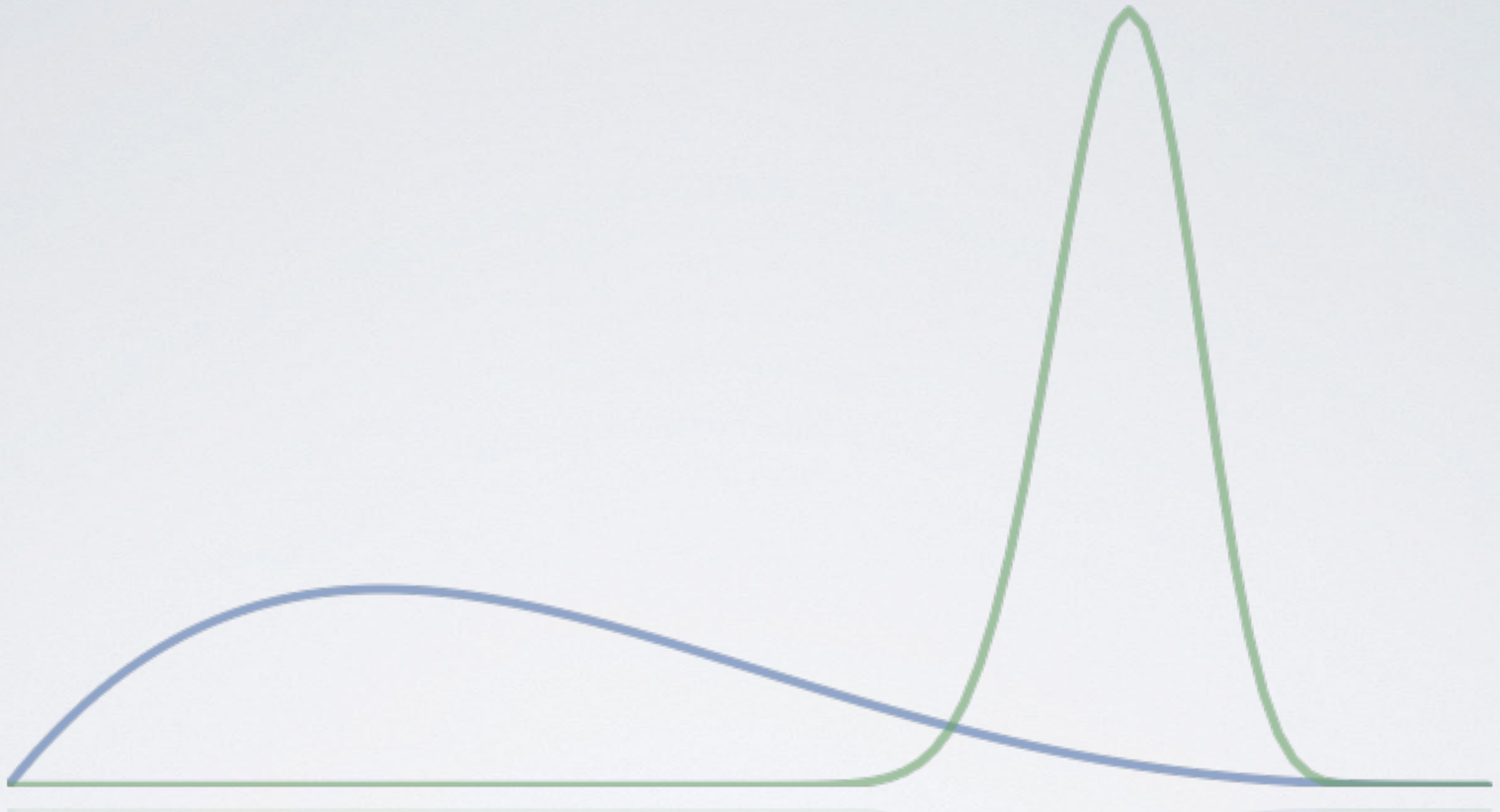


# MACHINE LEARNING FOR MARKETING OPTIMIZATION

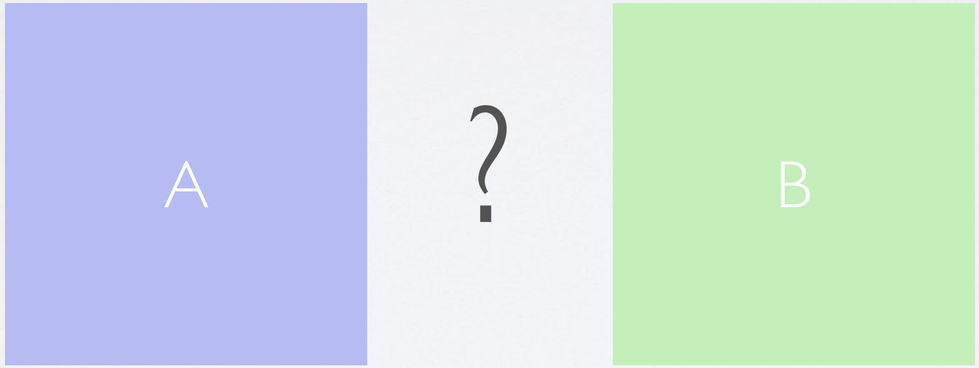
Multi-Armed Bandits - A Technical Overview

