How Facebook applies Probability and Statistics to empower its features

65011304 Inthat Sappipat

01006719 PROBABILITY AND STATISTICS 1

B.Eng. Software Engineering, KMITL

Abstract

In the realm of social media, Facebook stands as a pioneer, continually evolving its features to cater to the dynamic needs of its vast user base. This term paper delves into the intricate interplay of Probability and Statistics within the framework of Facebook's functionalities. By employing these mathematical tools, Facebook not only refines user experience but also optimizes content delivery, advertising strategies, and overall platform performance.

The paper explores key areas where Probability and Statistics play a pivotal role, such as personalized content and targeted advertising algorithms. Through a detailed examination of these applications, we unravel the complex algorithms that underlie Facebook's ability to understand user behavior, predict preferences, and enhance engagement.

This exploration not only contributes to a deeper understanding of Facebook's operational dynamics but also underscores the broader implications for the future of social media platforms and their role in shaping our interconnected world.

Table of contents

Chapter 1: Introduction

Facebook has become a pervasive influence in today's digital exchanges, fostering friendships on an unprecedented scale by overcoming geographical borders. Behind the scenes, careful use of probability and statistics rather than just mechanical skill is what makes Facebook's features successful. In order to understand user behavior, anticipate preferences, and improve content delivery, Facebook uses a complex tapestry of algorithms and approaches.

As I navigate through the realms of probability and statistics within Facebook's ecosystem, it becomes evident that these mathematical tools are not mere abstract concepts but rather the foundation upon which user experiences are crafted. From the moment a user logs in, Facebook is crunching vast amounts of data, probabilistically predicting what posts, ads, or recommendations would resonate most with them. This nuanced understanding allows Facebook to tailor its content delivery, creating a personalized digital space for each user.

Moreover, the integration of Probability and Statistics extends beyond individual user experiences to fuel the platform's overarching functionality. Whether it's the complex algorithms governing the News Feed or the precision of targeted advertising, Facebook's reliance on these mathematical principles is a testament to the platform's commitment to delivering value to its users.

In this term paper, we'll look closely at how Facebook applies probability and statistics to empower its features, concentrating on targeted advertising, user behaviors, and personalized ranking. Also, it will focus on the 2 algorithms used by Facebook that is A/B testing and Bayesian bandits. I hope that anyone who reads this paper will gain a lot of knowledge and understand the various processes of Facebook.

Chapter 2: User Behaviors

Facebook leverages Probability and Statistics to understand and predict user behaviors on its platform. Through the analysis of vast amounts of user data, Facebook employs statistical models to identify patterns and trends in how users interact with the platform. This includes:

1. Engagement Prediction: Probability models are used to predict the likelihood of user engagement with various types of content—posts, videos, and ads. This helps Facebook prioritize and display content that is more likely to capture and retain user attention.

2. Click-through Rates: Probability is applied to estimate the likelihood of a user clicking on a particular link or ad. By analyzing historical data, Facebook refines its algorithms to improve the accuracy of predicting which content users are more likely to engage with.

3. User Segmentation: Probability and statistics are instrumental in creating user segments based on behavior. This segmentation allows for more personalized content recommendations and targeted advertising, enhancing the overall user experience.

Chapter 3: Personalized Ranking

Facebook employs sophisticated personalized ranking algorithms to tailor each user's feed to their preferences. This involves:

1. Content Recommendation: By analyzing a user's past interactions, such as likes, comments, and shares, Facebook predicts the probability of a user engaging with similar content in the future. This information is then used to personalize the content shown on the user's feed.

2. Friend Suggestions: Probability models help Facebook suggest friends to users based on common connections, interests, and interaction patterns. The algorithms assess the likelihood of a successful friend request and recommend connections accordingly.

3. Ad Targeting: Probability and statistical models are crucial in targeted advertising. Facebook predicts the probability of a user being interested in a particular product or service based on their online behavior. Advertisements are

then personalized to match individual user preferences, increasing the likelihood of conversion.

