GENERAL ASSEMBLY

INTRO TO DATA SCIENCE
BAYESIAN A/B TESTING

AGENDA

I. Review Bayes Theorem:

Monty Hall Problem

II. A/B Headline Testing: Problem Statement

FREQUENTIST APPROACH

BAYESIAN APPROACH

CERTAINTY OF CONCLUSION

I. Review Bayes Theorem

Bayes Theorem: Monty Hall Problem

The Monty Hall Problem:

Bayes Theorem: Monty Hall Problem

The Monty Hall Problem:

- *Suppose you're on the game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats.*
- *You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.*
- *He then says to you, "Do you want to pick door No. 2?"*

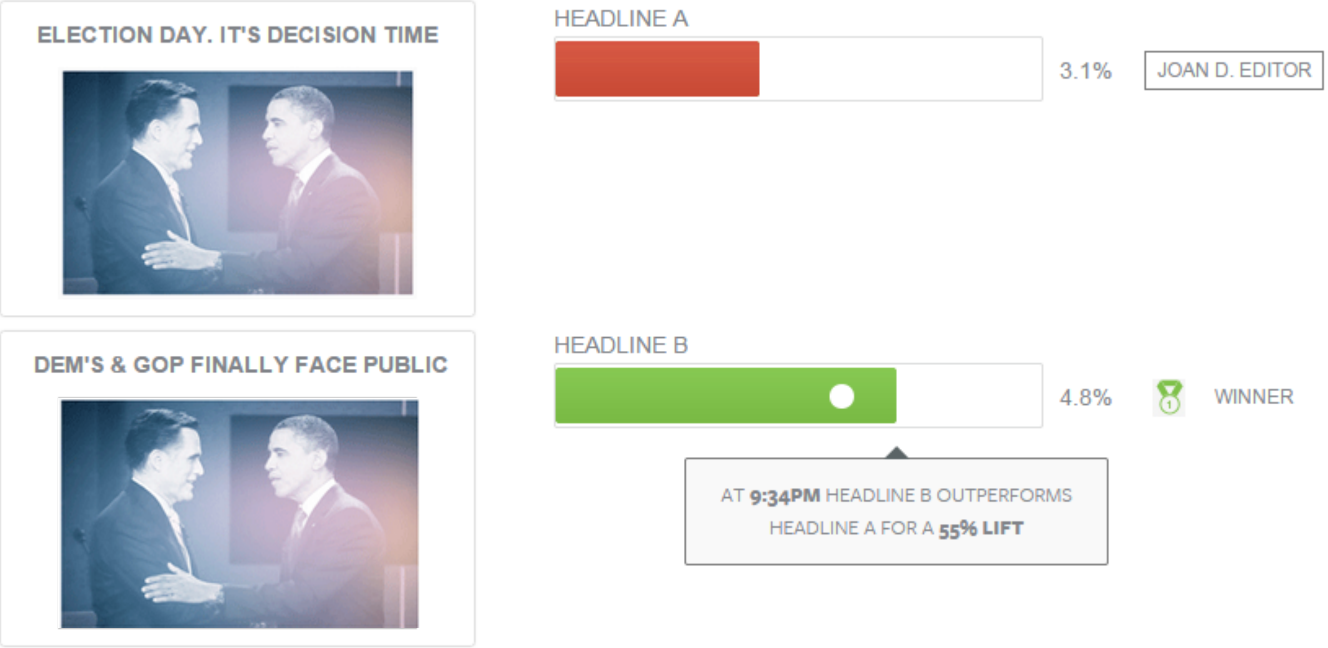
Is it to your advantage to switch your choice?

Bayes Theorem

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

I. PROBLEM STATEMENT

What is A/B Headline Testing?



What is A/B Headline Testing?

Headline A: "What Harbaugh regrets about Super Bowl" (3.06% CTR)

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

Constraints of Headline Testing

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

Constraints of Headline Testing

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

Constraints of Headline Testing

- 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 OF HEADLINE TESTING

- 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

II. 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*
  - the goal is to figure out if they are different or not

Statistical significance

- 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

STATISTICAL SIGNIFICANCE

- 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	*test stopped*	*test stopped*
End of experiment	Insignificant	**Significant**	**Significant!**	**Significant**

III. Bayesian approach

Frequentist vs. Bayesian

Recall:

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

BAYESIAN APPROACH

- CTR_A and CTR_B are no longer fixed numbers, but probability distributions.
- We model CTR_A and CTR_B with the ***Beta Distribution***
- Now the question becomes:
  What is the probability that CTR_A is larger than CTR_B given the data from the experiment?

$$P(\mathbf{CTR}_{A} > \mathbf{CTR}_{B} \mid \mathbf{data})$$

Bayes Theorem

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

RECALL...

