Comparative Analysis of Multi-Armed Bandit Algorithms in Stock Selection

ECE 8803
Vidya Muthukumar

Deepa Manogaran, Supriya Sundar Faculty: Vidya Muthukumar
School of Electrical and Computer Engineering • Georgia Institute of Technology • Atlanta GA, 30332 USA

Abstract

While the multi-armed bandit model enjoys popularity in studying the exploration-exploitation trade off in sequential decision-making, there remains ambiguity in the optimal algorithm with respect to regret, confidence and bounds when applying to a specific problem. The multi-armed bandit problem is defined as the determination of choice of which arm to pull in a 'K' armed slot machine to maximize profit in a sequence of trials. Many different algorithms exist in theory to this end, but no concrete comparative analysis can be comprehended with respect to a specific dataset or application. This project proposes to identify the optimal algorithm for stock market analysis using empirical evaluation of Upper Confidence Bound(UCB) and Thompson Sampling algorithm.

Upper Confidence Bound(UCB) is favored in financial analysis since it restricts the sampling over time to the actions showing the best performance to maximize the total reward. It progresses from significant exploration to significant exploitation ensuring highest mean reward selected in accordance with the confidence measure. Thompson Sampling is of significance as it gradually refines a model of the probability of the reward for each action and actions are chosen by sampling from this distribution. This gives an estimate for the mean reward value of an action with a measure of confidence for that estimate; allowing optimal action identification faster. We tested the proposed algorithms to determine which allows provision of better stable outcome.

Introduction

Any Portfolio Selection Problem (PSP) has the investor determining a way to allocate available, finite wealth among the available choice of assets. The realization of asset allocation methodology builds the portfolio. Multi-armed bandit problem models the exploration/exploitation trade-off inherent in sequential decision making.

The application of Multi Armed Bandit (MAB) algorithm to fiscal data helps mitigate loss and maximize profit. Though many generalizations and different versions of different bandit algorithms have been studied in this regard, the Upper Confidence Bound (UCB) and Thompson Sampling (TS) algorithms enjoy a superior standing owing to its optimism in face of uncertainty.

While Thompson Sampling follows a Bayesian approach and is essentially a randomized probability matching algorithm where prior ought to be determined, the Upper Confidence Bound (UCB), a deterministic algorithm can be utilized in a Bayesian environment with known priors.