Chapter 4: Understanding the Facebook Algorithm

The Facebook algorithm governs the personalized ranking of content on the platform. It employs a mathematical expression, R(u, c), where factors like user interaction (I), user preferences (P), and content engagement (E) are weighted by α, β, and γ, respectively. These weights are dynamically adjusted based on user interactions and trends.

$$\boxed{R(u, c) = \alpha * I(u, c) + \beta * P(u, t) + \gamma * E(c)}$$

The algorithm aims to prioritize content based on three main signals:

1. Who posted it: Content from frequently interacted sources is prioritized, and represented in a weighted graph of user and content source relationships.

2. Type of content: User preference vectors capture affinities for different content types, ensuring users see more of what they engage with.

3. Interactions with the post: Higher engagement posts, especially from close connections, receive priority based on an engagement score calculated from metrics like likes, comments, and shares.

Users can customize their feed through actions like selecting favorites, hiding posts, and providing feedback, influencing the algorithm's weights and preferences for future content ranking.

Chapter 5: News Feed ranking algorithm with machine learning

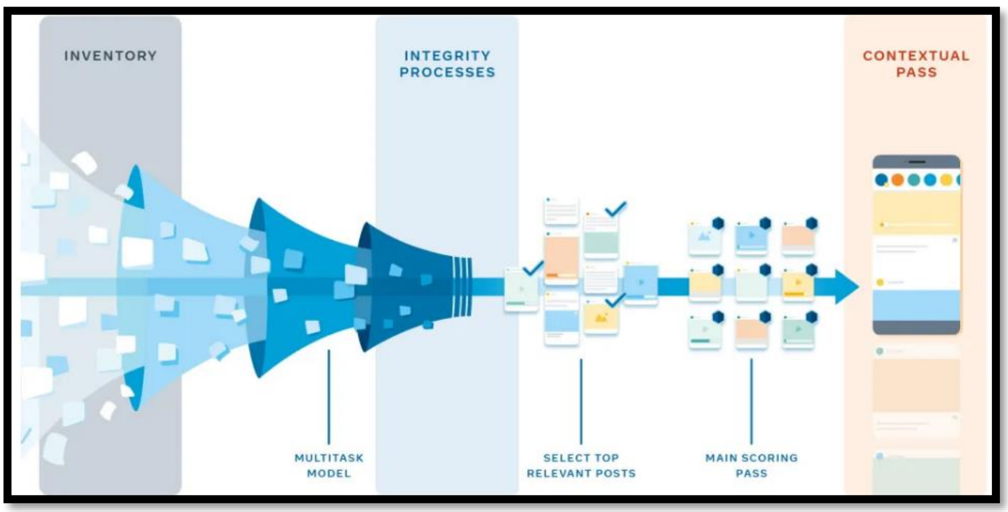
Facebook's News Feed ranking system is a massive undertaking, aiming to present personalized and relevant content to over 2 billion users. The system uses machine learning (ML) to predict what content matters most to each user, enhancing their experience. The ranking algorithm involves:

1. Objective Function: Facebook defines an objective function for each user, like Juan, considering factors such as user interactions, preferences, and content engagement. This is a single-objective optimization process.

2. User Preferences: ML models predict user engagement (e.g., likes, comments) for each post based on various attributes. For example, Juan's likelihood to engage with a running video from Saanvi is calculated using features like their interaction history and the type of content.

