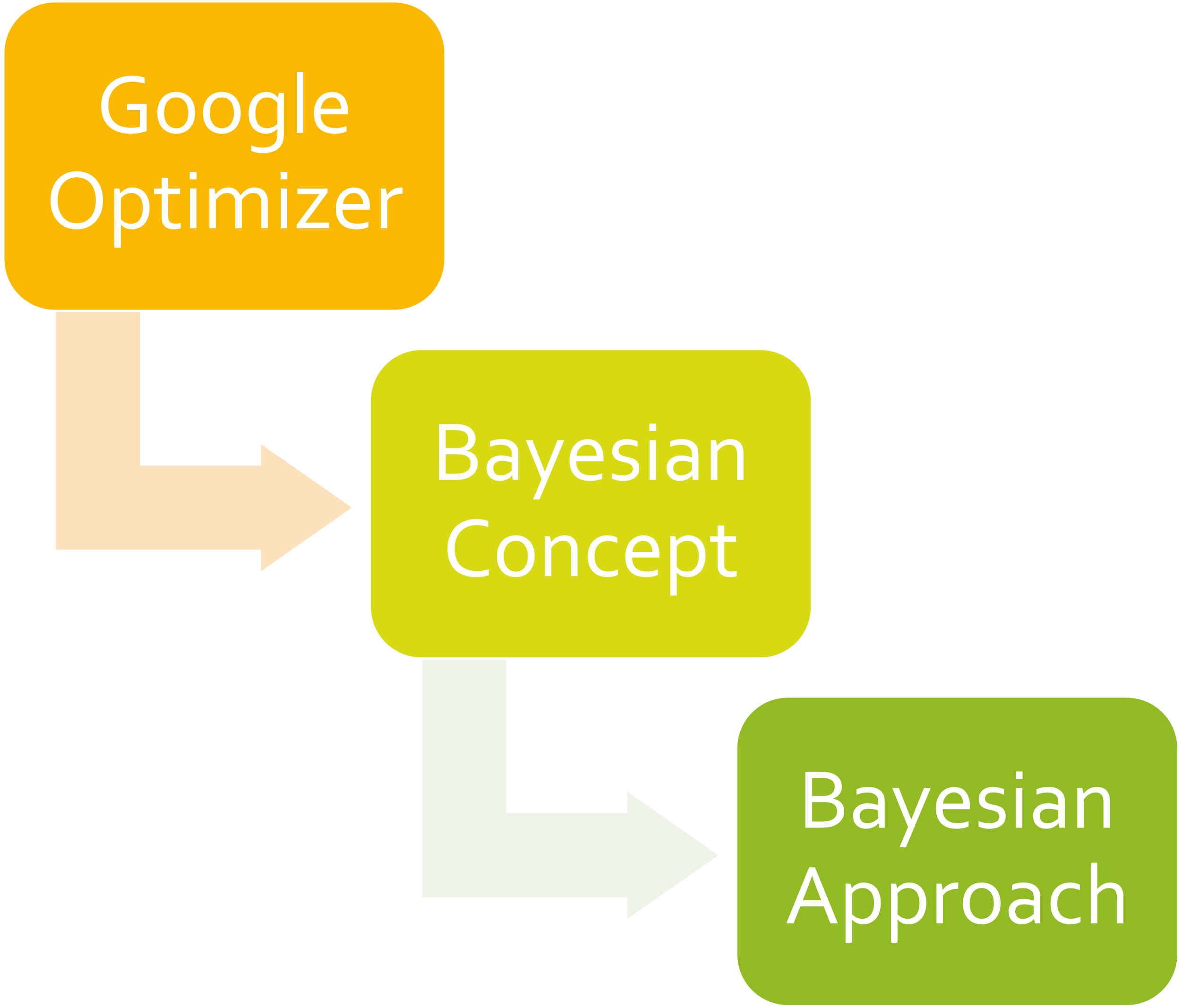
ECON7960 User Experience and A/B Test

Hong Kong Baptist University

Topic 10: Brandt Problem and its Algorithms

In topic 9, we have done



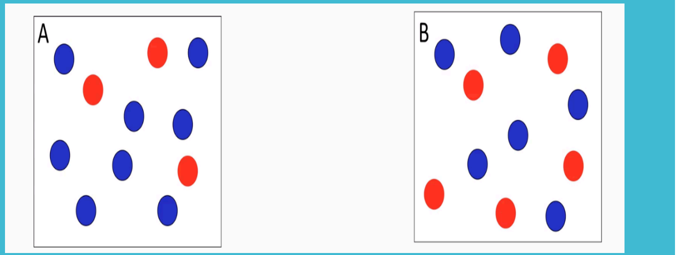
Google Optimizer

Bayesian Concept

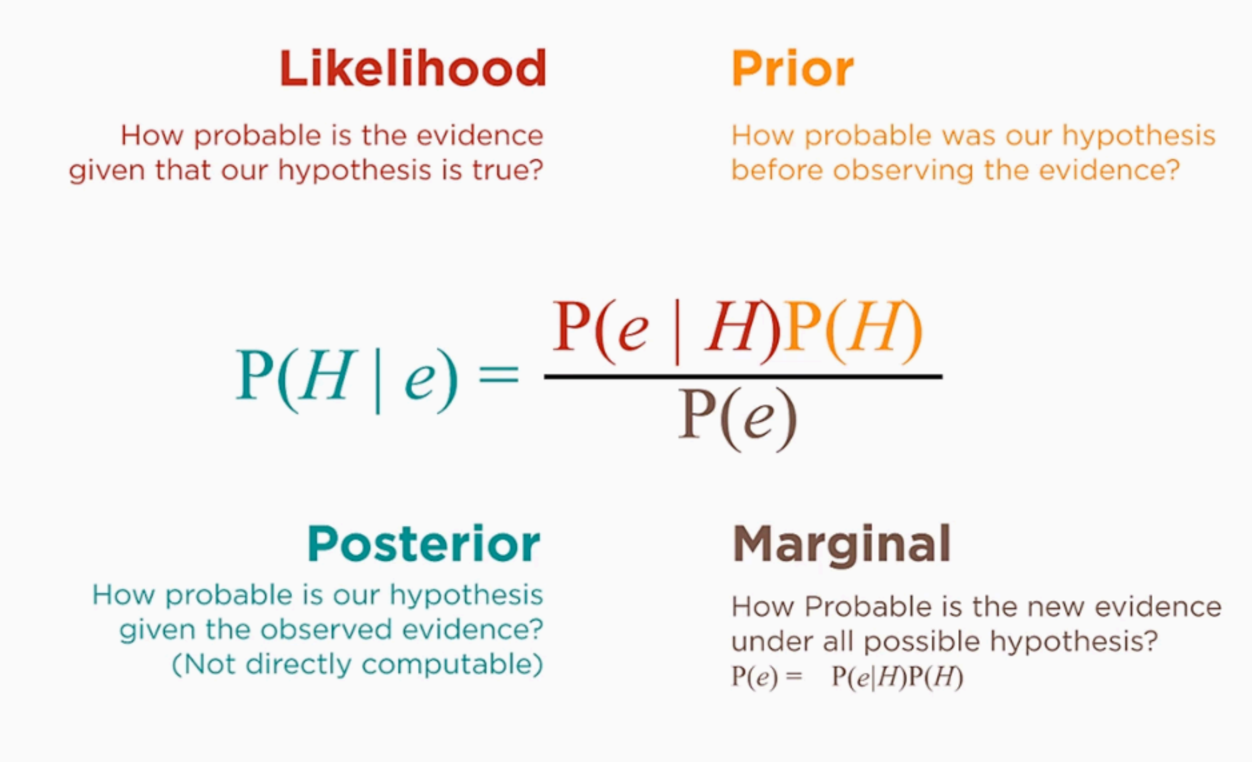
Bayesian Approach

Bayes Approach

Two groups of customers with persona A and persona B:
RED (converted)
BLUE (not converted)



- Using experiment outcomes "e" to predict thing (group A or B) is always a Bayesian concept, based on conditional probability, example: In finance, rate the risk of lending money and in medical field, determine the accuracy of medial test



Likelihood

How probable is the evidence given that our hypothesis is true?

Prior

How probable was our hypothesis before observing the evidence?

$$P(H \mid e) = \frac{P(e \mid H)P(H)}{P(e)}$$

Posterior

How probable is our hypothesis given the observed evidence? (Not directly computable)

Marginal

How Probable is the new evidence under all possible hypothesis?
$P(e) = \quad P(e|H)P(H)$

A Bayesian Problem

- Consider an example. Your company have two group of customers, YOUNG and ELDERLY. YOUNG contains 75 NOT LIKE YOUR PRODUCT and 25 LIKE YOUR PRODUCT, while ELDERLY contains 50 LIKE YOUR PRODUCT and 50 DISLIKE YOUR PRODUCT.
- They randomly visit your web site and you know they from one of these groups.
- Assume out of 10, only 1 convert.
- What are the probabilities for the hypotheses " H 1 : this sample is from YOUNG " and " H 2 :this sample is from ELDERLY," respectively?
- Before they randomly draw come to your site, both belief that these two groups are equally likely. After you observe the result with 10% convert, one needs to update the probability for both hypotheses according to Bayes' formula.
- Consider hypothesis , the probability from YOUTH group.

Calculation:

This result also makes sense intuitively. The probability for drawing non convert from YOUTH is twice as high as for the convert event happening with ELDERLY. Therefore, having drawn 9 non-convert visitor out of 10, the hypothesis YOUNG has with an updated probability higher than the updated probability for hypothesis ELDERLY.

