A/B HEADLINE TESTING AT OUTBRAIN: VISUAL REVENUE

Rob Doherty / @robdoherty2

AGENDA

AGENDA

- A/B Headline Testing: Problem Statement

AGENDA

- A/B Headline Testing: Problem Statement
- Frequentist approach

AGENDA

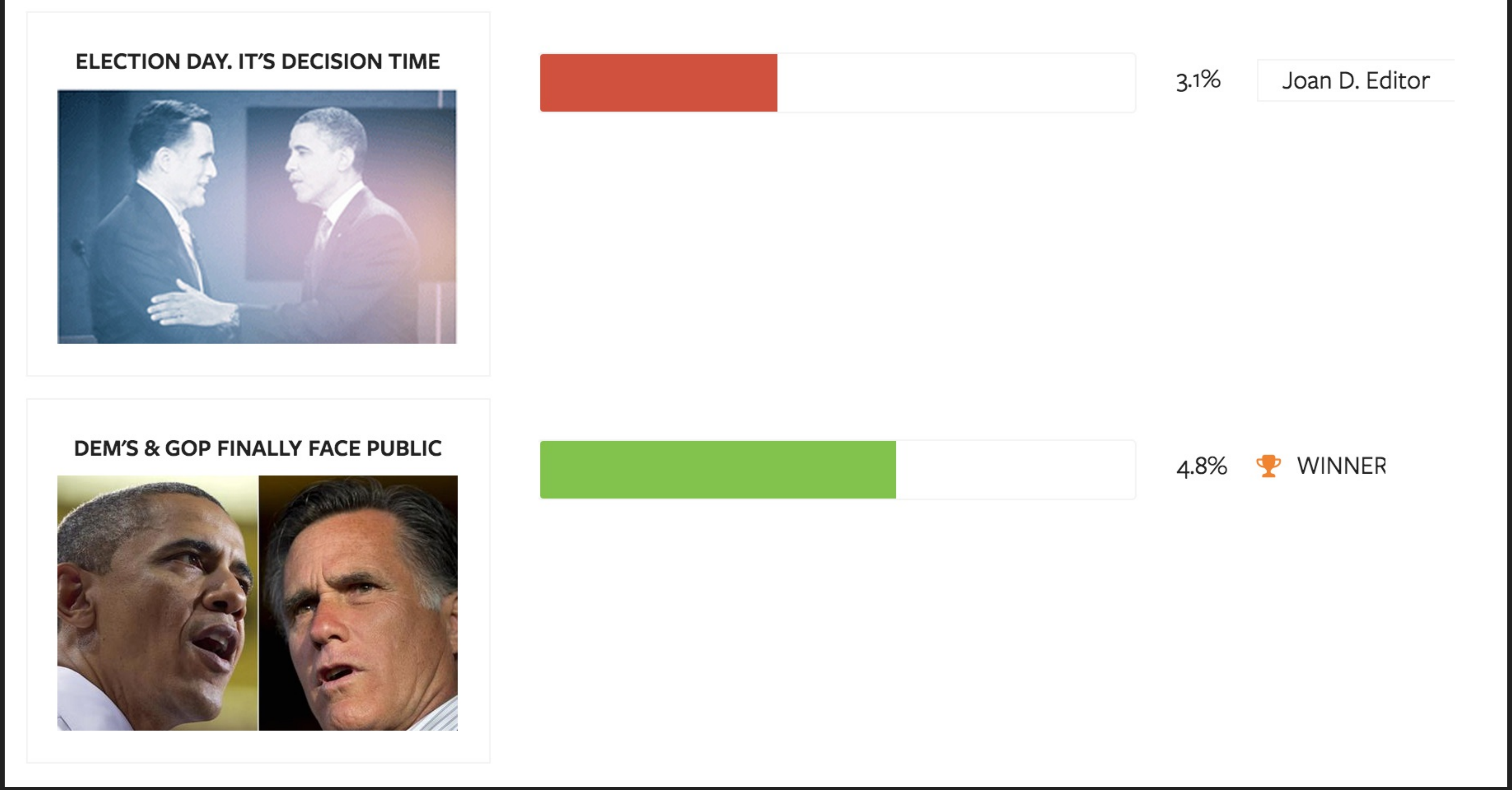
- A/B Headline Testing: Problem Statement
- Frequentist approach
- Bayesian approach

AGENDA

- A/B Headline Testing: Problem Statement
- Frequentist approach
- Bayesian approach
- Certainty of conclusion

PROBLEM STATEMENT

WHAT IS A/B HEADLINE TESTING?



WHAT IS A/B HEADLINE TESTING?