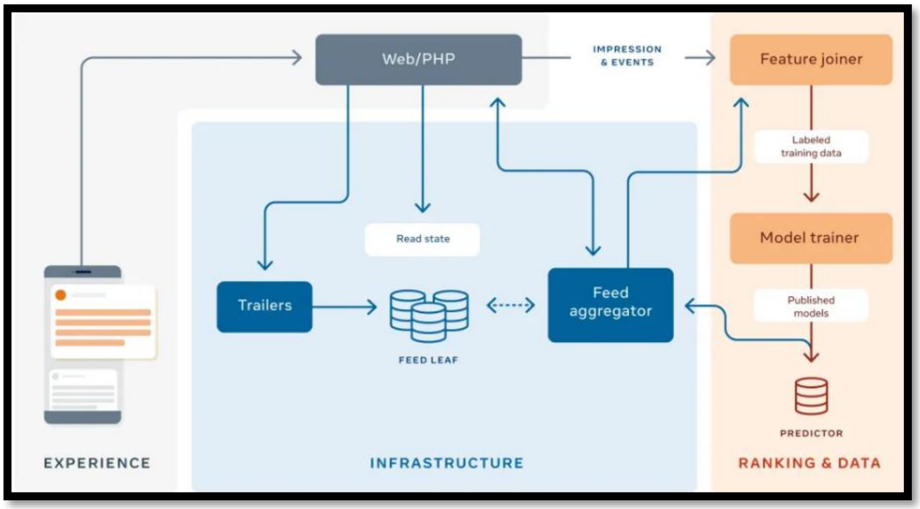
3. Multiple Objectives: The system optimizes for multiple objectives, such as likes, comments, and shares, aggregating them into a single value to create long-term value for users.

4. Survey-Based Metrics: To ensure alignment with user preferences, Facebook surveys users about meaningful interactions, guiding the algorithm to focus on what people find valuable.



The system's architecture involves a Web/PHP layer and a feed aggregator. The aggregator collects, analyzes, and ranks posts in real-time based on predictions from ML models. The ranking process includes multiple passes, considering content diversity and contextual features.

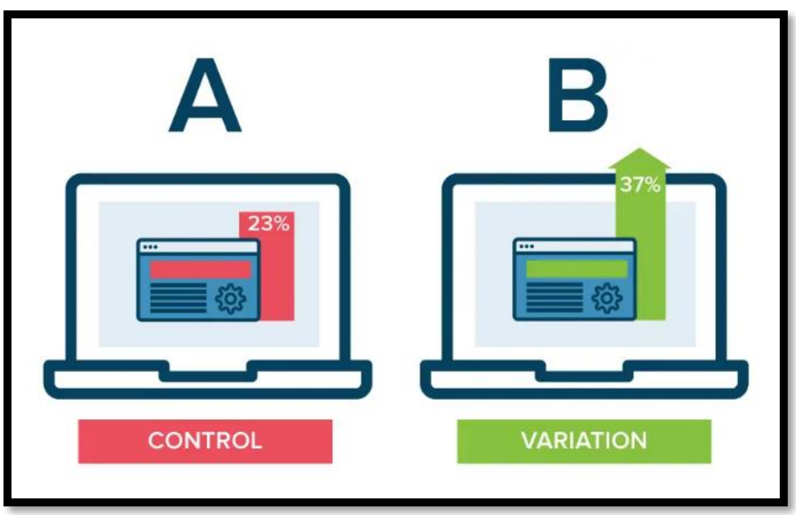
Facebook's News Feed ranking system leverages ML, neural networks, and sophisticated algorithms to predict and prioritize content, providing users with a more engaging and personalized experience. The process is complex but essential for delivering valuable content to billions of users worldwide.



Chapter 6: A/B testing

What is A/B testing?

A/B testing (also known as split testing or bucket testing) is a methodology for comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

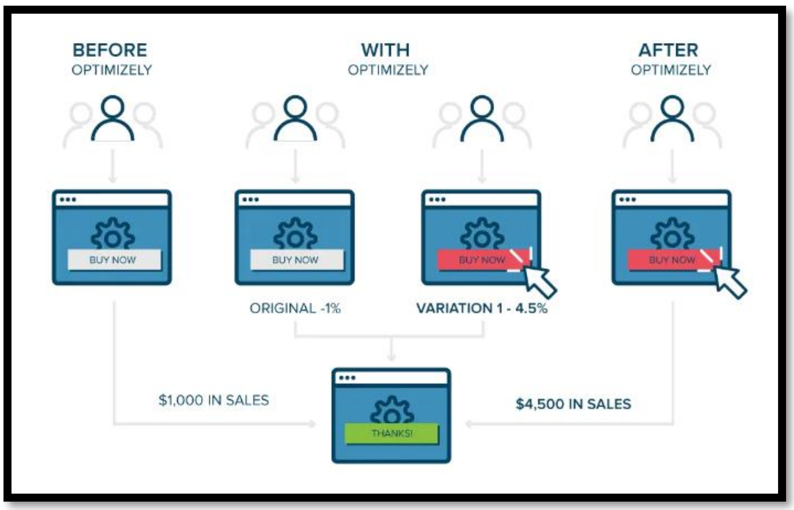


Running an A/B test that directly compares a variation against a current experience lets you ask focused questions about changes to your website or app and then collect data about the impact of that change.

