A Probabilistic Temporal Logic with Frequency Operators and Its Model Checking

Takashi Tomita Shigeki Hagihara Naoki Yonezaki

Dept. of Computer Science,
Graduate School of Information Science and Engineering,
Tokyo Institute of Technology
{tomita, hagihara, yonezaki}@fmx.cs.titech.ac.jp

Probabilistic Computation Tree Logic (PCTL) and Continuous Stochastic Logic (CSL) are often used to describe specifications of probabilistic properties for discrete time and continuous time, respectively. In PCTL and CSL, the possibility of executions satisfying some temporal properties can be quantitatively represented by the probabilistic extension of the path quantifiers in their basic Computation Tree Logic (CTL), however, path formulae of them are expressed via the same operators in CTL. For this reason, both of them cannot represent formulae with quantitative temporal properties, such as those of the form "some properties hold to more than 80% of time points (in a certain bounded interval) on the path." In this paper, we introduce a new temporal operator which expressed the notion of frequency of events, and define probabilistic frequency temporal logic (PFTL) based on CTL*. As a result, we can easily represent the temporal properties of behavior in probabilistic systems. However, it is difficult to develop a model checker for the full PFTL, due to rich expressiveness. Accordingly, we develop a model-checking algorithm for the CTL-like fragment of PFTL against finite-state Markov chains, and an approximate model-checking algorithm for the bounded Linear Temporal Logic (LTL) -like fragment of PFTL against countable-state Markov chains.

1 Introduction

To analyze probabilistic systems, probabilistic model checking is often used. In probabilistic model checking, the inputs are a probabilistic model and a probabilistic property described in a specification language, and the output is whether or not the model satisfies the property. Probabilistic Computation Tree Logic [8, 10] (PCTL) and Continuous Stochastic Logic [2, 3, 10] (CSL) are often used to describe specifications of probabilistic properties. PCTL and CSL are probabilistic extensions of Computation Tree Logic [6] (CTL) for discrete-time and continuous-time, respectively. In PCTL and CSL, the probabilistic path quantifier **P** is introduced in place of the universal path quantifier **A** (for all paths, ...) and the existential path quantifier **E** (there exists a path such that ...). As a result, we can quantitatively represent the possibility of executions satisfying temporal properties of interest. However, PCTL and CSL can only describe path formulae with temporal operators of the form "some properties hold in the next state" via the next-operator **X**; of the form "some properties eventually hold" via the eventually-operator **F** (or $\Diamond$); of the form "some properties always hold" via the always-operator **G** (or $\Box$); and of the form "some properties hold at a certain time point and other properties hold until that point" via the until-operator **U**. Thus, "property φ holds to more than 80% of time points (in the interval $[0, 10]$) on the path" cannot be represented in PCTL or CSL. To capture similar quantitative properties of an above example, CSL additionally has the steady-state operator **S** [3, 10], and there are also extensions of PCTL and CSL with reward (or cost) structure [10]. Even though, the steady-state operator **S** can only capture

F. Yu and C. Wang: The 13th International Workshop on
Verification of Infinite State Systems (INFINITY'11)
EPTCS 73, 2011, pp. 79–93, doi:10.4204/EPTCS.73.9

the expected steady-state probability of being states satisfying properties of interest, and PCTL/CSL with rewards can only express the properties of the expected value of cumulated reward associated with states or transitions.

To capture temporal properties of this kind, it is necessary to employ the integral of the duration of states, as in Duration Calculus [14] (DC). In DC, the above property is explicitly described by $\int_0^{10} \varphi(t)\mathrm{d}t \geq 8$. In this paper, we describe such properties using the concept of frequency and introduce probabilistic frequency temporal logic (PFTL) based on CTL* [6], for discrete-time/continuous-time. To this logic, we add the (conditional) frequency operator $\mathbf{Q}$. Using the frequency operator $\mathbf{Q}$, we describe the above path property by $\mathbf{Q}^{\leq 10}_{>0.8}\varphi$ in PFTL. PFTL has rich expressiveness, and hence it is difficult to develop a model checker for the full logic (see Section 4). However, we develop a numerical model-checking algorithm for the CTL-like fragment of PFTL against finite-state Markov chains (MCs), and a statistical model-checking algorithm for the bounded Linear Temporal Logic [6] (LTL) -like fragment of PFTL against infinite-state MCs. The outline of the numerical algorithm for the CTL-like fragment is similar to that of PCTL and CSL [8, 3, 10]. We compute transient and steady-state probabilities and reachability via matrix operations. The difference is that our technique requires the number of states satisfying the formulae of interest to be counted in terms of frequency. On the other hand, the statistical algorithm is an approximate one, based on statistical inference, and hence there are errors (although the significance level can be set according to our needs). However, we anticipate that it will provide useful information in many cases. We estimate whether or not "an input MC satisfies an input formula" using the sequential probability ratio test [11] (SPRT), as in [12] for CSL.

The remainder of this paper is organized as follows. In Section 2, we give the definitions of discrete-time/continuous-time MCs, and describe their probabilistic behavior. In Section 3, we define the syntax and semantics of PFTL and discuss the expressiveness of PFTL. In Section 4, we present the numerical model-checking algorithm for the CTL-like fragment of PFTL against finite-state MCs, and the statistical model-checking algorithm for the bounded LTL-like fragment of PFTL against infinite-state MCs. Our conclusions are stated in Section 5.

2 Markov chains

In this section, we present the definitions of discrete-time/continuous-time MCs and describe their probabilistic behavior. We fix a set *AP* of atomic propositions that expresses the properties of interest.

Definition 1. *A (labeled) discrete-time Markov chain (DTMC) $\mathscr{D}$ is a tuple $(S,\bar{s},P,L)$ such that: S is a countable set of states; $\bar{s}\in S$ is an initial state; $P : S^2 \to [0,1]$ is a transition probability matrix satisfying the condition that $\sum_{s'\in S} P(s,s') = 1$ and $\{s'|P(s,s') > 0\}$ is finite for all s; $L : S \to 2^{AP}$ is a labeling function that assigns to each state the set of valid atomic propositions in the state.*

$P(s,s')$ denotes the probability of a one-step transition from s to s'. An execution (or discrete-time path) of a DTMC $\mathscr{D}$ is represented by an infinite sequence of states $\omega = s_0 s_1 \ldots$, where $\forall i.P(s_i,s_{i+1}) > 0$ and $\Omega^{\mathscr{D}}_s$ is the set of all paths starting from state s in $\mathscr{D}$. For a path $\omega = s_0 s_1 \ldots$, we denote the i-th state s_i by $\omega(i)$ and the i-th suffix $s_i s_{i+1} \ldots$ by ω^i. Let $C^{\mathscr{D}}_{s_0}(s_0 \ldots s_n)$ be a cylinder set $\{\omega \in \Omega^{\mathscr{D}}_{s_0} | \forall i \leq n.\omega(i) = s_i\}$, and let $\Sigma_{\Omega^{\mathscr{D}}_{s_0}}$ be the smallest σ-algebra containing all the cylinder sets $C^{\mathscr{D}}_{s_0}(s_0,\ldots,s_n)$ in $\Omega^{\mathscr{D}}_{s_0}$. The probability measure $Pr^{\mathscr{D}}_{s_0}$ on the measurable space $(\Omega^{\mathscr{D}}_{s_0}, \Sigma_{\Omega^{\mathscr{D}}_{s_0}})$ is uniquely defined as follows:

$$Pr^{\mathscr{D}}_{s_0}(C^{\mathscr{D}}_{s_0}(s_0 \ldots s_n)) = \prod_{i=1}^{n} P(s_{i-1},s_i).$$

Definition 2. *A (labeled) continuous-time Markov chain (CTMC) $\mathscr{C}$ is a tuple $(S,\bar{s},Q,L)$ such that: S is a countable set of states; $\bar{s}\in S$ is an initial state; $Q:S^2\to\mathbb{R}$ is an infinitesimal generator matrix satisfying the condition that $\sum_{s'\in S\setminus\{s\}}Q(s,s')=-Q(s,s)$, $Q(s,s')\geq 0$ if $s\neq s'$ and $\{s'|Q(s,s')>0\}$ is finite for all s; $L:S\to 2^{AP}$ is a labeling function that assigns to each state the set of valid atomic propositions in the state.*

