# Part B Project Proposal

Ramsay Davis (Exeter)

## Proposal

Multi-armed bandits are a form of decision making problem in the field of reinforcement learning, where an agent must choose between multiple options, each with an unknown reward distribution, to maximize the total reward over time while balancing exploration and exploitation. [2]. In the case of stochastic linear bandit problems, one popular algorithm used is the Upper Confidence Bound, or UCB algorithm. The effectiveness of this algorithm relies on the creation of tight bounds for confidence sets. Good results already exist for bandits with finitely many choices (known as arms) and subgaussian noise [2]. Here, the confidence set of the penalised least squares estimator of $\theta_*$, denoted $\hat{\theta}_t$, is given by the equation:

$$\mathbb{P}(\|\hat{\theta}_t - \theta_*\|_{V_t(\lambda)} < \sqrt{\lambda}\|\theta_*\|_2 + \sqrt{2\log\frac{1}{\delta} + \log\frac{\det V_t(\lambda)}{\lambda^d}}) \geq 1 - \delta \tag{1}$$

The purpose of this project is to strengthen these concentration bounds in the infinite and Bernoulli case, answering an open problem proposed by Marco Mussi, Simone Drago and Alberto Maria Metelli in a short paper [3]. The aim will to be to explore two different ways of deriving a tighter bound in this situation. Firstly, to extend results from Yasin Abbasi-Yadkori's PhD thesis [1] regarding vector-valued martingale tail inequalities, which hold in the infinite armed case with subgaussian noise, to the Bernoulli setting. And secondly, by applying sequential likelihood ratios, which can easily be applied to the Bernoulli case but extending it to infinity may prove more challenging. I have worked with David Janz in creating this proposal, and he has agreed to supervise this project.

## References

[1] Yasin Abbasi-Yadkori. *Online Learning for Linearly Parametrised Control Problems*. PhD thesis, University of Alberta, 2012.

[2] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[3] Alberto Maria Metelli Marco Mussi, Simone Drago. Open problem: Tight bounds for kernelized multi-armed bandits with bernoulli rewards. paper, 2024.