Statistical Model Checking Past, Present, and Future (Track Introduction)

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 2014.

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

T. Margaria and B. Steffen (Eds.): ISoLA 2014, Part II, LNCS 8803, pp. 135–142, 2014.

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 $Pr(|p'-p| \leq \delta) \geq 1-\alpha$.

2.1 Rare Events

Statistical model checking avoids the exponential growth of states associated with probabilistic model checking by estimating probabilities from multiple executions of a system and by giving results within confidence bounds. Rare properties are often important but pose a particular challenge for simulation-based approaches, hence a key objective for SMC is to reduce the number and length of simulations necessary to produce a result with a given level of confidence. In the literature, one finds two techniques to cope with rare events: *importance sampling* and *importance splitting.*

In order to minimize the number of simulations, importance sampling (see e.g., [Rid10, DBNR00]) works by estimating a probability using weighted simulations that favour the rare property, then compensating for the weights. For importance sampling to be efficient, it is thus crucial to find good importance sampling distributions without considering the entire state space. In [CZ11] Zuliani and Clarke outlined the challenges for SMC and rare-events. A first theory contribution was then provided by Barbot et al. who proposed to use reduction techniques together with cross-entropy [BHP12]. In [JLS12], we presented a simple algorithm that uses the notion of cross-entropy minimisation to find an optimal importance sampling distribution. In contrast to previous work, our algorithm uses a naturally defined low dimensional vector of parameters to specify this distribution and thus avoids the intractable explicit representation of a transition matrix. We show that our parametrisation leads to a unique optimum and can produce many orders of magnitude improvement in simulation efficiency.

One of the open challenges with importance sampling is that the variance of the estimator cannot be usefully bounded with only the knowledge gained from simulation. Importance *splitting* (see e.g., [CG07]) achieves this objective by estimating a sequence of conditional probabilities, whose product is the required result. In [JLS13] motivated the use of importance splitting for statistical model checking and were the first to link this standard variance reduction technique [KM53] with temporal logical. In particular, they showed how to create *score functions* based on logical properties, and thus define a set of *levels* that delimit the conditional probabilities. In [JLS13] they also described the necessary and desirable properties of score functions and levels, and gave two importance splitting algorithms: one that uses fixed levels and one that discovers optimal levels adaptively.

2.2 Nondeterminism

Markov decision processes (MDP) and other nondeterministic models interleave nondeterministic *actions* and probabilistic transitions, possibly with rewards or costs assigned to the actions [Bel57, Put94]. These models have proved useful in many real optimisation problems (see [Whi85, Whi88, Whi93] for a survey of applications of MDPs) and are also used in a more abstract sense to represent concurrent probabilistic systems (e.g., [BDA95]). Such systems comprise probabilistic subsystems whose transitions depend on the states of the other subsystems, while the order in which concurrently enabled transitions execute is nondeterministic. This order may radically affect the expected reward or the probability that a system will satisfy a given property. Numerical model checking may be used to calculate the upper and lower bounds of these quantities, but a simulation semantics is not immediate for nondeterministic systems and SMC is therefore challenging.

SMC cannot be applied to nondeterministic systems without first resolving the nondeterminism using a *scheduler* (alternatively a *strategy* or a *policy*). Since nondeterministic and probabilistic choices are interleaved, schedulers are typically of the same order of complexity as the system as a whole and may be infinite.

In [LS14] Jegouret et al presented the basis of the first lightweight SMC algorithms for MDPs and other nondeterministic models, using an $\mathcal{O}(1)$ representation of history-dependent schedulers. This solution is based on pseudorandom number generators and an efficient hash function, allowing schedulers to be sampled using Monte Carlo techniques. Some previous attempts to apply SMC to nondeterministic models [BFHH11, LP12, HMZ$^+$12, HT13] have been memory-intensive (heavyweight) and incomplete in various ways. The algorithms of [BFHH11, HT13] consider only systems with 'spurious' nondeterminism that does not actually affect the probability of a property. In [LP12] the authors consider only memoryless schedulers and do not consider the standard notion of optimality used in model checking (i.e., with respect to probability). The algorithm of [HMZ$^+$12] addresses a standard qualitative probabilistic model checking problem, but is limited to memoryless schedulers that must fit into memory and does not in general converge to the optimal scheduler. Most recently[DJL$^+$], SMC – or reinforcement learning – has been applied to learn near-cost-optimal strategies for priced timed stochastic games subject to guaranteed worst-case time bounds. The method is implemented using a combination of Uppaal-TIGA (for timed games) and Uppaal SMC and provides three alternatives light-weight datastructures for representing stochastic strategies.

