Bayesian Statistical Model-Checking of Continuous Stochastic Logic

Autonomous systems are transforming the society by enabling sophisticated technologies such as robotic surgery and driverless cars. On one hand, increased automation through removal of the human-in-the-loop promises enhanced efficiency, while, on the other hand, the highly uncertain and safety critical environments, such as, varying weather and road conditions, and the presence of pedestrians on the road, pose challenge to the design of reliable autonomous systems. Hence, there is an immediate need for a robust framework for certifying the correctness of autonomous systems.

In this report, we explore verifying the correctness of uncertain autonomous systems modeled as discrete-time Markov chains (DTMCs) against correctness criteria provided as continuous stochastic logic (CSL) formulae. Statistical model-checking (SMC) is a paradigm for verification based on formulating the verification problem as a hypothesis testing problem. We propose a novel statistical model-checking algorithm based on Bayesian hypothesis testing. While Bayesian approaches for simpler logics without nested probabilistic operators and frequentist approaches for nested logic have been previously explored, the Bayesian approach for CSL that has nested probabilistic operators has not been addressed. The challenge in the nested case arises from the fact that unlike in probabilistic model-checking (PMC), where we obtain a definitive answer for the model-checking problem for the sub-formulae, we only obtain a correct answer with a certain confidence, which needs to be factored into the recursive SMC algorithm. We have implemented our algorithm in a Python toolbox and present our evaluation on some benchmark examples. We observe that while both the Bayesian and frequentist SMC perform well in terms of inference, Bayesian SMC is more efficient in terms of the number of samples and verification time. On several examples, it even beats the state-of-the-art probabilistic model-checker PRISM.