Zank Bennett

# CHALLENGE

We have two similar-content messages (A and B), to show to a large user pool

We know little or nothing about our users




Which message do we show to **maximize profits**, or to
**minimize regret** of showing the wrong message?

# MULTI-ARMED BANDITS

If you have 100 quarters and two slot machines, how do you select the right machine to play to maximize your winnings?

Do you *explore* different machines or *exploit* the one that seems to be paying off?


The multi-armed bandit algorithm is a near-optimal way to maximize payoff while balancing the explore/exploit tradeoff

# HOW IT WORKS: UNDERSTANDING PDF'S

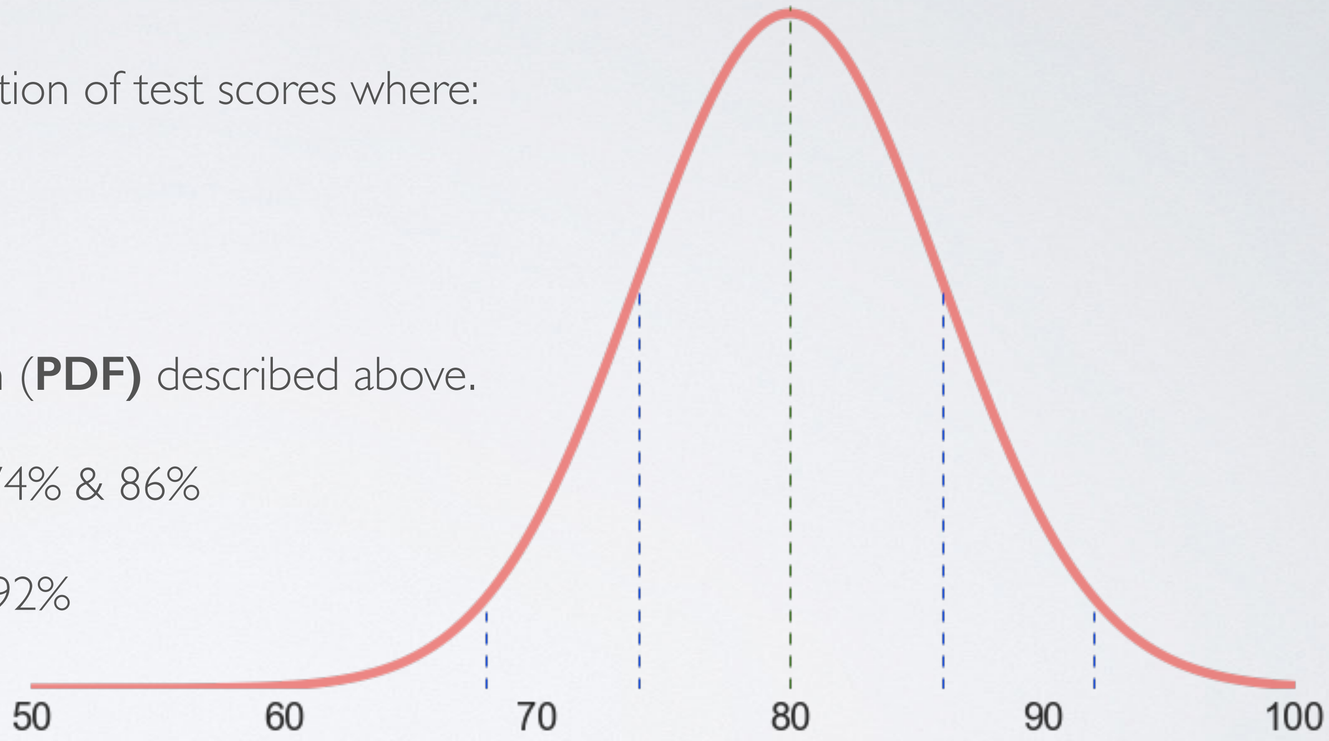
As an example, assume this Gaussian distribution of test scores where:

$\mu = 80\%$
$\sigma = 6\%$

To the right is the probability density function (**PDF)** described above.

