Bayesian A/B Testing Analysis for Marketing Experiment

By: Aditya Prabhu

Table of Contents

Objective

Problem:
Marketing and product teams often use A/B testing to determine which variant yields better results. Traditional frequentist approaches provide p-values, which are commonly misinterpreted and do not directly answer the question "How likely is variant B better than variant A?"

Goal:
In this project, I analysed a synthetic A/B test dataset using both frequentist and Bayesian approaches to:

- Compare interpretation differences
- Calculate the probability that the treatment variant is better
- Provide clearer, probability-based business recommendations

Business Impact:
Bayesian A/B testing provides intuitive, actionable insights, enabling confident decisions in campaign optimisation, UI changes, and product feature launches.

Data Understanding

Dataset Overview:
For this project, I generated a synthetic A/B testing dataset to simulate a real-world online experiment. The dataset reflects a typical marketing or product test scenario where two variants are tested to evaluate their impact on user conversions.

Key Variables:

- **user_id:**
  A unique identifier assigned to each user participating in the experiment. This ensures that each observation represents an individual user without duplication.

- **group:**
  Indicates the variant each user was exposed to. It has two categories:

  - **Control:** Users who were shown the existing version (baseline).

  - **Treatment:** Users who were shown the new variant intended to improve conversion.

- **converted:**
  A binary variable indicating whether the user converted (1) or not (0). In a business context, this could represent actions such as signing up for a service, clicking a CTA button, or completing a purchase.

Dataset Generation Details:

- **Total Users:** 10,000
  The dataset includes 10,000 users randomly assigned to control and treatment groups to simulate an adequately powered A/B test.

- **Group Distribution:**
  Users were assigned to control and treatment groups randomly to avoid selection bias and ensure comparable groups.

- **Conversion Rates:**

- **Control Group:** Conversion rate was simulated at 10%, representing typical baseline conversion for an online product or marketing campaign.
- **Treatment Group:** Conversion rate was simulated at 12%, indicating an uplift over the control to test detection capability.

Initial Exploratory Findings:

1. **Balanced Group Assignment:**
   The group assignment check confirmed that control and treatment groups had approximately equal sample sizes, as expected in a well-designed randomised experiment.
2. **Conversion Outcome Distribution:**
   Both groups showed binary conversion outcomes (0/1) with no missing values, confirming dataset integrity for analysis.
3. **Relevance to Business Context:**
   This dataset structure mirrors typical online A/B testing data used in marketing campaigns, landing page tests, UI experiments, or email conversion tests. It thus allows me to demonstrate analysis skills that directly translate to real-world business scenarios.

Data Cleaning

To ensure accurate and reliable analysis, I performed the following data cleaning and validation steps:

1. **Verified Group Assignment Distributions:**
   I began by checking the distribution of users across the control and treatment groups. This confirmed that the random assignment process produced roughly equal group sizes, which is essential to minimise allocation bias and ensure that observed differences in conversion rates are attributable to the treatment effect rather than imbalanced sample sizes.

2. **Checked for Missing Values:**
   I inspected all columns for missing data. Missing values in A/B test datasets can arise from incomplete user interactions or data collection issues. In this synthetic dataset, no missing values were found, ensuring that all records could be included in the analysis without imputation or removal.

3. **Validated Data Types and Formats:**
   I ensured that each column had the correct data type for analysis:

   - `user_id` as integers
   - `group` as categorical variables (control/treatment)
   - `converted` as binary integers (0/1)

4. This step prevents downstream errors during statistical tests and visualisation.

5. **Checked for Duplicate Records:**
   Duplicate user records can inflate conversion counts and bias results. I confirmed that each `user_id` was unique, ensuring that every observation represented a distinct user in the test.

6. **Ensured Correct Conversion Rate Simulation:**
   As this was a synthetic dataset, I validated that the simulated conversion rates reflected the intended design: 10% for control and 12% for treatment. This confirmed that the dataset generation process functioned as planned, making it suitable for both frequentist and Bayesian analyses.

Outcome of Data Cleaning:

The dataset was confirmed to be clean, well-structured, and ready for analysis, with no missing data, duplicates, or data type issues. This ensured the reliability of subsequent frequentist and Bayesian statistical tests, enabling me to derive valid and actionable business insights.

Frequentist Analysis

Performed a two-proportion z-test to assess statistical significance.