$Q(s,s')$ is the rate of a one-step transition from s to s' if $s\neq s'$. Otherwise, $-Q(s,s)$ is the exit rate from s and the spent time in s is exponentially distributed with parameter $-Q(s,s)$. An execution (or continuous-time path) of a CTMC $\mathscr{C}$ is represented by an infinite alternating sequence $\omega=s_0t_0s_1t_1\ldots$ or a finite and non-empty sequence $\omega=s_0t_0\ldots s_n\infty$, where $s_i\in S$ and $t_i\in\mathbb{R}_{>0}$ (this value represents the time spent in s_i) for all $i\geq 0$. $\Omega_s^{\mathscr{C}}$ is the set of all paths starting from state s of $\mathscr{C}$. For a path $\omega=s_0t_0s_1t_1\ldots(s_n\infty)$, we denote the i-th state s_i by $\omega(i)$, the i-th spent time t_i by $time(\omega,i)$ and the suffix $s_it_i's_{i+1}t_{i+1}\ldots$ after time point t by ω^t, where $i=\min\{i'|\sum_{j=0}^{i'}t_j>t\}$ and $t_i'=\sum_{j=0}^{i}t_j-t$. A path ω is called an infinite time-length path if $\sum_{i=0}^{\infty}time(\omega,i)=\infty$ (therefore, an infinite number of transitions do not occur in any bounded intervals of $\mathbb{R}_{\geq 0}$ on the path). For an interval I in $\mathbb{R}_{\geq 0}$, let $C_{s_0}^{\mathscr{C}}(s_0,I_0,s_1,I_1,\cdots,I_{n-1},s_n)$ be a continuous-time cylinder set $\{\omega\in\Omega_{s_0}^{\mathscr{C}}(s_0)|\omega(i)=s_i\wedge time(\omega,i)\in I_i\}$, and let $\Sigma_{\Omega_{s_0}^{\mathscr{C}}}$ be the smallest σ-algebra that contains all cylinder sets $C_{s_0}^{\mathscr{C}}(s_0,I_0,s_1,I_1,\cdots,I_{n-1},s_n)$ in $\Omega_{s_0}^{\mathscr{C}}$. The probability measure $Pr_{s_0}^{\mathscr{C}}$ on the measurable space $(\Omega_{s_0}^{\mathscr{C}},\Sigma_{\Omega_{s_0}^{\mathscr{C}}})$ is uniquely defined as follows:

$$Pr_{s_0}^{\mathscr{C}}(C_{s_0}^{\mathscr{C}}(s_0,I_0,s_1,I_1,\cdots,I_{n-1},s_n))=\prod_{i=0}^{n-1}\frac{Q(s_i,s_{i+1})}{-Q(s_i,s_i)}\cdot\int_{I_i}-Q(s_i,s_i)\cdot e^{Q(s_i,s_i)t}\mathrm{d}t.$$

We assume that CTMCs in this paper are not explosive, that is, almost all paths of them are infinite time-length.

In numerical computations for CTMCs, transient probabilities are tractable and the uniformization method is a standard technique for computing transient probabilities of CTMCs.

Definition 3. *For a CTMC $\mathscr{C}=(S,\bar{s},Q,L)$ such that $\sup\{-Q(s,s)|s\in S\}$ is finite, a uniformized DTMC $unif_\lambda(\mathscr{C})$ is $(S,\bar{s},\mathbf{I}+Q/\lambda,L)$, where λ is a uniformization rate greater than or equal to $\sup\{-Q(s,s)|s\in S\}$ and $\mathbf{I}$ is the unit matrix.*

If each transition time in $\mathrm{unif}_\lambda(\mathscr{C})$ is exponentially distributed with parameter λ, the behavior of $\mathrm{unif}_\lambda(\mathscr{C})$ is equivalent to that of $\mathscr{C}$ in a sense. For a uniformized DTMC $\mathrm{unif}_\lambda(\mathscr{C})=(S,\bar{s},P,L)$, the transient probability matrix $\Pi_k^{\mathscr{C}}$ ($\Pi_k^{\mathscr{C}}(s,s')$ is the probability of being in state s', k time-units after the current state s in $\mathscr{C}$) is computed as follows:

$$\Pi_k^{\mathscr{C}}=\sum_{n=0}^{\infty}\rho(n;\lambda k)\cdot P^n \text{ where } \rho(n;\lambda k) \text{ is the Poisson distribution } e^{-\lambda k}(\lambda k)^n/n!.$$

In the numerical computation, this infinite sum can be truncated. The truncation points can be determined by Fox-Glynn algorithm [7], which gives an $\mathscr{O}(\lambda k)$-size upper bound.

3 Probabilistic frequency temporal logic

In this section, we define the syntax and semantics of PFTL for discrete-time/continuous-time. We discuss the expressiveness of PFTL in Section 3.2.

3.1 Syntax and semantics

Definition 4. *Probabilistic Frequency Temporal Logic (PFTL) is defined as follows:*

$$\begin{array}{lll} \textit{state formula } \varphi & ::= & a \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \mathbf{P}_{\sim p}[\psi] \\ \textit{path formula } \psi & ::= & \varphi \mid \neg\psi \mid \psi_1 \wedge \psi_2 \mid \mathbf{X}\psi \mid \psi_1 \mathbf{U}^I \psi_2 \mid \mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle \end{array}$$

where $a \in AP$, $\sim, \bowtie \in \{<, >, \leq, \geq\}$, $p, q \in [0,1]$ and I is an interval of $\mathbb{N}$ for discrete time (or of $\mathbb{R}_{\geq 0}$ for continuous time).

Intuitively speaking, $\mathbf{P}_{\sim p}[\psi]$ means that the occurrence probability of paths starting from the given state and satisfying ψ obeys the bound $\sim p$; $\mathbf{X}\psi$ means that the suffix after the next state on the path satisfies ψ; $\psi_1 \mathbf{U}^I \psi_2$ means that ψ_2 holds at a certain time point in the interval I on the path and ψ_1 holds until that point is reached; $\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle$ means that the conditional frequency of time points satisfying ψ_1 under the condition ψ_2 in the interval I on the path obeys the bound $\bowtie q$. We allow the following abbreviations:

$$\begin{array}{rcl} \varphi_1 \vee \varphi_2 & \equiv & \neg(\neg\varphi_1 \wedge \neg\varphi_2) \\ \texttt{true} & \equiv & \varphi \vee \neg\varphi \\ \varphi_1 \rightarrow \varphi_2 & \equiv & \neg\varphi_1 \vee \varphi_2 \\ \mathbf{Q}^I_{\bowtie q}\psi & \equiv & \mathbf{Q}^I_{\bowtie q}\langle\psi|\texttt{true}\rangle \\ \mathbf{F}^I\psi & \equiv & \texttt{true}\mathbf{U}^I\psi \\ \mathbf{G}^I\psi & \equiv & \neg\mathbf{F}^I\neg\psi \end{array}$$

In the sequel, we often omit the time bound I if $I = [0,\infty)$ and denote a time bound $\{i | i \bowtie k\}$ by $\bowtie k$ and a time bound $\{j | j \pm i \in I\}$ by $I \mp i$.

We now describe the semantics for DTMCs. The frequency in a finite interval is simply defined as the ratio of the number of time points satisfying subformulae in the interval. For an unbounded $\mathbf{Q}$ formula, we write a semantics (called *limit semantics*) in terms of the limit superior and limit inferior of the global frequency on the path. In general, the occurrence frequency of states in a path may not converge. However, if the MC is finite, we can regard it as simply the limit of the global frequency of the path, because of the convergence property of the limit distribution of a finite-state MC.

Definition 5. *Let the DTMC $\mathscr{D} = (S, \bar{s}, P, L)$. For a state $s \in S$, a discrete-time path ω, a state formula φ and a path formula ψ, the satisfaction relation $\models$ is defined as follows:*

$$\begin{array}{rcl} \mathscr{D}, s \models a & \Leftrightarrow & a \in L(s) \\ \mathscr{D}, s \models \neg\varphi & \Leftrightarrow & \mathscr{D}, s \not\models \varphi \\ \mathscr{D}, s \models \varphi_1 \wedge \varphi_2 & \Leftrightarrow & \mathscr{D}, s \models \varphi_1 \textit{ and } \mathscr{D}, s \models \varphi_2 \\ \mathscr{D}, s \models \mathbf{P}_{\sim p}[\psi] & \Leftrightarrow & Pr^{\mathscr{D}}_s(\{\omega \in \Omega^{\mathscr{D}}_s | \mathscr{D}, \omega \models \psi\}) \sim p \\ \mathscr{D}, \omega \models \varphi & \Leftrightarrow & \mathscr{D}, \omega(0) \models \varphi \\ \mathscr{D}, \omega \models \neg\psi & \Leftrightarrow & \mathscr{D}, \omega \not\models \psi \\ \mathscr{D}, \omega \models \psi_1 \wedge \psi_2 & \Leftrightarrow & \mathscr{D}, \omega \models \psi_1 \textit{ and } \mathscr{D}, \omega \models \psi_2 \\ \mathscr{D}, \omega \models \mathbf{X}\psi & \Leftrightarrow & \mathscr{D}, \omega^1 \models \psi \\ \mathscr{D}, \omega \models \psi_1 \mathbf{U}^I \psi_2 & \Leftrightarrow & \exists i \in I.(\mathscr{D}, \omega^i \models \psi_2 \textit{ and } \forall j < i.\mathscr{D}, \omega^j \models \psi_1) \end{array}$$

$$\mathscr{D},\omega \models \mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle \quad \Leftrightarrow \quad \begin{cases} \texttt{true} & \text{if } \{i \in I|\mathscr{D},\omega^i \models \psi_2\} = \emptyset, \\ \dfrac{|\{i \in I|\mathscr{D},\omega^i \models \psi_1 \wedge \psi_2\}|}{|\{i \in I|\mathscr{D},\omega^i \models \psi_2\}|} \bowtie q & \text{if } \sup I \in \mathbb{N}, \\ \limsup\limits_{k\to\infty} \dfrac{|\{i \in \mathbb{N}_{\leq k} \cap I|\mathscr{D},\omega^i \models \psi_1 \wedge \psi_2\}|}{|\{i \in \mathbb{N}_{\leq k} \cap I|\mathscr{D},\omega^i \models \psi_2\}|} \bowtie q & \text{if } \sup I = \infty \text{ and } \bowtie \in \{<,\leq\}, \\ \liminf\limits_{k\to\infty} \dfrac{|\{i \in \mathbb{N}_{\leq k} \cap I|\mathscr{D},\omega^i \models \psi_1 \wedge \psi_2\}|}{|\{i \in \mathbb{N}_{\leq k} \cap I|\mathscr{D},\omega^i \models \psi_2\}|} \bowtie q & \text{otherwise.} \end{cases}$$

We define a semantics for CTMCs as follows.