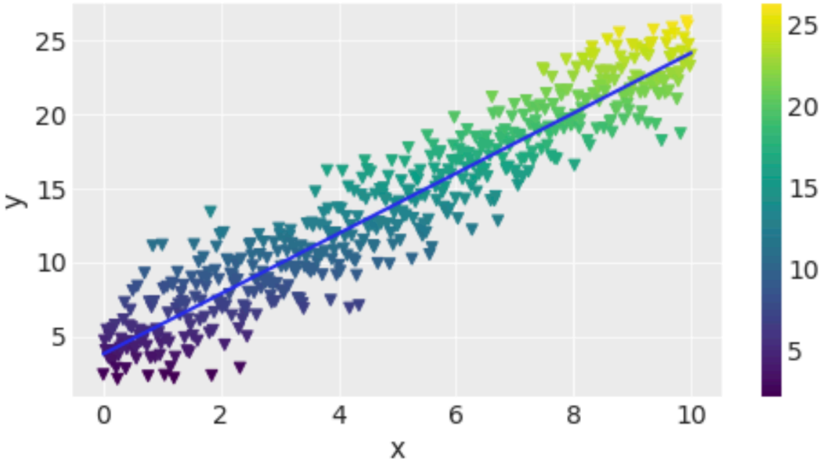
- Probability of experiment outcome D: p (1 convert, 9 not convert)
- Prior : p (YOUTH) = 0.5
- Likelihood : p (D | YOUTH) = 10 x (0.25) x $(0.75)^9$ = .188
- What is p(D)?
  - p(D) = p (D | YOUTH) p(YOUTH)+ p (D | ELDERLY) p(ELDERLY))
  - = 10 x (0.25) x $(0.75)^9$ x .5 + 10 x (.5) x$(0.5)^9$ x 0.5
  - =.094 + .005
  - =.099
- This gives for the updated probability of web visitor are
  - p(YOUTH|D) = {p(D|YOUTH)x p(YOUTH)}/p(D)

    = .5 x .188/.099

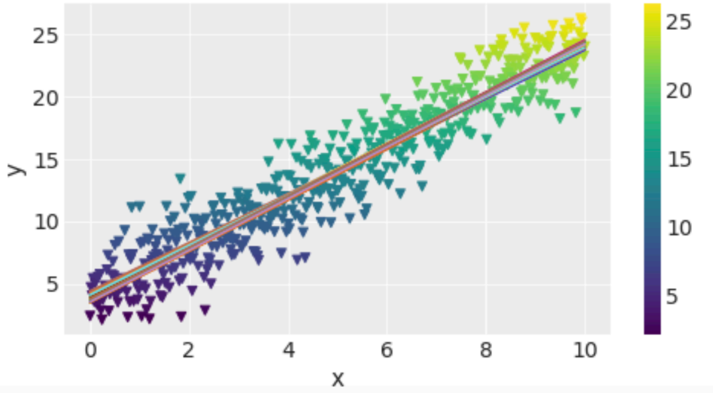
    =.95
- Concluded that this batch of visitors has this observation outcome of conversion that are not half YOUNG and half ELDERLY. It is 95% of YOUTH.

