Statistical Model Checking the 2018 Edition!

Kim Guldstrand Larsen[1,2](✉) and Axel Legay[1,2]

[1] Aalborg University, Aalborg, Denmark
kgl@cs.aau.dk
[2] Inria Rennes – Bretagne Atlantique, Rennes, 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 2018. This is the fourth edition of the track at Isola.

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

T. Margaria and B. Steffen (Eds.): ISoLA 2018, LNCS 11245, pp. 261–270, 2018.
https://doi.org/10.1007/978-3-030-03421-4_17

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

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: rare-events, application to railway and security, sampling minimization, nondeterminism, or real-time extensions. Summary of the contributions:

– In [DJS18], the authors propose schemes to reduce the number of samplings needed for SMC to converge. It is known that Bayesian and rare event techniques can be used to reduce the sample size but they can not be applied without prerequisite or knowledge about the system under scrutiny. Recently, sequential algorithms based on Monte Carlo estimations and Massart bounds have been proposed to reduce the sample size while providing guarantees on error bounds which has been shown to outperform alternative frequentist approaches. In this work, the authors discuss some features regarding the distribution and the optimisation of these algorithms.
– In [Str18], the author focuses on processor based systems with interruptible executions. Their predictability analysis becomes more difficult especially when interrupts may occur at arbitrary times, suffer from arrival and servicing jitters, are subject to priorities, or may be nested and un/masked at run-time. Such a behavior of interrupts and executions has stochastic aspects and leads to the explosion of the number of situations to be considered. To cope with such a behavior, the author proposes a simulation model that relies on a network of stochastic timed automata and involves the above-mentioned behavioral aspects related to interrupts and executions. For a system, modeled by means of the automata, the author shows that the problem of analyzing its predictability may be efficiently solved by means of the statistical model checking.
– In [ALR18], the authors study incomplete stochastic systems that are missing some parts of their design, or are lacking information about some components. For such systems, it is interesting to get early analysis results of the requirements, in order to adequately refine their design. The main contribution of the paper takes the form of a three-valued temporal logics for which authors offer SMC algorithms.

- In [ABK18], The maximum reachability probabilities in a Markov decision process can be computed using value iteration (VI). Recently, simulation-based heuristic extensions of VI have been introduced, such as bounded real-time dynamic programming (BRTDP), which often manage to avoid explicit analysis of the whole state space while preserving guarantees on the computed result. In this paper, the authors introduce a new class of such heuristics, based on Monte Carlo tree search (MCTS), a technique celebrated in various machine-learning settings. They also provide a spectrum of algorithms ranging from MCTS to BRTDP. Finally, the authors evaluate these techniques and show that for larger examples, where VI is no more applicable, our techniques are more broadly applicable than BRTDP with only a minor additional overhead.
- In [BBL18], the authors propose to synthesize defense configurations to counter sophisticated attack strategies minimizing resource usage while ensuring a high probability of success. For this, they combine Statistical Model Checking techniques with Genetic Algorithms. Experiments performed on real-life case studies show substantial improvements compared to existing techniques.
- In [FGP18], authors discuss the application of SMC on assisted/automated driving systems. For developers of assisted or automated driving systems, gaining specific feedback and quantitative figures on the safety impact of the systems under development is crucial. However, obtaining such data from simulation of their design models is a complex and often time-consuming process. Especially when data of interest hinge on extremely rare events, an estimation of potential risks is highly desirable but a non-trivial task lacking easily applicable methods. In this paper, the authors describe how a quantitative statement for a risk estimation involving extremely rare events can be obtained by guiding simulation based on reinforcement learning. The method draws on variance reduction and importance sampling, yet applies different optimization principles than related methods, like the cross-entropy methods against which they compare
- In [DHS18], authors extend their work on Lightweight scheduler sampling. It is known that this approach brings statistical model checking to nondeterministic formalisms with undiscounted properties, in constant memory. Its direct application to continuous-time models is rendered ineffective by their dense concrete state spaces and the need to consider continuous input for optimal decisions. In this paper the authors describe the challenges and state of the art in applying lightweight scheduler sampling to three continuous-time formalisms: After a review of recent work on exploiting discrete abstractions for probabilistic timed automata, they discuss scheduler sampling for Markov automata and apply it on two case studies, and offer a large discussion on potential future research and extensions.
- In [BBC18], the authors present an experience in modelling and statistical model checking a satellite-based moving block signalling scenario from the railway industry with Uppaal SMC. This demonstrates the usability and applicability of Uppaal SMC in the railway domain. The authors also propose

a promising direction for future work, in which they envision spatio-temporal analysis with Uppaal SMC.
- In [BPS18], the authors discuss the link between SMC and Randomized testing. The latter is a lightweight approach for searching for bugs. There is a tradeoff between the number of testing experiments performed and the probability to find errors. An important challenge in random testing is when the errors that we try to detect are scattered with very low probability among the different executions, forming a "rare event". In this paper, the authors suggest the use of a "biasing automaton", which observes the tested sequence and controls the distribution of the different choices of extending it. Authors also show how to find such a biasing automaton based on genetic programming techniques.