Definition 6. *Let the CTMC $\mathscr{C} = (S,\bar{s},Q,L)$. For a state $s \in S$, a continuous-time path ω with infinite time-length, a state formula φ and a path formula ψ, the satisfaction relation $\models$ is defined as follows:*

$$\begin{aligned} \mathscr{C},s \models a &\Leftrightarrow a \in L(s) \\ \mathscr{C},s \models \neg\varphi &\Leftrightarrow \mathscr{C},s \not\models \varphi \\ \mathscr{C},s \models \varphi_1 \wedge \varphi_2 &\Leftrightarrow \mathscr{C},s \models \varphi_1 \textit{ and } \mathscr{C},s \models \varphi_2 \\ \mathscr{C},s \models \mathbf{P}_{\sim p}[\psi] &\Leftrightarrow Pr^{\mathscr{C}}_s(\{\omega \in \Omega^{\mathscr{C}}_s|\mathscr{C},\omega \models \psi\}) \sim p \\ \mathscr{C},\omega \models \varphi &\Leftrightarrow \mathscr{C},\omega(0) \models \varphi \\ \mathscr{C},\omega \models \neg\psi &\Leftrightarrow \mathscr{C},\omega \not\models \psi \\ \mathscr{C},\omega \models \psi_1 \wedge \psi_2 &\Leftrightarrow \mathscr{C},\omega \models \psi_1 \textit{ and } \mathscr{C},\omega \models \psi_2 \\ \mathscr{C},\omega \models \mathbf{X}\psi &\Leftrightarrow time(\omega,0) \in \mathbb{R}_{>0} \textit{ and } \mathscr{C},\omega^{time(\omega,0)} \models \psi \\ \mathscr{C},\omega \models \psi_1 \mathbf{U}^I \psi_2 &\Leftrightarrow \exists t \in I.(\mathscr{C},\omega^t \models \psi_2 \textit{ and } \forall t' \in (0,t).\mathscr{C},\omega^{t'} \models \psi_1) \end{aligned}$$

$$\mathscr{C},\omega \models \mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle \quad \Leftrightarrow \quad \begin{cases} \texttt{true} & \text{if } \{t \in I|\mathscr{C},\omega^t \models \psi_2\} = \emptyset, \\ f^I_{\langle\psi_1|\psi_2\rangle}(0) \bowtie q & \text{if } \sup I \in \mathbb{R}, \\ \limsup\limits_{k\to\infty} f^{I\cap[0,k]}_{\langle\psi_1|\psi_2\rangle}(0) \bowtie q & \text{if } \sup I = \infty \textit{ and } \bowtie \in \{<,\leq\}, \\ \liminf\limits_{k\to\infty} f^{I\cap[0,k]}_{\langle\psi_1|\psi_2\rangle}(0) \bowtie q & \text{otherwise.} \end{cases}$$

where $f^I_{\langle\psi_1|\psi_2\rangle}(t)$ is the frequency of time points satisfying ψ_1 under ψ_2 in the interval $I+t$, for the Lebesgue measure $\mathscr{L}$, given by:

$$f^I_{\langle\psi_1|\psi_2\rangle}(t) = \begin{cases} \dfrac{|\{t' \in I+t|\mathscr{C},\omega^{t'} \models \psi_1 \wedge \psi_2\}|}{|\{t' \in I+t|\mathscr{C},\omega^{t'} \models \psi_2\}|} & \text{if } \sup I \neq \infty \text{ and } \{t' \in I+t|\mathscr{C},\omega^{t'} \models \psi_2\} \neq \emptyset \text{ and} \\ & \quad \mathscr{L}(\{t' \in I+t|\mathscr{C},\omega^{t'} \models \psi_2\}) = 0, \\ \dfrac{\mathscr{L}(\{t \in I+t|\mathscr{C},\omega^{t'} \models \psi_1 \wedge \psi_2\})}{\mathscr{L}(\{t' \in I+t|\mathscr{C},\omega^{t'} \models \psi_2\})} & \text{if } \sup I \neq \infty \text{ and } \mathscr{L}(\{t' \in I+t|\mathscr{C},\omega^{t'} \models \psi_2\}) > 0, \\ \texttt{undefined} & \text{otherwise.} \end{cases}$$

In continuous time, we must consider two cases: the number of time points satisfying subformulae in a finite interval is either only finite or continuously infinite. For finite time points, the frequency is defined in a manner similar to the discrete time. For continuously infinite time points, the frequency is defined as the ratio of the Lebesgue measure of the set of time points satisfying subformulae. By the following proposition (the proof is omitted from this paper), the set of time points satisfying subformulae is Lebesgue measurable. It is not necessary to consider the case in which there exists a countably infinite number of time points satisfying subformulae in a finite interval.

Proposition 7. *For a CTMC $\mathscr{C}$, a path formula ψ, a bound interval I and a continuous-time path ω with infinite time-length, the set $\{t \in I | \mathscr{C}, \omega^t \models \psi\}$ of time points satisfying ψ in I for ω can be expressed as a finite union of intervals.*

Note . For an unbounded formula $\mathbf{Q}^I$ ($\sup I = \infty$), we can define alternative semantics as follows:

$$\mathscr{D}, \omega \models \mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle \quad \Leftrightarrow \quad \exists i \in \mathbb{N}.\forall j > i.\frac{|\{j' \in I \cap [0,j] | \mathscr{D}, \omega^{j'} \models \psi_1 \wedge \psi_2\}|}{|\{j' \in I \cap [0,j] | \mathscr{D}, \omega^{j'} \models \psi_2\}|} \bowtie q$$

$$\mathscr{C}, \omega \models \mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle \quad \Leftrightarrow \quad \exists t \in \mathbb{R}.\forall t' > t.\frac{\mathscr{L}(\{t'' \in I \cap [0,t'] | \mathscr{C}, \omega^{t''} \models \psi_1 \wedge \psi_2\})}{\mathscr{L}(\{t'' \in I \cap [0,t'] | \mathscr{C}, \omega^{t''} \models \psi_2\})} \bowtie q$$

The above semantics (called *stable semantics*) would be tractable for analysis using automata-based methods, because it is captured by the co-Büchi condition. However, in the present paper, we use the limit semantics because it facilitates numerical model-checking.

3.2 Expressiveness

PFTL can flexibly express properties of paths via the frequency operator $\mathbf{Q}$. We present some examples and note the expressiveness of PFTL.