LAB2: Bayesian Regression

Classical Regression



Bayesian Regression

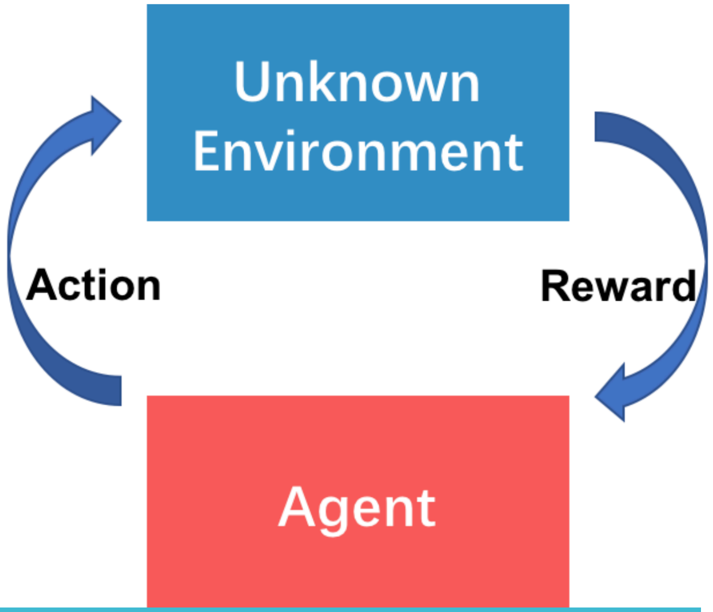


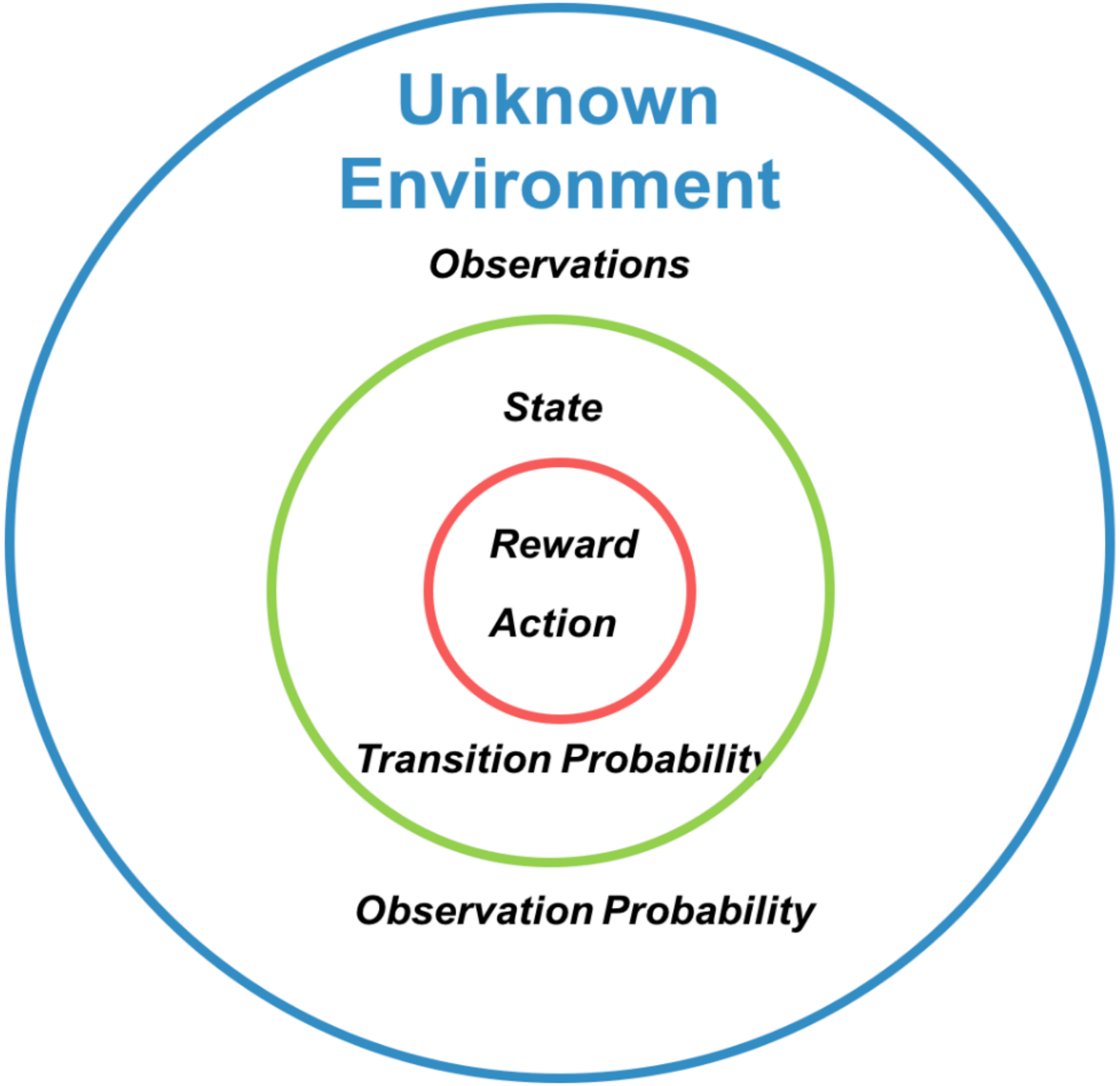
In topic 10, we will do,

Introducing Bandits Algorithms

- All these topics are related about how find a right redesign to achieve the "best" user experience
- In real time, it is a learning processes. But in principle, it is not machine learning or deep learning
- Four scenarios when reasoning with uncertainty

	Actions don't change the state of the world	Actions change state of the world
Learn Model of Outcomes	Multi-armed bandits	Reinforcement Learning
Given Model of Stochastic Outcomes	Decision theory (most classical economics)	Markov Decision Process (MDP)





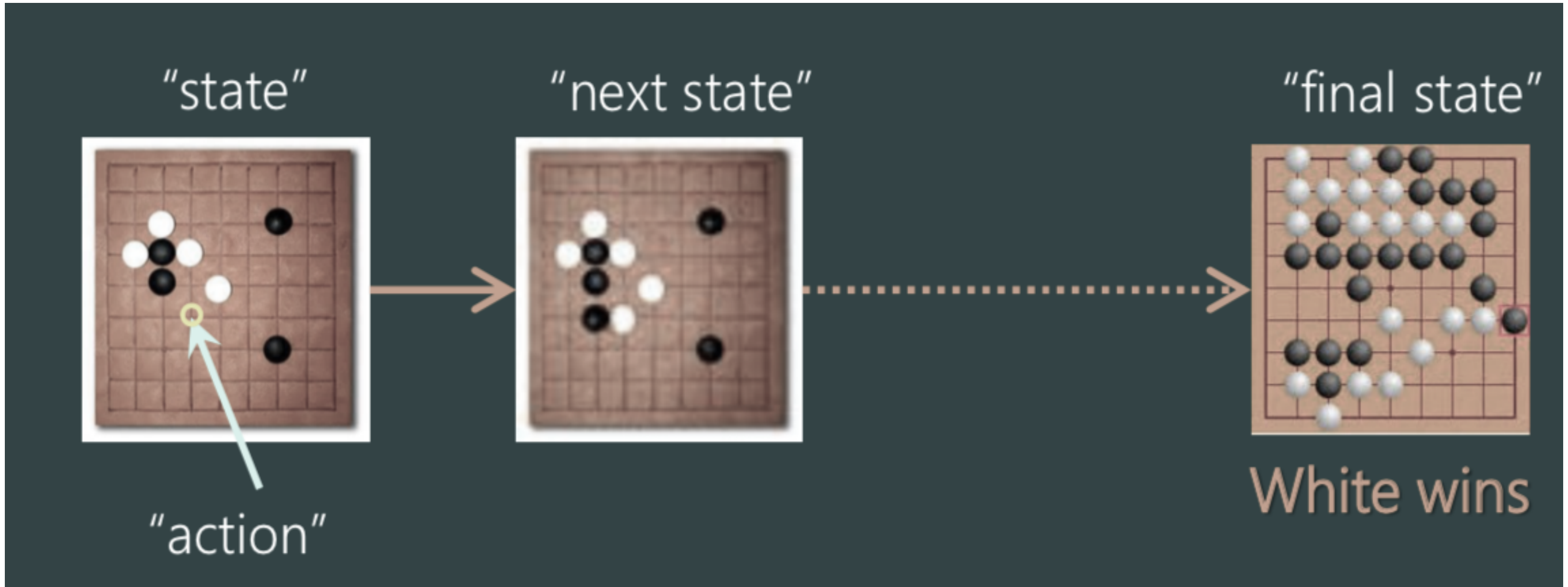
- Bandit Problem
- MDP
- POMDP
  Partially observable Markov Decision Process

