Implementing Bayesian A/B Testing

Janani Ravi
CO-FOUNDER, LOONYCORN
www.loonycorn.com

Overview

Understanding A/B testing

Choosing the right test based on target variable distributions

Understanding Bayes' theorem

Understanding Bayesian A/B testing

Representing priors as distributions

Implementing Bayesian A/B testing

"If this is coffee, please bring me some tea; but if this is tea, please bring me some coffee."

Abraham Lincoln

Blue Pill, Red Pill

Version A

Website with a blue 'Buy Now' button

Version B

Website with a red 'Buy Now' button

All else is identical. Which version works better?

A/B Testing

A randomized experiment used to compare two versions of a single variable by testing subjects' responses to the two variants to find which version is more effective.

A/B Testing

A randomized experiment used to compare **two versions of a single variable** by testing subjects' responses to the two variants to find which version is more effective.

A/B Testing

A randomized experiment used to compare two versions of a single variable by **testing subjects' responses to the two variants** to find which version is more effective.

A/B Testing

A randomized experiment used to compare two versions of a single variable by testing subjects' responses to the two variants to **find which version is more effective**.

A/B Testing

A **randomized experiment** used to compare two versions of a single variable by testing subjects' responses to the two variants to find which version is more effective.

A/B Testing

A randomized experiment used to compare two versions of a single variable by testing subjects' responses to the two variants to find which version is more effective.

Blue Pill, Red Pill

Version A

Website with a blue ‘Buy Now’ button

Version B

Website with a red ‘Buy Now’ button

All else is identical. Which version works better?

Blue Pill, Red Pill

Version A (Control)

Website with a blue ‘Buy Now’ button

Version B (Treatment)

Website with a red ‘Buy Now’ button

Run a statistical test to find out

A/B Testing

What statistical run to test?

Depends on the distribution of the target variable being measured

- Click-through rate
- Transactions per user
- Revenue per user
- Shopping basket (Product, Quantity)

Distributions and Statistical Tests

A/B Testing

What statistical run to test?

Depends on the distribution of the target variable being measured

- Click-through rate
- Transactions per user
- Revenue per user
- Shopping basket (Product, Quantity)

A/B Testing

What statistical run to test?

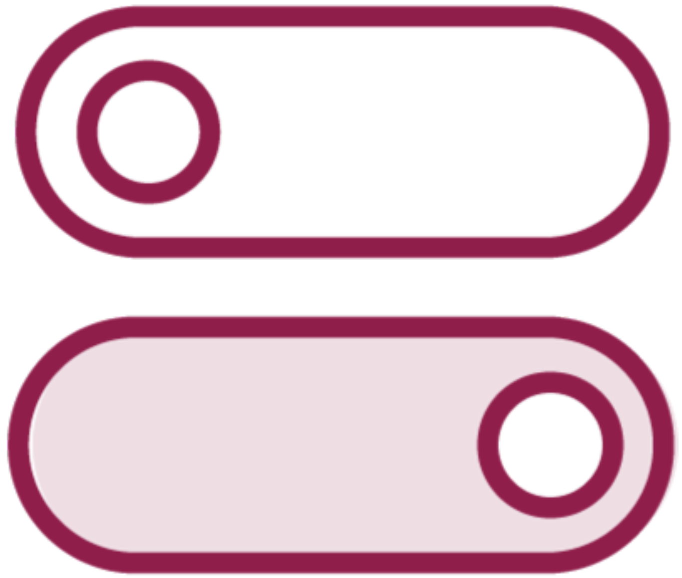
Depends on the distribution of the target variable being measured

- Click-through rate
- Transactions per user
- Revenue per user
- Shopping basket (Product, Quantity)

Binomial Distribution

Discrete probability distribution of number of successes in a sequence of independent experiments, each with two possible outcomes.