4 Conclusion

This fourth edition of the SMC track at ISOLA shows two interesting points. The first is that researchers keep on working and produce new results on SMC core topics. This includes, e.g., nondeterminism or rare events. In addition, new challenges have emerged. This includes, e.g., a study comparison with testing, or the introduction of uncertainty. This gives motivation to keep on working on the topic and get ready for the fifth edition which will take place in 2020 where we will celebrate twenty years of SMC!

References

[ABK18] Ashok, P., Brazdil, T., Kretinsky, J., Slamecka, O.: Monte Carlo tree search for verifying reachability in Markov decision processes. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 322–335. Springer, Cham (2018)

[ALR18] Arora, S., Traonouez, L.-M., Legay, A., Richmond, T.: Statistical model-checking of incomplete stochastic systems. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 354–371. Springer, Cham (2018)

[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)

[BBC18] ter Beek, M., Basile, D., Ciancia, V.: Statistical model checking a moving block railway signalling scenario with Uppaal SMC. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 372–391. Springer, Cham (2018)

[BBL18] Mediouni, B.L., Nouri, A., Bozga, M., Legay, A., Bensalem, S.: Mitigating security risks through attack strategies exploration. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 392–413. Springer, Cham (2018)

[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). https://doi.org/10.1007/978-3-642-40196-1_12

[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). https://doi.org/10.1007/3-540-60692-0_70

[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, Princeton (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). https://doi.org/10.1007/978-3-642-21461-5_4

[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). https://doi.org/10.1007/978-3-642-28756-5_23

[BPS18] Peled, D., Bu, L., Shen, D., Tzirulnikov, Y.: Chasing errors using biasing automata. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 271–286. Springer, Cham (2018)

[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). https://doi.org/10.1007/978-3-540-24611-4_5

[CG07] Cérou, F., Guyader, A.: Adaptive multilevel splitting for rare event analysis. Stoch. Anal. Appl. **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). https://doi.org/10.1007/978-3-642-24372-1_1

[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). https://doi.org/10.1007/978-3-642-38088-4_24

[DHS18] D'Argenio, P., Hartmanns, A., Sedwards, S.: Lightweight statistical model checking in nondeterministic continuous time. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 336–353. Springer, Cham (2018)

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

[DJS18] Jegourel, C., Sun, J., Dong, J.-S.: On the sequential Massart algorithm for statistical model checking. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 287–304. Springer, Cham (2018)

[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). https://doi.org/10.1007/978-3-642-22110-1_27

[FGP18] Puch, S., Fraenzle, M., Gerwinn, S.: Quantitative risk assessment of safety-critical systems via guided simulation for rare events. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 305–321. Springer, Cham (2018)

[HH13] Hermanns, H., Hartmanns, A.: An internet inspired approach to power grid stability. IT Inf. Technol. **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). https://doi.org/10.1007/978-3-540-24622-0_8

[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). https://doi.org/10.1007/978-3-642-38088-4_23

[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). https://doi.org/10.1007/978-3-642-31424-7_26

[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). https://doi.org/10.1007/978-3-642-39799-8_38

[KM53] Kahn, H., Marshall, A.W.: Methods of reducing sample size in Monte Carlo computations. Oper. Res. **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] Guldstrand Larsen, K., Skou, A., Kim Guldstrand Larsen and Arne Skou: Bisimulation through probabilistic testing. Inf. Comput. **94**(1), 1–28 (1991)

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

[Oka59] Okamoto, M.: Some inequalities relating to the partial sum of binomial probabilities. Ann. Inst. Stat. Math. **10**, 29–35 (1959)

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

[Rid10] Ridder, A.: Asymptotic optimality of the cross-entropy method for Markov chain problems. Procedia Comput. Sci. **1**(1), 1571–1578 (2010)

[Str18] Strnadel, J.: Statistical model checking of processor systems in various interrupt scenarios. In: Margaria, T., Steffen, B. (eds.) ISoLA 2018. LNCS, vol. 11245, pp. 414–429. Springer, Cham (2018)

[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). https://doi.org/10.1007/978-3-540-27813-9_16

[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). https://doi.org/10.1007/11513988_26

[Wal45] Wald, A.: Sequential tests of statistical hypotheses. Ann. Math. Stat. **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. J. Oper. Res. Soc. **44**(11), 1073–1096 (1993)

[You05a] Younes, H.L.S.: Verification and planning for stochastic processes with asynchronous events. Ph.D. 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). https://doi.org/10.1007/11513988_43

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