68 % of students received a score between 74% & 86%

95% of students received between a 68% & 92%




If you choose a student at random, 95% of the time, you'll choose a student who scored between 68% & 92%

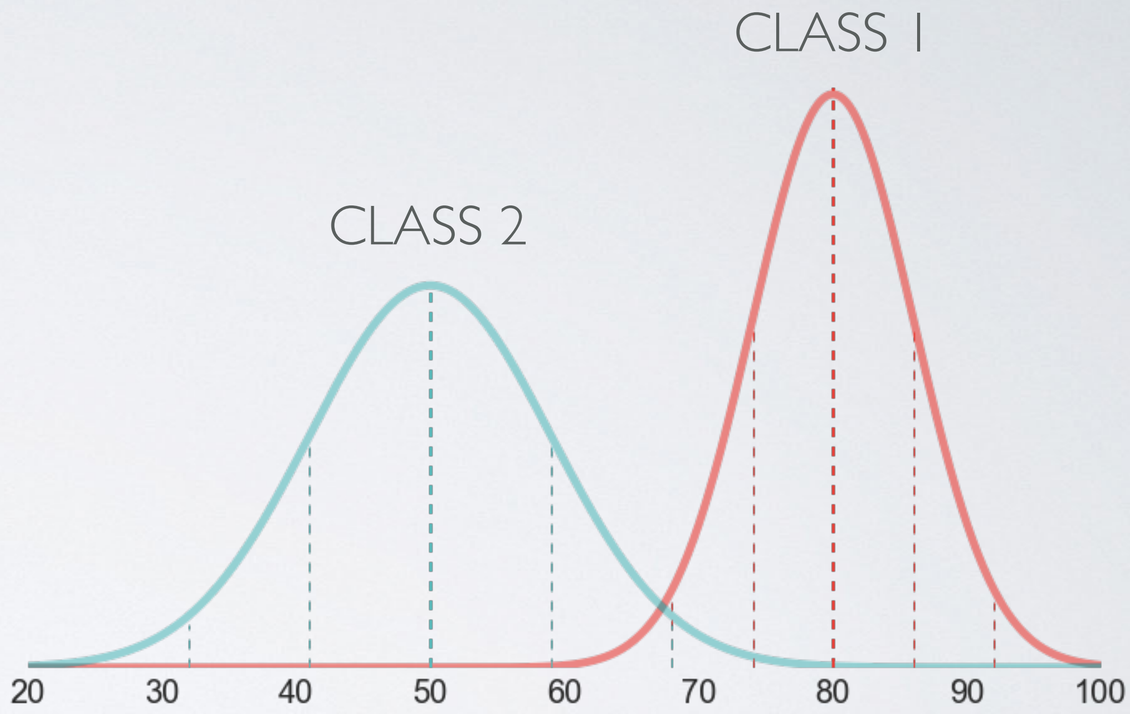
# HOW IT WORKS: SAMPLING

Assume two classrooms:

| | CLASS 1 | CLASS 2 |
|---|---|---|
| μ | 80 | 50 |
| σ | 6 | 9 |

Choose a student from each class.
What is the confidence that a student from class 2 will have a higher score than a student from class 1?

If each PDF represented the confidence in observed CTR's, which message would you show?
How often would you be wrong?



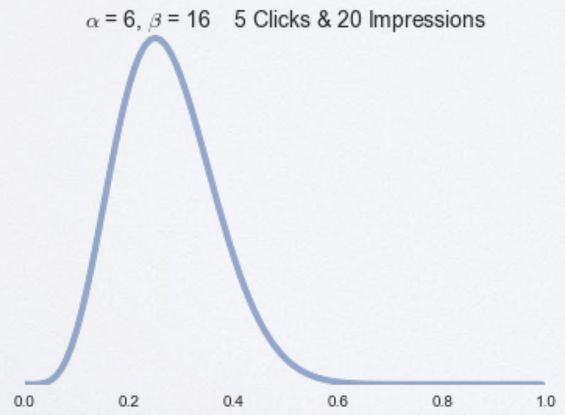

Measured CTR's are associated with some uncertainty.
PDF's allow assessment of that uncertainty.

