A/B Testing Project: Bayesian vs. Frequentist Inference

Andrés González Ortega

15 de Junio de 2025

What Was Done in This Notebook

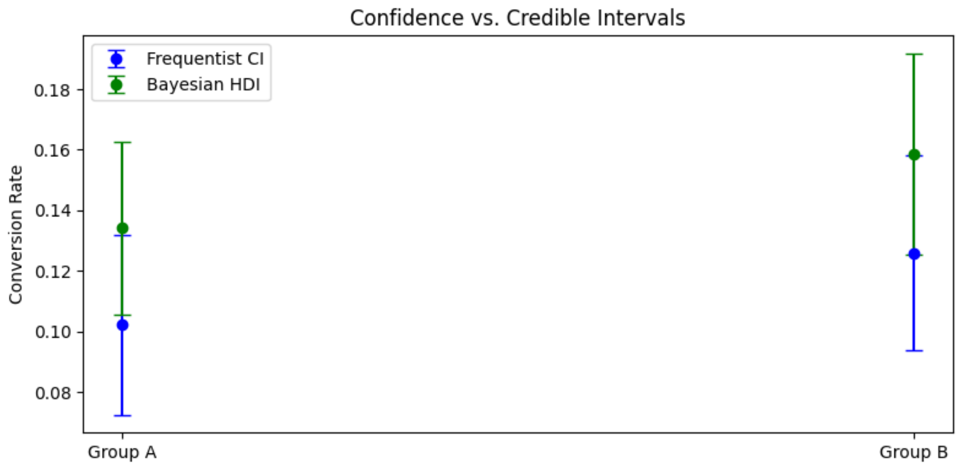
This notebook walks through a real-world-style A/B testing problem using two approaches to statistics: frequentist and Bayesian inference. The objective is to determine whether version B of a product or webpage performs better than version A based on simulated user conversion data.
In simpler terms, we generated fake data to simulate 1,000 users split equally between two groups. Each user either "converted" (like clicked a button or bought something) or didn't. The question is: which group has a higher conversion rate?

What the Code Did, Step by Step

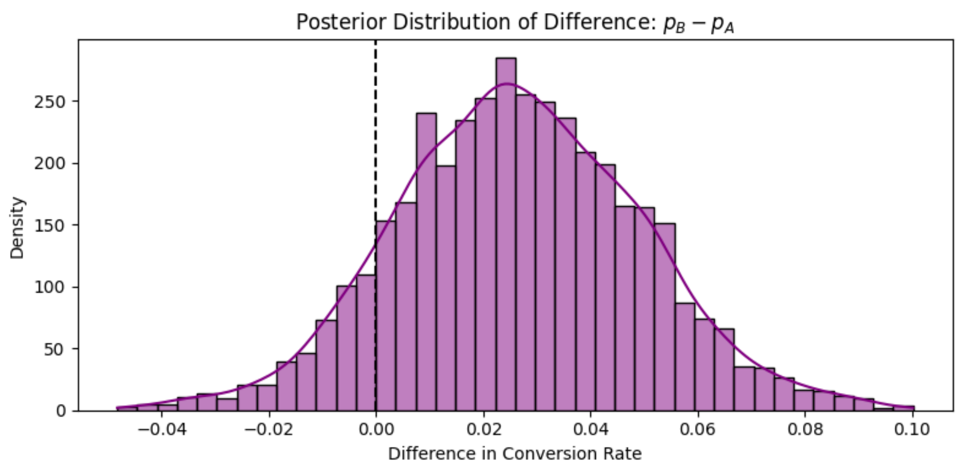
1. The data was simulated using `numpy`, with known true probabilities of conversion.
2. A summary was created to compute and visualize conversion rates per group.
3. A frequentist hypothesis test (z-test) and confidence intervals were calculated using `statsmodels`.
4. A Bayesian model was created using `PyMC`, defining priors with a Beta distribution and sampling from the posterior.
5. Posterior probabilities were extracted to check how likely it is that group B performs better than group A.
6. We visualized both frequentist and Bayesian intervals to compare interpretations.
7. Finally, a decision analysis was made using expected utility theory. We evaluated whether switching to group B would be worth it depending on the potential gains or losses.

Key Visual Outputs

Below is the visualization comparing frequentist confidence intervals with Bayesian credible intervals:



And here is the posterior distribution of the difference between the two groups, $p_B - p_A$, showing how much better B might be compared to A:



Conclusion in Plain Language

The frequentist analysis gave us a p-value of around 0.24, which by conventional standards isn't enough to say "B is better". However, the Bayesian analysis showed that there's an 88.4 % chance that group B is better than A. That's not definitive, but it's suggestive.
To help make a decision, we used the idea of expected utility. If switching to B gains you 100 points when you're right and costs you 50 when you're wrong, then the expected outcome is still positive. So even without "statistical significance", you might still want to switch to B.

This highlights how Bayesian methods can help in real-world decision-making, especially when you're not just looking for certainty but trying to weigh risks and rewards.