Headline A: "What Harbaugh regrets about Super Bowl"
(2.46% CTR)

Headline B: "John Harbaugh explains Super Bowl tirade"
(4.93% CTR)

CONSTRAINTS

CONSTRAINTS

- Headlines may be on the front page for a short time

CONSTRAINTS

- Headlines may be on the front page for a short time
- The number of readers varies greatly per front page

CONSTRAINTS

- Headlines may be on the front page for a short time
- The number of readers varies greatly per front page
- The CTR of a headline depends on frontpage position

CONSTRAINTS

- Headlines may be on the front page for a short time
- The number of readers varies greatly per front page
- The CTR of a headline depends on frontpage position
- Front pages are dynamic, so headlines can change position

FREQUENTIST APPROACH

FREQUENTIST APPROACH

- The parameter-of-interest is CTR
- Null hypothesis testing: assume that there are true-but-unknown CTRs for A and B

FREQUENTIST APPROACH

- The parameter-of-interest is CTR
- Null hypothesis testing: assume that there are true-but-unknown CTRs for A and B
  - the goal is to figure out if they are different or not

FREQUENTIST APPROACH

- Probability that an effect is not due to just chance alone
- Must define in advance the probability of a sampling error
- Sample size is an important component

FREQUENTIST APPROACH

- Probability that an effect is not due to just chance alone
- Must define in advance the probability of a sampling error
- Sample size is an important component

AN EXAMPLE

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
After 200 observations	Insignificant	Insignificant	Significant	Significant
After 500 observations	Insignificant	Significant	Insignificant	Significant
End of experiment	Insignificant	Significant	Insignificant	Significant

AN EXAMPLE

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
After 200 observations	Insignificant	Insignificant	Significant	Significant
After 500 observations	Insignificant	Significant	Insignificant	Significant
End of experiment	Insignificant	Significant	Insignificant	Significant

AN EXAMPLE

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
After 200 observations	Insignificant	Insignificant	Significant	Significant
After 500 observations	Insignificant	Significant	Insignificant	Significant
End of experiment	Insignificant	Significant	Insignificant	Significant

BAYESIAN APPROACH

BAYESIAN APPROACH

- Frequentist: probability measures a proportion of outcomes
- Bayesian: probability measures a degree of belief

BAYESIAN APPROACH

- $\mathrm{CTR_A}$ and $\mathrm{CTR_B}$ are no longer fixed numbers, but probability distributions
- Now the question becomes: What is the probability that $\mathrm{CTR_A}$ is larger than $\mathrm{CTR_B}$ given the data from the experiment?

$$P(\mathrm{CTR_A} > \mathrm{CTR_B} | \mathrm{data})$$

BAYES THEOREM

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

BAYES THEOREM

- Prior - a distribution that encodes your prior belief about the parameter-of-interest
- Likelihood - a function that encodes how likely your data is given a range of possible parameters
- Posterior - a distribution of the parameter-of-interest given your data, combining the prior and likelihood

BAYES THEOREM: CONJUGATE PRIOR

- Prior - a distribution that encodes your prior belief about the parameter-of-interest
- Conjugate Prior - for certain choices of the prior, the *posterior* has the same algebraic form as the prior (with updated parameters)

PROPORTIONAL BAYES THEOREM

$$P(A|B) \sim P(B|A)P(A)$$

the posterior is proportional to the liklihood times the prior

BETA DISTRIBUTION

We model $\mathrm{CTR_A}$ and $\mathrm{CTR_B}$ with the *Beta Distribution*

$$\mathrm{CTR} \sim \mathrm{Beta}(\alpha, \beta)$$

BETA DISTRIBUTION

We model CTR_A and CTR_B with the *Beta Distribution*

$$\mathrm{CTR} \sim \mathrm{Beta}(\alpha, \beta)$$

- Represents a distribution of probabilities

BETA DISTRIBUTION

We model CTR_A and CTR_B with the *Beta Distribution*

$$\mathrm{CTR} \sim \mathrm{Beta}(\alpha, \beta)$$

- Represents a distribution of probabilities
  - it represents all possible values of a probability when we don't know what that probability is

BETA DISTRIBUTION

We model CTR_A and CTR_B with the *Beta Distribution*

$$\mathrm{CTR} \sim \mathrm{Beta}(\alpha, \beta)$$

- Represents a distribution of probabilities
  - it represents all possible values of a probability when we don't know what that probability is
- The domain is $(0, 1)$, just like a probability

