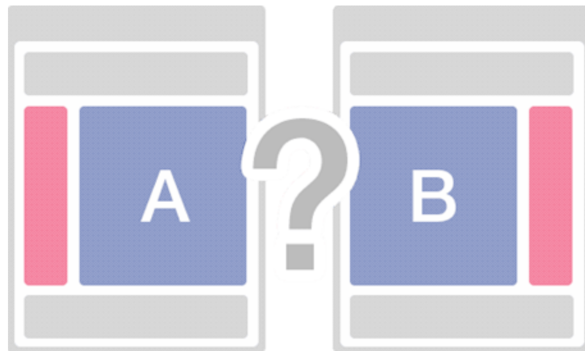


Bayesian Approach to A/B Testing

Varun Panicker

Varun Panicker
April, 2018

Flow

- Frequentist vs Bayesian Statistics
- Traditional A/B Testing
- Bayesian A/B Testing

PyData MUMBAI

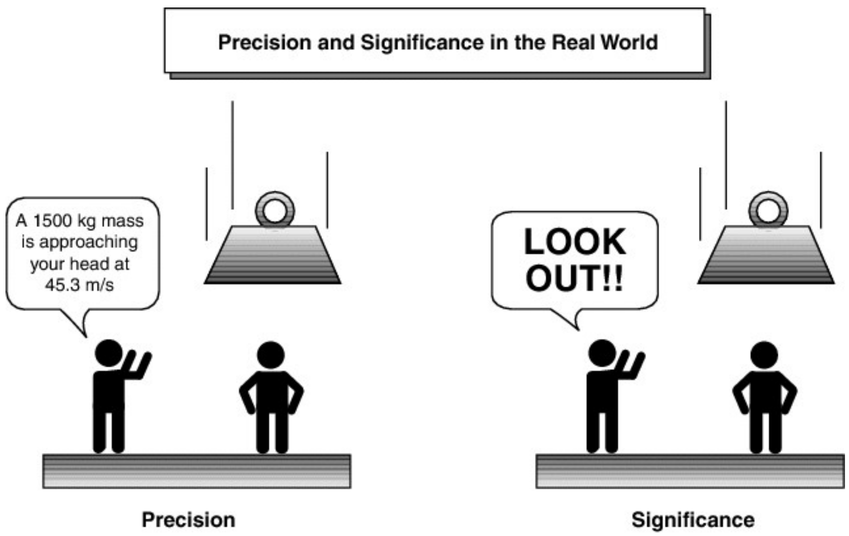
Varun Panicker
April, 2018

Frequentist vs Bayesian Approach

Problem: I have lost my phone but I know it is somewhere in my house

- Frequentist Reasoning

I can hear the phone beeping. I also have a mental model which helps me identify the area from which the sound is coming. Therefore, upon hearing the beep, I infer the area of my home I must search to locate the phone.



PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian Approach

- Bayesian Reasoning

I can hear the phone beeping. Now, apart from a mental model which helps me identify the area from which the sound is coming from, **I also know the locations where I have misplaced the phone in the past**. So, I combine my inferences using the beeps and my prior information about the locations I have misplaced the phone in the past to identify an area I must search to locate the phone.

PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

- Let's say John has been diagnosed with a rare disease that affects 0.1% of the population
- The test John underwent for diagnosis claims to have an accuracy of 99%

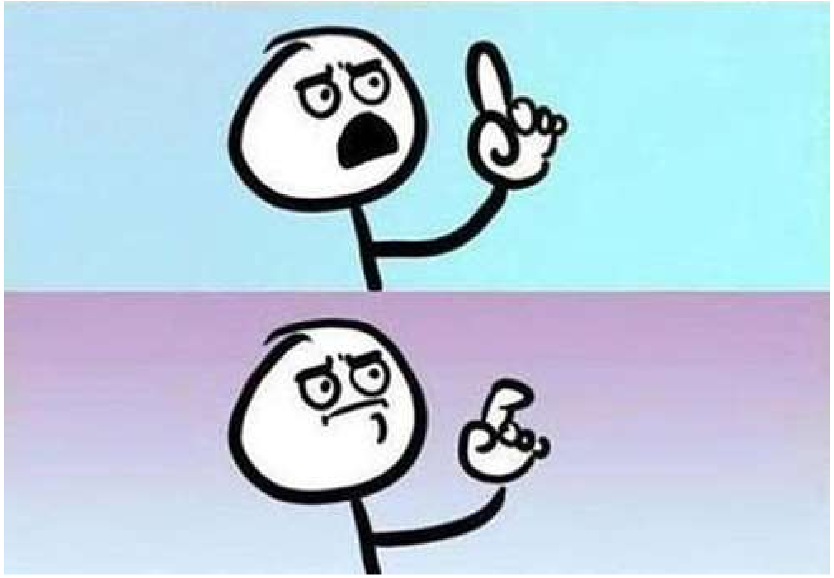
What are the chances that John has the disease?

PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

Did you guess 99%? Well, think Bayes!



Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

John actually has the disease, given

$$P(H \mid E) = \frac{P\,(E \mid H)\ \times P\,(H)}{P\,(H)\ \times P\,(E \mid H) + P\,(-H)\ \times P\,(E \mid -H)}$$

John tested positive for it

PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

John actually has the disease, given

PRIOR probability of having the disease

$$P(H \mid E) = \frac{P\,(E \mid H)\ \times P\,(H)}{P\,(H)\ \times P\,(E \mid H) + P\,(-H)\ \times P\,(E \mid\ -H)}$$

John tested positive for it

PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

John actually has the disease, given

P of event given the hypothesis is true .i.e. P that John tests positive if he has the disease

PRIOR probability of having the disease

$$P(H \mid E) = \frac{P(E \mid H) \times P(H)}{P(H) \times P(E \mid H) + P(-H) \times P(E \mid -H)}$$

John tested positive for it

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

John actually has the disease, given

P of event given the hypothesis is true .i.e. P that John tests positive if he has the disease

PRIOR probability of having the disease

$$P(H \mid E) = \frac{P\,(E \mid H)\ \times P\,(H)}{\underbrace{P\,(H)\ \times P\,(E \mid H)}_{\text{P that John has the disease and has been correctly diagnosed}} + P\,(-H)\ \times P\,(E \mid -H)}$$

John tested positive for it

P that John has the disease and has been correctly diagnosed

PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

John actually has the disease, given

P of event given the hypothesis is true .i.e.
P that John tests positive if he has the disease

PRIOR probability of having the disease

$$P(H \mid E) = \frac{P(E \mid H) \times P(H)}{\underbrace{P(H) \times P(E \mid H)}_{\text{P that John has the disease and has been correctly diagnosed}} + \underbrace{P(-H) \times P(E \mid -H)}_{\text{P that John does not have the disease and has been incorrectly diagnosed}}}$$

