

A large, faint, light brown network graph serves as the background for the text. It consists of a grid of nodes connected by various lines, creating a complex web of paths and intersections. The text is overlaid on this grid.

Data  
Intelligence  
Applications

# PRICING in E-COMMERCE

## E-Commerce Data

Electronic commerce is the main economic scenario where the dimension of pricing can lead to significant increase in revenue

→ in last 5 years the sales due to electronic-commerce increased a lot

## ECONOMIC PRELIMINARIES

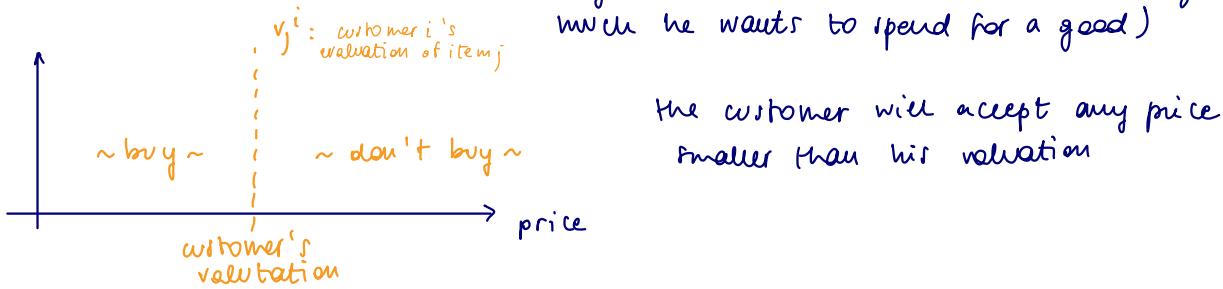
### Basic scenario

A single seller has to sell infinite units of a single good to multiple customers

- SELLER MODEL:
  - the seller has a cost over the good

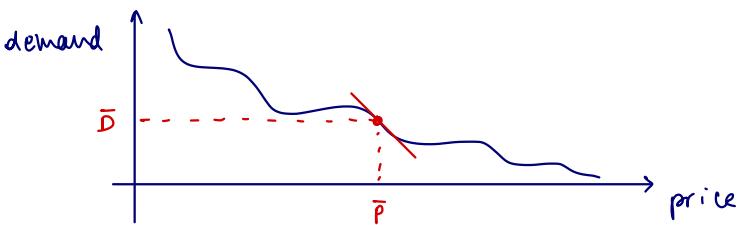
- all units are sold at same price

- CUSTOMER MODEL:
  - every customer has a valuation over the good (says how much he wants to spend for a good)



### demand curve

If we have many customers we construct the demand curve → different evaluation



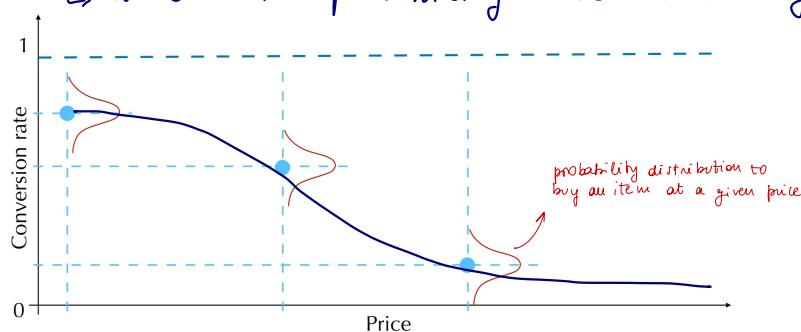
demand = number of customers that would buy an item at a given price

demand < 1 : the seller sells same number of items with all higher price

→ it corresponds to the normalization of the derivative of the demand wrt the price (look at the speed of change)

⇒ actually the demand curve is unknown ; we rely on conversion rate

↳ which is the probability to sell an item given a price ?



from conversion rate we build the average demand curve

## ■ Pricing

~ How to choose the price of a good with economical tools ~

- Scenarios

1. Monopoly → only one seller in the market
2. Oligopoly → few sellers in the market
3. Competitive market → many sellers in the market

- Definitions

$p$ : price

$q(p)$ : demand at price  $p$

$c(q(p))$ : cost of  $q(p)$  items for the seller (\*)

$p \cdot q(p)$ : revenue at price  $p$

$p \cdot q(p) - c(q(p))$ : profit at price  $p$

$q$ : demand

$p(q)$ : price of a demand

$c(q)$ : cost of  $q$  items for the seller

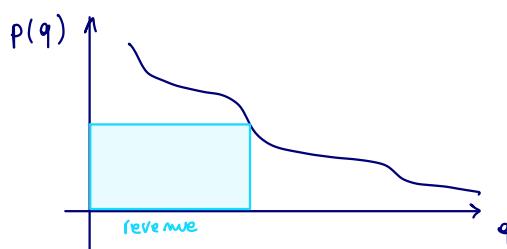
$q \cdot p(q)$ : revenue at demand  $q$

$q \cdot p(q) - c(q)$ : profit at demand  $q$

(\*) The cost of every item is the same.

If the number of sold items is different,  
also the cost of every item is different

⇒ the seller wants to maximize the profit



the price to apply depends on the quantity we want to sell

- profit maximization

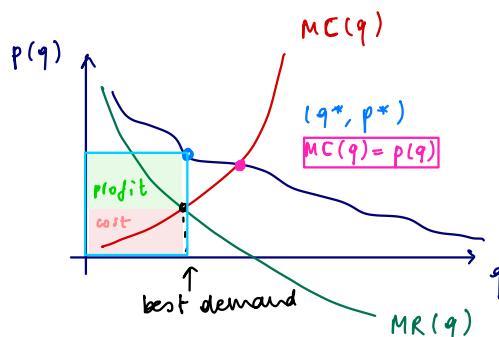
we need to find  $\bar{q} = \underset{q}{\operatorname{argmax}} \{ p(q) \cdot q - c(q) \}$

$$\frac{d}{dq} (p(q) \cdot q - c(q)) = 0 \rightarrow \frac{dp(q)}{dq} \cdot q + p(q) = \frac{dc(q)}{dq}$$

marginal revenue  $MR(q)$ :  
revenue provided by selling  
the last unit of good when  
 $\# \text{items} = q$

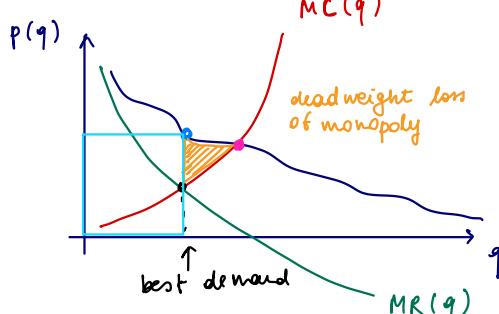
marginal cost  $MC(q)$ :  
cost due to selling the  
last unit of good when  
 $\# \text{items} = q$

best price in the demand curve → the seller sells  $q$  items with  
an higher price  
in order to sell more items (avoid negative profit)  
→ OPTIMAL SOCIAL PRICE



→ deadweight loss of monopoly: inefficiency of the market due to the seller

⇒ the seller is selfish and he doesn't allow  
the market to be efficient



the seller increases his profit wrt the choice of the best demand → make happier more customers

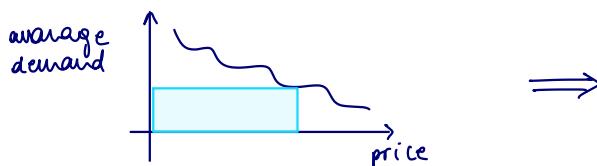
## ☒ Price discrimination

- Simple price inefficiency

- Deadweight loss of monopoly is due to the adoption of a simple pricing strategy  
every item is sold at the same price

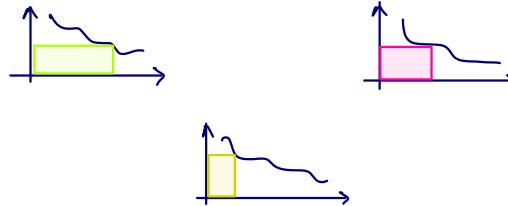
- If the seller can personalize the price for each customer, the social efficiency of the market can be improved

⇒ we build different demand curves according to different customer classes



### AGGREGATE DEMAND:

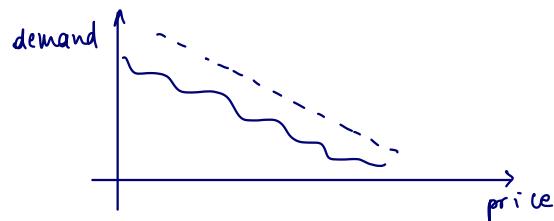
If we don't have information on customers we need to set the same price for each customer



DISAGGREGATE MODELS: If the seller knows the class of the customers, we can separate the problem and apply different prices for every class  
 ⇒ price discrimination increases the profit

## ☒ Single product demand curves

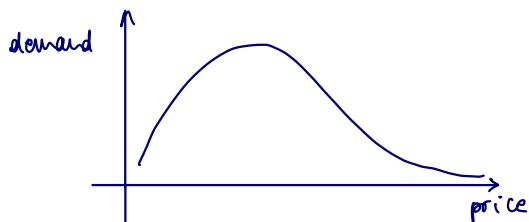
- Negative slope demand curve



the demand monotonically decreases as the price increases

→ It is referred to as the "law of demand": people will buy more items as its price falls

- Non negative slope demand curve



the demand does not monotonically decrease as the price increases  
 → goods for which the quantity demanded increases as the price increases

## ☒ Interdependent product demand curves

~ Correlated Goods ~

A seller has to sell an infinite number of A and B to multiple customers  
 → There is dependency between A and B

- Strategic dependency

### ① Strategic substitutes

idea: good A can substitute good B (and viceversa)

effect: if the price of A reduces then the customers' willingness to pay good B reduces

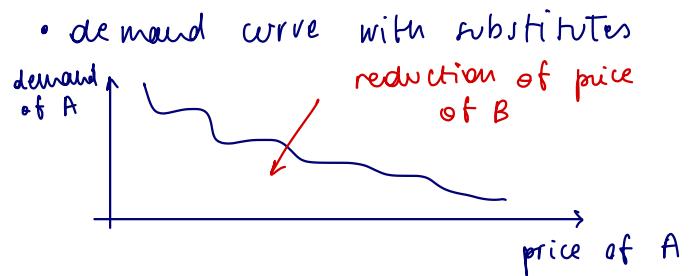
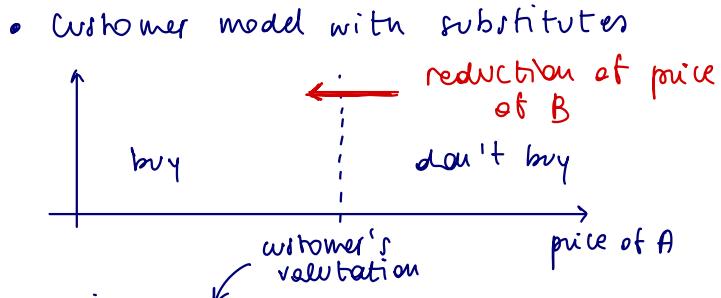
A similar to B  
 $A \uparrow \Leftrightarrow B \uparrow$

### ② Strategic complements

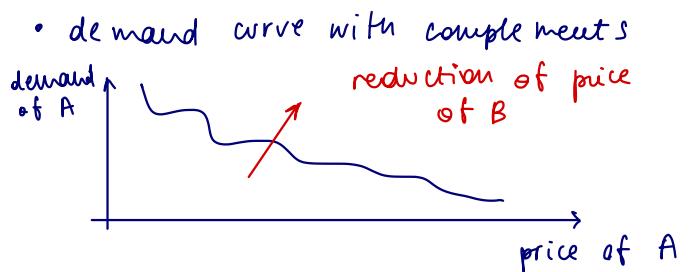
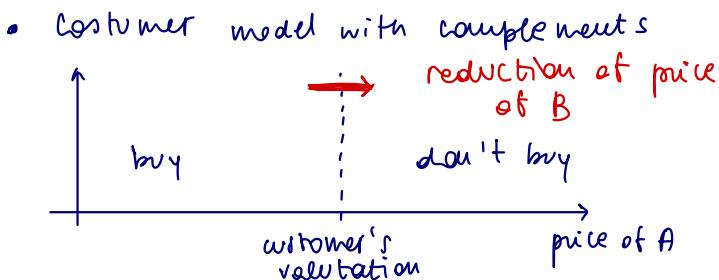
idea: good A can complement good B (and viceversa)

effect: if the price of A reduces, then the customers' willingness to pay good B increases

A and B used together  
 $A \downarrow \Leftrightarrow B \uparrow$   
 {  $A \downarrow$ : sell more items  
 $B \uparrow$ : sell more items }



$v_i^j(p_k)$ : customer  $i$ 's valuation of item  $j$  depending on price of good  $k$



### pricing interdependent goods

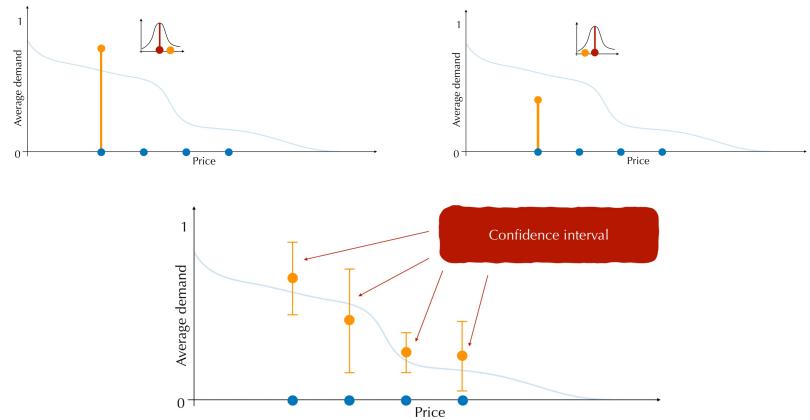
→ two or more goods that are interdependent (due to strategic complements or strategic substitutes) must be priced together in order to maximize the profit

## LEARNING THE DEMAND CURVE

The average expected demand curve in general is unknown  
→ we have to estimate it

### Sampling the demand curve

1. we fix some values of price and we estimate the demand
2. observe the demand drawn from a probability distribution (in general is gaussian)
3. collect samples and calculate the empirical mean (average)
4. Estimate the confidence intervals



### Setting up

given some candidates (prices) we associate an unknown expected reward (profit = price · demand)

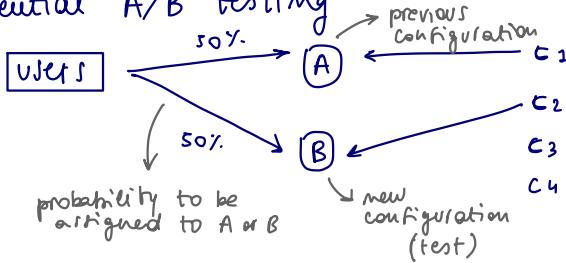
- hp
1. rewards are stationary : the probability distribution of rewards is stationary and the demand curve doesn't change
  2. we can use a single candidate per round
  3. The feedback is not delayed : given a price we can't observe the samples
  4. The number of rounds is bounded : finite time horizon

# INTRODUCTION TO A/B/m TESTING

② Non-technical introduction to A/B/m testing

~ How to learn the best prices ~

- Sequential A/B testing



we want to understand if B is better than A

- Several definitions of reward

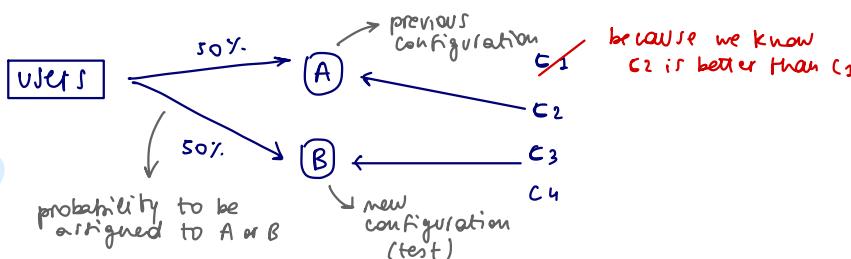
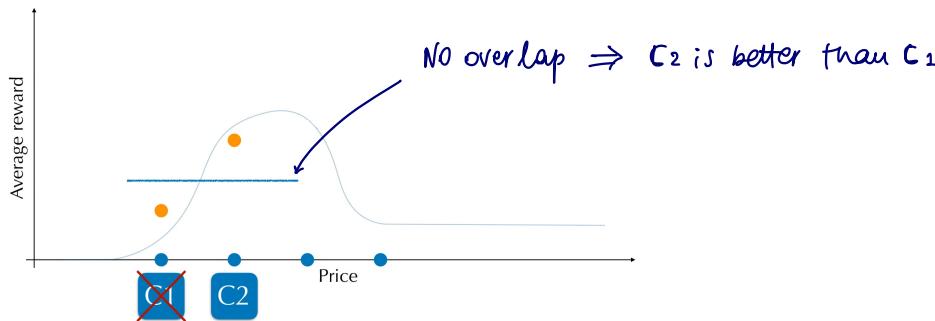
1. profit maximization

