Bandits in Auctions (& more)

Vianney Perchet joint work with P. Rigollet (MIT) and J. Weed (MIT)

CEMRACS 2017
July 20 2017

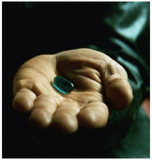
CMLA, ENS Paris-Saclay & Criteo Research

Motivations & Objectives

Classical Examples of Bandits Problems

- Size of data: *n* patients with some proba of getting cured
- Choose one of two treatments to prescribe

or



- Patients **cured** ♡ or **dead** ☠

1) **Inference:** Find the best treatment between the red and blue
2) **Cumul:** Save as many patients as possible

Classical Examples of Bandits Problems

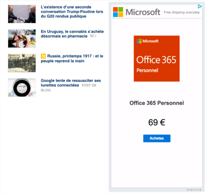
- Size of data: *n* banners with some proba of click
- Choose one of two ads to display

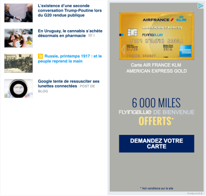
or

- Banner **clicked** or **ignored**

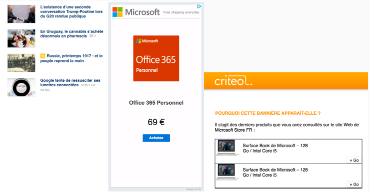
1) **Inference:** Find the best ad between the red and blue
2) **Cumul:** Get as many clicks as possible

Example of Repeated Auctions

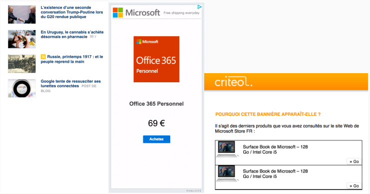




Example of Repeated Auctions



Example of Repeated Auctions



Ad slot sold by lemonde.fr. **2nd-price auctions**

- Several (marketing) companies places bids
- Highest bid wins (…), say criteo, **pays to lemonde** 2nd bid (…)
- criteo chooses ad of a client, Microsoft or Cdiscount or Booking
- criteo gets **paid by the client** if the user clicks on the ad

Main Problem: Repeated auctions with unknown private valuation

Learn valuations, find which ad to display & **good strategies**

Example of Repeated Auctions

Accepter ou refuser les sociétés une par une

Société	Activé / désactivé	Status	Infor
1plusX	Activé / Désactivé	?	
33Across	Activé / Désactivé	?	
4W MARKETPLACE SRL	Activé / Désactivé	?	
Accordant Media	Activé / Désactivé	✓	
Acxiom	Activé / Désactivé	?	
ad4mat	Activé / Désactivé	?	
ADARA	Activé / Désactivé	?	
Adbrain LTD	Activé / Désactivé	?	
AddThis (formerly Clearspring)	Activé / Désactivé	✓	
ADEX	Activé / Désactivé	✓	
Adform	Activé / Désactivé	✓	
ADITION	Activé / Désactivé	✓	
AdLantic	Activé / Désactivé	?	
Admedo	Activé / Désactivé	✓	
Adobe	Activé / Désactivé	✓	
AdRoll	Activé / Désactivé	?	
AdServerPub	Activé / Désactivé	?	
Adtaxi	Activé / Désactivé	✓	
AdUx	Activé / Désactivé	✓	
Affectv	Activé / Désactivé	?	
affilinet	Activé / Désactivé	?	
AggregateKnowledge	Activé / Désactivé	✓	
Amazon Ad System	Activé / Désactivé	✓	
AOL	Activé / Désactivé	✓	
Atlas Solutions, LLC	Activé / Désactivé	✓	
AudienceScience	Activé / Désactivé	✓	
BannerConnect	Activé / Désactivé	✓	
BlueKai	Activé / Désactivé	✓	
Brandcrumb	Activé / Désactivé	?	
Captify	Activé / Désactivé	✓	
Conversant	Activé / Désactivé	✓	
Crimtan	Activé / Désactivé	✓	
Criteo	Activé / Désactivé	✓	
DataXu, Inc.	Activé / Désactivé	✓	
Delta Projects	Activé / Désactivé	✓	
emetriq GmbH	Activé / Désactivé	?	
eXelate	Activé / Désactivé	✓	
Exponential Interactive	Activé / Désactivé	?	
eyeota	Activé / Désactivé	✓	
Ezakus	Activé / Désactivé	?	
Facebook	Activé / Désactivé	✓	
Flashtalking	Activé / Désactivé	?	
Fonecta Oy	Activé / Désactivé	?	
Gamned	Activé / Désactivé	✓	
Google	Activé / Désactivé	✓	
Gumgum	Activé / Désactivé	?	
IgnitionOne	Activé / Désactivé	✓	
Infectious Media	Activé / Désactivé	✓	
Intent Media, Inc.	Activé / Désactivé	?	
iotec	Activé / Désactivé	✓	
iPromote	Activé / Désactivé	?	
Knorex	Activé / Désactivé	?	
Lagardere Publicite	Activé / Désactivé	?	
Leiki	Activé / Désactivé	?	
Ligatus	Activé / Désactivé	?	
Light Reaction	Activé / Désactivé	✓	
LinkedIn	Activé / Désactivé	✓	
Lotame	Activé / Désactivé	✓	
Mapp	Activé / Désactivé	?	
MaxPoint Interactive, Inc.	Activé / Désactivé	✓	
MEDiA iQ	Activé / Désactivé	✓	
MediaMath	Activé / Désactivé	?	
Mediascale GmbH & Co Kg.	Activé / Désactivé	✓	
Metrigo	Activé / Désactivé	?	
Microsoft Advertising	Activé / Désactivé	✓	
mobile.de GmbH	Activé / Désactivé	✓	
mPlatform, Inc.	Activé / Désactivé	✓	
myThings	Activé / Désactivé	?	
Nano Interactive	Activé / Désactivé	?	
NEORY	Activé / Désactivé	?	
Netmining	Activé / Désactivé	✓	
NextPerformance	Activé / Désactivé	?	
nugg.ad	Activé / Désactivé	✓	
OpenX	Activé / Désactivé	✓	
Platform161	Activé / Désactivé	✓	
Plista	Activé / Désactivé	✓	
Programattik	Activé / Désactivé	?	
Public-Idees	Activé / Désactivé	?	
Publicis Media	Activé / Désactivé	?	
Quantcast	Activé / Désactivé	✓	
RadiumOne	Activé / Désactivé	✓	
Rocket Fuel	Activé / Désactivé	✓	
Rubicon Project	Activé / Désactivé	✓	
RUN	Activé / Désactivé	✓	
Scoota	Activé / Désactivé	?	
Semasio GmbH	Activé / Désactivé	✓	
Sizmek Inc.	Activé / Désactivé	✓	
Skimlinks	Activé / Désactivé	✓	
Smart AdServer	Activé / Désactivé	✓	
Sociomantic	Activé / Désactivé	?	
Sojern	Activé / Désactivé	✓	
Stailamedia AG	Activé / Désactivé	?	
Switch Concepts Ltd	Activé / Désactivé	?	
Taboola Europe Limited	Activé / Désactivé	✓	
Tapad	Activé / Désactivé	✓	
Temelio	Activé / Désactivé	✓	
The Trade Desk	Activé / Désactivé	✓	
TubeMogul	Activé / Désactivé	?	
TURBO	Activé / Désactivé	?	
Turn	Activé / Désactivé	✓	
twyn group	Activé / Désactivé	?	
Undertone	Activé / Désactivé	?	
United Internet Media GmbH	Activé / Désactivé	?	
Varick Media Management	Activé / Désactivé	?	
Ve Global	Activé / Désactivé	?	
Viant	Activé / Désactivé	?	
Vibrant Image	Activé / Désactivé	?	