John tested positive for it

P that John has the disease and has been correctly diagnosed

P that John **does not** have the disease and has been incorrectly diagnosed

Varun Panicker
April, 2018

PyData MUMBAI

Frequentist vs Bayesian – An Example

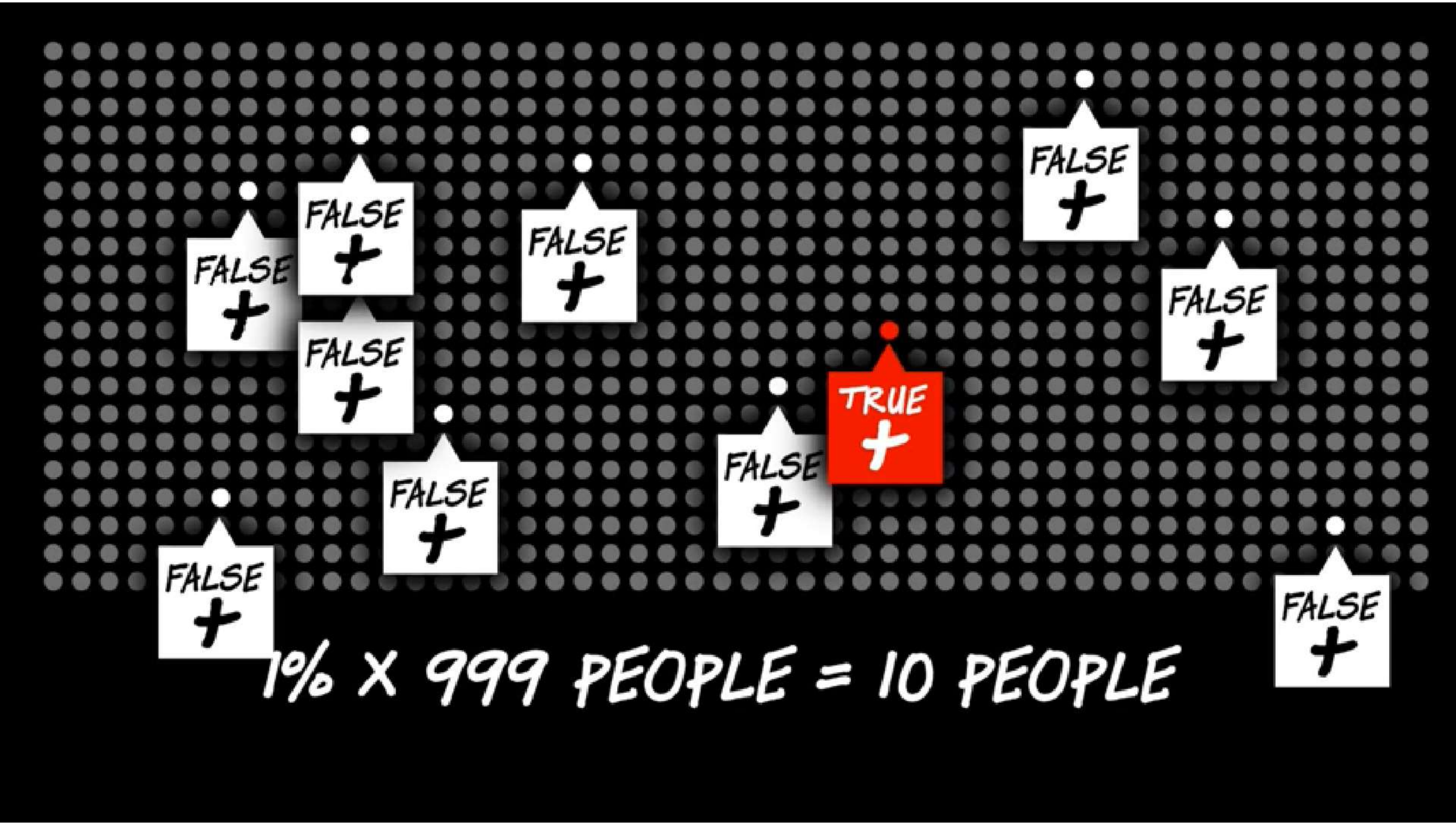
$$P(H \mid E) = \frac{P\,(E \mid H)\ \times P\,(H)}{P\,(H)\ \times P\,(E \mid H) + P\,(-H)\ \times P\,(E \mid -H)}$$

$$P(H \mid E) = \frac{0.99 \times 0.001}{0.001\ \times 0.99 +\ 0.999\ \times 0.01}$$

$$P(H \mid E) = 9\%$$

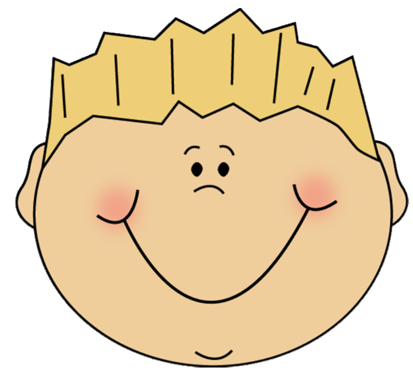
PyData MUMBAI

Varun Panicker
April, 2018


FALSE +
FALSE +
FALSE +
FALSE +
FALSE +
FALSE +
FALSE +
TRUE +
FALSE +
FALSE +
FALSE +
1% X 999 PEOPLE = 10 PEOPLE

Frequentist vs Bayesian – An Example

- Hence, we can say that John is a part of a group of 11 people out of which only 1 person has the disease.
- Therefore, the chances of him actually having it are $\frac{1}{11} = 9\%$
- John now decides to do a second test to conclude!



