsf{app}(\mathcal{G},r)(\omega)$ and $\mathsf{ver}(\mathcal{G},r)(\omega)$ count the number of those ground instances of r which are applicable respectively verified in ω and which additionally consist of ground atoms in $\mathcal{G}$ only. The function $\phi(\mathcal{G},r)(\omega)$ is a determining factor of the aforementioned nonlinear equation system and hence of the ME-distribution relative to $\mathcal{KB} = (\mathcal{F},\mathcal{B})$ itself, as it can be used to combine the logical information of a conditional $r \in \mathcal{B}$ provided by $\mathsf{app}(\mathcal{G},r)$ and $\mathsf{ver}(\mathcal{G},r)$, as well as its probabilistic information p.

We now illustrate the product representation of the ME-distribution.

Proposition 2. *Let $\mathcal{KB} = (\mathcal{F},\mathcal{B})$ be a knowledge base with $\mathcal{B} = \{r_1,\dots,r_n\}$. Further, let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms with* $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}$. *The maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})$ exists iff the vector function $\vec{\mathcal{S}}(\mathcal{G},\mathcal{KB})(x_1,\dots,x_n) = (\mathcal{S}_1,\dots,\mathcal{S}_n)$ with*

$$\mathcal{S}_i = \mathcal{S}_i(\mathcal{G},\mathcal{KB})(x_1,\dots,x_n) = \sum_{\omega \in \Omega(\mathcal{G})|_{\mathcal{F}}} \phi(\mathcal{G},r_i)(\omega) \cdot \prod_{j=1}^{n} x_j^{\phi(\mathcal{G},r_j)(\omega)}, \qquad i = 1,\dots,n, \tag{1}$$

has a root in $(\mathbb{R}_{>0})^n$.[2] *Once such a root $(\alpha_1,\dots,\alpha_n)$ is given, which we call the* effect *of $\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})$, the maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})$ is characterized by*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})(\omega) = \begin{cases} \alpha_0(\mathcal{G},\mathcal{KB}) \cdot \prod_{i=1}^{n} \alpha_i^{\phi(\mathcal{G},r_i)(\omega)} & \text{iff} \quad \omega \in \Omega(\mathcal{G})|_{\mathcal{F}} \\ 0 & \text{iff} \quad \omega \in \Omega(\mathcal{G}) \setminus \Omega(\mathcal{G})|_{\mathcal{F}} \end{cases},$$

where $\alpha_0(\mathcal{G},\mathcal{KB})$ is a normalizing constant defined by

$$\alpha_0(\mathcal{G},\mathcal{KB}) = \left(\sum_{\omega \in \Omega(\mathcal{G})|_{\mathcal{F}}} \prod_{i=1}^{n} \alpha_i^{\phi(\mathcal{G},r_i)(\omega)} \right)^{-1}.$$

We now consider the reasoning tasks of drawing nonmonotonic inferences at maximum entropy from ME-distributions and revising epistemic states represented by ME-distributions. For the first task, the principle of maximum entropy yields a *nonmonotonic inference relation*, which answers the question:

"With which probability should a conditional be believed based on a given knowledge base and according to the principle of maximum entropy?"

This inference relation satisfies important postulates for nonmonotonic inferences such as *inclusion*, *idempotence*, *cumulativity*, and *loop* [6] and is defined as follows.

Definition 6 (Nonmonotonic Inference Relation)**.** *Let $\mathcal{KB}$ be a $\mathcal{G}$-consistent knowledge base. We say that $\mathcal{KB}$* infers *a conditional r under the principle of maximum entropy, or $\mathcal{KB}$* ME-*infers r for short, which is written*

$$\mathcal{KB} \mathrel{|\!\sim}_{\mathsf{ME}} r, \quad \text{iff} \quad \mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB}) \models r. \tag{2}$$

Note that if $\mathcal{KB}$ ME-infers r for some set of ground atoms $\mathcal{G}$ with $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}$, then $\mathcal{KB}$ ME-infers r for every $\mathcal{G}$ with $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}$, which makes the inference relation $\mathrel{|\!\sim}_{\mathsf{ME}}$ independent of $\mathcal{G}$ (at least if $\mathcal{G}$ covers all ground atoms that may occur in $\mathcal{KB}$). The inference relation

[2] Finding a root of the vector function $\vec{\mathcal{S}}(\mathcal{G},\mathcal{KB})$ is mathematically the same as solving the nonlinear equation system $\vec{\mathcal{S}}(\mathcal{G},\mathcal{KB}) = \vec{0}$.

(2) is also potent enough to subsume the inference of factual knowledge, as the probability p in the query conditional is not restricted to $p \in (0,1)$, i.e., we may define for logical reasons

$$\mathcal{KB} \mid\!\sim_{\mathsf{ME}} F \quad \text{iff} \quad \mathcal{KB} \mid\!\sim_{\mathsf{ME}} (F|\top)[1], \qquad F \in \mathsf{FOL}.$$

As another reasoning task, we investigate *belief revision* respectively revising epistemic states. Belief revision is the process of integrating new information into existing knowledge without generating any inconsistency. Therefore, it answers the question:

"How should a probability distribution, which represents an agent's epistemic state, be revised according to recent knowledge of the agent?"

