The Complexity of the Consistency Problem in the Probabilistic Description Logic $\mathcal{ALC}^{\mathsf{ME}}$ *

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

[1] Dept. of Computer Science, TU Dresden, Dresden, Germany
[2] Dept. of Computer Science, TU Dortmund, Dortmund, Germany

Abstract. The probabilistic Description Logic $\mathcal{ALC}^{\mathsf{ME}}$ is an extension of the Description Logic $\mathcal{ALC}$ that allows for uncertain conditional statements of the form "if C holds, then D holds with probability p," together with probabilistic assertions about individuals. In $\mathcal{ALC}^{\mathsf{ME}}$, probabilities are understood as an agent's degree of belief. Probabilistic conditionals are formally interpreted based on the so-called aggregating semantics, which combines a statistical interpretation of probabilities with a subjective one. Knowledge bases of $\mathcal{ALC}^{\mathsf{ME}}$ are interpreted over a fixed finite domain and based on their maximum entropy (ME) model. We prove that checking consistency of such knowledge bases can be done in time polynomial in the cardinality of the domain, and in exponential time in the size of a binary encoding of this cardinality. If the size of the knowledge base is also taken into account, the combined complexity of the consistency problem is NP-complete for unary encoding of the domain cardinality and NExpTime-complete for binary encoding.

1 Introduction

Description Logics (DLs) [2] are a well-investigated family of logic-based knowledge representation languages, which can be used to represent *terminological knowledge* about concepts as well as *assertional knowledge* about individuals. DLs constitute the formal foundation of the Web Ontology Language OWL,[3] and they are frequently used for defining biomedical ontologies [9]. DLs are (usually decidable) fragments of first-order logic, and thus inherit the restrictions of *classical* logic: they cannot be used to represent uncertain knowledge. In many application domains (e.g., medicine), however, knowledge is not necessarily certain. For example, a doctor may not know definitely that a patient has influenza, but only believe that this is the case with a certain probability. This is an example for a so-called *subjective* probability. From a technical point of view, subjective probabilities are often formalized using probability distributions over possible worlds (i.e., interpretations). To obtain the probability of an assertion like "John has influenza," one then sums up the probabilities of the worlds that

* This work was supported by the German Research Foundation (DFG) within the Research Unit FOR 1513 "Hybrid Reasoning for Intelligent Systems".

[3] see `https://www.w3.org/TR/owl2-overview/`

satisfy the assertion. Another type of probability, called *statistical*, is needed to treat general statements like "humans have their heart on the left with probability p." In this setting, one wants to compare the number of individuals that are human and have their heart on the left with the number of all humans *within one world*, rather than summing up the probabilities of the worlds where all humans have their heart on the left. Thus, when defining a probabilistic DL, there is a need for treating assertional knowledge using subjective probabilities, and terminological knowledge using a statistical approach. More information on the distinction between statistical and subjective probabilities can be found in [8]. Most probabilistic extensions of DLs handle either subjective probabilities [12] or statistical ones [15], or are essentially classical terminologies over probabilistic databases [4].

The probabilistic DL $\mathcal{ALC}^{\mathsf{ME}}$ [23] was designed such that it can accommodate both points of view. In $\mathcal{ALC}^{\mathsf{ME}}$, the terminological part of the knowledge base consists of probabilistic conditionals, which are statements of the form $(D|C)[p]$, which can be read as "if C holds for an individual, then D holds for this individual with probability p." Such a probability should be understood as an agent's degree of belief. Formally, the meaning of probabilistic conditionals is defined using the so-called *aggregating semantics* [11]. This semantics generalizes the statistical interpretation of conditional probabilities by combining it with subjective probabilities based on probability distributions over possible worlds. Basically, in a fixed possible world, the conditional $(D|C)$ can be evaluated statistically by the relative fraction of those individuals that belong to both C and D measured against the individuals that belong to C. In the aggregating semantics, this fraction is not built independently for every possible world, but the single numerators and denominators of the fractions are respectively weighted with the probability of the respective possible world, and are summed up thereafter. Hence, the aggregating semantics mimics statistical probabilities from a subjective point of view. Assertions can then be interpreted in a purely subjective way by summing up the probabilities of the worlds in which the respective assertion holds. Due to this combination of statistical and subjective probabilities, the models of $\mathcal{ALC}^{\mathsf{ME}}$-knowledge bases are probability distributions over a set of interpretations that serve as possible worlds. These worlds are built over a fixed finite domain, which guarantees that this set of interpretations is also finite and constitutes a well-defined probability space.

The aggregating semantics defines what the models of an $\mathcal{ALC}^{\mathsf{ME}}$ knowledge base are. However, reasoning w.r.t. all these models is usually not productive due to the vast number of probabilistic models. For this reason, we choose as a single model of a knowledge base its *maximum entropy (*ME*) distribution* [14]. From a commonsense point of view, the maximum entropy distribution is a good choice as it fulfills a number of commonsense principles that can be subsumed under the main idea that "essentially similar problems should have essentially similar solutions" [13]. Moreover, the maximum entropy distribution is known to process conditional relationships particularly well according to conditional logic standards [10]. If the knowledge base is *consistent* in the sense that it has an aggregating semantics model, then it also has a unique maximum entropy model [10,14]. Hence, deciding whether an $\mathcal{ALC}^{\mathsf{ME}}$ knowledge base has a maximum entropy model is the same as deciding whether it has a model according to the aggregating semantics. For this reason, we restrict our attention to deciding the latter inference problems. This is relevant also if one wants to use the aggregating semantics without its combination with maximum entropy.

It should be noted that the general approach of using the aggregating semantics in combination with maximum entropy to define the semantics of probabilistic conditionals has been introduced and discussed before [20,11], and is not particular to probabilistic DLs. A detailed discussions of the aggregating semantics (plus ME) and comparisons with related approaches, in particular with approaches by Halpern and colleagues (see, e.g., [8,7]), can be found in [20]. The instantiation of this approach with the DL $\mathcal{ALC}$ was first considered in our previous work [23], and the investigation of the computational properties of the resulting logic $\mathcal{ALC}^{\mathsf{ME}}$ is continued in the present paper. To be more precise, we first show that checking consistency of an $\mathcal{ALC}^{\mathsf{ME}}$ knowledge base is possible in time polynomial in the cardinality of the finite domain used to construct the possible worlds, and in time exponential in the size of the binary encoding of this cardinality. The first of these two complexity results was already shown in [23] for $\mathcal{ALC}^{\mathsf{ME}}$ knowledge bases without assertions. An important tool for proving this result was the use of so-called types, which have also been employed to show complexity results for classical DLs and other logics [17,16]. In order to extend this result to $\mathcal{ALC}^{\mathsf{ME}}$ knowledge bases with probabilistic assertions, we need to modify the notion of types such that it can also accommodate individuals. The second contribution of the present paper is to determine the *combined complexity* of checking consistency in $\mathcal{ALC}^{\mathsf{ME}}$, i.e., the complexity measured w.r.t. the domain size *and* the size of the knowledge base. For unary encoding of the domain cardinality, we show that this problem is in NP, and for binary encoding that it is in NExpTime. Since fixed domain reasoning in classical $\mathcal{ALC}$ is already NP-complete in the unary case [18] and NExpTime-complete in the binary case [6] these complexity bounds are tight. These results show that the complexity of fixed-domain reasoning in $\mathcal{ALC}$ does not increase if probabilistic conditionals and probabilistic assertions with aggregating semantics are added.