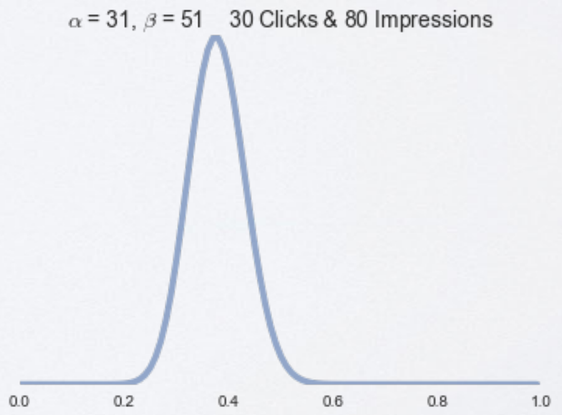
# PROBABILITY & THE BETA DISTRIBUTION

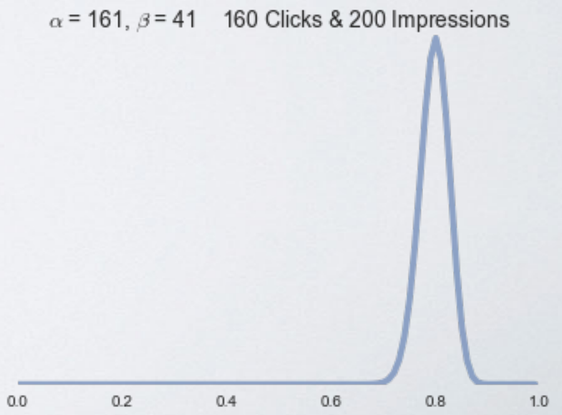
The beta distribution is the conjugate prior for the Bernoulli distribution. So, sampling values from the appropriate beta distribution allow us to model the variability associated with a given number of clicks and impressions. This concept is at the heart of the multi-armed bandit algorithm with Thompson sampling. The shape of the beta distribution is controlled entirely by two shape parameters, $\alpha$ and $\beta$.

Here are some beta distributions with $\alpha$ and $\beta$ values: with number of clicks and impressions. Priors (initial shape parameters that represent what we know before collecting any data) assumed to be 1.

$$\alpha = \text{prior} + \text{number of clicks}$$
$$\beta = \text{prior} + \text{number of impressions} - \text{number of clicks}$$

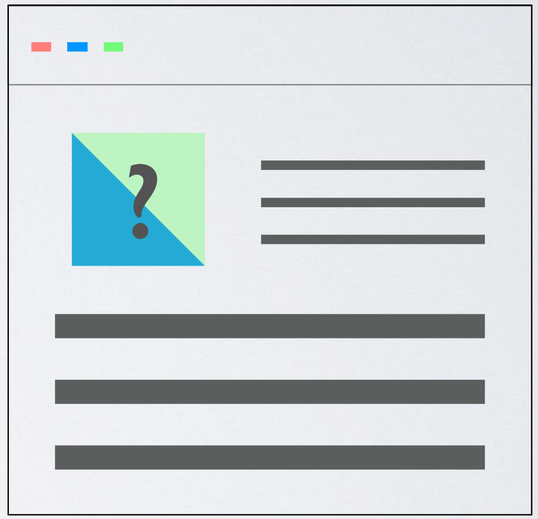



# MAB WITH THOMPSON SAMPLING

For two given messages with unknown click-through rates (CTR's), which message do we show?

**MAB Algorithm with Thompson Sampling. At each message impression, record the number of impression and clicks:**

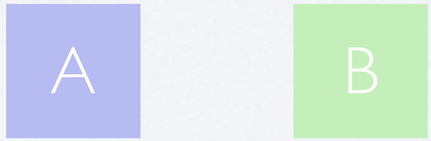
1. Initially,  nothing is known about the CTR of either message, so, show either message at equal probability (50/50 chance)
2. Using the number of impressions and clicks, generate the new *posterior probability* of choosing each message
3. Using those posteriors, sample each distribution and use those values as the CTR's
4. Show the message with the highest sampled CTR value




This process repeats each time an message is shown.