Some companies whose cookies can be controlled

Back to Classical Examples of Bandits Problems

- Size of data: n mails with some proba of spam
- Choose one of two actions: spam or ham

or

- Mail **correctly** or **incorrectly** classified

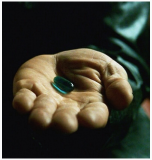
1) **Inference:** Find the best between the red and blue
2) **Cumul:** Minimize number of errors as possible

Back to Classical Examples of Bandits Problems

Back to Classical Examples of Bandits Problems

- Size of data: n patients with some proba of getting cured
- Choose one of two

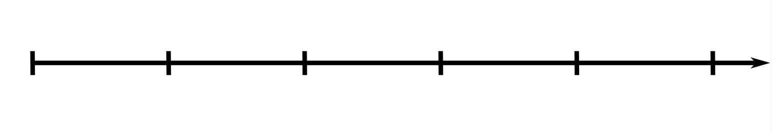
or



- Patients **cured** ♡ or **dead** ☠

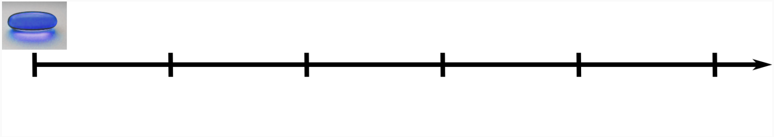
1) **Inference:** Find the best treatment between the red and blue
2) **Cumul:** Save as many patients as possible

Two-Armed Bandit



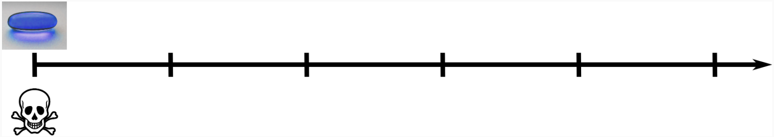
– Patients arrive and are treated sequentially.

Two-Armed Bandit



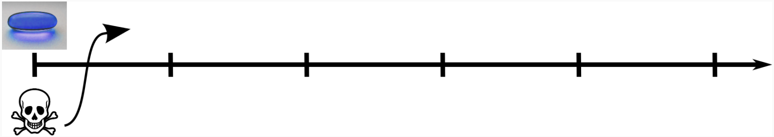
– Patients arrive and are treated sequentially.

Two-Armed Bandit



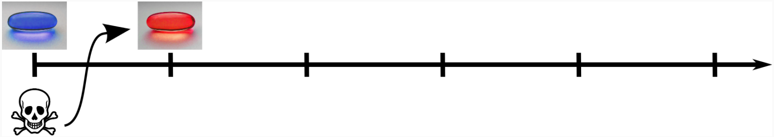
– Patients arrive and are treated sequentially.

Two-Armed Bandit



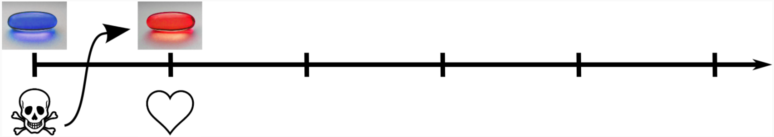
– Patients arrive and are treated sequentially.

Two-Armed Bandit



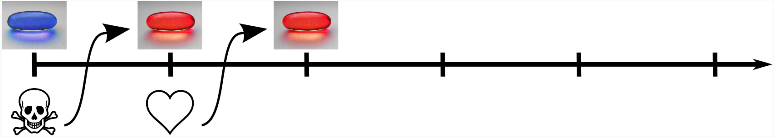
– Patients arrive and are treated sequentially.

Two-Armed Bandit



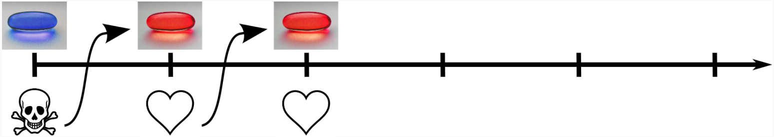
– Patients arrive and are treated sequentially.

Two-Armed Bandit



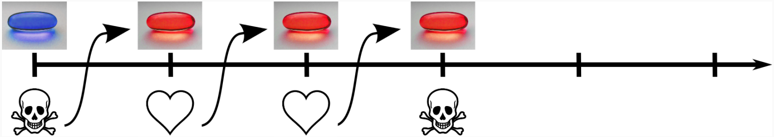
– Patients arrive and are treated sequentially.

Two-Armed Bandit



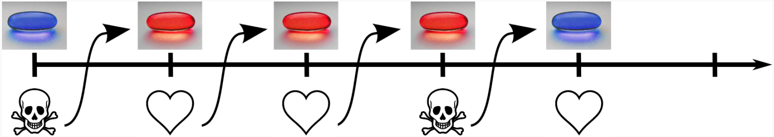
– Patients arrive and are treated sequentially.