3 Content of the Session

SMC has been implemented in several prototypes and 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 isola session discusses several aspects of SMC, which includes: non-determinism, rare-events, applications to biology/energy-centric/power grids, and runtime verification procedures suited for SMC. Summary of the contributions:

- In [JLS14], the authors propose new SMC techniques for rare-events. The main contribution is in extending [JLS13] with new adaptive level algorithms based on branching simulation, and to show that this permits to get a more precise estimate of the rare probability. In the authors stud
- In [BLT14], the authors propose to apply simulation-based techniques to systems whose number of configurations can vary at execution. They develop a new logic and new SMC techniques for such systems.
- In [BNB+14], the authors provide a complete and detailed comparison of several SMC model checkers, especially for the real-time setting. The authors present five semantic caveats and give a classification scheme for SMC algorithms. They also argue that caution is needed and believe that the caveats and classification scheme in this paper serve as a guiding reference for thoroughly understanding them.
- In [BGGM14], the authors propose an application of SMC to systems biology. More precisely, they consider the Wnt-beta-catenin signaling pathway that plays an important role in the proliferation of neural cells. They analyze the dynamics of the system by combining SMC with the Hybrid Automata Stochastic Logic. in
- In [WHL14], the authors consider wireless systems such as satellites and sensor networks are often battery-powered. The main contributions are (1) to show how SMC can be used to calculate an upper bound on the attainable number of task instances from a battery, and (2) to synthesize a battery-aware scheduler that wastes no energy on instances that are not guaranteed to make their deadlines.
- In [GPS+14], the authors consider CTL-based measuring on paths, and generalize the mea- surement results to the full structure using optimal Monte Carlo estimation techniques. To experimentally validate their framework, they present LTL-based measurements for a flocking model of bird-like agents.
- In [EHFJ+14], the authors explore the effectiveness and challenges of using monitoring techniques, based on Aspect-Oriented Programming, to block adware at the library level, on mobile devices based on Android using miAdBlocker. The authors also present the lessons learned from this experience,

and we identify some challenges when applying runtime monitoring techniques to real-world case studies.
- In [Hav14], the author presents a form of automaton, referred to as data automata, suited for monitoring sequences of data-carrying events, for example emitted by an executing software system. He also presents and evaluate an optimized external DSL for data automata, as well as a comparable unoptimized internal DSL (API) in the Scala programming language, in order to compare the two solutions.

References

[BBB+12] Basu, A., Bensalem, S., Bozga, M., Delahaye, B., Legay, A.: Statistical abstraction and model-checking of large heterogeneous systems. STTT 14(1), 53–72 (2012)

[BCLS13] Boyer, B., Corre, K., Legay, A., Sedwards, S.: PLASMA-lab: A flexible, distributable statistical model checking library. In: Joshi, K., Siegle, M., Stoelinga, M., D'Argenio, P.R. (eds.) QEST 2013. LNCS, vol. 8054, pp. 160–164. Springer, Heidelberg (2013)

[BDA95] Bianco, A., De Alfaro, L.: Model checking of probabilistic and nondeterministic systems. In: Thiagarajan, P.S. (ed.) FSTTCS 1995. LNCS, vol. 1026, pp. 499–513. Springer, Heidelberg (1995)

[BDD+11] Ballarini, P., Djafri, H., Duflot, M., Haddad, S., Pekergin, N.: Cosmos: A statistical model checker for the hybrid automata stochastic logic. In: QEST, pp. 143–144. IEEE Computer Society (2011)