The essential of belief revision is to tread the new information as more reliable than the prior knowledge and to adjust the prior knowledge such that integrating the new information is uncritical. Following the *principle of conditional preservation* [6], the revised probability distribution $\mathcal{P}^*(\mathcal{G},\mathcal{KB})$ should have *minimal cross-entropy* relative to the prior distribution $\mathcal{P}(\mathcal{G})$, i.e.,

$$\mathcal{P}^*(\mathcal{G},\mathcal{KB}) = \arg \min_{\substack{\mathcal{Q}(\mathcal{G})\models\mathcal{KB} \\ \forall\omega\in\Omega(\mathcal{G}).\mathcal{P}(\mathcal{G})(\omega)\Rightarrow\mathcal{Q}(\mathcal{G})(\omega)}} \sum_{\substack{\omega\in\Omega(\mathcal{G}) \\ \mathcal{Q}(\mathcal{G})(\omega)\neq 0}} \mathcal{Q}(\mathcal{G})(\omega) \cdot \log \frac{\mathcal{Q}(\mathcal{G})(\omega)}{\mathcal{P}(\mathcal{G})(\omega)}.$$

To maintain compatibility between the prior distribution $\mathcal{P}(\mathcal{G})$ and the posterior distribution $\mathcal{P}^*(\mathcal{G},\mathcal{KB})$, we assume the recent knowledge $\mathcal{KB}$ to be $\mathcal{P}(\mathcal{G})$-consistent.

Definition 7 ($\mathcal{P}(\mathcal{G})$-consistency)**.** *Let $\mathcal{P}(\mathcal{G}) : \Omega(\mathcal{G}) \to [0,1]$ be a probability distribution, and let $\mathcal{KB}$ be a knowledge base with $\mathsf{Supp}(\mathcal{KB}) \subseteq \mathcal{G}$. $\mathcal{KB}$ is called* $\mathcal{P}(\mathcal{G})$-consistent *iff there is a probability distribution $\mathcal{P}'(\mathcal{G}) : \Omega(G) \to [0,1]$ with $\mathcal{P}'(\mathcal{G}) \models \mathcal{KB}$ which is* $\mathcal{P}(\mathcal{G})$-continuous*, i.e., $\mathcal{P}(\mathcal{G})(\omega) = 0$ implies $\mathcal{P}'(\mathcal{G})(\omega) = 0$ for all $\omega \in \Omega(\mathcal{G})$.*

If $\mathcal{KB}$ is $\mathcal{P}(\mathcal{G})$*-consistent*, then $\mathcal{P}^*(\mathcal{G},\mathcal{KB})$ is guaranteed to exist (cf. [1]) and is called the *c-revision of $\mathcal{P}(\mathcal{G})$ with respect to $\mathcal{KB}$.*

Both reasoning tasks, drawing inferences and revising epistemic states, require to solve a nonlinear equation system, either to calculate the distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})$ as an agent's epistemic state, or the (posterior) distribution $\mathcal{P}^*_{\mathsf{ME}}(\mathcal{G},\mathcal{KB}')$ after revision with $\mathcal{KB}'$, for which also an equation system based representation similar to Proposition 2 exists (cf. [6]). As these calculations are expensive and non-transparent, it would be beneficial to restrict the calculations to those parts of the equation systems, and hence of the knowledge bases themselves, that really affect the output. In the next sections, we propose some independence results for maximum entropy distributions that accomplish this request, i.e., not only the distributions themselves decompose into independent parts but also the generating equation systems. Therewith, the problem sizes of maximum entropy calculations can be reduced dramatically, which we will illustrate by an example.

3 Independence Based on Syntax Splitting

In this section, we show that syntactically independent parts of a knowledge base can be treated independently when calculating the maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})$, which results in a product representation of $\mathcal{P}_{\mathsf{ME}}(\mathcal{G},\mathcal{KB})$. We further show how this product representation affects drawing inferences at maximum entropy as well as performing c-revision.

First of all, we explain the notion of syntactical independence.

Definition 8 (Syntactically Independent Knowledge Bases)**.** *Let $\mathcal{KB}_1$ and $\mathcal{KB}_2$ be knowledge bases. We call $\mathcal{KB}_1$ and $\mathcal{KB}_2$* syntactically independent *iff* $\mathsf{Supp}(\mathcal{KB}_1) \cap \mathsf{Supp}(\mathcal{KB}_2) = \emptyset$.

We introduce some further notations. For this, let $\mathcal{KB}_1 = (\mathcal{F}_1, \mathcal{B}_1)$ and $\mathcal{KB}_2 = (\mathcal{F}_2, \mathcal{B}_2)$ be knowledge bases.

1. We write $\mathcal{KB}_1 \sqsubseteq \mathcal{KB}_2$ iff both $\mathcal{F}_1 \subseteq \mathcal{F}_2$ and $\mathcal{B}_1 \subseteq \mathcal{B}_2$.
2. We abbreviate the knowledge base $(\mathcal{F}_1 \cup \mathcal{F}_2, \mathcal{B}_1 \cup \mathcal{B}_2)$ with $\mathcal{KB}_1 \sqcup \mathcal{KB}_2$ and with $\mathcal{KB}_1 \dot{\sqcup} \mathcal{KB}_2$ if both unions $\mathcal{F}_1 \cup \mathcal{F}_2$ and $\mathcal{B}_1 \cup \mathcal{B}_2$ are disjoint.