Two-Armed Bandit



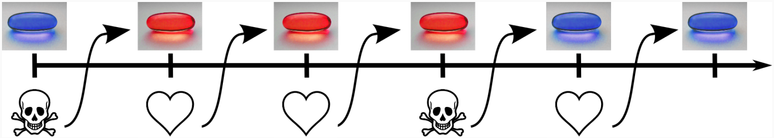
– Patients arrive and are treated sequentially.

Two-Armed Bandit



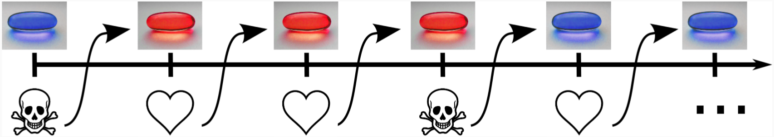
– Patients arrive and are treated sequentially.

Two-Armed Bandit



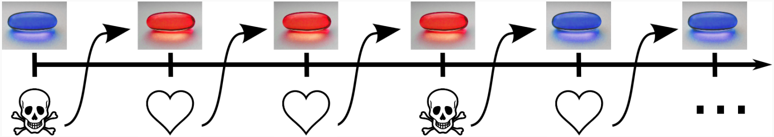
– Patients arrive and are treated sequentially.

Two-Armed Bandit



– Patients arrive and are treated sequentially.

Two-Armed Bandit



- Patients arrive and are treated sequentially.
- Save as many as possible.

A bit of theory

Stochastic Multi-Armed Bandit

K-Armed Stochastic Bandit Problems

- K actions $i \in \{1, \dots, K\}$, outcome $X_t^i \in \mathbb{R}$ (sub-)Gaussian, bounded

$$X_1^i, X_2^i, \dots, \sim \mathcal{N}(\mu^i, 1) \quad \textbf{i.i.d.}$$

- Non-Anticipative Policy: $\pi_t\left(X_1^{\pi_1}, X_2^{\pi_2}, \dots, X_{t-1}^{\pi_{t-1}}\right) \in \{1, \dots, K\}$
- Goal: Maximize expected reward $\sum_{t=1}^{T} \mathbb{E} X_t^{\pi_t} = \sum_{t=1}^{T} \mu^{\pi_t}$
- Performance: Cumulative Regret

$$R_T = \max_{i \in \{1,\dots,K\}} \sum_{t=1}^{T} \mu^i - \sum_{t=1}^{T} \mu^{\pi_t} = \Delta_i \sum_{t=1}^{T} \mathbb{1}\{\pi_t = i \neq \star\}$$

with $\Delta_i = \mu^\star - \mu^i$, the "gap" or cost of error i.

Most Famous algorithm [Auer, Cesa-Bianchi, Fisher, '02]

- UCB - "Upper Confidence Bound"

$$\pi_{t+1} = \arg\max_i \left\{ \overline{X}_t^i + \sqrt{\frac{2\log(t)}{T^i(t)}} \right\},$$

where $T^i(t) = \sum_{t=1}^{t} \mathbb{1}\{\pi_t = i\}$ and $\overline{X}_t^i = \frac{1}{T_t^i} \sum_{s:i_s=i} X_s^i$.

Regret:

$$\mathbb{E}\, R_T \lesssim \sum_k \frac{\log(T)}{\Delta_k}$$

Worst-Case:

$$\mathbb{E}\, R_T \lesssim \sup_\Delta K \frac{\log(T)}{\Delta} \wedge T\Delta$$
$$\approx \sqrt{KT\log(T)}$$

Ideas of proof $\pi_{t+1} = \arg\max_i \left\{ \overline{X}_t^i + \sqrt{\frac{2\log(t)}{T^i(t)}} \right\}$

- 2-lines proof:

$$\pi_{t+1} = i \neq \star \iff \overline{X}_t^\star + \sqrt{\frac{2\log(t)}{T^\star(t)}} \leq \overline{X}_t^i + \sqrt{\frac{2\log(t)}{T^i(t)}}$$

$$\text{“} \Longrightarrow \text{”} \Delta_i \leq \sqrt{\frac{2\log(t)}{T^i(t)}} \Longrightarrow T^i(t) \lesssim \frac{\log(t)}{\Delta_i^2}$$

- Number of mistakes grows as $\frac{\log(t)}{\Delta_i^2}$; each mistake costs Δ_i.

Regret at stage T $\lesssim \sum_i \frac{\log(T)}{\Delta_i^2} \times \Delta_i \approx \sum_i \frac{\log(T)}{\Delta_i}$

- “ $\Longrightarrow$ ” actually happens with overwhelming proba
- “optimal”: no algo always has a regret smaller than $\sum_i \frac{\log(T)}{\Delta_i}$

Other Algos

- ETC [Perchet,Rigollet]. pull in round-robin then eliminate

$$R_T \lesssim \sum_k \frac{\log(T\Delta^k)}{\Delta^k}, \text{ worst case } R_T \leq \sqrt{T\log(K)K}$$

- Other algo, MOSS [Audibert, Bubeck], variants of UCB

$$R_T \lesssim K\frac{\log(T\Delta^{\min}/K)}{\Delta^{\min}}, \text{ worst case } R_T \leq \sqrt{TK}$$

- Infinite number of actions $x \in [0,1]^d$ with $\Delta(x)$ 1 Lipschitz. Discretize + UCB gives

$$R_T \lesssim T\varepsilon + \sqrt{\frac{T}{\varepsilon}} \leq T^{2/3}$$

Adversarial Multi-Armed Bandit

K-Armed Adversarial Bandit Problems

- K actions $i \in [K] = \{1, \ldots, K\}$, outcome $X_t^i \in \mathbb{R}$ bounded in $[0, 1]$

No assumption on $X_1^i, X_2^i, \ldots$

- Non-Anticipative Policy: $\pi_t\left(X_1^{\pi_1}, X_2^{\pi_2}, \ldots, X_{t-1}^{\pi_{t-1}}\right) \in [K]$
- Performance: Cumulative Regret

$$R_T = \max_{i \in [K]} \sum_{t=1}^{T} X_t^i - \sum_{t=1}^{T} X_t^{\pi_t}$$

- Convex optimization of $p \mapsto \mathbb{E}_p \sum_{t=1}^{T} X_t^i$, from $\Delta([K])$ to $[0, 1]$