BETA DISTRIBUTION

We model CTR_A and CTR_B with the *Beta Distribution*

$$\mathrm{CTR} \sim \mathrm{Beta}(\alpha, \beta)$$

- Represents a distribution of probabilities
  - it represents all possible values of a probability when we don't know what that probability is
- The domain is $(0, 1)$, just like a probability
- Parameters:

$$\mathrm{Beta}(\alpha + \mathrm{hits}, \beta + \mathrm{misses})$$

BETA DISTRIBUTION

$$\mathrm{Beta}(\alpha, \beta)$$

$$\alpha = \text{pageviews} * \text{CTR}$$

$$\beta = \text{pageviews} * (1 - \text{CTR})$$

BETA DISTRIBUTION

Assumptions of prior

BETA DISTRIBUTION

Assumptions of prior

- Before we observe any clicks we assume all headlines are equally likely to be clicked on

BETA DISTRIBUTION

Assumptions of prior

- Before we observe any clicks we assume all headlines are equally likely to be clicked on
- That is, we start with a ***uniform prior***

BETA DISTRIBUTION

Assumptions of prior

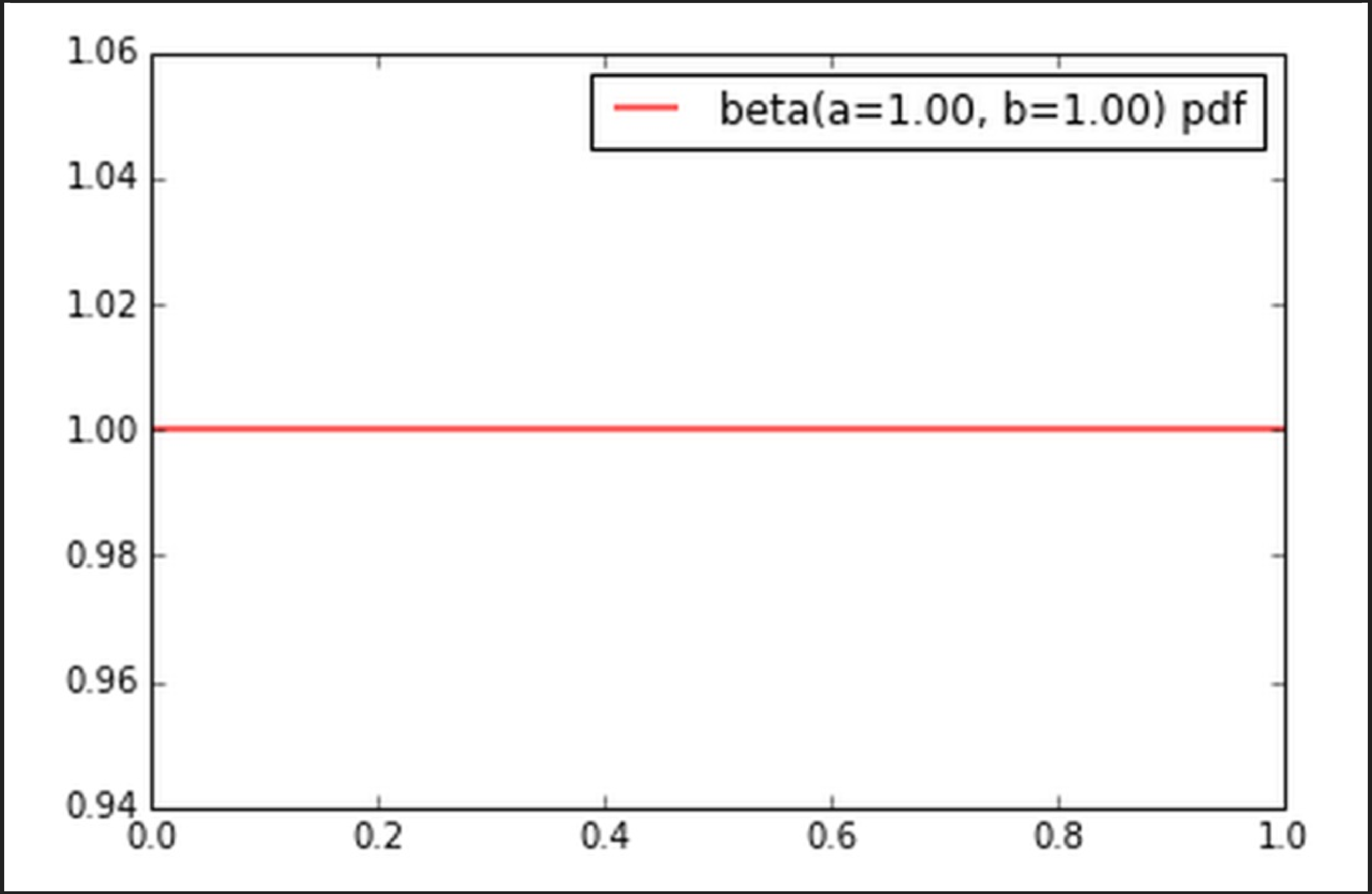
- Before we observe any clicks we assume all headlines are equally likely to be clicked on
- That is, we start with a ***uniform prior***