Definition 9 (Syntax Partition)**.** *Let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms, and let $\{\mathcal{G}_1, \ldots, \mathcal{G}_m\}$ be a partition of $\mathcal{G}$. Further, let $\mathcal{KB}$ be a knowledge base. If there are $\mathcal{KB}_1, \ldots, \mathcal{KB}_m \sqsubseteq \mathcal{KB}$ such that $\mathsf{Supp}(\mathcal{KB}_i) \subseteq \mathcal{G}_i$ for $i = 1, \ldots, m$ and $\mathcal{KB}_1 \dot{\sqcup} \ldots \dot{\sqcup} \mathcal{KB}_m = \mathcal{KB}$, we call $(\mathcal{KB}_1, \ldots, \mathcal{KB}_m)$ a* syntax partition *of $\mathcal{KB}$, and we say that $\mathcal{KB}$* (syntactically) splits over *$(\mathcal{G}_1, \ldots, \mathcal{G}_m)_{\mathcal{G}}$* into *$(\mathcal{KB}_1, \ldots, \mathcal{KB}_m)$.*

Note that $\{\mathcal{KB}_1, \ldots, \mathcal{KB}_m\}$ does not have to be a "partition" of $\mathcal{KB}$ in the proper sense as some $\mathcal{KB}' \in \{\mathcal{KB}_1, \ldots, \mathcal{KB}_m\}$ may be empty.

In the situation of Definition 9, the (partial) knowledge bases $\mathcal{KB}_1, \ldots, \mathcal{KB}_m$ are pairwise syntactically independent, and whenever pairwise syntactically independent knowledge bases $\mathcal{KB}_1, \ldots, \mathcal{KB}_m \sqsubseteq \mathcal{KB}$ with $\mathcal{KB}_1 \dot{\sqcup} \ldots \dot{\sqcup} \mathcal{KB}_m = \mathcal{KB}$ exist, one can find $\mathcal{G}_1, \ldots, \mathcal{G}_m \subseteq \mathcal{G}(\Sigma)$ such that $\mathcal{KB}$ splits over $(\mathcal{G}_1, \ldots, \mathcal{G}_m)_{\bigcup_{i=1}^m \mathcal{G}_i}$ into $(\mathcal{KB}_1, \ldots, \mathcal{KB}_m)$ (just choose the supports $\mathsf{Supp}(\mathcal{KB}_1), \ldots, \mathsf{Supp}(\mathcal{KB}_m)$). However, $\{\mathcal{G}_1, \ldots, \mathcal{G}_m\}$ does not have to be a partition of $\bigcup_{i=1}^m \mathcal{G}_i$ in this case, which is the reason why we prefer the "indirect" decomposition of $\mathcal{KB}$ into syntactically independent parts as in Definition 9.[3]

The next proposition states that the maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$ factorizes into the product over $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}_i)$ for $i = 1, \ldots, m$ if $\mathcal{KB}$ splits over $(\mathcal{G}_1, \ldots, \mathcal{G}_m)_{\mathcal{G}}$ into $(\mathcal{KB}_1, \ldots, \mathcal{KB}_m)$, which is the main result of this section.

Proposition 3. *Let $\mathcal{KB}$ be a $\mathcal{G}$-consistent knowledge base that splits over $(\mathcal{G}_1, \ldots, \mathcal{G}_m)_{\mathcal{G}}$ into $(\mathcal{KB}_1, \ldots, \mathcal{KB}_m)$. Then,*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})(\omega) = \prod_{i=1}^{m} \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}_i)(\omega|_{\mathcal{G}_i}), \qquad \omega \in \Omega(\mathcal{G}). \tag{3}$$

A proof of Proposition 3 can be found in the appendix.

As a benefit, the inference query *"Does $\mathcal{P}_{\mathsf{ME}}(\mathcal{KB}) \mid\!\sim_{\mathsf{ME}} (B|A)[p]$ hold?"* is only affected by those factors $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}_i)$ of $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$ which satisfy $\mathsf{Supp}((B|A)[p]) \cap \mathcal{G}_i \neq \emptyset$.

Proposition 4. *Let $\mathcal{KB}$ be a $\mathcal{G}$-consistent knowledge base that splits over $(\mathcal{G}_1, \mathcal{G}_2)_{\mathcal{G}}$ into $(\mathcal{KB}_1, \mathcal{KB}_2)$. Further, let $r = (B|A)\langle\mathcal{C}\rangle[p]$ be a conditional with $\mathsf{Supp}(r) \subseteq \mathcal{G}_1$. Then,*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB}) \mid\!\sim_{\mathsf{ME}} (B|A)\langle\mathcal{C}\rangle[p] \quad \textit{iff} \quad \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_1, \mathcal{KB}_1) \mid\!\sim_{\mathsf{ME}} (B|A)\langle\mathcal{C}\rangle[p].$$

Similarly, when revising a belief state, only the affected parts of the epistemic state have to be adjusted.

[3] The fact that all the sets of ground atoms $\mathcal{G}_1, \ldots, \mathcal{G}_m$ are non-empty simplifies the formalization of Proposition 3 and its proof.