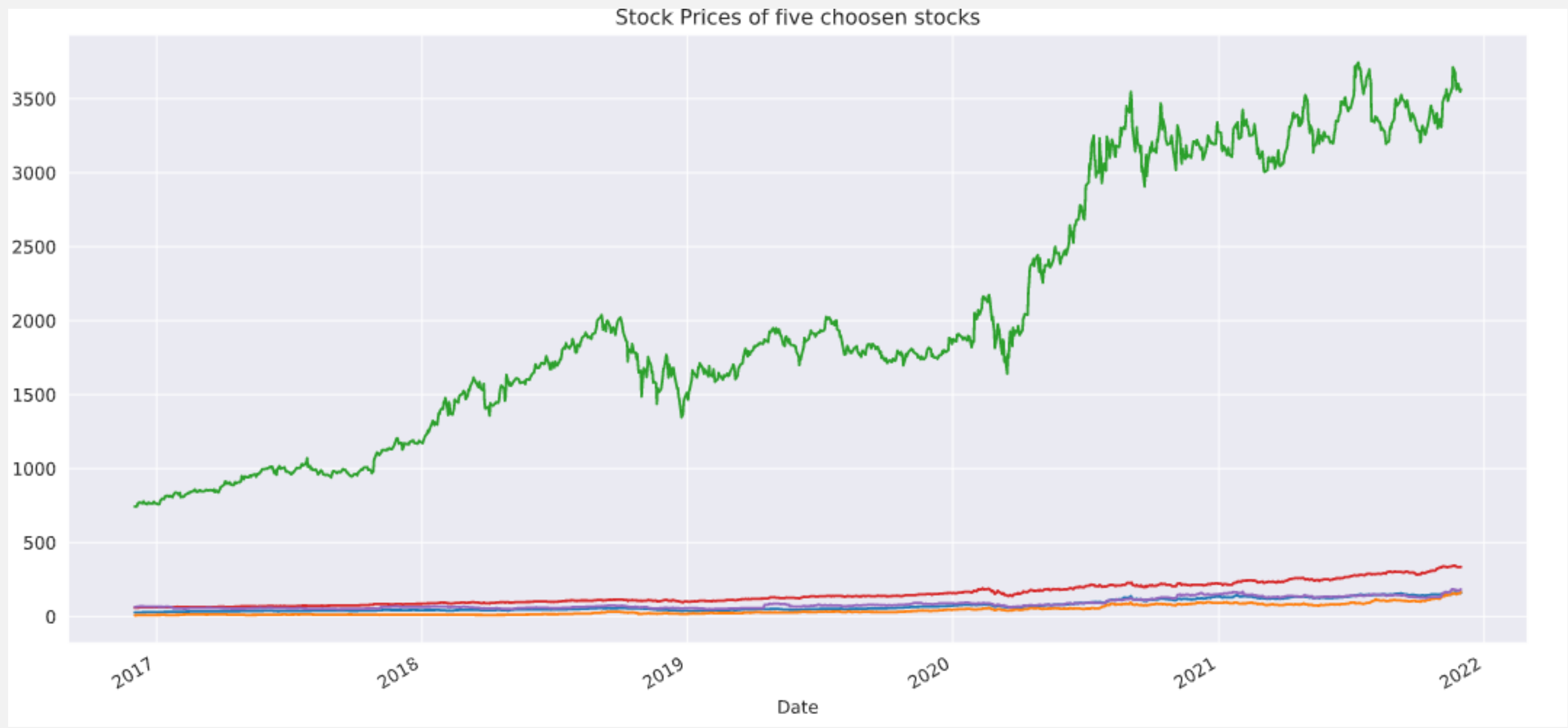
TS may have shorter exploration period than UCB, but both provide similar qualitative results.