BETA DISTRIBUTION

Assumptions of prior

- Before we observe any clicks we assume all headlines are equally likely to be clicked on
- That is, we start with a ***uniform prior***
- A uniform distribution is the same as Beta(1,1)

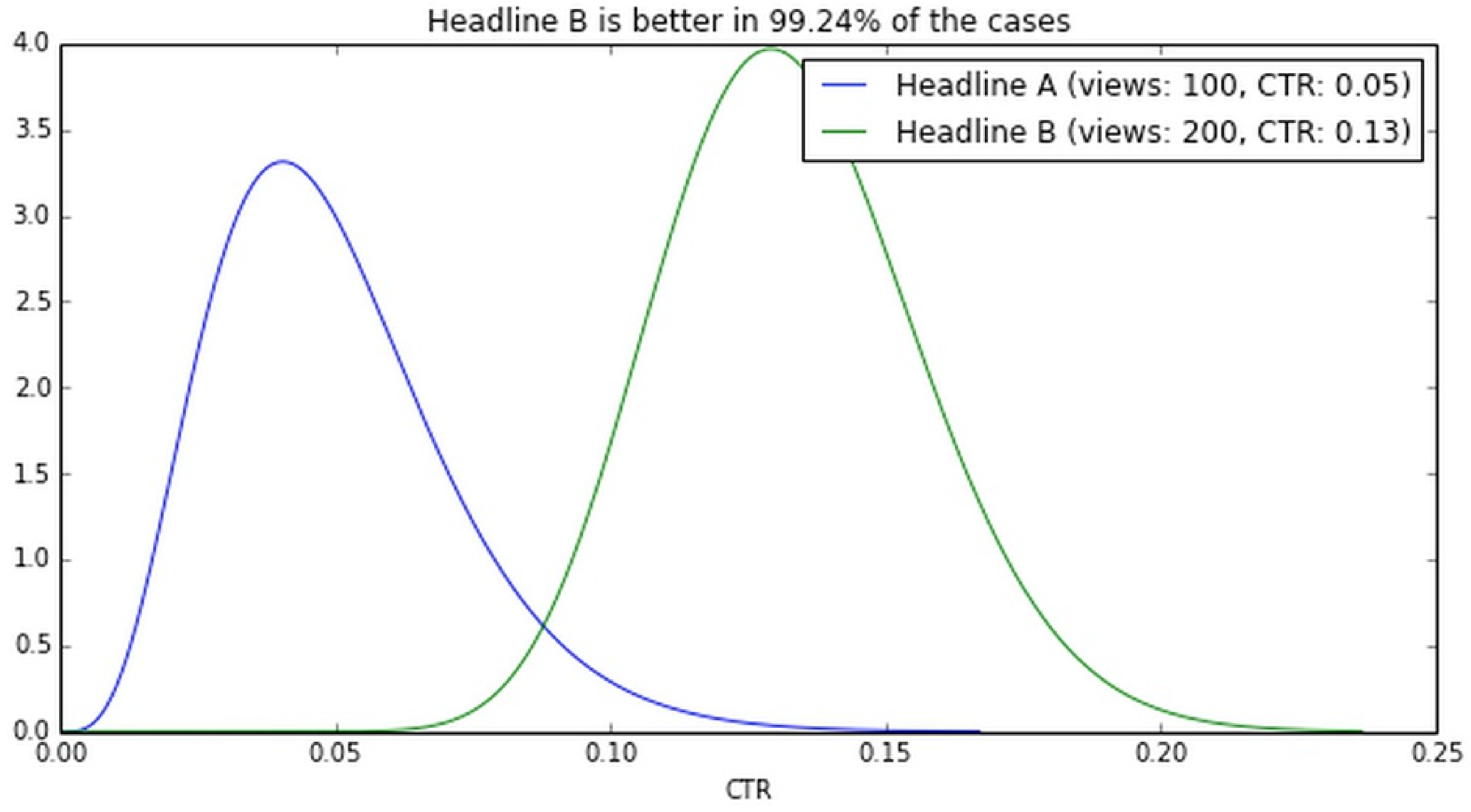
UNIFORM PRIOR



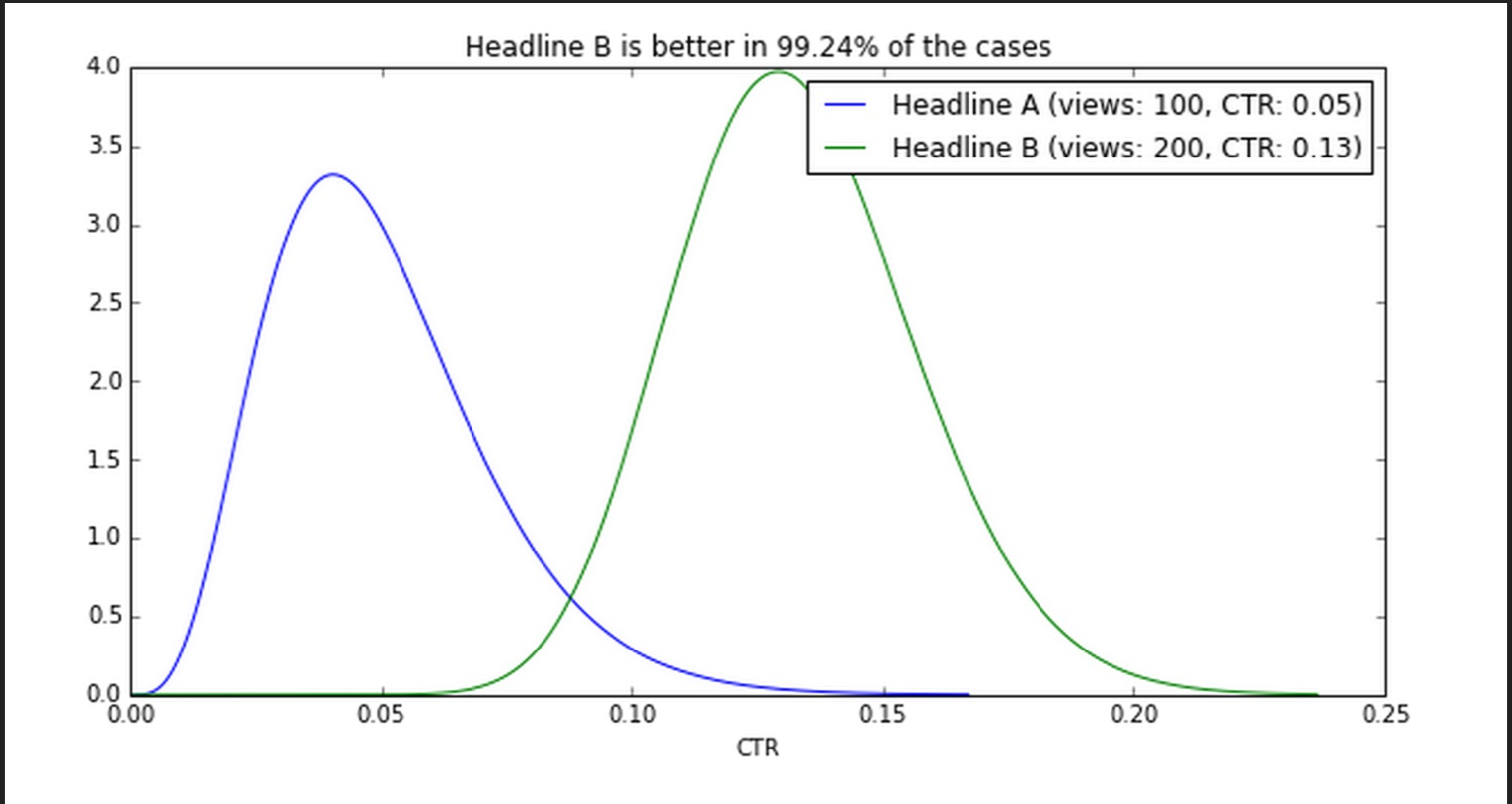
PICKING A WINNER: RANDOM SAMPLING

```
from scipy.stats import beta
def percent_better(a_views, b_views, a_ctr, b_ctr, size):
    ra = beta.rvs(a_views*a_ctr, a_views*(1-a_ctr), size=(size))
    rb = beta.rvs(b_views*b_ctr, b_views*(1-b_ctr), size=(size))
    return sum(ra >= rb) / (1.0*size)
```

```
fig = figure(figsize=(10,5))
demonstrate(100,200, 0.04969, 0.13287, size=1000000)
```