Proposition 5. *Let $\mathcal{P}^*(\mathcal{G}, \mathcal{KB})$ be the c-revision of the probability distribution $\mathcal{P}(\mathcal{G})$ with respect to the newly acquired knowledge in $\mathcal{KB}$. If $\mathcal{KB}$ splits over $(\mathcal{G}_1, \ldots, \mathcal{G}_m)_\mathcal{G}$ into $(\mathcal{KB}_1, \ldots, \mathcal{KB}_m)$ and $\mathcal{P}(\mathcal{G})$ factorizes over $\{\mathcal{G}_1, \ldots, \mathcal{G}_m\}$, i.e.,*

$$\mathcal{P}(\mathcal{G})(\omega) = \prod_{i=1}^{m} \mathcal{P}(\mathcal{G}_i)(\omega|_{\mathcal{G}_i}), \qquad \omega \in \Omega(\mathcal{G}),$$

then

$$\mathcal{P}^*(\mathcal{G}, \mathcal{KB})(\omega) = \prod_{i=1}^{m} \mathcal{P}^*(\mathcal{G}_i, \mathcal{KB}_i)(\omega|_{\mathcal{G}_i}), \qquad \omega \in \Omega(\mathcal{G}). \tag{4}$$

As a possible situation in which Proposition 5 could be applied, think about $\mathcal{P}(\mathcal{G})$ being a maximum entropy distribution based on a knowledge base that already split over $(\mathcal{G}_1, \ldots, \mathcal{G}_m)_\mathcal{G}$. Then, $\mathcal{P}(\mathcal{G})$ indeed factorizes over $\{\mathcal{G}_1, \ldots, \mathcal{G}_m\}$.

While the syntactical splitting of a knowledge base $\mathcal{KB}$ is a very natural and basic criterion for the independence of the maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$, we show a more advanced criterion in the next section which comprises semantical information in the notion of isomorphic operands. Beforehand, we conclude this section with an illustrating example.

Example 4. *We consider an agent who knows that every penguin is a bird and Tweety indeed is a penguin. Further, he believes that typical birds are usually able to fly, say with probability 0.9, but Tweety as a penguin is an atypical bird for which this statement is not true. Actually, the agent has never seen Tweety fly. Therefore, he assumes that Tweety can not fly with probability 0.99. Formally, the knowledge of the agent can be represented as follows:*

Let $\mathsf{Pred} = \{B/1, F/1, P/1\}$ *be the set of predicates with the meanings "x is a bird"* $(B(x))$*, "x is able to fly"* $(F(x))$*, and "x is a penguin"* $(P(x))$*. Further, let* Const *be the set of birds; in particular, tweety* $\in \mathsf{Const}$*. Then, the knowledge of the agent is* $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ *with*

$$\mathcal{F} = \{\forall x \in \mathsf{Const}. P(x) \Rightarrow B(x),\ P(tweety)\},$$
$$\mathcal{B} = \{(F(x)|B(x))\langle x \in \mathsf{Const} \setminus \{tweety\}\rangle[0.9],\ (\neg F(tweety)|\top)[0.99]\}.$$

Without loss of generality, we assume that there are k birds and the unnamed birds (= all birds except Tweety) are associated with the constants $a_1, \ldots, a_{k-1}$. The set of ground atoms over Σ,

$$\mathcal{G}(\Sigma) = \{B(tweety), F(tweety), P(tweety)\} \cup \bigcup_{i=1}^{k-1} \{B(a_i), F(a_i), P(a_i)\}$$

can be partitioned into $\{\mathcal{G}_t, \mathcal{G}'\}$ with

$$\mathcal{G}_t = \{B(tweety), F(tweety), P(tweety)\},$$
$$\mathcal{G}' = \bigcup_{i=1}^{k-1} \{B(a_i), F(a_i), P(a_i)\}.$$

Moreover, $(\mathcal{G}_t, \mathcal{G}')_{\mathcal{G}(\Sigma)}$ is a syntax partition for $\mathcal{KB}$, as $\mathcal{KB}$ splits over $(\mathcal{G}_t, \mathcal{G}')_{\mathcal{G}(\Sigma)}$ into $(\mathcal{KB}_t, \mathcal{KB}')$ with

$$\mathcal{KB}_t = (\{P(tweety) \Rightarrow B(tweety), P(tweety)\}, \{(\neg F(tweety)|\top)[0.99]\}),$$
$$\mathcal{KB}' = (\bigcup_{i=1}^{k-1} \{P(a_i) \Rightarrow B(a_i)\}, \{(F(x)|B(x))\langle x \in \mathsf{Const} \setminus \{tweety\}\rangle[0.9]\}).$$

Here, we used the fact that the formula $\forall x \in \mathsf{Const}.P(x) \Rightarrow B(x)$ is semantically equivalent to the set of formulas $\bigcup_{a\in\mathsf{Const}}\{P(a) \Rightarrow B(a)\}$. According to Proposition 3, it finally holds that

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}(\Sigma), \mathcal{KB})(\omega) = \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_t, \mathcal{KB}_t)(\omega|_{\mathcal{G}_t}) \cdot \mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')(\omega|_{\mathcal{G}'}), \quad \omega \in \Omega(\mathcal{G}(\Sigma)).$$

Hence, reasoning about Tweety is independent of the knowledge about the unnamed birds and vice versa.

4 Independence Based on Isomorphic Operands

While the focus in the previous Section 3 is on investigating independent parts of a knowledge base $\mathcal{KB}$ as a whole, we now want to separate single syntactically independent ground instances of the conditionals in $\mathcal{KB}$. However, the fact that the ground instances of the conditionals in $\mathcal{KB}$ can be grouped into syntactically independent sets is not sufficient to establish the desired factorization of the ME-distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$. Instead, we additionally need to postulate a certain isomorphism property between the syntactically independent ground instances.