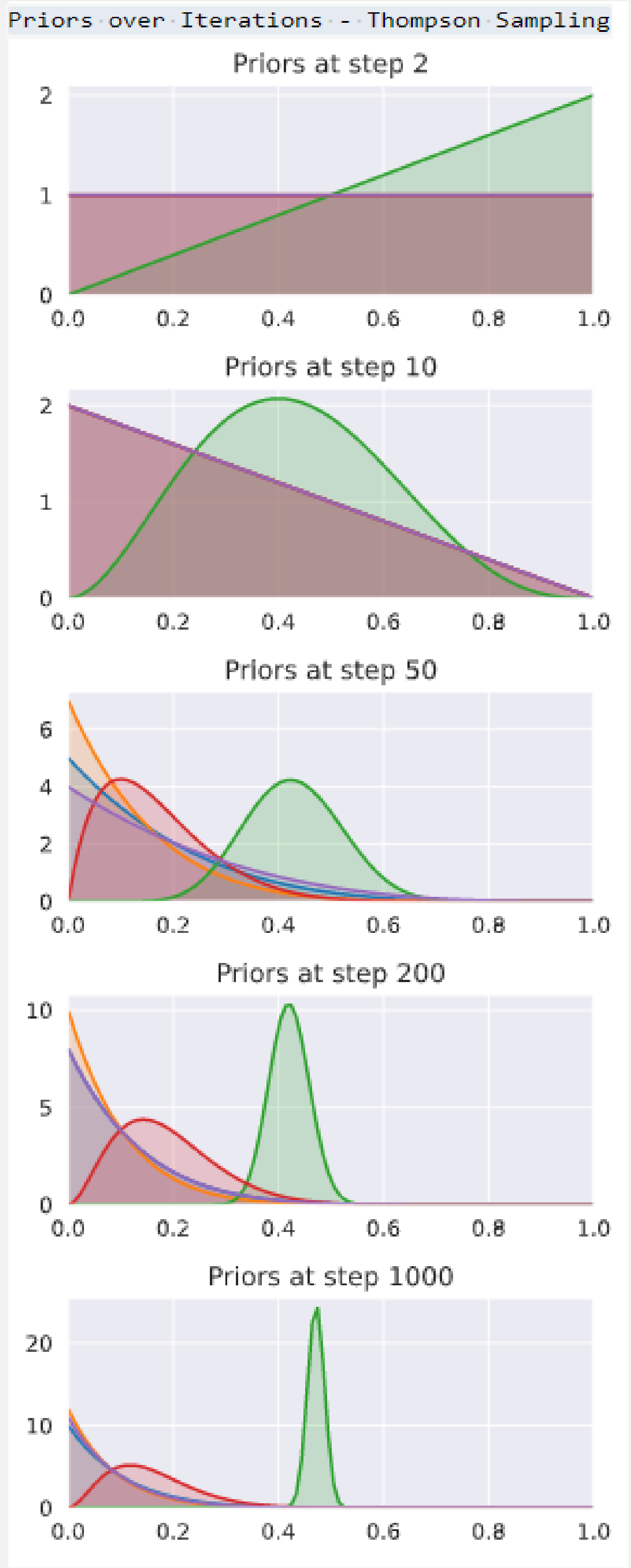
Experiment

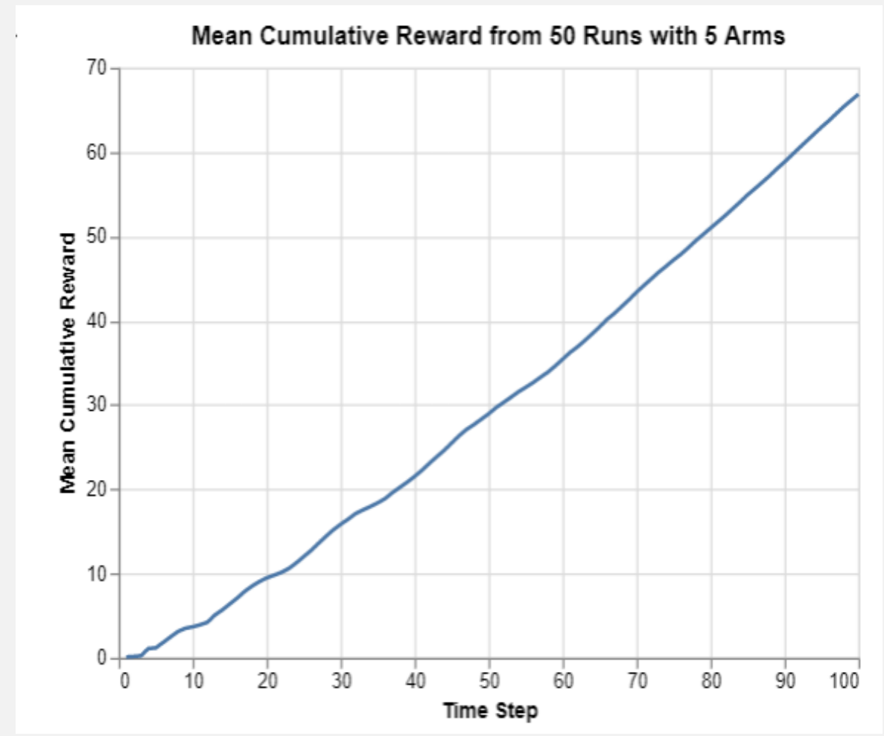
The exploration-exploitation trade off demonstrated by the two algorithms and its effect on the stock data of five companies spread over a period of five years is evaluated. We analyzed the algorithms on a set of generated data with 5 arms and random mean rewards/priors. This validated our implementation of the algorithms as the expected behavior of choosing the arm with higher mean reward or prior probability was observed for both the algorithms. Next, we choose five stocks of Fortune 500 companies – four semiconductor stocks and one software stock. The stocks vary in industry, volatility and trading range. The trend in opening prices for the stocks is plotted here.