[Bel57] Bellman, R.: Dynamic Programming. Princeton University Press (1957)

[BFHH11] Bogdoll, J., Ferrer Fioriti, L.M., Hartmanns, A., Hermanns, H.: Partial order methods for statistical model checking and simulation. In: Bruni, R., Dingel, J. (eds.) FMOODS/FORTE 2011. LNCS, vol. 6722, pp. 59–74. Springer, Heidelberg (2011)

[BGGM14] Ballarini, P., Gallet, E., Le Gall, P., Manceny, M.: Formal analysis of the wnt/β-catenin pathway through statistical model checking. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 193–207. Springer, Heidelberg (2014)

[BHHK03] Baier, C., Haverkort, B.R., Hermanns, H., Katoen, J.-P.: Model-checking algorithms for continuous-time markov chains. IEEE 29(6), 524–541 (2003)

[BHP12] Barbot, B., Haddad, S., Picaronny, C.: Coupling and importance sampling for statistical model checking. In: Flanagan, C., König, B. (eds.) TACAS 2012. LNCS, vol. 7214, pp. 331–346. Springer, Heidelberg (2012)

[BLT14] Boyer, B., Legay, A., Traonouez, L.-M.: A formalism for stochastic adaptive systems. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 160–176. Springer, Heidelberg (2014)

[BNB+14] Bohlender, D., Bruintjes, H., Junges, S., Katelaan, J., Nguyen, V.Y., Noll, T.: A review of statistical model checking pitfalls on real-time stochastic models. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 177–192. Springer, Heidelberg (2014)

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

[CG04] Ciesinski, F., Größer, M.: On probabilistic computation tree logic. In: Baier, C., Haverkort, B.R., Hermanns, H., Katoen, J.-P., Siegle, M. (eds.) Validation of Stochastic Systems. LNCS, vol. 2925, pp. 147–188. Springer, Heidelberg (2004)

[CG07] Cérou, F., Guyader, A.: Adaptive multilevel splitting for rare event analysis. Stochastic Analysis and Applications 25, 417–443 (2007)

[CZ11] Clarke, E.M., Zuliani, P.: Statistical model checking for cyber-physical systems. In: Bultan, T., Hsiung, P.-A. (eds.) ATVA 2011. LNCS, vol. 6996, pp. 1–12. Springer, Heidelberg (2011)

[DBNR00] De Boer, P.-T., Nicola, V.F., Rubinstein, R.Y.: Adaptive importance sampling simulation of queueing networks. In: Winter Simulation Conference, vol. 1, pp. 646–655 (2000)

[DDL+13] David, A., Du, D., Guldstrand Larsen, K., Legay, A., Mikučionis, M.: Optimizing control strategy using statistical model checking. In: Brat, G., Rungta, N., Venet, A. (eds.) NFM 2013. LNCS, vol. 7871, pp. 352–367. Springer, Heidelberg (2013)

[DJL+] David, A., Jensen, P.G., Larsen, K.G., Legay, A., Lime, D., Sorensen, M.G., Taankvist, J.H.

[DLL+11] David, A., Larsen, K.G., Legay, A., Mikučionis, M., Wang, Z.: Time for statistical model checking of real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 349–355. Springer, Heidelberg (2011)

[EHFJ+14] El-Harake, K., Falcone, Y., Jerad, W., Langet, M., Mamlouk, M.: Blocking advertisements on android devices using monitoring techniques. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 239–253. Springer, Heidelberg (2014)

[GPS+14] Grosu, R., Peled, D., Ramakrishnan, C.R., Smolka, S.A., Stoller, S.D., Yang, J.: Using statistical model checking for measuring systems. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 223–238. Springer, Heidelberg (2014)

[Hav14] Havelund, K.: Monitoring with data automata. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 254–273. Springer, Heidelberg (2014)

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

[HLMP04] Hérault, T., Lassaigne, R., Magniette, F., Peyronnet, S.: Approximate probabilistic model checking. In: Steffen, B., Levi, G. (eds.) VMCAI 2004. LNCS, vol. 2937, pp. 73–84. Springer, Heidelberg (2004)