- demand curve has to be estimated (thanks to the fact that for every price it behaves as a truncated Gaussian distribution) and it has to be multiplied by the marginal profit
- conversion rate has to be estimated (thanks to the fact that for every price it behaves as a binomial distribution (buy or not)) and it has to be multiplied by the marginal profit

2. volumes maximization (market invasion) → fight with another company

- demand has to be estimated
- conversion rate has to be estimated

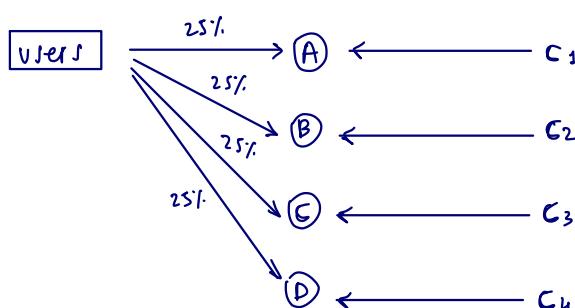
⇒ we collect many samples in order to have a small confidence interval



this method is very slow since we compare all couples

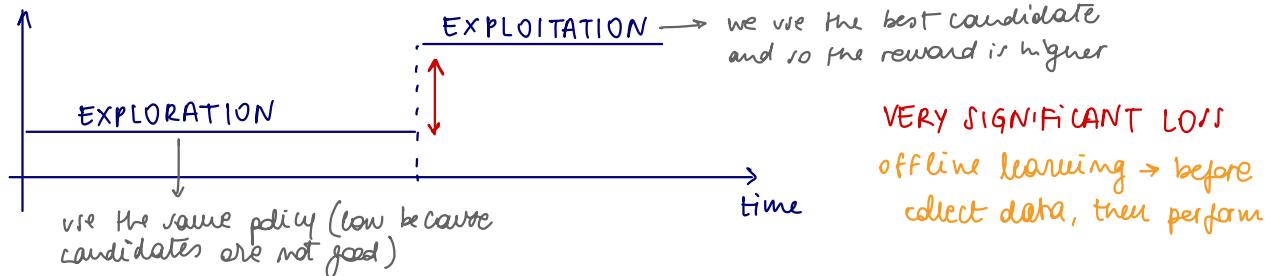
- Sequential A/B/m testing

we decide to evaluate all together



A candidate is chosen from a uniform probability because no prior information and we calculate the expected reward. We do many times for all candidates at the end of the experiment we select the winner

- pure exploration  $\rightarrow$  in the time horizon we consider all candidates with same weight
- pure exploitation  $\rightarrow$  we use only the winner to find the expected reward



$\rightarrow$  A/B/m Testing depends on the length of time horizon

- { LOW: lose a bit of expected reward
- { LONG: lose the ratio half expected reward

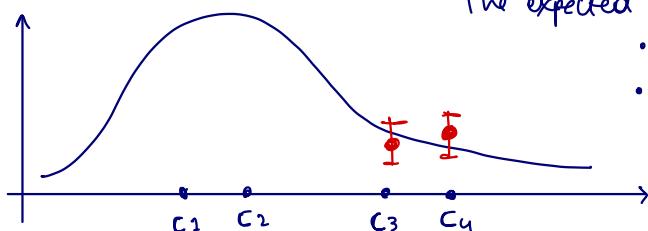
$\Rightarrow$  How long should an A/B/m test be?

- confidence to be guaranteed
- reward to be collected

#### • A/B/m weaknesses

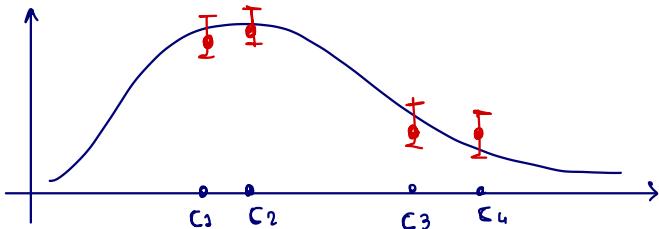
1. assumption of stationary process  $\rightarrow$  probability distribution doesn't change during time, but actually it can be good in future
2. long time to identify the optimal candidate  $\rightarrow$  time horizon is too short wrt the best solution
3. Discarding a potentially optimal candidate  $\rightarrow$  if the required confidence needs a larger number of samples wrt time horizon  $\sim$  high confidence

#### Example (A/B testing)



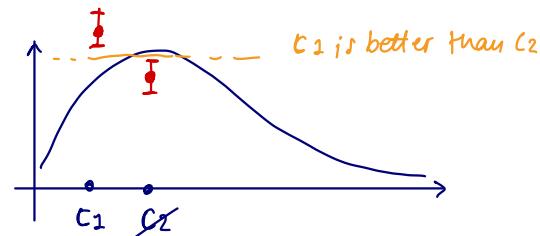
- The expected reward of these two suboptimal solutions is very close
- need a larger number of samples
- loss time to find the best between C3 and C4  
but maybe C1 and C2 are better

#### Example (A/B/m testing)



- This time we test all candidates together
- long time to discriminate C1 vs C2 and C3 vs C4
- we can exclude immediately C3 and C4 but we have to wait the end of the episode

- $\rightarrow$  in order to decide the optimal candidate we rely on small confidence intervals but it may happen that we exclude a better solution
- $\rightarrow$  once a candidate is discarded, it cannot be reconsidered in the future
- $\Rightarrow$  INACCURATE

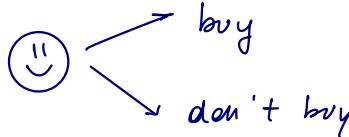
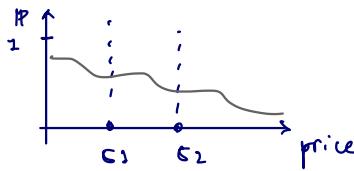


# A/B/m TESTING TECHNICALITIES

## Hypothesis testing

### ① Defining the problem

The data we collect is composed by customers who decide to buy or not an item given a price - we can model this decision with a Bernoulli distribution

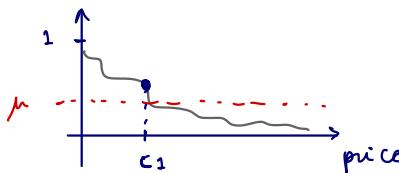


### ② Formulating an hypothesis to verify

#### 1 candidate

$$H_0: \mu_{c_1} = \mu \quad \text{vs} \quad H_1: \mu_{c_1} \neq \mu$$

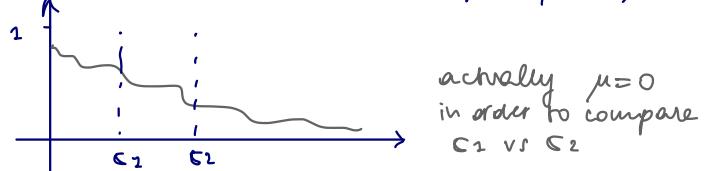
$$H_1: \mu_{c_1} > \mu$$



#### 2 candidates

$$H_0: \mu_{c_1} - \mu_{c_2} = \mu \quad \text{vs} \quad H_1: \mu_{c_1} - \mu_{c_2} \neq \mu$$

$$H_1: \mu_{c_1} - \mu_{c_2} > \mu$$



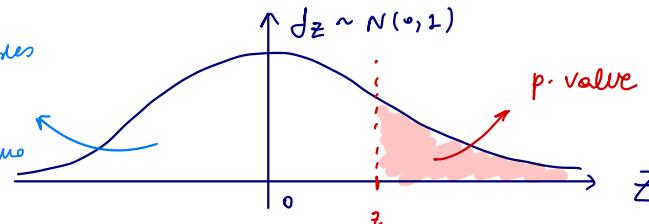
### ③ Selecting a test statistic

→ The test statistic defines the algorithm to use and verify the hypothesis

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\mu_0(1-\mu_0)}{m}}} \rightarrow \begin{array}{l} \text{test statistic} \\ \text{random variable} \\ \bar{X} \text{ sample mean rv} \end{array}$$

$$\xrightarrow{\text{data}} Z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\mu_0(1-\mu_0)}{m}}} \rightarrow \begin{array}{l} \text{test statistic} \\ \bar{x} \text{ empirical mean} \end{array}$$

1-p-value:  
probability that samples  
are drawn from a  
distribution whose  
mean is bigger than  $\mu_0$



$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\bar{Y}(1-\bar{Y})}{m_1 + m_2}}}$$

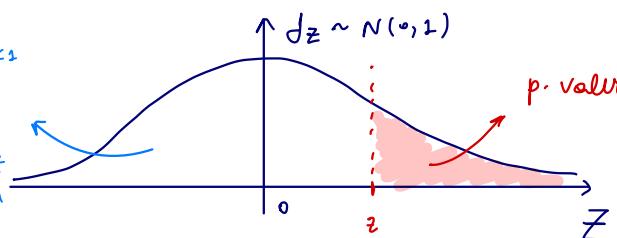
$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\bar{y}(1-\bar{y})}{m_1 + m_2}}}$$

$$Y = \frac{m_1 \bar{X}_1 + m_2 \bar{X}_2}{m_1 + m_2}$$

1-p-value:  
probability that samples  $x_1$   
are drawn from a  
distribution whose mean  
is bigger than the mean of  
the distribution from which  
samples  $x_2$  are drawn

$\xrightarrow{\text{data}}$

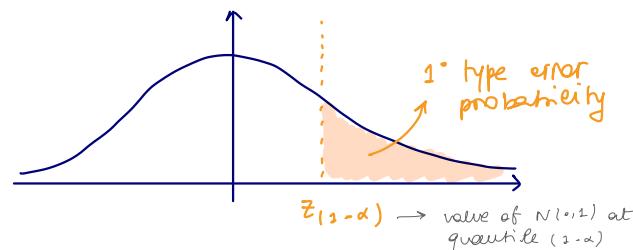
$$y = \frac{m_1 \bar{x}_1 + m_2 \bar{x}_2}{m_1 + m_2}$$



### ④ Selecting the accuracy of the test statistic (before collecting data)

$$\alpha = \text{P}(1^{\circ}\text{ type error})$$

reject  $H_0$  when  
it's true



⑤ Performing random samples  
Fix the number of samples, but how many?

	$H_0$ is true	$H_1$ is true
Not rejecting $H_0$	$1 - \alpha$	$\beta$
Rejecting $H_0$	$\alpha$	$1 - \beta$

We want to collect a small number of samples but at the same time it has to be enough to find the winner.

⇒ The minimum number of samples to collect for A/B testing is:

$$H_0: \mu_{\text{ctrl}} = 0, \gamma$$

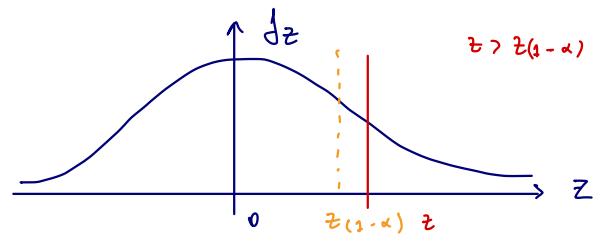
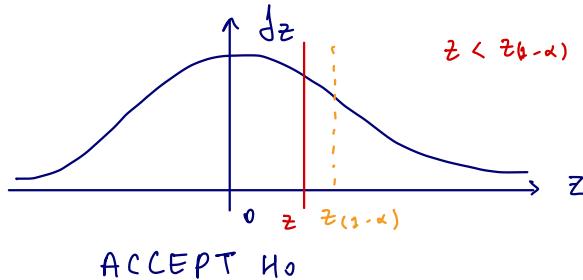
$$H_1: \mu_{\text{ctrl}} \geq 0, \gamma + \delta$$

$$n = \frac{(z_{(1-\alpha)} + z_{(\beta)})^2 \sigma^2}{\delta^2}$$

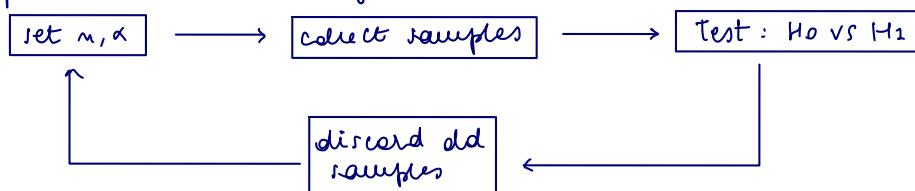
$\sigma \rightarrow$  standard deviation

$\delta \rightarrow$  minimal variation in the process

⑥ Calculating the test statistic on the given samples and Taking the decision



⇒ we repeat the test many times



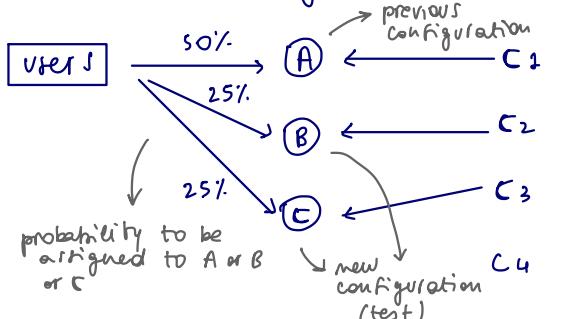
Rein A/B testing is not an iterative tool

- we have to collect all data before taking the decision

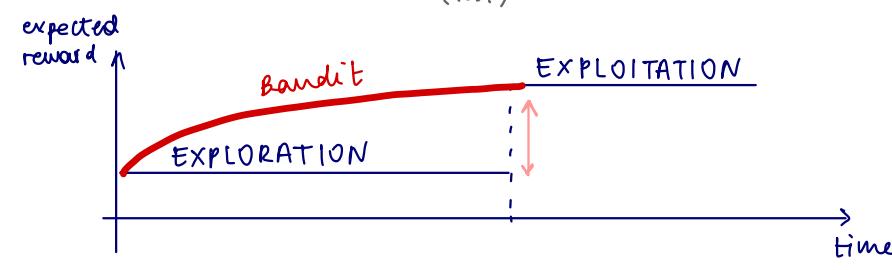
- Once we decide, the old samples are discarded and not used anymore

## INTRODUCTION TO BANDIT ALGORITHMS

• from A/B/m testing to Bandit Algorithm



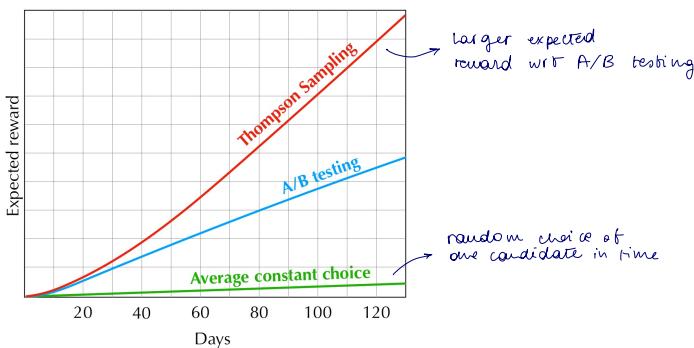
Bandit Algorithms don't use a uniform distribution to assign users to A, B, C → the probability changes wrt observations we collect



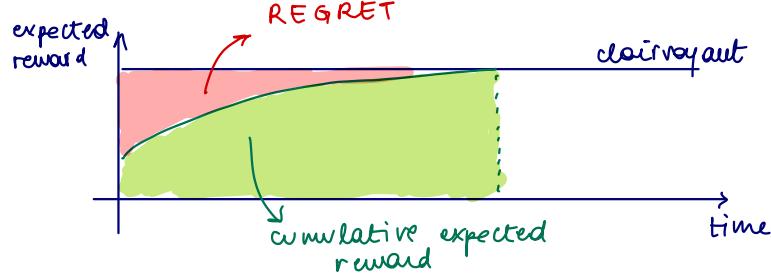
we perform better during the exploitation phase because we tune the probability wrt observations we get

online learning → use collected data while performing

## Scenarios: Pricing for E-commerce



## Regret Minimization



regret: what we're losing due to the fact we're learning (loss of reward)

CLAIRVOYANT ALGORITHM:

a priori the expected reward of all candidates is known and at each step the best candidate is chosen

~ perfect algorithm, not realizable ~

→ since a zero-regret algorithm doesn't exist (except clairvoyant), we want to provide a lower and upper bound to the regret  
 ⇒ BANDIT ALGORITHMS

obs the minimization of the regret corresponds to the maximization of the reward  
 regret  $\downarrow \Rightarrow$  reward  $\uparrow$