The rest of the paper is organized as follows. First, we start with a brief repetition of the classical DL $\mathcal{ALC}$. We extend $\mathcal{ALC}$ with probabilistic conditionals and assertions and introduce the aggregating semantics as a probabilistic interpretation of knowledge bases within the resulting probabilistic DL $\mathcal{ALC}^{\mathsf{ME}}$. Since the consistency problem for $\mathcal{ALC}^{\mathsf{ME}}$ knowledge bases does not depend on the ME distribution, we do not define this distribution formally in the present paper (see [23] for the exact definition), but illustrate its usefulness by an example. After that, we introduce our notion of types, and use it to give an alternative proof of the known ExpTime upper bound for consistency of classical $\mathcal{ALC}$ knowledge bases. Based on the approach used in this proof, we then show our complexity results for consistency in $\mathcal{ALC}^{\mathsf{ME}}$ using a translation into a system of linear equations over the real numbers, whose variables basically correspond to multisets types.

2 The Description Logics $\mathcal{ALC}$ and $\mathcal{ALC}^{\mathsf{ME}}$

We start with a brief introduction of the classical DL $\mathcal{ALC}$, and then introduce its probabilistic variant $\mathcal{ALC}^{\mathsf{ME}}$.

Classical $\mathcal{ALC}$ The basic building blocks of most DLs are the pairwise disjoint sets of concept names N_C, role names N_R, and individual names N_I. From these, the set of $\mathcal{ALC}$ concepts is defined inductively as follows:

- every concept name $A \in N_C$ is an $\mathcal{ALC}$ concept;
- $\top$ (top concept) and $\bot$ (bottom concept) are $\mathcal{ALC}$ concepts;
- if C, D are $\mathcal{ALC}$ concepts and $r \in N_R$ is a role name, then $\neg C$ (negation), $C \sqcap D$ (conjunction), $C \sqcup D$ (disjunction), $\exists r.C$ (existential restriction), and $\forall r.C$ (value restriction) are also $\mathcal{ALC}$ concepts.

An $\mathcal{ALC}$ concept inclusion (GCI) is of the form $C \sqsubseteq D$, where C and D are $\mathcal{ALC}$ concepts. A classical $\mathcal{ALC}$ TBox is a finite set of $\mathcal{ALC}$ concept inclusions. An $\mathcal{ALC}$ assertion is of the form $C(a)$ where C is an $\mathcal{ALC}$ concept and $a \in N_I$, or $r(a, b)$ with $r \in N_R$ and $a, b \in N_I$. A classical $\mathcal{ALC}$ ABox is a finite set of $\mathcal{ALC}$ assertions. Together, TBox and ABox form an $\mathcal{ALC}$ knowledge base (KB).

The semantics of $\mathcal{ALC}$ is based on interpretations. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty set of elements $\Delta^{\mathcal{I}}$, the domain, and an interpretation function that assigns to each concept name $A \in N_C$ a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, to each role name $r \in N_R$ a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and to each individual name $a \in N_I$ an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. The interpretation function is extended to $\mathcal{ALC}$ concepts as follows:

$$
\begin{aligned}
\top^{\mathcal{I}} &= \Delta^{\mathcal{I}}, \qquad \bot^{\mathcal{I}} = \emptyset, \qquad (\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}, \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}}, \qquad (C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}, \\
(\exists r.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}}.(d, e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}}\}, \\
(\forall r.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} \mid \forall e \in \Delta^{\mathcal{I}}.(d, e) \in r^{\mathcal{I}} \implies e \in C^{\mathcal{I}}\}.
\end{aligned}
$$

An interpretation $\mathcal{I}$ satisfies a concept inclusion $C \sqsubseteq D$ ($\mathcal{I} \models C \sqsubseteq D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. It is a model of a TBox $\mathcal{T}$ if it satisfies all concept inclusions occurring in $\mathcal{T}$. $\mathcal{I}$ satisfies an assertion $C(a)$ ($\mathcal{I} \models C(a)$) if $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and $r(a, b)$ ($\mathcal{I} \models r(a, b)$) if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. It is a model of an ABox $\mathcal{A}$ if it satisfies all assertions in $\mathcal{A}$. A KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is consistent if there exists a model that satisfies both $\mathcal{T}$ and $\mathcal{A}$.

Note that we do *not* employ the unique name assumption (UNA), i.e., we do not assume that different individual names are interpreted by different elements of the interpretation domain.

Probabilistic $\mathcal{ALC}^{\mathsf{ME}}$ In our probabilistic extension $\mathcal{ALC}^{\mathsf{ME}}$ of $\mathcal{ALC}$, we use probabilistic conditionals instead of concept inclusions. A probabilistic $\mathcal{ALC}$ conditional is of the form $(D|C)[p]$, where C and D are $\mathcal{ALC}$ concepts and $p \in [0, 1]$. We call a finite set of probabilistic conditionals a CBox. A probabilistic ABox (or pABox) contains assertions labeled with probabilities, i.e., probabilistic assertions of the form $C(a)[p]$ or $r(a, b)[p]$, where again $p \in [0, 1]$. A probabilistic knowledge base (pKB) consists of both a CBox and a pABox.[4]

Example 1. Using probabilistic $\mathcal{ALC}^{\mathsf{ME}}$, we can express that every person has at least one friend, on average one in two people are unhappy, and that people with only happy friends are much more likely to be happy themselves in the following CBox:

$$\mathcal{C} = \{(\exists \mathsf{friend.Person} \mid \mathsf{Person})[1], \quad (\neg\mathsf{Happy} \mid \mathsf{Person})[0.5], \\ (\mathsf{Happy} \mid \mathsf{Person} \sqcap \forall \mathsf{friend.Happy})[0.9]\}.$$

Additionally, let us introduce the persons Emma and Peter, for whom we state that Emma considers Peter a friend, and Peter is quite happy:

$$\mathcal{A} = \{\mathsf{Person}(\mathsf{peter})[1], \quad \mathsf{Person}(\mathsf{emma})[1], \\ \mathsf{Happy}(\mathsf{peter})[0.8], \quad \mathsf{friend}(\mathsf{emma}, \mathsf{peter})[0.9]\}.$$

The semantics of probabilistic conditionals and assertions is defined via probabilistic interpretations, which are probability distributions over classical interpretations. For this definition to be well-behaved, we consider a fixed, finite domain Δ and assume that the signature (i.e., the set of concept, role, and individual names) is finite. For the signature, we can simply restrict to those names that actually occur in a given pKB $\mathcal{K}$, i.e., to concept names $\mathrm{sig}_C(\mathcal{K}) = \{A \in N_C \mid A \text{ occurs in } \mathcal{K}\}$, role names $\mathrm{sig}_R(\mathcal{K}) = \{r \in N_R \mid r \text{ occurs in } \mathcal{K}\}$ and individual names $\mathrm{sig}_I(\mathcal{K}) = \{a \in N_I \mid a \text{ occurs in } \mathcal{K}\}$. Then, we denote the set of all interpretations $\mathcal{I} = (\Delta, \mathrm{sig}_C(\mathcal{K}) \to \mathcal{P}(\Delta), \mathrm{sig}_R(\mathcal{K}) \to \mathcal{P}(\Delta \times \Delta), \mathrm{sig}_I(\mathcal{K}) \to \Delta)$ as $\mathcal{I}_{\mathcal{K},\Delta}$. Since Δ and all $\mathrm{sig}_\star(\mathcal{K})$ are finite, $\mathcal{I}_{\mathcal{K},\Delta}$ is also finite. Then, a *probabilistic interpretation* is a probability distribution over $\mathcal{I}_{\mathcal{K},\Delta}$, i.e., a function $\mu : \mathcal{I}_{\mathcal{K},\Delta} \to [0, 1]$ such that $\sum_{\mathcal{I} \in \mathcal{I}_{\mathcal{K},\Delta}} \mu(\mathcal{I}) = 1$.

