Cutting Diamonds

Temporal DLs with Probabilistic Distributions over Data

Alisa Kovtunova and Rafael Peñaloza

KRDB Research Centre, Free University of Bozen-Bolzano, Italy.
{alisa.kovtunova,rafael.penaloza}@unibz.it

Abstract. Recent work has studied a probabilistic extension of the temporal logic LTL that refines the eventuality (or diamond) constructor with a probability distribution on when will this eventuality be satisfied. In this paper, we adapt this notion to a well established temporal extension of DL-Lite, allowing the new probabilistic constructor only in the ABox assertions. We investigate the satisfiability problem of this new temporal DL over equiparametric geometric distributions.

1 Introduction

Combinations of DLs with temporal formalisms have been widely investigated since the early work of [20]; we refer the reader to [2,8,9,16] for detailed surveys of the area. Despite the different visions of the problem presented, logical theories that encode a domain of interest are always represented by factual statements. However, speaking about the future by itself can imply (probabilistic) uncertainty.

For example, insurance companies estimate the probability of the insured event in the duration period of the policy based in a number of factors; e.g., for a life insurance, they consider health condition, number of children, habits, sun radiation in home region, etc. This defines the extent of monthly payment for a customer. If one uses a classical temporal DL, one can only express that everyone dies ($\mathsf{LivingBeing} \sqsubseteq \Diamond\mathsf{Dead}$), and miss the golden goose of insurers.

There is also a large pool of proposals for probabilistic DLs, e.g., [10,17–19], that differ widely in many fundamental aspects, like the way in which probabilities are used, in the syntax, in the chosen semantics, and in the possible application. We refer to [13] for a (now slightly outdated) survey. There are two main views of probability [11], statistical and subjective. While the statistical view considers a probability distribution over a domain that specifies the probability for an individual in the domain to be randomly picked, we choose the subjective view, which specifies the probability distribution over a set of possible worlds. In our case, a world would be a possible evolution of a system; that is, a standard temporal DL interpretation. The authors of [10] argue the subjective semantics provides an appropriate modelling for probabilistic statements about individuals, e.g., the statement "there's at least 80% chance of not having an earthquake tomorrow" implies that an earthquake will either occur or will not occur. Thus, in the set of possible worlds, there are some structures in which an earthquake happens and others in which it does not. This type of uncertainty can also be called epistemic, because it regards probabilities as the degree of our belief.

This paper presents the language *TLD-Lite*, which, to the best of our knowledge, is a first probabilistic extension of temporal description logics and aims at closing the gap between probabilistic and temporal extensions of lightweight DLs. We propose a new view on one of the pillars of the linear temporal logic LTL, the eventuality (or diamond) constructor, which expresses that some property will hold at some point in the future. The specification of the diamond operator is nondeterministic and, thus, can be too rough; indeed, with the basis of actual life experience one often has an idea—albeit uncertain—about *when* the property may hold. Employing basic notions from statistical analyses, we combine an ontology of a well established temporal extension of DL-Lite proposed in [5] with temporal data specified by the geometric distribution with parameter $p \geq \frac{1}{2}$.

This logic allows us to reason over time dimension with uncertain knowledge. For example, if one builds an earthquake-resistant house, one wants to be sure an earthquake does not ruin it before building bricks and concrete blocks are properly reinforced. Also, if local hospitals take vacation on Sundays, an earthquake in the construction camp this day of the week inflicts heavier losses.

For the resulting temporal lightweight description logic with distributions, aka *TLD-Lite*, we present the formal semantics underlying the language, introduce the probabilistic formalism realised by means of the probabilistic constraint $\Diamond_\delta$, called the *distribution eventuality*, and investigate the satisfiability problem. Consistency can be checked by a deterministic algorithm using exponential space in the size of input. An open question is whether the result can be improved to match the coNExpTime upper bound obtained in [14] for the propositional temporal formula with only one instance of the distribution eventuality $\Diamond_\delta$. This new refined diamond constructor includes a discrete probability distribution δ that can be used to specify the likelihood of observing the property of interest, for the first time, at each possible point in time.

Due to space restrictions, longer proofs of Lemmas 8 and 11 are deferred to the full version [15].

2 Preliminaries

We briefly introduce the basics of probability theory and temporal extensions of description logics.

2.1 Probability Theory

We start by providing the basic notions of probability needed for this paper. For a deeper study on probabilities, we refer the interested reader to [6]. Let Ω be a set called the *sample space*. A *σ-algebra* over Ω is a class $\mathcal{F}$ of subsets of Ω that contains the empty set, and is closed under complements and under countable unions. A *probability measure* is a function $\mu : \mathcal{F} \to [0, 1]$ such that $\mu(\Omega) = 1$, and for any countable collection of pairwise disjoint sets $E_i \in \mathcal{F}$, $i \geq 1$, it holds that $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$. The probability of a set $E \in \mathcal{F}$ is $\int_{\omega \in E} \mu d\omega$, where the integration is made w.r.t. the measure μ.

A usual case is when Ω is the set $\mathbb{R}$ or all real numbers, and $\mathcal{F}$ is the standard Borel σ-algebra over $\mathbb{R}$; that is, the smallest σ-algebra containing all open intervals in $\mathbb{R}$. In this case, μ is called a *continuous* probability measure, and the integration defining the probability of a set E corresponds to the standard Riemann integration.

If Ω is a countable (or finite) set, the standard σ-algebra is formed by the power set of Ω, and a probability measure μ is uniquely determined by a function $\mu : \Omega \to [0, 1]$. Given a set $E \subseteq \Omega$, $\mu(E) = \sum_{\omega \in E} \mu(\omega)$; that is, the probability of a set is the sum of the probabilities of the elements it contains. In this case, μ is called a *discrete* probability measure. In addition, if $\mu(\omega) > 0$ for all $\omega \in \Omega$, then μ is *complete*. In contrast to our definition, in a classical Kolmogorov probability space the completeness of μ only requires it to be necessarily defined for every $\omega \in \Omega$. When Ω is the set of all natural numbers $\mathbb{N}$, we specify the distribution μ as a function $\mu : \mathbb{N} \to [0, 1]$.