EXP-algo

- Main insight: $\pi_t \sim p_t \in \Delta([K])$, more weights on best actions

$$p_t^i = \frac{e^{\eta \sum_{s=1}^{t-1} X_s^i}}{\sum_{j \in [K]} e^{\eta \sum_{s=1}^{t-1} X_s^j}}, \quad \eta \text{ is a parameter}$$

- Only $X_t^{\pi_t}$ is observed, not X_t. Estimate X_t by $\widehat{X}_t$

$$\widehat{X}_t^i = 1 - \left(\frac{1 - X_t^i}{p_t^i}\right) \mathbb{1}\{\pi_t = i\} \text{ and run EXP on } \widehat{X}_t$$

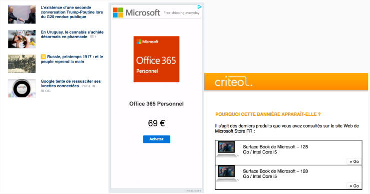
  - $\mathbb{E}\widehat{X}_t^i = 1 - (1 - p_t^i).0 + p_t^i \frac{1 - X_t^i}{p_t^i} = X_t^i$, unbiased estimator
  - $\mathbb{E} \sum_{i \in K} p_t^i (\widehat{X}_t^i)^2 \leq 1 + \sum_{i \in [K]} p_t^i \left(\frac{1 - X_t^i}{p_t^i}\right)^2 p_t^i \leq K + 1$ bounded variance

- Using this estimate we obtain that

$$\mathbb{E} R_T \leq \frac{\log(K)}{\eta} + \eta(K+1)T \leq 3\sqrt{\log(K) K T}$$

Bandits & Repeated Auctions

Back to Repeated Auctions



Ad slot sold by lemonde.fr. **2nd-price auctions**

- Several (marketing) companies places bids
- Highest bid wins (...), say criteo, **pays to lemonde** 2nd bid (...)
- criteo chooses ad of a client, Microsoft or Cdiscount or Booking
- criteo gets **paid by the client** if the user clicks on the ad

Main Problem: Repeated auctions with unknown private valuation

Learn valuations, find which ad to display & **good strategies**

2nd price Auctions

- A good is sold on second price auctions auction.
- Each buyer, with valuation $v^{(i)}$, puts a bet $b^{(i)}$
- The highest bidder wins and **pays second highest bid** $b^\sharp = \max_{i \neq \text{argmax}} b^{(i)}$ (ties broken arbitrarily)

Truthful auctions

optimal strategy bid its own valuation $b^{(i)} = v^{(i)}$

- Utility of bidder : $\left(v^{(i)} - b^\sharp\right)\mathbb{1}\{b^{(i)} \geq b^\sharp\}$
  - if $b^{(i)} > v^{(i)}$ might only pay too much
  - if $b^{(i)} > v^{(i)}$ might loose the auction

Reserve price

- Utility of highest value: $v^{\star} - b^{\sharp}$
- Utility of seller (value v_0): $b^{\sharp} - v_0$, can be negative !

Reserve price

A threshold c: if $b^* \geq c$; price $\max\{b^{\sharp}, c\}$ otherwise not sold

- Still truthful: c is a bid
- **Optimal reserve price** c^* max. $\mathbb{E}(\max\{v^{\sharp}, c\} - v_0)\mathbb{1}\{v^* \geq c\}$
- Depends on the (actually unknown) distributions of value.

Main model

- Learning optimal reserve price [Cesa-Bianchi, Gentile, Mansour]

From the point of view of a bidder ?

- At round $t = 1, \dots, T$:

  bidder bids $b_t \in [0, 1]$

  if $b_t > m_t$ (maximum other bids & reserve price)

  win good, observe value $v_t \in [0, 1]$

- Total utility: $\sum_{t=1}^{T}(v_t - m_t)\mathbb{1}\{b_t > m_t\}$
- Total **regret**:

$$\max_{b \in [0,1]} \sum_{t=1}^{T}(v_t - m_t)\mathbb{1}\{b > m_t\} - \sum_{t=1}^{T}(v_t - m_t)\mathbb{1}\{b_t > m_t\}$$

Data Assumptions - Stochastic vs Adversarial

- **Stochastic:** v_t i.i.d. $\mathbb{E}[v_t] = v \in [0, 1]$

  m_t stochastic (i.i.d. $\mathbb{E}[m_t] = m$), indpt. of v_t

  m_t adversarial (no assumptions), indpt. of v_t

  In both cases, **expected** regret attained at v.

$$\sum_{t=1}^{T} (\mathbf{v} - m_t)\mathbb{1}\{\mathbf{v} > m_t\} - \sum_{t=1}^{T} (\mathbf{v} - m_t)\mathbb{1}\{b_t > m_t\}$$

- **Adversarial:** no assumptions at all on v_t and m_t

Tools that we will use

Variants of stochastic & adversarial **multi-armed bandit**

Stochastic Repeated Auctions

Our policy: UCBid

- Auctions: infinite action space, but with a special structure.
- Round $t+1$ bid

$$b_{t+1} = \min\left(\overline{v}_{\omega_t} + \sqrt{\frac{3\log(t)}{2\omega_t}}, 1\right)$$

where ω_t number of auctions won.

- Our first main result

Theorem - Stochastic case

UCBid yields a regret bound of

$$\mathbb{E}R_T \leq 3 + 12\frac{\log(T)}{\Delta} \wedge 5\sqrt{T\log(T)}$$

where Δ is such that no bid m_t is in the interval $(v, v+\Delta)$

Fully stochastic case: UCBid

- If $m_t \sim \mu$ satisfies margin condition, parameter α (unknown):

Definition - margin condition

$$\forall u > 0,\ \mu\{(v, v+u)\} \le Cu^\alpha \text{ for some constant } C.$$

The bigger α, the easier.