# MAB WALKTHROUGH

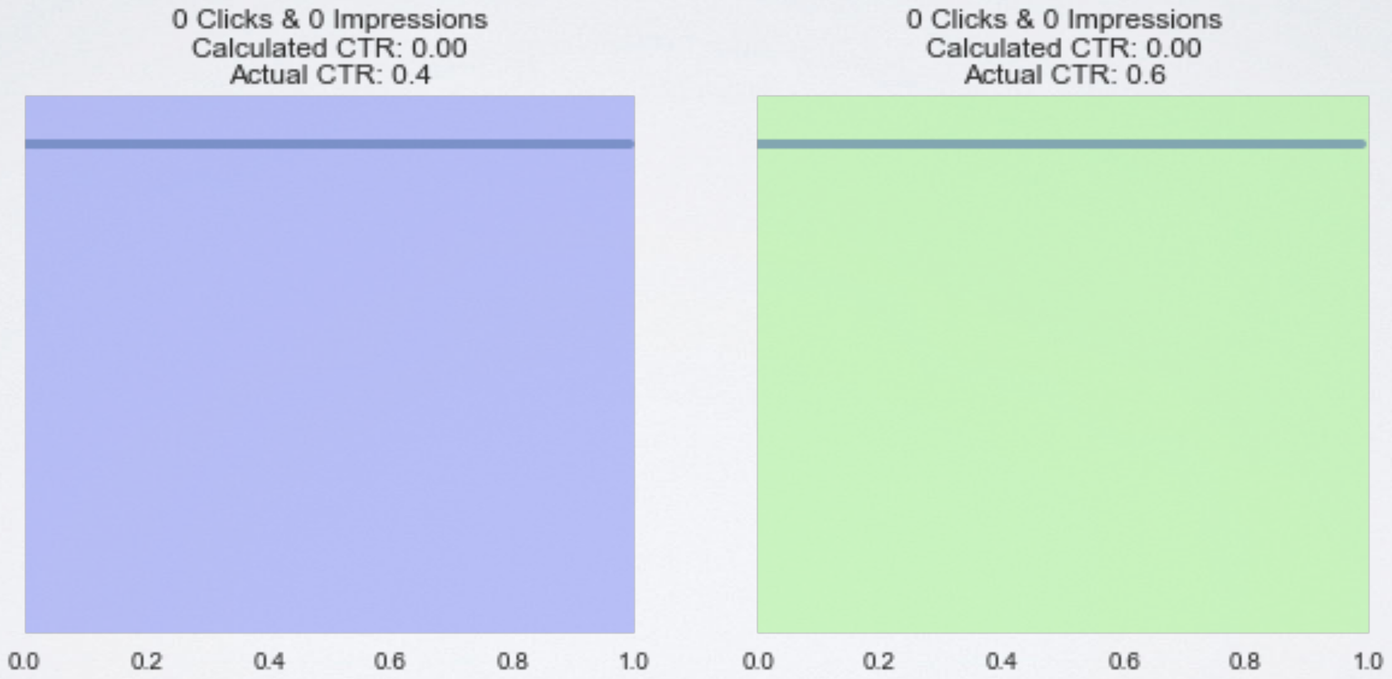
Throughout: assume two messages, A & B
with CTR's of 0.4 and 0.6 respectively




