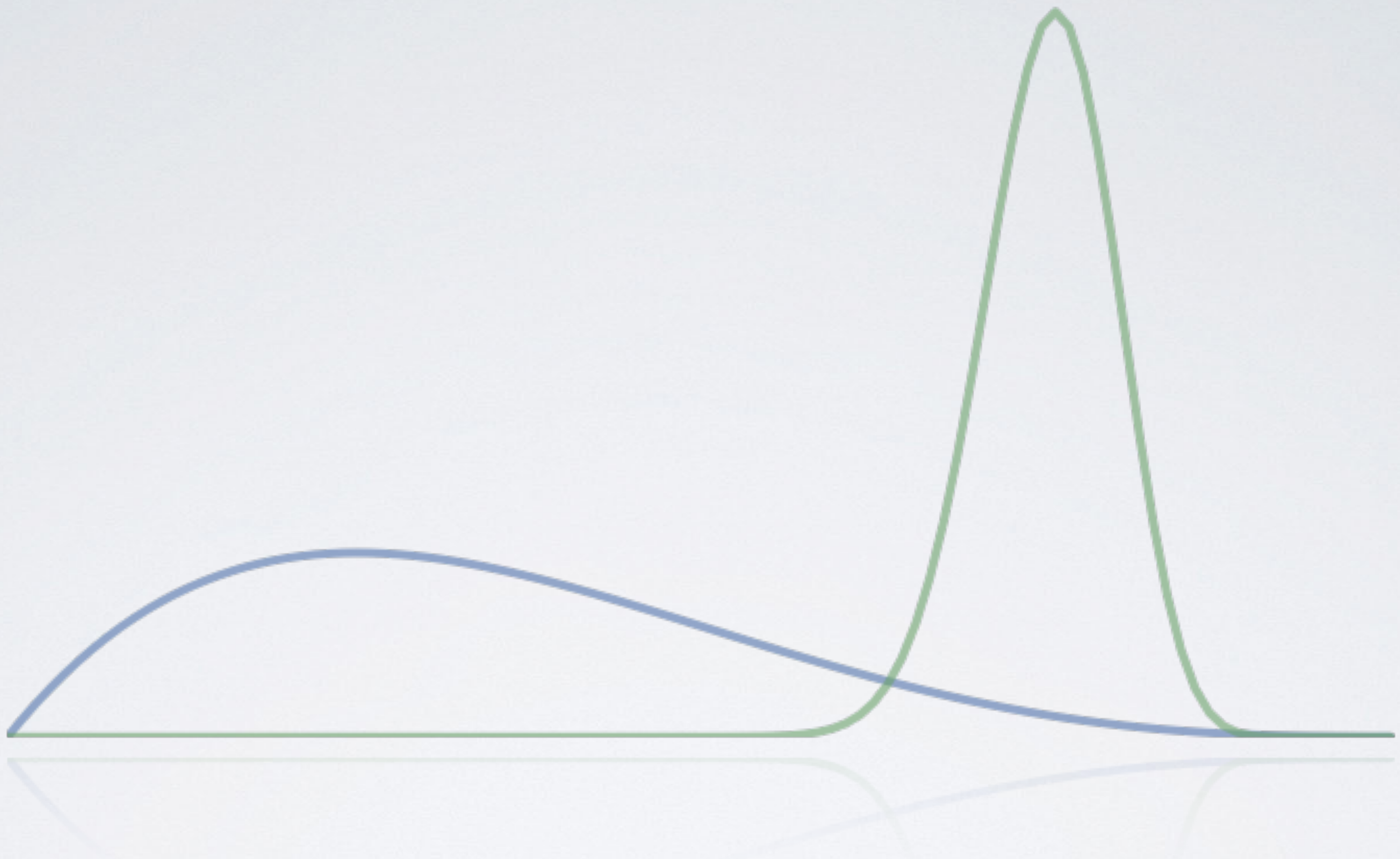


MACHINE LEARNING FOR MARKETING OPTIMIZATION

Multi-Armed Bandits with Thompson Sampling - A Technical Overview

Zank Bennett - Data Scientist

OVERVIEW

A ? B

PROBLEM:
As practitioners of predictive analytics, we are often tasked with presenting content quickly to large populations about which little or nothing is known.