What is a Bandits problem?

- Bandits Framework: understanding the concept of "regret"
- "Explore-exploit" algorithms
- "Contextual bandits"
- Implementing and extending bandit algorithms

Learning from the outcome of the previous learning experience, called the reinforcement learning

- In reinforcement learning, there are 4 fundamental questions
  - How to represent the process of this kind of learning?
  - Is the process problem specific or can it be generalized?
  - Aspects of the credit-assignment problem having to do with determining when the behaviour in the learning process that deserves credit?
  - When to start exploration again?



- https://www.ted.com/talks/jeff_hawkins_how_brain_science_will_change_computing/transcript?language=en#t-128578

Representation

- "The key to artificial intelligence has always been the representation" – Jeff Hawkins

Generalization

- What ability to behave well in hitherto unseen states or to behave at the lowest cost in a changing environment

Credit Assignment

- The reinforcement signal that the RL-agent receives is a numerical reward, which encodes the success of an action's outcome, and the agent seeks to learn to select actions that maximizes the accumulated reward over time.

The Concept of Regret: Making the Choice too earlier

In real world, many candidate variations one to try.

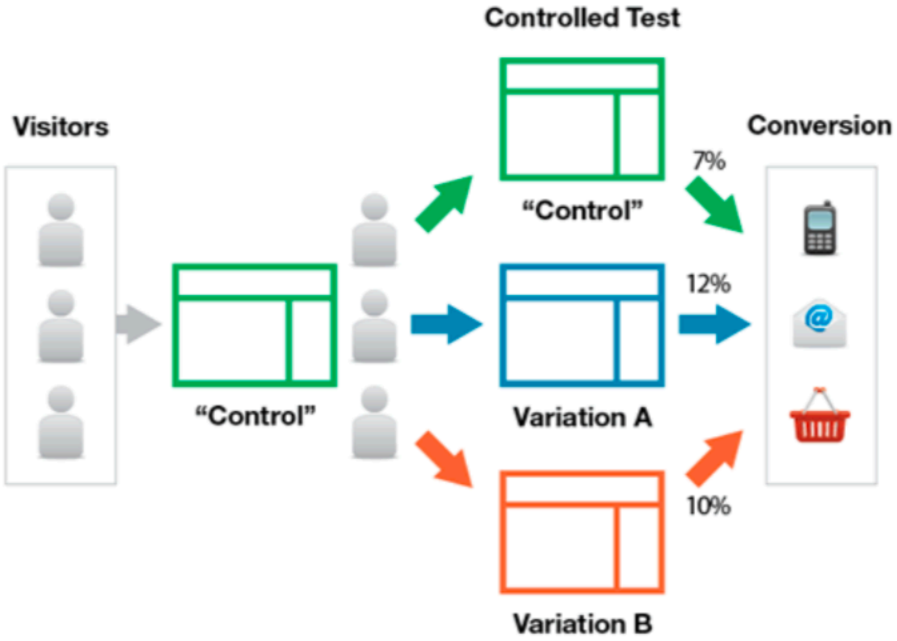
To deliver a quick win, settle on some variants as the candidate, but you are not sure

