MA4601/MAT061 Stochastic Search and Optimisation

Assignment 4: Multi-armed Bandits

Due 12:00 mid-day, Thursday 23rd April

The goal of this assignment is to explore the tradeoff between exploration and exploitation in multi-armed bandit heuristics.

You will need to submit two files: a programme file titled `YOUR_NAME_programme.r` (or `.py`, `.jl`, etc.) and a report as a pdf file titled `YOUR_NAME_report.pdf`. Submission by email to `joneso18@cardiff.ac.uk`. The report should be presented as a stand-alone document that can be understood without having to read your code. It should be no more than four pages long.

Consider the following modifications of the ϵ-greedy, UCB1, and Bayesian decision rules.

ϵ-greedy For some ρ, with probability $1 - \rho/t$ choose the bandit with highest $\hat{\theta}_i$, otherwise choose a bandit uniformly at random.

UCB(ρ)

$$i(t) = \arg\max_i \left(\hat{\theta}_i(t-1) + \sqrt{\frac{\rho \log t}{T_i(t-1)}} \right).$$

Bayesian Let $q(\Theta_i(t), \rho)$ be the 100ρ percentage point of $\Theta_i(t)$, then

$$i(t) = \arg\max_i q(\Theta_i(t), \rho).$$

Implement these decision rules and compare their performance using 10 multi-armed bandits with randomly chosen returns.

Use Bayesian Global Optimisation to find the optimal value of ρ in each case. You may use the function `BayesianOptimization` from the R package `rBayesianOptimization`.

Marks will be allocated on the following basis:

50% Code correctness (how well does it work).

25% Quality of analysis (what have we learnt about these decision rules).

25% Clarity of report.