- $\mathbf{Q}_{>0}\psi$: the global frequency of time points satisfying ψ on a path is greater than 0.
  - This formula is not equivalent to $\mathbf{GF}\psi$ representing "ψ is satisfied infinitely often on the path," because the global frequency on the path may converge to 0 even if ψ is satisfied infinitely often.
- $\mathbf{Q}^{[0,20]}_{>0.8}x = 10$: more than 80% of the time points in $[0,20]$ satisfy the proposition $x = 10$.
  - For probabilistic systems, states are often associated with numerical values as in MCs with rewards. This formula is different than both $\mathbf{G}^{[0,20]}x = 10$ and $\mathbf{G}^{[0,20]}8 \leq x \leq 12$. To capture behavior of a probabilistic system, we can write flexible expressions in PFTL.
- $\mathbf{P}_{=1}[\mathbf{Q}_{\bowtie q}\varphi]$: the global frequency of the time points satisfying φ obeys the bound $\bowtie q$ for almost all paths.
  - This formula is equivalent to the CSL formula $\mathbf{S}_{\bowtie q}[\varphi]$ if the given MC is irreducible (that is, it is possible to reach any state from any state). Otherwise, the $\mathbf{S}$ formula means that the expected value of the global frequency obeys the bound $\bowtie q$.
- $\mathbf{Q}_{>0.9}\langle\psi|\varphi\rangle$: more than 90% of time points satisfying φ on the path also satisfy ψ.
  - If we assume probabilistic fairness, this formula is similar to a path formula $\mathbf{G}(\varphi \to \mathbf{P}_{>0.9}[\psi])$ that means the probabilistic branching property $\mathbf{P}_{>0.9}[\psi]$ holds at all states satisfying φ on the path. Furthermore, a conditional frequency (in a sense, it can be interpreted as a conditional probability) between path formulae on a path can be expressed via the $\mathbf{Q}$ operator without path quantifications.
- $\neg\mathbf{Q}_{>0.1}\varphi \wedge \neg\mathbf{Q}_{<0.9}\varphi$: the frequency of time points satisfying φ becomes less than 0.1 and also greater than 0.9 infinitely often.
  - Roughly speaking, this formula describes a situation in which intervals where φ frequently holds and intervals where φ frequently does not hold appear alternately and become progressively longer in both the limit semantics and the stable semantics. However, it is not a property of the languages defined by ω-Kleene closure, e.g., ω-regular and ω-context free

languages. In the discrete-time stable semantics, for natural numbers q_1 and q_2 such that $0 < q_1 < q_2$, a single frequency formula $\mathbf{Q}_{>q_1/q_2}\langle\varphi_1|\varphi_2\rangle$ is a property of ω-context free. The class of ω-context free language is equivalent to the class of language accepted by ω-pushdown automata [5], and we can construct an ω-pushdown automaton which stores the value $n\cdot(q_2-q_1)-m\cdot q_1$ in the stack, where n and m are the numbers of visiting states satisfying $\varphi_1\wedge\varphi_2$ and $\neg\varphi_1\wedge\varphi_2$, respectively. Then $\mathbf{Q}_{>q_1/q_2}\langle\varphi_1|\varphi_2\rangle$ can be represented by the automaton with the co-Büchi condition "the stored value $n\cdot(q_2-q_1)-m\cdot q_1$ is non-positive at finitely many time-points."

4 Model checking

In this section, we introduce model-checking algorithms. The inputs are a DTMC $\mathscr{D}=(S,\bar{s},P,L)$ (or a CTMC $\mathscr{C}=(S,\bar{s},Q,L)$) and a formula φ. The output is whether or not $\mathscr{D},\bar{s}\models\varphi$ (or $\mathscr{C},\bar{s}\models\varphi$). Unfortunately, it is difficult to develop a model-checking algorithm for PFTL because of its high expressiveness of path formulae, which describes linear time properties. In the model checking of linear time logic against a (non-) probabilistic system, an automata-based approach is generally used. In this type of approach, a (non-) deterministic ω-automaton equivalent to (the negation of) the input path formula ψ is first constructed. Then the synchronized product system of the input system and the constructed ω-automata is analyzed. Because the synchronized product system captures the intersection of the behavior of the input system and that (out) of ψ, we can reduce the model checking to the reachability (or emptiness) problem. However, the language class of the path formulae in PFTL and its equivalent automata class are open in both the limit semantics and the stable semantics. The limit semantics does not primarily match existing automata, which do not have an concept of convergence.

The stable semantics also results in intractable problems. For discrete time, the language class of the path formulae in PFTL is at least a superclass of ω-regular, and includes ω-context free and non-ω-regular languages and also non- ω-Kleene closure languages. Hence, for model checking using an automata-based approach, we require a new type of automata to capture frequency. Such automata must have stack-like features, because they must be able to recognize some ω-context free languages. For continuous time, the set of the path formulae in PFTL is a superset of Metric Temporal Logic [9] (MTL), which is a real-time extension of LTL, in an interval-based semantics. Timed automata [1] are widely used as real-time automata, however, there exist MTL formulae (including bounded formulae [4]) for which there is no equivalent timed automata. We conjecture that the required automata to satisfy some frequency conditions in continuous time is some kind of extended timed automata and that it is also impossible to construct such a timed automaton to capture a property represented by a path formula in PFTL. It may be possible to obtain a synchronized product directly, or it may not be necessary to employ an automata-based approach, but there is currently no available method for model checking of an LTL-like fragment of PFTL.

Accordingly, we develop separate model-checking algorithms for two fragments of PFTL. The first is a strict numerical model-checking procedure for the CTL-like fragment of PFTL against finite-state MCs (Section 4.1). The second is a statistics-based approximation model-checking for the bounded LTL-like fragment of PFTL against infinite-state MCs (Section 4.2). The model checking for the bounded LTL-like fragment of PFTL against infinite-state DTMCs can be reduced to the model checking for LTL against finite-states DTMCs. Because, the number of reachable states from the initial state for bounded steps is finite and a bounded $\mathbf{Q}^I$ formula can be translated into a nested $\mathbf{X}$ formula. However, the translated formula has $\mathscr{O}(\inf I+2^{|I|})$-size and hence it is difficult to check exactly for the bounded

LTL-like fragment of PFTL in the viewpoint of complexity. In a statistics-based approach, we sample prefix sequences of paths of an input MC by probabilistic simulation and statistically determine whether or not an input model satisfies an input formula by using the sample. Thus, we can apply statistical methods to model checking for an infinite-state MC, because it is easy to generate prefix sequences of paths of an MC even if the MC has infinitely many states. For finite-state MCs, numerical techniques are often limited by the state explosion problem. Statistical methods can also overcome also this issue.

4.1 Model checking of the CTL-like fragment of PFTL

In this section, we introduce a model checking algorithm for the CTL-like fragment of PFTL:

$$\begin{array}{lll} \text{state formula } \varphi & ::= & a \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \mathbf{P}_{\sim p}[\psi] \\ \text{path formula } \psi & ::= & \mathbf{X}\varphi \mid \varphi_1 \mathbf{U}^I \varphi_2 \mid \mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle \end{array}$$

against finite-state MCs.

The outline of the algorithm is similar to that for PCTL/CSL [8, 3, 10]. We recursively compute a set $Sat(\varphi)$ of states satisfying φ from sets of states satisfying subformulae of φ.

$$\begin{array}{rcl} Sat(a) & = & \{s \in S | s \in L(a)\} \\ Sat(\neg\varphi) & = & S \setminus Sat(\varphi) \\ Sat(\varphi_1 \wedge \varphi_2) & = & Sat(\varphi_1) \cap Sat(\varphi_2) \\ Sat(P_{\sim p}[\psi]) & = & \{s \in S | Prob^{\mathscr{D}/\mathscr{C}}(\psi)(s) \sim p\} \end{array}$$

where $Prob^{\mathscr{D}/\mathscr{C}}(\psi)$ is the vector of occurrence probabilities of paths satisfying ψ for each starting state in discrete-time/continuous-time.

In this paper, we indicate only how to compute $Prob^{\mathscr{D}/\mathscr{C}}(\psi)$ for the case $\psi = \mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle$. For $\psi = \mathbf{X}\varphi$ or $\psi = \varphi_1 \mathbf{U}^I \varphi_2$, we can use procedure for PCTL/CSL. We assume that $Sat(\varphi_1)$ and $Sat(\varphi_2)$ are already computed, and that an interval I is either of the form $[k,k']$ ($k' \neq \infty$) or $[k,\infty)$, because all intervals of $\mathbb{N}$ can be represented in one of these forms, and $Prob^{\mathscr{C}}(\mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle)$ is equal to $Prob^{\mathscr{C}}(\mathbf{Q}^{[\inf I, \sup I]}_{\bowtie q}\langle\varphi_1|\varphi_2\rangle)$ for I such that $\inf I \neq \sup I$.

$\mathbf{P}_{\sim p}[\mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle]$ for DTMCs. If $I = [k,k']$ ($k' \in \mathbb{N}$), we compute the occurrence probability of a path by counting the number of states satisfying φ_2 and $\varphi_1 \wedge \varphi_2$ in the interval $[k,k']$ on the path. Let the vector $v^h_{j,i}(s)$ be the occurrence probability of a path starting from s, visiting states in $Sat(\varphi_2)$ i times and states in $Sat(\varphi_1) \cap Sat(\varphi_2)$ j times, within h steps:

$$v^0_{j,i}(s) = \begin{cases} 1 & \text{if } (i=0, j=0 \text{ and } s \notin Sat(\varphi_2)) \text{ or} \\ & \quad (i=1, j=0 \text{ and } s \in Sat(\varphi_2) \setminus Sat(\varphi_1)) \text{ or} \\ & \quad (i=1, j=1 \text{ and } s \in Sat(\varphi_1) \cap Sat(\varphi_2)), \\ 0 & \text{otherwise.} \end{cases}$$

$$v^h_{j,i}(s) = \begin{cases} P(s,-) \cdot v^{h-1}_{j-1,i-1} & \text{if } s \in Sat(\varphi_1) \cap Sat(\varphi_2), \\ P(s,-) \cdot v^{h-1}_{j,i-1} & \text{if } s \in Sat(\varphi_2) \setminus Sat(\varphi_1), \\ P(s,-) \cdot v^{h-1}_{j,i} & \text{otherwise.} \end{cases}$$

where $P(n,-)$ is the n-th row vector of the transition probability matrix P.