- ***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: Proportional

$$P(A|B) \propto P(B|A) \cdot P(A)$$

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* (generally with different parameter values)

BETA DISTRIBUTION

- Represents a distribution of probabilities-
  - That is, it represents all the possible values of a probability when we don't know what that probability is
- The domain of the Beta distribution is (0, 1), just like a probability
- Parameterization: Beta(alpha+***hits***, beta+***misses***)
  - note: mean = alpha / (alpha + beta)

BETA DISTRIBUTION

β(alpha, beta)

- Beta Distribution on Wikipedia
- Beta Distribution Notebook

BETA DISTRIBUTION

$$\beta(\text{alpha}, \text{beta})$$

alpha = views * CTR,

- these are the ***hits***

beta = views * (1 - CTR),

- these are the ***misses***

Beta distribution

β(alpha, beta)

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

BETA DISTRIBUTION

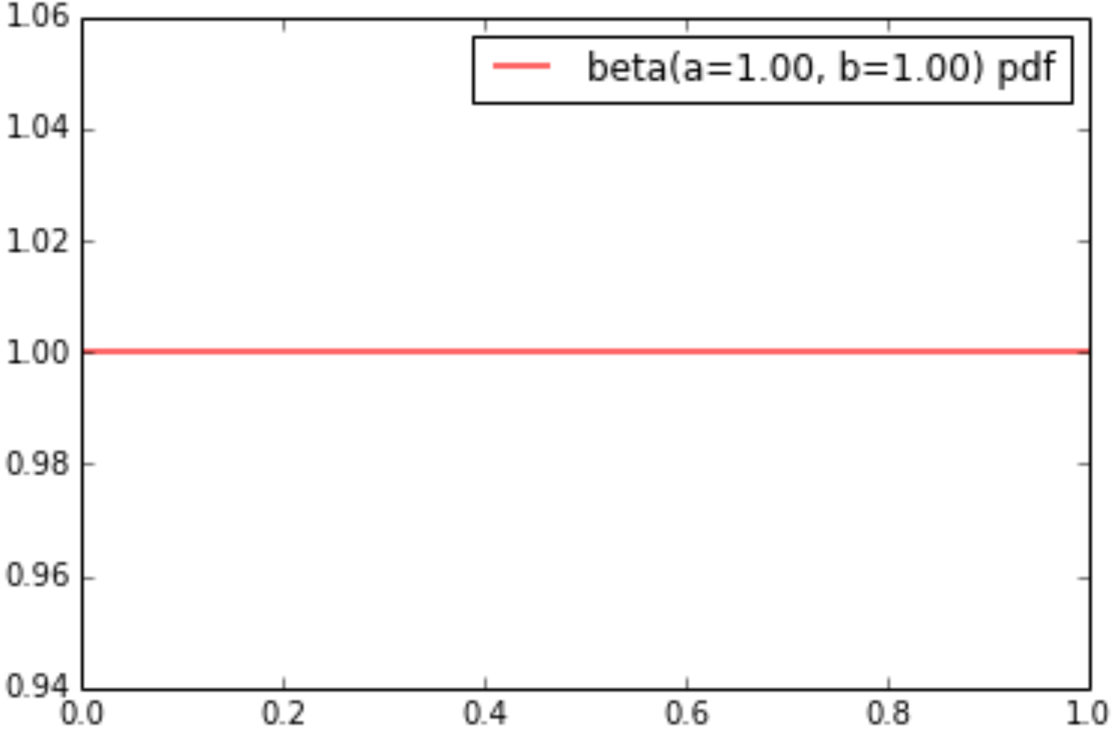
$$\beta(\text{alpha}, \text{beta})$$

Prior:

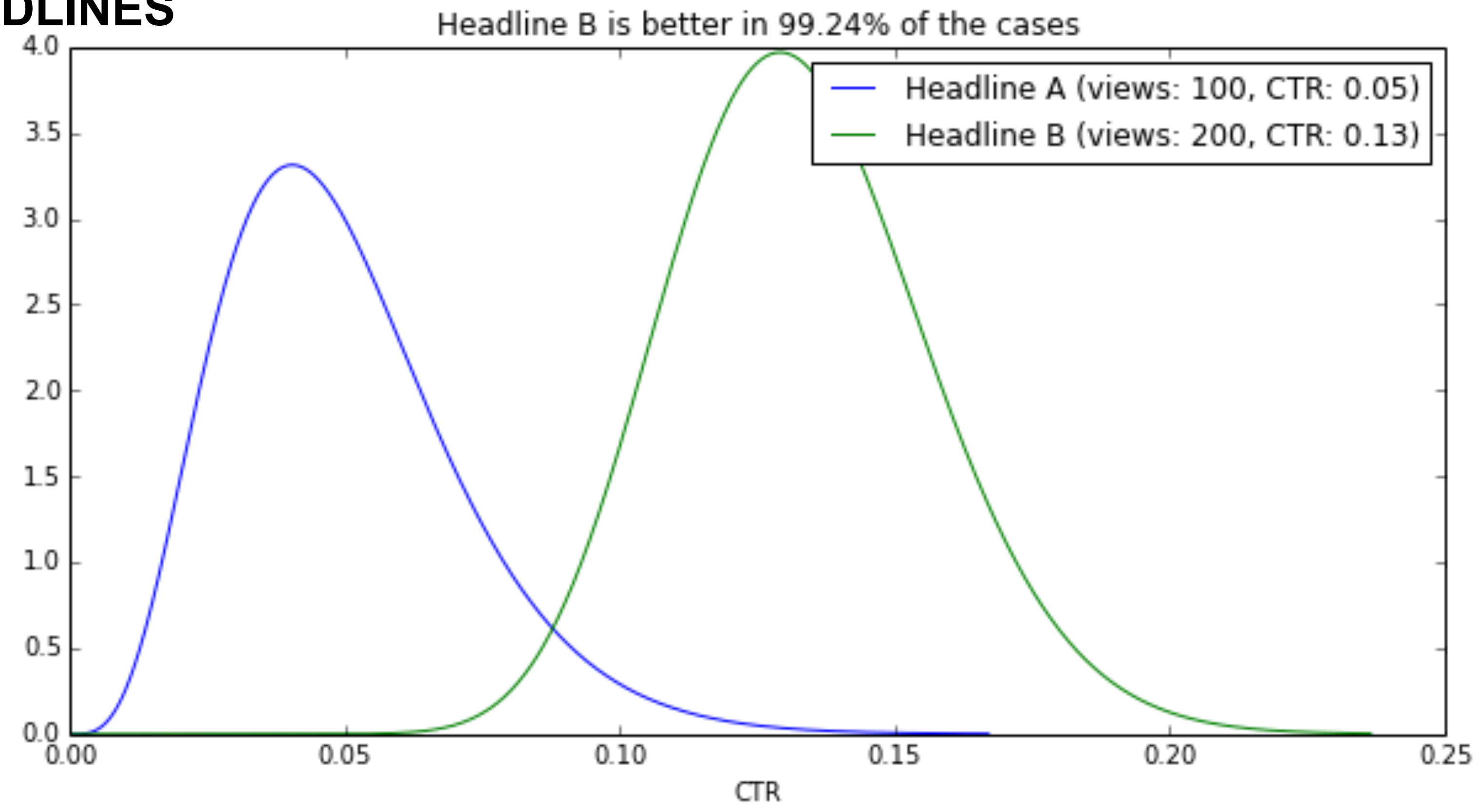
$CTR_A = \beta(1, 1)$,

$CTR_B = \beta(1, 1)$

Beta distribution: Uniform Prior



Beta distribution: Comparing Two Headlines

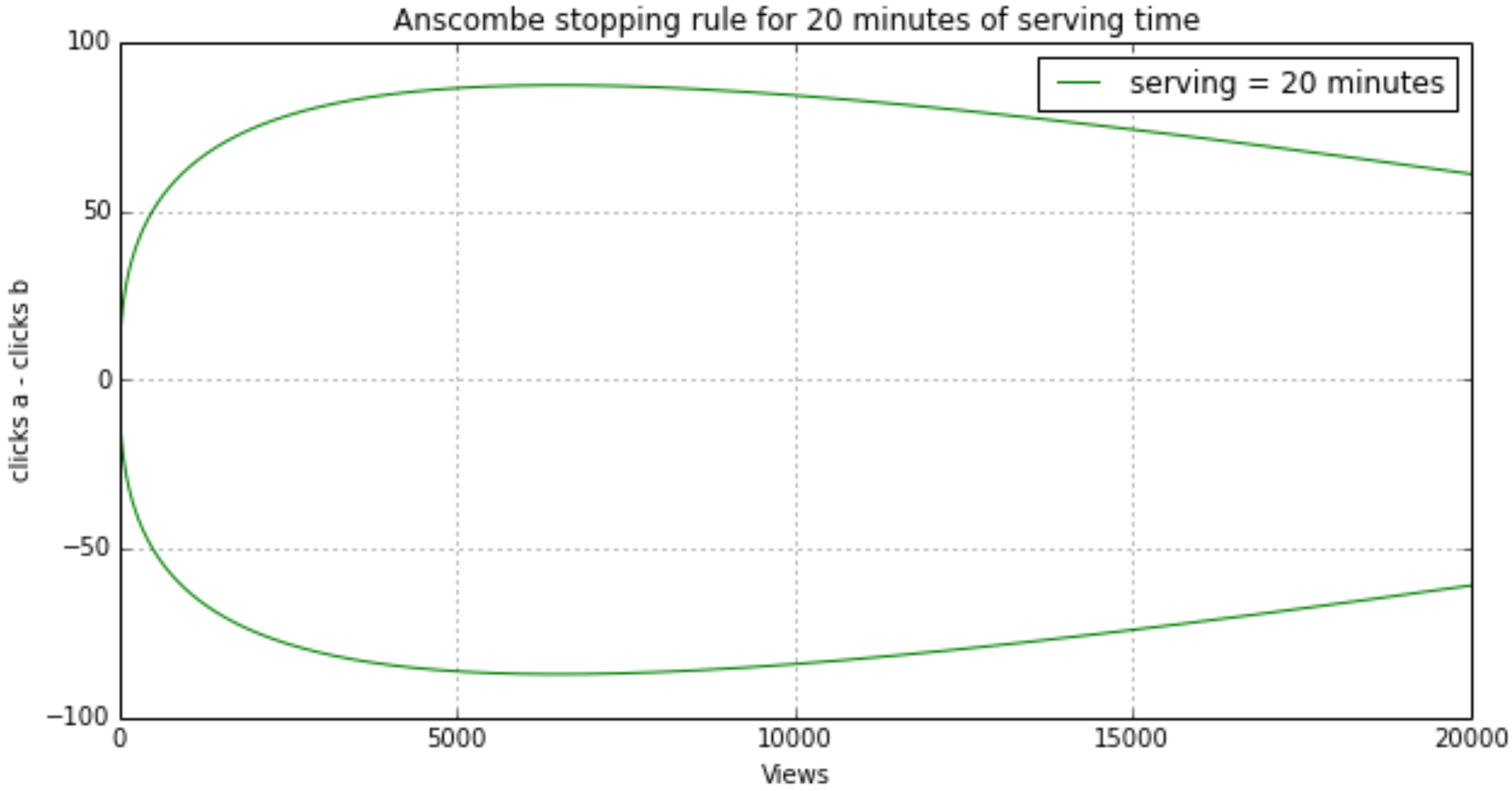


ANSCOMBE BOUNDARY

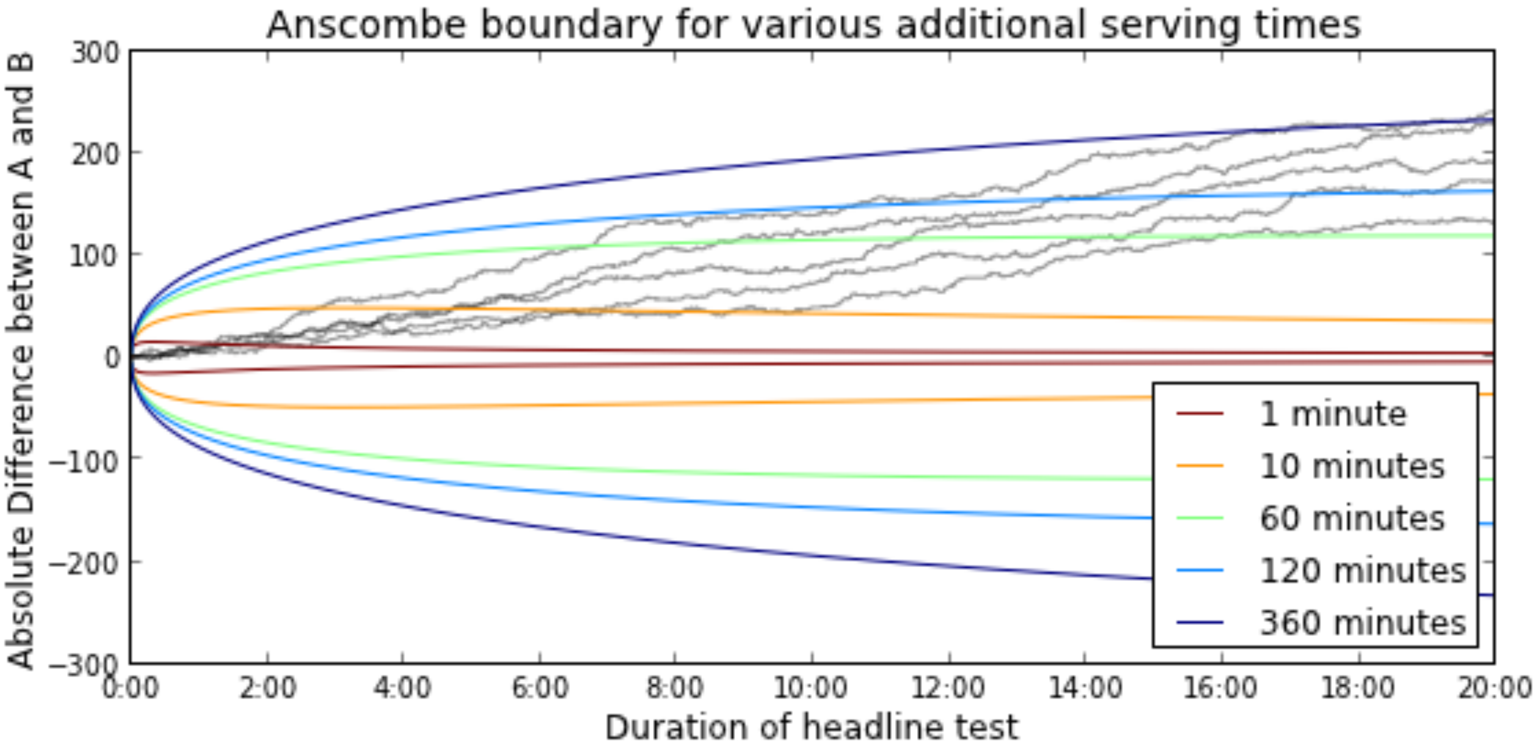
