Bayesian Multi-Armed Bandits for Online Experimentation

Group 5: Anastasiia Saenko, Dhaval Potdar

April 2024

Abstract

This paper presents a comprehensive exploration of Bayesian Multi-Armed Bandits (MABs) and their application in online experimentation. Through extensive simulations, we showcase the effectiveness of Bayesian MABs in adjusting to dynamic user preferences and optimizing traffic allocation in real-time, outperforming traditional A/B tests. Our report delves into the theoretical underpinnings of Bayesian methods, details the methodology behind MABs, and demonstrates their practical advantages in experiments. The results emphasize the superiority of Bayesian MABs in rapidly adapting to evolving user behaviors, confirming their value for businesses seeking agile and data-driven decision-making frameworks in online settings.

Introduction

Motivation

The rapid evolution of online marketing, conversion optimization, and user experience offers both challenges and opportunities for applied scientists. Businesses increasingly depend on computational methods to navigate decision-making under uncertainty. Critical decisions include choosing effective ad campaigns, prioritizing search results, and identifying optimal marketing investments to enhance user interactions, such as website purchases.

Online experimentation is crucial for optimizing business strategies. Developed frameworks help address various business scenarios and improve decision-making, enabling companies to quantify risks and plan strategically. For example, choosing what to display on recommendation feeds or selecting website designs that optimize user experience and ad revenue is essential. In our digital world, these decisions have significant financial implications, particularly for large companies with extensive user traffic.

Problem statement

Traditionally, A/B testing has been vital for optimizing online strategies, with major companies like Google, Amazon, and Facebook [1] using it to determine which webpage variations maximize user engagement or sales. However, it suffers from significant limitations due to its need for large sample sizes and lengthy durations to achieve statistical significance [2]. This often leads to delayed decision-making and poor utilization of continuous data in rapidly changing environments.

This paper introduces the **Bayesian multi-armed bandit** (MAB) framework as a superior alternative to traditional frequentist A/B testing [3]. Bayesian MABs incorporate prior knowledge and real-time feedback, enabling more adaptive and efficient strategies. Unlike traditional A/B tests, which maintain a fixed traffic distribution among variants, Bayesian MABs dynamically adjust traffic allocation based on real-time data. This flexibility is crucial in fast-paced environments where user preferences and behaviors may quickly change, requiring timely responses to maintain a competitive edge.

Objective

The Bayesian MAB method's strength lies in its probabilistic modeling approach. Bayesian MABs use prior distributions and update these beliefs as new data is observed, thereby optimizing the decision-making process. This approach not only provides a robust framework for dealing with uncertainty but also maximizes the learning from every user interaction.

Our objective with this report is to explore the theoretical foundations and practical implementations of Bayesian MABs in online experimentation. By integrating examples and comparing this approach with traditional A/B testing, we aim to demonstrate the substantial advantages of Bayesian MABs. These include their ability to make faster, data-driven decisions that can adapt to user preferences and behaviors dynamically, ultimately enhancing online marketing strategies.

Background

The Foundations of Online Experimentation

In the digital marketplace, online experimentation has proven critical for businesses seeking to enhance their web presence. Central to this practice is A/B testing, where businesses test different versions of web elements like pages or ads to identify which one yields better outcomes based on defined performance metrics [4][5]. This method typically relies on frequentist statistics, using hypothesis testing and calculating p-values to establish which version is statistically superior [6].

While A/B testing is a valuable tool, its efficacy is hampered by several constraints, particularly noticeable in the volatile landscape of online marketing:

- **Extensive Sample Sizes**: To obtain reliable results, A/B testing demands large sample sizes, which can significantly slow the pace at which insights are generated and applied. This delay often results in missed opportunities to optimize effectively in a timely manner [2].

- **Rigid Experimental Design**: Traditional A/B testing splits traffic evenly between options throughout the testing period. This fixed approach does not capitalize on the insights gained as user data accumulates, thus limiting the potential for mid-experiment adjustments [4].

Rise of Bayesian Experimental Methods

To mitigate these issues, there has been a shift toward Bayesian methods in experimentation. Unlike frequentist statistics that fix probabilities, Bayesian statistics treat probabilities as evolving degrees of belief, which are refined as new data becomes available [7]. This paradigm allows for more dynamic decision-making and traffic allocation based on ongoing analysis and insights.

Bayesian Multi-Armed Bandits and Online Experimentation

Drawing from the classic dilemma in the "multi-armed bandit" problem—a scenario in which a gambler must choose from multiple slot machines, each with unknown payout ratios to maximize returns—the Bayesian MAB approach adapts this challenge for online experimentation [3]. Each variant, be it a website design or ad campaign, acts like a slot

machine. As data about their performance accrues, the Bayesian MAB model updates its assessments and reallocates resources dynamically to better-performing variants. This method not only speeds up the experimentation process but also increases the accuracy of its outcomes by continuously optimizing the allocation of traffic to variants showing promising results [8].

Practical applications of Bayesian MABs

Bayesian MABs are widely used across various industries due to their ability to adapt and optimize in real-time. In e-commerce, they enhance website functionality by testing and optimizing layouts and product recommendations to boost conversions [3]. Marketing teams use these models to dynamically optimize ad campaigns by selecting the best-performing options based on current user engagement [9]. In user interface research, Bayesian MABs help determine the impact of different design elements on user experience by testing variations rigorously [7]. They are also crucial in recommendation systems, personalizing content like articles or videos for individual preferences [10]. Overall, Bayesian MABs provide a dynamic decision-making tool ideal for environments that change quickly and where traditional methods may falter due to their slow adaptation and high sample size requirements.

Methodology

How do Bayesian MABs work

The Multi-Armed Bandit (MAB) problem is a foundational concept within the field of reinforcement learning, which focuses on how agents can learn to make decisions by interacting with an environment. In the MAB framework, each decision (or "arm pull") provides new information that can be used to update the agent's understanding and strategy. The challenge lies in the dual need to explore (to test all arms to find the best one) and exploit (to use the best-known arm to maximize rewards). The MAB problem simplifies the broader reinforcement learning challenge by focusing on learning the optimal action selection without having to account for evolving environment states.