Here $Prob^{\mathscr{D}}(\mathbf{Q}_{\bowtie q}^{[k,k']}\langle\varphi_1|\varphi_2\rangle)$ is the probability of satisfying $\mathbf{Q}_{\bowtie q}^{[0,k-k']}\langle\varphi_1|\varphi_2\rangle$ after k-steps, and the k-step transition probability matrix is computed by P^k. Hence, we can compute $Prob^{\mathscr{D}}(\mathbf{Q}_{\bowtie q}^{[k,k']}\langle\varphi_1|\varphi_2\rangle)$ as follows:

$$Prob^{\mathscr{D}}(\mathbf{Q}_{\bowtie q}^{[k,k']}\langle\varphi_1|\varphi_2\rangle) = P^k \cdot \sum_{i=0}^{k'-k+1} \sum_{\substack{j\le i \\ i>0 \Rightarrow j \bowtie i\cdot q}} v_{j,i}^{k'-k}$$

If $I=[k,\infty)$ (unbounded), the basic idea is similar to the algorithm for the **S** operator in CSL [3, 10]. Each path in a finite-state MC has to reach one bottom strongly connected component (BSCC) B (B is a strongly connected component, and $s\in B$ cannot reach $s'\notin B$). BSCCs are computed by Tarjan's Algorithm ($\mathscr{O}(|S|)$). For a non-BSCC A and BSCCs $B_1,\ldots,B_n$, let the matrix P be reordered as

$$\begin{bmatrix} P_A & P_{AB_1} & \cdots & \cdots & P_{AB_n} \\ \mathbf{0} & P_{B_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & P_{B_n} \end{bmatrix}$$

where P_{XY} is a partial transition matrix from X to Y of the transition probability matrix P with $X,Y\subseteq S$ (we denote a partial transition matrix P_{XX} by P_X).

For each path reaching the BSCC B_i, the occurrence frequency of states converges to the limit distribution π_{B_i} depending on B_i. π_{B_i} can be computed as the unique solution of the system of linear equations:

$$\pi_{B_i}P_{B_i} = \pi_{B_i} \text{ and } \pi_{B_i}\vec{\mathbf{1}} = 1$$

where $\vec{\mathbf{1}}$ is the vector in which all elements are 1.

If the BSCC B_i has $s\in Sat(\varphi_2)$, the global frequency of φ_1 under φ_2 converges according to the limit distribution π_{B_i}. Otherwise, the global frequency is determined by the local frequency before reaching B_i. Then we compute the probability vector $r_A, r_{B_1},\ldots,r_{B_n}$ of reaching BSCCs for which the global frequency of φ_1 under φ_2 obeys the bound $\bowtie q$. For the BSCC B_i having state $s\in Sat(\varphi_2)$,

$$r_{B_i} = \begin{cases} \vec{\mathbf{1}} & \text{if } B_i\cap Sat(\varphi_2)\neq\emptyset \text{ and } \dfrac{\sum_{s\in B_i\cap Sat(\varphi_1)\cap Sat(\varphi_2)}\pi_{B_i}(s)}{\sum_{s\in B_i\cap Sat(\varphi_2)}\pi_{B_i}(s)} \bowtie q, \\ \vec{\mathbf{0}} & \text{otherwise.} \end{cases}$$

For the non-BSCC A, r_A can then be computed as the unique solution of the system of linear equations:

$$(P_A-\mathbf{I})r_A = -\sum_{0\le i\le n} P_{AB_i}r_{B_i}.$$

Finally, we compute the probability of reaching BSCCs having no state $s\in Sat(\varphi_2)$ and satisfying the bound $\bowtie q$. In a manner similar to the procedure used for $v_{j,i}^h$, we compute the occurrence probability of a path by counting the number of states satisfying ψ_2 and $\psi_1\wedge\psi_2$ until reaching BSCCs that have no state $s\in Sat(\varphi_2)$ on the path. Let the vector $u_{j,i}^h(s)$ be the occurrence probability of a path starting from s, visiting states in $Sat(\varphi_2)$ i times, states in $Sat(\varphi_1)\cap Sat(\varphi_2)$ j times, and states in $\bigcup_{B_i\cap Sat(\varphi_2)=\emptyset}B_i$ in h

steps the first time, within h steps.

$$u^0_{0,0}(s) = \begin{cases} 1 & \text{if } s \in \bigcup_{B_i \cap Sat(\varphi_2)=\emptyset} B_i, \\ 0 & \text{otherwise.} \end{cases}, \quad u^h_{j,i}(s) = \begin{cases} 0 & \text{if } s \notin A, \\ P(s,-)\cdot u^{h-1}_{j-1,i-1} & \text{if } s \in Sat(\varphi_1)\cap Sat(\varphi_2)\cap A, \\ P(s,-)\cdot u^{h-1}_{j,i-1} & \text{if } s \in (Sat(\varphi_2)\setminus Sat(\varphi_1))\cap A, \\ P(s,-)\cdot u^{h-1}_{j,i} & \text{otherwise.} \end{cases}$$

The reason $u^h_{j,i}(s) = 0$ if $s \notin A$ for $h > 0$ is that $s \notin A$ cannot reach BSCCs having no state $s \in Sat(\varphi_2)$ in h steps the first time.

The probability of reaching BSCCs having no state $s \in Sat(\varphi_2)$ and satisfying the bound $\bowtie q$ can be obtained analytically as the infinite sum of $u^h_{j,i}(s)$ for $h = 0$ to ∞, because the number of steps required to reach BSCCs from states in the non-BSCC A is unbounded. However, we can adequately approximate the true probability for large h (see Section 4.3.1). Thus we can compute $Prob^{\mathscr{D}}(\mathbf{Q}^{[k,\infty)}_{\bowtie q}\langle\varphi_1|\varphi_2\rangle)$ as follows:

$$Prob^{\mathscr{D}}(\mathbf{Q}^{[k,\infty)}_{\bowtie q}\langle\varphi_1|\varphi_2\rangle) = P^k \cdot \left(\left[r^T_A, r^T_{B_1}, \ldots, r^T_{B_n}\right]^T + \sum_{h=0}^{\infty}\sum_{i=0}^{h+1}\sum_{\substack{j\le i \\ i>0 \Rightarrow j \bowtie i\cdot q}} u^h_{j,i}\right).$$

where the superscript ^{T} means transposition of a vector.

$\mathbf{P}_{\sim p}[\mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle]$ for CTMCs. On a uniformized DTMC $\mathrm{unif}_\lambda(\mathscr{C}) = (S, \bar{s}, P, L)$ of the input CTMC $\mathscr{C}$, the occurrence probability of sequences $s_0 s_1 \ldots$ of states can be captured by the techniques for DTMCs. Therefore, the remainder is the occurrence probability of sequences $t_0 t_1 \ldots$ of spent times such that the ratio of the total spent time in states obeys the bound $\bowtie q$ on the path, for the uniformization rate λ.

Consider a simple case that i states ($s_0, \ldots, s_i$, $i-1$ transitions) are in $[0,k]$, the number of transitions is l ($j \le l < i$) in $[0, qk]$ and $i-l-1$ in the rest of the interval $(qk, k]$ on the path. In this case, the total of t_0 to t_{j-1} is less than $q \cdot k$ and the occurrence probability of a sequence of spent times $t_0 \ldots t_i$ is

$$\begin{aligned} \rho(l;\lambda qk)\cdot\rho(i-l-1;\lambda(1-q)k) &= e^{-\lambda qk}\cdot\frac{(\lambda qk)^l}{l!}\cdot e^{-\lambda(1-q)k}\cdot\frac{(\lambda(1-q)k)^{i-l-1}}{(i-l-1)!} \\ &= \rho(i-1;\lambda k)\cdot\frac{(i-1)!\cdot q^l\cdot(1-q)^{i-l-1}}{l!\cdot(i-l-1)!}. \end{aligned}$$

As above, the occurrence probability of a sequence of spent times obeying the given frequency bound depends on only the numbers of states satisfying subformulae in the interval of interest, and it can be computed using the binomial distribution, because each spent time is independent and exponentially distributed with parameter λ, and the Poisson probability $\rho(i-1;\lambda k)$ is the occurrence probability of $i-1$ transitions in $[0,k]$. Under the other conditions, we can obtain similar results. Hence, the conditional probability $B_{\bowtie q}(j,i)$ of satisfying the frequency bound $\bowtie q$, when the numbers of states satisfying φ_2 and $\varphi_1 \wedge \varphi_2$ in the interval I are i and j respectively, is given by:

$$B_{\bowtie q}(j,i) = \begin{cases} 1 & \text{if } i = 0 \text{ or } (i = j \text{ and } 1 \bowtie q) \text{ or } (j = 0 \text{ and } 0 \bowtie q), \\ \sum_{l=j}^{i-1}\frac{(i-1)!\cdot q^l\cdot(1-q)^{i-l-1}}{l!\cdot(i-l-1)!} & \text{if } 0 < j < i, 0 < q < 1 \text{ and } \bowtie \in \{<, \le\}, \\ \sum_{l=0}^{j-1}\frac{(i-1)!\cdot q^l\cdot(1-q)^{i-l-1}}{l!\cdot(i-l-1)!} & \text{if } 0 < j < i, 0 < q < 1 \text{ and } \bowtie \in \{>, \ge\}, \\ 0 & \text{otherwise.} \end{cases}$$

