**(Cover Page Image)**

**DSC5101 Group Assignment 3**

A Study on Historical Surveys Using Multi-Armed Bandit Techniques



**November 16, 2019**

Group Members

Utkarsh Chaturvedi ( )

Dasol Kim ( )

Shengnan Sang ( )

Nikhil Thomas ( )

Table of Contents

[**1** **Executive Summary** 2](#_Toc24827572)

[**2** **Multi-Armed Bandits** 2](#_Toc24827573)

[**3** **Part I: Campaign Background** 2](#_Toc24827574)

[**3** **Part I: Retailer Background** 3](#_Toc24827575)

[**4** **UCB1 Sampling vs. Thompson vs. Original** 3](#_Toc24827576)

[**5** **Summary & Recommendations** 4](#_Toc24827577)

[6 **Appendix** 5](#_Toc24827578)

# **Executive Summary**

This report aims to solve two business problems by implementing, as well as comparing, the Upper Confidence Bound Multi-Armed Bandit (UCB1) and Thompson Sampling deterministic algorithms.

* In the first scenario, the main issue of interest is whether A/B testing and website optimization used during Obama’s fundraising campaign can be further improved with the aforementioned testing methods. Ultimately, results are compared with the actual campaign estimates of an additional $60 million in donations.
* The second case involves a commonplace decision for retail stores: among their newly released product lines, which items should be promoted to each distinct customer segment? Ideally, this targeted marketing will improve website conversion rates significantly. In particular, the customer types are divided into 4 groups based on age: less than 30, between 30 and 45, between 46 and 60, and greater than 60. We found significant differences in the product to be marketed to each age-group with a UCB1 test.

# **Multi-Armed Bandits**

Similar to A/B testing, the Multi-Armed Bandit (“MAB”) testing approach utilizes reinforcement learning to narrow down among a set of competing choices, and determine which variants are the best in terms of realizing the user’s end goal. For instance, given both limited marketing budget and time constraint, the consumer-focused business aims to maximize expected profits through higher conversion rates and improved customer engagement.

The trade-off dilemma between exploration and exploitation is fundamental in the multi-arm bandit algorithm. Specifically, the hypothetical slot machine player performs “exploration” by testing new arms, while “exploitation” occurs when the gambler continues to play the most profitable arm in order to maximize expected return. This can be contrasted with A/B testing method that is wholly focused on “exploration” (in the initial phase of the experiment).

In brief, the epsilon-greedy strategy is commonly used to solve MAB problems. By definition, epsilon is a parameter between 0 and 1 and determined by the user as the total percentage of iterations for exploration. Alternatively, the Thompson Sampling (“TS”) approach commonly uses a beta distribution model to illustrate the expected returns for each arm. During each round, the distribution shape adjusts accordingly and ultimately converges to the actual rate of success, for example, the website click-through and conversion rates. TS critically relies on Bayes’ Theorem as exploitation dominates over time against exploration.

It can be argued that UCB and TS algorithms each have their advantages. To be more specific, UCB gives more importance to exploration when compared to the other methods. This ensures that the model is more robust towards changes in the customer profiles and shows higher sensitivity to time. On the other hand, exploitation is the main object for TS. Hence, as soon as a favourable option is found, then the algorithm continues to exploit in the subsequent trials, which leaves less likelihood for future exploration, unless a new arm with alpha & beta = 1 is introduced, and has a fair chance against the other arms.

# **Part I: Campaign Background**

With the goal of raising Obama’s presential rating and profile, members of the 2008 Obama Campaign performed A/B testing to decide which combination of the website would lead to the highest conversions in terms of click-through rate that eventually translates into donation funds. The test included a total combination of 24 test cases using 4 different buttons and 6 different media backgrounds. It was a randomized control trial conducted on approximately 310,000 unique visitors, with each iteration made visible to around 13,000 people. The A/B testing approach helped identify a certain website styling that had a significantly higher click-through rate than the campaign manager’s/teams favorite option. In fact, the winning variation had a sign-up rate of 11.6% compared to the average of 8.26%, which is an improvement of roughly 40.6%.

This above-mentioned approach that was adopted by the campaign team was highly successful since it led to additional 2.88 million email addresses and $60 million donations. However, there are a number of flaws that should be pointed out:

1. The A/B tests were performed during the start of the campaign. Yet as the presidential race advances, an alternative website setup would have been more favorable. This indicates that concentrated learning should be focused towards the start of the campaign.
2. With A/B testing, all the website versions were tested for 13,000 runs. This is not a very efficient approach since those under-performing websites are still being run. Hence, there is a waste of opportunity by running unsuccessful websites multiple times.

# **Part I: Retailer Background**

The retailer *“ForeverProfit”* is planning to roll out 6 new products with an objective of identifying customer segments and promoting specific products that benefit them. The consumer types have been divided into 4 groups based on age, and each group is assumed to have different probabilities of purchasing the new products. Moreover, the unique number of visitors to the retailer’s website is presumed to be 4 million, which is uniformly distributed among the 4 segments.

# **UCB1 Sampling vs. Thompson vs. Original**

To address the two flaws with A/B testing, alternative techniques including UCB1 and TS are adopted below:

To reiterate, the UCB1 model is built with the aim on exploitation, that is running the successful website. However, there is now a time component that encourages exploration of arms that have not been played for a long period. The TS model uses a beta distribution to decide which arm is optimal to choose. Both models have spread out learning throughout the entire campaign duration, and is able to pick up changes in audience sentiment with regards to their most favorable website design.

*Part I Results Table:*



According to the results from the table above, both UCB1 and TS has their own advantages after running for 100 million iterations. Note that UCB1 gave more importance to exploration when compared to the other two models. This addresses the concern of adjusting robustness in the case of changes in the targeted customers. Whereas for TS, exploitation is the main objective and once a favorable website is identified, then it continues to exploit with less probability of future exploration.

*Part II Results Table:*



The MAB approach is ideal when there are different customer segments, each with distinct affinities towards certain products. This model provides an ability to fine-tune which product should be promoted to each customer group.

To tackle the problem in part II, both UCB1 and TS methods were applied to determine which product has the highest conversion rates for each customer segment, which helps improve market basket analysis and limits the amount of time spent on advertising a product that has low affinity for the consumer.

Additionally, the information gained can be served as an input to numerous price discrimination models after adjusting the prices to ensure not only that the products are favorable to the customer, but also that revenue and profitability is maximized.

# **Summary & Recommendations**

After weighing the pros and cons of UCB1 and TS methods, we decide to implement both models to solve the two business problems. UCB1 is beneficial especially since customers may continuously update their needs, which is satisfied with this algorithm since it allows for exploration with a time aspect. Furthermore, it is advantageous to use TS when new products are released onto the market. In this event, learnings from previous runs are not lost and it is possible to incorporate new products into the model.

# **Appendix**

*Figure 1.1: UCB1 – u values*



*Figure 1.2: UCB1 – n values*



*Figure 2.1: Thompson – a values*



*Figure 2.2: Thompson – (a +b) values*



*Figure 2.3: Thompson – probabilities*

*Figure 4.1 UCB1 – u values*



*Figure 4.2: UCB1– n values*