PyData MUMBAI

Varun Panicker
April, 2018

Frequentist vs Bayesian – An Example

Posterior probability
(9 %)

$$P(H \mid E) = \frac{P\,(E \mid H)\ \times P\,(H)}{P\,(H)\ \times P\,(E \mid H) + P\,(-H)\ \times P\,(E \mid -H)}$$

$$P(H \mid E) = \frac{0.99 \times 0.09}{0.09\ \times 0.99 +\ 0.91\ \times 0.01}$$

$$P(H \mid E) \sim = 91\%$$

PyData
MUMBAI

Varun Panicker
April, 2018

Thomas Bayes

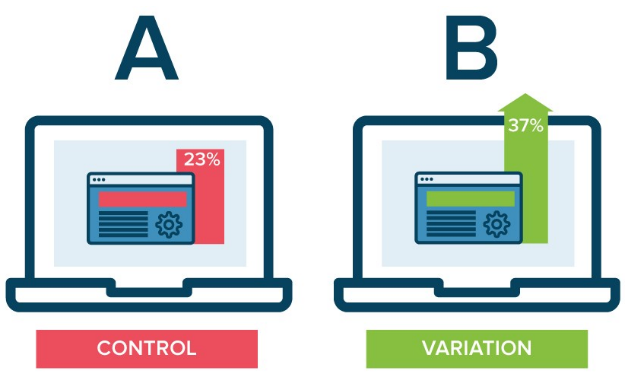
- An Essay towards solving a Problem in the Doctrine of Chances, 1763

http://www.stat.ucla.edu/history/essay.pdf

Varun Panicker
April, 2018

Traditional A/B testing

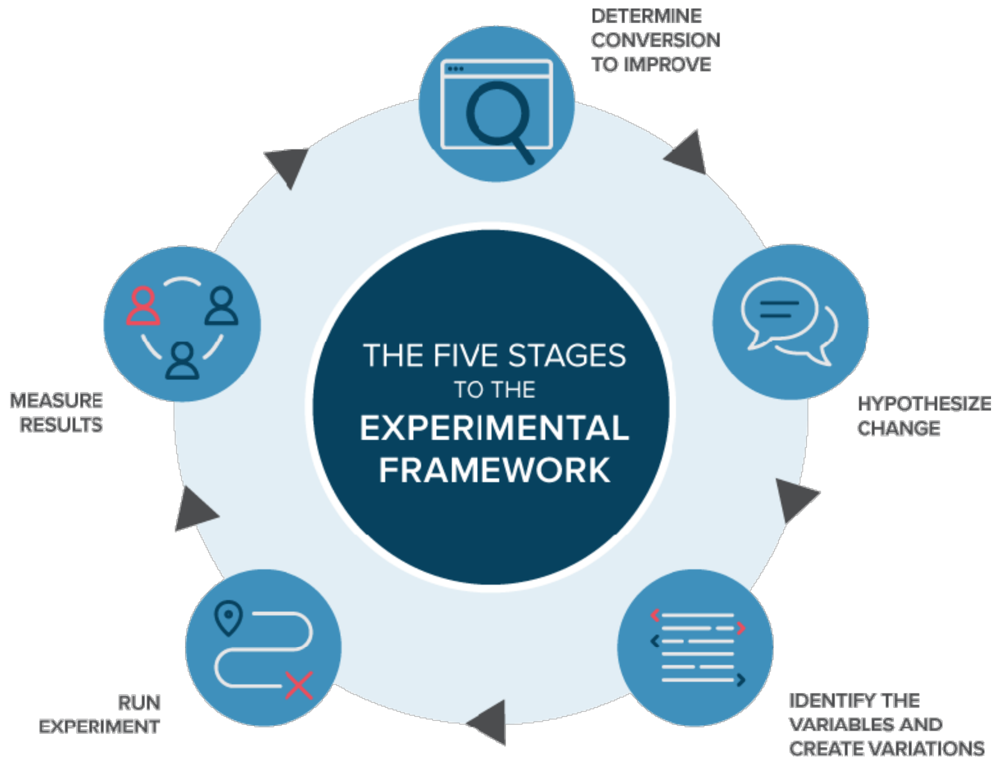
- A typical A/B test compares the values of a parameter across two variations (control and treatment)
- Uses standard frequentist parameter testing measures, i.e., p-values and confidence intervals.