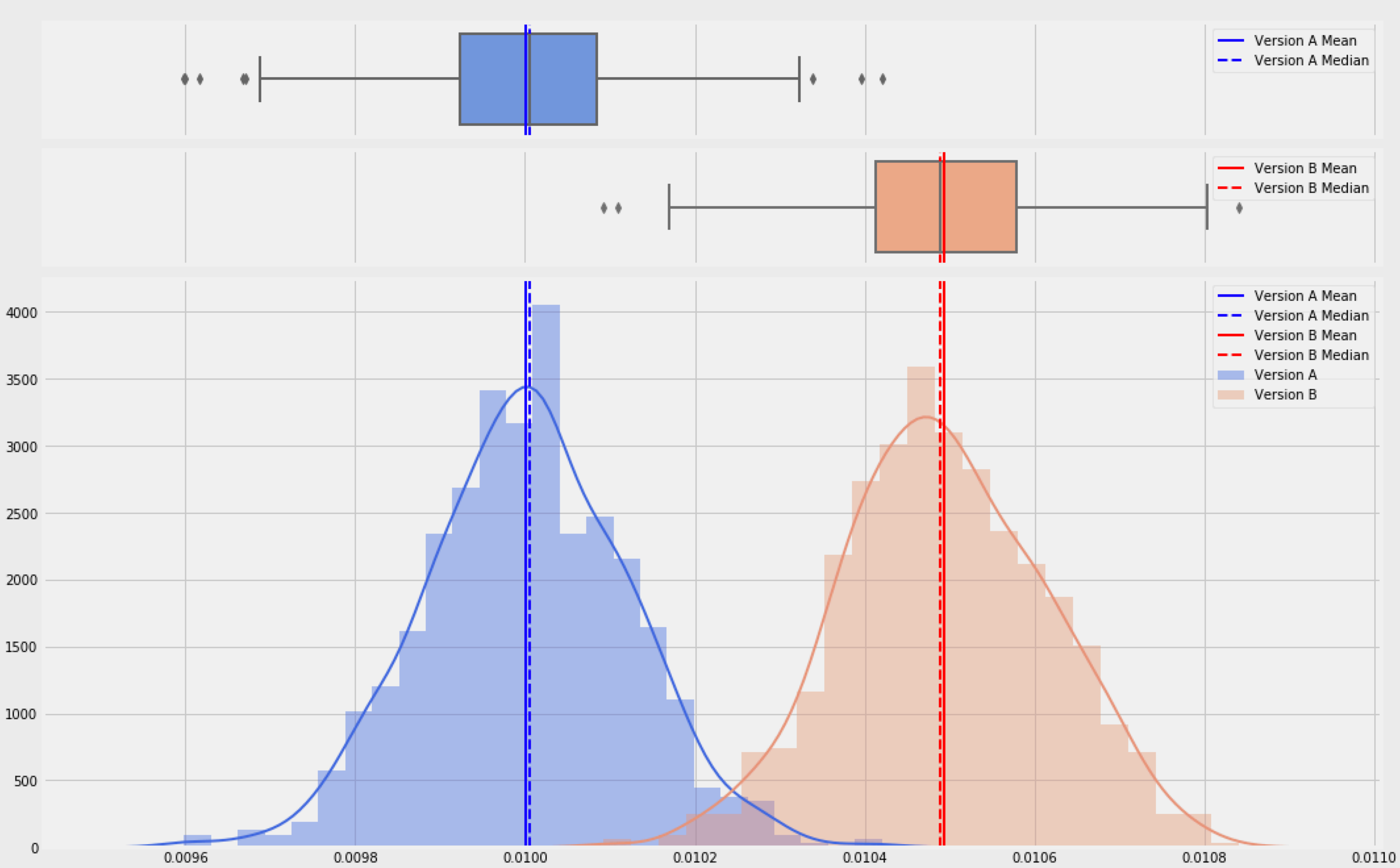
Will the positive result of your chosen A/B/n test leads to temporal benefits? But ultimately goes to suboptimal