0.1 Methodology

0.1.1 Stochastic and Multi-Armed Bandit Problems

In the realm of online experimentation, the decision-making process is addressed through the Multi-Armed Bandit (MAB) problem, a subset of stochastic bandits with distinct assumptions and notations. We will use the outline by A Slivkins [11].

Assumptions:

1. The rewards for each action are independent and identically distributed (IID), with distribution D_a for each action a.

2. Rewards per round are bounded within the interval [0, 1].

3. Focus is on the mean reward vector $\mu \in [0,1]^K$, where $\mu(a) = \mathbb{E}[D_a]$ is the mean reward for action a.

Notations:

- K: Number of actions or 'arms'.
- T: Number of rounds.
- R_t: Reward distribution for each arm a at time t.
- a_t: Action selected at time t.
- r_t: Observed reward at time t, sampled from R_{a_t}.

The algorithm seeks to maximize cumulative rewards over time by observing data, updating information, and making decisions that predict future performance.

0.1.2 Bayesian MAB Framework

The Bayesian MAB approach is characterized by its use of prior and posterior distributions, with Thompson Sampling as a strategy for balancing exploration and exploitation.

1. **Initial Prior:** Start with a Beta distribution with parameters α and β, representing prior beliefs.
2. **Sampling:** At each round t, sample from the posterior distribution $\theta_k \sim \text{Beta}(\alpha_k, \beta_k)$ for each arm.
3. **Action Selection:** Choose the arm with the highest sampled value, $a_t = \arg\max_k \hat{\theta}_k$.
4. **Posterior Update:** Update the Beta distribution parameters α_{a_t} and β_{a_t} with the observed reward r_t.
5. **Thompson Sampling:** Implement Thompson Sampling for action choice, which samples from the posterior to inform decision-making.

This methodology adapts dynamically to changes in user preferences, especially in rapidly evolving environments like online advertising.

Experimental setup

In our experimental simulation we are going to evaluate two ways to run online experiments: A/B tests (spit tests/frequentist tests) and Bayesian MAB experiments with dynamic allocation of traffic. We start with comparing how two differ in their experiments setup.

The traditional **A/B tests** generally follow a workflow similar to more general hypothesis testing. We illustrate that this process is very linear and not sequential:

In contrast to A/B testing, MAB workflow involves a dynamic assignment of users to variants. In traditional A/B testing, the experiment progresses linearly from hypothesis for-

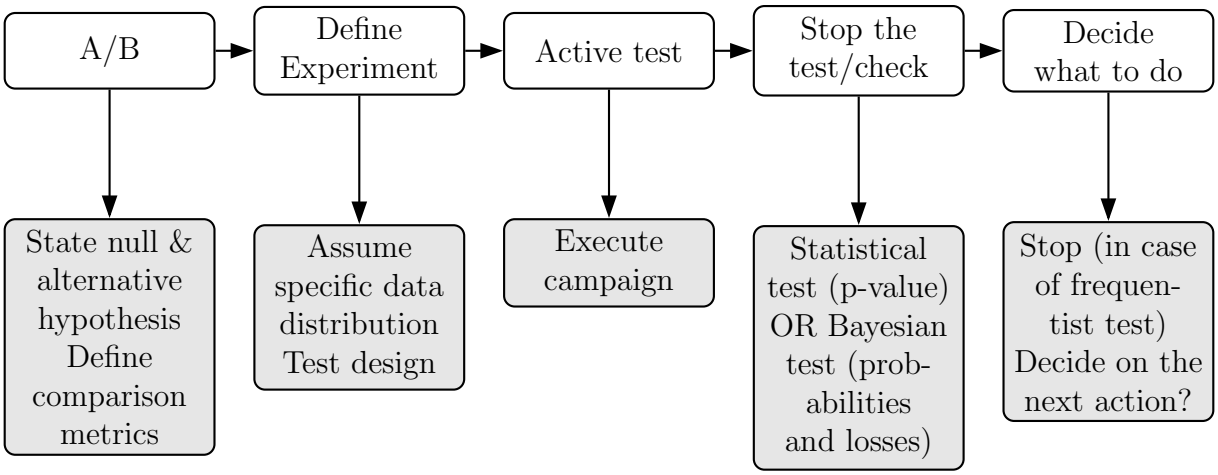


Figure 1: The general workflow of A/B testing methodology.

mulation to decision-making after statistical analysis. Traffic allocation remains unchanged until the test concludes, potentially limiting responsiveness in dynamic online environments.

In contrast, the Bayesian MAB workflow is inherently iterative. Following setup and experiment definition, it employs continuous adaptation during the active testing phase. Thompson Sampling dynamically adjusts traffic allocation in response to incoming data, establishing a cyclic and adaptive optimization cycle.

Bayesian allocation leverages Thompson Sampling as given by:

$$P(\theta_A > \theta_B|D) = \int_0^1 \int_0^{\theta_A} \pi(\theta_A|D)\pi(\theta_B|D)d\theta_B d\theta_A,$$

where θ_A and θ_B denote the conversion rates of variants A and B, respectively, and D signifies the observed data. This formula supports a real-time, data-responsive optimization strategy, boosting the efficiency of online experiments.

Simulation study

For our study, we simulate an A/B/C...n campaign where we not only evaluate multiple variants but also explore the performance of different bandit strategies.

Experiment objective: this experiment evaluates the effectiveness of Thompson Sampling within the Bayesian MAB framework for online decision-making. Simulations mimic real-world scenarios like online advertising, testing the method's ability to learn and adapt in order to maximize rewards. The goal is to demonstrate the advantages of Bayesian MABs over traditional A/B testing, specifically their accelerated decision-making guided by real-time data.

Hypothesis:

- H_0: All algorithms perform equally in terms of conversion rate optimization.
- H_1: There is a significant difference in the performance of at least one of the algorithms in optimizing conversion rates.