PyData MUMBAI

Varun Panicker
April, 2018

Traditional A/B testing



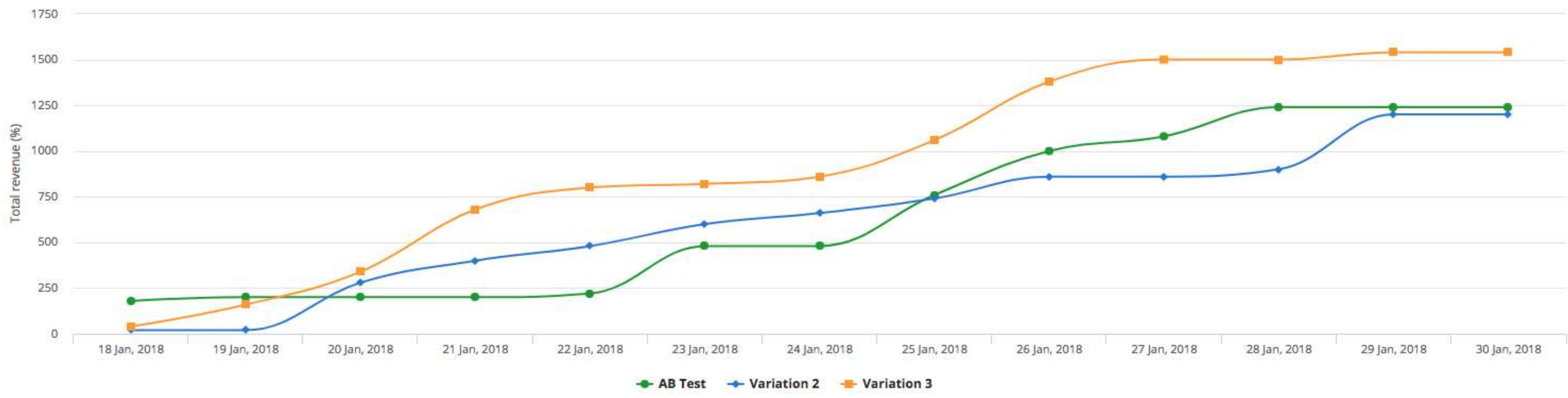
PyData MUMBAI

Varun Panicker
April, 2018

Revenue Increase Test

Testing how the Pareto principle influences our revenue based on small but effective changes.

Graph interval: Daily



Total revenue over time

$ 3980

Variation Name	Traffic	Visitors	Unique Visitors	Sales	Revenue	Revenue per Visitor	Conversion Rate	Improvement	Chance to beat Original
AB Test	34%	196	196	62	1240.00	6.33	31.63%	[This is the Control]	
Variation 2	33%	150	150	60	1200.00	8	40%	26.46%	94.63%
Variation 3	33%	141	141	77	1540.00	10.92	54.61%	72.65%	100.00%

Automatic winner settings disabled CHANGE

« BACK TO PAGE SETTINGS

STOP TEST AND CHOOSE WINNER

Traditional A/B testing

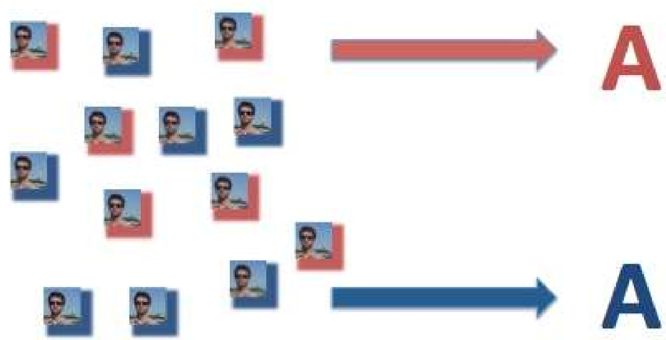
- Run Duration and “Peeking”

The only point at which you should evaluate significance is the endpoint that you predetermined for your test

Varun Panicker
April, 2018

A/A Tests

- A/A testing uses two identical versions of a page against each other.
- It is done to check that the tool being used to run the experiment is statistically fair
- In an A/A test, the tool should report no difference in conversions between the control and variation, if the test is implemented correctly.