Theorem - Fully stochastic case

$$\mathbb{E}R_T \le \begin{cases} c_\alpha T^{\frac{1-\alpha}{2}} \log^{\frac{1+\alpha}{2}}(T) & \text{if } \alpha < 1 \\ c_\alpha \log^2(T) & \text{if } \alpha = 1 \\ c_\alpha \log(T) & \text{if } \alpha > 1 \end{cases}$$

- Almost matching lower bound

$$\mathbb{E}R_T \ge \begin{cases} c_\alpha T^{\frac{1-\alpha}{2}} & \text{if } \alpha < 1 \\ c_\alpha \log(T) & \text{if } \alpha \ge 1 \end{cases}$$

Adversarial Repeated Auctions

Our policy: EXPTree

$$\max_{b\in[0,1]} \sum_{t=1}^{T} (v_t - m_t)\mathbb{1}\{b > m_t\} - \sum_{t=1}^{T} (v_t - m_t)\mathbb{1}\{b_t > m_t\}$$

- Main idea: **Nested** partitions $\mathcal{P}_t$ of $[0,1]$
  - $\mathcal{P}_t = \{[m^{(s)}, m^{(s+1)}), s = 0, \dots, t-1\}$
  - $m_t \in [m^{(s^*)}, m^{(s^*+1)})$: **split it** into $[m^{(s^*)}, m_t)$ and $[m_t, m^{(s^*+1)})$
- **Weights** of interval $\mathcal{I}$ is $\omega^{\mathcal{I}} = e^{\eta \sum_t \widehat{X}_t^s}$ where $\widehat{X}_{t+1}^s$ is unbiased est. of the value of a bid in $\mathcal{I}$ or in a parent of $\mathcal{I}$.
- At round $t+1$, pick an interval $\mathcal{I}_{t+1}$ in $\mathcal{P}_{t+1}$ with proba proportional to $|\mathcal{I}_{t+1}|\omega_{t+1}$.
- Finally, **bid** b_{t+1} **uniform in** $\mathcal{I}_{t+1}$

Performances of EXPTree

Theorem – Upper-bound

EXPTree yields a regret bounded as

$$\mathbb{E}R_T \leq 4\sqrt{T \log(1/\Delta^\circ)}$$

with Δ° the width of interval contains the best fixed bid.

Is the dependency in Δ° necessary ? yes

Theorem – Lower-bound

For any algo, there exists a sequence of m_t and v_t s.t.

$$\mathbb{E}R_T \geq \frac{1}{32}\sqrt{T \lfloor \log_2(1/2\Delta^\circ) \rfloor}$$

Summary

$$\max_{b\in[0,1]} \sum_{t=1}^{T}(v_t - m_t)\mathbb{1}\{b > m_t\} - \sum_{t=1}^{T}(v_t - m_t)\mathbb{1}\{b_t > m_t\}$$

- v_t stochastic, m_t stochastic: variant of UCB
  - $R_T \lesssim T^{\frac{1-\alpha}{2}} \log(T)^{\frac{1+\alpha}{2}}$
  - Interpolate between $\log(T)$ regret (easy pb), and $\sqrt{T}$ (hard pb)
- v_t stochastic, m_t adversarial: variants of UCB
  - $R_T \lesssim \min\left\{\sqrt{T\log(T)}, \frac{\log(T)}{\Delta}\right\}$
  - Logarithmic regret, even if parts of data are adversarial !
- v_t adversarial, m_t adversarial: variant of Exponential weights
  - $R_T \lesssim \sqrt{T\log(1/\Delta^\circ)}$
  - Same rates as with Δ°-discretization and full info !

Very (quite ?) interesting....

useful as it is?

not really...

Here is a list of reasons

On the basic assumptions

1. Stochastic: Data are not iid, patients are different

   ill-posedness, feature selection/model selection

2. Different Timing: several actions for one reward

   pomdp, learn trade bias/variance

3. Delays: Rewards not received instantaneously

   grouping, evaluations

4. Combinatorial: Several decisions at each stage

   combinatorial optimization, cascading

5. Non-linearity: concave gain, diminishing returns, etc

Few announcements

- Tim Roughgarden (Stanford) is giving a 10h lecture series on

  Data-Driven Optimal Auction Theory

  September 14-21, Polytechnique
- Criteo is organising

  Machine Learning in the Real World #3

  End of November (21 ?), Paris
- For both events (or any other info) do not hesitate !

Investigating (past/present/futur) them

Patients are different

- We assumed (implicitly ?) that all patients/users are identical
- Treatments efficiency (proba of clicks) depend on age, gender...
- Those covariates or **contexts** are observed/known **before** taking the decision of blue/red pill

  The decision (and regret...) should ultimately depend on it

General Model of Contextual Bandits

- **Covariates:** $\omega_t \in \Omega = [0,1]^d$, i.i.d., law μ (equivalent to) λ

  The cookies of a user, the medical history, etc.

- **Decisions:** $\pi_t \in \{1, .., K\}$

  The decision can (should) depend on the context ω_t

- **Reward:** $X_t^k \in [0,1] \sim \nu^k(\omega_t)$, $\mathbb{E}[X^k|\omega] = \mu^k(\omega)$

  The expected reward of action k depend on the context ω

- **Objectives:** Find the best decision given the request

  Minimize regret $R_T := \sum_{t=1}^{T} \mu^{\pi^\star(\omega_t)}(\omega_t) - \mu^{\pi_t}(\omega_t)$

Regularity assumptions

1. Smoothness of the pb: Every μ^k is β-hölder, with $\beta \in (0, 1]$:

$$\exists\, L > 0,\ \forall\, \omega, \omega' \in \mathcal{X},\ \|\mu(\omega) - \mu(\omega')\| \ \leq\ L\|\omega - \omega'\|^{\beta}$$

2. Complexity of the pb: (α-margin condition) $\exists C_0 > 0$,

$$\mathbb{P}_X \left[0 < \left|\mu^1(\omega) - \mu^2(\omega)\right| < \delta\right] \ \leq\ C_0\delta^{\alpha}$$

Regularity assumptions

1. Smoothness of the pb: Every μ^k is β-hölder, with $\beta \in (0, 1]$:

$$\exists\, L > 0,\ \forall\, \omega, \omega' \in \mathcal{X},\ \|\mu(\omega) - \mu(\omega')\| \ \leq\ L\|\omega - \omega'\|^{\beta}$$

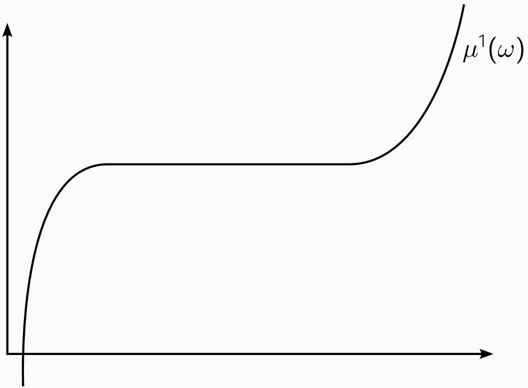
2. Complexity of the pb: (α-margin condition) $\exists C_0 > 0$,

$$\mathbb{P}_X\left[0 < \left|\mu^{\star}(\omega) - \mu^{\sharp}(\omega)\right| < \delta\right] \ \leq\ C_0\delta^{\alpha}$$