Definition 10 (Isomorphic Instance Separating Partition)**.** *Let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms, and let $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}$ be a partition of $\mathcal{G}$. Further let $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ be a knowledge base. We call $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ an* instance separating partition *for $\mathcal{KB}$ iff*

1. *there are $\mathcal{F}_1, \dots, \mathcal{F}_m \subseteq \mathcal{F}$ such that $\mathcal{F} = \mathcal{F}_1 \,\dot\cup\, \dots \,\dot\cup\, \mathcal{F}_m$ and $\mathsf{Supp}(\mathcal{F}_i) \subseteq \mathcal{G}_i$ for $i = 1, \dots, m$,*
2. $\bigcup_{k=1}^{m}\{r' \in \mathsf{Grnd}(r) \mid \mathsf{Supp}(r') \subseteq \mathcal{G}_k\} = \mathsf{Grnd}(r), \qquad r \in \mathcal{B}.$

We call $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ isomorphic *iff for all pairs $\mathcal{G}_i, \mathcal{G}_j \in \{\mathcal{G}_1, \dots, \mathcal{G}_m\}$, there is a bijection $\rho : \Omega(\mathcal{G}_i)|_{\mathcal{F}_i} \to \Omega(\mathcal{G}_j)|_{\mathcal{F}_j}$ such that*

$$\phi(\mathcal{G}_i, r)(\omega) = \phi(\mathcal{G}_j, r)(\rho(\omega)), \qquad \omega \in \Omega(\mathcal{G}_i)|_{\mathcal{F}_i},\ r \in \mathcal{B}.$$

If $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ is an instance separating partition for $\mathcal{KB}$, we say that $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ satisfies the *instance separating property*. This property is purely syntactical and enables one to group all conditionals (and facts) in $\mathcal{KB}$ into syntactically independent sets. The isomorphism property of $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ instead is semantically driven, as it exploits the evaluation of the ground instances of the conditionals in $\mathcal{KB}$. It states that the different ground instances somehow all "behave" the same. Informally, across all possible worlds, the ground instances are evaluated the same, however the evaluation with respect to a particular possible world may be different. Unfortunately, one can show that the isomorphy between the elements in $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ alone is not sufficient to prove the following central independence result of this section, i.e., one may not skip the instance separating property of $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_{\mathcal{G}}$ to obtain the factorization of the ME-distribution.

In the next proposition we use the following definitions: Let r be a conditional, and let $\mathcal{G} \subseteq \mathcal{G}(\Sigma)$ be a set of ground atoms. Then, $r_{\mathcal{G}}$ denotes the conditional r where the free variables in r may only be substituted with constants from $\mathcal{G}$. Further, let $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ be a knowledge base. Then, $\mathcal{KB}_{\mathcal{G}} = (\mathcal{F}_{\mathcal{G}}, \mathcal{B}_{\mathcal{G}})$ where

$$\mathcal{F}_{\mathcal{G}} = \{F \in \mathcal{F} \mid \mathsf{Supp}(F) \subseteq \mathcal{G}\},$$
$$\mathcal{B}_{\mathcal{G}} = \{r_{\mathcal{G}} \mid r \in \mathcal{B}\}.$$

Proposition 6. *Let $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ be a $\mathcal{G}$-consistent knowledge base, and let $\{\mathcal{G}_1, \dots, \mathcal{G}_m\}_\mathcal{G}$ be an isomorphic instance separating partition for $\mathcal{KB}$. Then,*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})(\omega) = \prod_{i=1}^{m} \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}_{\mathcal{G}_i})(\omega|_{\mathcal{G}_i}), \qquad \omega \in \Omega(\mathcal{G}).$$

Example 5. *We continue Example 4 and define $\mathcal{G}_i = \{B(a_i), F(a_i), P(a_i)\}$ as well as $\mathcal{F}_i = \{P(a_i) \Rightarrow B(a_i)\}$ for $i = 1, \dots, k-1$. Then, $\{\mathcal{G}_1, \dots, \mathcal{G}_{k-1}\}_{\mathcal{G}'}$ is an isomorphic instance separating partition for $\mathcal{KB}'$. In order to see this, consider the bijection $\rho : \Omega(\mathcal{G}_i)|_{\mathcal{F}_i} \to \Omega(\mathcal{G}_j)|_{\mathcal{F}_j}$ which simply replaces the constant a_i by a_j in every (partial) possible world in $\Omega(\mathcal{G}_i)|_{\mathcal{F}_i}$ for $1 \leq i, j \leq k-1$. Following Proposition 6,*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')(\omega) = \prod_{i=1}^{k-1} \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}'_{\mathcal{G}_i})(\omega|_{\mathcal{G}_i}), \qquad \omega \in \Omega(\mathcal{G}'),$$

i.e., even every unnamed bird in the szenario of Example 4 can be investigated on its own. A more in-depth analysis shows that the probability distributions $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}'_{\mathcal{G}_i})$ for $i = 1, \dots, k-1$ are identically up to a permutation of constants, which means, that only one of these distributions actually has to be calculated in oder to determine $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')$. In total, when taking both independence results into account, it is sufficient to calculate two probability distributions ($\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_t, \mathcal{KB}_t)$ and $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_i, \mathcal{KB}'_{\mathcal{G}_i})$ for any $i \in \{1, \dots, k-1\}$), each defined over six possible worlds only, in order to determine the entire maximum entropy distribution $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$ from Example 4. Following the naïve way of directly calculating $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$ instead means calculating a probability distribution defined over 2^{3^k} possible worlds and hence crucially depends on the size of Const.

We now discuss a problem class for which the precondition of Proposition 6 is satisfied, i.e., we indicate a class of knowledge bases for which there are isomorphic instance separating partitions. For this, we consider a fragment of FOL consisting of Boolean combinations of unary predicates.