A simple example of a complete discrete distribution is the *geometric* distribution. The geometric distribution $\mathsf{Geom}(p)$ with a parameter (the probability of success) $p \in (0, 1)$ is defined, for every $i \in \mathbb{N}$, by $\mu(i) = (1 - p)^{i-1}p$. This distribution describes the probability of observing the first success in a repeated trial of an experiment at time i. Returning back to the insurance example, according to [21], an attained integer age at death has the geometrical distribution if the force of mortality were constant at all ages.

2.2 Temporal DLs

Temporal *DL-Lite* logics are extensions of standard *DL-Lite* description logics introduced by [1, 7]. Similarly to [3], since we want to reason about the future, we also allow applications of the discrete unary future operators $\bigcirc$ ("in the next time point"), $\Box$ ("always in the future"), $\Diamond$ ("eventually in the future") and the binary operator $\mathcal{U}$ ("until") to basic concepts. We use the non-strict semantics for $\mathcal{U}$, $\Diamond$ and $\Box$ in the sense that their semantics includes the current moment of time.

Formally, *TLD-Lite* contains *individual names* $a_0, a_1, \ldots$, *concept names* $A_0, A_1, \ldots$, *flexible role names* $P_0, P_1, \ldots$ and *rigid role names* $G_0, G_1, \ldots$. *Roles* R, *basic concepts* B and *concepts* are defined by the grammar

$$R ::= S \mid S^-, \qquad S ::= P_i \mid G_i, \qquad B ::= \bot \mid A_i \mid \exists R,$$
$$C ::= B \mid \neg C \mid \bigcirc C \mid \Diamond C \mid \Box C \mid C \,\mathcal{U}\, C \mid C \sqcap C.$$

We denote the nesting of temporal operators as the superscription of the temporal operator from the set $\{\Box, \bigcirc\}$; that is, $\Box^0 A = \bigcirc^0 A = A$, and $\Box^{n+1} A = \Box\Box^n A$, $\bigcirc^{n+1} A = \bigcirc\bigcirc^n A$.

A temporal *concept inclusion* (CI) takes the form $C_1 \sqsubseteq C_2$, while temporal *role inclusions* (RI) are of the form $R_1 \sqsubseteq R_2$. As usual $C_1 \equiv C_2$ abbreviates $C_1 \sqsubseteq C_2$ and $C_2 \sqsubseteq C_1$. All CIs and RIs are assumed to hold globally (over the whole timeline). Note that an CI can express that a basic concept is *rigid*, i.e., interpreted in the same way at every point of time. A temporal *TBox* $\mathcal{T}$ (resp. *RBox* $\mathcal{R}$) is a finite set of temporal CIs (resp., RIs). Their union $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$ is called a temporal *ontology*. Since the non-strict operators are obviously definable in terms of the strict ones, which do not include the current moment, temporal CIs and RIs are expressible in terms of PSPACE-complete $T_{\mathcal{US}}DL\text{-}Lite^{\mathcal{N}}_{bool}$ logic [5].

A *temporal interpretation* is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}(n)})$, where $\Delta^{\mathcal{I}} \neq \emptyset$ and

$$\mathcal{I}(n) = (\Delta^{\mathcal{I}}, a_0^{\mathcal{I}}, \ldots, A_0^{\mathcal{I}(n)}, \ldots, P_0^{\mathcal{I}(n)}, \ldots, G_0^{\mathcal{I}}, \ldots,)$$

contains a standard DL interpretation for each time instant $n \in \mathbb{N}$ of the ordered set $(\mathbb{N}, <)$, that is, $a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, $A_i^{\mathcal{I}(n)} \subseteq \Delta^{\mathcal{I}}$ and $P_i^{\mathcal{I}(n)}, G_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The domain $\Delta^{\mathcal{I}}$ and the interpretations $a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ of the individual names and $G_i^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ of rigid role names are the same for all $n \in \mathbb{N}$, thus, we adopt the *constant domain assumption*. However, we do not assume the unique name, since neither functionality nor number restrictions are applied to this logic.

The DL and temporal constructs are interpreted in $\mathcal{I}(n)$ as follows:

$$\begin{aligned}
(R^-)^{\mathcal{I}(n)} &= \{(x,y) \mid (y,x) \in R^{\mathcal{I}(n)}\},\\
(\exists R)^{\mathcal{I}(n)} &= \big\{x \mid (x,y) \in R^{\mathcal{I}(n)},\ \text{for some } y\big\},\\
(\neg C)^{\mathcal{I}(n)} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}(n)},\\
(C_1 \sqcap C_2)^{\mathcal{I}(n)} &= C_1^{\mathcal{I}(n)} \cap C_2^{\mathcal{I}(n)},\\
(\Diamond C)^{\mathcal{I}(n)} &= \bigcup_{k \geq n} C^{\mathcal{I}(k)},\\
(\Box C)^{\mathcal{I}(n)} &= \bigcap\nolimits_{k \geq n} C^{\mathcal{I}(k)},\\
(\bigcirc C)^{\mathcal{I}(n)} &= C^{\mathcal{I}(n+1)},\\
(C_1 \,\mathcal{U}\, C_2)^{\mathcal{I}(n)} &= \bigcup_{k \geq n} \Big(C_2^{\mathcal{I}(k)} \cap \bigcap\nolimits_{k > l \geq n} C_1^{\mathcal{I}(l)}\Big).
\end{aligned}$$

As usual, $\bot$ is interpreted by $\emptyset$ and $\top$ by $\Delta^{\mathcal{I}}$ for concepts and by $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for roles. As mentioned before, CIs and RIs are interpreted in $\mathcal{I}$ *globally* in the sense that they hold in $\mathcal{I}$ if $C_1^{\mathcal{I}(n)} \subseteq C_2^{\mathcal{I}(n)}$ and $R_1^{\mathcal{I}(n)} \subseteq R_2^{\mathcal{I}(n)}$ hold for all $n \in \mathbb{N}$. Given an inclusion α and a temporal interpretation $\mathcal{I}$, we write $\mathcal{I} \models \alpha$ if α holds in $\mathcal{I}$.

