Statistical Model Checking
(an overview)

Paolo Ballarini
The Model Checking approach

- non-Probabilistic Model
- Labelled Transition System
- Temporal Logic formula
- Probabilistic MODEL CHECKER:
  - True
  - False (counter-example)
The Probabilistic Model Checking approach

Probabilistic Model

DTMC, CTMC, MDP

Probabilistic Temporal Logic formula

Threshold problem: to determine whether the probability of the formula to hold matches the given threshold

\[ P_{\leq 0.9} \phi \] threshold formula

True

False
The Probabilistic Model Checking approach

- Probabilistic Model
- DTMC, CTMC, MDP
- Quantitative formula: $P = \phi$
- Approximation problem: to determine an approximation for the probability of the formula to hold

**Probabilistic Model**

- **DTMC, CTMC, MDP**
- **Quantitative formula:** $P = \phi$

**Approximation problem:**

to determine an approximation for the probability of the formula to hold
Numerical vs Statistical

**Numerical model checking**: application of numerical methods for Model's solution (i.e. *transient analysis*), the approximation is in form of truncation error
- **+**: better when very accurate approximation are required
- **−**: large memory requirements
- **−**: many realistic models are numerically untreatable

**Statistical model checking**: application of discrete-event simulation to sampling model's executions used to produced an approximated output
- **+**: small memory requirements (no need to store the state-space)
- **+**: only option for verification of many realistic models
- **−**: very accurate approximations may be computationally harder
Outline

Temporal Logic for CTMC Models

Statistical Model Checking
  Threshold problem (Hypothesis Testing)
  Approximation problem

Current work, future directions
  Confidence Interval based Model Checking
Outline

Temporal Logic for CTMC Models

Statistical Model Checking
- Threshold problem (Hypothesis Testing)
- Approximation problem

Current work, future directions
- Confidence Interval based Model Checking
Temporal Queries for CTMC models

CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both *probability measures* and *time bounds*.

- **what is the probability that the system will FAIL (within time $T$)** ("unconditional" reachability)
- **what is the probability that the system will go from WORK to FAIL (within time $T$)** ("simple-conditional" reachability)
- **what is the probability that the system will go from WORK to DEGR to FAIL (within time $T$)** ("sequential" reachability)
- **what is the probability that the system will go from WORK to DEGR within time $T_1$ and then from DEGR to FAIL within time $T_2$** ("timed-sequential" reachability)
Temporal Queries for CTMC models

CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both probability measures and time bounds.

- **what is the probability that the system will FAIL (within time $T$)**
  (“unconditional” reachability)

- **what is the probability that the system will go from WORK to FAIL (within time $T$)**
  (“simple-conditional” reachability)

- **what is the probability that the system will go from WORK to DEGR to FAIL (within time $T$)**
  (“sequential” reachability)

- **what is the probability that the system will go from WORK to DEGR within time $T_1$ and then from DEGR to FAIL within time $T_2$**
  (“timed-sequential” reachability)
Temporal Queries for CTMC models

CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both probability measures and time bounds.

- **What is the probability that the system will FAIL (within time $T$)** ("unconditional" reachability)
- **What is the probability that the system will go from WORK to FAIL (within time $T$)** ("simple-conditional" reachability)
- **What is the probability that the system will go from WORK to DEGR to FAIL (within time $T$)** ("sequential" reachability)
- **What is the probability that the system will go from WORK to DEGR within time $T_1$ and then from DEGR to FAIL within time $T_2$** ("timed-sequential" reachability)
Temporal Queries for CTMC models

CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both probability measures and time bounds.

- what is the probability that the system will FAIL (within time $T$) (“unconditional" reachability)
- what is the probability that the system will go from WORK to FAIL (within time $T$) (“simple-conditional" reachability)
- what is the probability that the system will go from WORK to DEGR to FAIL (within time $T$) (“sequential" reachability)
- what is the probability that the system will go from WORK to DEGR within time $T_1$ and then from DEGR to FAIL within time $T_2$ (“timed-sequential " reachability)
Temporal Logic Taxonomy

**CSL**: Continuous Stochastic Logic

**asCSL**: action-state CSL

**CSL-TA**: CSL Timed Automata
Temporal Logic Taxonomy

**CSL**: Continuous Stochastic Logic

**asCSL**: action-state CSL

**CSL-TA**: CSL Timed Automata

The diagram illustrates the relationships between different temporal logics and automata. The logics include:

- **CSL**
- **LTL**
- **asCSL**
- **CSL-TA**

Each logic is represented as a distinct node, and the diagram shows how they are related and nested within each other.
Temporal Logic Taxonomy

**CSL**: Continuous Stochastic Logic

**asCSL**: action-state CSL

**CSL-TA**: CSL Timed Automata

---

Diagram:

- **CSL-TA**: Contains CSL and asCSL
- **CSL**: Contains Reachability
- **asCSL**: Contains Reachability
- **LTL**: Contains Reachability
CSL (Continuous Stochastic Logic)

- CSL: a branching time logic
  \[ \phi := a | \neg \phi | \phi \land \phi | S_{\sim p}(\phi) | \mathcal{P}_{\sim p}(\phi) \]
  \[ \varphi := X^I \phi | \phi U^I \phi \]

- \( \sim \in \{<, \leq, \geq, >, =, \neq\} \), \( I \subseteq \mathbb{R}_{\geq 0} \) (time-bound)

- characteristic of CSL: each temporal operator must be preceded by a probability bound \( \sim_p \)

\[ \phi \equiv \mathcal{P}_{\leq 0.9}(\text{WORK} U [\mathcal{P}_{< 0.5}(\text{DEGR} U \text{FAIL})]) \]
Outline

Temporal Logic for CTMC Models

Statistical Model Checking
  Threshold problem (Hypothesis Testing)
  Approximation problem

Current work, future directions
  Confidence Interval based Model Checking
Statistical MC: two formulations

1 **Threshold problem:** does the probability measure of query $Q$ meets the bound $\sim p$?
   - CSL
     - Younes, Simmons: YMER (sequential Acceptance sampling)
     - San et al: VESTA (simple hypothesis testing)
     - Katoen, Zapreev: MRMC (through Estimation)
   - LTL
     - Clarke et al.: (hypothesis Testing + Bayesian)

2 **Estimation problem:** what is the probability measure of query $Q$?
   - LTL
     - Peyronnet et al.: APMC (Chernoff-Hoeffding bound)
   - CSL
     - Parker et al.: PRISM (implementation of Peyronnet et al method)
     - Katoen, Zapreev: MRMC (?)
   - LTL
     - Ballarini et al: OCEAN (Wilson Confidence Interval)
Statistical MC: two formulations

1. **Threshold problem**: does the probability measure of query $Q$ meets the bound $\sim p$?
   - **CSL**
     - Younes, Simmons: YMER (sequential Acceptance sampling)
     - San et al: VESTA (simple hypothesis testing)
     - Katoen, Zapreev: MRMC (through Estimation)
   - **LTL**
     - Clarke et al.: (hypothesis Testing + Bayesian)

2. **Estimation problem**: what is the probability measure of query $Q$?
   - **LTL**
     - Peyronnet et al.: APMC (Chernoff-Hoeffding bound)
   - **CSL**
     - Parker et al.: PRISM (implementation of Peyronnet et al. method)
     - Katoen, Zapreev: MRMC (?)
   - **LTL**
     - Ballarini et al: OCEAN (Wilson Confidence Interval)
Statistical MC: two formulations

1. **Threshold problem**: does the probability measure of query $Q$ meets the bound $\sim p$?
   - CSL
     - Younes, Simmons: YMER (sequential Acceptance sampling)
     - San et al: VESTA (simple hypothesis testing)
     - Katoen, Zapreev: MRMC (through Estimation)
   - LTL
     - Clarke et al.: (hypothesis Testing + Bayesian)

2. **Estimation problem**: what is the probability measure of query $Q$?
   - LTL
     - Peyronnet et al.: APMC (Chernoff-Hoeffding bound)
   - CSL
     - Parker et al.: PRISM (implementation of Peyronnet et al. method)
     - Katoen, Zapreev: MRMC (?)
   - LTL
     - Ballarini et al: OCEAN (Wilson Confidence Interval)
Statistical MC: basic idea

- **Inputs:**
  - a model $M$ and property $\phi$
  - a bound for the desired level of approximation:
    - **hypothesis testing:** error probability bounds $\langle \alpha, \beta \rangle$ + width of indifference region $\delta$
    - **error bounded estimation:** the desired error $\epsilon$ + the confidence level $\delta$
    - **confidence interval estimation:** a confidence level $\alpha$ + and width of the confidence interval $\epsilon$