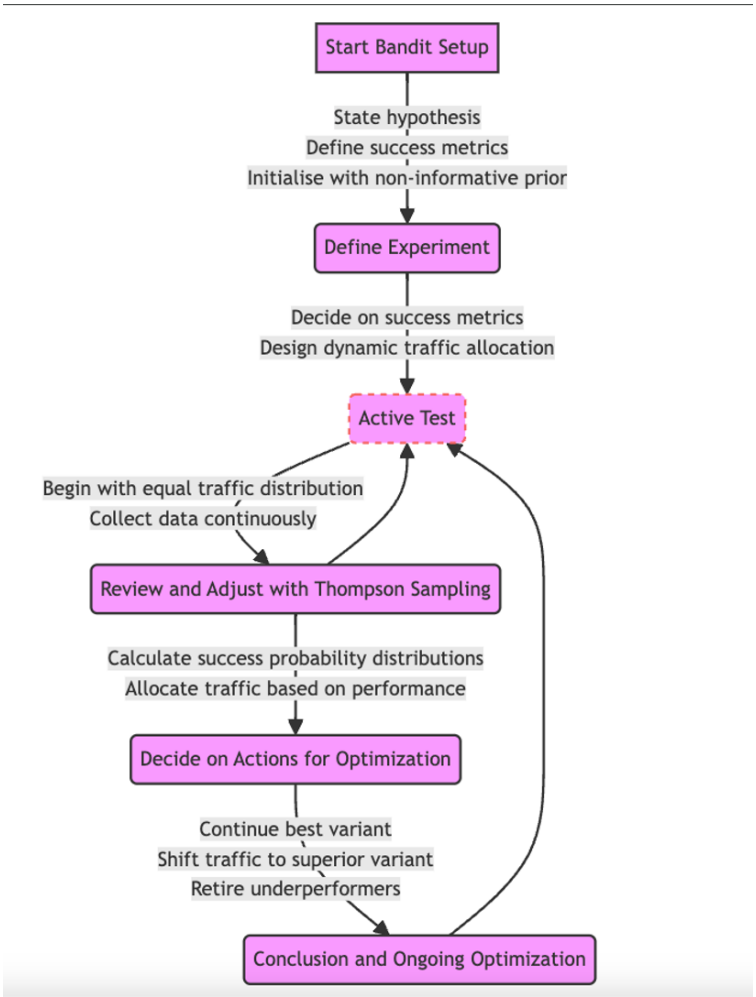


Figure 2: Bayesian MAB workflow for the experiment.

Result Analysis: The performance of the allocation algorithms is assessed on the following criteria:

- **Distribution Allocation:** We examine the evolution of traffic allocation across variants and its congruence with the true conversion rates. The traffic allocation $A_i(t)$ to each variant i at time t is compared against the actual conversion rate θ_i.

- **Convergence Metric:** The algorithms' convergence speed towards the real conversion probabilities is quantified using regret, defined as:

$$\text{Regret}(T) = \sum_{t=1}^{T} (\theta^* - \theta_{i_t}) \tag{1}$$

  where θ^* represents the optimal conversion rate among all variants, θ_{i_t} denotes the conversion rate of the chosen variant at time t, and T is the total number of trials.

- **Efficiency Metric:** The total reward (R) earned by each algorithm is compared to the

theoretical maximum reward ($R_{\max}$) achievable by always selecting the best variant:

$$\text{Efficiency} = \frac{R}{R_{\max}} \tag{2}$$

This metric measures the percentage of the maximum possible conversions captured by the algorithm.

Decisions are then made to either continue exploiting the current allocation strategy or to adjust it by shifting traffic from underperforming to better-performing variants, thereby optimizing the overall conversion rate.

Execution: For the simulation study, we re-constructed some commonly used Python classes[12] to simulate an A/B/C...n campaign environment where multiple advertising variants are tested simultaneously. This class, named simulation, encapsulates all the methods required to simulate, run, and plot the results of campaigns using different bandit strategies, including random allocation, ε-greedy, , and Thompson Sampling.

Results

Convergence to probabilities/conversions

The simulation study utilized the Thompson Sampling algorithm to elucidate the Bayesian Bandits approach within a Bernoulli framework. Actions within this setup correspond to different bandit arms, each associated with a binary reward output determined by a success probability θ_k. The true reward probabilities $\theta = (\theta_1, \ldots, \theta_K)$ remain fixed yet unknown to the algorithm.

Initially, the belief about each action's success probability was modeled with a Beta distribution, characterized by parameters α_k and β_k. The probability density function for action k is given by:

$$p(\theta_k) = \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \theta_k^{\alpha_k - 1} (1 - \theta_k)^{\beta_k - 1} \tag{3}$$

After each action and observed reward, parameters are updated via the Bayesian update rule, reflecting the Beta-Bernoulli conjugacy. This process iteratively refines the algorithm's belief model, optimizing the action selection strategy.

The Python implementation we used can be found at Appendix .1.

For the Python implementation, we assumed 3 bandits and did K=1000 iterations starting with a Beta(1,1) prior for each. For the experiment, we initialized the probabilities of 0.4,0.5,0.6 for the bandits. Note that in practice, we would not know these in advance. After 1000 iterations, we observed the following distributions.

The result of this simulation is shown in Figure 4.

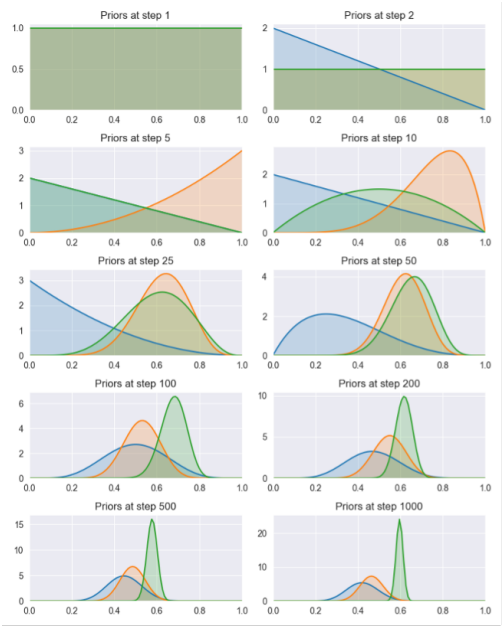


Figure 3: The observed distributions at periods 1, 2, 5, 10, 25, 50, 100, 200, 500 and 1000

The simulation iterates through the following steps:

1. Sample a probability model for each action.

2. Select the action with the highest probability sample.

3. Apply the action, observe the result, and update the model parameters.

Three bandits were assumed, each initialized with success probabilities of 0.4, 0.5, and 0.6, and iterated over 1000 trials. The algorithm demonstrated an adaptive preference, increasingly favoring the action with the highest probability, thus validating the effectiveness of the Bayesian method.