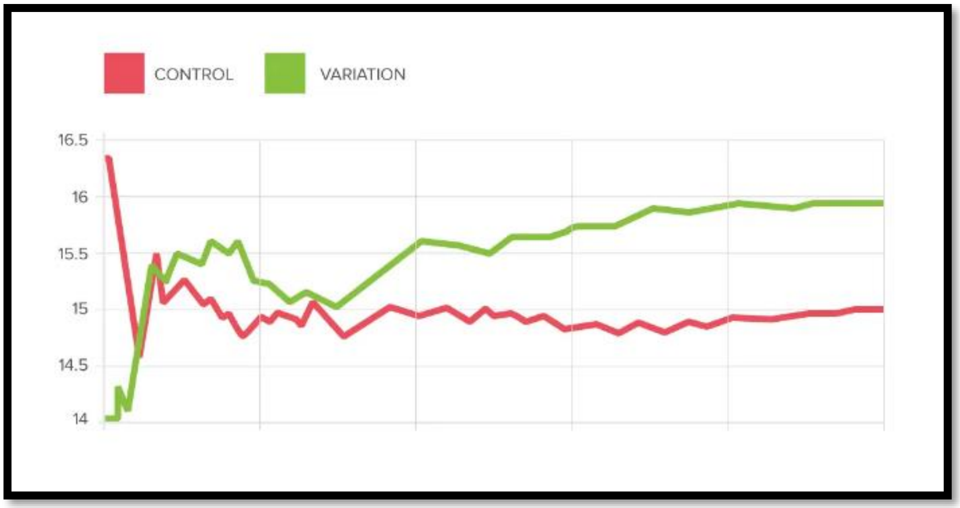
Testing takes the guesswork out of website optimization and enables data-informed decisions that shift business conversations from "we think" to "we know." By measuring changes' impact on your metrics, you can ensure that every change produces positive results.

How A/B testing works

In an A/B test, you modify a webpage or app screen to create a second version of the same page. This change can be as simple as a single headline, button, or complete page redesign. Then, half of your traffic is shown the original version of the page (known as control or A) and half is shown the modified version of the page (the variation or B).



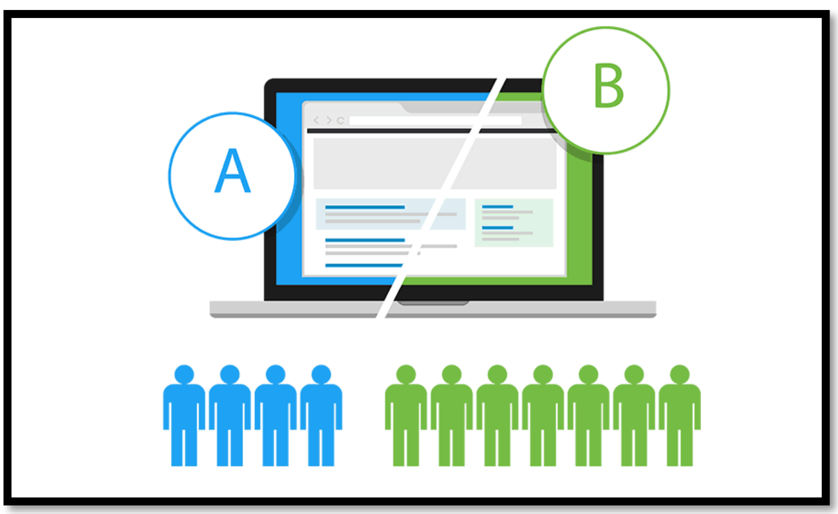
As visitors are served either the control or variation, their engagement with each experience is measured and collected in a dashboard and analyzed through a statistical engine. You can then determine whether changing the experience (variation or B) had a positive, negative, or neutral effect against the baseline (control or A).