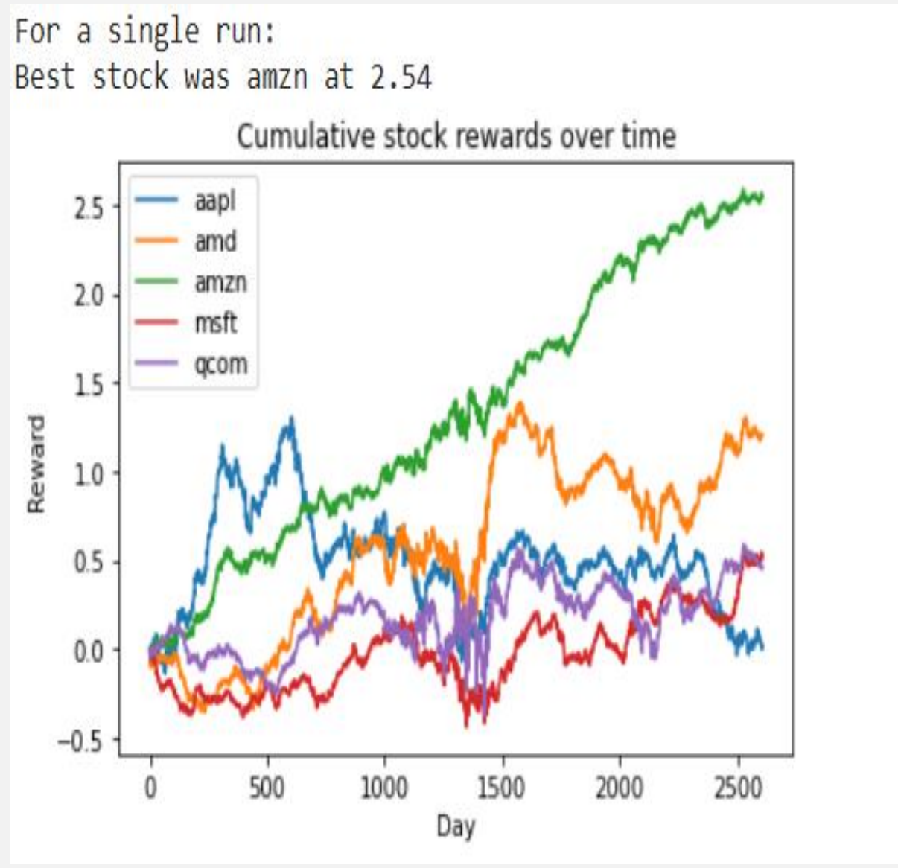
The problem is modelled as choosing a stock that would give at least one dollar return every day for five years. This is a design choice to keep the input parameters consistent across the two algorithms. The algorithms are expected to choose the stock in 'green'. A win is defined as the algorithm pulling an arm with at least one dollar return for a given trial step. Reward is computed as the total wins through the trial.