Experiments.

Traditional A/B testing's extended duration and requirement for live, continuously interacting variants can hinder the practical study of Bayesian Bandits. To address this, this report explores the use of Bayesian Multi-Armed Bandits through a simulated experiment[13], mitigating the need for a complex live testing environment.

We launched a 30-day simulated A/B test on two website variants with true click-through rates (CTR) of 10% and 30%. The simulation, which concluded in minutes, started with non-informative $\beta(1, 1)$ priors for both CTRs. Below are the results from the first 5 days.

Table 1: First 5 days for the A/B test for Variant A

Day	Impressions	Impressions w/ Decay	Clicks	Clicks w/ decay	CTR	α	β
0	0	0.0	0	0.0	0.0%	1	1
1	1934	1.0	644	1.0	100.0%	2	1
2	491	1161.4	165	387.4	33.4%	388	775
3	2960	991.8	901	331.8	33.5%	333	661
4	1206	2371.5	359	740	31.2%	741	1632

Table 2: First 5 days for the A/B test for Variant B

Day	Impressions	Impressions w/ Decay	Clicks	Clicks w/ decay	CTR	α	β
0	0	0.0	0	0.0	0.0%	1	1
1	1897	1.0	181	1.0	100.0%	2	1
2	521	1139.2	47	109.6	9.6%	111	1031
3	0	996.5	0	94.4	9.5%	95	903
4	0	598.3	0	57.0	9.5%	58	542

We use a daily recency decay factor of 0.95 to lessen the impact of past impressions and clicks. α and β are the priors for the next day's Bayesian update, which incorporates the Clicks/w decay. Here are the results after 300 days.

Table 3: Final day's results for Variant A

Day	Impressions	Impressions /w Decay	Clicks	Clicks /w decay	CTR	α	β
300	0	50.95	0	5.39	10.57%	6	47

Table 4: Final day's results for Variant B

Day	Impressions	Impressions /w Decay	Clicks	Clicks /w decay	CTR	α	β
300	3450	3225.32	1098	963.14	29.86%	964	2263

And below is the realized distribution for both variants in the end. As expected, the variance of variant B is lesser than that of variant A, and this tells us that the sampling algorithm directed most of the traffic to variant B, as it ascertained that the true probability is higher for variant B.

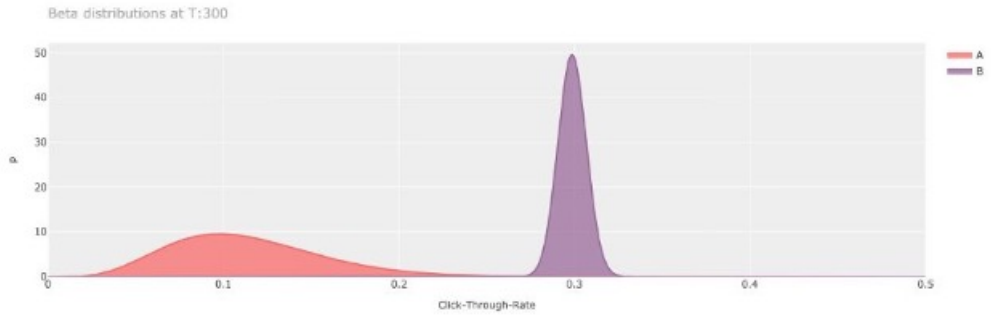


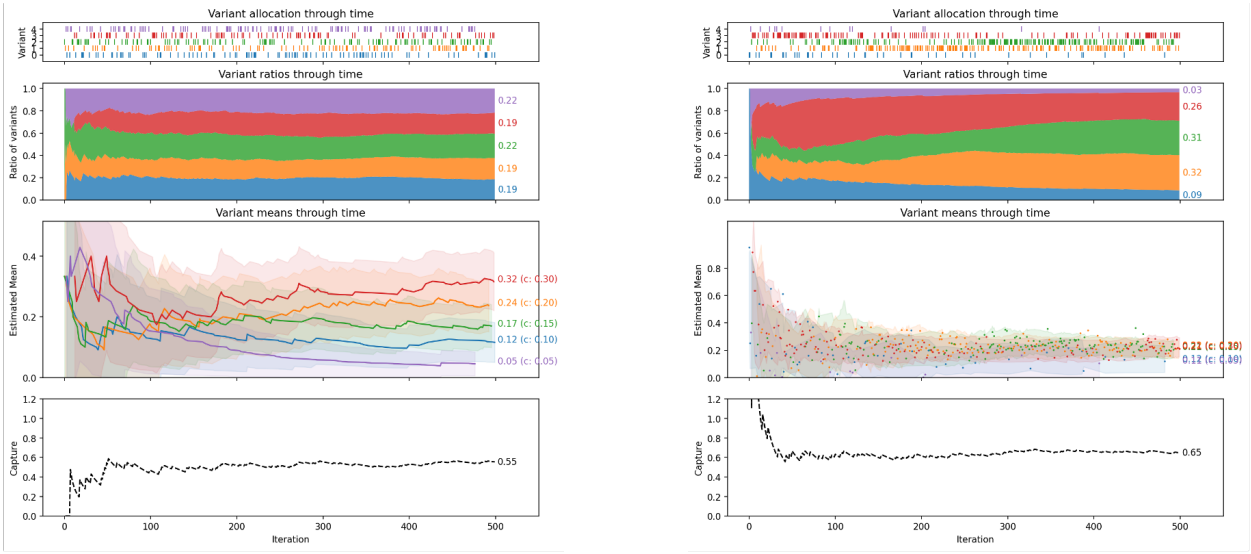
Figure 4: Realized Distributions for the A/B test

This iterative updating and the algorithm's adaptability to favor the most rewarding action exemplify the Bayesian method's efficacy. Dynamic allocation strategies, such as those utilized in this study, are crucial for real-time optimization in online environments subject to rapidly changing user behaviors.

These observations underscore the capability of Thompson Sampling not just to identify the most profitable variant but also to minimize regret by dynamically focusing efforts on the highest-return option.

Traffic Allocation comparison

Here we simulated two allocation strategies for our 5 ads: random selection (an even split) and Bayesian MAB.



(a) Random Selection

(b) Bayesian MAB

Figure 5: Traffic Allocation under different strategies.