A/B Testing Approach on Facebook

One of the methods for assessing and determining the optimal performance plan is A/B testing. Notwithstanding its potential benefits, an A/B test may not be the optimal approach for targeting online advertisements given its multiple drawbacks and inefficient use of resources.

Although this testing methodology can offer insightful information in a variety of scenarios, evaluating the numerical statistics for each algorithm strategy necessitates a set testing period and sample size. This implies that there will be an initial outlay of funds as well as time and opportunity for the advertising.



Challenges with simple A/B testing approaches

While winner-takes-all A/B tests may seem straightforward, they can lead to unnecessary expenditures for advertisers. There are three primary challenges associated with A/B tests:

1. Changing Consumer Preferences: Over time, consumer preferences evolve. Declaring a single creative as the perpetual winner assumes it should always be displayed. However, this may not hold true, as the A/B test might have been conducted during a specific season or on particular days. For example, a food delivery app might find a salad to be a winning creative in the summer but hot chocolate as the winner in winter. The concept of a clear 'winner' or 'loser' is not always accurate, introducing the possibility of false negatives and positives.

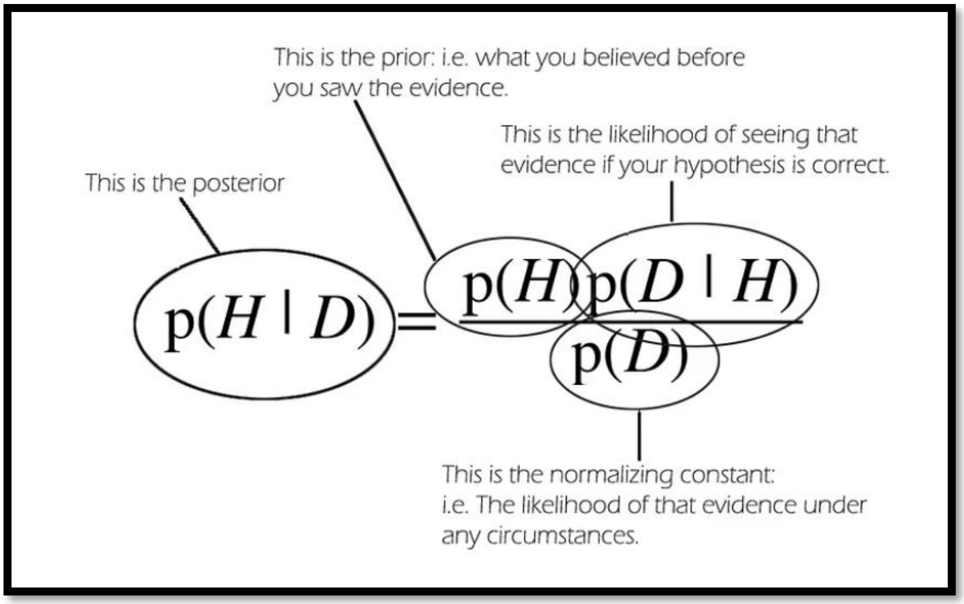
2. Probabilistic Effectiveness of Ads: In real-life scenarios, all ads have some level of effectiveness. The probabilistic nature of reality means that a winning creative isn't universally the best for 100% of users or impressions, as sometimes implied in A/B testing paradigms. The term 'winning' indicates better performance most of the time, but there could be instances where the 'losing' creative is more effective, such as in 10% of impressions or for specific audiences. Even 'bad' ads can lead to purchases, and 'good' ads may result in no purchases.

3. Optimal Allocation of Impressions: Equally distributing impressions can lead to inefficient spending. If a winning creative is superior 90% of the time and a losing creative is better 10% of the time, a 50-50 allocation of spend between them can result in wasted resources. The losing creative may end up being more expensive to run more frequently than its actual effectiveness justifies.