CTR = 0.4 CTR = 0.6

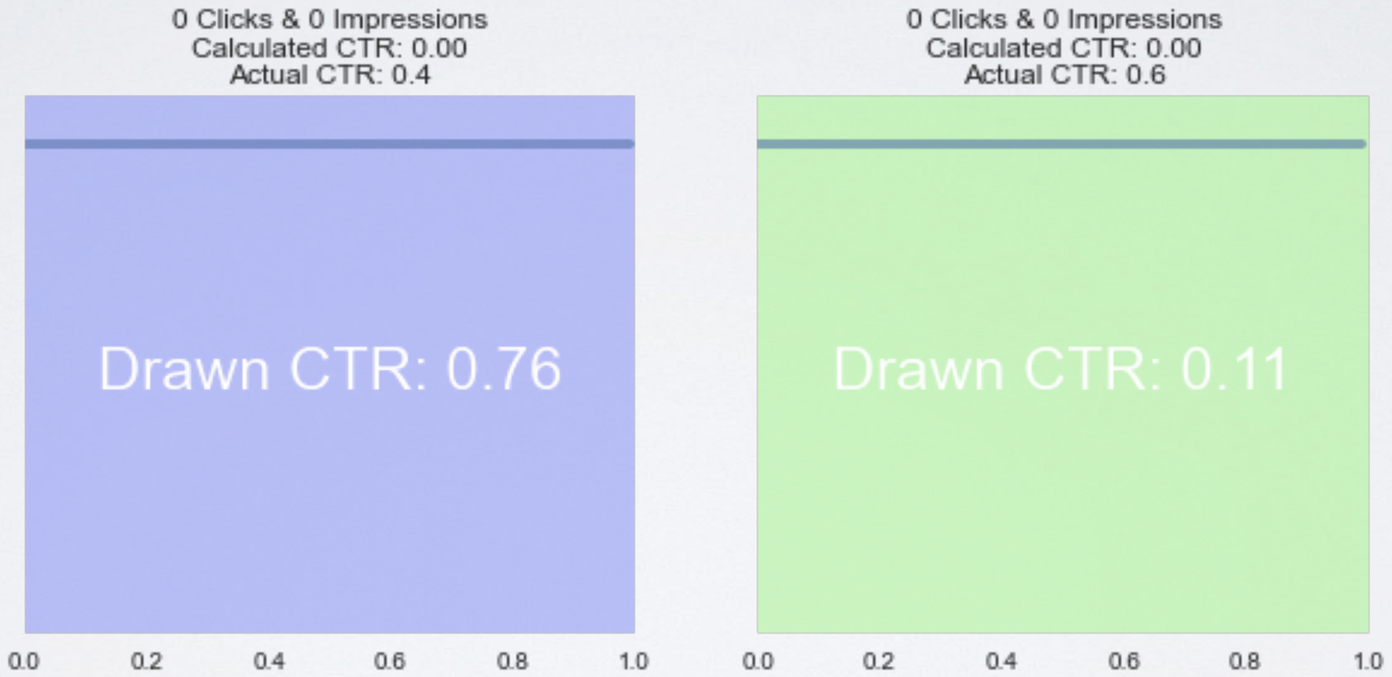
# MAB WALKTHROUGH: STEP 1

1. Initially,  nothing is known about the CTR of either message, so, show either message at equal probability (50/50 chance)



# MAB WALKTHROUGH: STEP2

2. Using the number of impressions and clicks, generate the new *posterior probability* of choosing each message




In this case, message A is shown. In the simulation, draw from a Bernoulli distribution with E[x] = 0.4 to see if A was clicked by the user

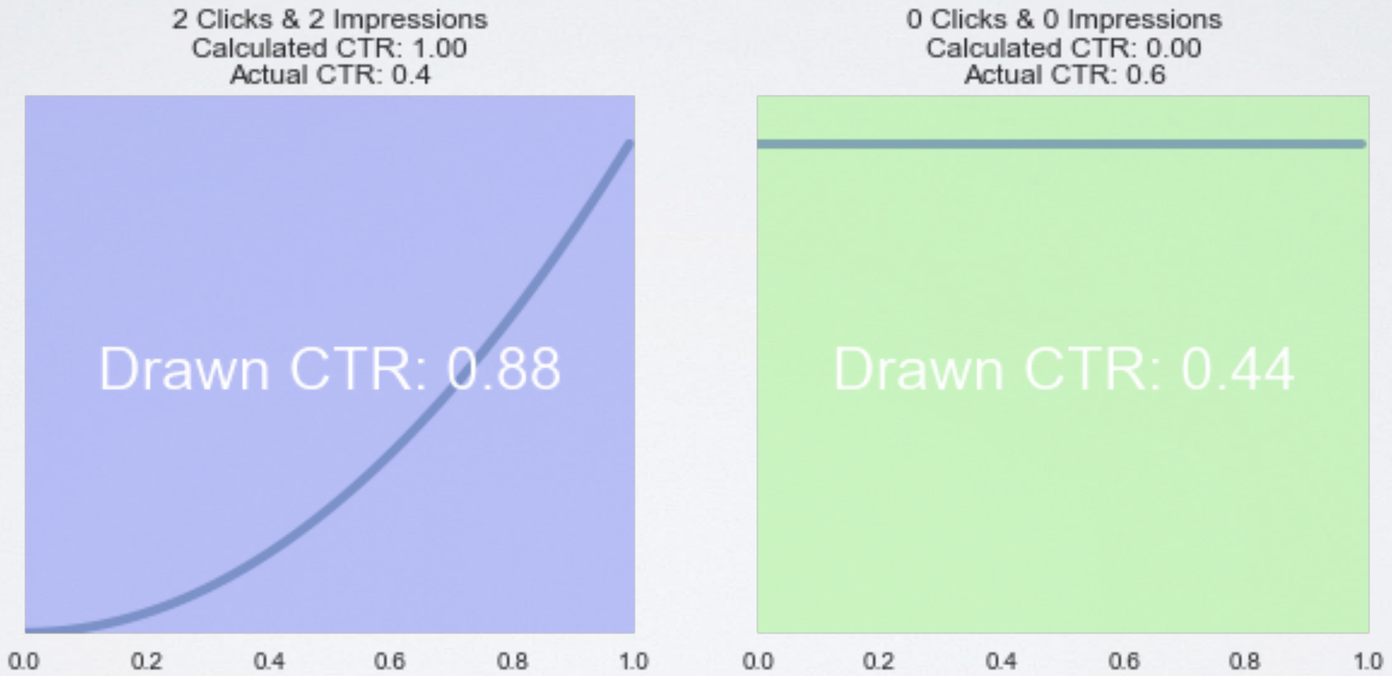
# MAB WALKTHROUGH: STEPS 2-4

2.-4. Using the number of impressions and clicks, generate the new *posterior probability* of choosing each message


In this case, message A *was* clicked. Increment the number of clicks and number of impressions, then plot new PDF. Notice how A's PDF is biased towards higher CTR's based on the evidence collected