- generate $N$ (finite) *sample trajectories* $\sigma_i$ through discrete-event stochastic simulation

- to each trajectory $\sigma_i$ corresponds a Bernoulli variable $X_i$:

\[
\begin{align*}
X_i = 1 & \iff (\sigma_i \models \phi) \\
X_i = 0 & \iff (\sigma_i \not\models \phi)
\end{align*}
\]

- **outcome:**
  - an approximated answer to $M \models \phi$ with a bounded error probability.
  - an approximated estimate of $Pr(\phi, M)$
Hypothesis-testing based Statistical MC

- **the problem:** decide whether $Pr(\phi, M) \sim \theta$, for example:

$$\mathcal{P}_{\leq 0.05}(true \ U^{[0,3]} \ FAIL)$$

(“there’s at most $\theta = 0.05\%$ chance that system will FAIL within time 3”)

- **principle:** there’s no need to get an accurate estimate of $Pr(\phi, M)$ in order to decide whether $Pr(\phi, M) \sim \theta$: a “rough” estimate (resulting from small samples) may be enough

- **hypothesis testing formulation:** the decision on $Pr(\phi, M) \sim \theta$ can be formulated in terms of an Hypothesis Testing problem, with

$$H : p \geq \theta$$
$$K : p < \theta$$

where $p = Pr(\phi, M)$ is the actual probability.

- $\alpha$ : probability of accepting $K$ when $H$ holds (false negative)
- $\beta$ : probability of accepting $H$ when $K$ holds (false positive)

$\langle \alpha, \beta \rangle$ is referred to as the strength of the test
Acceptance Sampling [Younes, Simmons]

(a) Prob. of accepting $H$ (ideal)

(b) Prob. of accepting $H$ (with indifference region)

- **a)** problems: requires exhaustive sample

- **b)** an indifference region is introduced and hypothesis testing is reformulated:

$$
H_0 : p \geq p_0 \quad p_0 = \theta + \delta \\
H_1 : p < p_1 \quad p_1 = \theta - \delta
$$
Acceptance Sampling (1): fixed-size samples

The number of successes \( Y = \sum_i X_i \) is Binomial

\[
Pr(Y \leq c) = F(c, n, p) = \sum_{i=0}^c \binom{n}{i} p^i (1 - p)^{n-i}
\]

fixed-size samples

- \( H_0 \) is accepted if \( \sum_i X_i > c \) (\( c \) is a constant)
- the acceptance problem corresponds to an optimization problem: “find \( n \) and \( c \) such that;”

\[
F(c, n, p_0) \leq \alpha \quad \quad 1 - F(c, n, p_1) \leq \beta
\]

given the strength \( \langle \alpha, \beta \rangle \) and the indifference region \( IR = [p_1, p_0] \).

approximated sample size is:

\[
\tilde{n} = \Phi^{-1}(\alpha) \sqrt{p_0(1 - p_0)} + \Phi^{-1}(\beta) \sqrt{p_1(1 - p_1)} \frac{1}{(p_0 - p_1)^2}
\]

Note that: for a fixed width of \( IR \) the sample size is largest if \( IR \) is centered in \( p = 1/2 \) and decreases when moving towards the extremes 0 and 1.
Acceptance Sampling (2): sequential test

- **Wald sequential test:** a fixed size sample $n$ may be a waist
  - if after $m < n$ observations $Y > c$ then we can accept $H_0$
  - if after $m < n$ observations $Y$ cannot exceed $n$ then we can accept $H_1$

$$f_m = \prod_{i=1}^{m} \frac{Pr[X_i = x_i|p = p_1]}{Pr[X_i = x_i|p = p_0]} = \frac{p_1^{d_m} (1 - p_1)^{m-d_m}}{p_0^{d_m} (1 - p_0)^{m-d_m}}$$

$$d_m = \sum_{i=1}^{m} x_i$$

- $H_0$ is accepted if $f_m \leq B$ with $B = \frac{\beta}{1 - \alpha}$
- $H_1$ is accepted if $f_m \geq A$ with $A = \frac{1 - \beta}{\alpha}$
CSL Model Checking through acceptance sampling