Chapter 7: The Bayesian Bandits Paradigm

Bayesian Bandits

For precision in targeted ads, Facebook utilizes a Bayesian approach to adjust its algorithm. With an expansive user base and a wealth of data, Facebook holds a strategic advantage. This approach ensures that the algorithm optimizes targeting performance, maximizing profits while safeguarding assets.



Using pre-collected data, the Bayesian Bandits Paradigm methodically assesses which strategy results in the highest purchase rate. The algorithm cycles through these phases, dynamically updating the probability of each option, during the advertising journey.

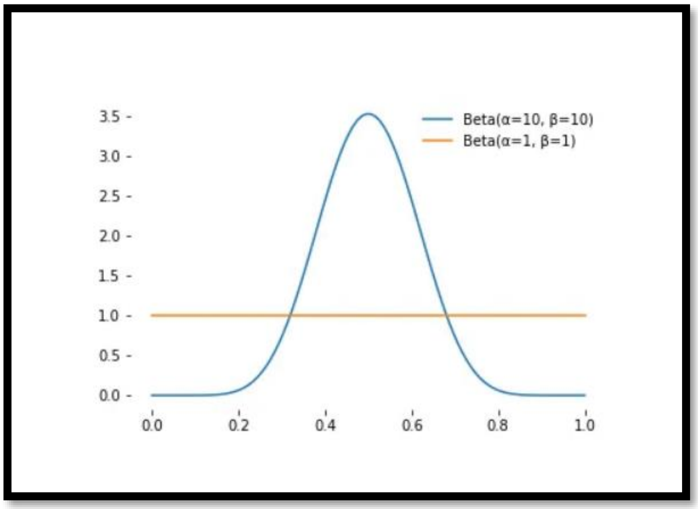
This approach provides real-time flexibility, skillfully balancing the investigation of novel ad variants with the utilization of preexisting knowledge. Through continuous optimization of its comprehension of ad success, the algorithm quickly adapts to changes in user behavior or market conditions.

Beta Distribution

Normally, the beta distribution is a continuous probability distribution defined on the interval [0, 1] parametrized by two positive shape parameters, denoted by α and β. The PDF of the beta distribution is:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

And, the pdf looks like the below for different values of α and β:



Updating the Prior

The reason that the beta distribution is used for modeling the probabilities is that this distribution has a great mathematical property.

$$\underbrace{\text{Beta}}_{\text{prior}} \cdot \overbrace{\text{Binomial}}^{\text{data}} = \overbrace{\text{Beta}}^{\text{posterior}}$$

If the prior is f(α,β), then the posterior distribution is again beta, given by f(α+#success, β+#failures)

Chapter 8: Conclusion

Facebook's strategic utilization of Probability and Statistics, particularly through Bayesian approaches and the Bayesian Bandits Paradigm, underscores its commitment to optimizing the user experience and maximizing advertising efficiency. By employing sophisticated algorithms and machine learning models, Facebook not only predicts user behaviors but also tailors content and advertisements to individual preferences in real time.

The A/B testing methodology, while valuable in certain scenarios, presents challenges in the dynamic realm of online advertising, where consumer preferences evolve, and the probabilistic effectiveness of ads comes into play. Recognizing these challenges, Facebook's adoption of Bayesian Bandits provides a more adaptive and efficient alternative, allowing for continuous optimization and quick adaptation to changes in user behavior and market dynamics.

The incorporation of the beta distribution further enhances the modeling of probabilities, showcasing Facebook's commitment to employing robust mathematical tools for precise decision-making. Overall, Facebook's multifaceted approach, combining statistical methods, machine learning, and adaptive algorithms, contributes to the platform's success in delivering personalized and engaging content to its vast user base.

Chapter 9: References

1. https://mobiledevmemo.com/bayesian-bandits-behind-the-scenes-of-facebooks-spend-allocation-decisioning/

2. https://www.optimizely.com/optimization-glossary/ab-testing/

3. https://towardsdatascience.com/bayesian-bandits-explained-simply-a5b43d9d5e38

4. https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/

5. https://www.linkedin.com/pulse/mastering-facebook-algorithm-comprehensive-guide-insights-patel/