## Bandit Algorithm

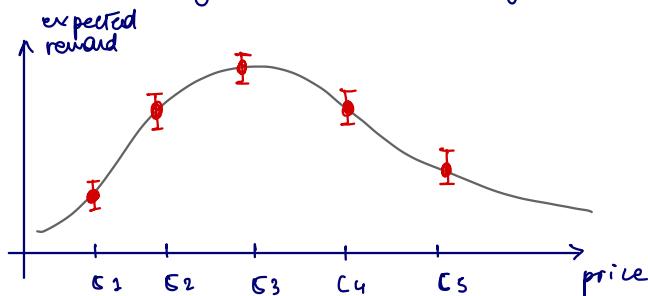


## Offline vs Online learning

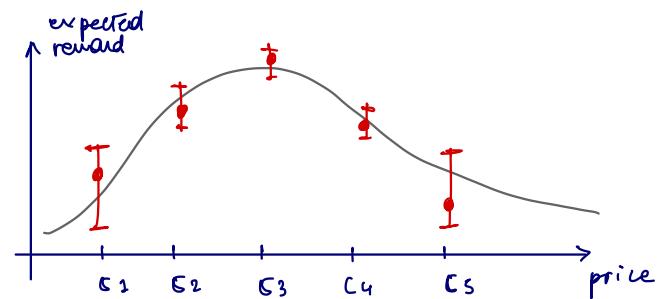
→ A/B/m testing is an offline learning → collect all data and then decide which candidate to play

→ Bandit Algorithm is an online learning → decide at every step using the previous data

## A/B/m Testing vs Bandit Algorithms



A/B/m Testing does an accurate estimation of all candidates



Bandit algorithms do an accurate estimation only for good candidates

# BANDIT TECHNICALITIES

## VB1

- every arm is associated with an upper confidence bound  
→ optimistic estimation of the reward provided by an arm
- the arm with the highest upper confidence bound is chosen at each step
- update of the rewards → update of the upper confidence bounds

frequentist and deterministic approach

hyp : 1. the rv of every arm is a bernoulli with unknown mean  
2. every arm has a different mean  
3. The model captures the scenario in which one needs to learn a conversion rate  
4. reward = conversion rate - net margin (unk)

## algorithm

1° step : play once every arm  $a \in A$

2° step : at every time  $t$  play arm  $a_t$ :

$$a_t = \operatorname{argmax}_{a \in A} \left\{ \bar{x}_a + \sqrt{\frac{2 \log t}{m_a(t-1)}} \right\}$$

UCB (uncertainty)

$a^*$ : optimal arm

$a_t$ : arm  $a$  chosen at time  $t$

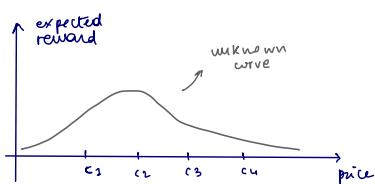
$X_a$ : rv of reward of arm  $a$

$x_a$ : realization of  $X_a$

$\bar{x}_a$ : empirical mean of  $x_a$

$m_a(t)$ : number of times we chose arm  $a$  from 0 to  $t$

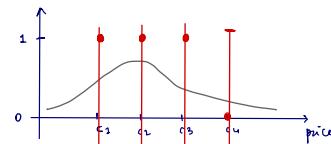
## Example



1° step : we pull  $c_1 \rightarrow x_{c1} = 1$     we pull  $c_2 \rightarrow x_{c2} = 1$   
we pull  $c_3 \rightarrow x_{c3} = 1$     we pull  $c_4 \rightarrow x_{c4} = 0$

2° step : compute the empirical mean  $x_a = \bar{x}_a$   
and compute UCB

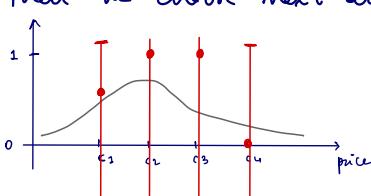
$$\rightarrow \text{choose next } a_t = \operatorname{argmax} \left\{ \bar{x}_a + \sqrt{\frac{2 \log t}{m_a(t-1)}} \right\}$$



② we can choose among  $c_1, c_2, c_3$

we pull  $c_1 \rightarrow x_{c1} = 0$

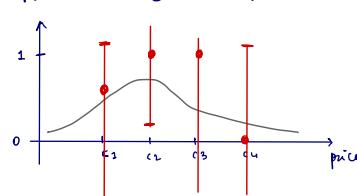
then we choose next  $a_t$



③ we can choose among  $c_2, c_3$

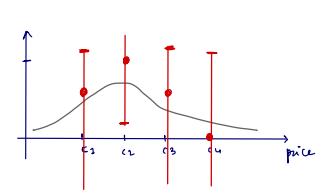
we pull  $c_2 \rightarrow x_{c2} = 1$ ,

then we choose next  $a_t = c_3$

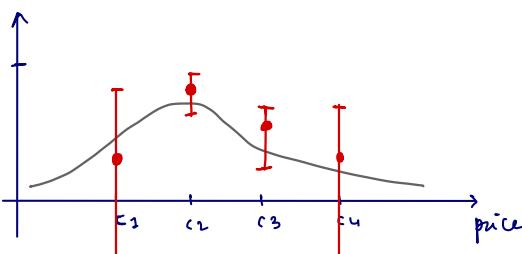


④ we pull  $c_3 \rightarrow x_{c3} = 0$

then we choose  $c_2$



⇒ after many samples



we have different confidence intervals

## Regret

$$\text{Def (regret)} \quad R_T(V) = T \mu_{a^*} - \sum_{t=1}^T \mu_{a_t}$$

expected regret  
after  $T$  steps

expected reward  
of clairvoyant

expected regret of  
algorithm  $V$

Def (VLCB1 Regret)

$$Q_T(\text{VLCB1}) \leq \sum_{a: \mu_a < \mu^*} \frac{4 \log T}{\mu^* - \mu_a} + 8(\mu^* - \mu_a)$$

- it increases logarithmically as  $T$  increases
- it increases linearly as  $|A|$  increases

## Thompson Sampling

- For every arm, we have a prior on its expected value  
→ if arms' rewards are Bernoulli then priors are Beta
- For every arm we draw a sample according to Beta distribution
- the arm with the best sample is chosen at every step
- we update the Beta distribution according to the observation
 
$$\begin{cases} \text{observe 1} & \text{Beta}(\alpha, \beta) \rightarrow \text{Beta}(\alpha+1, \beta) \\ \text{observe 0} & \text{Beta}(\alpha, \beta) \rightarrow \text{Beta}(\alpha, \beta+1) \end{cases}$$

Bayesian and randomized approach

### algorithm

1° Step : start with  $\alpha=1, \beta=1$

2° Step : at every time  $t$ , for any arm

$\vartheta_a = \text{Sample } (\mathbb{P}(\mu_a = \vartheta_a)) \text{ from } \text{Beta}(\alpha, \beta)$

3° Step : at every time  $t$  play arm  $a_t$  such that

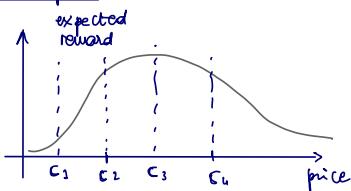
$$a_t = \underset{a \in A}{\operatorname{argmax}} \{ \vartheta_a \}$$

4° Step : update Beta distribution of arm  $a_t$

$$(\alpha_{a_t}, \beta_{a_t}) = (\alpha_{a_t}, \beta_{a_t}) + (x_{a_t, t}, 1 - x_{a_t, t})$$

$x_{a, t}$  realization of  $X_a$  at time  $t$   
 $\mathbb{P}(\mu_a = \vartheta_a)$  : prior on  $\mathbb{E}[X_a]$   
 $\mathbb{P}(\mu_a = \vartheta_a) = \text{Beta}(\alpha_a, \beta_a)$

### Example



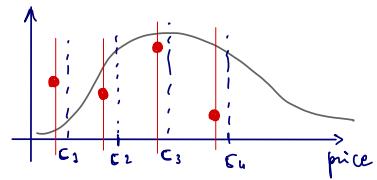
at the beginning  $\alpha=1, \beta=1$   
 No prior information :  $\text{Beta}(1, 1) = \text{Uniform}$  !

① Sample all candidates from Beta

then choose the arm with the highest reward

→  $c_3$  has the highest expected reward  $\Rightarrow x_{c_3, t=1}$

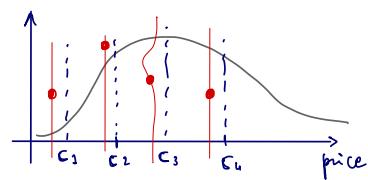
→ update  $c_3$  beta distributions accordingly



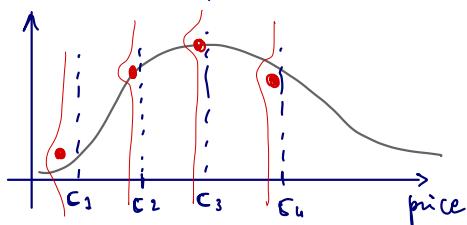
② Sample all candidates from Beta and choose the best

→  $c_2$  has the highest expected reward  $\Rightarrow x_{c_2, t=0}$

→ update  $c_2$  beta distributions accordingly



⇒ after many samples



we can see that almost the probability is close to the expected value wrt the most promising candidates (because they were pulled more)

Def (TS Regret)

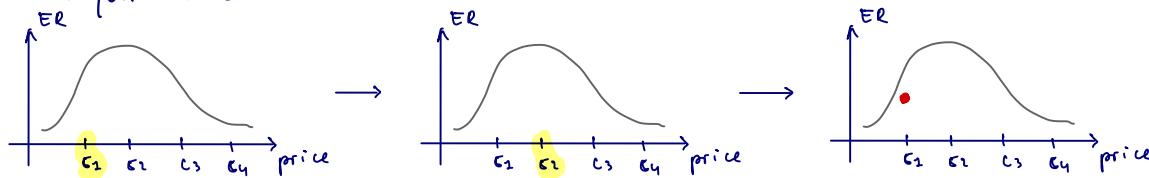
$$R_T(\text{TS}) \leq (1+\epsilon) \sum_{a: \mu_a < \mu^*} \frac{(\mu^* - \mu_a)(\log T + \log(\log T))}{K \geq (\mu^*, \mu_a)} + C(\epsilon, \mu_1, \dots, \mu_A)$$

- it increases logarithmically as  $T$  increases
- it increases linearly as  $|A|$  increases

# DELAYED FEEDBACK

Bandit algorithm are very fast but the realization needed for the decisions are observed after some time

## delayed - UCB1



we pull  $c_1$ , then we pull  $c_2$  and we observe the realization of  $c_1$

### algorithm

1° step : play once every arm  $a \in A$

2° step : at every time  $t$  play arm at such that

$$a_t = \underset{a: \hat{n}(t-1) > 0}{\operatorname{argmax}} \left\{ \bar{x}_a + \frac{2 \log T}{\sqrt{\hat{n}(t-1)}} \right\}$$

### regret

$\hat{n}(t)$ : number of feedbacks for arm  $a$  at time  $t$

$\bar{G}_{a,T} = \max_{t \leq T} G_{a,t}$  max number of missing feedbacks for all arm  $a$   $[0, t]$

$$R_T(\text{UCB1}) \leq \sum_{a: \mu_a^* < \mu_a} \left[ \frac{8 \log T}{\mu_a^* - \mu_a} + 3.5 (\mu_a^* - \mu_a) \right] + \sum_a (\mu_a^* - \mu_a) \mathbb{E} [\bar{G}_{a,T}]$$

- it increases logarithmically as  $T$  increases
- it increases linearly as  $|A|$  increases
- $\Rightarrow$  it increases linearly as the delay increases

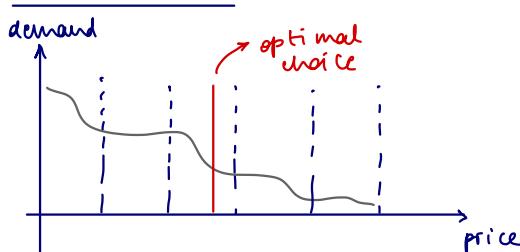
## delayed - TS

It has the same idea of UCB1, but it's not discussed in literature.  
we expect that the regret increases linearly as the delay increases

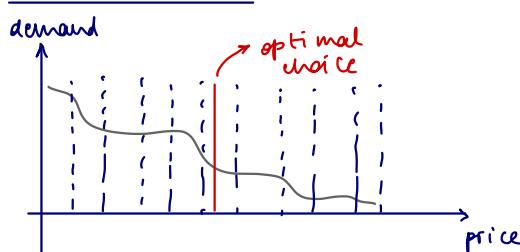
# HOW MANY CANDIDATES?

We need different candidates in order to do a discretization of the price space

### 5 candidates



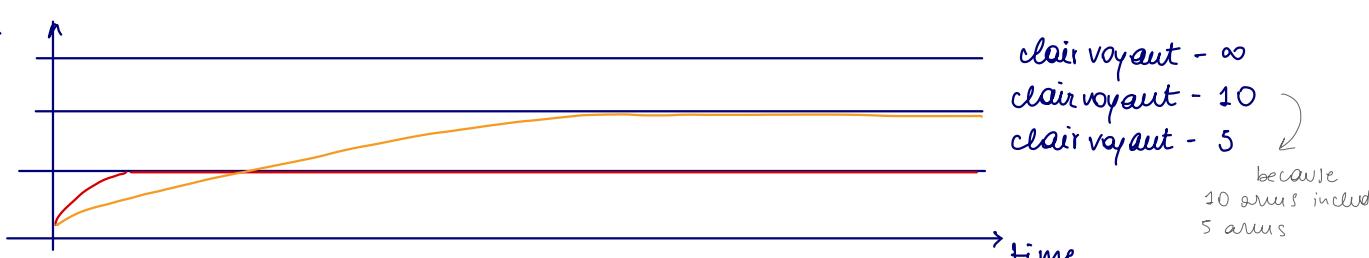
### 10 candidates



### • Expected reward

When we evaluate the regret we have to evaluate the expected reward achieved by the clairvoyant algorithm (the infinite case and the finite case)

### expected reward



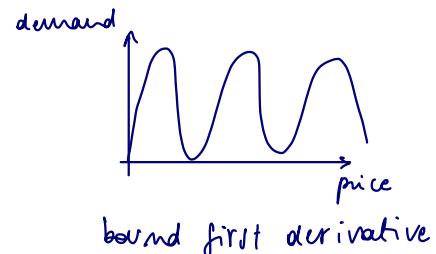
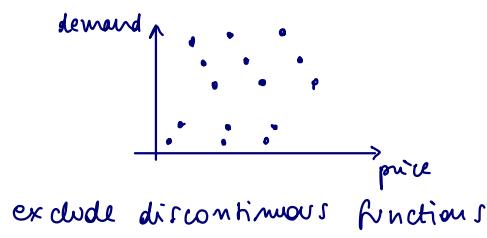
The scenario with more arms reaches a higher expected reward after a long time because we need to collect more data

- if we can collect few data we need a small number of arms
- if we can collect many samples we need an higher number of arms
- ⇒ The optimal number of candidates depends on the length of the horizon over which we want to maximize the expected reward

### • Discretization

In finite time we can only discretize the space of arms, so we need some assumptions to bound the cumulative regret

→ the reward function needs to have some regularities



→ when the second derivative of the reward is  $< 0$ , then the regret is minimized  
The number of candidates should be:

$$\text{start from this number of arms} \quad k = \lceil (T \log T)^{1/4} \rceil \quad \longrightarrow \quad Q_T(\text{UCB1}) \leq \sqrt{T \log T}$$

## NON STATIONARY ENVIRONMENTS

Now we want to consider the change of rewards' probability distribution during time.