2.3 Distributions of Data Instances over Time

In temporal variants of *DL-Lite*, instances from the ABox can be associated with temporal constructors as well. In our logic *TLD-Lite*, in addition to the standard constructors used also in the ontology, we allow a probabilistic constructor that provides a distribution of the time needed until the assertion is observed. Formally, a *TLD-Lite ABox* (or *data instance*) is a finite set $\mathcal{A}$ of atoms of the form

$$\begin{array}{ccc} \bigcirc^n A(a), & \bigcirc^n \Diamond_\delta A(a), & \bigcirc^n \neg A(a), \\ \bigcirc^n R(a,b), & \bigcirc^n \Diamond_\delta R(a,b) & \bigcirc^n \neg R(a,b), \end{array}$$

where a, b are individual names, $n \in \mathbb{N}$ and δ is a complete distribution over $\mathbb{N}$. The new constructor $\Diamond_\delta \theta(\boldsymbol{a})$, for $\theta = A$, $\boldsymbol{a} = a$ and $\theta = R$, $\boldsymbol{a} = (a,b)$, expresses that the time until the event $\theta(\boldsymbol{a})$ is first observed has distribution δ. We denote by $\mathsf{ind}(\mathcal{A})$ the set of individual names in $\mathcal{A}$. A *TLD-Lite knowledge base* (KB) $\mathcal{K}$ is a pair $(\mathcal{O}, \mathcal{A})$, where $\mathcal{O}$ is a temporal ontology and $\mathcal{A}$ a *TLD-Lite* ABox.

To render the probabilistic properties, the semantics of *TLD-Lite* is based on the *multiple-world* approach. A *TLD-Lite* interpretation is a pair $\mathcal{P} = (\mathfrak{I}, \mu)$, where $\mathfrak{I}$ is a set of temporal interpretations $\mathcal{I}$ and μ is a probability distribution over $\mathfrak{I}$. Given a set of temporal interpretations $\mathfrak{I}$, a concept name or a role θ, individual names $\boldsymbol{a}$ and $n \in \mathbb{N}$, let $\mathfrak{I}^\theta_{\boldsymbol{a},n} := \{\mathcal{I} \in \mathfrak{I} \mid \boldsymbol{a}^\mathcal{I} \in \theta^{\mathcal{I}(n)}\}$. For $\mathcal{P} = (\mathfrak{I}, \mu)$, we interpret the probabilistic construct in $\mathcal{I} \in \mathfrak{I}$ at the time point n over the set of individual names as

$$(\Diamond_\delta \theta)^{\mathcal{I}(n)} = \{\boldsymbol{a}^\mathcal{I} \mid \mu(\mathfrak{I}^\theta_{\boldsymbol{a},n+i} \setminus \bigcup_{j=0}^{i-1} \mathfrak{I}^\theta_{\boldsymbol{a},n+j}) = \delta(i) \text{ for all } i \geq 0\}. \tag{1}$$

In contrast to [14], we do not require the unique constant domain for all interpretations in the set $\mathfrak{I}$.

$\mathcal{P} = (\mathfrak{I}, \mu)$ is a *model* of $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ and write $\mathcal{P} \models \mathcal{K}$ if, for any $\mathcal{I} \in \mathfrak{I}$,

- all concept and role inclusions from $\mathcal{O}$ hold in $\mathcal{I}$, i.e. $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{O}$;
- $a^\mathcal{I} \in A^{\mathcal{I}(n)}$ for $\bigcirc^n A(a) \in \mathcal{A}$, and $(a^\mathcal{I}, b^\mathcal{I}) \in R^{\mathcal{I}(n)}$ for $\bigcirc^n R(a,b) \in \mathcal{A}$;
  $a^\mathcal{I} \notin A^{\mathcal{I}(n)}$ for $\bigcirc^n \neg A(a) \in \mathcal{A}$, and $(a^\mathcal{I}, b^\mathcal{I}) \notin R^{\mathcal{I}(n)}$ for $\bigcirc^n \neg R(a,b) \in \mathcal{A}$;
- $\boldsymbol{a}^\mathcal{I} \in (\Diamond_\delta \theta)^{\mathcal{I}(n)}$, for all $\bigcirc^n \Diamond_\delta \theta(\boldsymbol{a}) \in \mathcal{A}$.

Similarly to the standard case, a KB $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ is *consistent* if it has a model. As it is obvious from our semantics, we are using the standard open-world assumption from DLs.

There are several reasons why we allow the probabilistic operator only in ABox instances: (i) semantic: each model of $\mathcal{K}$ can differ in the anonymous part, while the definition (1) requires common object names for all temporal interpretations in $\mathfrak{I}$; (ii) computational: *DL-Lite* allows infinitely many anonymous objects; bounding the probability to the ABox objects ensures existence of a model with a finite number of terms, i.e., concept names or roles, prefixed with $\Diamond_\delta$; (iii) even leaving out the anonymous part of *TLD-Lite*, CIs and RIs can also express infinitely many times repeated events for ABox objects, e.g., $C \sqsubseteq \bigcirc^2 C$.