The *semantics of probabilistic assertions* is defined as one would expect: a probabilistic interpretation μ satisfies a probabilistic assertion of the form $C(a)[p]$ or the form $r(a, b)[p]$ if

$$\sum_{\substack{\mathcal{I} \in \mathcal{I}_{\mathcal{K},\Delta} \\ s.t.\ a^{\mathcal{I}} \in C^{\mathcal{I}}}} \mu(\mathcal{I}) = p \qquad \text{or} \qquad \sum_{\substack{\mathcal{I} \in \mathcal{I}_{\mathcal{K},\Delta} \\ s.t.\ (a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}}} \mu(\mathcal{I}) = p.$$

Defining the *semantics of probabilistic conditionals* is more involved since here we need to consider not only all possible worlds, but also all elements of

[4] We will see later (proof of Corollary 14) that setting all probabilities to 1 in a pKB basically yields a classical KB, and thus $\mathcal{ALC}^{\mathsf{ME}}$ indeed is an extension of $\mathcal{ALC}$.

the domain. There are multiple possibilities for how to combine these two dimensions. In this work, we use the aggregating semantics to define the semantics of our probabilistic extension of $\mathcal{ALC}$. Under the aggregating semantics [11], a probabilistic interpretation μ satisfies a probabilistic conditional $(D|C)[p]$, denoted $\mu \models (D|C)[p]$, if

$$\frac{\sum_{\mathcal{I}\in\mathcal{I}_{\mathcal{K},\Delta}} |C^{\mathcal{I}} \cap D^{\mathcal{I}}| \cdot \mu(\mathcal{I})}{\sum_{\mathcal{I}\in\mathcal{I}_{\mathcal{K},\Delta}} |C^{\mathcal{I}}| \cdot \mu(\mathcal{I})} = p. \tag{1}$$

A probabilistic interpretation μ is a *model of a CBox* $\mathcal{C}$ $(\mu \models \mathcal{C})$ if it satisfies all probabilistic conditionals in $\mathcal{C}$, and a *model of a pABox* $\mathcal{A}$ $(\mu \models \mathcal{A})$ if it satisfies all probabilistic assertions in $\mathcal{A}$. It is a *model of a pKB* $\mathcal{K}$ if it is a model of both its CBox and pABox.

Equation (1) formalizes the intuition underlying conditional probabilities by weighting the probabilities $\mu(\mathcal{I})$ with the number of individuals for which the conditional $(D|C)[p]$ is *applicable* $(|C^{\mathcal{I}}|)$ or *verified* $(|C^{\mathcal{I}} \cap D^{\mathcal{I}}|)$ in $\mathcal{I}$. Hence, the aggregating semantics mimics statistical probabilities from a subjective point of view, and probabilities can be understood as an agent's degrees of belief. If, on the one hand, μ is the distribution that assigns the probability 1 to a single interpretation $\mathcal{I}$, which means that the agent is certain that $\mathcal{I}$ is the real world, then the aggregating semantics boils down to counting relative frequencies in this world. On the other hand, if μ is the uniform distribution on those interpretations that do not contradict facts (conditionals or assertions with 0/1-probability), which means that the agent is minimally confident in her beliefs, then the aggregating semantics means counting relative frequencies spread over all interpretations.

Consistency is the question whether a given pKB has a model (for a given domain size). In previous work [23], we were concerned with the model of a pKB with maximal entropy, as this ME-model has several nice properties. In particular, reasoning with respect to all probabilistic models instead of solely the ME-model leads to monotonic and often uninformative inferences, as demonstrated in the next example.

Example 2. Consider the CBox $\mathcal{C} = \{(\mathsf{Happy}|\mathsf{Wealthy})[0.7],\ (\mathsf{Happy}|\mathsf{Parent})[0.9]\}$. Then $\mathcal{C}$ has a model in which wealthy parents are happy with probability 0, as well as a model in which wealthy parents are happy with probability 1. This is the case since the marginal probabilities of wealthy persons and of parents, respectively, as stated in $\mathcal{C}$, do not limit the probabilities of wealthy parents. Hence, when reasoning over all probabilistic models of $\mathcal{C}$, it is impossible to make a statement about the happiness of wealthy parents although it is obviously reasonable to assume that wealthy parents are happy with at least probability 0.7.

In the ME-approach, instead, it holds that the maximum entropy probability of wealthy parents being happy is $\mathcal{P}^{\mathsf{ME}}(\mathsf{Happy}|\mathsf{Wealthy} \sqcap \mathsf{Parent}) \approx 0.908$. Note that this holds independently of the domain size $|\Delta| > 0$ (see [22] for details).

However, as mentioned before, if we are only interested in consistency, then this distinction is irrelevant: A pKB has an ME-model iff it has a model at all. For this reason, it is not necessary to introduce the principle of maximum entropy and the definition of the ME-model here.

Example 3. We can now reconsider the pKB in Example 1, and see how its interpretation under aggregation semantics differs from the one under other probabilistic formalisms. For instance, the assertion $\mathsf{Happy}(\mathsf{peter})[0.8]$ does not contradict the conditional $(\mathsf{Happy} \mid \mathsf{Person})[0.5]$. Indeed, the aggregating semantics implies that, on average, people are happy with a probability of 0.5, not that every person needs to have a subjective probability of exactly 0.5 of being happy. Thus, Peter being happy with an above-average probability only means that, for other people, the average probability to be happy will be slightly below 0.5, so that the total average can be 0.5.

Similarly, this pKB is consistent with Emma being unhappy, even if all her friends, like Peter, are happy. Again, the conditional probability of people being happy if all their friends are happy quantifies over all people, so one outlier will not necessarily lead to a contradiction.

3 Checking Consistency Using Types

Types classify individuals into equivalence classes depending on the concepts they satisfy. In this paper, we extend the notion of types for $\mathcal{ALC}$ found in the DL literature (see, e.g., [3,17]) such that named individuals and their relationships with other named individuals, as stated in an ABox, are taken into account. After introducing our notion of types, we will first use it to reprove the ExpTime upper bound for consistency in classical $\mathcal{ALC}$. The constructions and results used for this purpose are important for our treatment of consistency in $\mathcal{ALC}^{\mathsf{ME}}$. Type notions that can deal with individuals have been considered before in the DL literature, but usually in the more complicated setting of DLs that are considerably more expressive than $\mathcal{ALC}$ (see, e.g., [1], where such types are considered in the context of temporal extensions of DLs). Our results for the probabilistic case crucially depend on the exact notion of types introduced in the present paper, and in particular on the model construction employed in the proof of Theorem 7 below.