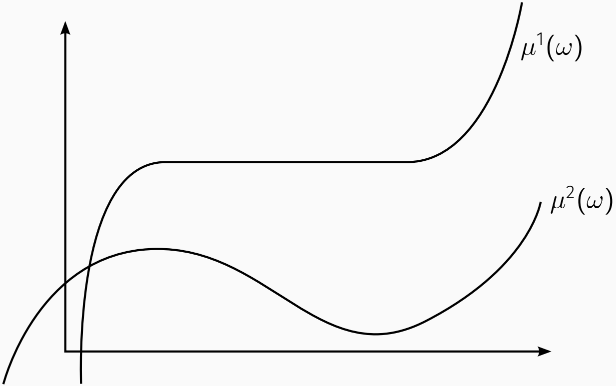
where $\mu^{\star}(\omega) = \max_k \mu^k(\omega)$ is the maximal μ^k and
$\mu^{\sharp}(\omega) = \max\left\{\mu^k(\omega)\, s.t.\, \mu^k(\omega) < \mu^{\star}(\omega)\right\}$ is the second max.

With $K > 2$: $\mu^{\star}$ is β-Hölder but $\mu^{\sharp}$ is not continuous.

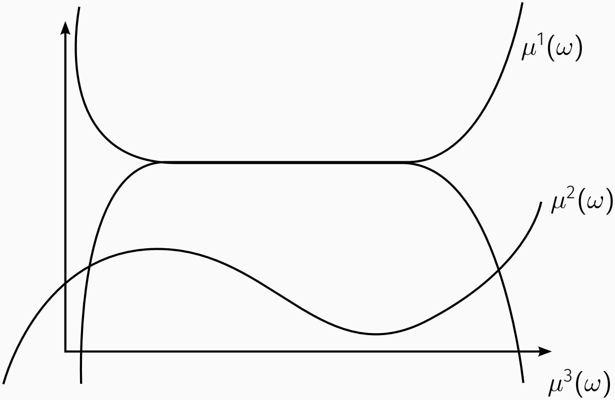
Regularity: an easy example (α big)



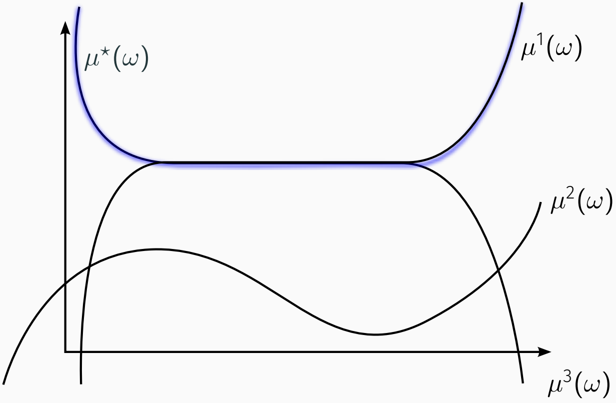
Regularity: an easy example (α big)



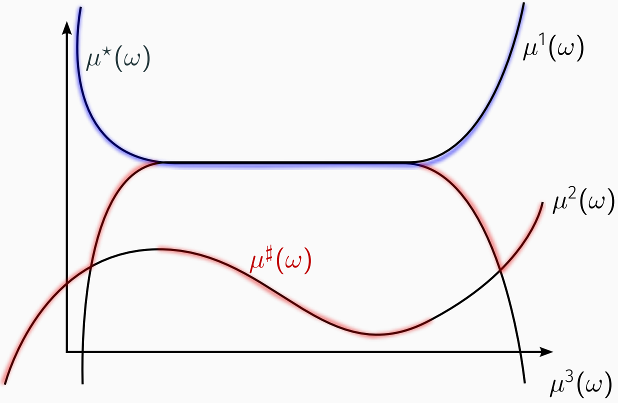
Regularity: an easy example (α big)



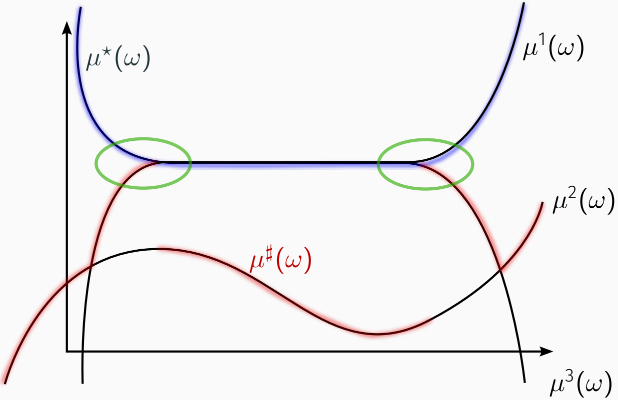
Regularity: an easy example (α big)



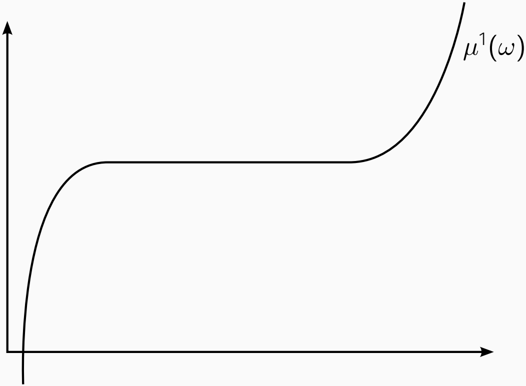
Regularity: an easy example (α big)



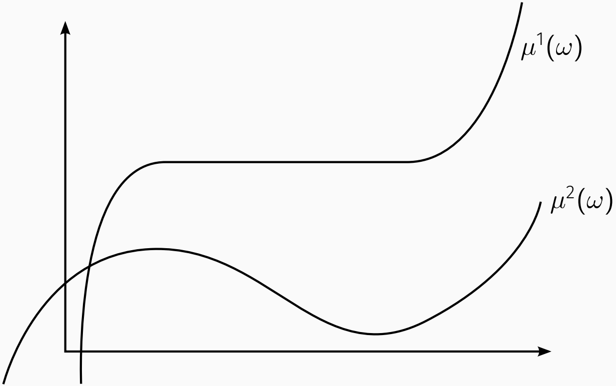
Regularity: an easy example (α big)



