Developing Three-Risk-Proof Sequential Decision Making for Agricultural Decisions under Environmental Risks

Debabrota Basu and Odalric-Ambrym Maillard

debabrota.basu@inria.fr and odalric-ambrym.maillard@inria.fr

Where: Scool (previously SequeL), Inria Lille- Nord Europe, Villeneuve d'Ascq, France

Expected background: Master in CS, specialised in machine learning and mathematical statistics. This work will require the candidate to be comfortable with mathematical proofs. Knowledge in Python is a plus.

Keywords: Markov decision processes, Contextual bandits, Quantification of uncertainty, Risk measures, Regret bounds, Environmental risks.

Objective: In the machine learning literature on sequential decision making, taking into account some notions of risk has received increasing attention over the past decade (Geibel and Wysotzki, 2005; Maillard, 2013; Chow and Ghavamzadeh, 2014; Garcıa and Fernández, 2015; Prashanth and Fu, 2018; Leurent et al., 2020; Eriksson et al., 2021; Baudry et al., 2021).

One of the research direction is to consider an objective that is not to maximize an expected score, but rather take into account aversion about bad situations (Howard and Matheson, 1972; Prashanth and Fu, 2018; Baudry et al., 2021). Let us consider recommendation to harvest a crop (formalized under the framework of contextual bandits). Instead of a crop that causes the highest yield in expectation but is very sensitive to change in weather conditions, one would prefer a crop that has less yield in expectation but low probability of a bad yield under weather flucftutations. Hence, risk-averse criterion (e.g. Conditional Value at Risk (CVaR) (Rockafellar and Uryasev, 2000)) have received attention, due to this motivation and also the challenges opened by working with such criterion (Garcıa and Fernández, 2015; Prashanth and Fu, 2018).

Accommodating risk in decision making pose two main problems. Firstly, quantifying the uncertainties involved in the inherent dynamics of decision making and the numerical models driving the decision making algorithm into coherent risk measures (Mihatsch and Neuneier, 2002; Eriksson and Dimitrakakis, 2020; Eriksson et al., 2021). Secondly, the risk measures are often non-linear and thus the generic principles from classical (risk-neutral) literature are often suboptimal in risk-sensitive decision making. These issues demand significant effort to coherently quantify relevant uncertainties, design novel algorithms adapted to this notions of risk, and to reinvent the corresponding proof techniques. We aim to design algorithms that can take into account such risk quantifications and perform sequential decision making with optimal performance guarantees. This is an active research area with multitude of open problems to solve.

This project is organized around three complementary perspectives on the notion of risk in sequential decision making under uncertainty. The first notion is mainly depending on the agent's perception of risk, considering that different agents may have different risk-preferences upon facing similar situations. The second one is more inherent to the system, that is subject to changes, model misspecification and possibly corrupted input signal. This risk is typically caused by external causes not directly linked with the decision maker. Last, the third one is about the risk created by the actions on the system. Indeed, some actions may modify the system and lead it to states that are riskier than others.

Our goal is to provide novel and theoretically sound sequential decision making algorithms to handle such diverse notions of risks, that can be used in the application context of forest management and possibly beyond.

We will commence our investigation with the contextual bandit framework (Li et al., 2010) of sequential decision making. If we are successful in designing such risk-proof algorithms for contextual bandits, the next goal will be to extend this work to Markov decision processes (Garcıa and Fernández, 2015). Realisation of this part of the project will be decided depending on the developments in the project.

Project Outcome: In this project, we follow the philosophy of open science. We aim to publish the research results generated in this process in open-access platforms (arXiv, hal) and as well as in premier machine learning (AAAI, AISTATS, ICML, IJCAI, NeurIPS, COLT) venues with open-access proceedings. Along with the research papers, we expect that the different phases of the project will lead to open-source software. As the problem under investigation is practical, an open-source code demonstrating the usefulness of the developed algorithm is desired. All the softwares will be developed and available through the gitlab platform hosted by Inria.

Special Remark: This project is a collaboration between équipe Scool of Inria Lille and Applied Mathematics and Computer Science laboratory of INRAE Toulouse. Thus, the candidate will collaborate actively with researchers in INRAE Toulouse.

References

Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. (2021). Optimal thompson sampling strategies for support-aware cvar bandits. In *International Conference on Machine Learning*, pages 716–726. PMLR.

Chow, Y. and Ghavamzadeh, M. (2014). Algorithms for cvar optimization in mdps. In *Advances in neural information processing systems*, pages 3509–3517.

Eriksson, H., Basu, D., Alibeigi, M., and Dimitrakakis, C. (2021). Sentinel: Taming uncertainty with ensemble-based distributional reinforcement learning. *arXiv preprint arXiv:2102.11075.*

Eriksson, H. and Dimitrakakis, C. (2020). Epistemic risk-sensitive reinforcement learning. In *ESANN*, pages 339–344.

Garcıa, J. and Fernández, F. (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480.

Geibel, P. and Wysotzki, F. (2005). Risk-sensitive reinforcement learning applied to control under constraints. *Journal of Artificial Intelligence Research*, 24:81–108.

Howard, R. A. and Matheson, J. E. (1972). Risk-sensitive markov decision processes. *Management science*, 18(7):356–369.

Leurent, E., Efimov, D., and Maillard, O.-A. (2020). Robust-adaptive control of linear systems: beyond quadratic costs. In *Advances in Neural Information Processing Systems 33*. Curran Associates, Inc.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670.

Maillard, O.-A. (2013). Robust risk-averse stochastic multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pages 218–233. Springer.

Mihatsch, O. and Neuneier, R. (2002). Risk-sensitive reinforcement learning. *Machine learning*, 49(2-3):267–290.

Prashanth, A. and Fu, M. (2018). Risk-sensitive reinforcement learning: A constrained optimization viewpoint. *arXiv e-prints*, pages arXiv–1810.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42.