Types For the sake of simplicity, we will only consider concepts using the constructors negation, conjunction, and existential restriction. Due to the equivalences $C \sqcup D \equiv \neg(\neg C \sqcap \neg D)$, $\forall r.C \equiv \neg(\exists r.\neg C)$, $\top \equiv A \sqcup \neg A$, and $\bot \equiv A \sqcap \neg A$, any concept can be transformed into an equivalent concept in this restricted form. We also assume that all double negations have been eliminated. For such a concept C, we define the set of its subconcepts as

$$\mathrm{sub}(C) = \{C\} \cup \begin{cases} \mathrm{sub}(C') & \text{if } C = \neg C' \text{ or } C = \exists r.C' \\ \mathrm{sub}(C') \cup \mathrm{sub}(D') & \text{if } C = C' \sqcap D' \end{cases}$$

Similarly, for a pKB $\mathcal{K}$ consisting of a CBox $\mathcal{C}$ and pABox $\mathcal{A}$, the set of all subconcepts is $\mathrm{sub}(\mathcal{K}) = \bigcup_{(D|C)[p]\in\mathcal{C}} \mathrm{sub}(D) \cup \mathrm{sub}(C) \cup \bigcup_{C(a)[p]\in\mathcal{A}} \mathrm{sub}(C)$. The set of subconcepts is defined in an analogous way for a classical $\mathcal{ALC}$ KB $\mathcal{K}$.

For convenience, we also want to include the negation of each concept. Thus, we define the closure of the set of subconcepts under negation as

$$\mathrm{sub}_\neg(\mathcal{K}) = \mathrm{sub}(\mathcal{K}) \cup \{\neg C \mid C \in \mathrm{sub}(\mathcal{K})\},$$

where we again assume that double negation is eliminated. In the presence of assertions, types also need to keep track of individual names and their connections. Basically, we achieve this by employing individuals names from the ABox as nominals [21] within existential restrictions. To be more precises, we use the set of existential restrictions to an individual:

$$\mathrm{EI}_\mathcal{K} = \{\exists r.a, \neg\exists r.a \mid a \in \mathrm{sig}_I(\mathcal{K}), r \in \mathrm{sig}_R(\mathcal{K})\}$$

Then, we can define a type as a set of concepts, existential restrictions to named individuals, and individual names:

Definition 4 (Type). *Given a KB $\mathcal{K}$, a type t for $\mathcal{K}$ is a subset $t \subseteq sub_\neg(\mathcal{K}) \cup sig_I(\mathcal{K}) \cup EI_\mathcal{K}$ such that*

1. *for every $\neg X \in sub_\neg(\mathcal{K}) \cup EI_\mathcal{K}$, either X or $\neg X$ belongs to t;*
2. *for every $C \sqcap D \in sub_\neg(\mathcal{K})$, we have $C \sqcap D \in t$ iff $C \in t$ and $D \in t$.*

We use types to characterize elements of an interpretation. In particular, we want to identify domain elements d of an interpretation $\mathcal{I}$ with the type that contains exactly those concepts the element is an instance of. In addition, we also need to keep track which individual name is interpreted as d, and to which individuals d is related to via a role. This motivates the following definition:

$$\begin{aligned}\tau(\mathcal{I}, d) := &\left\{C \in \mathrm{sub}_\neg(\mathcal{K}) \,\middle|\, d \in C^\mathcal{I}\right\} \cup \left\{a \in \mathrm{sig}_I(\mathcal{K}) \,\middle|\, a^\mathcal{I} = d\right\} \\ &\cup \left\{\exists r.a \,\middle|\, (d, a^\mathcal{I}) \in r^\mathcal{I}\right\} \cup \left\{\neg\exists r.a \,\middle|\, (d, a^\mathcal{I}) \notin r^\mathcal{I}\right\}\end{aligned}$$

It is easy to see that the type of an individual is indeed a type in the sense of Definition 4. Due to Definition 4, each type is compatible with the semantics of conjunction and negation. However, the satisfaction of existential restrictions depends on the presence of other types. Given a type t, an existential restriction $\exists r.X \in t$ with X being an individual name or concept, and the set of all negated existential restrictions $\{\neg\exists r.X_1, \ldots, \neg\exists r.X_k\} \subseteq t$ for role r, we say that a *type t' satisfies $\exists r.X$ in t* if $X \in t'$ and $X_i \notin t'$ for $i = 1, \ldots, k$.

Definition 5 (Consistency of a set of types). *A set of types T is* consistent *if (i) $T \neq \emptyset$, (ii) for every $t \in T$ and every $\exists r.X \in t$ there is a type $t' \in T$ that satisfies $\exists r.X$ in t, and (iii) every $a \in sig_I(\mathcal{K})$ occurs in exactly one $t \in T$.*

Condition (iii) says that, for every individual, there is exactly one type. Note that we do not require that a type contains at most one individual since we do not impose the UNA.

Consistency in classical $\mathcal{ALC}$ Before we prove complexity bounds for consistency in $\mathcal{ALC}^{\mathsf{ME}}$, we recall how to use types for classical $\mathcal{ALC}$. First, we show that consistent sets of types correspond to $\mathcal{ALC}$ interpretations. On the one hand, for every interpretation $\mathcal{I}$ we can construct a consistent set of types $\tau(\mathcal{I})$ as set of types of its domain elements: $\tau(\mathcal{I}) = \{\tau(\mathcal{I}, d) \mid d \in \Delta^{\mathcal{I}}\}$. On the other hand, given a consistent set of types T, we can build an interpretation $\mathcal{I}_T = (\Delta^{\mathcal{I}_T}, \cdot^{\mathcal{I}_T})$:

$$\begin{aligned}
\Delta^{\mathcal{I}_T} &:= T,\\
A^{\mathcal{I}_T} &:= \{t \in T \mid A \in t\},\\
r^{\mathcal{I}_T} &:= \{(t, t') \mid \exists X \in \mathrm{sub}_\neg(\mathcal{K}) \cup \mathrm{sig}_I(\mathcal{K}) : t' \text{ satisfies } \exists r.X \text{ in } t\},\\
a^{\mathcal{I}_T} &\text{ is the unique type } t \in T \text{ with } a \in t.
\end{aligned}$$

Lemma 6. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a classical $\mathcal{ALC}$ knowledge base consisting of TBox $\mathcal{T}$ and ABox $\mathcal{A}$ and T be a consistent set of types. Then, for every $t \in T$ and $C \in sub_\neg(\mathcal{K})$ we have $C \in t$ iff $t \in C^{\mathcal{I}_T}$.*

Proof. This can be proved by a simple induction on the structure of C. Here we consider only the most interesting case, which is the case where C is an existential restriction. If $C = \exists r.D$, then $\exists r.D \in t$ implies that there is $t' \in T$ such that t' satisfies $\exists r.D$ in t. This implies that $D \in t'$, and thus by induction $t' \in D^{\mathcal{I}_T}$. By construction of $\mathcal{I}_T$ we also have $(t, t') \in r^{\mathcal{I}_T}$, and thus $t \in (\exists r.D)^{\mathcal{I}_T}$.

Conversely, if $\exists r.D \notin t$, then $\neg\exists r.D \in t$. We need to show that $t \notin (\exists r.D)^{\mathcal{I}_T}$, i.e., if $(t, t') \in r^{\mathcal{I}_T}$, then $t' \notin D^{\mathcal{I}_T}$. However, $(t, t') \in r^{\mathcal{I}_T}$ implies that t' satisfies $\exists r.X$ in t for some X. Since $\neg\exists r.D \in t$, this can only be the case if $D \notin t'$. Induction now yields $t' \notin D^{\mathcal{I}_T}$ as required. □

There is a known correspondence that states that an $\mathcal{ALC}$ TBox $\mathcal{T}$ is consistent iff there exists a consistent set of types T that satisfies all GCIs, i.e., for each $C \sqsubseteq D \in \mathcal{T}$ and each $t \in T$ we have $C \in t$ implies $D \in t$ [3]. We can extend this result to KB consistency as follows:

Theorem 7. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a classical $\mathcal{ALC}$ KB. Then $\mathcal{K}$ is consistent if, and only if, there exists a consistent set of types T such that*

- *for all GCIs $C \sqsubseteq D \in \mathcal{T}$ and types $t \in T$ we have $C \in t$ implies $D \in t$;*
- *for all assertions $C(a) \in \mathcal{A}$ and types $t \in T$ with $a \in t$ we have $C \in t$; and*
- *for all assertions $r(a, b) \in \mathcal{A}$ and types $t \in T$ with $a \in t$ we have $\exists r.b \in t$.*

Proof. If $\mathcal{I}$ is a model of $\mathcal{K}$, then $\tau(\mathcal{I})$ is a consistent set of types, and it is easy to see that this set satisfies the three conditions of the theorem.