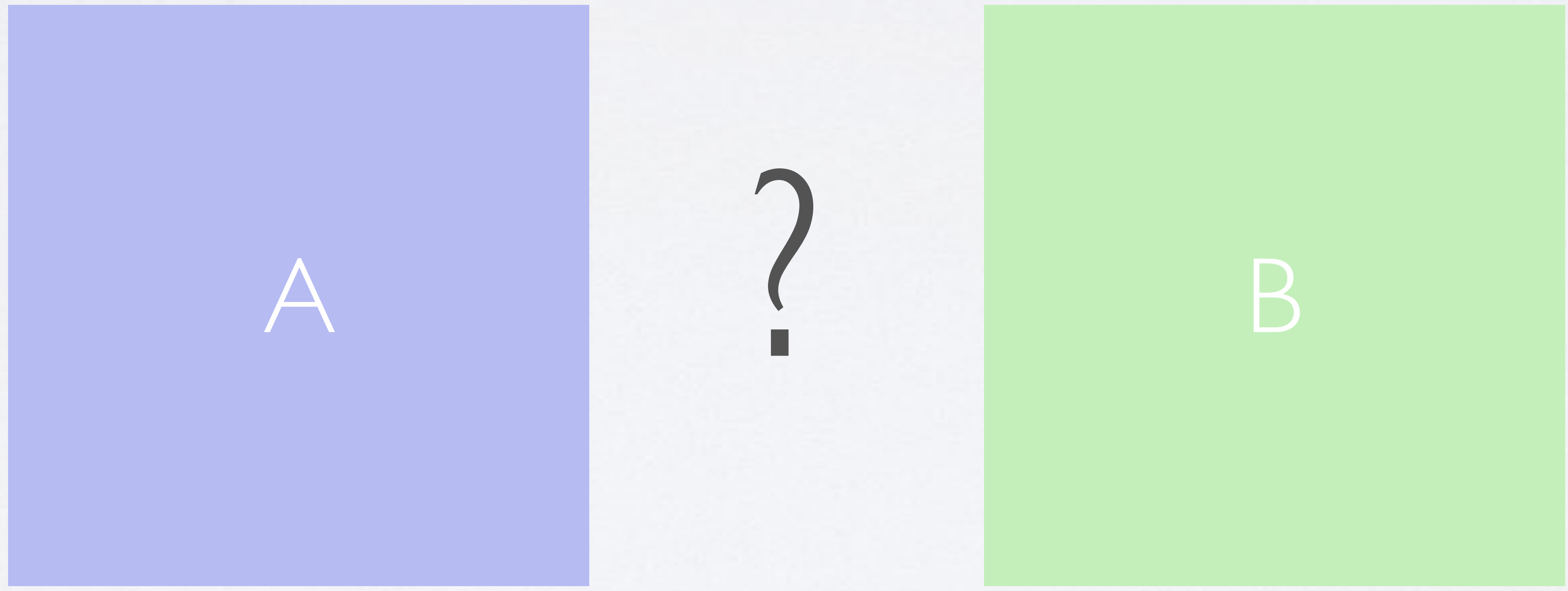
CHALLENGE:
Given a myriad of choices of what to present to a user (A/B/n), how can we train a machine to adaptively learn and exploit the option that will provide the greatest benefit to the most users?
Can we do this in a mathematically optimal or near optimal manner?

APPLICATIONS:

- Optimize email open rates by choosing the correct subject line out of several options
- Maximize ad-unit CTR
- Minimize illness during medical trials by minimizing regret of non-optimal treatments
- Adaptively choose the optimal blog post title to maximize CTR, given a myriad of options

CHALLENGE

- We have two items (A and B), to present to a large group of users
- There is some reward associated with consumption of each option
- We know little or nothing about our users



Which item do we present to unknown users to **maximize reward**, or to **minimize regret** of presenting the wrong item?

MULTI-ARMED BANDITS

If you have 100 coins and two slot machines, how do you select the correct 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

WHAT'S WRONG HERE?

HOW IT WORKS: UNDERSTANDING PDF'S

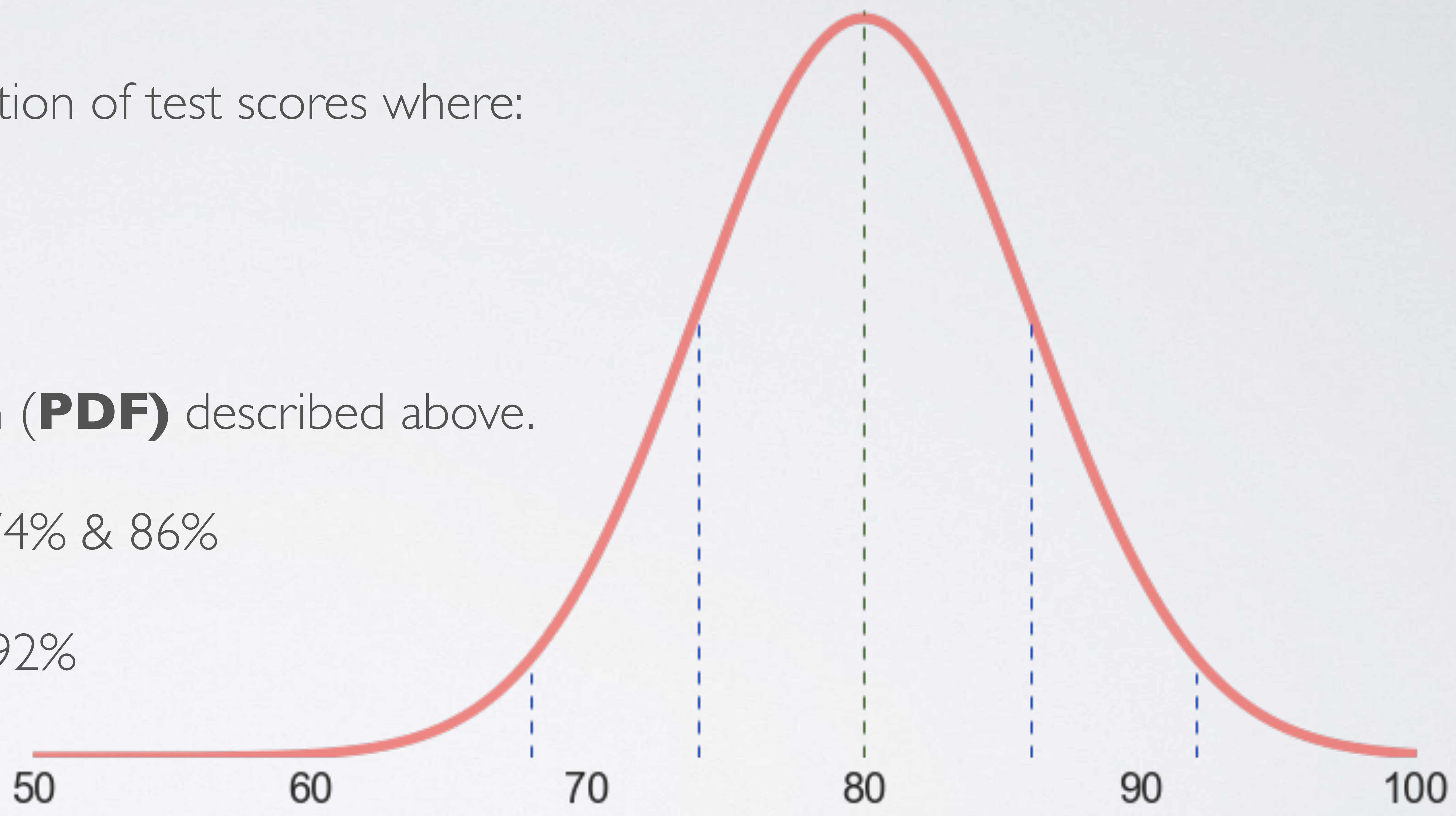
As an example, assume this Gaussian distribution of test scores where:

μ = 80%
σ = 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%

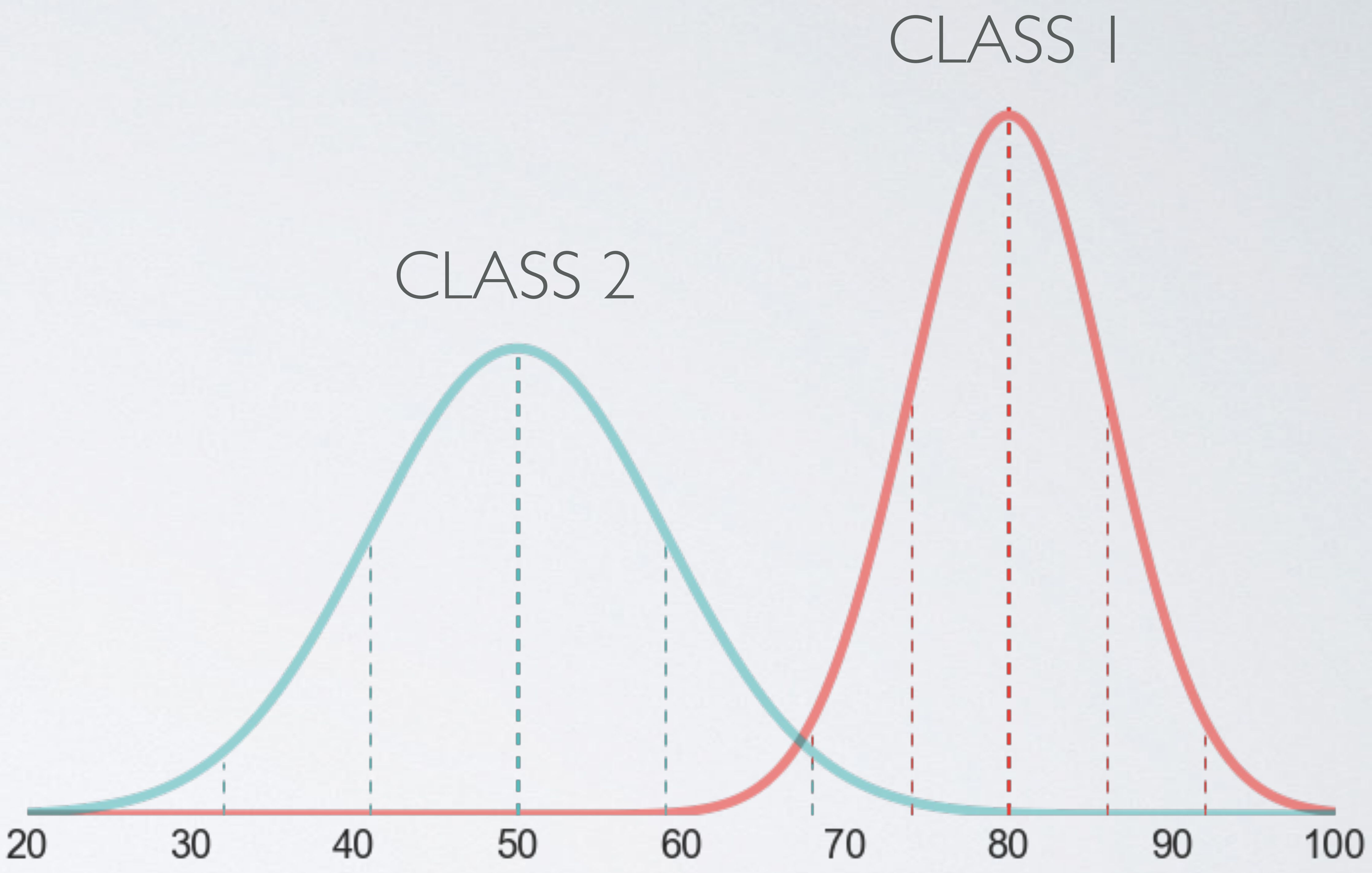
HOW IT WORKS: SAMPLING

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 ad CTR's, which ad would you show?
How often would you be wrong?