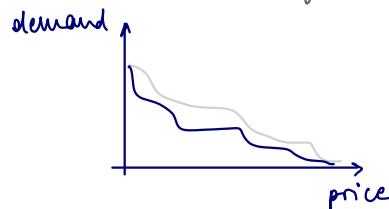
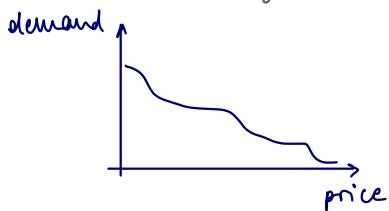
→ the mean expected value changes during time

→ without any information about the change of the expected value, a non-stationary environment is equivalent to an adversarial environment  
↳ non-stationary environments provide better guarantees in some situations (zero-sum game)

### • Non-stationary environments

① Abrupt changes ~ a new product enters the market

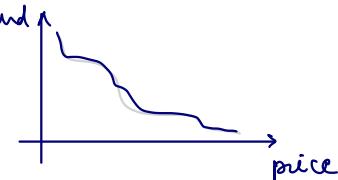
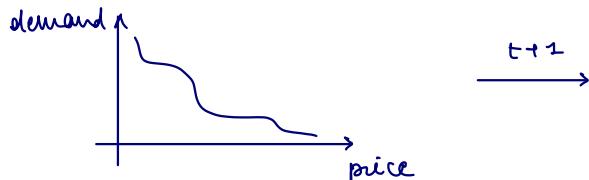
- time is divided in phases
- in every phase the reward functions are constant  
(the change is between the end of a phase and the beginning of a new one)



the length of phases may change and it's described by a breakpoint.  
phases can also be periodic

② Smooth change ~ a product loses attractiveness during time

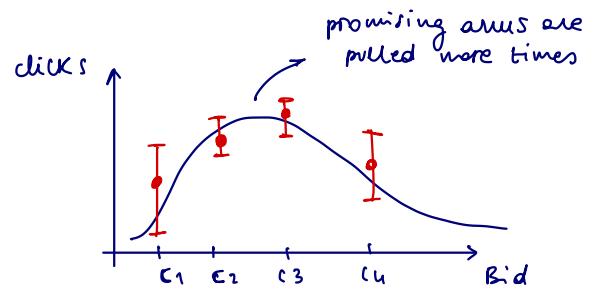
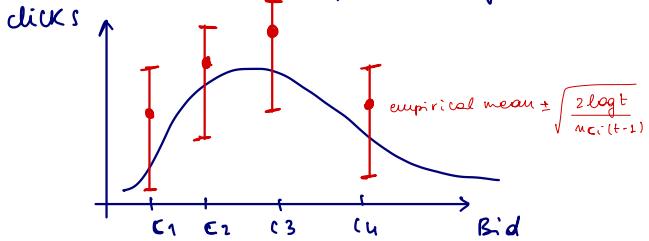
- Reward functions continuously change during time



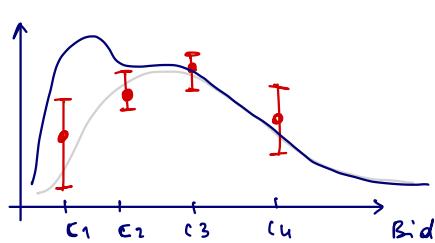
## B) Failure of standard algorithms

- standard bandit algorithms fail in non-stationary environments because they exploit confidence bounds which reduce monotonically
- The main reason is that their bounds reduce: bounds get closer linearly as the number of observed samples increases and logarithmically in time

$\Rightarrow$  UCB1 and abrupt changes



clicks



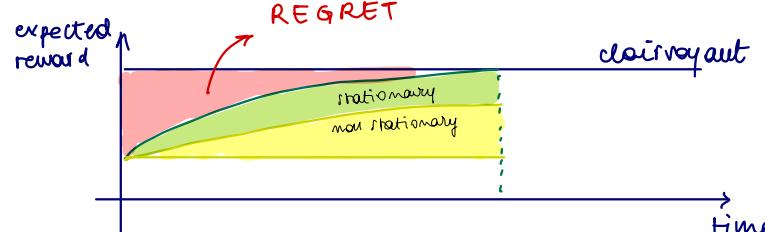
if there is a change of the demand wave, then UCB1 needs a long time to understand that  $c_1$  is the best and during all this time we continuously collect regret  
 $\rightarrow$  exponential time, because the confidence bound increases logarithmically

$\Rightarrow$  We need that confidence bounds don't monotonically reduce in time.

- A solution could be that we can use a sliding window to forget old samples
- $\hookrightarrow$  confidence bounds are computed by looking at last samples
  - $\hookrightarrow$  we get less accurate estimates and suboptimal decisions

### • Regret lower bound

The regret lower bound is larger than in stationary environments



### SW-UCB1 algorithm

- Step 1: play once every arm  $a \in A$   
 Step 2: at  $t$  play arm  $a_t$  with  $m_{at}(t-1, z) = 0$ , otherwise play arm  $a_t$  such that  

$$a_t = \operatorname{argmax}_{a \in A} \left\{ \bar{x}_{a,t,z} + \sqrt{\frac{2\log t}{m_{a,t,z}}} \right\}$$

$\bar{x}_{a,t,z}$ : empirical mean of  $x_a$  at time  $t$   
 computed with last  $z$  samples  
 $m_{a,t,z}$ : number of samples of arm  $a$  in  $z$  steps

### SW-TS algorithm

- Step 1: start with  $\alpha = 1, \beta = 1$   
 Step 2: at every time, for every arm  $a$   
 $\tilde{\theta}_a = \text{sample } (\text{IP}(\mu_a = \theta_a))$   
 Step 3: at every time  $t$  play arm  $a$  such that  

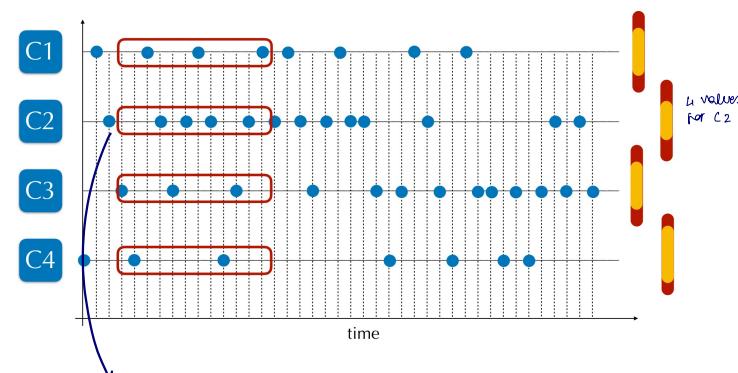
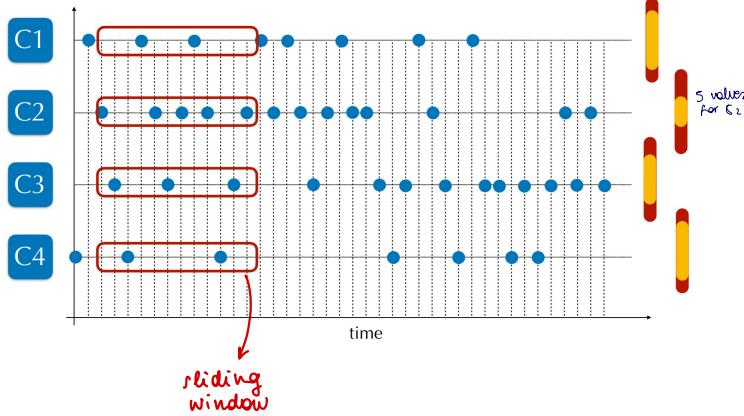
$$a_t = \operatorname{argmax}_{a \in A} \{ \tilde{\theta}_a \}$$

- Step 4: update Beta parameters of arm  $a_t$

$$t \leq z \quad (\alpha_{a_t}, \beta_{a_t}) = (\alpha_{a_t}, \beta_{a_t}) + (x_{a,t}, 1 - x_{a,t})$$

$$t > z \quad (\alpha_{a_t}, \beta_{a_t}) = \max \{ (1, 1), (\alpha_{a_t}, \beta_{a_t}) + (x_{a,t}, 1 - x_{a,t}) - (x_{a_{t-z}(t-z)}, 1 - x_{a_{t-z}(t-z)}) \}$$

- Sliding window and confidence bounds



### • Abrupt change

- the number of breakpoints must be small wrt the time horizon

$$m = |\{b_1, \dots, b_m\}| = \mathcal{O}(T^\alpha) \quad \alpha \in [0, 1]$$

- the regret is of the order  $\mathcal{O}(|A| T^{\frac{\alpha+1}{2}})$  → linear in the number of arms  
 $\alpha=1$  regret linear in  $T$   
 $\alpha=0$  number of breakpoints constant in  $T$

### • Smooth change

- the change is sufficiently smooth  
 $\delta > 0 : |\mu_{it} - \mu_{it'}| \leq \delta |t - t'|$

- The separation between the expected rewards of two arms is arbitrarily small only for a limited number of rounds

- the regret is of the order  $\mathcal{O}(|A| T^{\frac{\alpha+1}{2}})$

⇒ we can consider also the case of abrupt + smooth change

- time horizon is divided in phases (corresponding to abrupt changes)
- in every phase the change is smooth
- the regret is of the order  $\mathcal{O}(|A| T^{\frac{\alpha+1}{2}})$

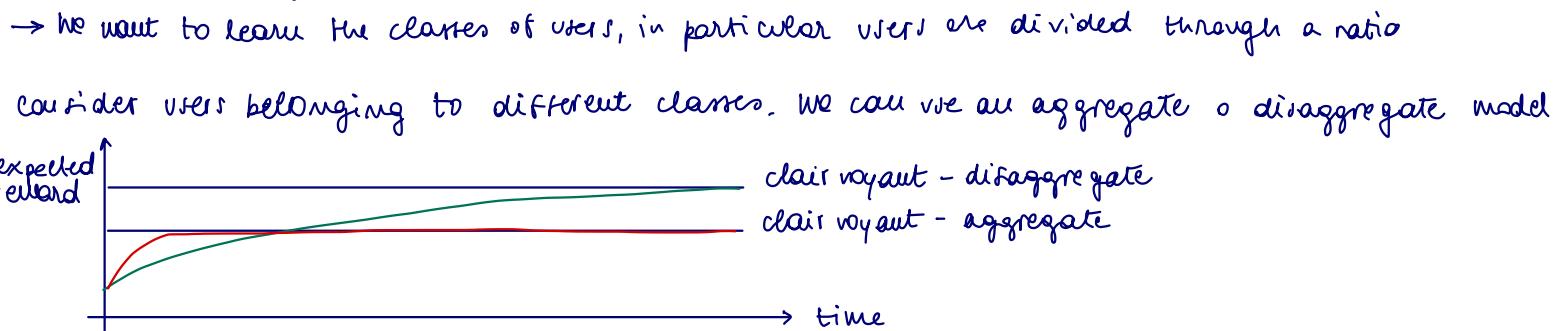
→ in practice we don't know

- number of breakpoints
- coefficient of smoothness
- the size of the sliding window cannot depend on unknown parameters

$$\boxed{T \propto \sqrt{T}} \quad \text{length of } T$$

- Ren • sliding windows force every candidate to be chosen repeatedly during time and can be re-evaluated in order to set the optimal one -
- Non stationary environments require permanent exploration (otherwise the algorithm suffers from a constant regret)  
 $\Rightarrow$  the regret is much higher

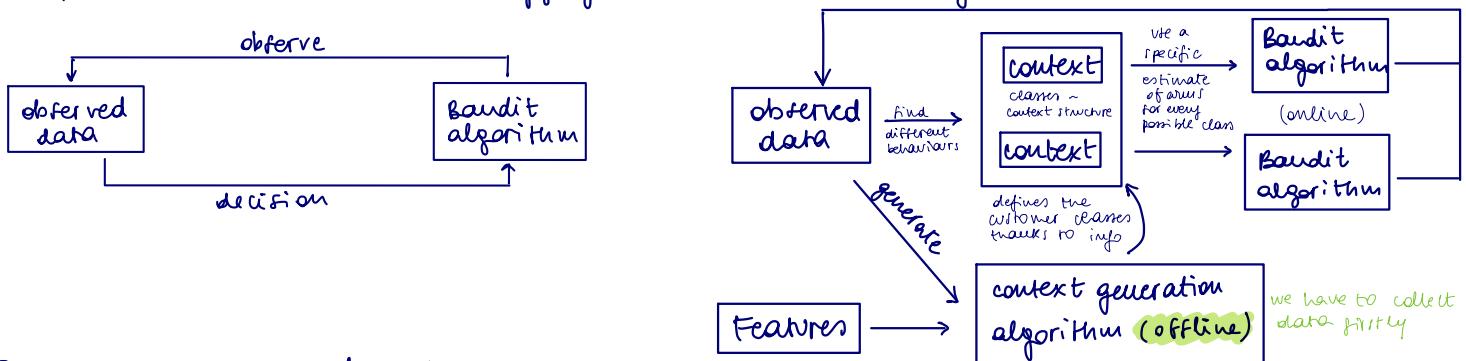
# CONTEXT GENERATION



The performance depends on the amount of data we have : { many data → disaggregate model  
few data → aggregate model

The best we can do is to start from an aggregate model and when we collect many samples we adopt a disaggregate model.

⇒ we have to learn the disaggregate model (so how many classes) from data



## 2 Context generation algorithm

Suppose to collect a lot of data and to be able to characterize data with different features. We're able to partition the space of features and to compare different partitions in order to evaluate the best one (male vs female → 2 different models ⇒ we split)  
→ the algorithm learns the classes!

- How many partitions?

The number of partitions is really huge ⇒ all the possible partitions can't be evaluated  
↳ some heuristics guiding the search is needed

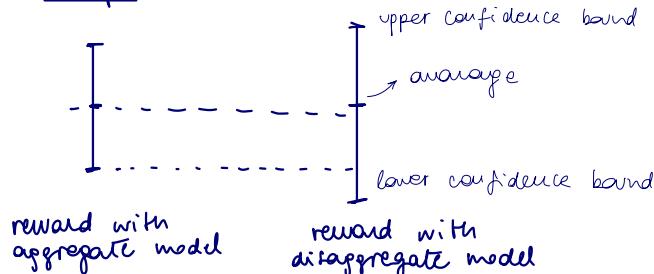
- How to compare the partitions?

If we have many partitions we need a lot of data in order to learn. The idea instead is to generate the context according to the data we get

Actually with few data, the convergence of the bandit algorithms is too slow

⇒ partitioning makes sense when we can provide a strong evidence that is better to do  
⇒ the comparison is made by looking a lower confidence bound

### Example



Here we have two possible context structures.

A: one context → we have an accurate confidence bound

B: two contexts → we have more uncertainty in the context, so we have a larger confidence interval

→ the best we can do is to split many times in order to have a bigger average needed, but splitting too much means the learning slower

⇒ we look at lower confidence bound (average and confidence intervals are useless from a computational point of view)

- Formal model

space of attributes :  $F \subseteq \{0,1\}^t$

Context :  $C \subseteq F$

Context structure :  $P = \{C : C \subseteq F, \bigcap_{i \in C} i = \emptyset \quad \bigcup_{i \in C} i = F\}$  → partition

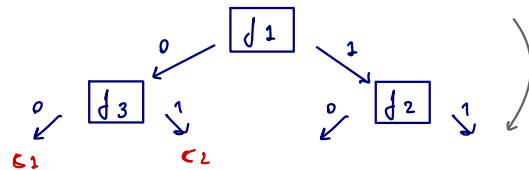
## • How many contexts?

Since the number of possible contexts is exponential, then the number of possible context structures is huge as well

→ the expected value of a context structure is  $\sum_{c \in \Omega} p_c \cdot \mu_{a^*, c}$  → sum of expected rewards of pulling the best arm in each context  
 probability to choose context  $c$

## ■ A greedy approach

Since we have an exponential number of contexts, we iteratively identify the most promising feature and we split the rest accordingly with a feature tree (here binary features)



every step we choose a feature and we decide if splitting is worth

SPLIT  $\iff$

The lower bound of the expected reward when we split



lower bound of the expected reward when we don't split

if we look at lower bound we're sure that the whole performance is better

1<sup>st</sup> context 2<sup>nd</sup> context

$$p_{c_1} \mu_{a^*, c_1} + p_{c_2} \mu_{a^*, c_2} > \mu_{a^*, c_0}$$

probability = lower bound on the user belongs to context  $c_1$   
 lower bound on the best expected reward for  $c_1$  choosing the best arm

probability = lower bound on the user belongs to context  $c_2$   
 lower bound on the best expected reward for  $c_2$  choosing the best arm

lower bound of original context

The choice of lower bound depends on the distribution. If it is finite, we can use the Hoeffding bound (with Bernoulli distribution)

$$\bar{x} - \sqrt{-\frac{\log \delta}{2|z|}}$$

where  $\delta$  is the confidence and  $z$  the set of data

## algorithm

- Step 1: for every feature evaluate the expected value after the split by using the lower bound
- Step 2: select the feature with the maximum marginal value if larger than the non-split case

→ properties:

- No optimality is guaranteed if we use a greedy algorithm
- The tree provides an immediate interpretation about the most important feature that affects the problem

# MATCHING IN ONLINE ENVIRONMENTS

The online advertising data is an example of matching a user with an advertisement

## MATCHING PROBLEMS AND ALTERNATING-PATH ALGORITHM

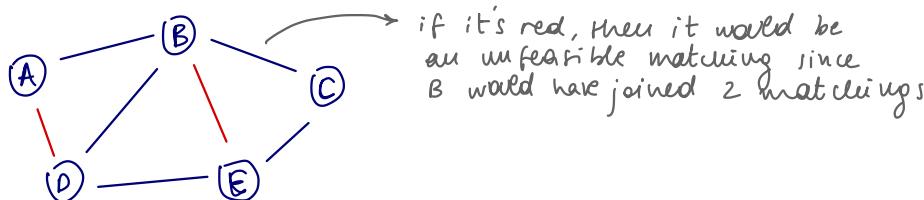
### Matching problem formulation

We are given a graph composed by nodes and edges.