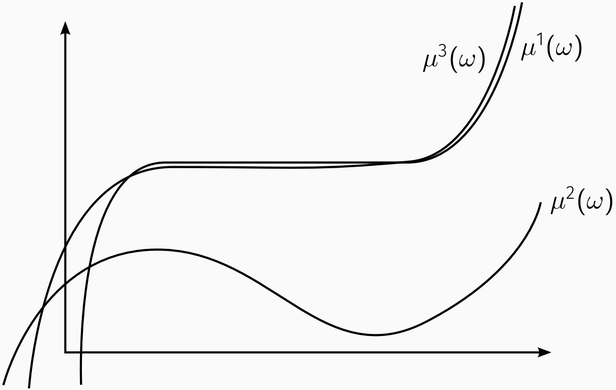
Regularity: a hard example (α small)



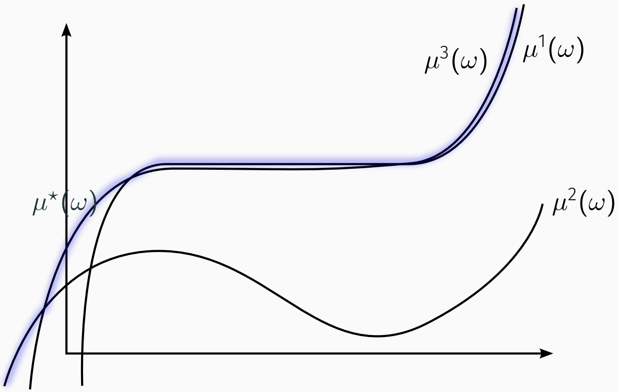
Regularity: a hard example (α small)



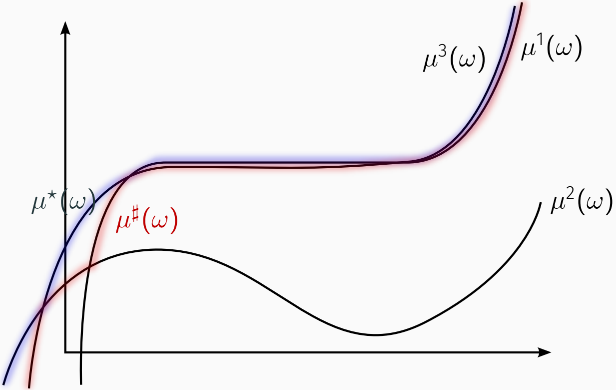
Regularity: a hard example (α small)



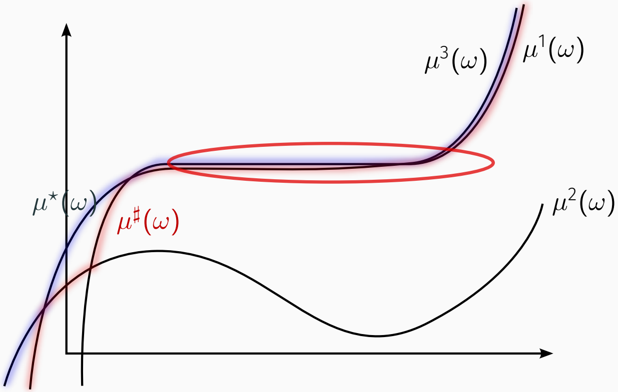
Regularity: a hard example (α small)



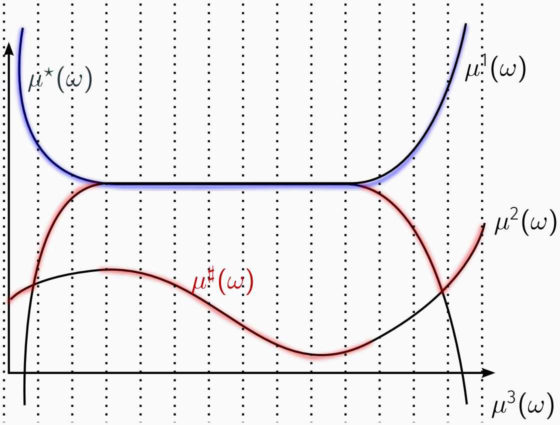
Regularity: a hard example (α small)



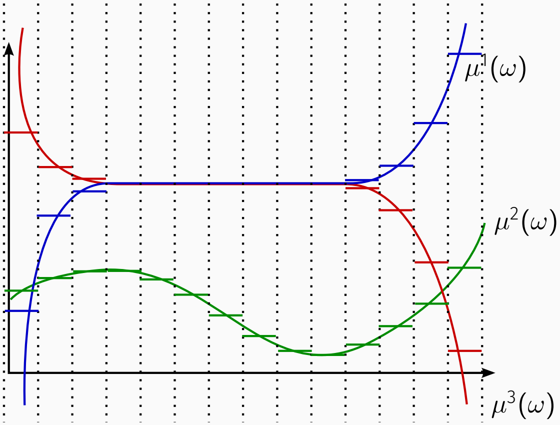
Regularity: a hard example (α small)



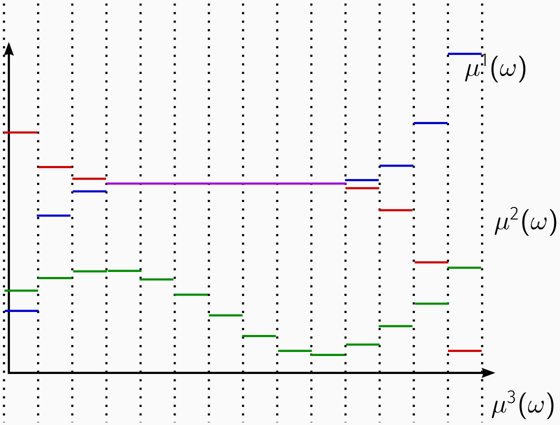
Binned policy



Binned policy



Binned policy



Binned Successive Elimination (BSE)

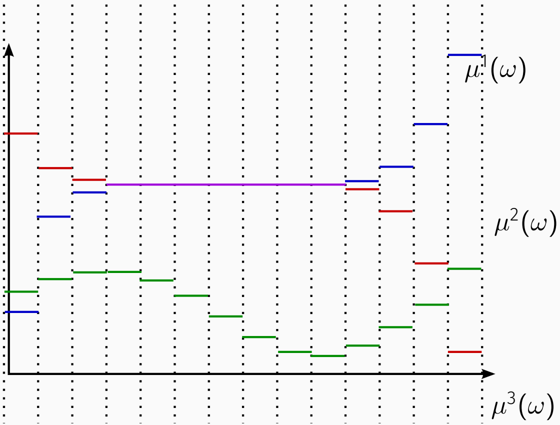
Theorem [P. and Rigollet ('13)]

If $\alpha < 1$, $\mathbb{E}[R_T(\mathrm{BSE})] \lesssim T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$, bin side $\left(\frac{K\log(K)}{T}\right)^{\frac{1}{2\beta+d}}$.