Definition 11 (BOOL)**.** *Let $\Sigma_{\mathsf{BOOL}} = (\mathsf{Pred}_1, \mathsf{Const})$ be a first-order signature consisting of a finite set of unary predicates Pred_1 and a finite set of constants* Const*. Further, let x be a variable. Then,* BOOL *is the smallest set such that*

1. $A(x) \in \mathsf{BOOL}$ *for every unary predicate* $A \in \mathsf{Pred}_1$*,*
2. $\neg A(x) \in \mathsf{BOOL}$ *for every* $A(x) \in \mathsf{BOOL}$*,*
3. $A(x) \wedge B(x) \in \mathsf{BOOL}$ *for every* $A(x), B(x) \in \mathsf{BOOL}$*,*
4. $A(x) \vee B(x) \in \mathsf{BOOL}$ *for every* $A(x), B(x) \in \mathsf{BOOL}$*.*

The definition of BOOL implies that formulas in BOOL are Boolean combinations of unary predicates with one free variable and without constants. In particular, for every formula $F \in \mathsf{BOOL}$, there are $|\mathsf{Const}|$ many ground instances, and different ground instances of the same formula are syntactically independent. Furthermore, every ground instance of F can be converted into any other just by exchanging the substituted constant. This leads to the following proposition.

Proposition 7. *Let $\mathcal{KB} = (\emptyset, \mathcal{B})$ be a $\mathcal{G}(\Sigma_{\mathsf{BOOL}})$-consistent knowledge base with $\mathcal{B} = \{r_1, \ldots, r_n\}$ and $r_i = (B_i|A_i)[p_i]$ such that $A_i, B_i \in \mathsf{BOOL}$ for $i = 1, \ldots, n$. Then,*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}(\Sigma_{\mathsf{BOOL}}), \mathcal{KB}) = \prod_{c \in \mathsf{Const}} \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_c, \mathcal{KB}_{\mathcal{G}_c})(\omega|_{\mathcal{G}_c}), \qquad \omega \in \Omega(\mathcal{G}(\Sigma_{\mathsf{BOOL}})),$$

where $\mathcal{G}_c = \{A(c) \mid A \in \mathsf{Pred}_1\}$.

Hence, Proposition 7 states that calculating the ME-distribution relative to $\mathcal{KB}$ can be reduced to calculating $|\mathsf{Const}|$ many ME-distributions, each relative to a distinct ground instantiation $\mathcal{KB}_{G_c}$ of $\mathcal{KB}$ including all the facts and believes about the particular constant c, when the original knowledge base $\mathcal{KB}$ is a knowledge base without facts and with conditionals built upon formulas in BOOL. In other words, the influence of a single domain element on the ME-distribution is independent from the other domain elements.

We conclude this section with a proposition which states how the independence result in Proposition 6 facilitates drawing inferences from knowledge bases built upon formulas in BOOL.

Proposition 8. *Let $\mathcal{KB} = (\emptyset, \mathcal{B})$ be a $\mathcal{G}(\Sigma_{\mathsf{BOOL}})$-consistent knowledge base with $\mathcal{B} = \{r_1, \ldots, r_n\}$ and $r_i = (B_i|A_i)[p_i]$ such that $A_i, B_i \in \mathsf{BOOL}$ for $i = 1, \ldots, n$. Further, let $(B|A)[p]$ be a conditional with $A, B \in \mathsf{BOOL}$ and $|\mathsf{Const}| = k$. Then,*

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB}) \mathrel{|\!\!\sim}_{\mathsf{ME}} (B|A)[p] \quad \textit{iff} \quad \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_c, \mathcal{KB}_{G_c}) \mathrel{|\!\!\sim}_{\mathsf{ME}} (B(c)|A(c))[p]$$

for an arbitrary constant $c \in \mathsf{Const}$.

Proposition 8 makes it possible to transfer knowledge about a certain domain element to the entirety of all domain elements and vice versa, since all the domain elements are interchangeable. In particular, the inferred probability p of the query conditional $(B|A)[p]$ with $A, B \in \mathsf{BOOL}$ is independent of $|\mathsf{Const}|$, i.e., of the domain size.

5 Conclusion and Future Work

We investigated maximum entropy reasoning (ME-reasoning) based on the aggregating semantics for relational conditional knowledge bases. We showed that under certain circumstances the ME-distribution factorizes into independent parts. One sufficient precondition for this is present when the knowledge base itself splits into syntactically independent parts. As a consequence, ME-reasoning tasks such as drawing inferences at maximum entropy and revising epistemic states by using c-revision may be performed on the single factors of the ME-distribution. This can help to reduce computational costs and the frequently criticized intricacy of ME-reasoning. Another more complex precondition for the factorization of the ME-distribution is fulfilled if the ground instances of the conditionals in the knowledge base satisfy a certain isomorphism property. As a concrete instance of knowledge bases which show this property, we discussed knowledge bases consisting of conditionals whose premises and conclusions are Boolean combinations of unary predicates.

In future work, we want to intensify our studies on independence results for maximum entropy distributions, in particular with regard to conditional independencies.

Acknowledgments

This research was supported by the German National Science Foundation (DFG), Research Unit FOR 1513 on Hybrid Reasoning for Intelligent Systems.

References

[1] Imre Csiszar. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3(1):146–158, 1975.

[2] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, 1988.

[3] Lise Getoor and Ben Taskar, editors. *Introduction to Statistical Relational Learning.* MIT Press, 2007.

[4] Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46(3):311–350, 1990.