$$|y| > \phi^{-1}\left(\frac{n}{(k+2n)}\right)\sqrt{n}$$

y = absolute difference between clicks for both headlines
n = number of page views so far
k = number of future readers (i.e. page views) who will be exposed to test, given a maximum time

ANSCOMBE BOUNDARY



ANSCOMBE BOUNDARY

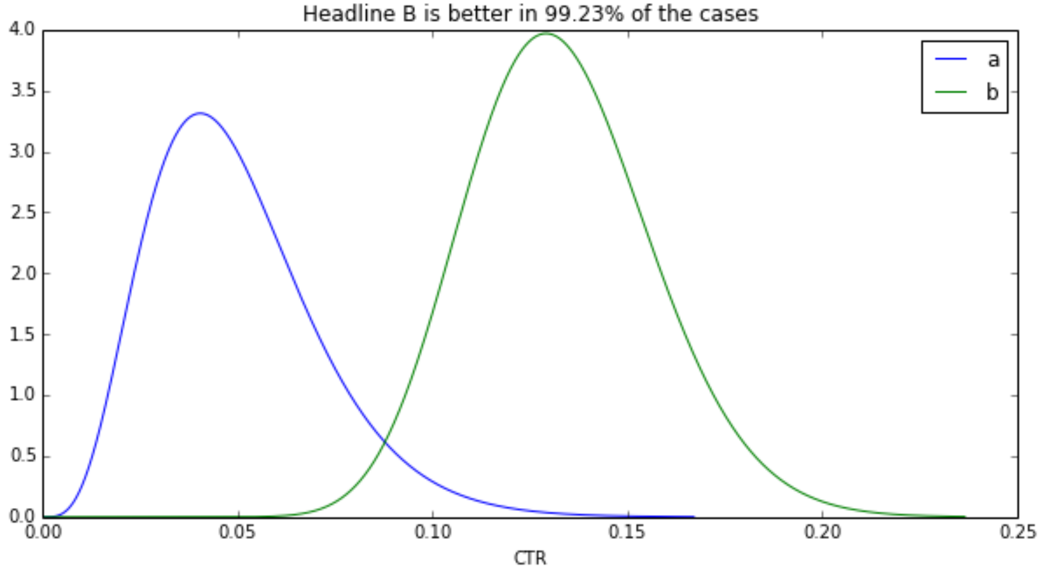


IV. Certainty of conclusion

Method to Check Certainty of Conclusion

```
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)
```

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



BETA DISTRIBUTION