**so far:** given a simple formula \( \phi \) we set the error bounds \( \alpha \) and \( \beta \) and the method estimates whether \( \phi \) holds with errors \( \alpha \) and \( \beta \)

\[ P \leq 0.05(WORK \; U^{[0,3]} \; FAIL) \]

**problem:** how do we set the error bounds for “complex formulae”?

\[ \cdot P \leq 0.05(WORK \; U^{[0,3]} \; FAIL) \land P \geq 0.6(WORK \; U^{[1,2]} \; DEGR) \; \text{(conjunction)} \]
\[ \cdot P \leq 0.5(WORK \; U^{[0,3]} \; P \geq 0.1(DEGR \; U^{[1,2]} \; FAIL)) \; \text{(nested UNTIL)} \]

- **non-nested path formulae:**
  - **conjunction:** to verify \( \bigwedge_{i=1}^{n} \Phi_i \) with strenght \( \langle \alpha, \beta \rangle \) it is sufficient to verify each conjunct with strength \( \langle \alpha/n, \beta/n \rangle \).

- **nested path formulae:**
  - the strength of the test for a property with nested prob. operators is a function on the strength of the test of its nested prob. operators
  - in order to verify \( P \sim_{\theta}(\varphi) \) with strength \( \langle \alpha, \beta \rangle \) set the IR to \( p_0 = (\theta + \delta(\theta)(1 - \alpha')) \) and \( p_1 = 1 - (1 - (\theta - \delta(\theta)))(1 - \beta') \), where \( \langle \alpha', \beta' \rangle \) is the strength of the test for \( \varphi \).
Statistical CSL threshold problem: summary

- **Approach:** statistical solution to the "threshold problem" for (subset of) CSL

- **Featured properties:**
  - time-bounded CSL (both simple and nested)

- **Non-Featured properties:**
  - unbounded CSL (a tentative-solution has been proposed by Sen *et al.* [2005])
  - steady-state formulae (on-going work by Pekergin, El Raib [2009])

- **Tools:**
  - YMER (Younes, Simmons)
  - VESTA (Sen *et al.*)
Statistical MC as an Approximation Problem

**Problem:** given a formula $\phi$ and probabilistic model $M$, calculate an approximation of $p = Pr(\phi, M)$ based on $N$ sampled trajectories of $M$.

- the point estimator is $\hat{p} = \frac{\sum_i X_i}{N}$

there are 2 approaches:

1. **error bounded estimation (Chernoff-Hoeffding bound):**
   - $\epsilon$: error
   - $\delta$: confidence-level
   $$\Pr\left((p - \epsilon) \leq \hat{p} \leq (p + \epsilon)\right) \geq 1 - \delta$$
   the probability that estimate $\hat{p}$ is farther than $\epsilon$ from actual value is less than $\delta$

2. **confidence-interval estimation:**
   - $\alpha$: confidence-level
   - $\epsilon$: interval-width
   $$\Pr\left(u(\hat{p}) \leq p \leq v(\hat{p})\right) \geq 1 - \alpha$$
   $$v(\hat{p}) - u(\hat{p}) = \epsilon$$: functions of the sample
   the probability that the actual value is farther than $\delta$ from the estimate $\hat{p}$ is less than $\alpha$
Approximate Probabilistic Model Checking

- APMC tool [Peyronnet et al.]: approximate LTL model checking tool

- it calculates approximated probability measure for:
  - general time-bounded LTL
  - time-unbounded monotone LTL (i.e. formulae of the positive fragment of LTL, i.e. negation only applied to atomic-propositions)

- it supports two error bounded estimation schemes:
  - Randomized approximation with additive error
    - fixed sample-size: \( N = \ln\left(\frac{2}{\delta}\right)/2\epsilon^2 \)
  - Randomized approximation with multiplicative error
    - complex 3-steps estimation procedure
Outline

Temporal Logic for CTMC Models

Statistical Model Checking
   Threshold problem (Hypothesis Testing)
   Approximation problem

Current work, future directions
   Confidence Interval based Model Checking
Confidence Interval based Model Checking

- OCEAN [Ballarini et al.]: a prototype tool for confidence-interval estimation of the probability of time-bounded LTL formulae vs CTMC biological models expressed in SBML

- it is based on the *Wilson-score interval* (an alternative to the most popular *Wald interval*) which guarantees a better *coverage-probability* than Wald's

\[
[L, U] = \hat{p} + \frac{1}{2N} z_{1-\frac{\alpha}{2}}^2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z_{1-\frac{\alpha}{2}}^2}{4N^2}}
\]

