MSc Project: ϵ-Best Mixture Identification

Supervisor: Wouter M. Koolen

Keywords: Sequential Decision Making, Bandits, Pure Exploration, Sample Complexity, Mixture Policy Identification

Background: Pure Exploration is an active area of Machine Learning and Statistics. Its central problem of Best Arm Identification has been studied intensively since (Even-Dar, Mannor, and Mansour, 2002), and worst-case optimal methods have been developed for the fixed confidence, fixed budget and simple regret settings (Bubeck, Munos, and Stoltz, 2011). After a long and respectable series of papers establishing worst-case optimality, a revolutionary new approach called Track-and-Stop was pioneered by Garivier and Kaufmann (2016) that delivers instance-optimal methods. Since then, several aspects of Track-and-Stop have been generalised and refined: tighter stopping thresholds were constructed by Kaufmann and Koolen (2021), computational efficiency was improved using saddle-point methods by Degenne, Koolen, and M´enard (2019), and problems with multiple answers were analysed by Degenne and Koolen (2019). A recent extension with subpopulations was proposed by Russac et al. (2021).

Academic Content: The typical pure exploration task is identification of the best arm from samples. Letting μk denote the mean of arm k, the task is to identify arg $\max_k \mu k$. The sample complexity for this problem is well studied. One may reduce the sample complexity — at the cost of incurring some approximation error ϵ — by asking for identification of any ϵ-best arm $k \in \{k \mid \mu k \geq \max k\ \mu k - \epsilon\}$.

The starting point of this project is the realisation that for many applications it is enough to identify an ϵ-optimal mixture policy. That is, we are happy with any probability distribution p

$$\left\{ p \in \triangle_K \middle| \sum_{k=1}^{K} p_k \mu_k \geq \max_k \mu_k - \epsilon \right\}.$$

One may think of p as a randomised policy that is close to optimal. This could find application e.g. when applying bandit techniques to reinforcement learning problems, in particular near-optimal policy identification in MDPs.

This project will investigate these questions: Is the sample complexity of ϵ-best mixture identification strictly lower than that of ϵ-best arm identification? If so, what is the sample complexity of ϵ-best mixture identification? And how can one design efficient algorithms for ϵ-best mixture identification? Possible algorithms include elimination methods based on confidence-intervals, Track-and-Stop and Bayesian-flavoured approaches (Kaufmann, Koolen, and Garivier, 2018).

The project is envisaged to consist of mostly theoretical work, with only minor computational and empirical components.

References

Bubeck, S., R. Munos, and G. Stoltz (2011). “Pure Exploration in Finitely Armed and Continuous Armed Bandits”. In: Theoretical Computer Science 412, 1832-1852 412, pp. 1832–1852.

Degenne, R. and W. M. Koolen (Dec. 2019). “Pure Exploration with Multiple Correct Answers”. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 14591–14600.

Degenne, R., W. M. Koolen, and P. M´enard (Dec. 2019). “Non-Asymptotic Pure Exploration by Solving Games”. In: Advances in Neural Information Processing Systems (NeurIPS) 32, pp. 14492–14501.

Degenne, R., H. Shao, and W. M. Koolen (July 2020). “Structure Adaptive Algorithms for Stochastic Bandits”. In Proceedings of the 37th International Conference on Machine Learning (ICML).

Even-Dar, E., S. Mannor, and Y. Mansour (2002). “PAC Bounds for Multi-armed Bandit and Markov Decision Processes”. In: Computational Learning Theory, 15th Annual Conference on Computational Learning Theory, COLT 2002, Sydney, Australia, July 8 10, 2002, Proceedings. Vol. 2375. Lecture Notes in Computer Science, pp. 255–270.

Garivier, A. and E. Kaufmann (2016). “Optimal Best arm Identification with Fixed Confidence”. In: Proceedings of the 29th Conference On Learning Theory (COLT).

Katariya, S., B. Kveton, C. Szepesv´ari, C. Vernade, and Z. Wen (2017). “Stochastic Rank-1 Bandits”. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA. Vol. 54. Proceedings of Machine Learning Research, pp. 392–401.

Kaufmann, E. and W. M. Koolen (Nov. 2021). “Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals”. In: Journal of Machine Learning Research 22.246, pp. 1–44.

Kaufmann, E., W. M. Koolen, and A. Garivier (Dec. 2018). “Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling”. In: Advances in Neural Information Processing Systems (NeurIPS) 31, pp. 6333–6343.

Russac, Y., C. Katsimerou, D. Bohle, O. Capp´e, A. Garivier, and W. M. Koolen (Dec. 2021). “A/B/n Testing with Control in the Presence of Subpopulations”. In: Advances in Neural Information Processing Systems (NeurIPS) 34.