[5] Edwin T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

[6] Gabriele Kern-Isberner. *Conditionals in nonmonotonic reasoning and belief revision.* Springer, 2001.

[7] Gabriele Kern-Isberner and Thomas Lukasiewicz. Combining probabilistic logic programming with the power of maximum entropy. *Artificial Intelligence*, 157(1-2):139–202, 2004.

[8] Gabriele Kern-Isberner and Matthias Thimm. Novel semantical approaches to relational probabilistic conditionals. In *Proc. of 12th KR conference.* AAAI Press, 2010.

[9] Jeff B. Paris. *The uncertain reasoner's companion – A mathematical perspective.* Cambridge University Press, 1994.

[10] Jeff B. Paris. Common sense and maximum entropy. *Synthese*, 117(1):75–93, 1999.

[11] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers Inc., 1988.

[12] Matthew Richardson and Pedro M. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.

[13] John E. Shore and Rodney W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory*, 26(1):26–37, 1980.

[14] Matthias Thimm and Gabriele Kern-Isberner. On probabilistic inference in relational conditional logics. *Logic Journal of the IGPL*, 20(5):872–908, 2012.

Appendix

We exemplarily prove Proposition 3. The basic strategy of this proof also applies to the proofs of the further propositions in Section 3 and Section 4. We stick to Convention 1 when it is appropriate, and we assume that the preconditions of Proposition 3 hold.

The following lemma is preparatory work for the proof of Proposition 3.

Lemma 1. *Let $\mathcal{KB} = (\mathcal{F}, \mathcal{B})$ be a $\mathcal{G}$-consistent knowledge base that splits over $(\mathcal{G}_1, \dots, \mathcal{G}_m)_{\mathcal{G}}$ into $(\mathcal{KB}_1, \dots, \mathcal{KB}_m)$. Further let $r \in \mathcal{B}$. Then,*

$$\phi(\mathcal{G}, r)(\omega) = \sum_{k=1}^{m} \phi(\mathcal{G}_k, r)(\omega|_{\mathcal{G}_k}).$$

Proof. It is

$$\phi(\mathcal{G}, r)(\omega) = \phi(\bigcup_{k=1}^{m} \mathcal{G}_k, r)(\bigwedge_{k=1}^{m} \omega|_{\mathcal{G}_k}) = \sum_{k=1}^{m} \phi(\mathcal{G}_k, r)(\bigwedge_{k=1}^{m} \omega|_{\mathcal{G}_k}) = \sum_{k=1}^{m} \phi(\mathcal{G}_k, r)(\omega|_{\mathcal{G}_k}). \qquad \square$$

Proof of Proposition 3. Let $\mathcal{G}' = \bigcup_{i=1}^{m-1} \mathcal{G}_i$ and $\mathcal{KB}' = \bigsqcup_{i=1}^{m-1} \mathcal{KB}_i$. We prove that

$$\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})(\omega) = \mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')(\omega|_{\mathcal{G}'}) \cdot \mathcal{P}_{\mathsf{ME}}(G_m, \mathcal{KB}_m)(\omega|_{\mathcal{G}_m}) \qquad \text{for all} \quad \omega \in \Omega(\mathcal{G}).$$

Then, Equation (4) follows by induction. Let $\mathcal{F} = \{F_1, \dots, F_k\}$. Without loss of generality, we assume that $\mathcal{KB}' = (\mathcal{F}', \mathcal{B}')$ with $\mathcal{F}' = \{F_1, \dots, F_s\}$ and $\mathcal{B}' = \{r_1, \dots, r_t\}$ and $\mathcal{KB}_m = (\mathcal{F}_m, \mathcal{B}_m)$ with $\mathcal{F}_m = \{F_{s+1}, \dots, F_k\}$ and $\mathcal{B}_m = \{r_{t+1}, \dots, r_n\}$. Let $i \in \{1, \dots, t\}$. Let $(\alpha_1, \dots, \alpha_n)$ be the effect of $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$. Then, according to Equation (1),

$$\begin{aligned}
0 &= \sum_{\omega \in \Omega(\mathcal{G})|_{\mathcal{F}}} \phi(\mathcal{G}, r_i)(\omega) \cdot \prod_{j=1}^{n} \alpha_j^{\phi(\mathcal{G}, r_j)(\omega)} \\
&= \sum_{\omega' \in \Omega(\mathcal{G}')|_{\mathcal{F}'}} \sum_{\omega_m \in \Omega(\mathcal{G}_m)|_{\mathcal{F}_m}} \Big(\phi(\mathcal{G}', r_i)(\omega') + \underbrace{\phi(\mathcal{G}_m, r_i)(\omega_m)}_{=0} \Big) \cdot \prod_{j=1}^{n} \alpha_j^{\phi(\mathcal{G}', r_j)(\omega') + \phi(\mathcal{G}_m, r_j)(\omega_j)} \\
&= \sum_{\omega' \in \Omega(\mathcal{G}')|_{\mathcal{F}'}} \sum_{\omega_m \in \Omega(\mathcal{G}_m)|_{\mathcal{F}_m}} \phi(\mathcal{G}', r_i)(\omega') \cdot \prod_{j=1}^{t} \alpha_j^{\phi(\mathcal{G}', r_j)(\omega')} \cdot \prod_{j=t+1}^{n} \alpha_j^{\phi(\mathcal{G}_m, r_j)(\omega_m)} \\
&= \Big(\sum_{\omega' \in \Omega(\mathcal{G}')|_{\mathcal{F}'}} \phi(\mathcal{G}', r_i)(\omega') \cdot \prod_{j=1}^{t} \alpha_j^{\phi(\mathcal{G}', r_j)(\omega')} \Big) \cdot \underbrace{\Big(\sum_{\omega_m \in \Omega(\mathcal{G}_m)|_{\mathcal{F}_m}} \prod_{j=t+1}^{n} \alpha_j^{\phi(\mathcal{G}_m, r_j)(\omega_m)} \Big)}_{>0} \\
&= \sum_{\omega' \in \Omega(\mathcal{G}')|_{\mathcal{F}'}} \phi(\mathcal{G}', r_i)(\omega') \cdot \prod_{j=1}^{t} \alpha_j^{\phi(\mathcal{G}', r_j)(\omega')} \\
&= \mathcal{S}_i(\mathcal{G}', \mathcal{KB}')(\alpha_1, \dots, \alpha_t).
\end{aligned}$$