Remark 1. The restrictions in the syntax for avoiding uncountable models are mainly of a technical nature: describing the distribution μ over an uncountable set of temporal interpretations needs measure-theoretic notions; and verifying the existence of uncountable models requires more advanced machinery.

The interpretation $\mathcal{P} = (\mathfrak{I}, \mu)$ is *countable* if the set $\mathfrak{I}$ contains countably many temporal interpretations. In [14] it was shown that the combination $\Box\diamondsuit_\delta$ can only be satisfied by an uncountable interpretation. *TLD-Lite* KBs allow the constructor $\diamondsuit$ only in ABox instances, and these occurrences need only a finite number of time points to be satisfied. Hence, *TLD-Lite* has the countable-model property.

Theorem 2. *If a TLD-Lite KB $\mathcal{K}$ is satisfiable, it has a countable model.*

Proof. Let $\mathcal{P} = (\mathfrak{I}, \mu)$ be a model of $\mathcal{K}$, and $\boldsymbol{d}$ be the number of all $\diamondsuit$-data instances appearing in $\mathcal{A}$. If $\boldsymbol{d} = 0$, then $\mathcal{K}$ does not contain any *TLD-Lite* instances, and, for every $\mathcal{I} \in \mathfrak{I}$, an interpretation $\mathcal{P}_\mathcal{I} = (\{\mathcal{I}\}, \mu_\mathcal{I})$, where $\mu_\mathcal{I}(\{\mathcal{I}\}) = 1$, is a model of $\mathcal{K}$. Otherwise, by semantics, for every instance $\bigcirc^{n_i}\diamondsuit_{\delta_i}\theta_i(\boldsymbol{a}_i) \in \mathcal{A}$, where $1 \leq i \leq \boldsymbol{d}$, we have

$$\mu(\mathfrak{I}^{\theta_i}_{\boldsymbol{a}_i, n_i+k} \setminus \bigcup_{j=0}^{k-1} \mathfrak{I}^{\theta_i}_{\boldsymbol{a}_i, n_i+j}) = \delta_i(k), \tag{2}$$

for any $k \geq 0$. Since the domain of probability functions, as a σ-algebra, is closed under countable *intersections*, the joint function $\mu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}})$ is also defined for the *TLD-Lite* model $\mathcal{P} = (\mathfrak{I}, \mu)$, where

$$\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}} = \bigcap_{i=1}^{\boldsymbol{d}} \left(\mathfrak{I}^{\theta_i}_{\boldsymbol{a}_i, n_i+k_i} \setminus \bigcup_{j=0}^{k_i-1} \mathfrak{I}^{\theta_i}_{\boldsymbol{a}_i, n_i+j} \right). \tag{3}$$

Note that $\sum_{\{k_1,\ldots,k_{\boldsymbol{d}}\} \in \mathbb{N}^{\boldsymbol{d}}} \mu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}}) = 1$.

Now from a (possibly) uncountable $\mathcal{P}$ we build a new countable *TLD-Lite* interpretation $\mathcal{P}' = (\mathfrak{I}', \nu)$ by assigning an appropriate weight to a representative interpretation $\mathcal{I}$ of a set $\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}}$ for every $k_1, \ldots, k_{\boldsymbol{d}} \geq 0$.

Initially we assume $\mathfrak{I}' = \emptyset$. For all $k_1, \ldots, k_{\boldsymbol{d}} \geq 0$, we consider a subset $\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}} \subseteq \mathfrak{I}$ defined by (3). If $\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}} = \emptyset$ and $\mu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}}) = 0$, then we assign $\nu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}'}) = \nu(\emptyset) = 0$. Otherwise, we pick any interpretation $\mathcal{I} \in \mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}}$ as a representative and set $\mathfrak{I}' = \mathfrak{I}' \cup \mathcal{I}$ with $\nu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}'}) = \nu(\mathcal{I}) = \mu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}})$. In the general case, the last equality, $\nu(\mathcal{I})$, can be equal to 0.

As $\boldsymbol{d}$ is finite and at each step of the procedure we add at most one interpretation, $\mathcal{P}'$ is countable. In order to show $\mathcal{P}' \models \mathcal{K}$, we notice that, for any axiom $\alpha \in \mathcal{O}$, we have $\mathcal{I} \models \alpha$, for any $\mathcal{I} \in \mathfrak{I}$. Since $\mathfrak{I}' \subseteq \mathfrak{I}$, we have the statement, $\mathcal{P}' \models \alpha$. A similar argument can be applied to the $\diamondsuit$-free ABox assertions.

Consider an instance $\bigcirc^n\diamondsuit_\delta\theta(\boldsymbol{a}) \in \mathcal{A}$. By construction of $\mathcal{P}' = (\mathfrak{I}', \nu)$, for any $k \geq 0$, it holds that

$$\begin{aligned}\nu(\{\mathcal{I} \in \mathfrak{I}' | \boldsymbol{a}^\mathcal{I} \in \theta^{\mathcal{I}(n+k)}\} \setminus \bigcup_{j=0}^{k-1} \{\mathcal{I} \in \mathfrak{I}' | \boldsymbol{a}^\mathcal{I} \in \theta^{\mathcal{I}(n+j)}\}) &= \sum_{\{k_1,\ldots,k_{\boldsymbol{d}}\} \setminus \{k\} \in \mathbb{N}^{\boldsymbol{d}-1}} \nu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}'}) \\ &= \sum_{\{k_1,\ldots,k_{\boldsymbol{d}}\} \setminus \{k\} \in \mathbb{N}^{\boldsymbol{d}-1}} \mu(\mathcal{J}^{k_1,\ldots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}}) = \delta(k).\end{aligned}$$

By semantics, $\mathcal{P}' \models \bigcirc^n \blacklozenge_\delta \theta(\boldsymbol{a})$. Thus, the *TLD-Lite* interpretation $\mathcal{P}' = (\mathfrak{I}', \nu)$ is a countable model of $\mathcal{K}$. □

3 Deciding Satisfiability

We now focus on the problem of deciding whether a given *TLD-Lite* KB $\mathcal{K}$ is satisfiable. The semantics of the $\blacklozenge$-operator given by (1) in a combination with a nondeterministic temporal operator can give interesting results.