Binomial Distribution



Use to model number of successes

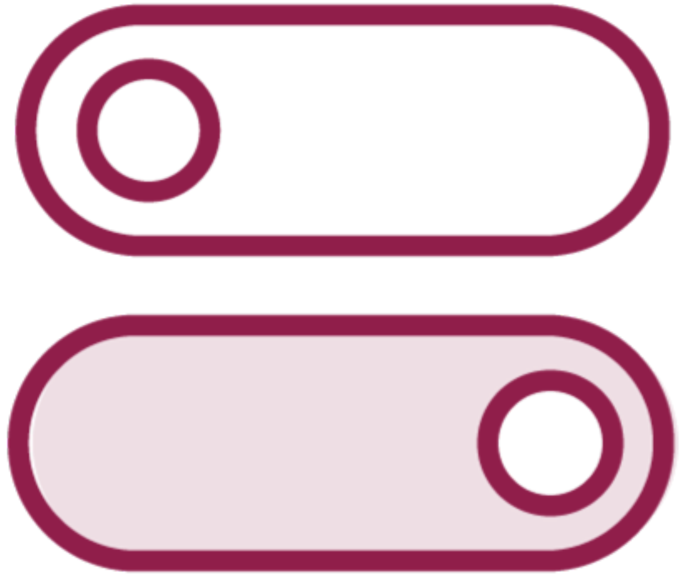
From N trials

Each trial outcome is binary, independent

Each trial is called a Bernoulli trial

Also called Bernoulli Distribution

Binomial Distribution



Number of heads in n coin tosses

Coin need not be fair

Probability of success = p

Probability of failure = 1-p

Defined by two parameters n, p

Multinomial Distribution

Discrete probability distribution of number of outcomes of each type in a sequence of independent experiments, each with k possible outcomes.

Multinomial Distribution

Discrete probability distribution of number of outcomes of each type in a sequence of independent experiments, each with **k possible outcomes**.

Multinomial Distribution

Generalization of Binomial distribution

Binomial

- Each trial like tossing a coin

Multinomial

- Each trial like rolling a dice

Poisson Distribution

Discrete probability distribution of number of events occurring within fixed time interval; events occur with known mean rate and independently of time of last event.

Poisson Distribution

Discrete probability distribution of **number of events occurring within fixed time interval**; events occur with known mean rate and independently of time of last event.

Poisson Distribution

Discrete probability distribution of number of events occurring within fixed time interval; events occur with known **mean rate and independently of time of last event**.

Poisson Distribution

Great for modeling rates of discrete events

Defined by single parameter - mean rate

Rates are often per unit time

Can also be per unit area, volume, ...

Poisson Distribution

Number of call center queries per hour

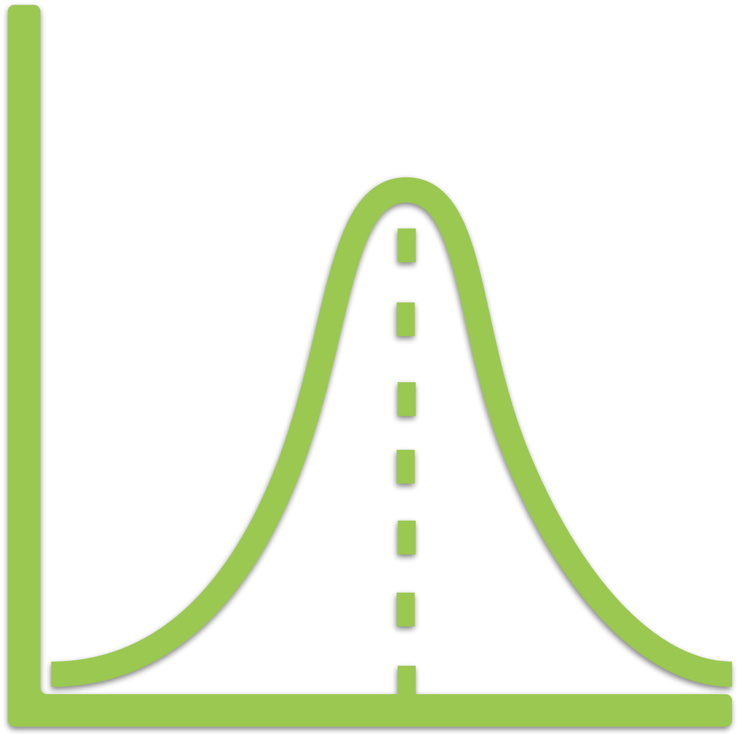
Number of buses at stop per hour

Number of meteorites striking in given interval

Gaussian (Normal) Distribution

Continuous probability distribution for a real-valued (continuous) random variable.