Each node represents a person and a link exists if 2 people share something

### Def (matching)

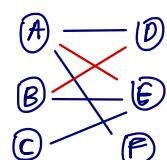
A matching is a subset of edges of the graph such that no pair of edges share vertices  
⇒ every vertex can occur in the matching only once



→ Actually most of our problems are explained by bipartite graphs

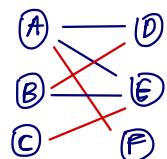
### Def (maximal matching)

A maximal matching is a matching which cannot be enlarged  
⇒ we cannot add a new edge



### Def (maximum matching)

A maximum matching is a matching with largest cardinality  
⇒ largest possible number of edges



### Def (perfect matching)

A perfect matching is a matching in which all vertices are matched  
⇒ all nodes are in the matching

⇒ a maximum matching always exists  
a perfect matching may not exist

Matching problem: Given a graph, the goal is finding a maximum matching

### • Alternating path algorithm

The most effective algorithms are

→ Hopcroft - Karp algorithm for bipartite graphs  $O(m\sqrt{m})$

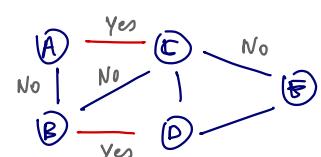
- Edmunds algorithm for arbitrary graph  $O(mn^2)$  // works with weighted graphs and

→ Micali - Vazirani algorithm for arbitrary graph  $O(m\sqrt{m})$  it's an extension of HK

All these algorithms are based on the concept of alternating path

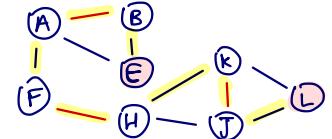
### Def (alternating path)

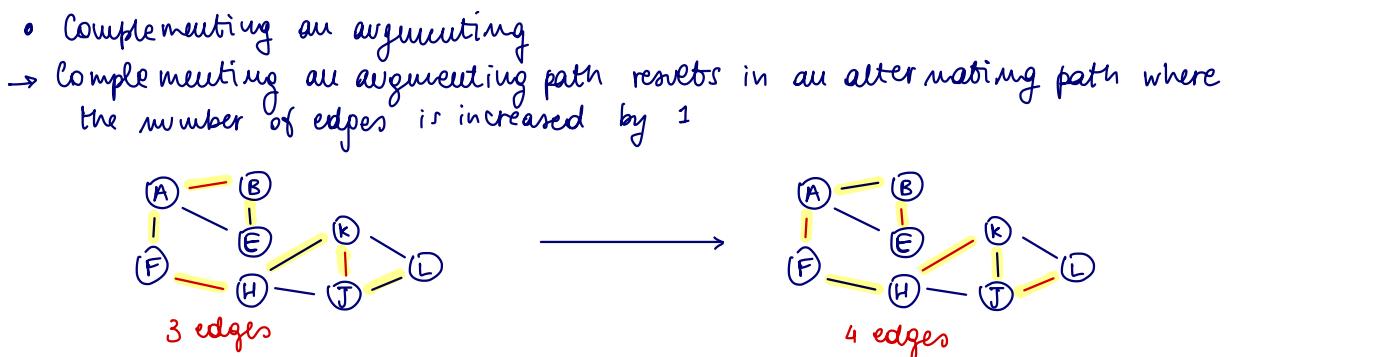
An alternating path is a path such that edges of the matching and edges of the complementary set (edges not in the matching) alternate



### Def (augmenting path)

An augmenting path is an alternating path starting and ending in unmatched vertices





→ A matching is maximum if it does not admit any augmenting path

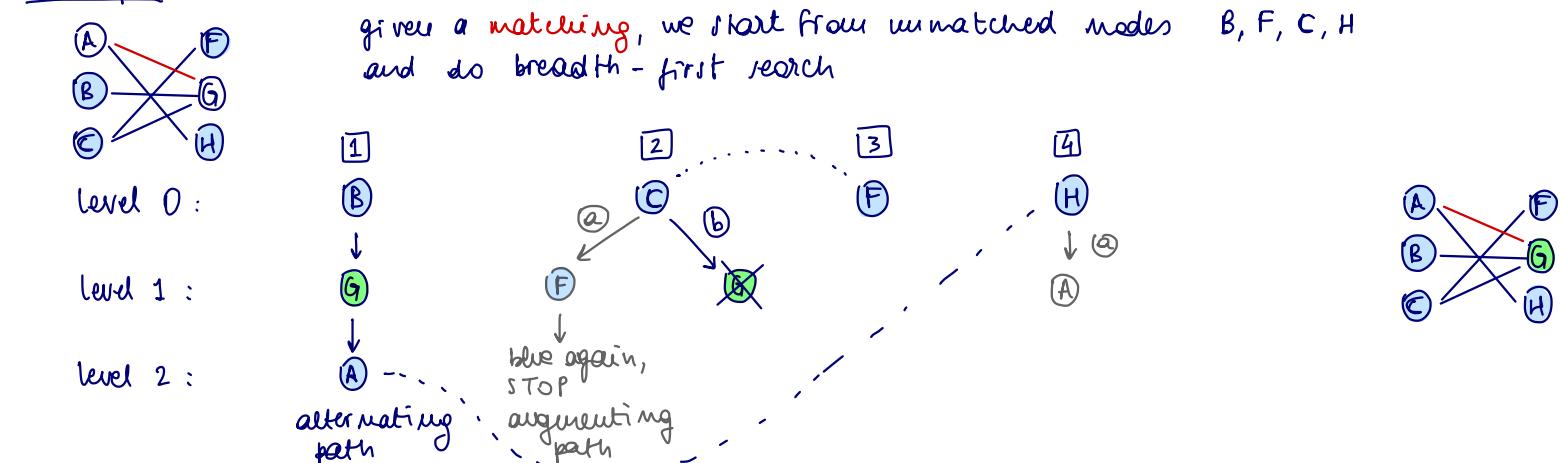
### Alternating-path algorithm

1. look for an augmenting path
  2. make the complementary
- ↑ repeat until no augmenting path exists  
↓ we find a maximum matching

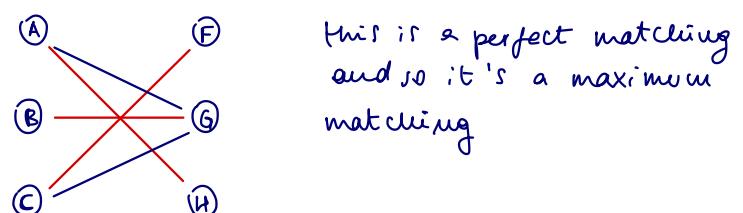
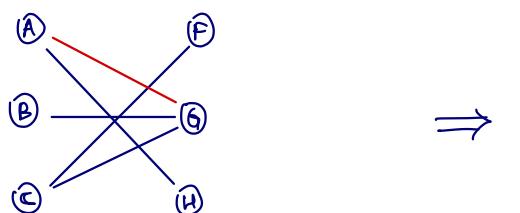
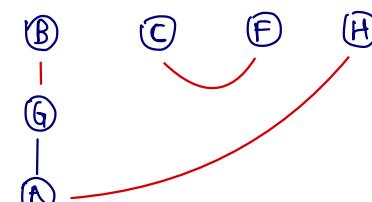
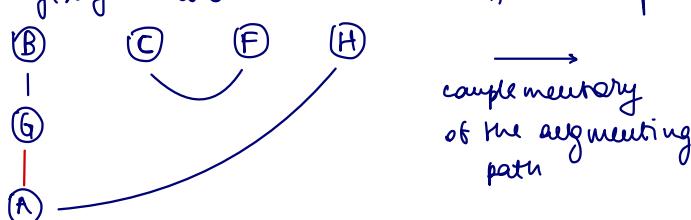
→ problem: looking for an augmenting path - we are given a matching + graph and we look for all the shortest augmenting path

1. consider every unmatched vertex
2. make a breadth-first search starting from every unmatched vertex
3. Given the unmatched node, make an expansion adding that path and the connected path in the matching if the vertex is matched
4. label vertices as blue/green according to the level of search tree
5. If expanding up the path we visit some already visited nodes then we stop the branch and we make an action according to the specific case

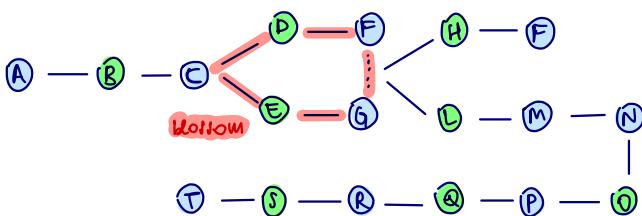
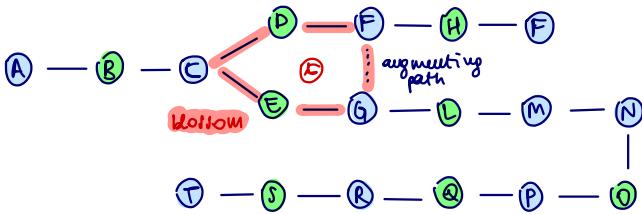
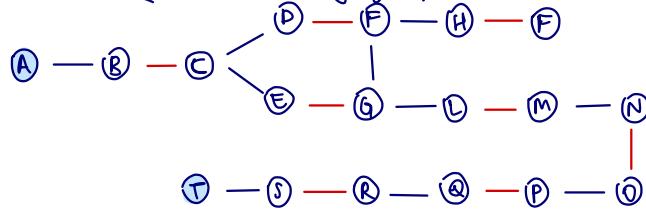
### Example



- ⓐ If the expanding even node is directly connected to another even node (previously labeled) then we have found an augmenting path  
⇒ such edge can be added to the matching
- ⓑ If the expanding even node is directly connected to an odd node (previously labeled) belonging to another subtree (root), then stop the search

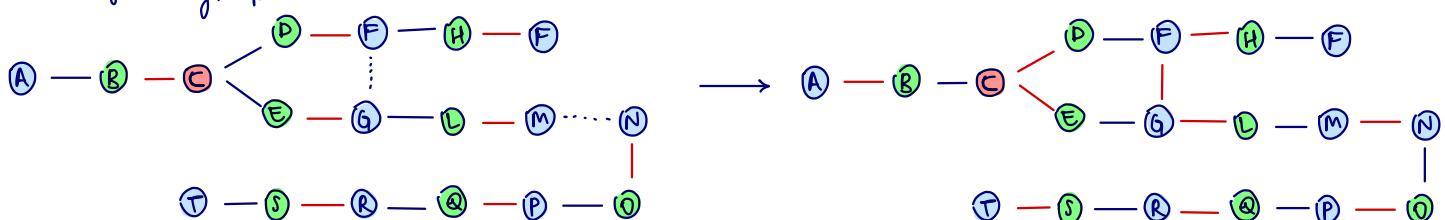


### Example (arbitrary graph)

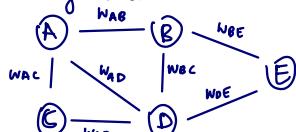


- How to deal with a blossom?

If an augmenting path is found, the blossom is expanded and we do the complementarity of the graph



### Weighted matching problem ~ assignment problem



each edge has a weight:  $w_{ij} > 0 \forall ij$

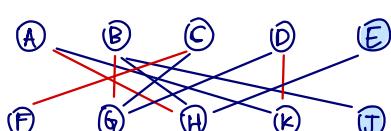
Def a maximum weighted matching is a matching with the largest cardinality, so the one that maximizes the sum of weights

- alternating path algorithm

- Hungarian algorithm for bipartite graphs  $\Theta(m^2n)$
- Edmunds algorithm for arbitrary graphs  $\Theta(mn^2)$

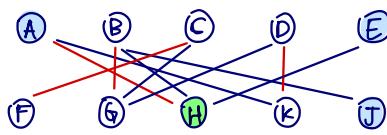
## MATCHING IN BIPARTITE GRAPHS

~ Alternating path algorithm applied to a bipartite graph ~

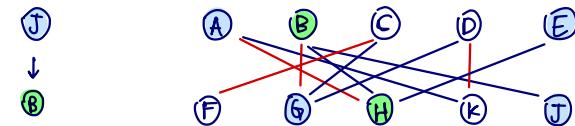


$$M = \{(A, H), (B, G), (C, F), (D, K)\}$$

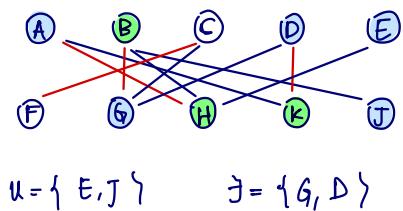
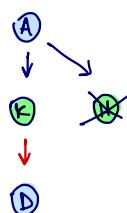
$$U = \{E, J\} \quad \bar{J} = \{E, J\}$$



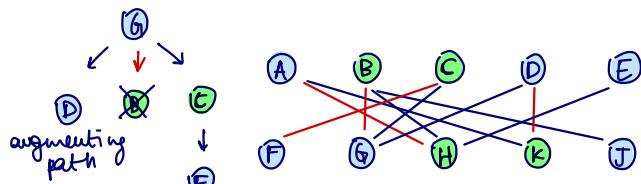
$$U = \{E, J\} \quad \bar{J} = \{J, A\}$$



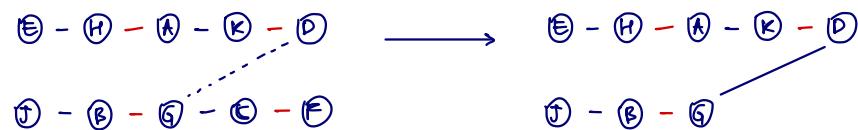
$U = \{E, J\}$        $Y = \{A, G\}$



$D \rightarrow \text{stop}$   
 $F \rightarrow C \rightarrow \text{stop}$



$U = \{E, J\}$        $Y = \{D, F\}$



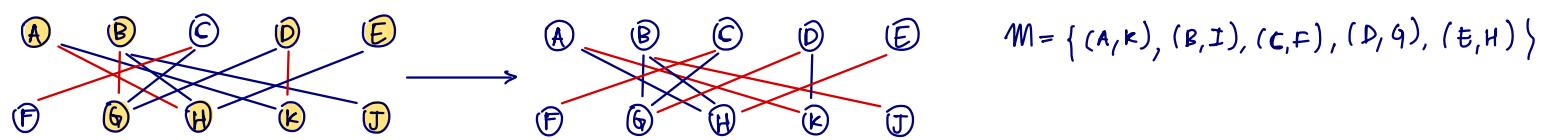
$U = \{E, J\}$        $Y = \{D, F\}$



complementary

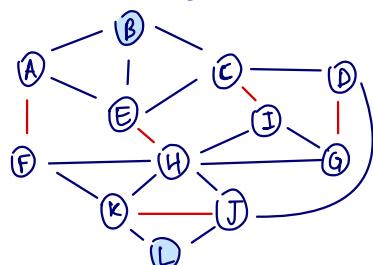
$\rightarrow$   $U = \{H, I, K, D\}$

and so we find



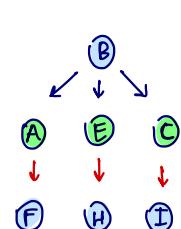
## MATCHING IN ARBITRARY GRAPHS

~ Alternating path algorithm applied to an arbitrary graph ~



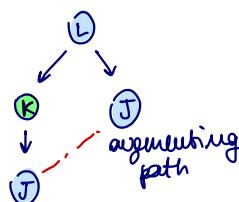
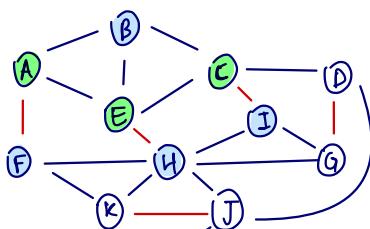
$M = \{(A, F), (C, I), (D, G), (K, J), (E, H)\}$

$U = \{B, L\}$        $Y = \{B, L\}$



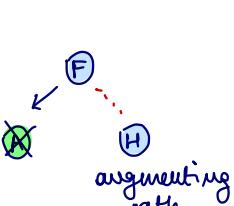
$U = \{B, L\}$

$Y = \{L, F, H, I\}$



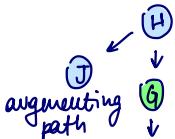
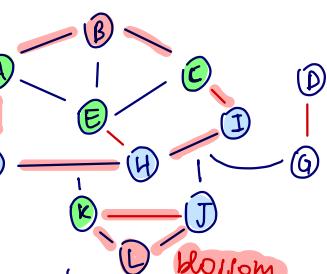
$U = \{B, L\}$

$Y = \{J, F, H, I\}$



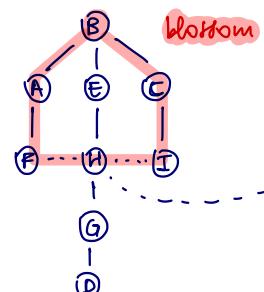
$U = \{B, L\}$

$Y = \{H, I, J\}$

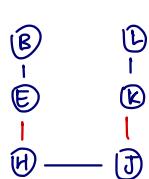


$U = \{B, L\}$

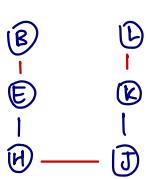
$Y = \{I, J, D\}$

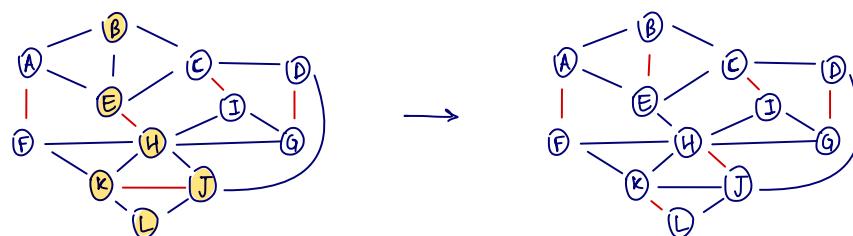


$\rightarrow$



$\rightarrow$





## COMBINATORIAL BANDITS AND MATCHING PROBLEMS

### • Combinatorial constraints

Until now we said that candidates (arms) are drawn from an unknown probability distribution. Now we add combinatorial constraints by pulling a subset of arms simultaneously.

