MSc Project: $\epsilon$-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 $\mu_k$ 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 $\epsilon$ — by asking for identification of any $\epsilon$-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 $\epsilon$-optimal mixture policy. That is, we are happy with any probability distribution $p$ such that

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

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 $\epsilon$-best mixture identification strictly lower than that of $\epsilon$-best arm identification? If so, what is the sample complexity of $\epsilon$-best mixture identification? And how can one design efficient algorithms for $\epsilon$-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


