30 years of Statistical Model Checking

Kim. G. Larsen and Axel Legay

[1] Aalborg University, Denmark
[2] INRIA Rennes – Bretagne Atlantique, France

Abstract. This short note introduces statistical model checking and gives a brief overview of the *Statistical Model Checking, past present and future* session at ISOLA 2020.

1 Context

Quantitative properties of stochastic systems are usually specified in logics that allow one to compare the measure of executions satisfying certain temporal properties with thresholds. The model checking problem for stochastic systems with respect to such logics is typically solved by a numerical approach [BHHK03,CG04] that iteratively computes (or approximates) the exact measure of paths satisfying relevant subformulas; the algorithms themselves depend on the class of systems being analysed as well as the logic used for specifying the properties.

Another approach to solve the model checking problem is to *simulate* the system for finitely many runs, and use *hypothesis testing* to infer whether the samples provide *statistical* evidence for the satisfaction or violation of the specification. This approach was first applied in [LS91], where it was shown that hypothesis testing could be used to settle probabilistic modal logic properties with arbitrary precision, leading in the limit to probabilistic bisimulation. More recently [You05a] this approach has been known as statistical model checking (SMC) and is based on the notion that since sample runs of a stochastic system are drawn according to the distribution defined by the system, they can be used to obtain estimates of the probability measure on executions. Starting from time-bounded PCTL properties [You05a], the technique has been extended to handle properties with unbounded until operators [SVA05b], as well as to black-box systems [SVA04,You05a]. Tools, based on this idea have been built [HLMP04,SVA05a,You05a,You05b,BDD$^+$11,DLL$^+$11,BCLS13], and have been used to analyse many systems that are intractable numerical approaches.

The SMC approach enjoys many advantages. First, the algorithms require only that the system be simulatable (or rather, sample executions be drawn according to the measure space defined by the system). Thus, it can be applied to larger class of systems than numerical model checking algorithms, including black-box systems and infinite state systems. In particular, SMC avoids the ‘state explosion problem’ [CES09]. Second the approach can be generalized to a larger class of properties, including Fourier transform based logics. Third, SMC requires many independent simulation runs, making it easy to parallelise and scale to industrial-sized systems.

While it offers solutions to some intractable numerical model checking problems, SMC also introduces some additional problems. First, SMC only provides probabilistic guarantees about the correctness of the results. Second, the required sample size grows quadratically with respect to the required confidence of the result. This makes rare properties difficult to verify. Third, only the simulation of purely probabilistic systems is well defined. Nondeterministic systems, which are common in the field of formal verification, are especially challenging for SMC.

2 On Statistical Model Checking

Consider a stochastic system $\mathcal{S}$ and a logical property φ that can be checked on finite executions of the system. Statistical Model Checking (SMC) refers to a series of simulation-based techniques that can be used to answer two questions: (1) *Qualitative*: Is the probability for $\mathcal{S}$ to satisfy φ greater or equal to a certain threshold? and (2) *Quantitative*: What is the probability for $\mathcal{S}$ to satisfy φ? In contrast to numerical approaches, the answer is given up to some correctness precision.

In the sequel, we overview two SMC techniques. Let B_i be a discrete random variable with a Bernoulli distribution of parameter p. Such a variable can only take 2 values 0 and 1 with $Pr[B_i = 1] = p$ and $Pr[B_i = 0] = 1 - p$. In our context, each variable B_i is associated with one simulation of the system. The outcome for B_i, denoted b_i, is 1 if the simulation satisfies φ and 0 otherwise.

Qualitative Answer The main approaches [You05a,SVA04] proposed to answer the qualitative question are based on *sequential hypothesis testing* [Wal45]. Let $p = Pr(\varphi)$. To determine whether $p \geq \theta$, we can test $H : p \geq \theta$ against $K : p < \theta$. A test-based solution does not guarantee a correct result but it is possible to bound the probability of error. The *strength* of a test is determined by two parameters, α and β, such that the probability of accepting K (respectively, H) when H (respectively, K) holds, called a Type-I error (respectively, a Type-II error) is less or equal to α (respectively, β). A test has *ideal performance* if the probability of the Type-I error (respectively, Type-II error) is exactly α (respectively, β). However, these requirements make it impossible to ensure a low probability for both types of errors simultaneously (see [Wal45,You05a] for details). A solution is to use an *indifference region* $[p_1, p_0]$ (given some δ, $p_1 = \theta - \delta$ and $p_0 = \theta + \delta$) and to test $H_0 : p \geq p_0$ against $H_1 : p \leq p_1$. We now sketch the Sequential Probability Ratio Test (SPRT). In this algorithm, one has to choose two values A and B ($A > B$) that ensure that the strength of the test is respected. Let m be the number of observations that have been made so far. The test is based on the following quotient:

$$\frac{p_{1m}}{p_{0m}} = \prod_{i=1}^{m} \frac{Pr(B_i = b_i \mid p = p_1)}{Pr(B_i = b_i \mid p = p_0)} = \frac{p_1^{d_m}(1 - p_1)^{m - d_m}}{p_0^{d_m}(1 - p_0)^{m - d_m}},$$

where $d_m = \sum_{i=1}^{m} b_i$. The idea is to accept H_0 if $\frac{p_{1m}}{p_{0m}} \geq A$, and H_1 if $\frac{p_{1m}}{p_{0m}} \leq B$. The algorithm computes $\frac{p_{1m}}{p_{0m}}$ for successive values of m until either H_0 or H_1 is satisfied. This has the advantage of minimizing the number of simulations required to make the decision.