Def superarm: collection of candidates (arms)

feasible superarm: superarm satisfying the constraint  $\mathbb{E}(g) = 0$   
reward of a superarm: sum of the rewards included in a superarm

→ Goal: maximize the cumulative in time expected reward

### • Assumption

The random variable of every arm is Bernoulli in  $\{0,1\}$  with unknown mean

### Thompson Sampling pseudocode

1. at every time  $t$  and for every arm  $a$

$$\tilde{\alpha}_a \leftarrow \text{sample } (\mathbb{P}(\mu_a = \theta_a))$$

$\underline{a}$ : superarm

2. at every time  $t$ , play superarm  $\underline{a}_t$

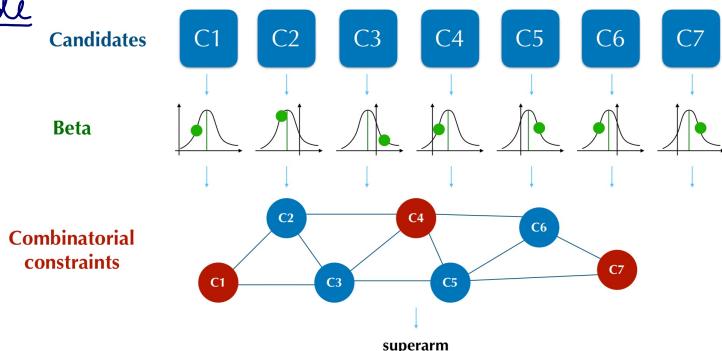
$$\underline{a}_t \leftarrow \underset{\sum_{a \in \underline{a}} \theta_a = 0}{\operatorname{argmax}}$$

solve a combinatorial problem

3. Update the Beta distribution

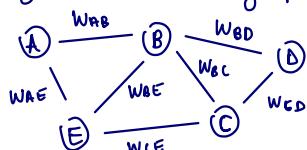
$$(\alpha_{at}, \beta_{at}) \leftarrow (\alpha_{at}, \beta_{at}) + (x_{at,t}, 1 - x_{at,t})$$

### Example



Def (TS Regret)  $Q_T(TS) = \mathcal{O}\left(|\mathcal{A}| \frac{\log T}{\Delta_{\min}}\right)$

### • Weighted matching problem and TS



$w_{ij} > 0$   $w_{ij}$  unknown  $\rightarrow$  The value of each edge is rv whose expected value is unknown

$\rightarrow$  matching  $\sim \text{Be}(p)$

$p$  = probability to pair

### Steps:

1. choose a superarm (edges)
2. draw samples
3. collect rewards
4. update Beta parameters

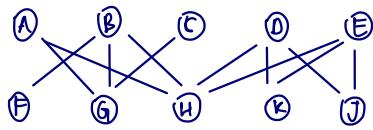


iterate for  $T$  rounds

we converge to the best matching because combinatorial bandit guarantees to minimize the regret

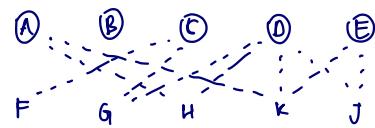
# ONLINE MATCHING PROBLEM

~ Matching problems which are online in time ~  
classical matching problem

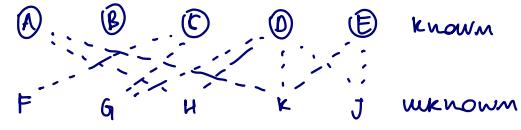


graph is completely known a priori

## Basic bipartite online matching



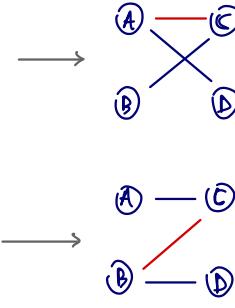
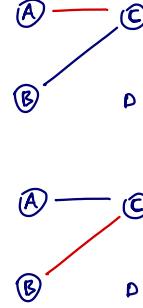
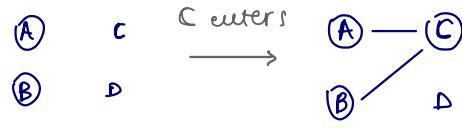
## Online matching problem



graph is discovered during time

1. One side of the bipartite graph is known a priori
2. The other side and edges are unknown at the beginning
3. at each round a single node enters the problem and its edges are discovered
4. at each round we decide if the entering node is matched with another node or it leaves the problem

• Online matching is inefficient



since A/C is in another  
matching, we can't do  
A-D / C-D

(we can't do the best: A-D, B-C)

now we need to decide what to match

C-B, C-A, none

inefficiency: the optimal choice

depends on what happens in the future

## Clairvoyant vs online optimal solution

Clairvoyant optimal solution  $\rightarrow$  best solution we have with complete information about the problem

Def competitive factor of an algorithm  $\alpha_L = \min_P \frac{\Gamma_L(P)}{\Gamma^*(P)} = \frac{\# \text{matchings of the algorithm}}{\# \text{matchings of clairvoyant solution}}$

Key • No deterministic algorithm can have a competitive factor  $> 1/2$  in online bipartite matching problems

• No randomized algorithms can have an expected competitive factor  $> \frac{e-1}{e}$  in online bipartite matching problem

## Deterministic algorithm

Match a node with an arbitrary unmatched node  
 $\rightarrow$  competitive factor =  $1/2$

### proof

1<sup>o</sup> step: the cardinality of any maximal matching is at least  $1/2$  of the cardinality of the maximum matching

let  $M^*$  the maximum matching (largest possible matching)

let  $M$  the maximal matching (cannot be enlarged)

$\forall e = (u, v)$  edge in  $M^*$ , either  $u$  or  $v$  is matched in  $M$ , otherwise  $M$  is not maximal  
 $\Rightarrow$  # matched nodes in  $M$   $\geq |M^*|$  more edges involved

since # matched nodes in  $M^*$  is  $\geq |M^*|$   $\Rightarrow |M| \geq |M^*|/2$

2<sup>o</sup> step: the greedy algorithm always finds a maximal matching

Let  $M$  be maximal

$\forall e = (u, v)$  edge not in  $M$ , consider  $v$ :

{  $v$  has been matched to some node other than  $u$

{  $v$  has not been matched since all the connected nodes (also  $u$ ) were already been matched

→ in both cases  $(u, v)$  cannot be added to  $M$

### 3 Randomized algorithm

1. Generate randomly an ordering over the nodes that are initially available

2. Match a node with first unmatched node in the ordering that we generated initially  
→ competitive factor of  $\frac{1}{1-e}$

## GENERALIZATION OF ONLINE MATCHING PROBLEMS

### 4 Problem

1. The graph is discovered during time

2. Every node may wait for at most  $k$  rounds before leaving

3.  $k$  may be node specific

### Prop (Negative result)

No deterministic competitive algorithm even when  $k$  is the same for all nodes

hp: graph is weighted  
graph could be arbitrary or bipartite

competitive:  
 $\alpha_k > 0$

Proof Suppose all nodes stay in the problem for 2 rounds

1. A enters the problem A

2. B enters the problem → Do we match A and B?

↳ match: A — B → C       $\frac{\text{APX}}{\text{OPT}} = \frac{1}{2} > 1$

↳ don't match: B       $\frac{\text{APX}}{\text{OPT}} = \frac{0}{1} = 0$

match A-B  
best clairvoyant sol B-C

don't match  
best clairvoyant sol A-B

⇒ independently from the choice we do, we have the worst case in which the competitive factor = 0, so there's no deterministic algorithm returning a strictly competitive factor.

Positive result → If we resort randomized algorithm we can get a strictly competitive factor which is an expectation wrt the randomization of the algorithm

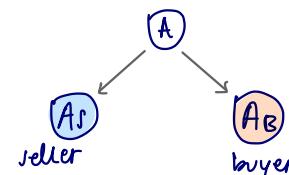
→ Postponed Dynamic Deferred Acceptance (randomized) algorithm has an expected competitive factor of  $\frac{1}{4}$  when nodes stay in the problem for the same number of rounds

### 5 Postponed Dynamic Deferred Acceptance

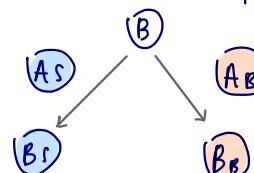
idea of the algorithm: 2-sided market (explore with a market approach)

$\frac{1}{4}$  for arbitrary  
 $\frac{1}{2}$  for bipartite

1<sup>o</sup> step: every time a node enters the problem, the algorithm generates a fictitious bipartite graph in which we have buyers and sellers



2<sup>o</sup> step: suppose A doesn't leave the problem - Then node B enters the problem



→ If A-B in original  $\Rightarrow$  As - Bb and Bs - Ab and we have to find the best matching among those 2

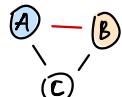
3<sup>rd</sup> step: suppose A and B don't leave the problem.  
Then node C enters the problem

4<sup>th</sup> step: A gets its deadline, so  
 a. we match it  
 b. otherwise it leaves without being matched  
 $\Rightarrow$  find the best matching

a) select A as seller ( $IP = 1/2$ )



A and B are matched



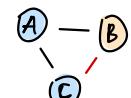
b) select A as buyer ( $IP = 1/2$ )



since the optimal matching is the red one, then we remove A without matching it

5<sup>th</sup> step: No other nodes enters the problem.

B gets its deadline.



### ■ Worst case example



a) select A as seller



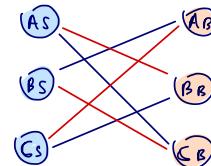
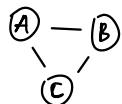
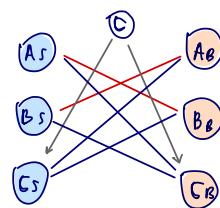
match A-B  
with  $IP = 1/2$

the algorithm returns a matching  
of value 1 with probability  $1/2$

$$\Rightarrow \text{competitive factor: } 1/2 (1 + \gg 1) / \gg 1 \approx 1/2$$

$\rightarrow$  The competitive factor is about  $1/2$  (both if C enters or not the problem)

$\Rightarrow$  In worst case example we get a strictly positive competitive factor, here we have a strictly competitive factor in expectation wrt randomization of the algorithm



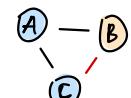
b) select A as buyer ( $IP = 1/2$ )



since the optimal matching is the red one, then we remove A without matching it

5<sup>th</sup> step: No other nodes enters the problem.

B gets its deadline.



b) select A as buyer

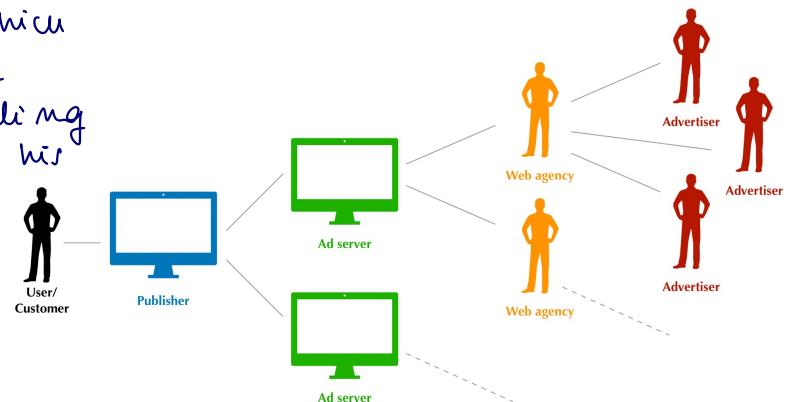


match B-C  
with  $IP = 1/2$

the algorithm returns a matching  
of value  $\gg 1$  with probability  $1/2$

# ONLINE ADVERTISING DATA

- A user / customer visits a webpage proposed by a publisher (owner of website) in which some advertisers want to show an ad.
- The advertiser wants to publish an appealing ad, so he pays the web agency to make his ad optimized.
- The web agency tries to optimize the target and the budget of each ad



## • Performance indices

- number of impressions → awareness
- number of clicks → traffic
- number of leads → potential sales
- number of conversions → sales

The goal is to maximize the conversions

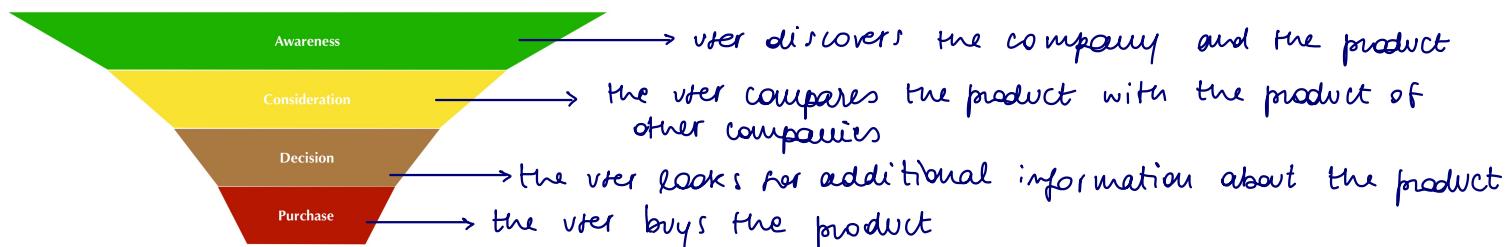
## • Format

- Display advertising → based on cost per impression
- Search advertising → based on pay-per-click
- Social advertising → based on both

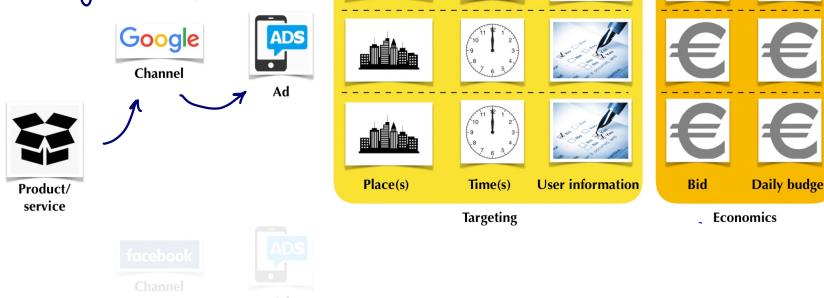
- Payment scheme: when an advertiser pays the publisher he has to pay for
  - cost per impression
  - cost per click

→ in order to evaluate the payments we need to consider all performance indices

## ☒ Marketing funnel



## ☒ Advertising campaign



- choose the channel (FB or Google)
- choose the ads for each channel
- choose the targeting → we get place, time and user information  
(for same ad we have multiple targeting)
- choose the bid and the daily budget
  - bid: amount that the advertiser will pay for a single click
  - daily budget: max amount of money that an advertiser would pay in a day for a given sub campaign

Channel	Ad	Targeting			Economics	
		Place	Time	User's information	Bid	Daily budget
Google	Ad1	Milan	Morning	car	1 Euro	10 Euro
Google	Ad1	Italy	Afternoon	car	2 Euro	1 Euro
Google	Ad2	...	...	sport	...	...