Here $Prob^{\mathscr{D}}(\mathbf{Q}_{\bowtie q}^{[k,k']}\langle\varphi_1|\varphi_2\rangle)$ is the probability of satisfying $\mathbf{Q}_{\bowtie q}^{[0,k-k']}\langle\varphi_1|\varphi_2\rangle$ after k time units, analogous to the DTMC case, and the transient probability for k time units is $\Pi_k^{\mathscr{C}}$. Therefore, for a bounded interval $I=[k,k']$,

$$Prob^{\mathscr{C}}(\mathbf{Q}_{\bowtie q}^{[k,k']}\langle\varphi_1|\varphi_2\rangle) = \begin{cases} \Pi_k^{\mathscr{C}} \cdot (v_{0,0}^0 \cdot B_{\bowtie q}(0,0) + v_{0,1}^0 \cdot B_{\bowtie q}(0,1) + v_{1,1}^0 \cdot B_{\bowtie q}(1,1)) & \text{if } k=k', \\ \Pi_k^{\mathscr{C}} \cdot \sum_{h=0}^{\infty} \rho(h;\lambda\cdot(k'-k)) \cdot \sum_{i=0}^{h+1}\sum_{j=0}^{i} v_{j,i}^h \cdot B_{\bowtie q}(j,i) & \text{otherwise.} \end{cases}$$

In the numerical computation, this infinite sum for $h=0$ to ∞ can also be truncated as the computation of the transient probability $\Pi_k^{\mathscr{C}}$.

For an unbounded interval $I=[k,\infty)$, we can apply a routine similar to that used for DTMCs. The difference is that we must consider the cumulative binomial probability $B_{\bowtie q}(j,i)$ for $u_{j,i}^h$, and the transient probability $\Pi_k^{\mathscr{C}}$ for k time units instead of P^k.

$$Prob^{\mathscr{C}}(\mathbf{Q}_{<q}^{[k,\infty)}\langle\varphi_1|\varphi_2\rangle) = \Pi_k^{\mathscr{C}} \cdot ([r_A^T, r_{B_1}^T, \ldots, r_{B_n}^T]^T + \sum_{h=0}^{\infty}\sum_{i=0}^{h+1}\sum_{j=0}^{i} u_{j,i}^h \cdot B_{\bowtie q}(j,i)).$$

4.2 Model checking the bounded LTL-like fragment of PFTL

In this section, we introduce a statistical model-checking algorithm for infinite-state MCs and the bounded LTL-like fragment of PFTL:

$$\begin{array}{lll} \text{state formula } \varphi & ::= & \mathbf{P}_{\sim p}[\psi] \\ \text{path formula } \psi & ::= & a \mid \neg\psi \mid \psi_1 \wedge \psi_2 \mid \psi_1 \mathbf{U}^I \psi_2 \mid \mathbf{Q}_{\bowtie q}^I\langle\psi_1|\psi_2\rangle \end{array}$$

where $p \in (0,1)$ and I is a bounded interval of $\mathbb{N}$ for discrete time (or of $\mathbb{R}_{\geq 0}$ for continuous time).

Because it is difficult to check exactly for a bounded LTL-like fragment formula in PFTL, we develop a statistics-based approximation model-checking algorithm. This techniques will provide us with useful information in many cases, even if it is not a strict model-checking procedure. In this approach, we sample finite prefix sequences of the paths of an input MC by probabilistic simulation and statistically determine whether or not the input MC satisfies an input formula by using the sample. We apply the sequential probability ratio test (SPRT) [11] to model checking, as was done in [12] for CSL.

4.2.1 Sequential probability ratio test

The SPRT is a sequential hypothesis test developed by Wald [11]. In a sequential test, the sample size is not fixed: observations are sequentially generated until the sample data indicate which hypothesis to supported under predesigned conditions. In SPRT, we preset the type I error rate $\alpha > 0$, the type II error rate $\beta > 0$, and the indifference region width $2\delta > 0$. For a formula $\mathbf{P}_{\sim p}[\psi]$ ($p \pm \delta \in (0,1)$), we test the null hypothesis H_0: $\hat{p} > p+\delta$ against the alternative hypothesis H_1: $\hat{p} < p-\delta$, where $\hat{p}$ is the true value of the occurrence probability of paths satisfying ψ. If the hypothesis $\hat{p}=\theta$ is true, the number m of paths satisfying ψ for a sample size n is binomially distributed $n!\theta^m(1-\theta)^{n-m}/(m!(n-m)!)$. Conversely, this value represents the likelihood of the hypothesis $\hat{p}=\theta$ if we observe that m paths satisfy ψ for a sample size n. Therefore, the likelihood ratio Λ of H_0 to H_1 for a sample $\{\omega_1,\ldots,\omega_n\}$ is:

$$\Lambda(\{\omega_1,\ldots,\omega_n\}) = \frac{(p+\delta)^m(1-(p+\delta))^{n-m}}{(p-\delta)^m(1-(p-\delta))^{n-m}}$$

where $m = |\{\omega_i | \omega_i \models \psi\}|$.

Here H_0 is more likely than H_1 for a given sample if the likelihood ratio is greater than 1 and H_1 is more likely than H_0 for the sample if the likelihood ratio is less than 1. For an observed sample $\{\omega_1, \ldots, \omega_n\}$ and error rates α and β, the next action is determined as follows:

$$\begin{cases} \text{Accept } H_0 & \text{if } \Lambda(\{\omega_1, \ldots, \omega_n\}) > (1-\beta)/\alpha, \\ \text{Accept } H_1 & \text{if } \Lambda(\{\omega_1, \ldots, \omega_n\}) < \beta/(1-\alpha), \\ \text{Observe and add } \omega_{n+1} \text{ to the sample} & \text{otherwise.} \end{cases}$$

As a result, the probability of accepting the hypothesis H_0 is at least $1-\alpha$ if $\hat{p} > p+\delta$, and at most β if $\hat{p} < p-\delta$. If $|\hat{p}-p| < \delta$, the hypotheses are indifferent at error rates α and β.

4.2.2 Satisfaction checking for bounded path formulae against paths

To carry out a test, we must check $\omega \models \psi$ for a sample path ω and a bounded formula ψ with the total boundary k_{total}. Whether or not $\omega \models \psi$ does not depend on the suffix after k_{total} steps/time-units of ω. For the finite prefix on $[0, k_{total}]$ of ω, we recursively compute an ordered set $SatInt_\omega(\psi)$ of subintervals satisfying ψ in $[0, k_{total}]$, using ordered sets of subintervals satisfy subformulae of ψ. We can then derive $\omega \models \psi$ if there exists $I \in SatInt_\omega(\psi)$ such that $0 \in I$.

We assume that $SatInt_\omega(\psi_1)$ and $SatInt_\omega(\psi_2)$ are already computed and merged. By writing $SatInt_\omega(\psi) = \{I_i, \ldots, I_n\}$, we mean that the set $\{I_i, \ldots, I_n\}$ satisfies $I_i \cap I_{i+1} = \emptyset$, $\sup I_i \leq \inf I_{i+1}$ and $\sup I_i = \inf I_{i+1} \Rightarrow \sup I_i \notin I_i, I_{i+1}$. In this paper, we do not include an algorithm for DTMCs, because the structure of a discrete-time path is simple, and it is not worthwhile to pursue the matter.

$a \in AP$ for CTMCs. For an atomic proposition $a \in AP$, the set of intervals satisfying a is determined immediately by the labeling function L. Therefore, $SatInt_\omega(a) = \{[\sum_{j=0}^{i-1} time(\omega, j), \sum_{j=0}^{i} time(\omega, j)) | a \in L(\omega(i))\}$.

$\neg\psi_1$ for CTMCs. $SatInt_\omega(\neg\psi_1)$ is a set of intervals complementary to the union of intervals in $SatInt_\omega(\psi_1)$ in $[0, k_{total}]$. Therefore, for $SatInt_\omega(\psi_1) = \{I_1, \ldots I_n\}$, $SatInt_\omega(\neg\psi_1) = \{[0, \inf I_1] \setminus I_1, [\sup I_n, k_{total}] \setminus I_n\} \cup \bigcup_{i=1}^{n-1} \{([\sup I_i, \inf I_{i+1}] \setminus I_i) \setminus I_{i+1}\}$.