Results

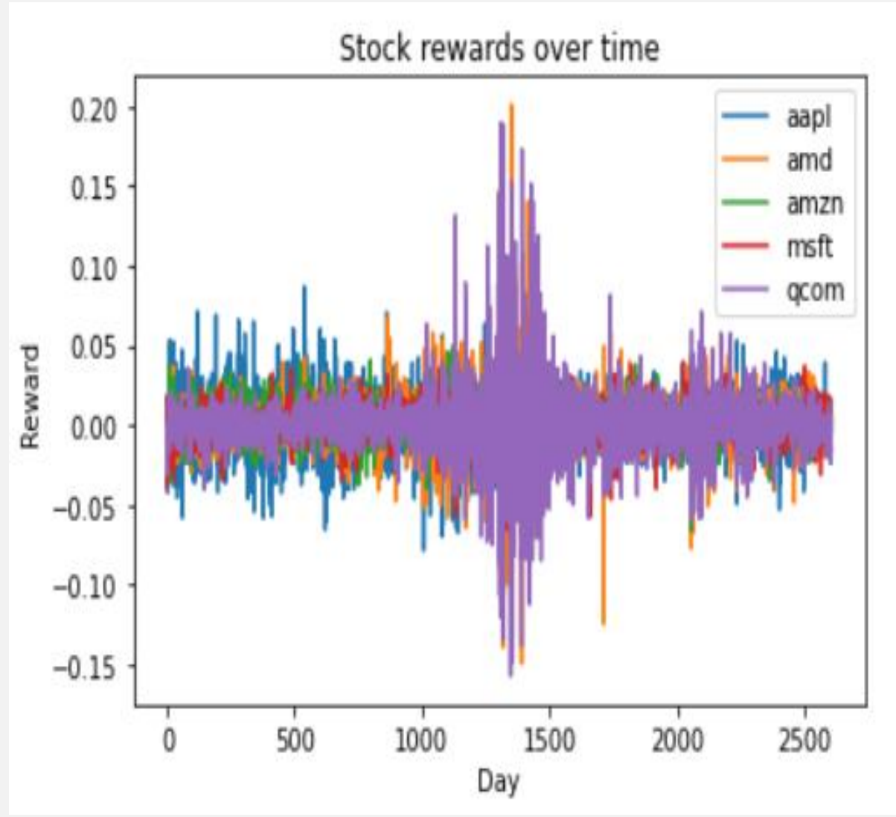




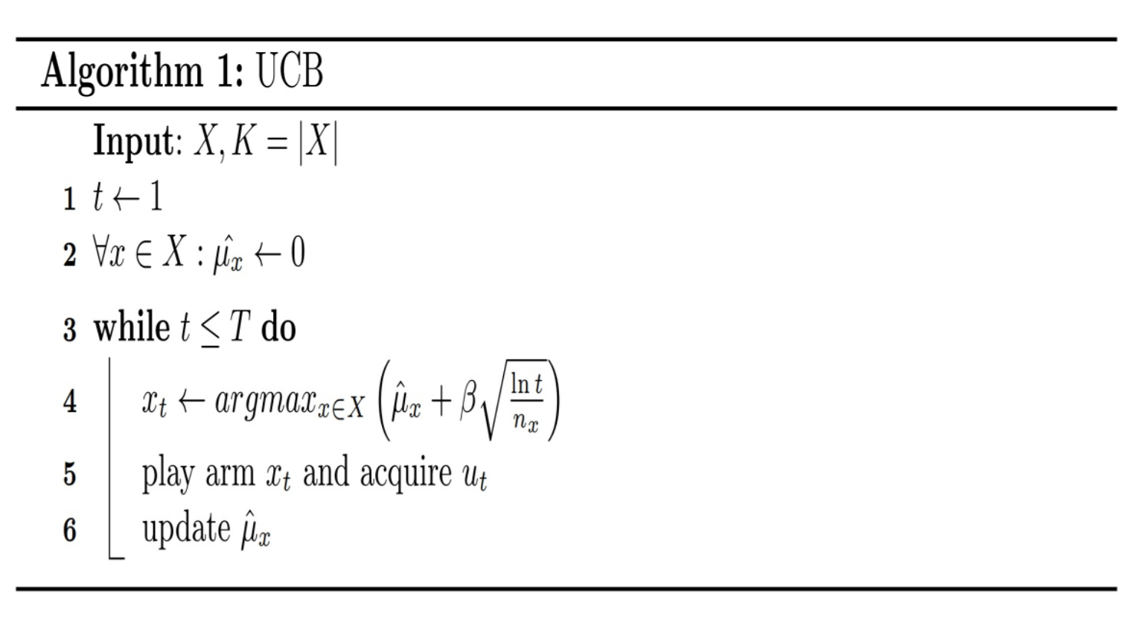




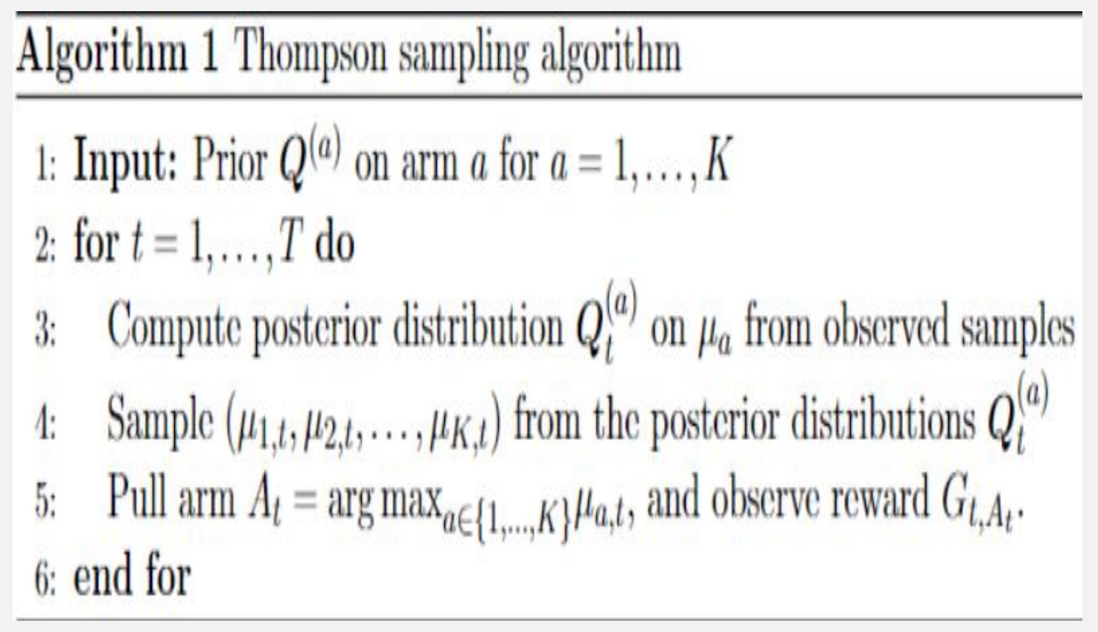
```
For a single run:
Best stock was amzn at 2.54
```



Discussion



```
Algorithm 1: UCB
  Input: $X, K = |X|$
  $t \leftarrow 1$
  $\forall x \in X : \hat{\mu_x} \leftarrow 0$
  while $t \leq T$ do
      $x_t \leftarrow argmax_{x \in X} \left(\hat{\mu}_x + \beta\sqrt{\frac{\ln t}{n_x}}\right)$
      play arm $x_t$ and acquire $u_t$
      update $\hat{\mu}_x$
```



```
Algorithm 1 Thompson sampling algorithm
  Input: Prior $Q^{(a)}$ on arm $a$ for $a = 1, \ldots, K$
  for $t = 1, \ldots, T$ do
      Compute posterior distribution $Q_t^{(a)}$ on $\mu_a$ from observed samples
      Sample $(\mu_{1,t}, \mu_{2,t}, \ldots, \mu_{K,t})$ from the posterior distributions $Q_t^{(a)}$
      Pull arm $A_t = \arg\max_{a \in \{1,\ldots,K\}} \mu_{a,t}$, and observe reward $G_{t,A_t}$.
  end for
```

Conclusions

The opening and closing, low and high value of each individual stock was analyzed over a period of five years. This individual pricing of the stock was first modeled as a curved graph to analyze the growth or fall in pricing and understand the general trend of the data.

Thompson Sampling (TS) was then applied to this and the priors and posteriors were generated. Posteriors were sampled at random and an arm was picked. Based on the choice made repetitively in accordance with the profit or loss possible among the five defined arms, the optimal stock was identified.

Upper Confidence Bound (UCB) when applied to these stocks similarly and the optimal stock was computed by finding the least regret and maximal reward data with pre-determined priors. The average award and the confidence interval of the stocks were determined, from which the upper or maximal bound possible was computed.

The comparison of the two algorithms for the given data showcased upper confidence bound to have performed better and identified Amazon as the company with optimal stock value in light of low regret.

References

1.Agrawal, S., & Goyal, N. 2012. Analysis of Thompson Sampling for the Multi-armed Bandit Problem, 25th Annual Conference on Learning Theory

2.Vermorel, J., & Mohri, M., Multi-Armed Bandit Algorithms and Empirical Evaluation, New York University

3.P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite Time Analysis of the Multiarmed Bandit Problem. Machine Learning, 47(2/3):235–256, 2002

4. S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in multi-armed bandits problems. In Algorithmic Learning Theory, pages 23–37. Springer, 2009.