Metric	Control	Treatment
Conversion Count	5013	4987
Sample Size	0.100339	0.118709
Conversion Rate	10%	12%

Results:

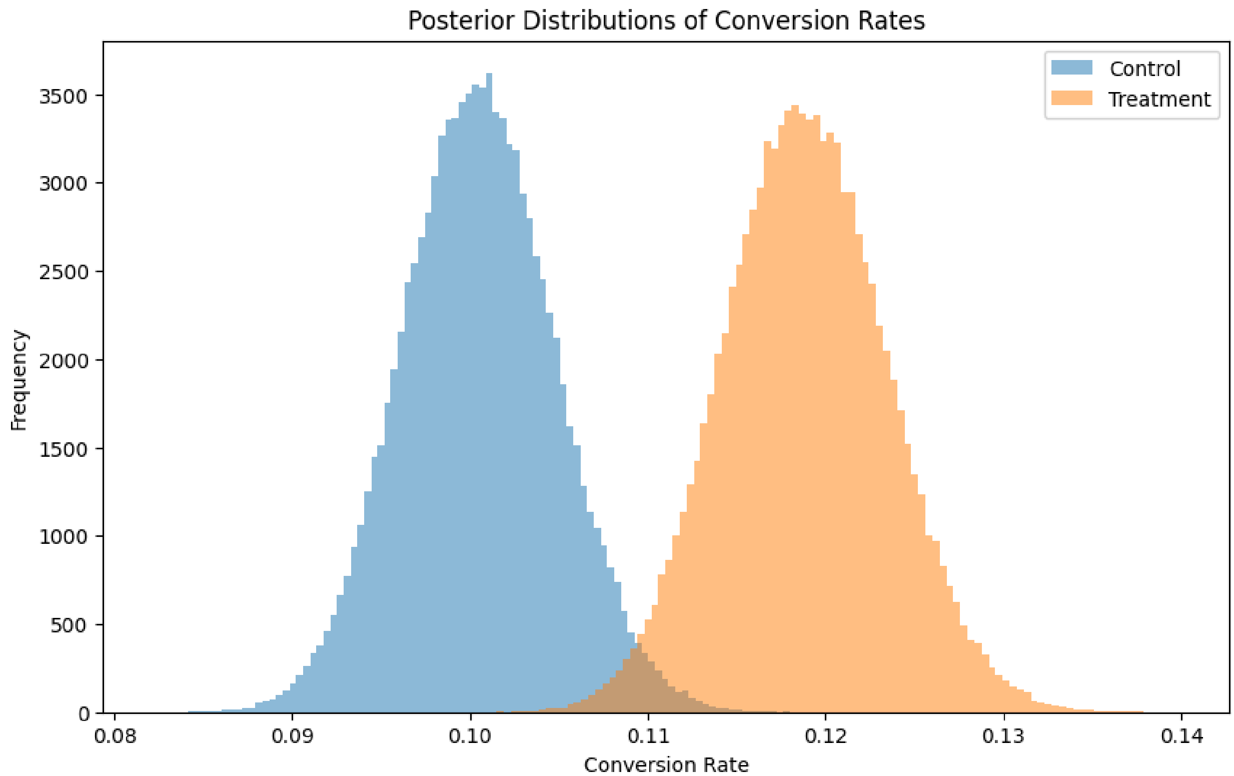
- **Z-test Statistic:** 2.94
- **P-value:** 0.0016

Interpretation:
With $p < 0.05$, the treatment conversion rate is statistically significantly higher than the control.

Bayesian Analysis

Applied Bayesian inference using Beta distributions to estimate posterior probabilities of conversion rates.

Posterior Distribution Visualisation



Key Result:

- **Probability that Treatment is Better than Control:** 99.86%

Business Recommendations

1. **Implement Treatment Variant Confidently**
   Why: Bayesian analysis indicates a 99.86% probability that the treatment outperforms control, suggesting a clear uplift in conversion.
   Recommendation: Roll out treatment variant to all users to maximise conversion and revenue.

2. **Adopt Bayesian Framework for Future Experiments**
   Why: Provides intuitive, actionable probability statements rather than abstract p-values.
   Recommendation: Incorporate Bayesian analysis into standard A/B testing procedures for clearer communication with stakeholders.

3. **Monitor Long-term Impact**
   Why: While the short-term test shows strong uplift, continuous monitoring ensures sustained performance.
   Recommendation: Track post-implementation KPIs to validate uplift and identify further optimisation opportunities.

Conclusion

In this project, I compared frequentist and Bayesian approaches to A/B testing using a synthetic dataset. The Bayesian framework demonstrated clearer, probability-based insights, enabling confident business decisions with quantifiable certainty.

This project demonstrates my ability to apply advanced statistical methodologies to derive strategic, business-focused recommendations.