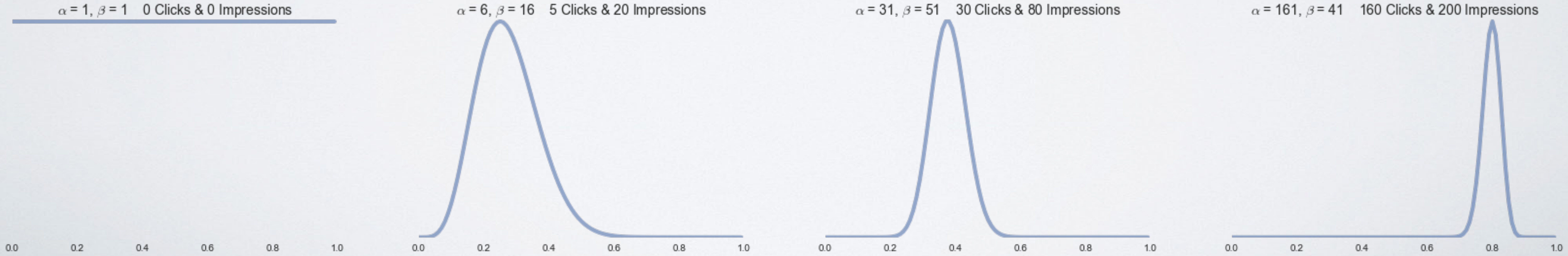
Measured CTR's are associated with some uncertainty.
PDF's allow assessment of that uncertainty.

PROBABILITY & THE BETA DISTRIBUTION

The beta distribution is the conjugate prior for the Bernoulli distribution. So, sampling values from the appropriate beta distribution allows us to model the variability associated with a given number of successes and attempts. 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, α and β.

Assuming ad clicks and impressions, here are some beta distributions with α and β values with number of clicks and impressions noted. Priors (initial shape parameters that represent what we know before collecting any data) are assumed to be 1.0.

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

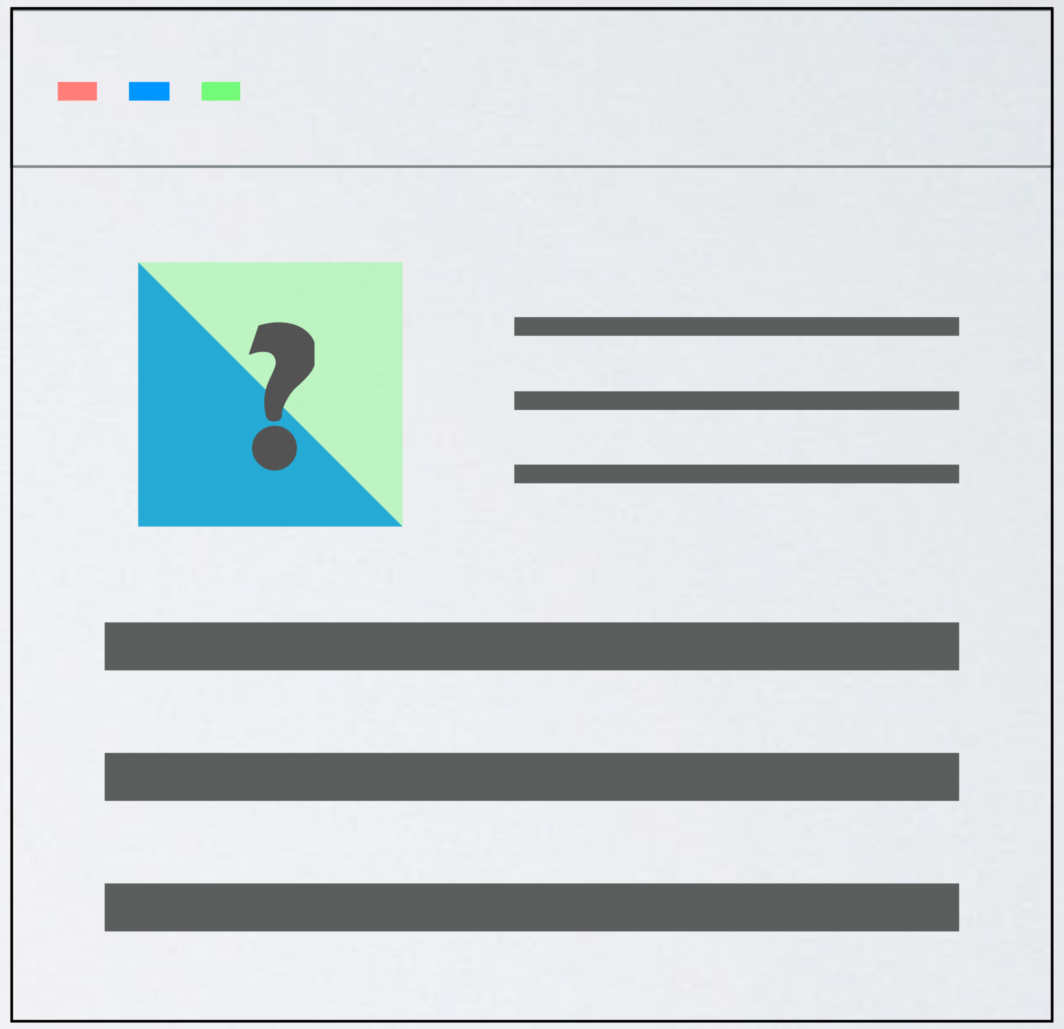


MAB WITH THOMPSON SAMPLING

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

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

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



This process repeats each time an ad is shown. A full walkthrough follows.