Random Selection results indicate a fairly even distribution of traffic across variants, as expected from a non-adaptive A/B/C test strategy. The results show consistent allocation regardless of variant performance, thus lacking in strategic optimization.

Bandit Algorithm Testing demonstrates a more intelligent allocation, with the system learning over time to direct traffic towards the better-performing variants. Thompson Sampling, in particular, exhibited a more pronounced ability to adapt allocation in favor of the variant with the highest conversion rates, as evidenced by the convergence of traffic towards a specific variant over time.

Additionally, we simulated a 6-week trial of this experiment which illustrates the same result. See Appendix.2 for more details.

Effective Conversion Rate

Effective Conversion Rate was used as a key performance metric. It represents how close the algorithm gets to the best possible conversion rate at each point in time. Thompson Sampling consistently achieved a conversion rate closer to the optimal, with less variability between simulations compared to ε-greedy and random selection. The ε-greedy algorithm showed learning and improvement over time but with more variability. Random selection remained consistently far from the optimal, highlighting the importance of a sophisticated approach in bandit algorithms for conversion rate optimization. Illustrated in the Figure 6.

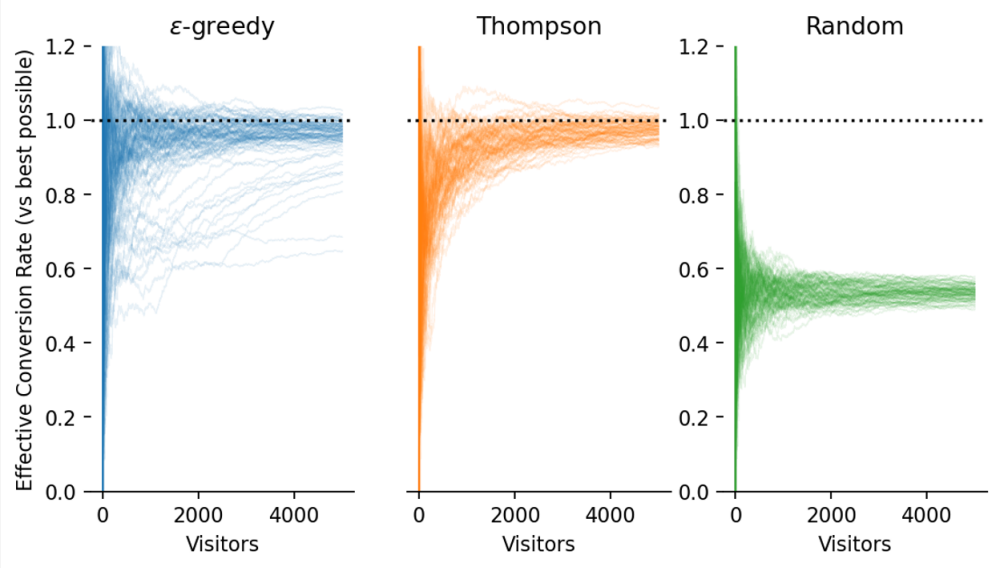


Figure 6: Repeated simulations vs max possible rewards (CR).

From these findings, Thompson Sampling is validated as a good alternative method for dynamic allocation in online experiments when the goal is to maximize conversion rates efficiently. Its capacity to balance exploration and exploitation demonstrates the practical utility of Bayesian MABs in real-world scenarios where rapid adaptation to user preferences is crucial.

Additional experiments

There are many situations where Bayesian MABs are better than traditional A/B tests because of the additional functionality the method provides. For example, Bayesian MABs

are robust to adding more new variations into the experiment with time which is impossible for traditional frequentist framework.

We concentrated on an additional use-case to illustrate this principle: on the case when user preferences in an experiment are not fixed over time. In this case, traditional A/B tests lack mainly because of two reasons:

- Bayesian MABs can adapt to changes in user preferences more effectively because they continuously update the estimated success probabilities of each variant in real time. This is in contrast to traditional A/B tests, which generally operate under the assumption that user preferences remain stable over the test period.

- Static Sampling: Traditional A/B tests "sample" user preferences at fixed intervals, essentially providing a snapshot of behavior during the test period. If preferences change after these intervals, the A/B test won't detect it until the next test iteration, if at all.

We ran a simulation where we iterated over user preferences (changed the favorite ad for n+1 user over the period).

Our findings show that Bayesian Multi-Armed Bandits are more effective than traditional A/B tests in environments with changing user preferences. They adapt in real-time, constantly updating to optimize for current user behavior, while A/B tests can miss shifts in preferences, leading to outdated conclusions. Bayesian MABs are thus better for responsive and agile decision-making in dynamic settings.

The results of this simulation can be found at Appendix .3.

Discussion and Conclusion

Discussion

Bayesian Multi-Armed Bandits (MABs) have proven to be a dynamic and adaptive approach to online experimentation, offering an alternative to traditional A/B tests. With their real-time adaptability, they allow for ongoing learning and strategy adjustment that matches the fluid nature of user interactions online. Our research highlights the efficient learning mechanism of Bayesian MABs, enabling swift and effective allocation of traffic to respond to the ever-changing preferences of users—a capability that traditional A/B tests, with their static frameworks, typically lack.

However, the application of Bayesian MABs is not without challenges. These include interpretative complexities for practitioners unfamiliar with Bayesian principles and the potential for premature data analysis leading to less than optimal decision-making [14]. Ensuring the robustness of Bayesian MABs also involves addressing the fairness of methodological comparisons. Despite these limitations, the strengths of Bayesian MABs, as demonstrated in our simulations, mark them as a valuable method for online decision-making and user experience optimization.

Conclusion

Our simulations reveal that Bayesian MABs stand out for their adaptive prowess in online settings, tailoring experiences by continuously updating with live user feedback—a stark contrast to the static nature of traditional A/B tests. This adaptability proves crucial when user preferences evolve, allowing for a more nuanced approach to traffic allocation and engagement.

The insights gleaned from our experiments highlight the promise of Bayesian MABs, though they come with a learning curve and certain analytical complexities. As we consider future work, refining our understanding of these systems will bolster their practical application, ensuring that businesses can leverage the full spectrum of benefits offered by Bayesian methods in optimizing user interactions online.

Bibliography

[1] VWO, "7 a/b testing examples [updated 2024]," 2024. Blog post.