For the other direction, assume that T be a consistent set of types that satisfies the three conditions from above. We show that $\mathcal{I}_T$ is a model of $\mathcal{K}$:

- Let $C \sqsubseteq D \in \mathcal{T}$ and assume that $t \in C^{\mathcal{I}_T}$. Then Lemma 6 yields $C \in t$, which implies $D \in t$ by the first condition. Lemma 6 thus yields $t \in D^{\mathcal{I}_T}$, which shows that $\mathcal{I}_T \models C \sqsubseteq D$.

– For every $C(a) \in \mathcal{A}$, we have $C \in t$ for the unique type t that contains a. By the definition of $\mathcal{I}_T$ and Lemma 6, this implies $a^{\mathcal{I}_T} = t \in C^{\mathcal{I}_T}$, and thus $\mathcal{I}_T \models C(a)$.
– For every $r(a, b) \in \mathcal{A}$, we have $\exists r.b \in t$ for the unique type t that contains a. Since T is consistent, there is a type $t' \in T$ that satisfies $\exists r.b$ in t. Consequently, $b \in t'$ and $(t, t') \in r^{\mathcal{I}_T}$. Since $t = a^{\mathcal{I}_T}$ and $t' = b^{\mathcal{I}_T}$, this shows $\mathcal{I}_T \models r(a, b)$.

This completes the proof of the theorem. □

Based on this theorem, consistency of a classical $\mathcal{ALC}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ can be decided using type elimination as follows:

1. Construct the set T of all types $t \subseteq \text{sub}_\neg(\mathcal{K}) \cup \text{EI}_\mathcal{K}$ for $\mathcal{K}$ that do not contain individual names, and for which $C \in t$ implies $D \in t$ for all $C \sqsubseteq D \in \mathcal{T}$.
2. Consider all extensions T' of T with types $t \cup I$ with $t \in T$ and $I \subseteq \text{sig}_I(\mathcal{K})$ such that
   – each individual name $a \in \text{sig}_I(\mathcal{K})$ occurs in exactly one type $t' \in T'$;
   – for all $C(a) \in \mathcal{A}$ and $t' \in T'$, $a \in t'$ implies $C \in t'$; and
   – for all $r(a, b) \in \mathcal{A}$ and $t' \in T'$, $a \in t'$ implies $\exists r.b \in t'$.
3. For each such set T', successively remove all types from T' with unsatisfied existential restrictions until no more such types remain.
4. Return "consistent" if at least one of the sets T' obtained this way is non-empty and contains for each $a \in \text{sig}_I(\mathcal{K})$ a type t with $a \in t$.

Corollary 8. *Consistency of $\mathcal{ALC}$ KBs can be decided in ExpTime.*

Proof. We need to show that the above algorithm is sound and complete, and runs in exponential time. We sketch how to show each claim:

Soundness follows directly from Theorem 7 since it is easy to see that a set T' that leads the algorithm to answer "consistent" is a consistent set of types that satisfies the three conditions of the theorem.

For completeness, assume that $\mathcal{K}$ is consistent, and thus by Theorem 7 a consistent set S of types with the stated properties exists. This consistent set of types S must be a subset of the set T' constructed by the algorithm after step 2 for some guess of the types for each individual name. However, then step 3 will never remove any type of S from T', and thus step 4 of the algorithm will return that $\mathcal{K}$ is indeed consistent.

Regarding runtime, note that the set T constructed in the first step contains at most exponentially many types. In the second step, at most exponentially many extensions T' of T are constructed since this step basically amounts to looking at all possible ways of choosing exactly one type for each of the (linearly many) individuals, and then removing choices that do not satisfy the stated conditions. Since each of the sets T' constructed this way is of at most exponential size, type elimination applied to T' takes at most exponential time. □

Consistency in $\mathcal{ALC}^{\mathsf{ME}}$ For probabilistic $\mathcal{ALC}$, we need to consider multiple worlds, each corresponding to a classical interpretation. Additionally, the aggregating semantics takes into account how many individuals verify or falsify a conditional, even if those individuals are indistinguishable (i.e., have the same type). Thus, instead of sets of types, we need to consider multisets of types.

Formally, a multiset on a domain X is a function $M : X \to \mathbb{N}$. We denote multisets as mappings, such as $M = \{\!\{x_1 \mapsto 3, x_2 \mapsto 1\}\!\}$. We say that an element $x \in X$ occurs $M(x)$ times in M, and that it occurs in M if $M(x) > 0$. The cardinality of the multiset M is given by the sum of the number of occurrences of each element, i.e. $|M| = \sum_{x \in X} M(x)$.

As said above, we are interested in multisets of types, and in particular multisets of types with a given cardinality $k = |\Delta|$. These correspond to $\mathcal{ALC}$ interpretations with a domain of size k. We define the (multiset-)type $\tau_M(\mathcal{I})$ of an interpretation $\mathcal{I}$ as follows:

$$\tau_M(\mathcal{I})(t) := |\{d \in \Delta \mid \tau(\mathcal{I}, d) = t\}|.$$

It is easy to see that $|\tau_M(\mathcal{I})| = |\Delta|$.

Consistency of multisets of types is defined analogously to the set case: every existential restriction in every type occurring in the multiset M must be satisfied by some other type occurring in M, and every individual name must occur in exactly one type that occurs exactly once in M. We denote the set of all consistent multisets of types with cardinality k with $\mathcal{M}_{\mathcal{K},k}$.

Similarly to the classical case, we build an interpretation $\mathcal{I}_M = (\Delta^{\mathcal{I}_M}, \cdot^{\mathcal{I}_M})$ from a multiset M of types, except now we take $M(t)$ copies for each element in M, to ensure that the interpretation domain has the same cardinality as M:

$$\begin{aligned}
\Delta^{\mathcal{I}_M} &:= \{(t, i) \mid 1 \leq i \leq M(t)\} \\
A^{\mathcal{I}_M} &:= \{(t, i) \in \Delta^{\mathcal{I}_M} \mid A \in t\}, \\
r^{\mathcal{I}_M} &:= \{((t, i), (t', j)) \in \Delta^{\mathcal{I}_M} \times \Delta^{\mathcal{I}_M} \mid \\
&\quad \exists X \in \mathrm{sub}_\neg(\mathcal{K}) \cup \mathrm{sig}_I(\mathcal{K}) : t' \text{ satisfies } \exists r.X \text{ in } t\}, \\
a^{\mathcal{I}_M} &:= (t, 1) \text{ where } t \text{ is the unique type occurring in } M \text{ with } a \in t.
\end{aligned} \tag{2}$$

It is easy to show that this construction achieves the same as in the classical case (see Lemma 6):

Lemma 9. *Let $\mathcal{K} = (\mathcal{C}, \mathcal{A})$ be a pKB consisting of a CBox $\mathcal{C}$ and a pABox $\mathcal{A}$, and let $M \in \mathcal{M}_{\mathcal{K},k}$ be a consistent multiset of types. Then $|\Delta^{\mathcal{I}_M}| = k$, and for every $t \in M$, $1 \leq i \leq M(t)$, and $C \in sub_\neg(\mathcal{K})$, we have $C \in t$ iff $(t, i) \in C^{\mathcal{I}_M}$.*

In order to use this lemma to obtain a characterization of consistent pKBs, we need to take into account that, in $\mathcal{ALC}^{\mathsf{ME}}$, models are probability distributions over classical interpretations. Consequently, we need to consider probability distributions over the set of all multisets of types of a given cardinality. The aggregating semantics depends on counting instances of concepts. Thus, we need to show that counting instances can be reduced to summing up the number of occurrences of the corresponding types.