MAB WALKTHROUGH

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

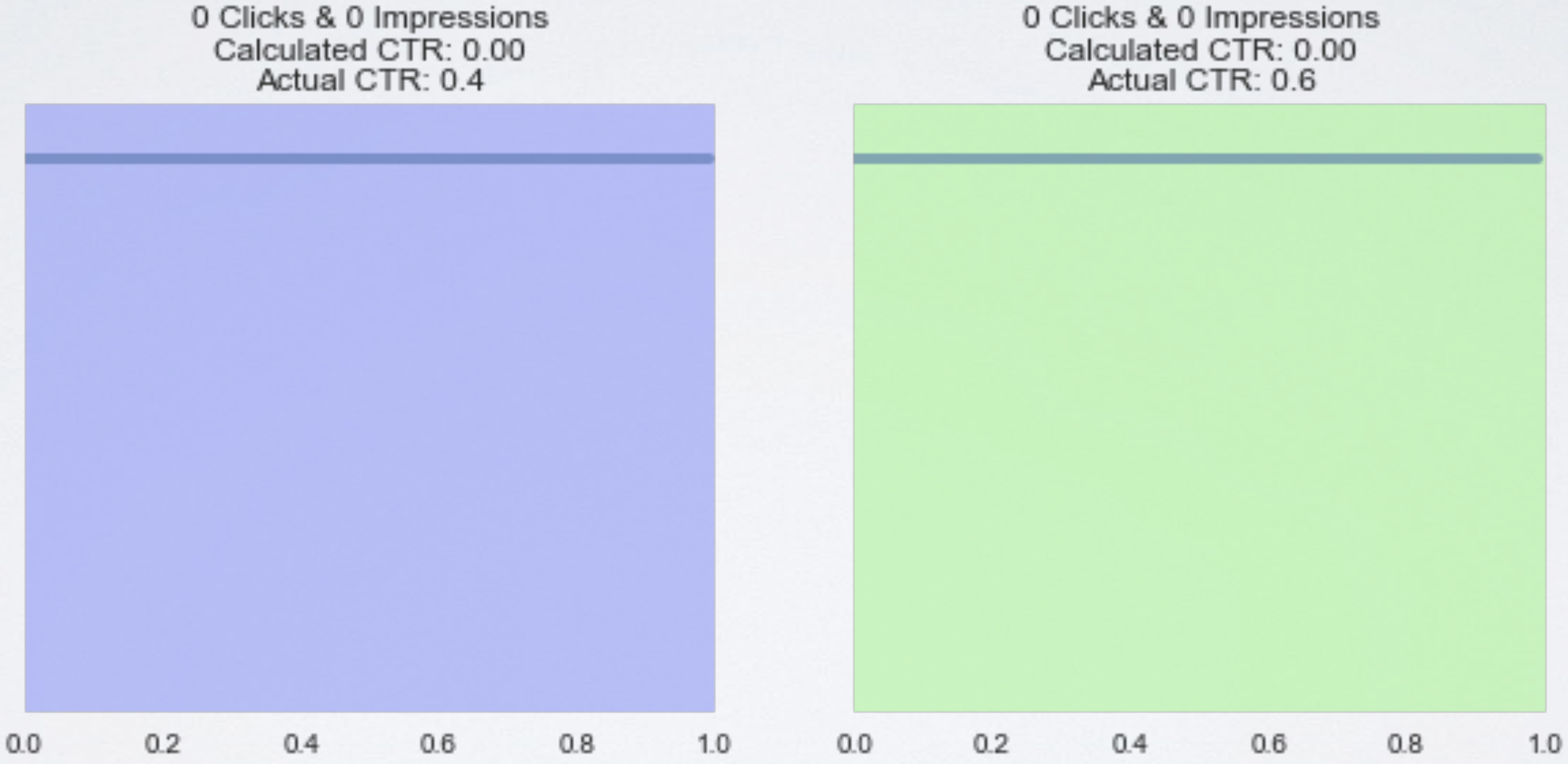
A	B
CTR = 0.4	CTR = 0.6

Every time ad A is shown, there is a known regret in CTR (0.6 - 0.4 = 0.2)

GOAL: minimize this regret (the number of impressions given to ad A)