Case Study: Dollar Shave Club

https://youtu.be/zNcLAKmFffY

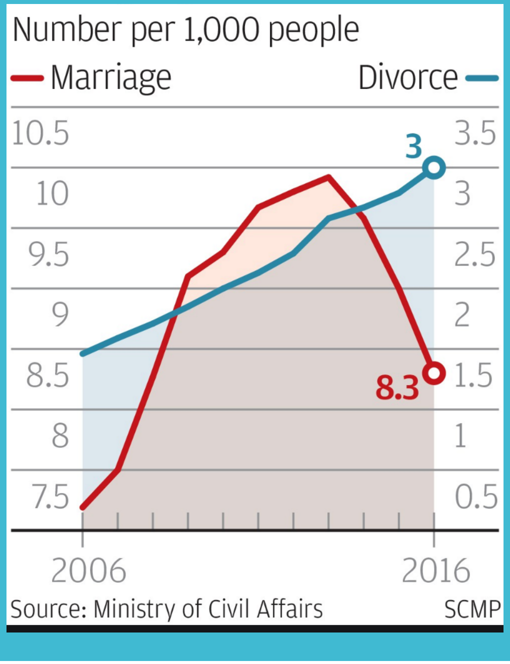
https://www.dollarshaveclub.com/





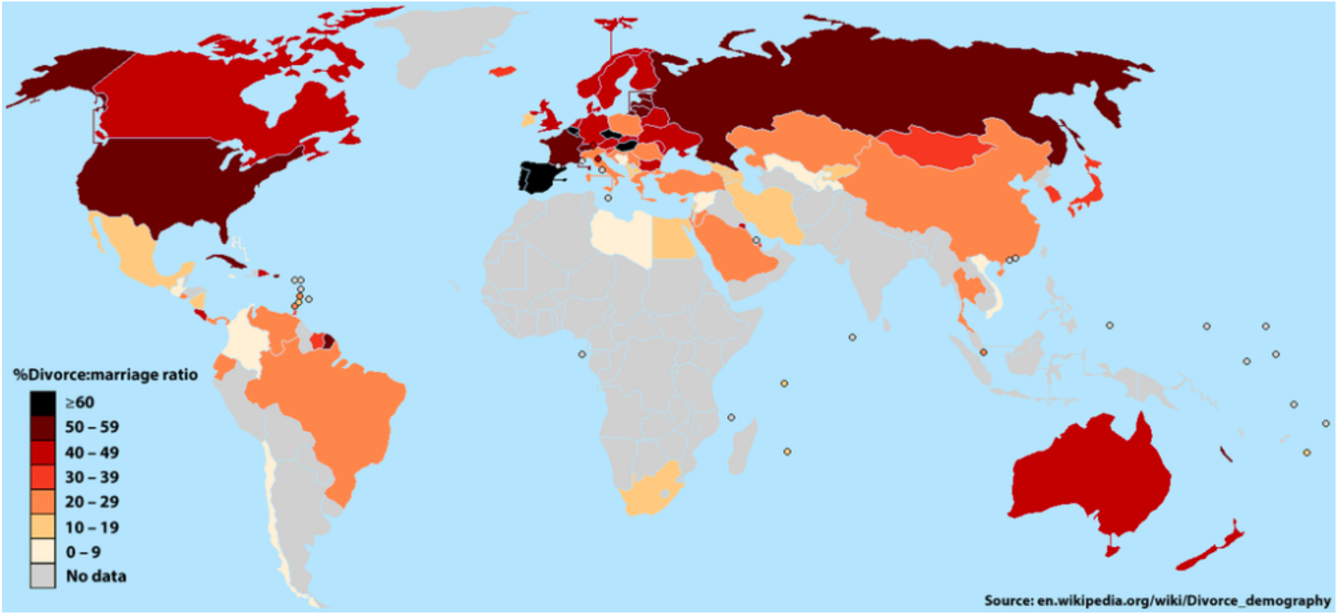
Data Distribution of 1000 Bootstrapped Means

The Concept of Regret: Making the Choice too earlier.



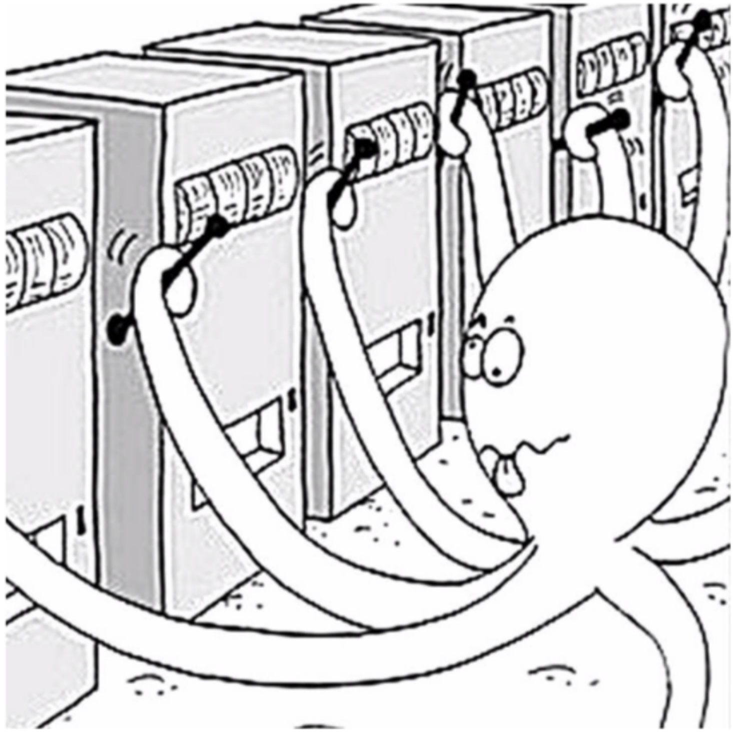
Source: Ministry of Civil Affairs SCMP

- Is there an action we have not yet tried that could lead to an overall better outcome?
- Do we too earlier to settle to a suboptimal choice?
- If the decision cannot be reversed and number of trial is limited, the sequence of actions matters.



Source: en.wikipedia.org/wiki/Divorce_demography

The Concept of Regret: Making the Choice too earlier.