subcampaign

## >User's information

channels distinguish user's information available

- display advertising → cookies (behaviour of users)
- social advertising → interests, friends, previous behaviour
- search advertising → keywords, IP, query

# PAY PER CLICK ADVERTISING

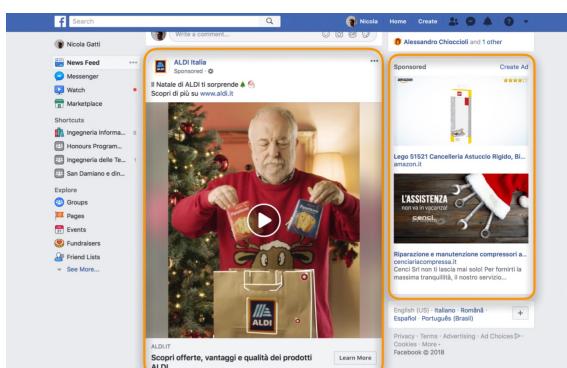
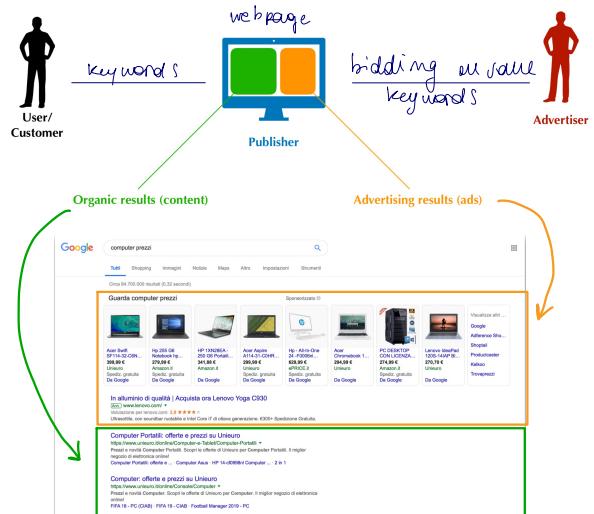
## Search advertising

The advertiser submits an ad to the publisher

The publisher sets the bid (on some keywords)

The user makes the query

Once the user makes a click on the ad, the advertiser pays for such a click

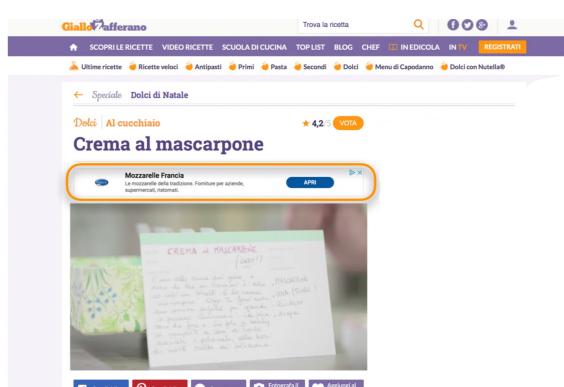


## Social advertising

It's similar to search advertising but it uses social networks as Facebook

## Contextual advertising

It scans a page, looks for keywords and generate a set of pertinent ads



## ■ Publisher side problem

We focus on the optimization problem of the publisher

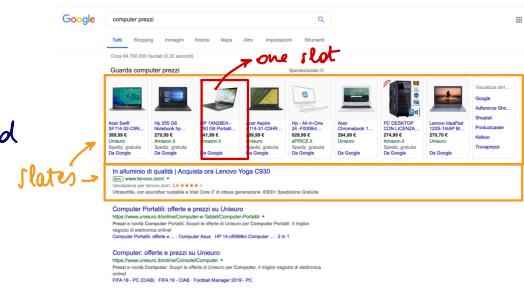
→ The publisher must produce the webpage including the content and ads and he defines the pay-per-click payment

## • Webpage generation

The publisher can display ads in different slots and each slot can be divided in one or more slots.

In each slot there's a single ad.

→ search engines allow a single ad to span over multiple slots



## ■ Mathematical model

The publisher has to maximize the best location of ads in slots

→ the optimization problem is to maximize the value of displaying an ad into a specific slot wrt the budget and bid

$$\text{expected value of having an ad } 'a' \text{ in slot } 's' = \lambda_s q_a v_a$$

probability that the user observes the slot observed

slot prominence

probability to click the ad "a" observed in slot "s" by the user

ad quality

$\lambda_s q_a \rightarrow$  probability to click an ad "a"

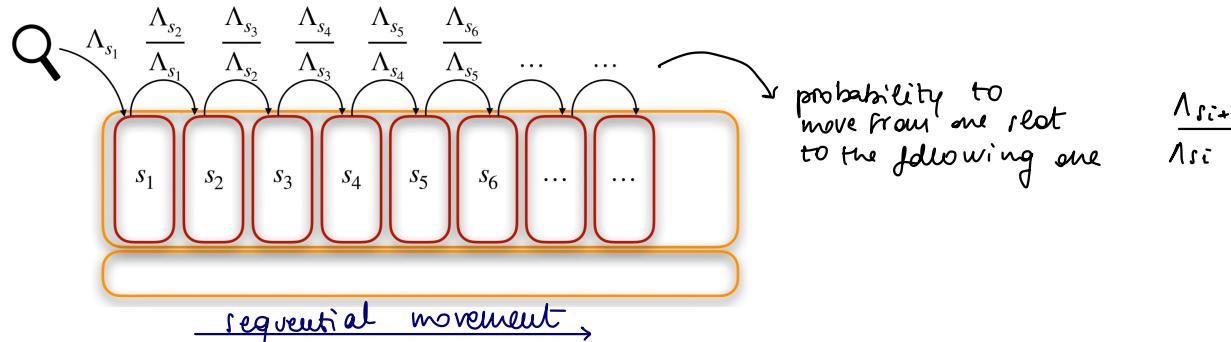
ma è il contrario?

$$\rightarrow \text{expected value of an allocation} \sum_a \lambda_{s(a)} q_a v_a$$

$s(a)$ : function that returns the slot in which each ad is allocated (otherwise  $s(a)=0$ )

## \* cascade

The idea behind the prominence of the slot is that the user follows a cascade model



→ Goal: find a location that maximizes the cumulative value of the slot

In order to find the best allocation we have to maximize the expected value

$$\arg \max_{s(a)} \left\{ \sum_a \lambda_{s(a)} q_a v_a \right\}$$

## \* algorithm

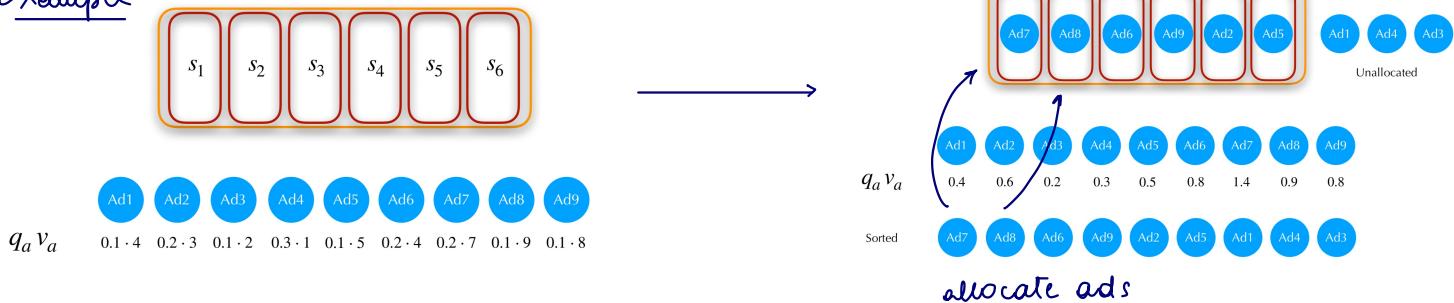
1° step: sort the ads in decreasing order in  $q_a v_a$

2° step: allocate according to such order

order of  $O(n \log k)$

- linear in the number of ads
- logarithmic in number of slots

## Example



- known and unknown parameters

$q_a v_a$  is crucial to find the best allocation, but in general allocation mechanism doesn't know it

$q_a \rightarrow$  estimated by search engine using data of all allocated ads

$q_a \rightarrow$  estimated by the search engine by means of bandit algorithms (with upper bounds)

$v_a \rightarrow$  private information of the advertiser

The advertiser communicates  $v_a$  by making a bid to search engine which can be the true evaluation or not

⇒ The advertiser cannot have a correct estimation of the true value because  $v$  is estimated by the behaviour of activities of website of advertiser

## ■ Auction mechanism

When the search engine generates a page, it runs an auction in order to find the best allocated ads and the pay-per-click-payments

- every ad has a different advertiser
- every advertiser makes a bid which determines the max amount of money the advertiser will pay for a click
- the auctioneer chooses → paymeets (different ways)
  - allocation using bids  $v_a$

### \* paymeets

1° method → GPS (Generalized Second Price)

$$p_a = \frac{q_{a+1} v_{a+1}}{q_a}$$

→ the ads are sorted in decreasing order in the expected value

$$p_a < \frac{q_a}{q_a} v_a = v_a : \text{pay-per-click-paymeet} < \text{bid}$$

⇒ if an advertiser doesn't make a bid larger than its value he will never pay more than its value

2° method → VCG (Vickrey-Clarke-Groves)

$$p_a = \frac{1}{\lambda_s(a) q_a} (x_a - y_a)$$

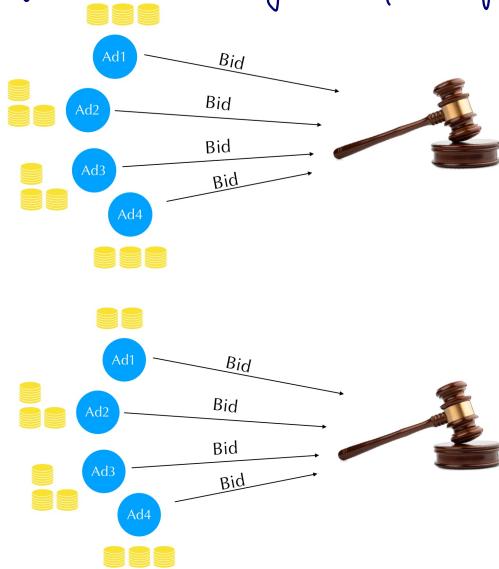
$$x_a = \max_{s(a')} \sum_{a' \neq a} \lambda_{s(a')} q_{a'} v_{a'}$$

$$y_a = \sum_{a' \neq a} \lambda_{\bar{s}(a')} q_{a'} v_{a'}$$

$$\bar{s}(a') = \arg \max_{s(a')} \sum_{a'} \lambda_{s(a')} q_{a'} v_{a'}$$

## Repeated auctions

In practice every advertiser has a daily budget and take part in auctions only if the remaining daily budget is not expired



every time a new user enters

all advertisers who have budget make a bid  
and the auctioneer chooses allocation and payments

the user clicks on ad 1 → ad 1 is the winner and  
it pays the click ⇒ its daily budget reduces  
until it ends

- During the day
  1. the bid for a subcampaign can be changed during the day (with an unknown delay)
  2. the budget for a subcampaign can be set once per day

### properties

GSP → bidding the actual value may not be the optimal strategy even without any budget constraint

VCG → bidding the actual value is the optimal strategy without any budget constraint  
⇒ truthful when played once but in repeated scenario is not truthful anymore  
→ bidding the actual value may be not the optimal strategy with budget constraints

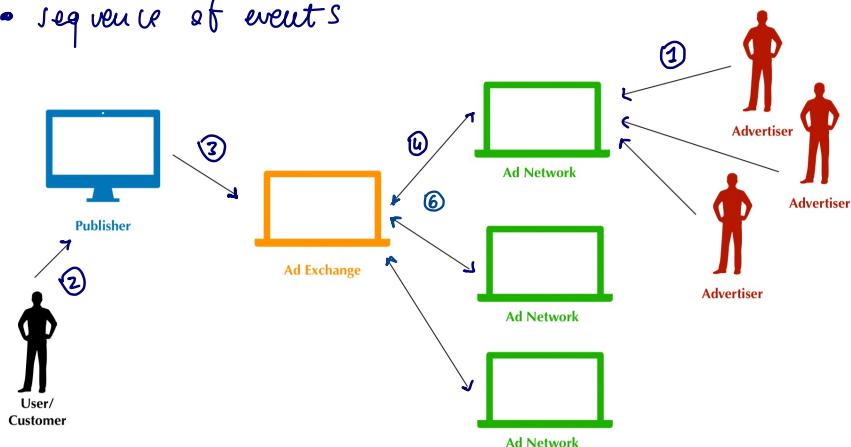
⇒ optimization problem:

Given a daily budget for a sub-campaign, what is the best (constant) bid?

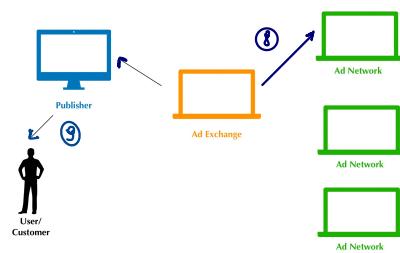
|| Given a daily budget for a campaign, what is the best constant bid and daily budget for every sub-campaign?

## DISPLAY ADVERTISING

### sequence of events



- the advertisers communicate their ads to the ad network with all data in terms of targeting bids
- an user "v" visits webpage "w" of publisher p(w)
- before generating the webpage, the publisher p(w) contacts ad exchange E with  $(w, I(u), r) = (\text{webpage}, \text{info(user)}, \text{context})$
- Ad exchange E contacts the ad network with  $(I(w), I(u)) = \text{users}$
- Every ad network runs an auction to determine the winning ad
- Ad network "i" communicates the ad with the corresponding bid and the budget to the exchange  $\rightarrow (b_i, d_i) = (\text{bid}, \text{budget})$
- Ad exchange determines the winner via an auction
- Ad exchange communicates the winner  $(b_i, d_i)$  to the publisher and the price to the winning ad network
- The publisher serves the webpage to the user with the impression of the ad

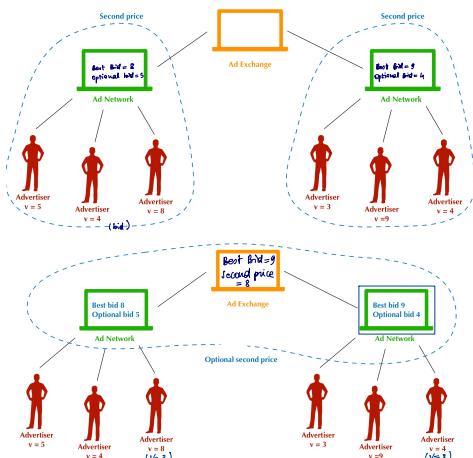


## Ad network auction mechanism

Every ad network runs a second price auction to determine the winner of the network and its payment

- every ad network communicates to the ad exchange
  - The winning bid and the ad (of the auction)
  - Another optional bid (potentially the second price)
- the ad exchange runs a second price auction to determine the winner and the payment

## Example



both ad networks run second price option  
 → the winner is the one with the highest bid and it pays for the second highest bid

the ad exchange runs the optional second price and it selects the best bid and the second price

# OPTIMIZATION PROBLEM

## Assumptions

- The performance of every subcampaign is independent from the performance of the other subcampaigns
  - Not true in general, good approximation
- The values of bid and daily budget are finite and given
  - Not true, but reasonable

Channel	Ad	Targeting			Economics	
		Place	Time	User's information	Bid	Daily budget
Google	Ad1	Milan	Morning	...	?	?
Google	Ad1	Italy	Afternoon	...	?	?
Google	Ad2	...	...	...	?	?
Google	Ad2	...	...	...	?	?
Bing	Ad3	...	...	...	?	?
Bing	Ad4	...	...	...	?	?
Facebook	Ad5	...	...	...	?	?
Facebook	Ad5	...	...	...	?	?
Facebook	Ad6	...	...	...	?	?
Facebook	Ad6	...	...	...	?	?

$$\max \sum_{j=1}^N v_j m_j(x_{jt}, y_{jt})$$

max value per click • number of clicks  
(impression) (depending on bid and budget)

$$\sum_{j=1}^N y_{jt} \leq y_t \quad \forall t \rightarrow \text{daily budget}$$

$$x_{jt} < x_{jt} < \bar{x}_{jt} \quad \forall t, j \rightarrow \text{range of bid for each subcampaign}$$

$$y_{jt} < y_{jt} < \bar{y}_{jt} \quad \forall t, j \rightarrow \text{range of budget for each daily budget}$$

$j = \text{subcampaign}$   
 $x_{jt} \rightarrow \text{bid of subcampaign } j \text{ at day } t$   
 $y_{jt} \rightarrow \text{budget of subcampaign } j \text{ at day } t$   
 $v_j \rightarrow \text{value per click}$   
 $m_j \rightarrow \text{number of clicks}$

## Algorithm

Step 1: remove the dependencies of the values of the bid (?)

