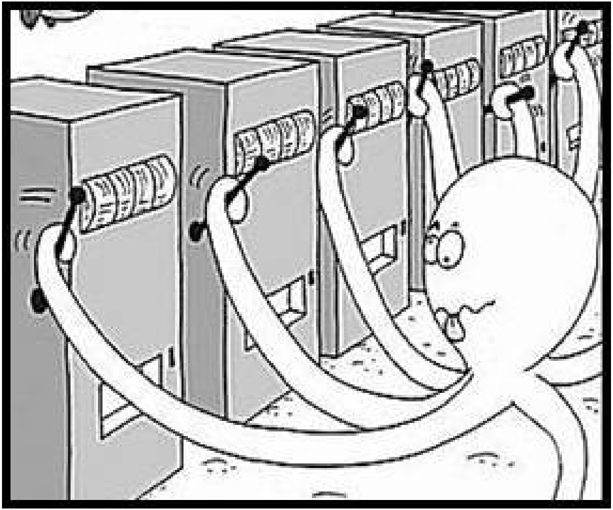
Multi-Armed Bandits

Michael Thomas

Multi-Armed Bandits (MAB)



Multi-Armed Bandits (MAB) Overview

- MABs get their name from a gambling scenario.
  - Note: A single slot machine is called a "one-armed bandit."
- Imagine a slot machine with 8 arms and each arm gives a different probability of winning.
- You don't know which arm is best, so you have to experiment.
- As you learn which is best, you'll want to experiment less and focus more on the arm with the best payout.
- Moving from the **experiment** phase to the **exploit** phase is difficult.
  - Spend too much time **experimenting** and you miss out on the chance to profit from the winning arm.
  - Start **exploiting** too soon and you might be focused on an arm that is not the best.
- Many algorithms exist to handle this, all of which are "multi-armed bandits."

Multi-Armed Bandits (MAB) Overview

- MABs get their name from a gambling scenario.
  - Note: A single slot machine is called a "one-armed bandit."
- Imagine a slot machine with 8 arms and each arm gives a different probability of winning.
- You don't know which arm is best, so you have to experiment.
- As you learn which is best, you'll want to experiment less and focus more on the arm with the best payout.
- Moving from the **experiment** phase to the **exploit** phase is difficult.
  - Spend too much time **experimenting** and you miss out on the chance to profit from the winning arm.
  - Start **exploiting** too soon and you might be focused on an arm that is not the best.
- Many algorithms exist to handle this, all of which are "multi-armed bandits."

Multi-Armed Bandits (MAB) Key Characteristics of Problem

- Sequential decisions.
- Choosing from two or more different options (e.g, two or more arms).
- Uncertain about the payout from different options.
- Will have to chose many more times than there are options (arms).

Marketing Application – Selecting Ad Copy

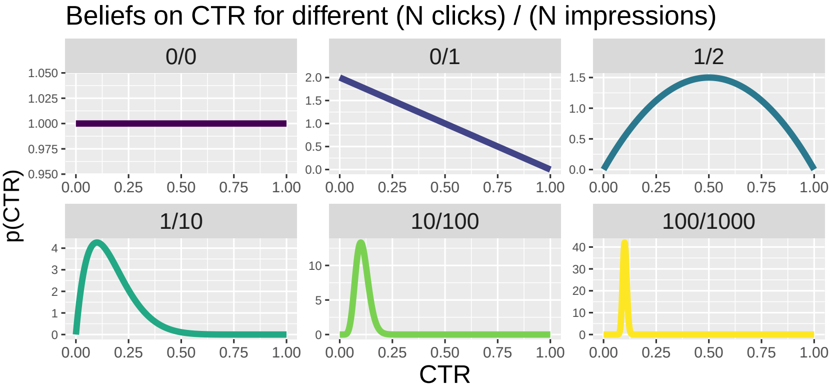
- Which ad copy should we use?
- The one with the highest click through rate (CTR)?
- When are we certain that one of them has the highest CTR?
  - 1 click? 100 clicks? 10,000 clicks?

Quantifying our Beliefs about the CTR

- To answer the questions on the previous slide we need to quantify how much we know about the CTR of the different ads based on our experiments.
- Bayesian statistics offers a helpful and intuitive approach to this problem.
  - Specifies our "beliefs" about the CTR for each ad type.
  - Allows us to update our beliefs each time a new ad is shown.
  - Allows our beliefs to become more precise as more data is acquired.