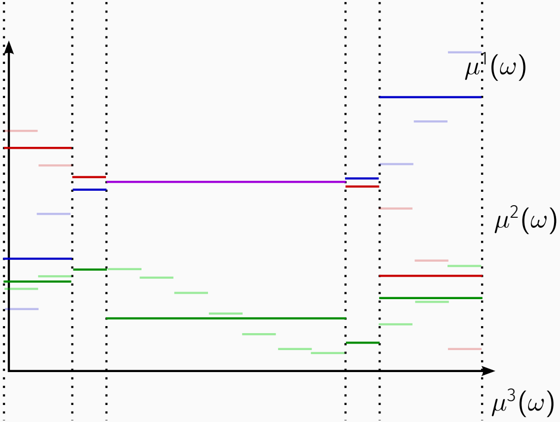
For $K = 2$, matches lower bound: minimax optimal w.r.t. T.

- Same bound with full monit [Audibert and Tsybakov, '07]
- No $\log(T)$: difficulty of nonparametric estimation washes away the effects of exploration/exploitation.
- $\alpha < 1$: cannot attain fast rates for easy problems.
- Adaptive partitioning !

Suboptimality of (BSE) for $\alpha \geq 1$



Suboptimality of (BSE) for $\alpha \geq 1$



Adaptive BSE (ABSE)

Theorem [P. and Rigollet ('13)]

For all α, $\mathbb{E}[R_T(\mathrm{ABSE})] \lesssim T \left(\frac{K \log(K)}{T} \right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$.

For $K = 2$, matches lower bound: minimax optimal w.r.t. T.

- Same bound than (BSE) even for easy problems $\alpha \geq 1$.

This is not the solution

1. **dimensions** dependent bound: $T^{1-\frac{\beta}{2\beta+d}}$

    $d = +\infty$ **and** $\beta = 0$, lots of contexts, no regularity

    Online selection of models ?

    Ill-posed pb $\mu(\cdot)$ not β-holder

    Estimation/Approx errors

    Performance = Approx Error + Regret(β, d, T)

2. Non-stationarity of **arms**: Value are not i.i.d., evolve with time. Ex. ads for movies.

    Cumulative objectives clearly not the solution.

    Discount ? How, why, at which speeds ?

3. Non-stationarity of **sets** of arms:

    Arms arrive and disappears

    How incorporate a new arm ? which index ?

This was really not the solution

1. Non-stationarity of **sets** of arms:

   Arms arrive and disappears

   How incorporate a new arm ? which index ?

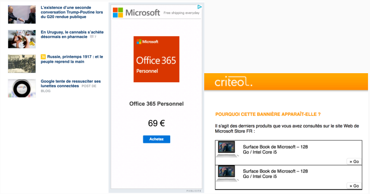
2. Contexts (covariates) are not in $\mathbb{R}^d$

   Rather descriptions, texts, id, images...How to embed ?

   training set is influenced by algorithms...

Different Timing

Example of Repeated Auctions



Ad slot sold by lemonde.fr. **2nd-price auctions**

- Several (marketing) companies places bids
- Highest bid wins (...), say criteo, **pays to lemonde** 2nd bid (...)
- criteo chooses ad of a client, Microsoft or Cdiscount or Boooking
- criteo **paid by the client** if the user clicks on the ad

Main Problem: Repeated auctions with unknown private valuation

Learn valuations, find which ad to display & **good strategies**

Repeated auctions

1. Can be modeled as a bandit pb with **Extra Structure**
2. Actually, Criteo (Google, Facebook) paid if **the user buys something after the click**

Needs several ”costly” auctions to seal a deal

Auctions lost can also help to seal deal (competitor displays ad for free)

Optimal strategy in repeated auctions, **learn it** ? (POMDP ?)

Reward timing per user,
decision timing by opportunities

Other examples - repeated A/B tests

- Companies test new technologies (algo, hardware, etc.) before putting in productions. Sequences of AB tests

  Timing of Decisions: each day, continue, stop or validate the current AB test

  Timing of Rewards: Total improvements of implemented techno.
- The longer AB test are, the more confident (reduces variance) but less and less implementation

Online tradeoff risks/performances

Delays

Rewards are not observed immediately

- Clinical trials: have to wait 6 months to see results.

    A trial length is 3 year : 6 phases

    Regret is still $\sqrt{T}$

- Marketing (ad displays), only see if users buy

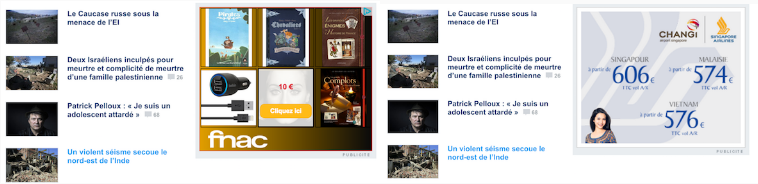
    No feedback is either no sale (forever) or no sale **yet**

    Build estimators with censured/missing data

        Feasible with iid data... but they are not!

Combinatorial Structure

Large Decision spaces



- Choose not to display 1 ad, but 4, 6, 10...
- Paid if sales after click (even if unrelated)

  Lots of correlations (between products, positions, colors/style of banner, **time**, etc.)

  Some products are seen, other are not (carrousels...)
- Too many possibilities of (almost) equal performances

  Compete with the best $R_T \leq \sqrt{KT}$

  but at least top 5%, $R_T \leq \sqrt{\log(K)\frac{1}{5\%}T}$??

Bandit theory is quite neat

To be ”applied”, or relevant, need LOTS of work

Anybody is welcome to join & collaborate!

Model selection, Feature extractions, Missing Data, Censured Data,

Combinatorial Optimization, New techniques estimators..