Example 3. Consider the KB $\mathcal{K} = (\{B \equiv \Diamond A\}, \{\blacklozenge_\delta B(a)\})$ for a complete distribution δ. $\mathcal{K}$ is unsatisfiable, since, for any temporal interpretation $\mathcal{I}$, the statement $\mathcal{I}, 1 \models \Diamond A$ implies $\mathcal{I}, 0 \models \Diamond A$, which contradicts the semantics of $\blacklozenge$-operator (1) with $i = 1$, by which, for any model $(\mathfrak{I}, \mu)$ of the KB,

$$\mu(\mathfrak{I}^{\Diamond A}_{a,1} \setminus \mathfrak{I}^{\Diamond A}_{a,0}) = \delta(1) > 0.$$

Namely, one should not be able to say that there is any positive probability of satisfying $\neg\Diamond A(a)$ in a time point 0, and $\Diamond A(a)$ in any later time $m > 0$.

3.1 Multidimensional Matrix

To represent a model $\mathcal{P} = (\mathfrak{I}, \mu)$ of a *TLD-Lite* KB $(\mathcal{O}, \mathcal{A})$ with $\boldsymbol{d}$ instances of $\blacklozenge_p$-data assertions in the ABox, we introduce a $\boldsymbol{d}$-dimension infinite matrix $M_\mathcal{P}$ with elements from the range of μ such that $M_\mathcal{P}(k_1, \dots, k_{\boldsymbol{d}}) = \mu(\mathcal{J}^{k_1,\dots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}})$, where $\mathcal{J}^{k_1,\dots,k_{\boldsymbol{d}}}_{\mathcal{A},\mathfrak{I}}$ is defined by (3). Notice that the mapping from a model to a matrix is surjective: similarly to the proof of Theorem 2, the mapping merges equivalent (from the point of view of $\blacklozenge$ unravelling) temporal interpretations together. Stepping away from exact *TLD-Lite* interpretations, by the same argument as in Theorem 2, we show the following result.

Theorem 4. *A TLD-Lite KB $\mathcal{K}$ is satisfiable iff there is a matrix M of elements from the interval $[0, 1]$ such that*

- *for any $\{k_1, \dots, k_{\boldsymbol{d}}\} \in \mathbb{N}^{\boldsymbol{d}}$, if $M(k_1, \dots, k_{\boldsymbol{d}}) > 0$ then the temporal KB $(\mathcal{O}, \mathcal{A}_{k_1,\dots,k_{\boldsymbol{d}}})$, where $\mathcal{A}_{k_1,\dots,k_{\boldsymbol{d}}}$ is the $\blacklozenge$-free ABox*

$$\bigcup_{i=1}^{\boldsymbol{d}} \{ \bigcup_{j=0}^{k_i - 1} \bigcirc^{n_i + j} \neg\theta_i(\boldsymbol{a}_i) \cup \bigcirc^{n_i + k_i} \theta_i(\boldsymbol{a}_i)\} \cup \mathcal{A} \setminus \{\bigcup_{i=1}^{\boldsymbol{d}} \bigcirc^{n_i} \blacklozenge_{\delta_i} \theta_i(\boldsymbol{a}_i)\}, \qquad (4)$$

  is satisfiable; and
- *for any $1 \le i \le \boldsymbol{d}$,*

$$\sum_{\{k_1,\dots,k_{\boldsymbol{d}}\} \setminus \{k_i\} \in \mathbb{N}^{\boldsymbol{d}-1}} M(k_1, \dots, k_{\boldsymbol{d}}) = \delta_i(k_i) \qquad (5)$$

We consider matrix entries starting with zero; i.e., $M(0,\dots,0)$ is the first element of the matrix M.

Theorem 4 allows us to avoid providing explicitly a model for a satisfiable *TLD-Lite* KB, since the matrix ensures it existence. But it does not provide an efficient solution or even an algorithm for the satisfiability problem, since it requires an infinite matrix. However, as each non-zero element corresponds to a classical temporal KB, we use properties of the probabilistic distribution and establish a periodical property of matrix entries to bound the size of the matrix.

3.2 Bernoulli Processes

So far we have introduced the *TLD-Lite* KB in general terms. In the following we focus on the special case where all used distributions describe a Bernoulli process; that is, we consider only the geometric distribution, $\mathsf{Geom}(p)$ with $0 < p < 1$; notice that this distribution is complete. To simplify the notation, we will simply write $\diamondsuit_p$ for this distribution. For example, the set $\{\diamondsuit_{\frac{1}{2}} H(a), \bigcirc\diamondsuit_{\frac{1}{2}} T(a)\}$ describes two experiments observing a repeated flip of the coin a, where H means that the coin landed heads, and $T \equiv \neg H$ that it landed tails.

The geometric distribution $\mathsf{Geom}(p)$ has some valuable to us properties:

1. for any $j > i \geq 0$, we have $\mathsf{Geom}(p)(i) > \mathsf{Geom}(p)(j)$; and
2. for any $p > \frac{1}{2}$ and $i \geq 0$, $\mathsf{Geom}(p)(i) > \sum_{j>i} \mathsf{Geom}(p)(j)$. If $p = \frac{1}{2}$, then this inequality becomes an equality.

We also assume that, for all ABox instances, the parameter of the geometric distribution p is unique, $\frac{1}{2} \leq p < 1$. To simplify presentation, particularly matrix-wise, we consider only the case of $\boldsymbol{d} = 2$. For an arbitrary finite $\boldsymbol{d}$, the reasoning we provide below is the same, but requires the use of more cumbersome notation.