[HMZ+12] Henriques, D., Martins, J.G., Zuliani, P., Platzer, A., Clarke, E.M.: Statistical model checking for Markov decision processes. In: 2012 Ninth International Conference on Quantitative Evaluation of Systems, pp. 84–93. IEEE (2012)

[HT13] Hartmanns, A., Timmer, M.: On-the-fly confluence detection for statistical model checking. In: Brat, G., Rungta, N., Venet, A. (eds.) NFM 2013. LNCS, vol. 7871, pp. 337–351. Springer, Heidelberg (2013)

[JLS12] Jegourel, C., Legay, A., Sedwards, S.: Cross-Entropy Optimisation of Importance Sampling Parameters for Statistical Model Checking. In: Madhusudan, P., Seshia, S.A. (eds.) CAV 2012. LNCS, vol. 7358, pp. 327–342. Springer, Heidelberg (2012)

[JLS13] Jegourel, C., Legay, A., Sedwards, S.: Importance splitting for statistical model checking rare properties. In: Sharygina, N., Veith, H. (eds.) CAV 2013. LNCS, vol. 8044, pp. 576–591. Springer, Heidelberg (2013)

[JLS14] Jegourel, C., Legay, A., Sedwards, S.: An effective heuristic for adaptive importance splitting in statistical model checking. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 143–159. Springer, Heidelberg (2014)

[KM53] Kahn, H., Marshall, A.W.: Methods of Reducing Sample Size in Monte Carlo Computations. Operations Research 1(5), 263–278 (1953)

[LP12] Lassaigne, R., Peyronnet, S.: Approximate planning and verification for large Markov decision processes. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing, pp. 1314–1319. ACM (2012)

[LS91] Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. Inf. Comput. 94(1), 1–28 (1991)

[LS14] Legay, A., Sedwards, S.: Lightweight Monte Carlo verification of Markov decision processes (submitted, 2014)

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

[Put94] Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley-Interscience (1994)

[Rid10] Ridder, A.: Asymptotic optimality of the cross-entropy method for markov chain problems. Procedia Computer Science 1(1), 1571–1578 (2010)

[SVA04] Sen, K., Viswanathan, M., Agha, G.: Statistical model checking of black-box probabilistic systems. In: Alur, R., Peled, D.A. (eds.) CAV 2004. LNCS, vol. 3114, pp. 202–215. Springer, Heidelberg (2004)

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

[SVA05b] Sen, K., Viswanathan, M., Agha, G.: On statistical model checking of stochastic systems. In: Etessami, K., Rajamani, S.K. (eds.) CAV 2005. LNCS, vol. 3576, pp. 266–280. Springer, Heidelberg (2005)

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

[Whi85] White, D.J.: Real applications of Markov decision processes. Interfaces 15(6), 73–83 (1985)

[Whi88] White, D.J.: Further real applications of Markov decision processes. Interfaces 18(5), 55–61 (1988)

[Whi93] White, D.J.: A survey of applications of Markov decision processes. Journal of the Operational Research Society 44(11), 1073–1096 (1993)

[WHL14] Wognsen, E.R., Hansen, R.R., Larsen, K.G.: Battery-aware scheduling of mixed criticality systems. In: Margaria, T., Steffen, B. (eds.) ISoLA 2014, Part II. LNCS, vol. 8803, pp. 208–222. Springer, Heidelberg (2014)

[You05a] Younes, H.L.S.: Verification and Planning for Stochastic Processes with Asynchronous Events. PhD thesis, Carnegie Mellon (2005)

[You05b] Younes, H.L.S.: Ymer: A statistical model checker. In: Etessami, K., Rajamani, S.K. (eds.) CAV 2005. LNCS, vol. 3576, pp. 429–433. Springer, Heidelberg (2005)

[Zul14] Zuliani, P.: Statistical model checking for biological applications. CoRR, abs/1405.2705 (2014)