Normal Distribution



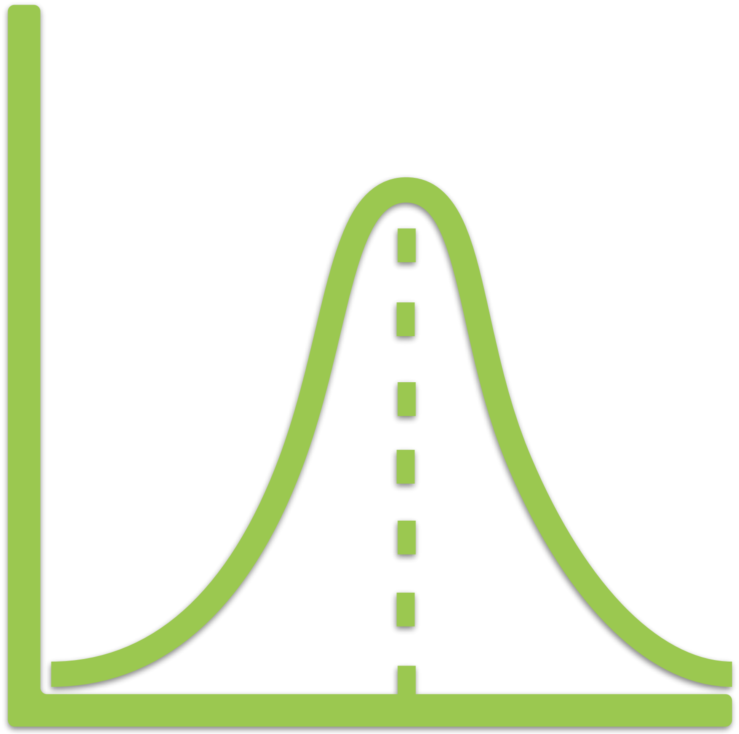
Bell curve - widely occurs in nature

Distribution of heights, weights

Many special properties

- Average of IID non-normals tends to normal (Central Limit Theorem)

Normal Distribution



Defined by two parameters: mean and standard deviation

Great for modeling averages (sums) of continuous quantities

A/B Testing

What statistical run to test?

Depends on the distribution of the target variable being measured

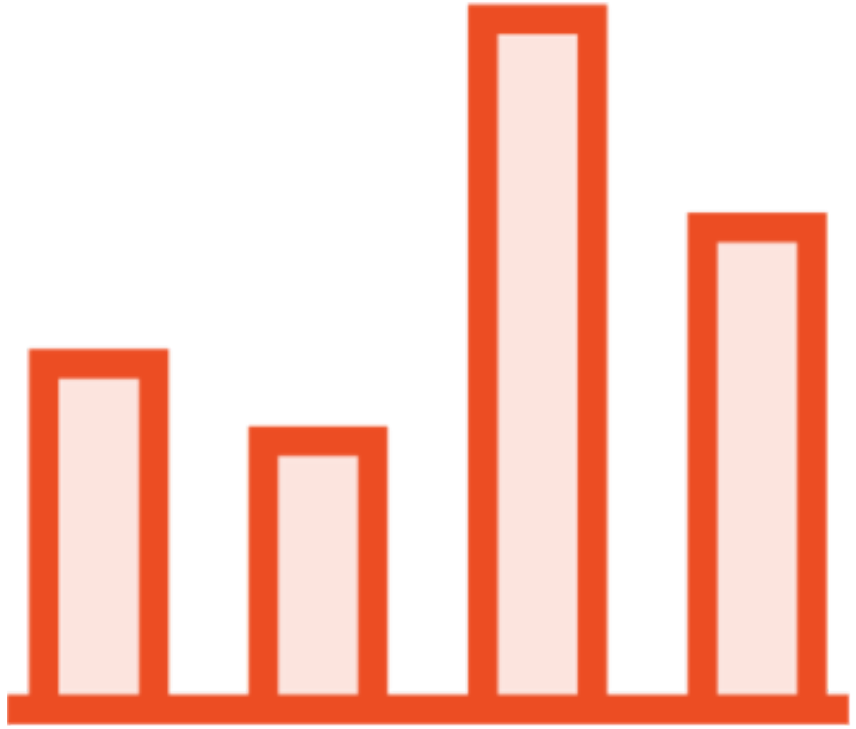
- Click-through rate
- Transactions per user
- Revenue per user
- Shopping basket (Product, Quantity)

A/B Testing

Use Case	Distribution	Statistical Test
Click-through Rates	Binomial	Fisher’s test
Transactions per session	Poisson	E-test
Revenue per user	Gaussian	Student’s t-test or Welch’s test
Shopping basket analysis	Multinomial	Chi-squared test

This approach to A/B testing using hypothesis testing is known as the **Frequentist Approach**

Frequentist Approach



Considers only samples for a certain time interval

Does not take into account prior experience or judgement

Does not answer the question “How likely is it that B is better than A and by how much”?