$\psi_1 \wedge \psi_2$ for CTMCs. $SatInt_\omega(\psi_1 \wedge \psi_2)$ is a set of intervals intersecting each element of $SatInt_\omega(\psi_1)$ and each element of $SatInt_\omega(\psi_2)$. Therefore, for $SatInt_\omega(\psi_1) = \{I_1, \ldots I_n\}$ and $SatInt_\omega(\psi_2) = \{J_1, \ldots J_m\}$, $SatInt_\omega(\psi_1 \wedge \psi_2) = \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} \{I_i \cap J_j\}$.

$\psi_1 \mathbf{U}^I \psi_2$ for CTMCs. Let $SatInt_\omega(\psi_1) = \{I_1, \ldots I_n\}$ and $SatInt_\omega(\psi_2) = \{J_1, \ldots J_m\}$. For time points $t \in I_i \cup \{\inf I_i\}$ and $t' \in J_j$ such that $t < t'$, there exists $I' \in SatInt_\omega(\psi_1 \mathbf{U}^I \psi_2)$ such that $t \in I'$ if $(t, t') \subseteq I_i$ and $t' \in I + t$. In this case, t' is in $(I_i \cup \sup I_i) \cap J_j$ $(= Y_{i,j})$ and t is in $X_{i,j}$ where $\inf X_{i,j} = \inf Y_{i,j} - \sup I$, $\sup X_{i,j} = \sup Y_{i,j} - \inf I$, $(\inf Y_{i,j} \in Y_{i,j} \wedge \sup I \in I) \Leftrightarrow \inf X_{i,j} \in X_{i,j}$ and $(\sup Y_{i,j} \in Y_{i,j} \wedge \inf I \in I) \Leftrightarrow \sup X_{i,j} \in X_{i,j}$. In addition, $SatInt_\omega(\psi_2) \subseteq SatInt_\omega(\psi_1 \mathbf{U}^I \psi_2)$ if $0 \in I$. Therefore,

$$SatInt_\omega(\psi_1 \mathbf{U}^I \psi_2) = \bigcup_{j=1}^{m} \bigcup_{i=1}^{n} \{X_{i,j} \cap (I_i \cup \{\inf I_i\})\} \cup \begin{cases} SatInt_\omega(\psi_2) & \text{if } 0 \in I, \\ \emptyset & \text{otherwise.} \end{cases}$$

$\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle$ for CTMCs. If $I=[0,0]$, $\mathbf{Q}^{[0,0]}_{\bowtie q}\langle\psi_1|\psi_2\rangle$ is just a conditional statement. Therefore, $SatInt_\omega(\mathbf{Q}^{[0,0]}_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ is equal to $SatInt_\omega(\psi_2\to\psi_1)$ if $1\bowtie q$, $SatInt_\omega(\psi_2\to\neg\psi_1)$ otherwise.

If $\inf I>0$, $\mathscr{C},\omega\models\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle\Leftrightarrow\mathscr{C},\omega^{\inf I}\models\mathbf{Q}^{I-\inf I}_{\bowtie q}\langle\psi_1|\psi_2\rangle$ by Definition 6. Therefore, if $\inf I>0$, $SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)=\{J-\inf I|J\in SatInt_\omega(\mathbf{Q}^{I-\inf I}_{\bowtie q}\langle\psi_1|\psi_2\rangle)\}$.

If $\inf I=0$ and $\sup I=k>0$, $SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ satisfy the property $J\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)\Leftrightarrow J\in\{t|f^I_{\langle\psi_1|\psi_2\rangle}(t)\bowtie q\}$ for any interval J. Therefore, we determine $SatInt_\omega(\mathbf{Q}^{I-\inf I}_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ by analyzing $f^I_{\langle\psi_1|\psi_2\rangle}(t)$. First, for $SatInt_\omega(\psi_1\wedge\psi_2)=\{I_1,\dots I_n\}$ and $SatInt_\omega(\psi_2)=\{J_1,\dots J_m\}$, we compute a set $\mathrm{nondif}_\omega(\psi_1|\psi_2)$ of candidates for non-differentiable points of $f^I_{\langle\psi_1|\psi_2\rangle}(t)$.

$$\mathrm{nondif}_\omega(\psi_1|\psi_2)=\{0,k_{total}-k\}\cup\{\inf I'-k,\inf I',\sup I'-k,\sup I'|I'\in\{I_1,\dots,I_n,J_1,\dots,J_m\}\}.$$

Let $\{t_1,\dots,t_l\}$ be the ordered elements of $\mathrm{nondif}_\omega(\langle\psi_1|\psi_2\rangle)$. The truth values of ψ_2 and $\psi_1\wedge\psi_2$ are unchanged in each interval (t_i,t_{i+1}) and $(t_i,t_{i+1})+k$, because if their truth values did change, there would have to be other non-differentiable points between t_i and t_{i+1}. Thus $f^I_{\langle\psi_1|\psi_2\rangle}(t)$ is monotonically increasing, monotonically decreasing, fixed, or undefined in the interval (t_i,t_{i+1}). In addition, $f^I_{\langle\psi_1|\psi_2\rangle}(t)$ is equal to $f^{(\inf I,\sup I)}_{\langle\psi_1|\psi_2\rangle}(t)$ for $t\in(t_i,t_{i+1})$.

Hence, for a non-differentiable time point t_i, $[t_i,t_i]\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ if $f^I_{\langle\psi_1|\psi_2\rangle}(t_i)\bowtie q$. For an interval (t_i,t_{i+1}) between non-differentiable time points, we determine whether or not (t_i,t_{i+1}), or a subinterval of it, is in $SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ as follows. For $\mathscr{L}^J_\psi(t)=\mathscr{L}(\bigcup_{I'\in SatInt(\psi)}I'\cap(J+t))$:

1. If $\mathscr{L}^{(0,k)}_{\psi_2}(t_i)=0$ and $\mathscr{L}^{(0,k)}_{\psi_2}(t_{i+1})=0$, ψ_2 and $\psi_1\wedge\psi_2$ do not hold on an interval $\subseteq(t_i,t_{i+1})$ with positive time length. Therefore, if $f^{(0,k]}_{\langle\psi_1|\psi_2\rangle}(t_i)$ $(=f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t')$ for $t_i<t'<t_{i+1})$ is undefined or obeys the bound $\bowtie q$, then $(t_i,t_{i+1})\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$.
2. If $\mathscr{L}^{(0,k)}_{\psi_2}(t_i)=0$ and $\mathscr{L}^{(0,k)}_{\psi_2}(t_{i+1})>0$, $f^I_{\langle\psi_1|\psi_2\rangle}(t)$ is fixed and equal to $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_{i+1})$ in the interval (t_i,t_{i+1}). Therefore, if $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_{i+1})$ obeys the bound $\bowtie q$, then $(t_i,t_{i+1})\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$.
3. If $\mathscr{L}^{(0,k)}_{\psi_2}(t_i)>0$ and $\mathscr{L}^{(0,k)}_{\psi_2}(t_{i+1})=0$, $f^I_{\langle\psi_1|\psi_2\rangle}(t)$ is fixed and equal to $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_i)$ in the interval (t_i,t_{i+1}). Therefore, if $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_i)$ obeys the bound $\bowtie q$, then $(t_i,t_{i+1})\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$.
4. If $\mathscr{L}^{(0,k)}_{\psi_2}(t_i)>0$ and $\mathscr{L}^{(0,k)}_{\psi_2}(t_{i+1})>0$, $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t)$ is monotonically increasing, monotonically decreasing, or fixed in (t_i,t_{i+1}). Therefore, if both $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_i)$ and $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_{i+1})$ obey the bound $\bowtie q$, then $(t_i,t_{i+1})\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$. Moreover, if either $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_i)$ or $f^{(0,k)}_{\langle\psi_1|\psi_2\rangle}(t_{i+1})$ obeys the bound $\bowtie q$, then (t_i,t'_i) or $(t'_i,t_{i+1})\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ where t'_i satisfies:

$$q=\frac{\mathscr{L}^{(0,k)}_{\psi_1\wedge\psi_2}(t_i)+a(t'_i-t_i)}{\mathscr{L}^{(0,k)}_{\psi_2}(t_i)+b(t'_i-t_i)}$$

with $a=(\mathscr{L}^{(0,k)}_{\psi_1\wedge\psi_2}(t_{i+1})-\mathscr{L}^{(0,k)}_{\psi_1\wedge\psi_2}(t_i))\ /(t_{i+1}-t_i)$ and $b=(\mathscr{L}^{(0,k)}_{\psi_2}(t_{i+1})-\mathscr{L}^{(0,k)}_{\psi_2}(t_i))\ /(t_{i+1}-t_i)$.

$$t'_i=t_i+\frac{q\cdot\mathscr{L}^{(0,k)}_{\psi_2}(t_i)-\mathscr{L}^{(0,k)}_{\psi_1\wedge\psi_2}(t_i)}{a-qb}.$$

In addition, because $f^I(t'_i)=q$, $[t'_i,t'_i]\in SatInt_\omega(\mathbf{Q}^I_{\bowtie q}\langle\psi_1|\psi_2\rangle)$ if $\bowtie\in\{\leq,\geq\}$.