Analogously, for $i \in \{t+1, \dots, n\}$, it follows that

$$0 = \sum_{\omega_m \in \Omega(\mathcal{G}_m)|_{\mathcal{F}_m}} \phi(\mathcal{G}_m, r_i)(\omega_m) \cdot \prod_{j=t+1}^{n} \alpha_j^{\phi(\mathcal{G}_m, r_j)(\omega_m)} = \mathcal{S}_i(\mathcal{G}_m, \mathcal{KB}_m)(\alpha_{t+1}, \dots, \alpha_n).$$

Hence, the maximum entropy distributions $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')$ and $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_m, \mathcal{KB}_m)$ exist, and they have the effects $(\alpha_1, \ldots, \alpha_t)$ resp. $(\alpha_{t+1}, \ldots, \alpha_n)$. Further, there is the following connection between the normalizing constants of $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})$, $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')$, and $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_m, \mathcal{KB}_m)$:

$$\begin{aligned}
\alpha_0(\mathcal{G}, \mathcal{KB}) &= \Big(\sum_{\omega \in \Omega(\mathcal{G})} \prod_{j=1}^{n} \alpha_j^{\phi(\mathcal{G}, r_j)(\omega)} \Big)^{-1} \\
&= \Big(\sum_{\omega' \in \Omega(\mathcal{G}')} \sum_{\omega_m \in \Omega(\mathcal{G}_m)} \prod_{j=1}^{t} \alpha_j^{\phi(\mathcal{G}', r_j)(\omega')} \cdot \prod_{j=t+1}^{n} \alpha_j^{\phi(\mathcal{G}_m, r_j)(\omega_m)} \Big)^{-1} \\
&= \Big(\sum_{\omega' \in \Omega(\mathcal{G}')} \prod_{j=1}^{n} \alpha_j^{\phi(\mathcal{G}', r_j)(\omega')} \Big)^{-1} \cdot \Big(\sum_{\omega_m \in \Omega(\mathcal{G}_m)} \prod_{j=1}^{n} \alpha_j^{\phi(\mathcal{G}_m, r_j)(\omega_m)} \Big)^{-1} \\
&= \alpha_0(\mathcal{G}', \mathcal{KB}') \cdot \alpha_0(\mathcal{G}_m, \mathcal{KB}_m)
\end{aligned}$$

For $\omega \in \Omega(\mathcal{G})|_{\mathcal{F}}$ it eventually follows that

$$\begin{aligned}
\mathcal{P}_{\mathsf{ME}}(\mathcal{G}, \mathcal{KB})(\omega) &= \alpha_0(\mathcal{G}, \mathcal{KB}) \cdot \prod_{j=1}^{n} \alpha_j^{\phi(\mathcal{G}, \mathcal{KB})(\omega)} \\
&= \alpha_0(\mathcal{G}', \mathcal{KB}') \cdot \alpha_0(\mathcal{G}_m, \mathcal{KB}_m) \cdot \prod_{j=1}^{t} \alpha_j^{\phi(\mathcal{G}', \mathcal{KB}')(\omega|_{\mathcal{G}'})} \cdot \prod_{j=t+1}^{n} \alpha_j^{\phi(\mathcal{G}_m, \mathcal{KB}_m)(\omega|_{\mathcal{G}_m})} \\
&= \alpha_0(\mathcal{G}', \mathcal{KB}') \cdot \prod_{j=1}^{t} \alpha_j^{\phi(\mathcal{G}', \mathcal{KB}')(\omega|_{\mathcal{G}'})} \cdot \alpha_0(\mathcal{G}_m, \mathcal{KB}_m) \cdot \prod_{j=t+1}^{n} \alpha_j^{\phi(\mathcal{G}_m, \mathcal{KB}_m)(\omega|_{\mathcal{G}_m})} \\
&= \mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')(\omega|_{\mathcal{G}'}) \cdot \mathcal{P}_{\mathsf{ME}}(\mathcal{G}_m, \mathcal{KB}_m)(\omega|_{\mathcal{G}_m}).
\end{aligned}$$

In the case where $\mathcal{F} \neq \emptyset$, for every $\omega \in \Omega(\mathcal{G}) \backslash \Omega(\mathcal{G})|_{\mathcal{F}}$, at least one of the factors $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}', \mathcal{KB}')(\omega')$ and $\mathcal{P}_{\mathsf{ME}}(\mathcal{G}_m, \mathcal{KB}_m)(\omega_m)$ is zero, as every $F \in \mathcal{F}$ is either in $\mathcal{KB}'$ or $\mathcal{KB}_m$, which completes the proof. □