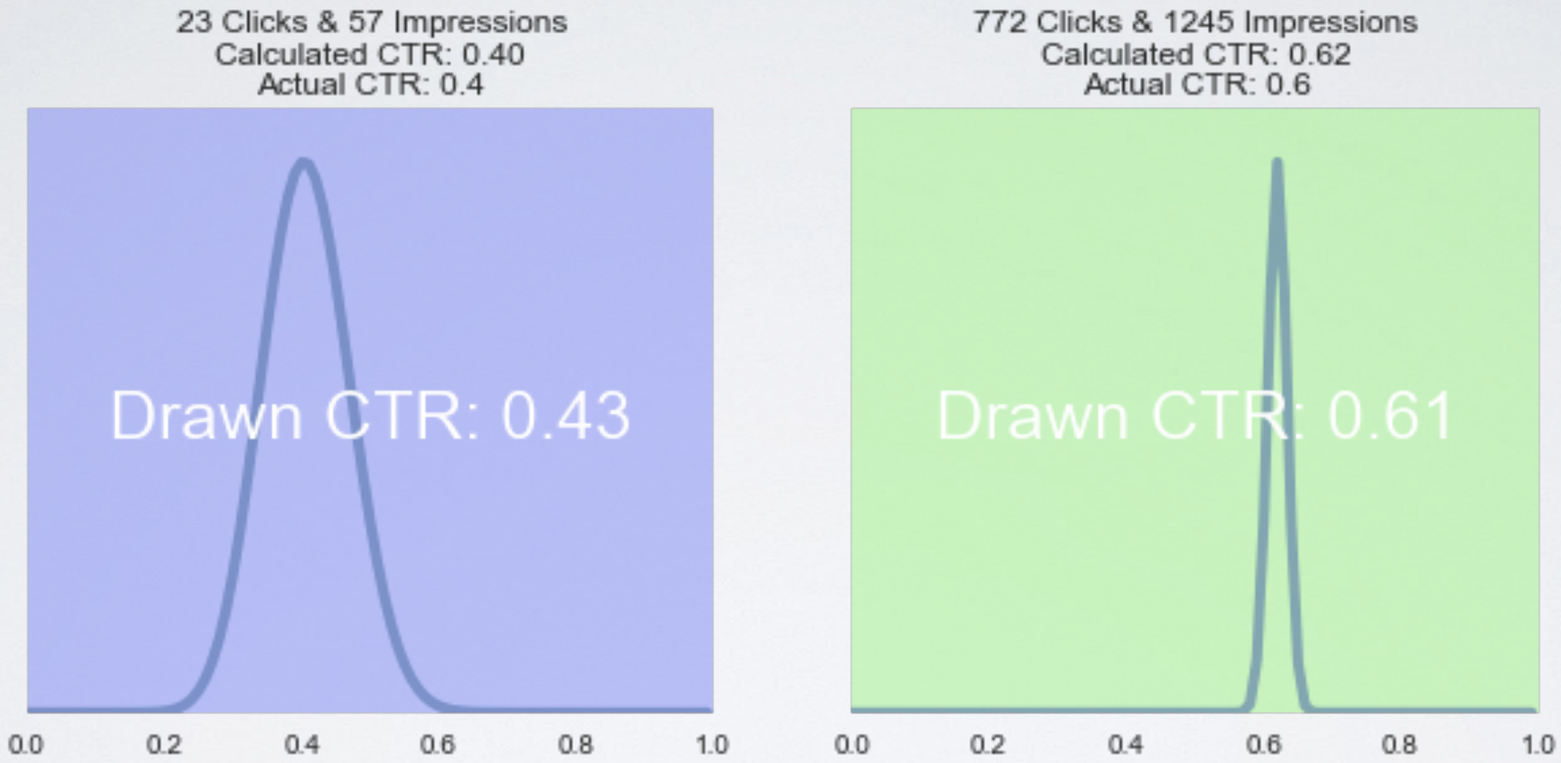
# MAB WALKTHROUGH: REPEAT STEPS 2-4

2.-4. Using the number of impressions and clicks, generate the new *posterior probability* of choosing each message




Message A *was* again clicked. Increment the number of clicks and number of impressions, then plot new PDF. Notice how A's PDF is biased non-linearly towards higher CTR's based on the evidence collected