An alternative approach is often superior: the **Bayesian Approach**

Bayes' Theorem

Binary Classification Problem

Runner

Police Officer

Classify a person who jogs past you on the street

A Priori Probabilities

Items	Occurence
Runners	9
Police officers	1
Total	10

Observation 1: Today is the city marathon, more runners than police officers out on the streets

A Priori Probabilities

P(Runner) = 9/10

P(Police Officer) = 1/10

These are ***a priori* probabilities:** before anything specific about the person is known

Conditional Probabilities

Handcuffs

Walkie-Talkie

Running Shoes

Observation 2: Specific items appear more often with one category than with the other

Conditional Probabilities

Item	Occurrences with Police Officers	Occurrences with Runners
Handcuffs	6	0
Running Shoes	2	8
Gun	9	0
Badge	8	0
Walkie-Talkie	8	3

Upon Closer Examination

Handcuffs

Badge

The person that zipped past carried these two items

Applying Bayes' Theorem

P(Runner/ Handcuffs,Badge)	=	Probability that a person carrying handcuffs and a badge is a runner

Step 1: Find probability that this person is a runner

Applying Bayes' Theorem

P(Police Officer/ Handcuffs,Badge)	=	Probability that a person carrying handcuffs and a badge is a police officer

Step 2: Find probability that this person is a police officer

Applying Bayes' Theorem

Compare

P(Police Officer/
Handcuffs,Badge)

and

P(Runner/
Handcuffs,Badge) =

Step 3: Pick the label with the higher probability

Jogger Is a Police Officer

P(Police Officer/ Handcuffs,Badge) > P(Runner/ Handcuffs,Badge)

Jogger Is a Marathon Runner

P(Police Officer/ Handcuffs,Badge) < P(Runner/ Handcuffs,Badge)

Bayesian A/B Testing

Blue Pill, Red Pill

Version A

Website with a blue ‘Buy Now’ button

Version B

Website with a red ‘Buy Now’ button

All else is identical. Which version works better?

Blue Pill, Red Pill

Version A (Control)

Website with a blue 'Buy Now' button

Version B (Treatment)

Website with a red 'Buy Now' button

Run a statistical test to find out

A/B Testing

Show Version A to N_a users

Show Version B to N_b users

Frame null hypothesis that target variable has same value for both samples

Perform statistical hypothesis test

Accept or reject null hypothesis

This approach to A/B testing using hypothesis testing is known as the **Frequentist Approach**

An alternative approach is often superior: the **Bayesian Approach**

Frequentist Inference

Type of statistical inference that draws conclusions from samples using frequencies (proportions).

Bayesian Inference

Type of statistical inference that uses Bayes' Theorem to calculate probability of a hypothesis being true.

Bayesian Inference

Type of statistical inference that uses Bayes' Theorem to **calculate probability of a hypothesis being true**.

Approaches to A/B Testing

Frequentist A/B Testing	Bayesian A/B Testing
Simple	Relatively complex
Looks back to fixed interval	Can be updated, so works with live data
No role for prior experience	Explicitly considers prior experience
Does not require calibration of probability distribution of prior	Explicitly requires prior probability distribution to be calibrated (to include prior experience)

Blue Pill, Red Pill

Version A

Website with a blue ‘Buy Now’ button

Version B

Website with a red ‘Buy Now’ button

All else is identical. Which version works better?