Lemma 10. *Let $\mathcal{K}$ be a pKB and $\mathcal{I}$ be an $\mathcal{ALC}$ interpretation with finite domain Δ and $\tau_M(\mathcal{I}) = M$. Then for all $(D|C)[p]$ occurring in $\mathcal{K}$ we have $|C^{\mathcal{I}}| = \sum_{t \in M \text{ s.t. } C \in t} M(t)$ and $|C^{\mathcal{I}} \cap D^{\mathcal{I}}| = \sum_{t \in M \text{ s.t. } \{C,D\} \subseteq t} M(t)$. Additionally, for any $a, b \in sig_I(\mathcal{K})$, we have $a^{\mathcal{I}} \in C^{\mathcal{I}}$ iff there is $t \in M$ with $\{a, C\} \subseteq t$, and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ iff there is $t \in M$ with $\{a, \exists r.b\} \subseteq t$.*

Proof. $M = \tau_M(\mathcal{I})$ implies that

$$\begin{aligned} |C^{\mathcal{I}} \cap D^{\mathcal{I}}| &= \left|\left\{d \in \Delta \;\middle|\; d \in C^{\mathcal{I}} \wedge d \in D^{\mathcal{I}}\right\}\right| \\ &= |\{d \in \Delta \mid \{C, D\} \subseteq \tau_M(\mathcal{I}, d)\}| \\ &= \sum_{t \in M \text{ s.t. } \{C,D\} \subseteq t} M(t). \end{aligned}$$

The same argument can be used to show $|C^{\mathcal{I}}| = \sum_{t \in M \text{ s.t. } C \in t} M(t)$.

Let $t \in \tau_M(\mathcal{I})$ be the unique type with $a \in t$. Then $a^{\mathcal{I}} \in C^{\mathcal{I}}$ iff $C \in t$ and thus $\{a, C\} \subseteq t$. If $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$, then $\exists r.b \in t = \tau(\mathcal{I}, a^{\mathcal{I}})$, and thus $\{a, \exists r.b\} \subseteq t$. Conversely, if $\exists r.b \in t = \tau(\mathcal{I}, a^{\mathcal{I}})$, then $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. □

Note that Lemma 10 implies that, for interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ with the same type $\tau_M(\mathcal{I}_1) = \tau_M(\mathcal{I}_2)$, we have $|C^{\mathcal{I}_1}| = |C^{\mathcal{I}_2}|$ and $|C^{\mathcal{I}_1} \cap D^{\mathcal{I}_1}| = |C^{\mathcal{I}_2} \cap D^{\mathcal{I}_2}|$ for all $(D|C)[p]$ occurring in $\mathcal{K}$, as well as $a^{\mathcal{I}_1} \in C^{\mathcal{I}_1}$ iff $a^{\mathcal{I}_2} \in C^{\mathcal{I}_2}$ and $(a^{\mathcal{I}_1}, b^{\mathcal{I}_1}) \in r^{\mathcal{I}_1}$ iff $(a^{\mathcal{I}_2}, b^{\mathcal{I}_2}) \in r^{\mathcal{I}_2}$. This means that the aggregating semantics cannot distinguish between interpretations with the same type. Thus, these types allow us to simplify equation (1): instead of summing over all interpretations $\mathcal{I}_{\mathcal{K},\Delta}$, we only have to consider those interpretations with different types. Based on these ideas, the following theorem characterizes consistency of pKBs in $\mathcal{ALC}^{\mathsf{ME}}$.

Theorem 11. *Let $\mathcal{K}$ be a pKB and Δ be a finite domain with $|\Delta| = k$. Then $\mathcal{K}$ is consistent if, and only if, the equation system* (3) *in Fig. 1 has a non-negative solution $\boldsymbol{p_M} \in \mathbb{R}_{\geq 0}^{\mathcal{M}_{\mathcal{K},k}}$.*

Proof. For each $M \in \mathcal{M}_{\mathcal{K},k}$, let $\mathcal{I}(M) = \{\mathcal{I} \in \mathcal{I}_{\mathcal{K},\Delta} \mid \tau_M(\mathcal{I}) = M\}$ be the set of interpretations with type M. It is easy to see that, for $M \neq M'$, we have $\mathcal{I}(M) \cap \mathcal{I}(M') = \emptyset$. Using Lemma 10 together with this fact,[5] we can translate between models of $\mathcal{K}$ and solutions of the system of equations (3) as follows.

First, assume that $\mathcal{K}$ is consistent, i.e., there exists a model $\mu : \mathcal{I}_{\mathcal{K},\Delta} \to [0, 1]$ of $\mathcal{K}$. Then it is easy to see that setting $p_M := \sum_{\mathcal{I} \in \mathcal{I}(M)} \mu(\mathcal{I})$ yields a solution of (3).

[5] More precisely, this fact is used in the identities marked with $*$ below.

$$
\begin{aligned}
\frac{\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\substack{t\in M\\ \text{s.t. } C\in t\wedge D\in t}} M(t)\cdot p_M}{\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\substack{t\in M\\ \text{s.t. } C\in t}} M(t)\cdot p_M} &= p, && \text{for } (D|C)[p]\in\mathcal{C},\\
\sum\limits_{\substack{M\in\mathcal{M}_{\mathcal{K},k}\\ \text{s.t. } \exists t\in M:\{a,C\}\subseteq t}} p_M &= p, && \text{for } C(a)[p]\in\mathcal{A},\\
\sum\limits_{\substack{M\in\mathcal{M}_{\mathcal{K},k}\\ \text{s.t. } \exists t\in M:\{a,\exists r.b\}\subseteq t}} p_M &= p, && \text{for } r(a,b)[p]\in\mathcal{A},\\
\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} p_M &= 1.
\end{aligned}
\tag{3}
$$

Fig. 1. The system of equations that characterizes consistency of pKBs in $\mathcal{ALC}^{\mathsf{ME}}$.