We start with important properties of $\mathsf{Geom}(p)$ matrices for $\frac{1}{2} \leq p < 1$.

Lemma 5. *For a satisfiable TLD-Lite KB $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ with two $\diamondsuit_p$-instances of $\mathsf{Geom}(p)$, $\frac{1}{2} \leq p < 1$ in the ABox, and any matrix M satisfying Conditions* (4) *and* (5) *of $\mathcal{K}$, we have*

1. *if $p > \frac{1}{2}$, then, for any $k \in \mathbb{N}$, the set*

$$L(k) = \{M(0,k), M(1,k), \dots, M(k,k), M(k,k-1), \dots M(k,0)\} \qquad (6)$$

contains at least one non-zero element,

2. *if $p = \frac{1}{2}$, then there exists at most one $k \in \mathbb{N}$ such that all elements $L(k)$ are zeroes.*

Proof. By Property 2 of Geom and the fact that, for any matrix M of $\mathcal{K}$ satisfying (4) and (5), the elements $M(k,l)$ and $M(l,k)$, for all $l > k$, are bounded with $\mathsf{Geom}(p)(l)$, the statement of item 1 is trivial.

Item 2 follows from the fact $\mathsf{Geom}(\frac{1}{2})(i) = \sum_{j>i} \mathsf{Geom}(\frac{1}{2})(j)$ for all $i \in \mathbb{N}$. Since $\sum_{l\in\mathbb{N}} M(k,l) = \frac{1}{2^{k+1}}$, if $L(k)$ contains only zeros for some k, i.e., $\sum_{0\leq l\leq k} M(k,l) = 0$, then $M(k,k+1) = \frac{1}{2^{k+2}}$, $M(k,k+2) = \frac{1}{2^{k+3}}$, etc. Thus, for all $m > k$ the set $L(m)$ contains at least two positive elements. □

The following example confirms that satisfiability of a KB depends on the chosen parameter p.

Example 6. The *TLD-Lite* KB $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ with $\mathcal{A} = \{\Diamond_p H(a), \Diamond_p T(a)\}$ and $\mathcal{O} = \{T \equiv \neg H\}$ is satisfiable if $p = \frac{1}{2}$. The matrix in this case has the form

$$M = \begin{bmatrix} \times & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \cdots \\ \frac{1}{4} & \times & & & \\ \frac{1}{8} & & \times & & \\ \frac{1}{16} & & & \times & \\ \cdots & & & & \cdots \end{bmatrix},$$

where the positions $\times$ correspond to unsatisfiable temporal KBs (4), and the matrix entries are trivially equal to 0.

However, the same *TLD-Lite* KB with $p > \frac{1}{2}$ does not have any model, since $M(0,0)$ is unsatisfiable and the value of $\mathsf{Geom}(p)(0) > \sum_{j>0} \mathsf{Geom}(p)(j)$ cannot be spread on the rest part of the matrix.

With these basic properties we can develop an (infinite) iterative process of building a matrix for a given *TLD-Lite* KB $\mathcal{K}$. A finite $(\ell+1) \times (\ell+1)$ matrix M_ℓ is called *partial*, for $\ell \in \mathbb{N}$, if

- for any $k_1, k_2 \in [0, \dots, \ell]$, if $M(k_1, k_2) > 0$ then the KB $(\mathcal{O}, \mathcal{A}_{k_1,k_2})$, where $\mathcal{A}_{k_1,k_2}$ is defined by (4), is satisfiable; and
- for any $k \in [0, \dots, \ell]$, $\sum_{0 \leq \{k_1,k_2\} \setminus \{k\} \leq \ell} M(k_1, k_2) = \mathsf{Geom}(p)(k)$.

Clearly, if we can prove there is a partial matrix M_ℓ for all $\ell \in \mathbb{N}$, we can conclude that *TLD-Lite* $\mathcal{K}$ is satisfiable.

Definition 7. *A pair of matrix entries $M(i,\ell), M(\ell,j)$, for $i, j \leq \ell$, is called* chained *if there is an odd chain of elements,*

$$\{M(i,\ell), M(i,k_1), M(m_1,k_1), M(m_1,k_2), \dots, M(m_h,j), M(\ell,j)\}, \qquad (7)$$

where $m_c, k_c < \ell$, for all $1 \leq c \leq h \in \mathbb{N}$, such that, for every element of this chain $M(s,t)$, the temporal KB $(\mathcal{O}, \mathcal{A}_{s,t})$ is consistent, and every even element is in the chain, e.g., $M(i,k_1), M(m_1,k_2), \dots, M(m_h,j)$, is chained. Trivially, the diagonal element $M(\ell,\ell)$ is chained with itself, if $(\mathcal{O}, \mathcal{A}_{\ell,\ell})$ is consistent.

Now we are ready to prove that a pair of chained elements for every $0 \leq k \leq \ell$ ensures the finite matrix M_ℓ is partial.

Lemma 8. *Any TLD-Lite KB $\mathcal{K} = (\mathcal{O}, \mathcal{A})$ with two $\Diamond_p$-instances of $\mathsf{Geom}(p)$, $\frac{1}{2} \leq p < 1$, i.e., $\bigcirc^{n_1} \Diamond_p \theta_1(\boldsymbol{a}_1), \bigcirc^{n_2} \Diamond_p \theta_2(\boldsymbol{a}_2) \in \mathcal{A}$ is satisfiable iff there is a matrix M such that either*

1. *for any $k \in \mathbb{N}$, there is a chained pair in $L(k)$; or*
2. *if $p = \frac{1}{2}$ and there is $k \in \mathbb{N}$ with no chained elements in $L(k)$, then a submatrix M_{k-1} is a partial matrix and $M(i,k) = M(k,i) = \frac{1}{2^{i+1}}$ for all $i > k$.*

The matrix building process is shown in the following example.