Quantitative Answer In [HLMP04] Peyronnet et al. propose an estimation procedure to compute the probability p for $\mathcal{S}$ to satisfy φ. Given a *precision* δ, the *Chernoff bound* of [Oka59] is used to compute a value for p' such that $|p'-p|\leq\delta$ with *confidence* $1-\alpha$. Let $B_1 \ldots B_m$ be m Bernoulli random variables with parameter p, associated to m simulations of the system considering φ. Let $p' = \sum_{i=1}^{m} b_i/m$, then the Chernoff bound [Oka59] gives $Pr(|p'-p| \geq \delta) \leq 2e^{-2m\delta^2}$. As a consequence, if we take $m = \lceil \ln(2/\alpha)/(2\delta^2) \rceil$, then $Pr(|p'-p|\leq\delta) \geq 1-\alpha$.

3 Content of the session

SMC has been implemented in prototypes/tools, which includes Uppaal-SMC [DLL+11], Plasma [BCLS13], Ymer [You05b], or COSMOS [BDD+11]. Those tools have been applied to several complex problems coming from a wide range of areas. This includes systems biology (see e.g., [Zul14]), automotive and avionics (see e.g., [BBB+12]), energy-centric systems(see e.g., [DDL+13]), or power grids(see e.g., [HH13]).

This year, the session is mostly focused on real-world applications of SMC on emerging industrial and societal topics such as automotive or COVID19. Some papers consider the integration of blackbox aspects, while others focus on combining the SMC approach with machine learning. It is worth observing that SMC reached a level of maturity that is sufficient for the approach to be integrated to real-life projects such as WABLIEFT or BEOCOVID presented hereafter.The program includes the following contributions.

- In [AKW20], the authors discuss the advantages gained when SMC is applied to white-box systems, utilizing the knowledge of their internals. The authors focus on the setting of unbounded-horizon properties such as reachability or LTL. The suggested approach is compared to other statistical and numerical techniques both conceptually as instantiations of the same framework, and experimentally. It not only clearly preserves scalability advantages of black-box SMC compared to classical model checking (while providing high level of guarantees), but it also scales yet better than either of the two for a wide class of models.
- In [GESL20], the authors focus on Mission planning. This is one of the crucial problems in the design of Multi-Agent Systems (MAS) because it requires the agents to calculate collision-free paths and efficiently schedule their tasks. The complexity of this problem greatly increases when realistic assumptions are includied, e.g. such as increase in number of agents, and timing requirements, as well as the stochastic behavior of the agents. In the paper, the authors propose a novel method that integrates statistical

model checking and reinforcement learning to overcome those difficulties. Additionally, the authors employ hybrid automata to model the continuous movement of agents and moving obstacles, and estimate the possible delay of the agents' traveling time when facing unpredictable obstacles, in order to synthesize mission plans that are statistically optimal. The authors show the result of synthesizing mission plans, analyzing bottlenecks, and the re-planning ability of the method in case of the sudden appearance of pedestrians by modeling and verifying a real industrial use case using UPPAAL SMC.

- In [BtBG+20], the authors provide a brief comparison of the modelling and analysis capabilities of two different formalisms and their associated simulation-based tools, acquired from experimenting with these methods and tools on one specific case study. The case study is a cyber-physical system from an industrial railway project, namely a railroad switch heater, and the quantitative properties concern energy consumption and reliability. The authors model and analyse the case study with stochastic activity networks and Möbius on the one hand and with stochastic hybrid automata and Uppaal SMC on the other hand. The authors give an overview of the performed experiments and highlight specific features of the two methodologies. This yields usefull pointers for future research and improvements.
- In [JLM+20], the authors present the BEOCOVID project. During the spring of 2020, the BEOCOVID project has been funded to investigate the use of stochastic hybrid models, statistical model checking and machine learning to anlyse, predict and control the rapid spreading of COVID19. The overall aim of BEOCOVID is to support government in decision making. In the paper the authors focus on the SEIRH epidemiological model instance of COVID19 pandemics and show how the risk of viral exposure, the impact of super-spreader events as well as other scenarios can be modelled, estimated and controlled in a variety of ways using the tool UPPAAL SMC.
- In [BLGW20], the authors present the WABLIEFT project. This project explores how to improve medical service delivery through a shared marketplace for service providers. This shared marketplace allows patients to choose services from providers and so support improved service delivery and patient satisfaction. Having a shared marketplace raises some service reliability and correctness challenges, as well as creates opportunities for improved information gathering. This work formalises the shared marketplace to prove correct behaviour and properties of the marketplace behaviour. The information available to the shared marketplace is also used to improve predictions of medical scenarios such as pandemics, and thus improve service delivery.