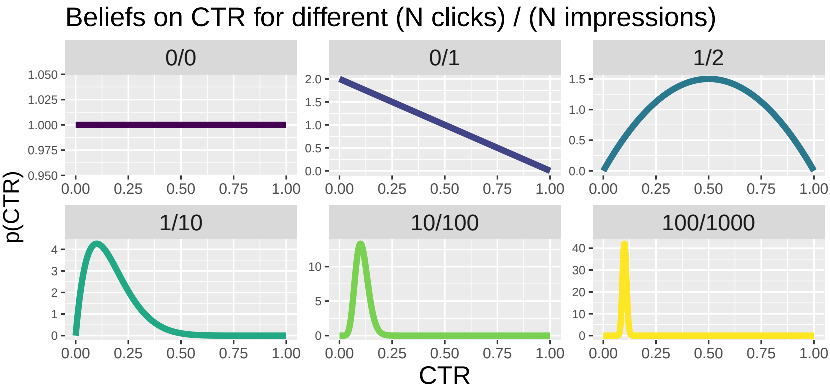
Quantifying Our Beliefs on CTR

- For CTR the value must lie somewhere on [0,1].
- The beta distribution is convenient.
  - Beta has two shape parameters: α and β.
  - Set $\alpha = N_{\text{clicked impressions}} + 1$ and $\beta = N_{\text{not clicked impressions}} + 1$
  - This describes beliefs on the value of the true CTR.



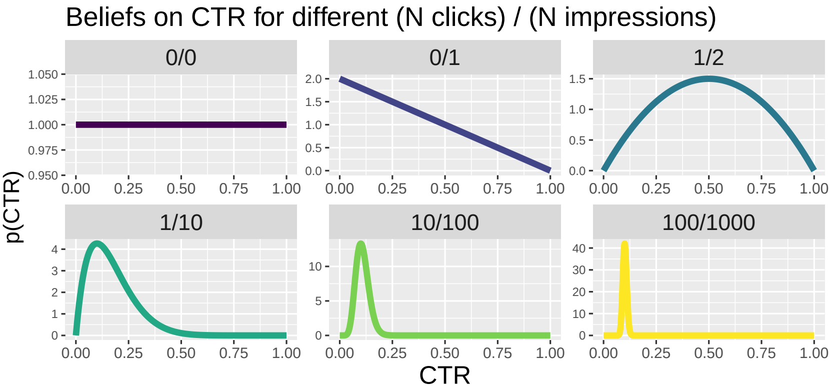
Quantifying Our Beliefs on CTR

- At 0/0 beliefs are "flat." We have no data and therefore no idea what the CTR is. We give the same weight to every value on [0,1].
- At 0/1 we have just one impression delivered but no clicks. This gives a higher likelihood of lower CTRs, but without much nuance.
- At 1/2 we have one click out of two impressions. Our best guess for CTR is 0.5, but it could be many values around there.



Quantifying Our Beliefs on CTR

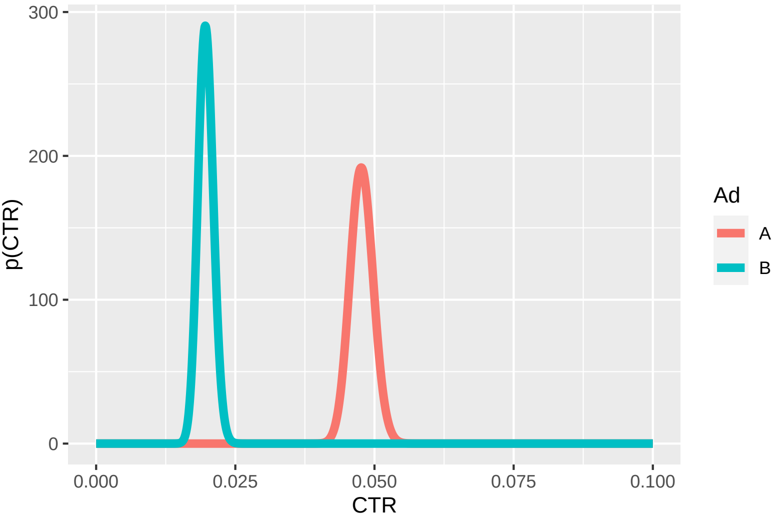
- At 1/10 we have one click out of ten impressions. We are much more confident that the CTR is less than 0.5, with most of the mass around 0.1.
- At 10/100 and 100/1000 we are increasingly confident that the CTR is close to 0.1.