[2] H. H. Olsson, J. Bosch, and A. Fabijan, "Experimentation that matters: a multi-case study on the challenges with a/b testing," in *Software Business: 8th International Conference, ICSOB 2017, Essen, Germany, June 12-13, 2017, Proceedings 8*, pp. 179–185, Springer, 2017.

[3] S. L. Scott, "A modern bayesian look at the multi-armed bandit," *Applied Stochastic Models in Business and Industry*, vol. 26, no. 6, pp. 639–658, 2010.

[4] R. Kohavi, R. Longbotham, D. Sommerfield, and R. M. Henne, "Controlled experiments on the web: survey and practical guide," *Data mining and knowledge discovery*, vol. 18, pp. 140–181, 2009.

[5] R. Kohavi and R. Longbotham, "Online controlled experiments and a/b tests," *Encyclopedia of machine learning and data mining*, pp. 1–11, 2015.

[6] A. Deng, Y. Xu, R. Kohavi, and T. Walker, "Improving the sensitivity of online controlled experiments by utilizing pre-experiment data," in *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 123–132, 2013.

[7] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," *Advances in neural information processing systems*, vol. 24, 2011.

[8] E. Kaufmann, N. Korda, and R. Munos, "Thompson sampling: An asymptotically optimal finite-time analysis," in *International conference on algorithmic learning theory*, pp. 199–213, Springer, 2012.

[9] A. Gopalan, S. Mannor, and Y. Mansour, "Thompson sampling for complex online problems," in *International conference on machine learning*, pp. 100–108, PMLR, 2014.

[10] J. Kawale, H. H. Bui, B. Kveton, L. Tran-Thanh, and S. Chawla, "Efficient thompson sampling for online matrix-factorization recommendation," *Advances in neural information processing systems*, vol. 28, 2015.

[11] A. Slivkins *et al.*, "Introduction to multi-armed bandits," *Foundations and Trends® in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[12] P. Stubley, "A visual exploration of multi-armed bandit experiments," 2020.

[13] Arngren, "Arngren/bayesian-ab-test." `https://github.com/Arngren/bayesian-ab-test`, 2024.

[14] M. Loecher, "Are multi-armed bandits susceptible to peeking?," *Zagreb International Review of Economics & Business*, vol. 21, no. 1, pp. 95–104, 2018.

Appendices

.1 Python implementation for probabilities

the Beta-Bernoulli bandit proceeds as per follows:

Algorithm 1 Thompson Sampling for Bernoulli Bandits (AlgorithmBernTS)

Input: Number of bandits K, prior parameters α, β
for $t = 1, 2, \ldots$ **do**
    # Sample model:
    for $k = 1, \ldots, K$ **do**
        Sample $\hat{\theta}_k \sim \text{Beta}(\alpha_k, \beta_k)$
    end for
    # Select and apply action:
    $x_t \leftarrow \arg\max_k \hat{\theta}_k$
    Apply action x_t and observe reward r_t
    # Update the distribution:
    $(\alpha_{x_t}, \beta_{x_t}) \leftarrow (\alpha_{x_t} + r_t, \beta_{x_t} + (1 - r_t))$
end for

.2 Simulated Visitor Allocation

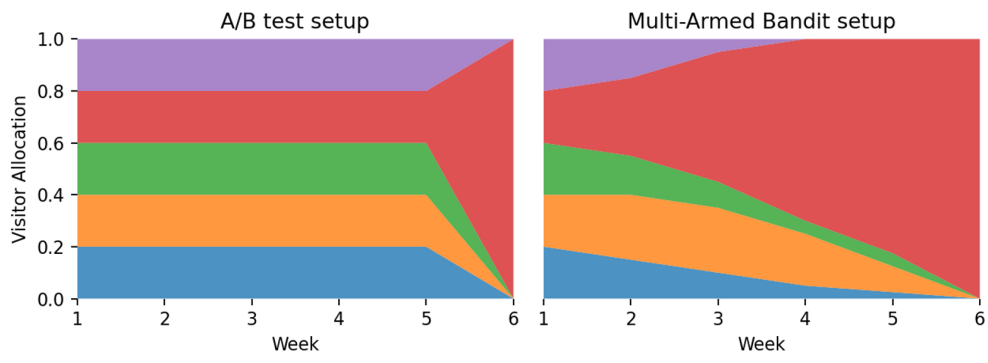


Figure 7: Six weeks experiment simulation

Simulated Visitor Allocation over a 6-week period showcased the stark contrast between the A/B test setup and the Multi-Armed Bandit setup. The latter was more dynamic,

with significant shifts in traffic allocation observable as the system learned which variants performed best.

.3 Simulation: different user preferences

Experiment set-up:

Objective: To assess the efficacy of Bayesian MABs in dynamically adjusting to changes in user ad preferences and to compare their performance to that of traditional A/B testing.

Hypothesis: Bayesian MABs will more accurately reflect evolving user preferences in ad variant selection, leading to optimized conversion rates over time, unlike traditional A/B testing which may lag in adaptation due to its static nature.

Methodology

Initialization

- Ad Variants Setup: Develop several ad variants, each with an associated conversion probability.
- User Preference Model: Design a model to simulate evolving user preferences impacting ad variant conversion probabilities.

Simulation Phases

1. Conduct a baseline A/B test to measure initial conversion rates without user preference adaptation.
2. Implement Bayesian MAB to allocate traffic dynamically based on real-time data.
3. Continue the simulation over multiple periods to observe Bayesian MAB adaptability.

Traffic Allocation

- A/B Test Traffic: Evenly distribute traffic across all variants for a fixed period.
- Bayesian MAB Traffic: Adjust traffic allocation based on updated success probabilities using Bayesian updating.

Data Collection

- Track conversions for each ad variant.
- Record the Bayesian MAB's traffic adjustments in response to preference changes.

Analysis and Evaluation

- Compare cumulative conversion rates and analyze the Bayesian MAB's rate of adjustment.
- Use heatmaps and graphs to display the distribution of variant preferences and traffic allocation over time.

Results Interpretation

- Discuss Bayesian MAB advantages in adapting to dynamic user preferences.
- Summarize potential limitations of traditional A/B testing in dynamic environments.