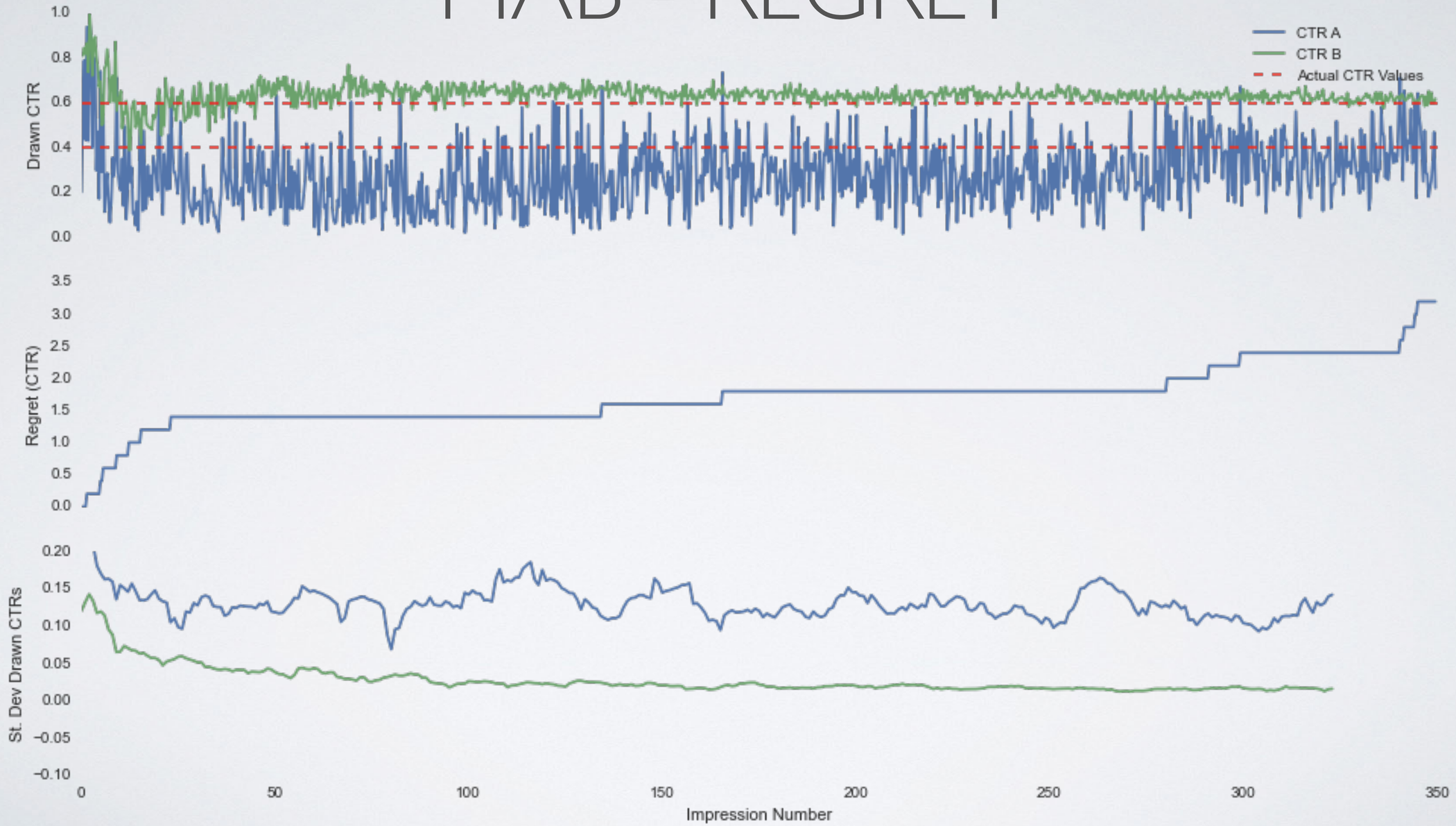
# MAB WALKTHROUGH: REPEAT 1000X




Message B has received most of the impressions. Its distribution is very narrow: CTR samples chosen from B will hover close to the actual CTR. In other words, **the MAB *learned* the CTR of the winning message.**

The MAB shows CTR's for message B that are very close to 0.62 and CTR's for message A within the range of 0.2 to 0.6.

# MAB - REGRET


CTR A
CTR B
Actual CTR Values
Drawn CTR
Regret (CTR)
St. Dev Drawn CTRs
Impression Number

# MAB

PROS:

- Useful when you don't know much/anything about the audience
- Ease of implementation
- Easy to plot/interpret usage statistics
- Not very computationally expensive
- It converges on a winner; you don't have to pick a winner at the end
- It's fast

CONS:

- Can be difficult to explain clearly to stakeholders
- There are better techniques when user/item information is known

# THANK YOU


Download slides and all supporting IPython code:
https://github.com/zankbennett/multi_armed_bandit_tech_overview

Contat me at
zankbennett@gmail.com

REFERENCES / READING:
Online simulator: https://e76d6ebf22ef8d7e079810f3d1f82ba1e5f145d5.googledrive.com/host/0B2GQktu-wcTiWDB2R2t2a2tMUG8/
Beta distribution: https://en.wikipedia.org/wiki/Beta_distribution