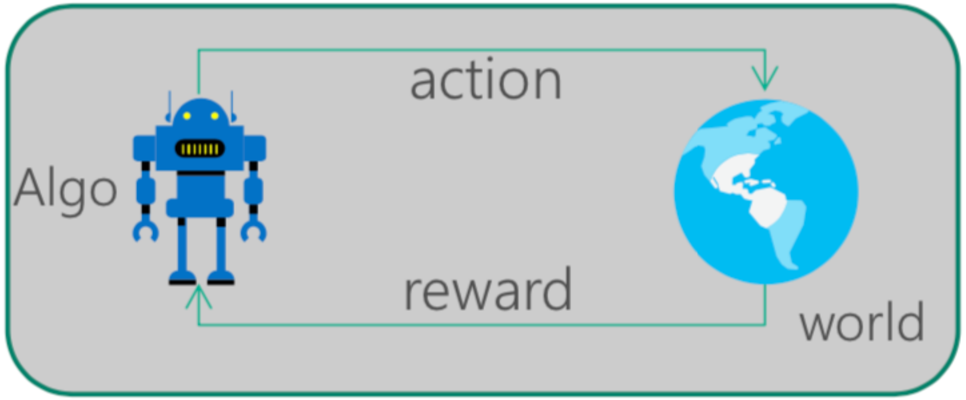
Slot machine Problem:

Bandit Framework

- Multi-armed bandit problem:
- Stochastic bandits: Problem formulation then to algorithms
- Adversarial bandits: Problem formulation then to algorithms
- Contextual bandits

Sequential Decision w/ Incomplete Information

- Exploration-Exploitation Dilemma
  - Exploration: Gather information
  - Exploitation: Optimal decision using current information
- Fundamental tradeoff between exploration and exploitation



Balancing Exploration and Exploitation

- Exploration:
  - Try out each action/option to find the best one, gather more information for long term benefit
- Exploitation:
  - Take the best action/option believed to give the best reward/payoff, get the maximum immediate reward given current information.
- Questions:
  - Select a restaurant for dinner
  - Medical treatment in clinical trials
  - Online advertisement
  - Oil drilling

The Naïve Algorithm

- A/B/n Testing

- Assign T/K patients to each action a for each arm i , $1 \leq i \leq K$.
- Implement: e.g. round-robin (In sports competitive matches per season, is usually double round-robins. Most association football leagues in the world are organized on a double round-robin basis, in which every team plays all others in its league once at home and once away.) through available actions
- *Implement the round-robin algorithm A/B tests*

- Can we do better to maximize $\Sigma_t r_t$

Multi-armed bandit problem

- Rewards $r_i(t)$ at each arm *i* are drawn i.i.d, with an expectation/mean u_i, unknown to the agent/gambler
- $r_i(t)$ is a bounded real-valued reward.
- Goal :
    - maximize the return(the accumulative reward.)
    - or
    - minimize the expected regret:
    - Regret = $u^* T - \Sigma^T_{t=1} E[r_{it}(t)]$, where
    - $u^* = \max_i[u_i]$, expectation from the best action

Regret, thought experiment to quantify “price of information”

- Suppose we know all reward distributions p(r|a)
- Optimal policy is to always play a* = argmax E[r|a] where a is the feasible action

Regret: $L_T = T\boldsymbol{E}[r|a^*] - \sum_t \boldsymbol{E}[r|a_t]$

Maximize $\sum_t r_t$ $\equiv$ Minimize regret L_T

Regret Minimization Principle

Optimism in the face of uncertainty

*To achieve low regret, we only need to identify an optimal arm a**

Good algorithm should not play sub-optimal arms too often...

So:

- *Use collected data to eliminate arms that "very likely" are sub-optimal*
- *However, choose the most optimistic remaining an good option*

Multi-armed bandit problem: Algorithm

- Stochastic bandits:
- Example: 10-armed bandits
- Question: what is your strategy? Which arm to pull at each time step t?

- Greedy method:
  - At time step t, estimate a value for each action
  - $Q_t(a)$=$sum\ of\ rewards\ when\ a\ taken\ prior\ to\ t$
    $number\ of\ times\ a\ taken\ prior\ to\ t$
  - Select the action with the maximum value. A_t= argmax$_a$ Qt(a)
  - Weaknesses of the greedy method
    - Always exploit current knowledge, no exploration
    - Can stuck with a suboptimal action

Algorithm 1: The ϵ- Greedy Algorithm

- Consider algorithm estimate $\hat{r}_a \approx E(r|a)$
- Optimistic-Greedy: Initialize $\hat{r}_a$ to a large initial value Q
- Then play Greedy algorithm
  - ϵ-Greedy:
    With probability ϵ, pick a uniformly random action
    With probability 1 – ϵ, play Greedy algorithm
- Question: How should we set ϵ?

Experiments LAB4

- Set up: (one run)
- 10-armed bandits
- Draw U_i from Gaussian(0,1), i = 1,…,10
  - the expectation/mean of rewards for action i
  - Rewards of action i at time t: $x_i(t)$
  - $x_i(t) \sim$ Gaussian(u_i, 1)
  - Play 5000 trials
    Average return at each time step

Observations

- Greedy method improved faster at the very beginning, but level off at a lower level.
  ε- Greedy methods continue to Explore and eventually perform better.
- Theε = 0.1 method improves slowly, but eventually performs better than the ε = 0.2 or 0.3 method.

Observations

- Big initial Q values force the Greedy method to explore more in the beginning.
- No exploration afterwards.
- Improve ε-greedy algorithm with decreasing ε over time. Improves faster in the beginning, also outperforms fixed ε-greedy methods in the long run.

Weaknesses of ε-Greedy methods: With probability ε, select an action randomly from all the actions with equal probability.

- Randomly selects an action to explore, does not explore more “promising” actions.
- Does not consider confidence interval. If an action has been taken many times, no need to explore it.