The experiment result:

Bayesian MABs demonstrate a clear advantage in adapting to and capitalizing on fluctuating user preferences. Unlike traditional A/B testing, which assumes a degree of preference stability, Bayesian MABs continuously learn and adjust the traffic distribution based on real-time data. This feature allows them to respond promptly to changes, optimizing for the most preferred user variations as they evolve.

In scenarios where user preferences are dynamic, traditional A/B testing may fail to capture the nuanced shifts that occur during the testing period. This limitation can result in suboptimal decision-making, as the test outcomes might not reflect the current state of user preferences. On the other hand, Bayesian MABs inherently account for variability and uncertainty, making them more robust and reliable for real-time optimization.

The simulation results support the conclusion that in real-world applications, where user behavior and preferences are subject to change due to trends, seasonality, or other external factors, Bayesian MABs can provide a strategic advantage.

The simulated different preferences model:

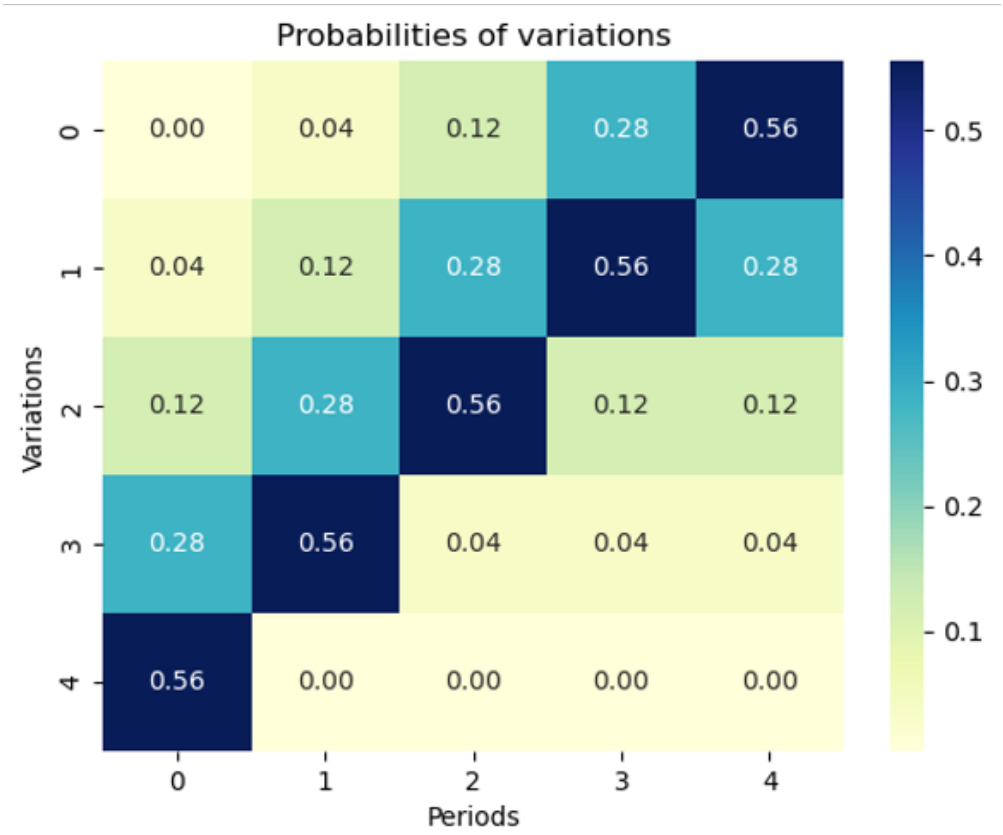


Figure 8: Changing preferences over time

Using the heatmap provided on the graph, we visualize the probability distributions of user preferences across different periods. The color intensity on the heatmap represents the magnitude of preference probability for each variant, with darker shades indicating higher probabilities.

Simulation phases:

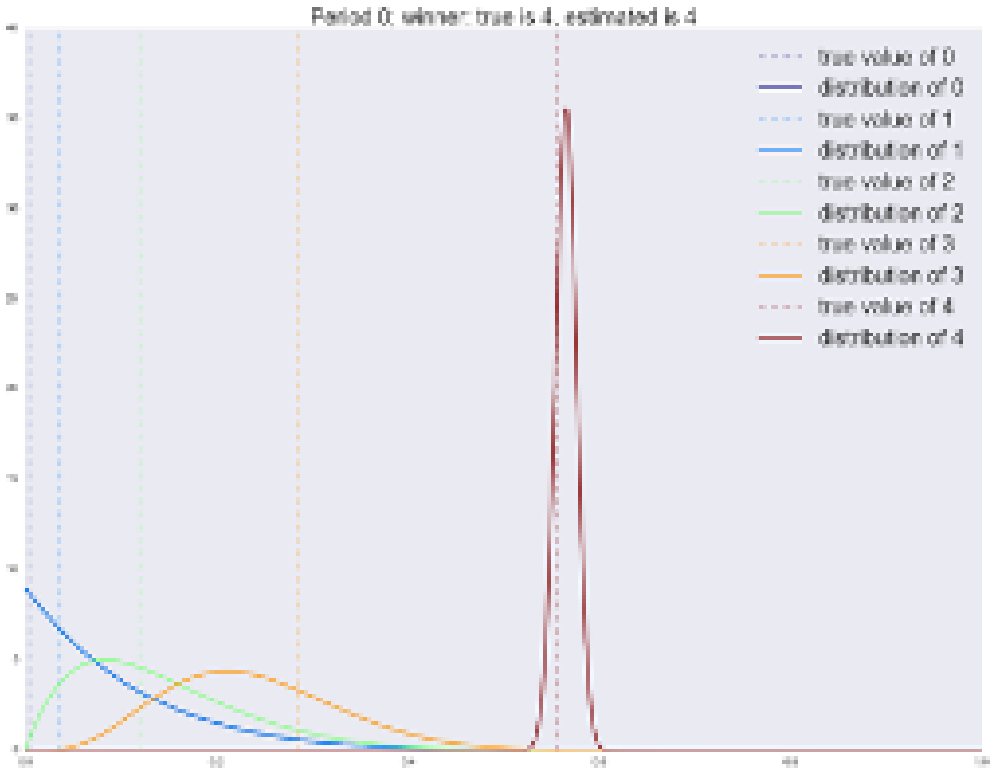


Figure 9: Bayesian MAB adjustment: period 1

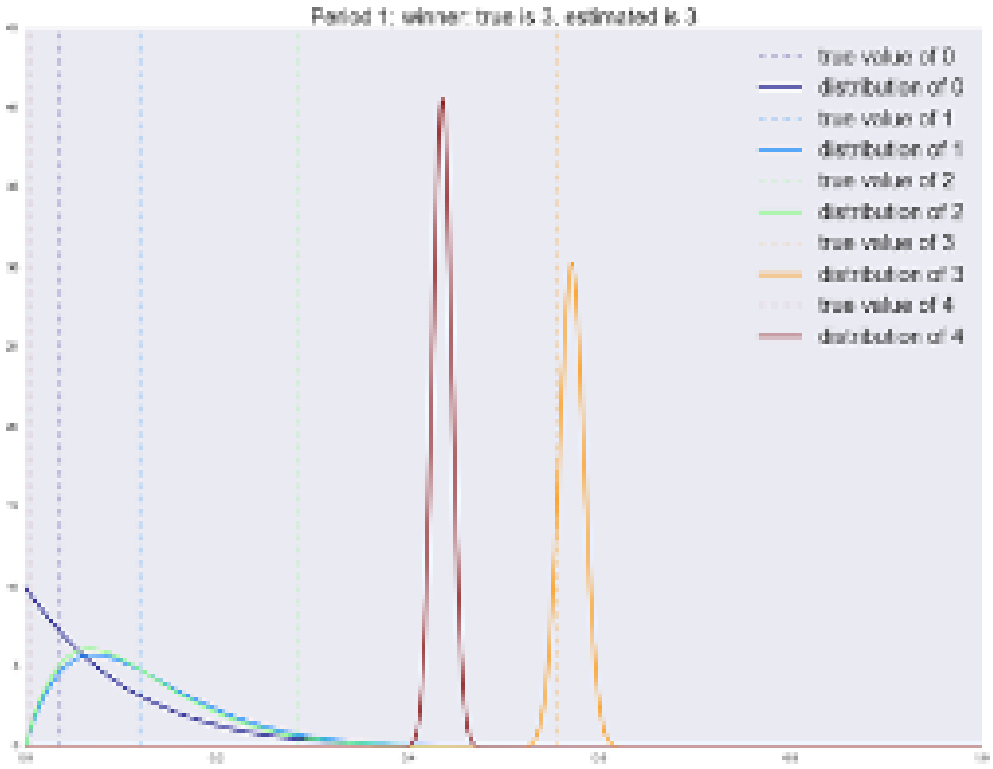


Figure 10: Bayesian MAB adjustment: period 2

Simulation with Preferences

The two graphs outline a simulation designed to evaluate how Bayesian Multi-Armed Bandits adapt to changing user preferences over time, particularly within the context of ad variant performance.

1. **First Period Analysis:** The process starts by identifying the ad variant with the highest success probability in the initial period, distinguished by a peak on the graph.

2. **Graph Interpretation:** Each graph reflects the expectation distribution for the ad variants, indicating the certainty of user preferences, with wider spreads suggesting higher uncertainty.

3. **Transition to Second Period:** Utilizing the data collected, the simulation advances to a new period, improving the accuracy of the ad variant performance predictions.

4. **Traffic Distribution:** Observations are made regarding how user traffic is distributed, noting that the variant identified as most successful initially receives a significant proportion of traffic.

5. **Challenges for New Variations:** The simulation considers the potential challenges faced by new ad variants in attracting user preference due to the gradual changes in user behavior.

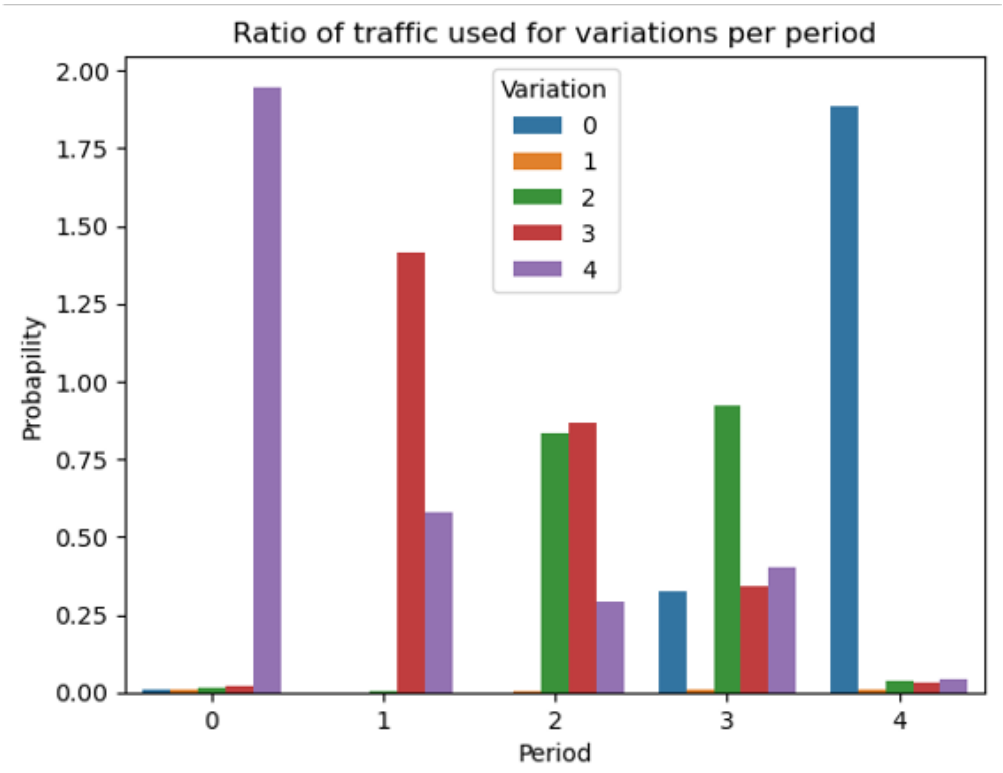


Figure 11: Ratio of traffic used for variations per period

The graph exemplifies the Bayesian MAB's data-driven approach to traffic allocation and provides a visual representation of the algorithm's ability to converge towards the optimal variant selection over time.