PICKING A WINNER



TEST STOPPING CRITERIA

Anscombe Boundary

TEST STOPPING CRITERIA

Anscombe Boundary

TEST STOPPING CRITERIA

Anscombe Boundary

$$|c_a - c_b| > \Phi^{-1}\left(\frac{n}{k + 2n}\right)\sqrt{n}$$

TEST STOPPING CRITERIA

Anscombe Boundary

$$|c_a - c_b| > \Phi^{-1}\left(\frac{n}{k + 2n}\right)\sqrt{n}$$

$|c_a - c_b|$ - absolute difference between clicks for both headlines

TEST STOPPING CRITERIA

Anscombe Boundary

$$|c_a - c_b| > \Phi^{-1}\left(\frac{n}{k + 2n}\right)\sqrt{n}$$

$|c_a - c_b|$ - absolute difference between clicks for both headlines

Φ^{-1} - the quantile function of the standard normal (or ppf, or inverse cdf)

TEST STOPPING CRITERIA

Anscombe Boundary

$$|c_a - c_b| > \Phi^{-1}\left(\frac{n}{k + 2n}\right)\sqrt{n}$$

$|c_a - c_b|$ - absolute difference between clicks for both headlines

Φ^{-1} - the quantile function of the standard normal (or ppf, or inverse cdf)

n - the number of page views so far

TEST STOPPING CRITERIA

Anscombe Boundary

$$|c_a - c_b| > \Phi^{-1}\left(\frac{n}{k+2n}\right)\sqrt{n}$$

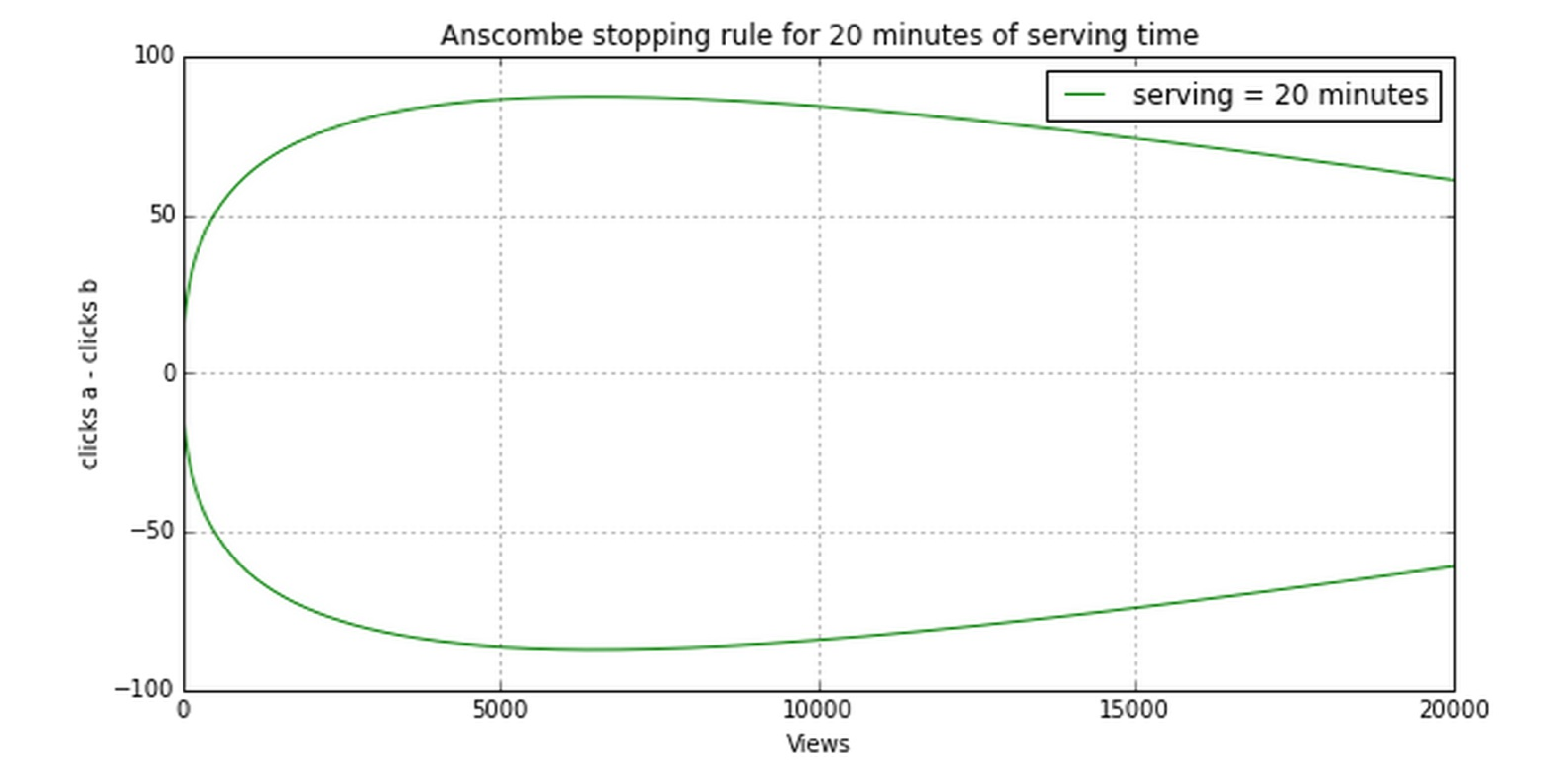
$|c_a - c_b|$ - absolute difference between clicks for both headlines