In fact, for $(D|C)[p]\in\mathcal{C}$ we have

$$
\begin{aligned}
\frac{\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\substack{t\in M\\ \text{s.t. } C\in t\wedge D\in t}} M(t)\cdot p_M}{\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\substack{t\in M\\ \text{s.t. } C\in t}} M(t)\cdot p_M}
&= \frac{\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\substack{t\in M\\ \text{s.t. } C\in t\wedge D\in t}} M(t)\cdot \sum\limits_{\mathcal{I}\in\mathfrak{I}_M}\mu(\mathcal{I})}{\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\substack{t\in M\\ \text{s.t. } C\in t}} M(t)\cdot \sum\limits_{\mathcal{I}\in\mathfrak{I}_M}\mu(\mathcal{I})}\\
&= \frac{\sum_{M\in\mathcal{M}_{\mathcal{K},k}} \sum_{\mathcal{I}\in\mathfrak{I}_M} \sum_{\substack{t\in M\\ C\in t\wedge D\in t}} M(t)\cdot\mu(\mathcal{I})}{\sum_{M\in\mathcal{M}_{\mathcal{K},k}} \sum_{\mathcal{I}\in\mathfrak{I}_M} \sum_{\substack{t\in M\\ C\in t}} M(t)\cdot\mu(\mathcal{I})}
\end{aligned}
$$

Using Lemma 10 we see that this sum is equal to

$$
\begin{aligned}
&= \frac{\sum_{M\in\mathcal{M}_{\mathcal{K},k}} \sum_{\mathcal{I}\in\mathfrak{I}_M} \left|C^{\mathcal{I}}\cap D^{\mathcal{I}}\right|\cdot\mu(\mathcal{I})}{\sum_{M\in\mathcal{M}_{\mathcal{K},k}} \sum_{\mathcal{I}\in\mathfrak{I}_M} |C^{\mathcal{I}}|\cdot\mu(\mathcal{I})}\\
&=^{*} \frac{\sum_{\mathcal{I}\in\mathfrak{I}_{\mathcal{K},\Delta}} \left|C^{\mathcal{I}}\cap D^{\mathcal{I}}\right|\cdot\mu(\mathcal{I})}{\sum_{\mathcal{I}\in\mathfrak{I}_{\mathcal{K},\Delta}} |C^{\mathcal{I}}|\cdot\mu(\mathcal{I})} = p.
\end{aligned}
$$

For $C(a)[p]\in\mathcal{A}$ we have

$$
\sum\limits_{\substack{M\in\mathcal{M}_{\mathcal{K},k}\\ \text{s.t. } \exists t\in M:\{a,C\}\subseteq t}} p_M = \sum\limits_{\substack{M\in\mathcal{M}_{\mathcal{K},k}\\ \text{s.t. } \exists t\in M:\{a,C\}\subseteq t}} \sum\limits_{\mathcal{I}\in\mathfrak{I}_M}\mu(\mathcal{I}) =^{*} \sum\limits_{\substack{\mathcal{I}\in\mathfrak{I}_{\mathcal{K},\Delta}\\ a^{\mathcal{I}}\in C^{\mathcal{I}}}}\mu(\mathcal{I}) = p.
$$

The assertions $r(a,b)\in\mathcal{A}$ can be treated analogously, and finally we have

$$
\sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} p_M = \sum\limits_{M\in\mathcal{M}_{\mathcal{K},k}} \sum\limits_{\mathcal{I}\in\mathfrak{I}_M}\mu(\mathcal{I}) =^{*} \sum\limits_{\mathcal{I}\in\mathfrak{I}_{\mathcal{K},\Delta}}\mu(\mathcal{I}) = 1.
$$

For the other direction, let $\boldsymbol{p_M} \in \mathbb{R}_{\geq 0}^{\mathcal{M}_{\mathcal{K},k}}$ be a solution to (3). Then, for every $M \in \mathcal{M}_{\mathcal{K},k}$, $\mathcal{I}(M)$ is not empty since $\tau_M(\mathcal{I}_M) = M$. Thus, we can choose a function $\mu : \mathcal{I}_{\mathcal{K},\Delta} \to [0,1]$ such that $\sum_{\mathcal{I}\in\mathcal{I}_M} \mu(\mathcal{I}) = p_M$ for every $M \in \mathcal{M}_{\mathcal{K},k}$, e.g.,

$$\mu(\mathcal{I}) = \frac{p_{\tau_M(\mathcal{I})}}{|\mathcal{I}_{\tau_M(\mathcal{I})}|}.$$

Then, analogously to the proof above, we can show that μ is indeed a probability distribution and satisfies equation (1). □

4 Complexity Bounds for Consistency in $\mathcal{ALC}^{\mathsf{ME}}$

In this section, we use the characterization of consistency given in Theorem 11 to determine the complexity of the consistency problem in $\mathcal{ALC}^{\mathsf{ME}}$. We will start with domain size complexity (where the complexity is measured in terms of the size of the domain Δ only), and then determine the combined complexity (measured in terms of the size of the domain and the knowledge base). In both settings, we will distinguish between unary and binary encoding of the domain size.

Domain size complexity Given a pKB $\mathcal{K}$ and a domain Δ with $|\Delta| = k$, we know that the number n of types can grow exponentially with the size of $\mathcal{K}$, i.e., $n \in \mathcal{O}(2^{|\mathcal{K}|})$. Then, the number of different multisets [19] over those n types of cardinality k is

$$|\mathcal{M}_{\mathcal{K},k}| = \left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k! \cdot (n-1)!}.$$

Interestingly, this can be simplified to both $|\mathcal{M}_{\mathcal{K},k}| = \frac{(n+k-1)(n+k-2)\cdots n}{k(k-1)\cdots 1} \in \mathcal{O}(n^k)$ and $|\mathcal{M}_{\mathcal{K},k}| = \frac{(n+k-1)(n+k-2)\cdots(k+1)}{(n-1)(n-2)\cdots 1} \in \mathcal{O}(k^n)$.

Since (3) is a linear equation system with $\mathcal{O}(|\mathcal{K}|)$ equations and $|\mathcal{M}_{\mathcal{K},k}|$ variables, and linear equation systems over the real numbers can be solved in polynomial time [5], this yields the following complexities.

Corollary 12 (Domain size complexity). *Let $\mathcal{K}$ be a fixed pKB (which is not part of the input) and Δ be a finite domain with $|\Delta| = k$. Then the consistency of $\mathcal{K}$ w.r.t. Δ can be decided in*

- *P in $|\Delta| = k$ (unary encoding),*
- *ExpTime in $\log(k)$ (binary encoding).*

This result extends an existing P-time result for domain size complexity for unary encoding given in [23] from CBoxes to the case of general probabilistic KBs also including assertional knowledge. It should be noted that the approach used in [23] to show the "in P" result also uses types, but is nevertheless quite different from the one employed here.

Combined complexity Both the number of interpretations in equation (1) and the number of multisets of types in (3) will usually grow exponentially with the size of the pKB $\mathcal{K}$, and thus will the number of variables. However, the number of linear equations in both systems will always be the number of probabilistic conditionals and probabilistic assertions plus one, i.e., it will at most grow linearly with the size of $\mathcal{K}$. We can exploit this fact using the following "sparse solution lemma" from linear programming:

Lemma 13 ([5], Theorem 9.3). *If a system of m linear equations has a non-negative solution in $\mathbb{R}$, then it has a solution with at most m variables positive.*

Thus, we can solve the consistency problem for a given pKB $\mathcal{K}$ with m conditionals and assertions (where $m \in \mathcal{O}(|\mathcal{K}|)$) and a domain Δ non-deterministically, by guessing a set $\mathcal{M}$ of $m+1$ distinct multisets $M_1, \ldots, M_{m+1}$ of types with cardinality $k = |\Delta|$, and checking whether these multisets are consistent and yield a solvable system of equations:

$$\begin{aligned}
\frac{\sum\limits_{M \in \mathcal{M}} \sum\limits_{\substack{t \in M \\ \text{s.t. } C \in t \wedge D \in t}} M(t) \cdot p_M}{\sum\limits_{M \in \mathcal{M}} \sum\limits_{\substack{t \in M \\ \text{s.t. } C \in t}} M(t) \cdot p_M} &= p, && \text{for } (D|C)[p] \in \mathcal{C}, \\
\sum_{\substack{M \in \mathcal{M} \\ \text{s.t. } \exists t \in M : \{a, C\} \subseteq t}} p_M &= p, && \text{for } C(a)[p] \in \mathcal{A}, \\
\sum_{\substack{M \in \mathcal{M} \\ \text{s.t. } \exists t \in M : \{a, \exists r.b\} \subseteq t}} p_M &= p, && \text{for } r(a,b)[p] \in \mathcal{A}, \\
\sum_{M \in \mathcal{M}} p_M &= 1.
\end{aligned} \tag{4}$$