	$vn(x_i, y_j)$	bid							max
		$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	
budget	$y_1$	100	110	115	100	max(100, 95, 100, 100)	100	115	115
	$y_2$	120	130	140	132	125	120	115	115
	$y_3$	135	140	143	145	138	130	125	120
	$y_4$	148	153	156	160	152	145	137	130
	$y_5$	152	155	158	163	165	157	140	135
	$y_6$	158	160	163	165	168	165	150	145
	$y_7$	164	168	170	174	178	170	163	158
	$y_8$	166	170	174	178	181	184	171	160

we find the max bid for each daily budget

If these values are larger than the upper bound or smaller than the lower we have an unfeasible case

Step 2: find the best allocation for each budget

		budget								we have C1 and C2
		0	10	20	30	40	50	60	70	
	0	0	0	0	0	0	0	0	0	
	+ C1	-∞	90	100	105	110	-∞	-∞	-∞	
	+ C2	-∞	90	172	192					
	+ C3									
	+ C4									
	+ C5									
C2	0	82	90	92	-∞	-∞	-∞	-∞		

best solution and we take the max

$\text{budget} = 10 \quad \left\{ \begin{array}{l} C_1 \text{ has budget} = 10 \rightarrow 90 \\ C_2 \text{ has budget} = 0 \end{array} \right.$        $\text{budget} = 20 \quad \left\{ \begin{array}{l} C_1 \text{ has budget} = 20 \rightarrow 100 \\ C_2 \text{ has budget} = 0 \end{array} \right.$

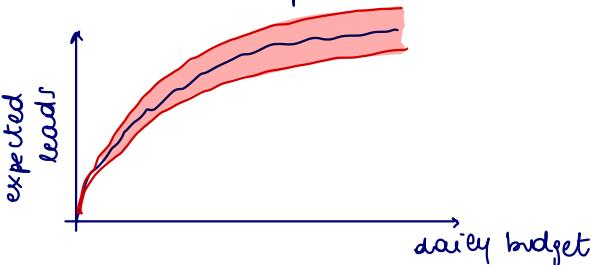
$\left\{ \begin{array}{l} C_2 \text{ has budget} = 10 \rightarrow 82 \\ C_2 \text{ has budget} = 0 \end{array} \right.$        $\left\{ \begin{array}{l} C_1 \text{ has budget} = 10 \rightarrow 90 \rightarrow 172 \\ C_2 \text{ has budget} = 10 \rightarrow 82 \end{array} \right.$

$\left\{ \begin{array}{l} C_1 \text{ has budget} = 20 \rightarrow 90 \rightarrow 192 \\ C_2 \text{ has budget} = 20 \rightarrow 90 \end{array} \right.$        $\left\{ \begin{array}{l} C_1 \text{ has budget} = 0 \\ C_2 \text{ has budget} = 20 \rightarrow 90 \end{array} \right.$

$\left\{ \begin{array}{l} C_2 \text{ has budget} = 30 \rightarrow 105 \\ C_2 \text{ has budget} = 0 \end{array} \right.$        $\left\{ \begin{array}{l} C_1 \text{ has budget} = 20 \rightarrow 100 \rightarrow 182 \\ C_2 \text{ has budget} = 10 \rightarrow 82 \end{array} \right.$

$\left\{ \begin{array}{l} C_1 \text{ has budget} = 20 \rightarrow 100 \rightarrow 182 \\ C_2 \text{ has budget} = 20 \rightarrow 92 \end{array} \right.$

## Real-world example



it shows how many leads the advertiser can get for every cumulative daily budget that is allocated on a campaign  
→ a campaign can decide to increase or decrease the daily budget spent wrt the improvements of the number of leads

→ complexity:  $O(Nk^2)$

- linear in the number of subcampaigns
- quadratic in the number of values for the daily budget

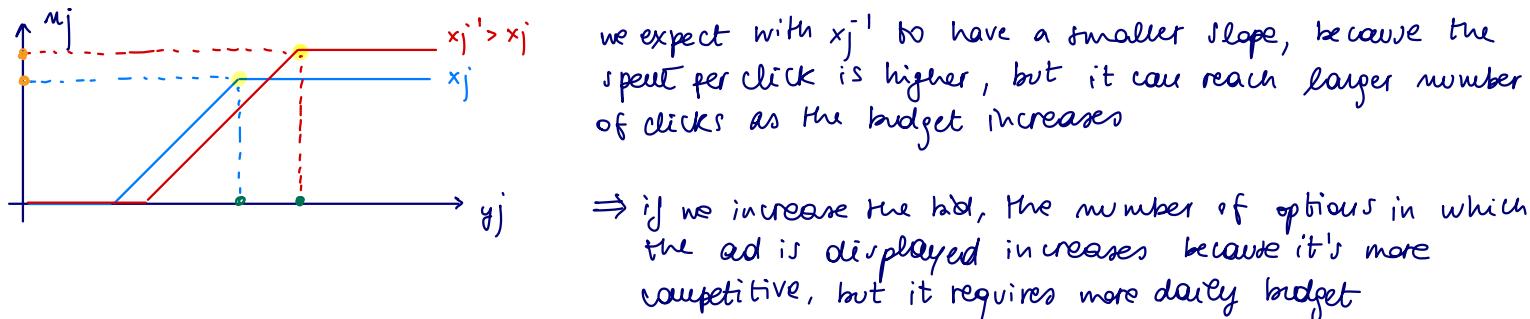
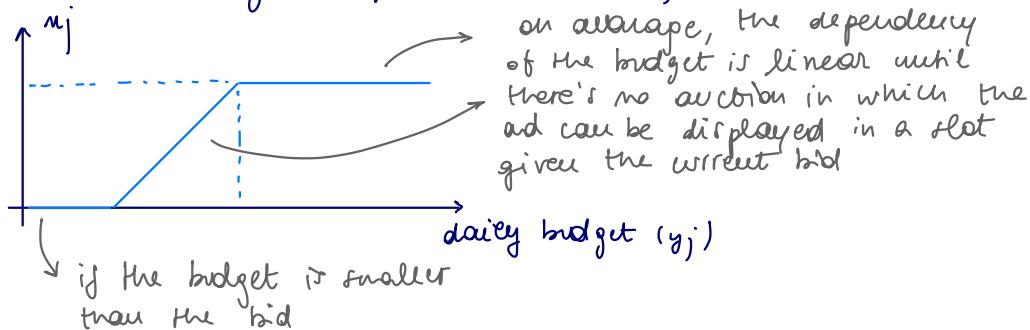
# REGRESSION BY GAUSSIAN PROCESS

The optimization problem can be solved once we know  $n_j(x_{jt}, y_{jt}) \forall x_{jt}, y_{jt}$   
 which are the number of clicks given a value of bid and daily budget  
 → learning every possible combination is unfeasible (it requires too many samples)

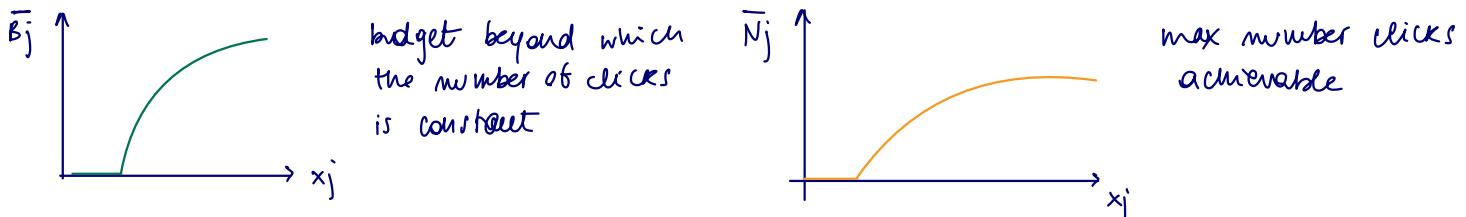
- Approach:

- the function is smooth
- The dependency of the function on bid and budget can be factorized

- Understanding the phenomenon ( $x_j = \text{constant}$ )



→ we want to learn the daily budget which makes the number of clicks constant



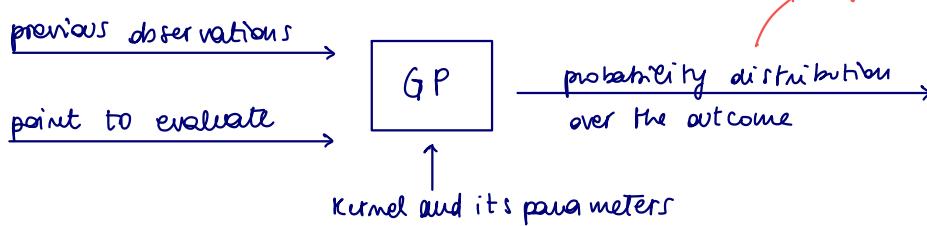
$$\Rightarrow n_j = \max \left\{ 0, \min \left\{ \delta_j, \bar{B}_j(x_j) \right\} - x_j \right\} \frac{\bar{N}_j(x_j)}{\bar{B}_j(x_j) - x_j}$$

- Regression problem

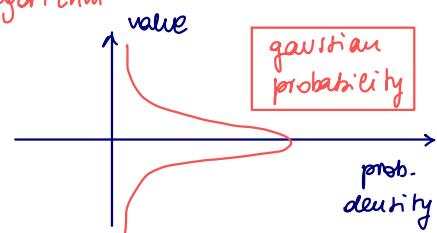
we want to estimate  $\bar{B}_j(x_j)$  and  $\bar{N}_j(x_j)$ , but even if we discretize  $x_j$  we need too many samples

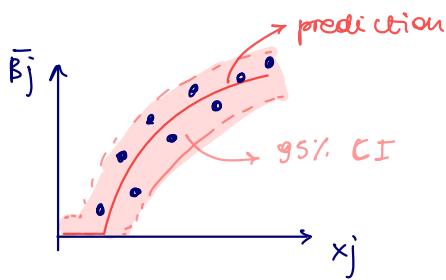
- we assume smoothness (correlation among points)
- we use a Gaussian Process (GP)

- Gaussian process



it assures the convergence of bandit algorithm





by increasing the number of samples the CI has more accuracy

## COMBINATORIAL GP BANDITS

- standard bandit algorithms
  - we are given a set of arms
  - we can pull a single arm
  - we observe the reward of the pulled arm

- GP bandit algorithms
  - arms are correlated
  - the reward of an arm provides information on the reward of the arms closest to it

### \* GP-TS algorithm

1<sup>o</sup> step : at every time t, for every arm a

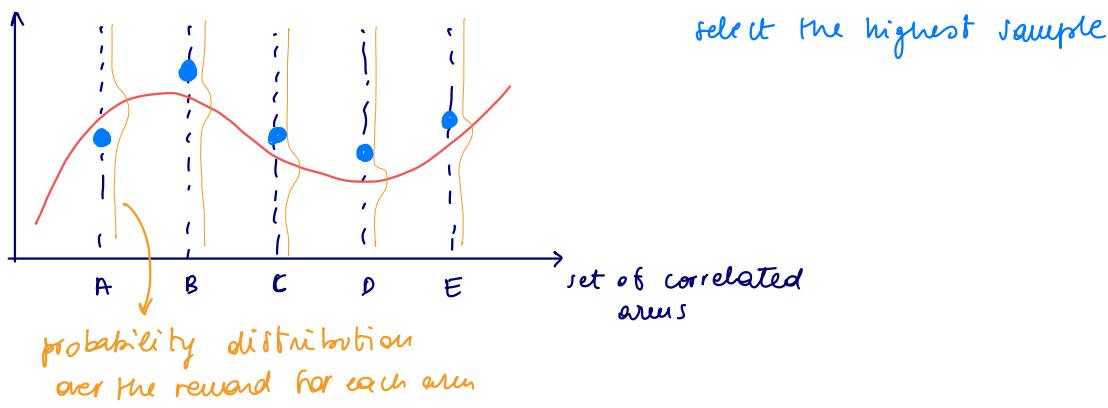
$$\tilde{\vartheta}_a = \text{sample } (\mathbb{P}(\mu_a = \vartheta_a)) \quad (\text{by GP})$$

2<sup>o</sup> step : at every time t, play arm a such that

$$a_t = \underset{a \in A}{\operatorname{argmax}} \{ \tilde{\vartheta}_a \}$$

3<sup>o</sup> step : update the GP according to the observed reward

### Example (Non combinatorial)



### • Regret bound

$$RT \leq \sqrt{\frac{8}{\log(1 + \frac{1}{\sigma^2})} T B \gamma_T} \quad \text{where } B = 8 \log\left(\frac{T^4 M}{6 \delta}\right)$$

$\gamma_T$ : info gained at time t  
 $\sigma^2$ : variance of GP  
 M: number of arms

### • Why GP regression

- The GP plays a crucial role in bandit algorithms because it returns a measure of uncertainty of the regression (CI) (wrt linear regression)
- The probability distribution is necessary for the exploration and convergence to the optimal solution

## Combinatorial GP bandits

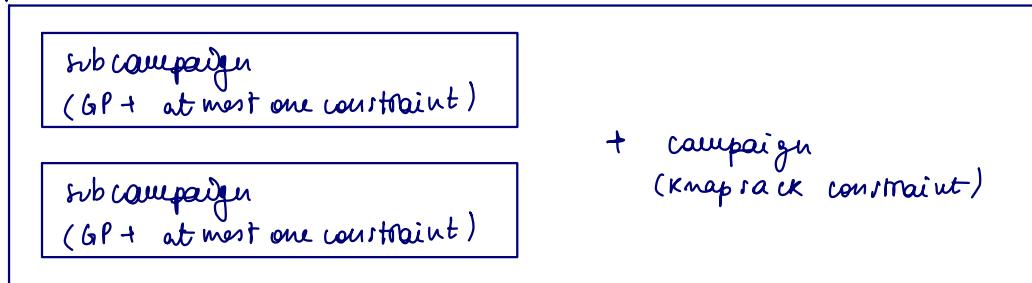
We can pull any set of arms satisfying some combinatorial constraint

- Regret

$$RT \leq \sqrt{\frac{2\Lambda^2}{\log(1 + \frac{1}{\delta^2})} T B \bar{Y}_T M} \quad \text{where } B = 8 \log\left(\frac{2 T^2 M}{\delta}\right) \quad \Lambda: \text{lipschitz constant}$$

- Combinatorial GP bandits for advertising

Every subcampaign is modelled by means of a GP with the constraint that at most a pair of (arm, budget) can be chosen



1. Sample for every pair the reward
2. solve the optimization problem

- Regret

$$RT \leq \sqrt{\frac{2\Lambda^2}{\log(1 + \frac{1}{\delta^2})} C T B \sum_{k=1}^C \bar{Y}_{k,T}} \quad \text{where } B = 8 \log\left(\frac{2 T^2 M C}{\delta}\right) \quad | \quad C: \text{number of GP}$$

# SOCIAL

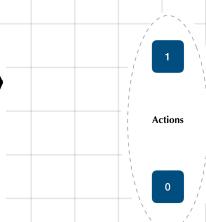
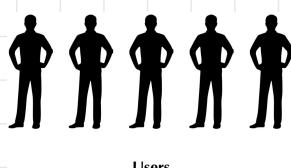
## Information cascades

We focus on scenarios in which a social influence phenomenon is due to information effects.

### Scenario

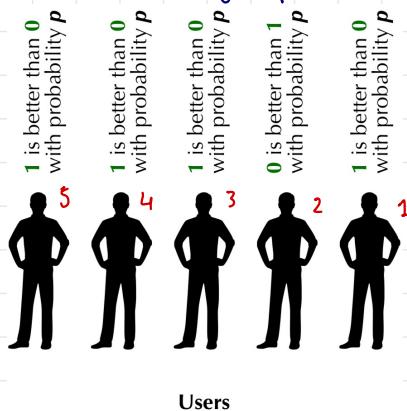
An user has to choose among different proposals (restaurants), but he doesn't have any information about the quality, so he looks at other users' recommendations  
→ IDEA: the user builds his beliefs looking at others.

### Model



We are given a set of users and every user can make one action among a given set of actions

1. Users don't know which action is the best, at the beginning all actions are equal
2. Each user at the beginning receives a signal saying which is the best action with a probability  $p > 1/2$
3. A part from the first user, all users can observe the taken actions by previous users and they update their belief accordingly



user 1: his belief depends only on the signal he received  
→ he chooses 1

user 2: he sees that user 1 chose 1 but the signal he received is that 0 is better than 1. The two beliefs nullify: ties are broken in favour of the received signal  
→ he chooses 0

⇒ all users update their beliefs

The belief of users depends on the number of users that in the past selected action 1 or 0  
⇒ it depends on the difference between the number of observed 1 and 0.