References

[AKW20] P Ashok, J. Kretinsky, and M. Weininger. Statistical model checking: Black or white? 2020. In this volume.

[BBB+12] Ananda Basu, Saddek Bensalem, Marius Bozga, Benoît Delahaye, and Axel Legay. Statistical abstraction and model-checking of large heterogeneous systems. *STTT*, 14(1):53–72, 2012.

[BCLS13] Benoit Boyer, Kevin Corre, Axel Legay, and Sean Sedwards. PLASMA-lab: A flexible, distributable statistical model checking library. In Kaustubh Joshi, Markus Siegle, Marielle Stoelinga, and PedroR. D'Argenio, editors, *Quantitative Evaluation of Systems*, volume 8054 of *Lecture Notes in Computer Science*, pages 160–164. Springer Berlin Heidelberg, 2013.

[BDD+11] Paolo Ballarini, Hilal Djafri, Marie Duflot, Serge Haddad, and Nihal Pekergin. Cosmos: A statistical model checker for the hybrid automata stochastic logic. In *QEST*, pages 143–144. IEEE Computer Society, 2011.

[BHHK03] Christel Baier, Boudewijn R. Haverkort, Holger Hermanns, and Joost-Pieter Katoen. Model-checking algorithms for continuous-time markov chains. *IEEE*, 29(6):524–541, 2003.

[BLGW20] E. Baranov, A. Legay, and T. Given-Wilson. Improving secure and robust patient service delivery. 2020. In this volume.

[BtBG+20] D. Basile, M. ter Beek, F. Di. Giandomenico, S. Gnesi, A. Fantechi, and G. Spagnolo. 30 years of simulation-based quantitative analysis tools: a comparison experiment between möbius and uppaal smc. 2020. In this volume.

[CES09] E. M. Clarke, E. A. Emerson, and J. Sifakis. Model checking: algorithmic verification and debugging. *Commun. ACM*, 52(11):74–84, November 2009.

[CG04] F. Ciesinski and M. Größer. On probabilistic computation tree logic. In *Validation of Stochastic Systems*, LNCS, 2925, pages 147–188. Springer, 2004.

[DDL+13] Alexandre David, Dehui Du, Kim Guldstrand Larsen, Axel Legay, and Marius Mikucionis. Optimizing control strategy using statistical model checking. In *NASA Formal Methods*, volume 7871 of *Lecture Notes in Computer Science*, pages 352–367. Springer, 2013.

[DLL+11] Alexandre David, Kim G. Larsen, Axel Legay, Marius Mikucionis, and Zheng Wang. Time for statistical model checking of real-time systems. In *CAV*, volume 6806 of *Lecture Notes in Computer Science*, pages 349–355. Springer, 2011.

[GESL20] R. Gu, E. Enoiu, C. Seceleanu, and K. Lundqvist. Probabilistic mission planning and analysis for multi-agent systems. 2020. In this volume.

[HH13] Holger Hermanns and Arnd Hartmanns. An internet inspired approach to power grid stability. *it - Information Technology*, 55(2):45–51, 2013.

[HLMP04] Thomas Hérault, Richard Lassaigne, Frédéric Magniette, and Sylvain Peyronnet. Approximate probabilistic model checking. In *VMCAI*, volume 2937 of *LNCS*, pages 73–84. Springer, 2004.

[JLM+20] P. Jensen, K. G. Larsen, M. Mikuconis, M. Muniz, D. Poulsen, and K. Jorgensen. Fluid model-checking in uppaal for covid-19. 2020. In this volume.

[LS91] Kim Guldstrand Larsen and Arne Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94(1):1–28, 1991.

[Oka59] Masashi Okamoto. Some inequalities relating to the partial sum of binomial probabilities. *Annals of the Institute of Statistical Mathematics*, 10:29–35, 1959.

[SVA04] Koushik Sen, Mahesh Viswanathan, and Gul Agha. Statistical model checking of black-box probabilistic systems. In *CAV*, LNCS 3114, pages 202–215. Springer, 2004.

[SVA05a] K. Sen, M. Viswanathan, and G. A. Agha. Vesta: A statistical model-checker and analyzer for probabilistic systems. In *QEST*, pages 251–252. IEEE Computer Society, 2005.

[SVA05b] Koushik Sen, Mahesh Viswanathan, and Gul Agha. On statistical model checking of stochastic systems. In *CAV*, LNCS 3576, pages 266–280, 2005.

[Wal45] A. Wald. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16(2):117–186, 1945.

[You05a] Håkan L. S. Younes. *Verification and Planning for Stochastic Processes with Asynchronous Events*. PhD thesis, Carnegie Mellon, 2005.

[You05b] Håkan L. S. Younes. Ymer: A statistical model checker. In *CAV*, volume 3576 of *LNCS*, pages 429–433. Springer, 2005.

[Zul14] Paolo Zuliani. Statistical model checking for biological applications. *CoRR*, abs/1405.2705, 2014.