Φ^{-1} - the quantile function of the standard normal (or ppf, or inverse cdf)

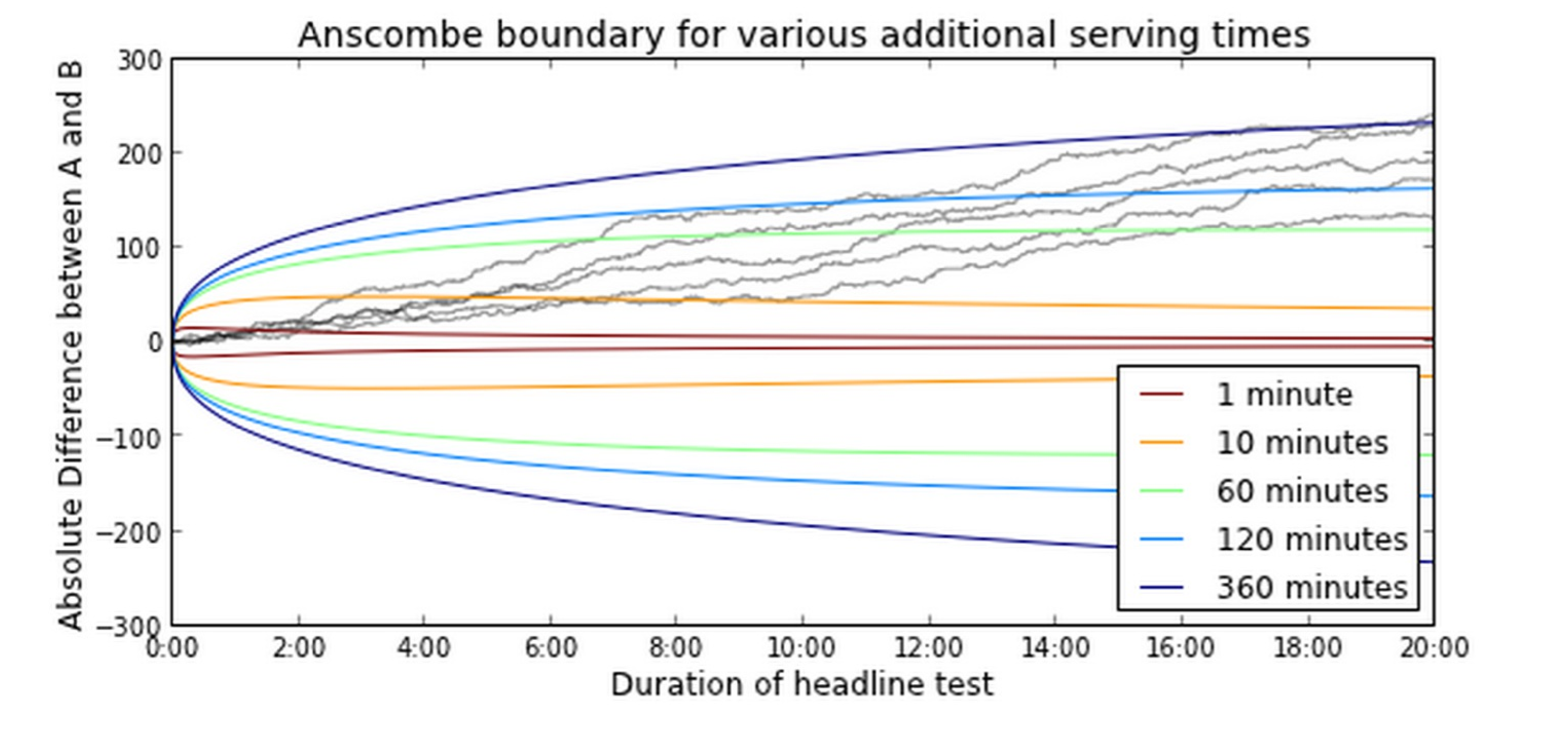
n - the number of page views so far

k - the number of future readers who will be exposed to test, given a maximum time