**sample size**

\[
N \geq z_{1-\frac{\alpha}{2}}^2 \frac{\hat{p}(1 - \hat{p}) - 2\epsilon^2}{\sqrt{\hat{p}^2 (1 - \hat{p})^2 + 4\epsilon^2 (\hat{p} - 0.5)^2}}
\]
The Iterative Wilson Procedure

**Wilson Procedure**

1. $N_{tot} = 0$
2. $N = \text{Wilson}_\text{sample}(\hat{p}_0, \epsilon, \alpha)$
3. Perform $N$ experiments ($\sigma \models \phi$)
4. $N_{tot} = N_{tot} + N; \hat{p} = \frac{|YES|}{N_{tot}}$
5. if $\hat{p} \leq 0.5$
6. $p' = \hat{p} + \epsilon$
7. else $p' = \hat{p} - \epsilon$
8. $N' = \text{Wilson}_\text{sample}(p', \epsilon, \alpha); N_{new} = N' - N_{tot}$
9. if $N_{new} > 0$
10. $N = N_{new};$ goto 3
11. else return $\hat{p}$ and $\text{Wilson}_\text{interval}(\hat{p}, N_{tot}, \alpha)$

**property:** the number of iterations depends on the chosen initial estimate $\hat{p}_0$

- **Conservative approach** ($\hat{p}_0 = 0.5$): worst case
- **Iterative approach** ($\hat{p}_0 = 1$ or $\hat{p}_0 = 0$): better case
- **Minimum sample size approach** ($\hat{p}_0 =$ actual $p$): best case
Performance of Wilson interval method

result: the iterative approach outperforms the (most popular) conservative approach particularly with extreme (actual) probability (up to 88% fewer iterations required)
Software Prototype: parallel on-the-fly verification

(c) Wilson-driven computational (d) Inter-process and client-server loop communications
Case study: the Cell-Cycle regulatory network

**CELL CYCLE**

- **G1** (growth)
- **S** (synthesis)
- **G2** (growth)
- **M** (mythosis) division

**CELL-CYCLE control network**

**Simulation of X, Y, and A species**

<table>
<thead>
<tr>
<th>Cdk/CycB (X)</th>
<th>APC/Cdh1 (Y, Y_in)</th>
<th>Cdc20 (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>k_1α X</td>
<td>∅</td>
</tr>
<tr>
<td>X</td>
<td>k_2 X → ∅</td>
<td>k_5α A</td>
</tr>
<tr>
<td>X + Y</td>
<td>k_2'' α X + Y → Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_in → Y</td>
<td>X → X + A</td>
</tr>
<tr>
<td></td>
<td>k_3 Y_in → Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_in + a → Y + A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>k_3'' Y_in → Y + A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>k_4 X + Y → X + Y_in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>k_6 A → ∅</td>
<td></td>
</tr>
</tbody>
</table>
Statistical Verification of Cell-Cycle model

- Relevant property: “the influence of $A$ (Cdc20) on $Y(\text{APC})$ during the $S/G2/M$ transition"

$$\phi \equiv (A \leq 4) \cup_{\leq t} (Y \geq 5)$$

![Graph showing probability vs. time with exact values and point estimates]
Threshold Problem vs Approximation Problem

point: what statistical MC approach is better, Threshold Problem (Hypothesis Testing) or Approximation Problem (Error bounded estimation/Confidence Interval)?

- Hypothesis Testing MC is generally cheaper to run (smaller sample size)
- however Hypothesis Testing MC does not provide a probability measure

it depends on the application

- Hypothesis Testing is preferable if strict Safety/Dependability constraints are known
  - e.g. “the prob. of a breakdown must not exceed $10^{-2}$”

- Approximation based MC is preferable for applications which require a lot of SENSITIVITY ANALYSIS
  - e.g. “to check the effect of a model’s parameter (e.g. a rate of a transition) has on a given property $\phi$”

this is the case for example in Biological Modeling
Future Work, Open Issues

- CSL-TA statistical model checking (starting now)
  - Approximation problem
  - Threshold problem?

- Error bound approximation vs Confidence Interval approximation
  - how do they compare?

- Time-unbounded properties: verification through finite length experiments

- Combination of Numerical and Statistical methods as a way to improve performances