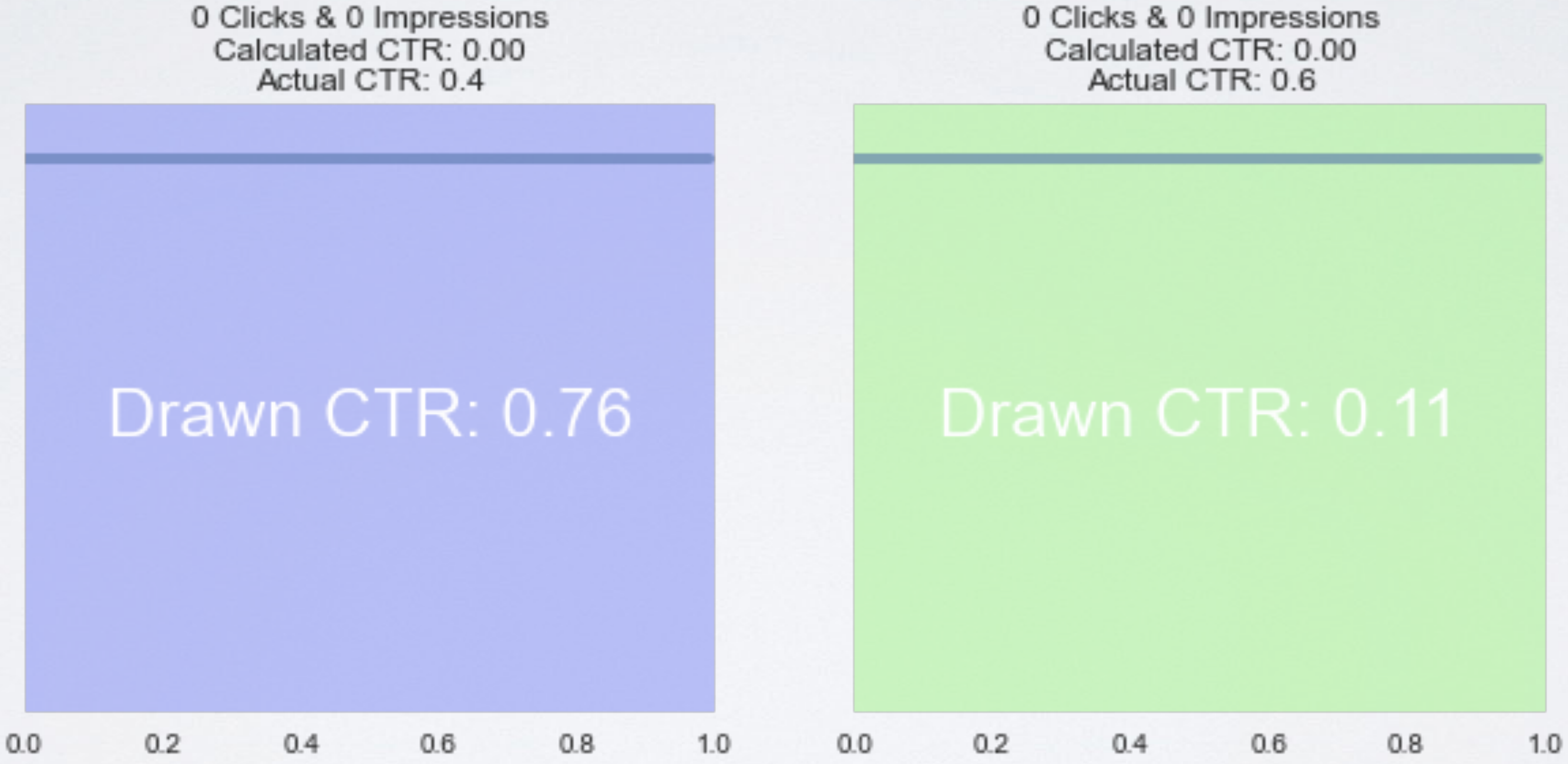
MAB WALKTHROUGH: STEP 1

1. Initially, nothing is known about the CTR of either ad, so, show either one 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 ad