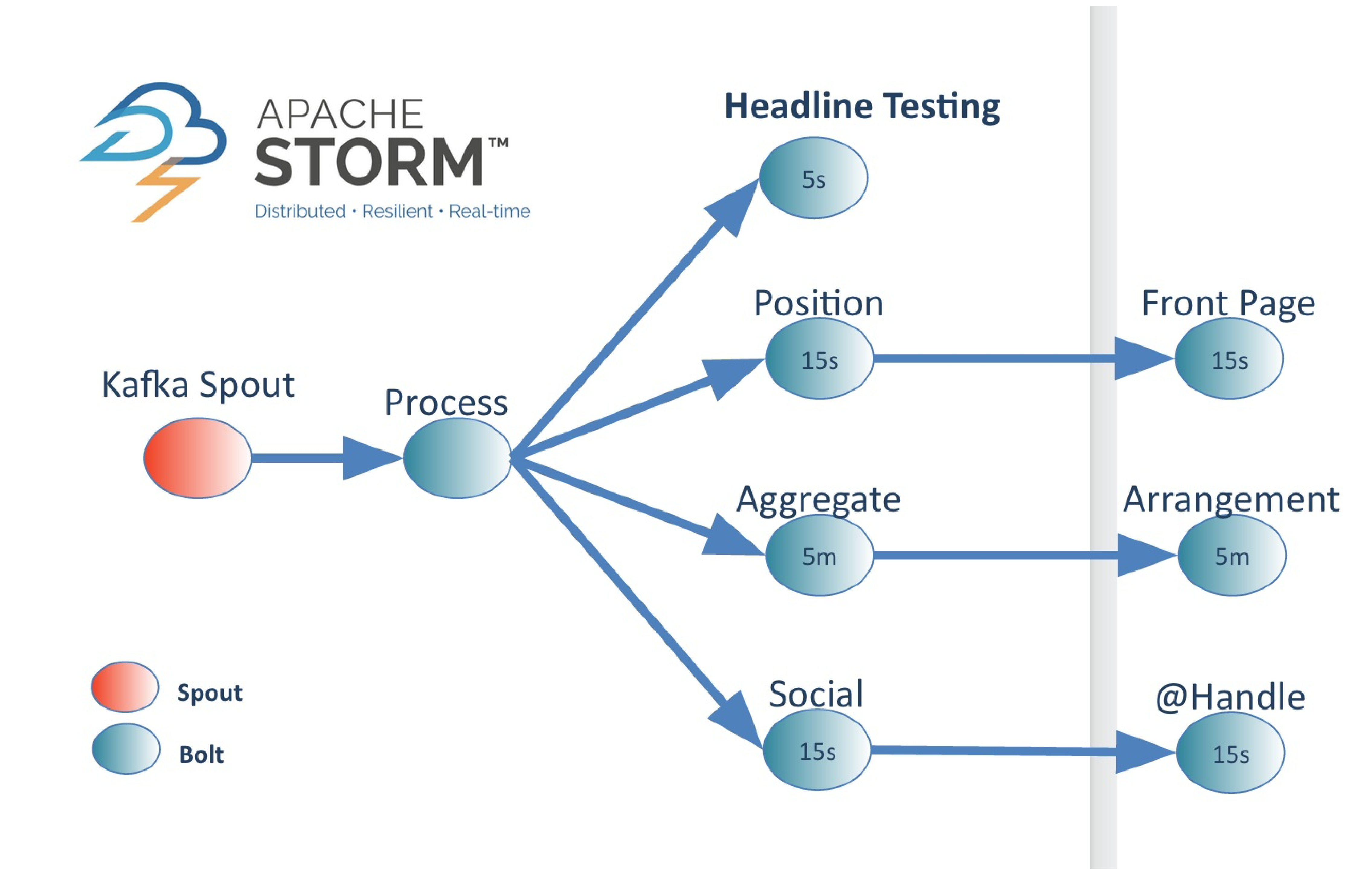
TEST STOPPING CRITERIA



TEST STOPPING CRITERIA



IMPLEMENTATION



FURTHER READING

- Bayesian headline testing at Visual Revenue, Jeroen Janssens
- A/B Testing with Hierarchical Models in PythonManojit Nandi
- Bayesian A/B TestsSergey Feldman
- Bayesian AB TestingMaciej Kula