Poll Questions

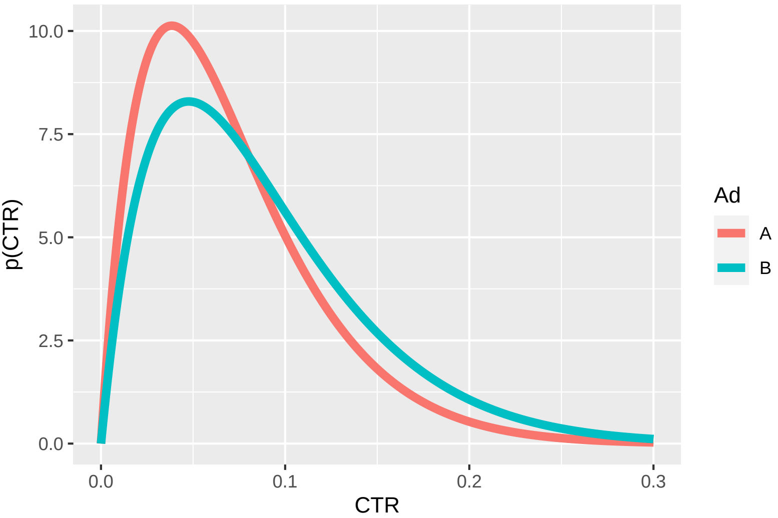
Go to www.PollEv.com/mthomas

Compare Results from Ads



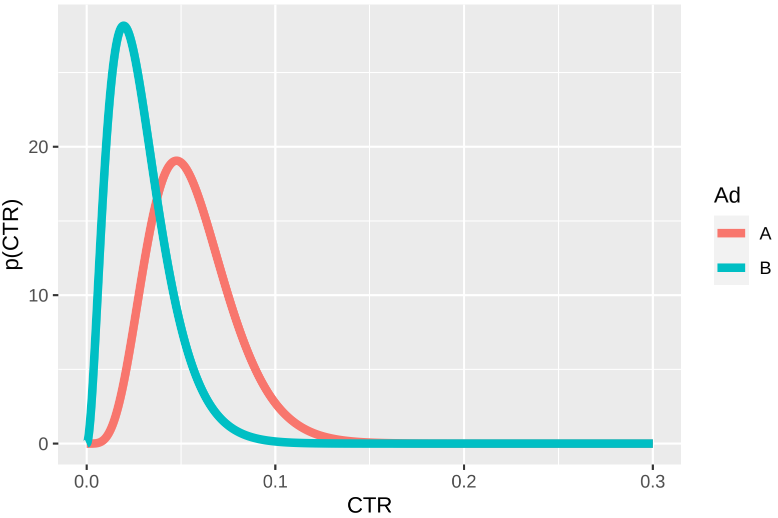
Explore or exploit?

Compare Results from Ads



Explore or exploit?

Compare Results from Ads

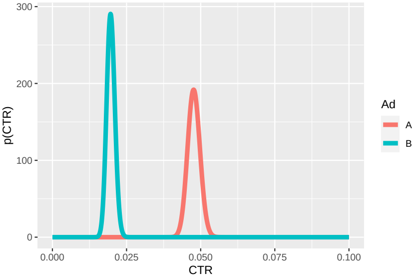


Explore or exploit?

Thompson Sampling

- Simple method for balancing explore and exploit:
  1. Randomly draw a sample from each distribution.
  2. Whichever sample is larger, show that ad next.
- Intuition:
  - When the distributions are similar then each ad has similar probability of being show next. This allows us to collect more information.
  - When the distributions are different, the one with the more favorable distribution is more likely to be chosen.

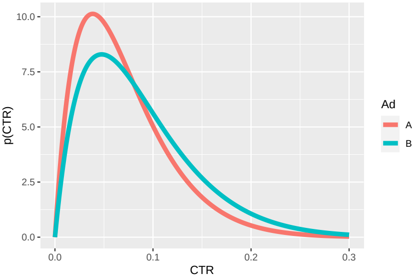
Practice Problem



Given these beliefs about the CTR of two ads, should the firm

- Ans: Mostly explore
- Mostly exploit
- Neither
- About equal amounts of explore and exploit

Practice Problem



Given these beliefs about the CTR of two ads, should the firm

- Mostly explore
- Ans: Mostly exploit
- Neither
- About equal amounts of explore and exploit

Practice Problem

Multi-armed Bandit problems have each of the following characteristics, except:

- It requires a sequence of decisions.
- It requires many options to chose from at each decision point.
- Ans: It requires that we are certain about the outcome associated with each decision.
- It involves an experimental phase.

Practice Problem

Which of the following is *not* true about Thompson Sampling?

- It relies on beliefs about the CTR of different ads
- Ans: It runs faster when there are more ads to compare
- It is framed using Bayesian statistics
- It tends to find the optimal ad if given enough time.