In this case, ad 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

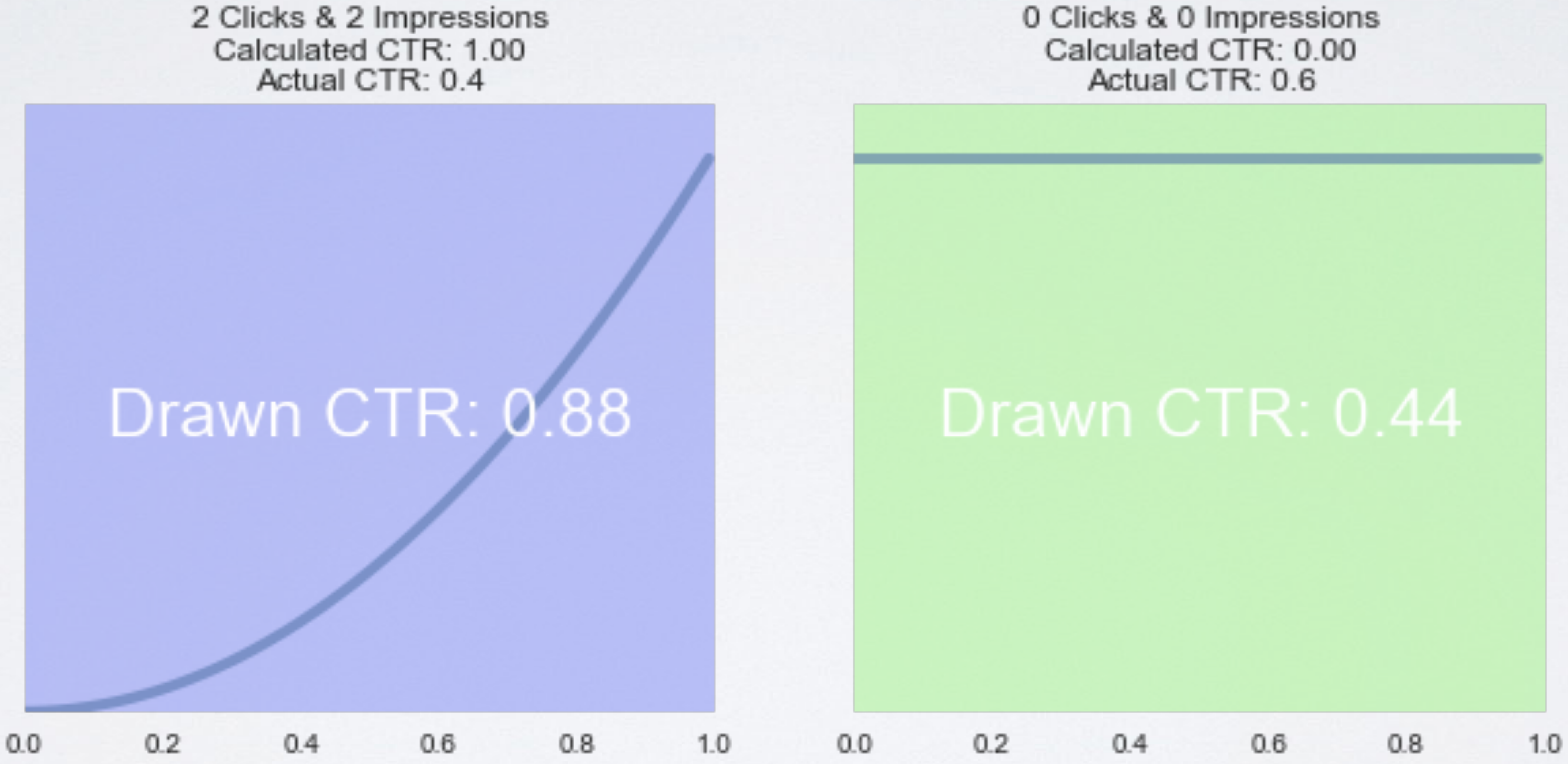
MAB WALKTHROUGH: STEPS 2-4

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

In this case, ad 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

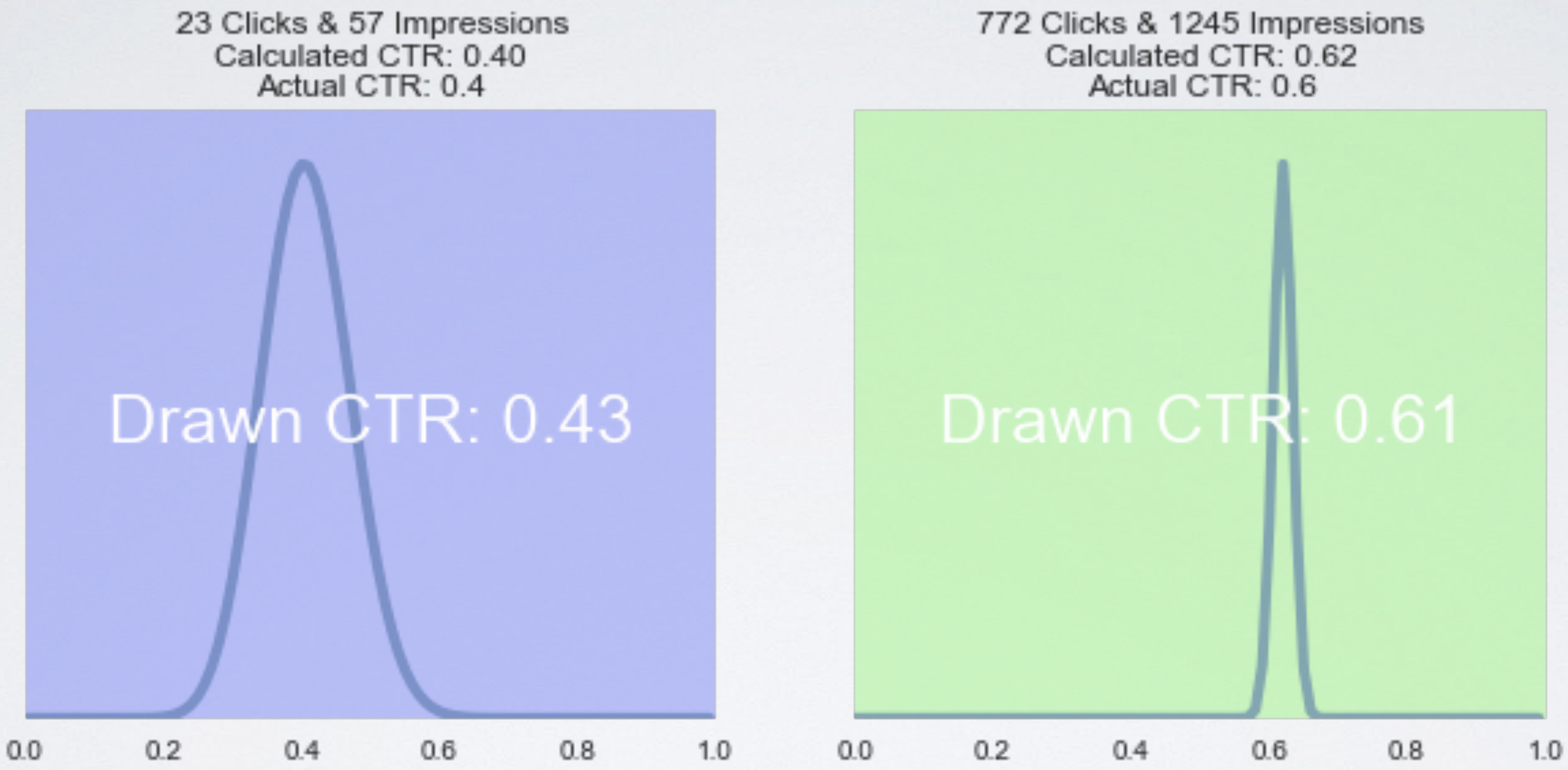
MAB WALKTHROUGH: REPEAT STEPS 2-4

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



Ad A was clicked again. 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

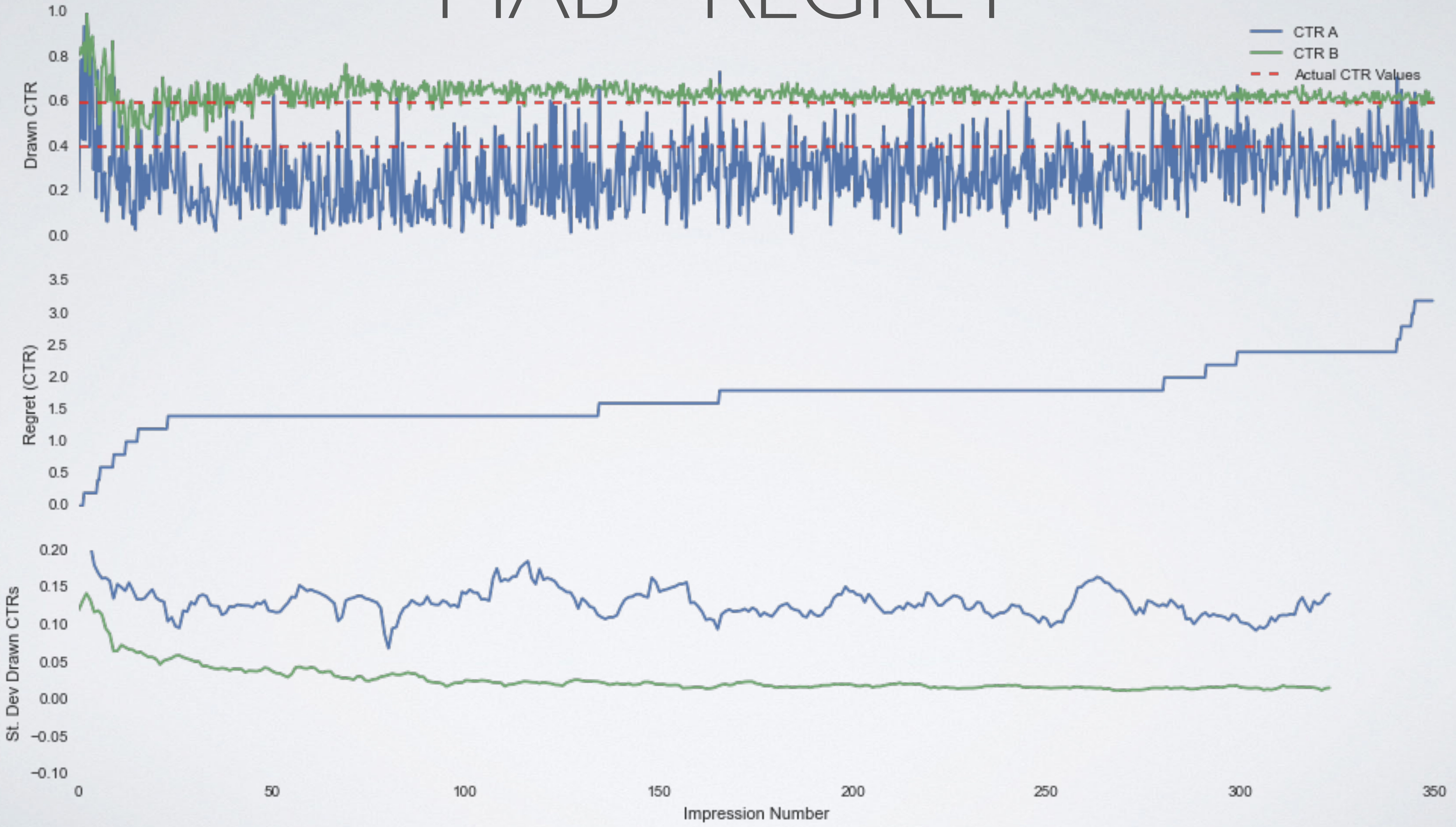
MAB WALKTHROUGH: REPEAT >1000X



Ad 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 ad.**

The MAB shows CTR's for ad B that are very close to 0.62 and CTR's for ad A within the range of 0.2 to 0.6.
This is because the 'winning' ad received many more impressions and clicks, increasing the model's confidence in the expected value.

MAB - REGRET



CTR A
CTR B
Actual CTR Values
Drawn CTR
1.0
0.8
0.6
0.4
0.2
0.0
Regret (CTR)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
St. Dev Drawn CTRs
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
0
50
100
150
200
250
300
350
Impression Number

MAB

PROS:
- Useful when you don’t know much/anything about the audience
- Ease of implementation
- Easy to plot/interpret/convey 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

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