Example 9. Consider a *TLD-Lite* ABox $\{\Diamond_p H(a), \bigcirc\Diamond_p T(a)\}$, where $p = \frac{1}{2}$, with a TBox $\{T \equiv \neg H\}$. The matrix building process, according to the proof of Lemma 8, runs as follows:

$$M_0 = \begin{bmatrix}\frac{1}{2}\end{bmatrix}, \quad M_1 = \begin{bmatrix}\frac{1}{2} & 0 \\ \times & \frac{1}{4}\end{bmatrix}, \quad M_2 = \begin{bmatrix}\frac{3}{8} & 0 & \frac{1}{8} \\ \times & \frac{1}{4} & 0 \\ \frac{1}{8} & \times & \times\end{bmatrix}, \quad M_3 = \begin{bmatrix}\frac{5}{16} & 0 & \frac{1}{8} & \frac{1}{16} \\ \times & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & \times & \times & \times \\ \frac{1}{16} & \times & \times & \times\end{bmatrix}, \dots$$

The sign $\times$ denotes matrix entries of unsatisfiable temporal KBs with the corresponding ABoxes of the form (4), and these matrix entries are trivially equal to 0. Iterating this process to infinity, the element $M(0,0)$, as the head of chained pairs, stays positive as $(\frac{1}{2} - \sum_{n\geq 3} \frac{1}{2^n}) = \frac{1}{4}$. Thus, (4) and (5) hold and the *TLD-Lite* KB is consistent.

It is worth noting that the condition $\frac{1}{2} \leq p < 1$ is crucial for the building process in Lemma 8.

Example 10. Now we let p from Example 9 be $\frac{1}{5}$. Despite the KBs for matrix entries remaining the same, and, for all $L(k)$, $k \geq 2$, we have a pair of chained elements, this *TLD-Lite* KB is unsatisfiable.

In the next subsection we demonstrate how to finitise the process in Lemma 8.

3.3 Periodical Properties

One can notice that, for a KB $\mathcal{K}$, there are constants $s = s(|\mathcal{K}|)$ and $p = p(|\mathcal{K}|)$ such that, starting from some $\ell > s$ time point, the formulas corresponding to the elements from $L(\ell)$ are equisatisfiable with some elements from $L(\ell+p)$ and, moreover, we show the following result.

Lemma 11. *For any satisfiable TLD-Lite KB $\mathcal{K}$ over two $\Diamond_p$-data instances, $\frac{1}{2} \leq p < 1$, and any matrix M satisfying* (4) *and* (5)*, there exist two integers $s, p < 2^{\mathcal{O}(|\mathcal{K}|)}$ such that, for any $\ell \geq s$ and any pair of chained elements $M(i,\ell)$ and $M(\ell,j)$, $i,j \leq \ell$, their translations $M(i',\ell+p)$ and $M(\ell+p,j')$ are also chained, where*

$$i' = \begin{cases} i, & \text{if } i < s, \\ i+p, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad j' = \begin{cases} j, & \text{if } j < s, \\ j+p, & \text{otherwise.} \end{cases} \tag{8}$$

Proof. The idea is based on the polynomially big (in the size of $\mathcal{K}$) translation from [4,5] of temporal $\Diamond_p$-free KBs into equisatisfiable LTL formulas. Then, we can use the same reasoning as for the periodic property of an LTL formula. □

Theorem 12. *Satisfiability of TLD-Lite KBs with* $\mathsf{Geom}(p)$, $\frac{1}{2} \leq p < 1$ *can be decided in* ExpSpace.

Proof. To check if the conditions of Lemma 8(1) hold, we "guess" the positions of unsatisfiable temporal KBs in the finite matrix M_{s+p}, where $s + p < 2^{f(\mathcal{K})+1}$ and $f(\mathcal{K})$ is a polynomial function. Then, for every $k < s + p$, among $L(k)$ we choose a pair of chained entries. If it is possible, Lemma 11 ensures that the partial matrix M_{s+p} can be extended to infinity.

If $p = \frac{1}{2}$ and there is a number $k \in \mathbb{N}$ with no chained elements in $L(k)$ as in Lemma 8(2), then $k < s + p$. Otherwise, if $k \geq s + p$, it contradicts Lemma 11 which guarantees existence chained elements in $L(k)$, if they exist for all $L(i)$, $i < s + p$. Thus, we need to find a partial matrix M_{k-1}, which is in NExpSpace as in the case $\frac{1}{2} < p < 1$, and then check the satisfiability of two exponentially big LTL formulas with an only one $\blacklozenge_p$-instance. A coNExpTime oracle for verifying this has been presented in [14].

Finally, in both cases, by Savitch's theorem, the satisfiability problem belongs to ExpSpace. □

The restriction to the geometric distribution is not essential for most results. The main theorems hold for any (even not complete) distribution δ with the property $\delta(i) > \sum_{j>i} \delta(j)$, for all $i, j \in dom(\delta)$. However, the case $\delta(i) = \sum_{j>i} \delta(j)$ relies on the procedure from [14] which exploits complete distribution properties.

4 Conclusions

We have proposed a probabilistic temporal DL that is derived from a temporal *DL-Lite* by specifying an exact distribution for a concept or a role in the data to be observed. We have also provided a first but substantial analysis of the complexity of reasoning in this logic, considering the standard reasoning task of KB consistency. The importance of our formalism arises from the fact that temporal observations of events can usually be predicted with a probabilistic distribution over time; e.g., through an analysis of historical data.