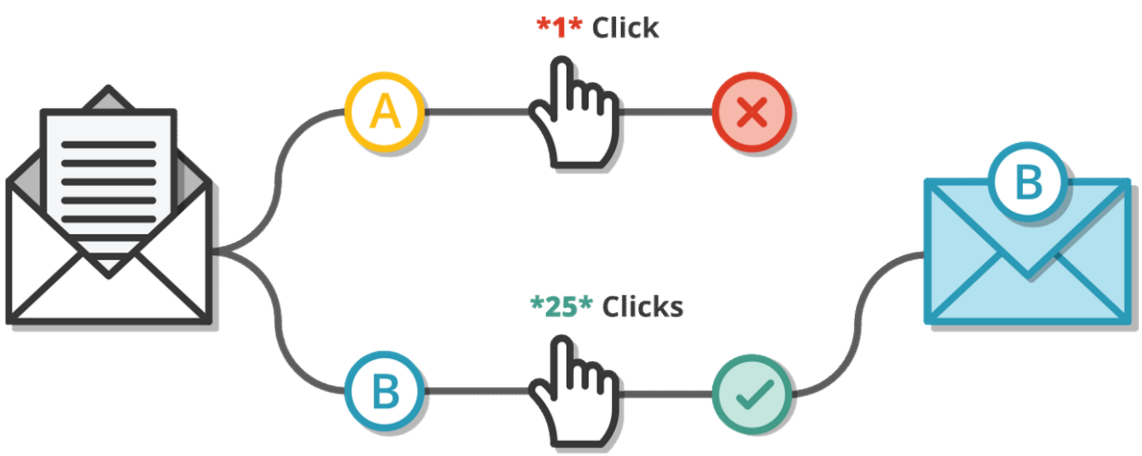
Traditional A/B testing

Drawbacks-

- Does not consider priors (Eg: Prior knowledge of Variant A)
- If a change is not significant enough, we fall back and continue with our control
- Not conclusive due to vagueness around significance

Bayesian A/B Testing

- Send an email newsletter to a group of 300 recipients
  - Variant A- Contains a big picture/logo like it always does
  - Variant B- Does not contain the picture



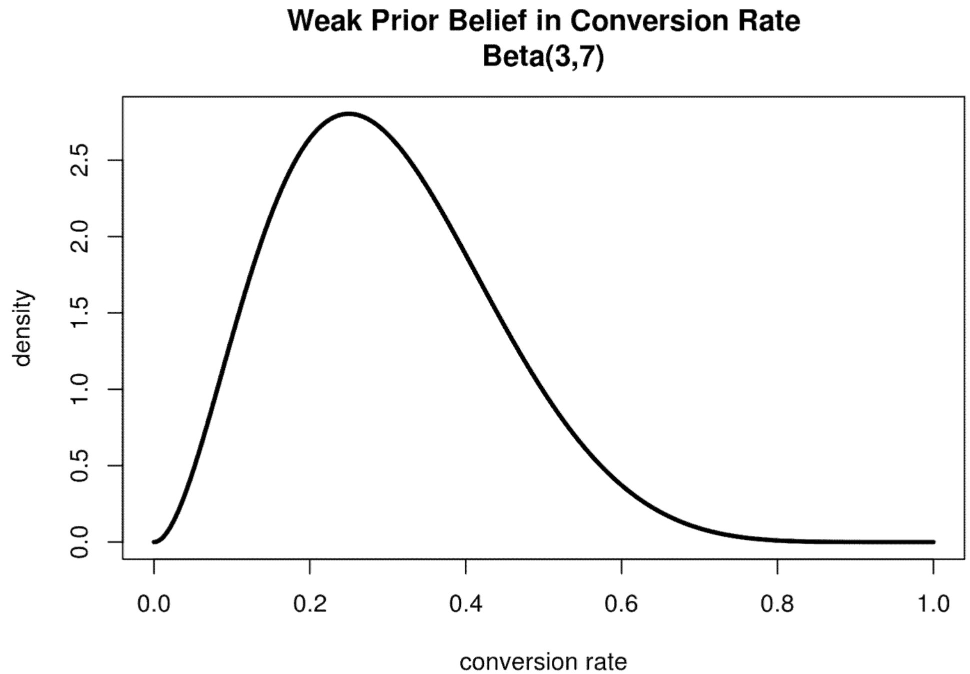
Varun Panicker
April, 2018

Bayesian A/B Testing

- Establish prior probability
- We've run email campaigns earlier so we expect that the probability of the recipient clicking the link to blog on any given email should be around 30%
- For simplicity, we'll use the same prior for A and B
- $Beta(\alpha,\beta)$
  - α = *Number of Times Success Observed*
  - β = *Number of Times Failure Observed*
- Therefore, **Beta(3,7)**