Probabilities in A/B Testing

CTR

Click-through rate before (prior) we know which version

CTR/A

Click-through rate after (posterior) we know version A was viewed

CTR/B

Click-through rate after (posterior) we know version B was viewed

Bayesian A/B Testing

CTR

Click-through rate before (prior) we know which version

Known from experience

CTR/A

Click-through rate after (posterior) we know version A was viewed

CTR/B

Click-through rate after (posterior) we know version B was viewed

Bayesian A/B Testing

CTR

Click-through rate before (prior) we know which version

CTR/A

Click-through rate after (posterior) we know version A was viewed

CTR/B

Click-through rate after (posterior) we know version B was viewed

Bayesian A/B Testing

CTR

Click-through rate before (prior) we know which version

CTR/A

Click-through rate after (posterior) we know version A was viewed

CTR/B

Click-through rate after (posterior) we know version B was viewed

Bayesian A/B Testing

CTR

Click-through rate before (prior) we know which version

CTR/A

Click-through rate after (posterior) we know version A was viewed

CTR/B

Click-through rate after (posterior) we know version B was viewed

Need to estimate

Bayesian A/B Testing

CTR

Click-through rate before (prior) we know which version

CTR/A

Click-through rate after (**posterior**) we know version A was viewed

CTR/B

Click-through rate after (**posterior**) we know version B was viewed

Conjugate Prior

In Bayesian inference, prior probability is not a constant

- Rather, itself follows a probability distribution
- Conjugate Prior Distribution

Important difference from frequentist approach

- Where prior is assumed constant (no probability distribution)

Conjugate Prior

Can calibrate Conjugate Prior probability distribution based on experience

Standard references give distribution of prior for different types of target variables

e.g. if target is Binomial, conjugate prior follows Beta Distribution

Choice of Conjugate Prior

Target Variable	Distribution of Target Variable	Distribution of Prior (Conjugate Prior)
Click-through Rates	Binomial	Beta Distribution
Transactions per session	Poisson	Gamma Distribution
Revenue per user	Gaussian	Gaussian
Shopping basket analysis	Multinomial	Dirichlet

Calibrating Conjugate Prior

Use past experience to calibrate conjugate prior

For instance, say we know from experience that CTR = 0.25

Tweak Beta distribution to get a peak (average) at 0.25

Feed this calibrated conjugate prior into procedure

Calibrating Conjugate Prior

Well-calibrated conjugate prior encapsulates past experience

Procedure can now generate scenarios based on this conjugate prior

Monte Carlo simulation used under the hood by R package

```
Average Click-through Rate = 0.25
```

Based on prior knowledge

Before any information about Version A or Version B incorporated into model

`P(Click-through)`

Distribution of Conjugate Prior

Given by Beta Distribution (since target variable is Binomial). Average of this Beta Distribution = 0.25 (from prior knowledge)

```
P(Click-through/Version A)
P(Click-through/Version B)
```

Posterior Probabilities - Need to find

Can use Bayes' Theorem to estimate these given distribution of conjugate prior

$$P(\text{Click-through}/\text{Version A}) = \frac{P(\text{Version A}/\text{Click-through}) \times P(\text{Click-through})}{P(\text{Version A})}$$

Bayes' Theorem

Can use Bayes' Theorem to estimate these given distribution of conjugate prior

$$P(\text{Click-through}/\text{Version B}) = \frac{P(\text{Version B}/\text{Click-through}) \times P(\text{Click-through})}{P(\text{Version B})}$$

Bayes' Theorem

Can use Bayes' Theorem to estimate these given distribution of conjugate prior

$$P(\text{Click-through}/\text{Version A}) = \frac{P(\text{Version A}/\text{Click-through}) \times \mathbf{P(Click\text{-}through)}}{P(\text{Version A})}$$

Conjugate Prior Probabilities Required

This is why we need to pass in a calibrated Conjugate Prior probability distribution

$$P(\text{Click-through}/\text{Version A}) = \frac{P(\text{Version A}/\text{Click-through}) \times P(\text{Click-through})}{P(\text{Version A})}$$

Beta Distribution B(α,β)

Defined by parameters α and β. Mean = α / (α + β)

$$P(CTR = p) = \frac{p^{(\alpha-1)}\ (1-p)^{(\beta-1)}}{\int_{x=0}^{x=1} x^{(\alpha-1)}\ (1-x)^{(\beta-1)}}$$

Beta Distribution B(α,β)

Defined by parameters α and β. Mean = α / (α + β)