This work is a first step towards a full formalism of uncertain temporal evolution of events, based on DLs. Our work extends previous results [14] developed for LTL, which can be seen as a special case of *TLD-Lite* where only one individual name exists. This paper provides new results, where more than one occurrence of the distribution eventuality $\blacklozenge_\delta$ may be observed in a model.

Following the footsteps of [12], an interesting direction for future research is to consider query answering under temporal ontologies in data-centric applications with uncertain temporal data. Along with a possible extension of temporal ontologies with interval-valued probabilistic constraints, for future work we also want to obtain effective methods for computing probabilities of different events and answer different types of probabilistic queries. Another possible line for research is the computation of the *expected* (essentially, the average) time required until a desired property is observed, as it was previously done in [14].

References

1. Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The DL-Lite family and relations. J. Artif. Intell. Res. (JAIR) 36, 1–69 (2009), http://dx.doi.org/10.1613/jair.2820
2. Artale, A., Kontchakov, R., Kovtunova, A., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: Ontology-mediated query answering over temporal data: A survey (invited talk). In: 24th International Symposium on Temporal Representation and Reasoning, TIME 2017, October 16-18, 2017, Mons, Belgium. pp. 1:1–1:37 (2017), https://doi.org/10.4230/LIPIcs.TIME.2017.1
3. Artale, A., Kontchakov, R., Lutz, C., Wolter, F., Zakharyaschev, M.: Temporalising tractable description logics. In: 14th International Symposium on Temporal Representation and Reasoning (TIME 2007), 28-30 June 2007, Alicante, Spain. pp. 11–22. IEEE Computer Society (2007), http://dx.doi.org/10.1109/TIME.2007.62
4. Artale, A., Kontchakov, R., Ryzhikov, V., Zakharyaschev, M.: DL-Lite with temporalised concepts, rigid axioms and roles. In: Frontiers of Combining Systems, 7th International Symposium, FroCoS 2009, Trento, Italy, September 16-18, 2009. Proceedings. pp. 133–148 (2009)
5. Artale, A., Kontchakov, R., Ryzhikov, V., Zakharyaschev, M.: A cookbook for temporal conceptual data modelling with description logics. ACM Trans. Comput. Log. 15(3), 25:1–25:50 (2014), http://doi.acm.org/10.1145/2629565
6. Billingsley, P.: Probability and Measure. John Wiley and Sons, third edn. (1995)
7. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The DL-Lite family. J. Autom. Reasoning 39(3), 385–429 (2007), http://dx.doi.org/10.1007/s10817-007-9078-x
8. Demri, S., Goranko, V., Lange, M.: Temporal Logics in Computer Science. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press (2016), http://www.cambridge.org/core_title/gb/434611
9. Gabbay, D.M., Kurucz, A., Wolter, F., Zakharyaschev, M.: Many-Dimensional Modal Logics: Theory and Applications. Elsevier North Holland (2003)
10. Gutiérrez-Basulto, V., Jung, J.C., Lutz, C., Schröder, L.: Probabilistic description logics for subjective uncertainty. J. Artif. Intell. Res. 58, 1–66 (2017), https://doi.org/10.1613/jair.5222
11. Halpern, J.Y.: An analysis of first-order logics of probability. Artif. Intell. 46(3), 311–350 (1990), https://doi.org/10.1016/0004-3702(90)90019-V
12. Jung, J.C., Lutz, C.: Ontology-based access to probabilistic data. In: Informal Proceedings of the 26th International Workshop on Description Logics, Ulm, Germany, July 23 - 26, 2013. pp. 258–270 (2013), http://ceur-ws.org/Vol-1014/paper_70.pdf
13. Klinov, P.: Practical Reasoning in Probabilistic Description Logic. Ph.D. thesis, The University of Manchester, Manchester, UK (2011)
14. Kovtunova, A., Peñaloza, R.: Cutting diamonds: A temporal logic with probabilistic distributions. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, October 30-November 2, 2018 (2018), accepted
15. Kovtunova, A., Peñaloza, R.: Cutting diamonds: Temporal DLs with probabilistic distributions over data. In: ArXiv e-prints. vol. 1810.01516[cs.LO] (2018), http://arxiv.org/abs/1810.01516

16. Lutz, C., Wolter, F., Zakharyaschev, M.: Temporal description logics: A survey. In: Demri, S., Jensen, C.S. (eds.) 15th International Symposium on Temporal Representation and Reasoning, TIME 2008, Université du Québec à Montréal, Canada, 16-18 June 2008. pp. 3–14. IEEE Computer Society (2008), http://dx.doi.org/10.1109/TIME.2008.14
17. Peñaloza, R., Potyka, N.: Towards statistical reasoning in description logics over finite domains. In: Scalable Uncertainty Management - 11th International Conference, SUM 2017, Granada, Spain, October 4-6, 2017, Proceedings. pp. 280–294 (2017), https://doi.org/10.1007/978-3-319-67582-4_20
18. Riguzzi, F., Bellodi, E., Lamma, E., Zese, R.: Probabilistic description logics under the distribution semantics. Semantic Web 6(5), 477–501 (2015), https://doi.org/10.3233/SW-140154
19. Sazonau, V., Sattler, U.: Tbox reasoning in the probabilistic description logic shiqp. In: Proceedings of the 28th International Workshop on Description Logics, Athens,Greece, June 7-10, 2015. (2015), http://ceur-ws.org/Vol-1350/paper-61.pdf
20. Schmiedel, A.: Temporal terminological logic. In: Shrobe, H.E., Dietterich, T.G., Swartout, W.R. (eds.) Proceedings of the 8th National Conference on Artificial Intelligence. Boston, Massachusetts, July 29 - August 3, 1990, 2 Volumes. pp. 640–645. AAAI Press / The MIT Press (1990), http://www.aaai.org/Library/AAAI/1990/aaai90-096.php
21. Slud, E.V.: Actuarial mathematics and life-table statistics (2006), Lecture Notes