4.3 Complexity

4.3.1 Complexity of model checking for the CTL-like fragment of PFTL

For a DTMC $\mathscr{D} = (S, \bar{s}, P, L)$ or a CTMC $\mathscr{C} = (S, \bar{s}, Q, L)$ (and its uniformized DTMC $\mathrm{unif}_\lambda(\mathscr{C}) = (S, \bar{s}, P, L)$) and a CTL-like fragment formula φ, the time complexity of model checking is linear in $|\varphi|$, which is the number of operators in φ, and polynomial in $|S|$, which is the complexity of the recursive procedure for each operators. Except for a **Q** formula, the time complexity is the same to that for PCTL/CSL [8, 3]. For each bounded $\mathbf{P}_{\sim p}[\mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle]$ ($\sup I < \infty$), computing the sum of vectors $v^h_{j,i}$ takes $\mathscr{O}(|S|^2 \cdot k^3)$ time, where $k = \sup I$ for DTMCs or $k = \lambda \cdot \sup I$ for CTMCs. For each unbounded $\mathbf{P}_{\sim p}[\mathbf{Q}^I_{\bowtie q}\langle\varphi_1|\varphi_2\rangle]$, computing the limit distributions $\pi_{B_1}, \ldots, \pi_{B_n}$ and the reachability vectors $r_A, r_{B_1}, \ldots, r_{B_n}$ takes $\mathscr{O}(|S|^3)$ time, where A and $B_1, \ldots, B_n$ are a non-BSCC and BSCCs of P, respectively. If the input MC is reducible, an additional computation of the transient probability $\Pi^{\mathscr{C}}_k$ and the sum of vectors $u^h_{j,i}$ takes $\mathscr{O}(|S|^2 \cdot k' + |S|^2 \cdot |\log e|^{-3})$ time, where $k' = \inf I$ for DTMCs or $k' = \lambda \cdot \inf I$ for CTMCs, and e is the maximum absolute value of the eigenvalues of the partial matrix P_A consisting the non-BSCC A of P. This is because the probability vector of reaching BSCCs within $\mathscr{O}(|\log e|^{-1})$-steps is sufficiently close to the probability vector of reaching BSCCs within an unbound number of steps.

4.3.2 Complexity of model checking for the LTL-like fragment of PFTL

The complexity of model checking for the LTL-like fragment of PFTL is divided into two parts, the complexity of the sample used in the testing and the complexity of the observations and satisfaction checking for a sample trace of path. Regarding the sample size, approximations for the expected sample size are provided in [11, 13]. This size depends on the chosen significance level and the difference between the query value p and the true probability $\hat{p}_\psi$ for an input formula $\mathbf{P}_{\sim p}[\psi]$. However, this is not specific to our method, and the details of the expected size are omitted from this paper. The observation of a sample path is just a probabilistic simulation, and its time complexity is $\mathscr{O}(k_{total} \cdot \log|E|)/\mathscr{O}(\lambda \cdot k_{total} \cdot \log|E|)$ where k_{total} is the total boundary of the input formula, $|E|$ is the number of transition choices of the input MC and λ is the average exit rate of the input CTMC, for the input DTMC/CTMC. For an input formula $\mathbf{P}_{\sim p}[\psi]$ and a DTMC, we need only count states satisfying subformulae for each operator. Therefore, the satisfaction checking takes $\mathscr{O}(k_{total} \cdot |\psi|)$ time, where $|\psi|$ is the size of ψ. However, on a CTMC, the size of the set of intervals satisfying formulae is at worst twice that of the set of intervals satisfying subformulae, for each **Q** operator. Thus, the satisfaction checking takes $\mathscr{O}(\lambda \cdot k_{total} \cdot |\psi| \cdot 2^{|\psi|_{\mathbf{Q}}})$ time, where $|\psi|_{\mathbf{Q}}$ is the number of **Q** operators in ψ. In practice, many intervals satisfying formulae are merged, because each spent time on a state is exponentially distributed with the exit rate of the state and the probability of generating a bad sample path by probabilistic simulation is negligible.

5 Conclusions and future directions

We introduced the frequency operator **Q** and defined the syntax and semantics of PFTL. PFTL has rich expressiveness, and it is difficult to develop a model checker for the full logic. However, we developed a numerical model-checking algorithm for the CTL-like fragment of PFTL against finite-state MCs, and a statistical model-checking algorithm for the bounded LTL-like fragment of PFTL against infinite-state MCs. The statistical model-checking is not strict, but we anticipate that it will provide useful information in many cases. Especially, it is worth noting that the **Q** operator can, in a sense, express a conditional probability between path formulae, without path quantifications.

Our extension is based on an intuitive idea for describing a property of a behavior, especially in a probabilistic system. Although, it is difficult to strictly check a model for the logic, and also the non-probabilistic version of PFTL, because it is intractable from the viewpoint of automata theory. Therefore, it will be necessary to find treatable and useful fragments of the logic and classes of restricted models. This is one future direction of our research. Another future direction is to provide approximate model-checking against more complex systems, or for further extended logics. In this paper, we have assumed that our model is an MC. Nevertheless, we can apply this type of approximate model-checking via statistical methods to more general stochastic processes, e.g., systems of stochastic ordinary differential equations (continuous states and continuous transitions), because we can directly use discretized traces of paths obtained from stochastic simulations. Also, it is not difficult to check whether or not a sample path satisfies a bounded property such as "φ_2 holds in the interval $[0,10]$ and φ_1 holds to more than 90% of the time points until that point" (frequently φ_1 until φ_2).

References

[1] R. Alur & D. L. Dill (1994): *A theory of timed automata*. *Theoretical Computer Science* 126(2), pp. 183–235, doi:10.1016/0304-3975(94)90010-8.

[2] A. Aziz, K. Sanwal, V. Singhal & R. Brayton (1996): *Verifying continuous time Markov chains*. *Computer Aided Verification, LNCS* 1102, pp. 269–276, doi:10.1007/3-540-61474-5_75.

[3] C. Baier, B. Haverkort, H. Hermanns & J. P. Katoen (2003): *Model-checking algorithms for continuous-time Markov chains*. *IEEE Transactions on software engineering* 29(6), pp. 524–541, doi:10.1109/TSE.2003.1205180.

[4] P. Bouyer, N. Markey, J. Ouaknine & J. Worrell (2008): *On expressiveness and complexity in real-time model checking*. In: *Automata, Languages and Programming, LNCS* 5126, pp. 124–135, doi:10.1007/978-3-540-70583-3_11.

[5] R. S. Cohen & A. Y. Gold (1977): *Theory of ω-languages I: characterizations of ω-context-free languages*. *Journal of Computer and System Sciences* 15(2), pp. 169–184, doi:10.1016/S0022-0000(77)80004-4.

[6] E. A. Emerson (1990): *Temporal and modal logic*. In: *Handbook of theoretical computer science, volume B: formal models and semantics*, MIT Press, pp. 995–1072.

[7] B. L. Fox & P. W. Glynn (1988): *Computing Poisson probabilities*. *Communications of the ACM* 31(4), pp. 440–445, doi:10.1145/42404.42409.

[8] H. Hansson & B. Jonsson (1994): *A logic for reasoning about time and reliability*. *Formal Aspects of Computing* 6(5), pp. 512–535, doi:10.1007/BF01211866.

[9] R. Koymans (1990): *Specifying real-time properties with metric temporal logic*. *Real-Time Systems* 2, pp. 255–299, doi:10.1007/BF01995674.

[10] M. Kwiatkowska, G. Norman & D. Parker (2007): *Stochastic model checking*. In: *Formal Methods for Performance Evaluation, LNCS* 4486, pp. 220–270, doi:10.1007/978-3-540-72522-0_6.

[11] A. Wald (1945): *Sequential tests of statistical hypotheses*. *The Annals of Mathematical Statistics* 16(2), pp. 117–186, doi:10.1214/aoms/1177731118.

[12] H. Younes & R. Simmons (2002): *Probabilistic verification of discrete event systems using acceptance sampling*. *Computer Aided Verification, LNCS* 2404, pp. 23–39, doi:10.1007/3-540-45657-0_17.

[13] H. L. S. Younes, M. Kwiatkowska, G. Norman & D. Parker (2006): *Numerical vs. statistical probabilistic model checking*. *International Journal on Software Tools for Technology Transfer* 8(3), pp. 216–228, doi:10.1007/s10009-005-0187-8.

[14] C. Zhou, C. A. R. Hoare & A. P. Anders P. Ravn (1991): *A calculus of durations*. *Information Processing Letters* 40(5), pp. 269–276, doi:10.1016/0020-0190(91)90122-X.