Bayesian A/B Testing

- PDF for our Beta (3,7)



Bayesian A/B Testing

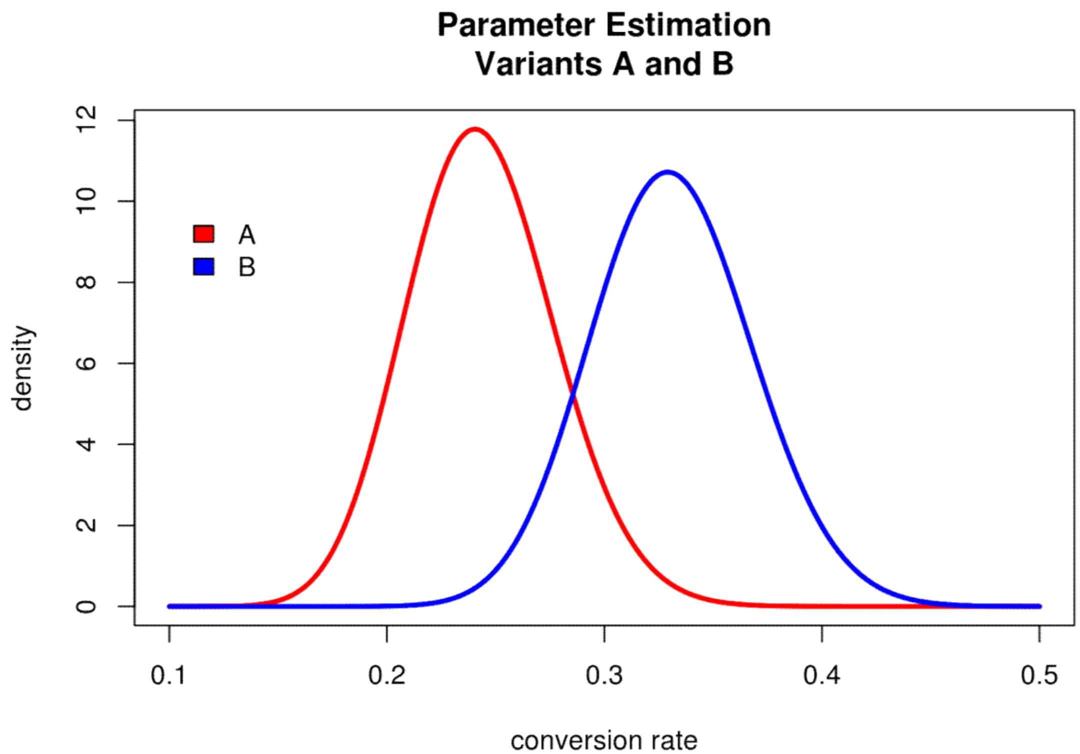
- Moment of truth!

	Clicked	Not Clicked	Observed Conversion Rate
Variant A	36	114	0.24
Variant B	50	100	0.33

- Variant A is going to be represented by Beta(36+3,114+7) and Variant B by Beta(50+3,100+7)

Bayesian A/B Testing

- Estimated distributions for Variants A and B



Varun Panicker
April, 2018

Bayesian A/B Testing

- Monte Carlo Simulation

```
n.trials <- 100000
prior.alpha <- 3
prior.beta <- 7
a.samples <- rbeta(n.trials,36+prior.alpha,114+prior.beta)
b.samples <- rbeta(n.trials,50+prior.alpha,100+prior.beta)
p.b_superior <- sum(b.samples > a.samples)/n.trials
```

Output: *p.b_superior* = 0.96

Bayesian A/B Testing

- **Magnitude is more important than Significance**
- Frequentist statistics tells us Significance, but what we're really after is Magnitude!
- This can be confirmed by Bayesian A/B Testing

Varun Panicker
April, 2018

Sources

- Vwo.com
- stats.stackexchange.com/users/100906/flimzy
- *Veritasium*
- pngtree.com
- *Peeking at A/B Tests*- Ramesh Johari, Pete Koomen, Leonid Pekelis, David Walsh
- conversionsciences.com/blog/ab-testing-statistics
- thrivethemes.com/optimize
- www.optimizely.com/optimization-glossary
- drjasondavis.com/blog
- countbayesie.com/blog

PyData MUMBAI

Thank You!

Varun Panicker
April, 2018