Interpreting Bayesian A/B Test

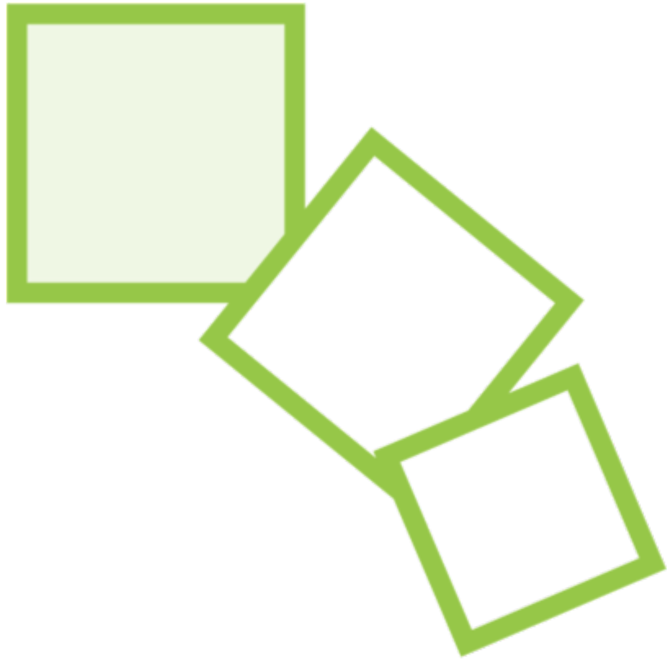
Test uses Monte Carlo to simulate

- P(Click-through/Version A)
- P(Click-through/Version B)

Each scenario estimates both probabilities

Obtains large number, say 10^5, scenarios

Interpreting Bayesian A/B Test



Now easy to find P(A > B) and P(B > A)

P(A > B) = % of scenarios where A > B

P(B > A) = % of scenarios where B > A

Interpreting Bayesian A/B Test

Calculate % difference between A and B for each scenario

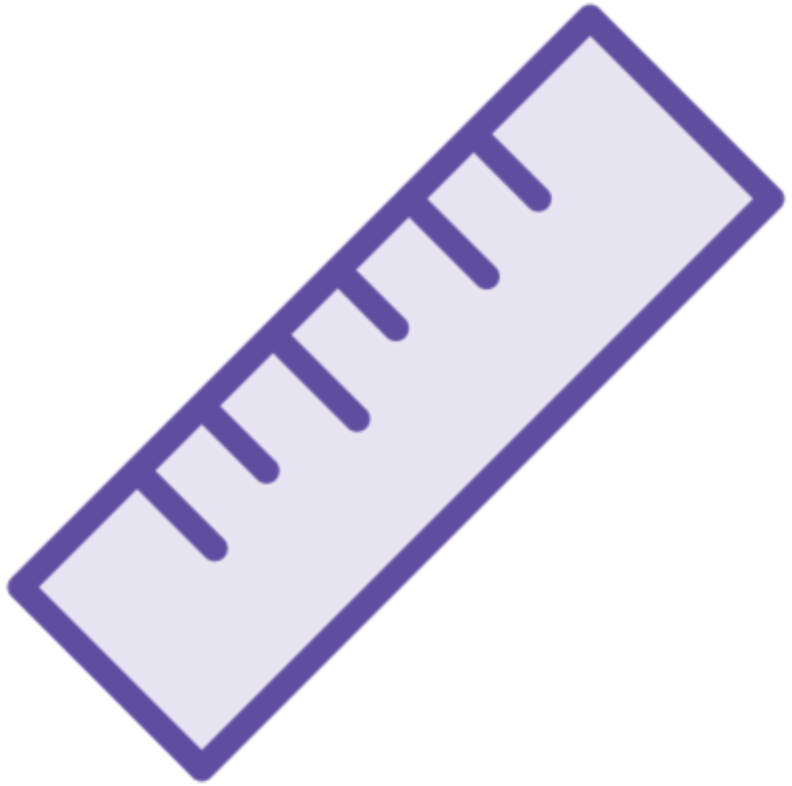
(A - B)/B

Find 90% credible interval for this % difference

“We can be 90% confident that (A - B)/B lies between _____ and _____”

Confidence Intervals: Frequentist Inference ::
Credible Intervals: Bayesian Inference

Credible and Confidence Intervals



In Frequentist inference, parameter has a true, fixed value

90% Confidence Interval: 90% confidence that true, fixed value lies in that range

In Bayesian inference, parameter is a random variable

90% Credible Interval: 90% posterior probability that parameter lies in that range

Demo

Implementing Bayesian A/B testing

Demo

BayesianABTesting

Summary

Understanding A/B testing

Choosing the right test based on target variable distributions

Understanding Bayes' theorem

Understanding Bayesian A/B testing

Representing priors as distributions

Implementing Bayesian A/B testing

Related Courses

Solving Problems with Numerical Methods

Interpreting Data Using Descriptive Statistics with Python