This provides us with the following complexity results:

Corollary 14 (Combined Complexity). *Let $\mathcal{K}$ be a pKB and Δ be a finite domain with $|\Delta| = k$. Then consistency of $\mathcal{K}$ w.r.t. Δ is*

- *NP-complete in $|\mathcal{K}| + k$ (unary encoding of k),*
- *NExpTime-complete in $|\mathcal{K}| + log(k)$ (binary encoding of k).*

Proof. Guessing a multiset of size k can be done by guessing k types (of size at most quadratic in $|\mathcal{K}|$). Thus, in total guessing can be done in non-deterministic time $\mathcal{O}(m \cdot k \cdot |\mathcal{K}|^2) = \mathcal{O}(|\mathcal{K}|^3 \cdot k)$. Evaluating the corresponding linear equation system (4) (of size polynomial in $|\mathcal{K}|$) can then be done in polynomial time. The complexity upper bounds follow directly from this observation.

According to [6,18], fixed-domain reasoning in classical $\mathcal{ALC}$ is already NP-complete for unary encoding of the domain size, and NExpTime-complete for binary encoding of the domain size. There is an easy reduction from fixed-domain consistency in classical $\mathcal{ALC}$ to $\mathcal{ALC}^{\mathsf{ME}}$ consistency: Simply exchange GCIs $C \sqsubseteq$

D with conditionals $(D|C)[1]$ and add probability 1 to all assertions. It is easy to see that each model $\mathcal{I}$ of the original KB can be translated into a model of the new pKB by setting the probability of $\mathcal{I}$ to 1, and of all other interpretations to 0. Similarly, for each model of the pKB all interpretations with non-zero probability must also be models of the original KB. Thus, the original classical KB has a model with domain Δ iff the constructed pKB is consistent w.r.t. Δ, which transfers the hardness results for fixed-domain consistency in classical $\mathcal{ALC}$ to consistency in $\mathcal{ALC}^{\mathsf{ME}}$. □

Note that the results for combined complexity cannot be shown using the approach employed in [23]. There, the constructed equation system not only has exponentially many variables, but also exponentially many equations. Thus, the sparse solution lemma cannot be used to reduce the complexity.

5 Conclusion

In this paper, we have determined the complexity of the consistency problem in the probabilistic Description Logic $\mathcal{ALC}^{\mathsf{ME}}$, considering both domain size and combined complexity and distinguishing between unary and binary encoding of the domain size. Our results are based on the notion of types, but to use this notion in a setting with assertions, we had to extend it such that it also takes named individuals and their relationships into account. Basically, these results show that probabilities do not increase the complexity of the consistency problem since we obtain the same results as for fixed domain reasoning in $\mathcal{ALC}$. Note that our results can be transferred easily to a variant of $\mathcal{ALC}^{\mathsf{ME}}$ in which probabilistic conditionals are provided with interval probabilities instead of point probabilities.

In future work, we want to extend our complexity results to other reasoning tasks and to DLs other than $\mathcal{ALC}$. In [23] we have already considered drawing inferences, but have only investigated the domain size complexity. More challenging is to go from fixed domain reasoning to finite domain reasoning, i.e., checking whether there is some finite domain Δ such that the pKB is consistent w.r.t. Δ. Finally, if a pKB is consistent, then we know that it has a unique ME-model, but the complexity of computing (an approximation of) this distribution is unclear, though [23] contains some preliminary results in this direction, but again restricted to domain size complexity.

References

1. Baader, F., Borgwardt, S., Koopmann, P., Ozaki, A., Thost, V.: Metric temporal description logics with interval-rigid names. In: Proc. FroCoS'17. Springer LNCS 10483, pp. 60–76 (2017)
2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)

3. Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017).
4. Baader, F., Koopmann, P., Turhan, A.: Using ontologies to query probabilistic numerical data. In: Proc. FroCoS'17. Springer LNCS 10483, pp. 77–94 (2017)
5. Chvatal, V.: Linear Programming. W. H. Freeman (1983)
6. Gaggl, S.A., Rudolph, S., Schweizer, L.: Fixed-domain reasoning for description logics. In: Proc. ECAI'16. pp. 819–827. IOS Press (2016)
7. Grove, A., Halpern, J., Koller, D.: Random worlds and maximum entropy. J. of Artificial Intelligence Research **2**, 33–88 (1994)
8. Halpern, J.Y.: An analysis of first-order logics of probability. Artif. Intell. **46**(3), 311–350 (1990)
9. Hoehndorf, R., Schofield, P.N., Gkoutos, G.V.: The role of ontologies in biological and biomedical research: A functional perspective. Brief. Bioinform. **16**(6), 1069–1080 (2015)
10. Kern-Isberner, G.: Conditionals in Nonmonotonic Reasoning and Belief Revision. Springer LNCS 2087, (2001)
11. Kern-Isberner, G., Thimm, M.: Novel semantical approaches to relational probabilistic conditionals. In: Proc. KR'10. pp. 382–392. AAAI Press (2010)
12. Lutz, C., Schröder, L.: Probabilistic description logics for subjective uncertainty. In: Proc. KR'10. pp. 393–403. AAAI Press (2010)
13. Paris, J.B.: Common sense and maximum entropy. Synthese **117**(1), 75–93 (1999)
14. Paris, J.B.: The Uncertain Reasoner's Companion: A Mathematical Perspective. Cambridge University Press (2006)
15. Peñaloza, R., Potyka, N.: Towards statistical reasoning in description logics over finite domains. In: Proc. SUM'17. Springer LNCS 10564, pp. 280–294, (2017)
16. Pratt, V.R.: Models of program logics. In: Proc. FOCS'79. pp. 115–122. IEEE Computer Society (1979)
17. Rudolph, S., Krötzsch, M., Hitzler, P.: Type-elimination-based reasoning for the description logic SHIQbs using decision diagrams and disjunctive datalog. Logical Methods in Computer Science **8**(1) (2012)
18. Rudolph, S., Schweizer, L.: Not too big, not too Small... Complexities of fixed-domain reasoning in first-order and description logics. In: Proc. EPIA'17. Springer LNCS 10423, pp. 695–708, (2017)
19. Stanley, R.: Enumerative Combinatorics:. Cambridge Studies in Advanced Mathematics, Cambridge University Press (1997)
20. Thimm, M., G.Kern-Isberner: On probabilistic inference in relational conditional logics. Logic Journal of the IGPL **20**(5), 872–908 (2012)
21. Tobies, S.: The complexity of reasoning with cardinality restrictions and nominals in expressive description logics. J. Artif. Intell. Res. **12**, 199–217 (2000)
22. Wilhelm, M., Kern-Isberner, G., Ecke, A.: Basic independence results for maximum entropy reasoning based on relational conditionals. In: Proc. GCAI'17. EPiC Series in Computing, vol. 50, pp. 36–50. EasyChair (2017)
23. Wilhelm, M., Kern-Isberner, G., Ecke, A., Baader, F.: Counting Strategies for the Probabilistic Description Logic $\mathcal{ALC}^{\mathsf{ME}}$ Under the Principle of Maximum Entropy. In: Proc. JELIA'19. Springer LNCS 11468, pp. 434–449, (2019)