The multi armed-bandit problem
(with covariates if we have time)

Vianney Perchet & Philippe Rigollet

LPMA
Université Paris – Diderot

ORFE
Princeton University

Algorithms and Dynamics for Games and Optimization

October, 14-18th 2013

LPMA

ORFE Operations Research & Financial Engineering

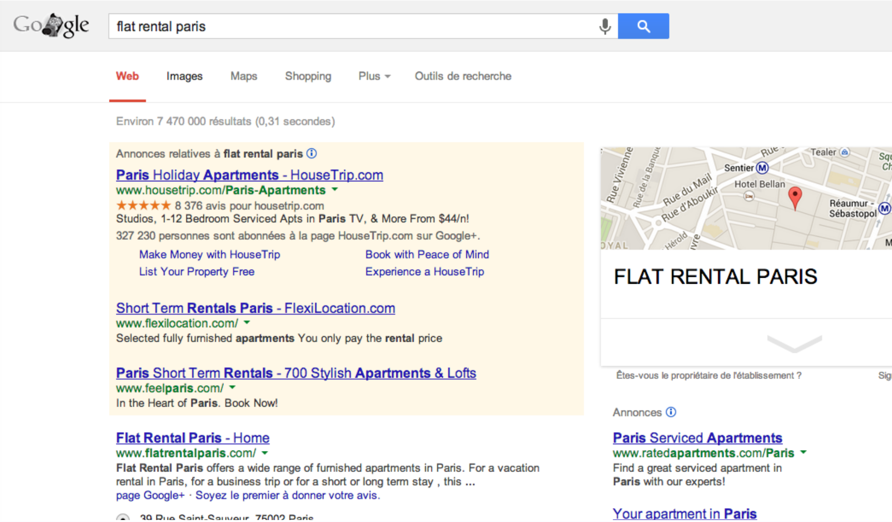
Introduction

Boring and useless definitions:

- **Bandits:** Optimization of a noisy function.
  - Observations: $f(x) + \varepsilon_x$ where ε_x is random variable
  - **Statistics**: lack of information (exploration)
  - **Optimization**: maximize $f(\cdot)$ (exploitation)
  - **Games**: cumulative loss/payoff/reward
- **Covariates:** Some additional side observations gathered
- **Start "easy":** f is maximized over a finite set

Concrete, simple and understandable examples follow.

Some real world examples



Some real world examples

Nightlife Tongoy | Search

Web Show options...

Did you mean: ***Nightlife Cominetti Sorin***

Environ 35 600 résultats (0,41 secondes)

Conseil : Recherchez des résultats uniquement en **français**. Vous pouvez indiquer votre langue de recherche sur la page Préférences.

Tongoy - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Tongoy** - Traduire cette page
Tongoy is a Chilean coastal town in the commune of Coquimbo in Elqui Province, Coquimbo Region. It is located 42 km (26 mi) to the south of Chile's second ...

Villa Chena **Tongoy** - San Bernardo - **Nightlife** | Facebook
https://www.facebook.com/pages/Villa...**Tongoy**/573596072662029
Villa Chena **Tongoy**, San Bernardo. 0 likes · 0 talking about this · 17 were here. Local Business.

Voyages Et Transport **Tongoy** - Foursquare
https://fr.foursquare.com/explore?q...near=**Tongoy**
Recommandations de Foursquare pour Voyages Et Transport dans **Tongoy**. Lieux comme ... Sinon, essaie :food, **nightlife**, coffee, shops, arts, outdoors. Afficher :.

Restaurants **Tongoy** : lire les avis sur des restaurants - **Tongoy**, Chili ...
www.tripadvisor.fr › ... › Chili › Coquimbo Region › Tongoy
Note Restaurants - cuisine Fruits de mer/Poisson à **Tongoy**, Coquimbo ... Restaurants **Tongoy** .... Belambra **Clubs**- Arena Bianca à Propriano, Corse.

Some real world examples

Rechercher | Livres anglais et étrangers | bandit

Recherche détaillée | Nos rubriques | Nouveautés | Meilleures ventes | Bonnes affaires

Livres anglais et étrangers › Relié › Plus de 50 EUR › "bandit"

Résultats 1 - 12 sur 48

1.


Bandits in the Roman Empire: Myth and Reality de Thomas Grunewald et John Drinkwater (**Relié** - 22 avril 2004)
Acheter neuf: **EUR 79,18**
6 neufs à partir de EUR 79,18   2 d'occasion à partir de EUR 71,26
Plus que 2 ex. Commandez vite !
Livraison gratuite possible (voir fiche produit).

2.


Multi-armed Bandit Allocation Indices de John C. Gittins, Richard Weber et Kevin Glazebrook (**Relié** - 11 mars 2011)
Acheter neuf: **EUR 75,18**
10 neufs à partir de EUR 65,45   1 d'occasion à partir de EUR 52,25
Recevez votre article le **mercredi 21 septembre**, si vous commandez dans les **6 heures** et choisissez la livraison en 1 jour ouvré.
Plus que 1 ex. Commandez vite !
Livraison gratuite possible (voir fiche produit).

3.


Bandit Territories: British Outlaw Traditions de Helen Phillips (**Relié** - 25 septembre 2008)
Acheter neuf: **EUR 85,76**
4 neufs à partir de EUR 82,77   1 d'occasion à partir de EUR 91,59
Plus que 3 ex. Commandez vite !
Livraison gratuite possible (voir fiche produit).

4.


Bandits at Sea: A Pirates Reader de C.R. Pennell (**Relié** - 31 août 2000)
Acheter neuf: **EUR 53,33**
3 neufs à partir de EUR 53,33   1 d'occasion à partir de EUR 97,86

Simplified decision problem of Google

- Different firms go to Google and offer

  if you put my ad after the keywords "Flat Rental Paris", every time a customer clicks on it, I'll give you b_i's euros

- A given ad i has some exogenous but **unknown** probability of being clicked p_i.

- Displaying ad i gives in expectation $p_i.b_i$ to Google.

- Objective of Google... maximize cumulated payoff as fast as possible.

Simplified decision problem of Google

- Different firms go to Google and offer

  if you put my ad after the keywords "Flat Rental Paris", every time a customer clicks on it, I'll give you b_i's euros
- A given ad i has some exogenous but **unknown** probability of being clicked p_i.
- Displaying ad i gives in expectation $p_i.b_i$ to Google.
- Objective of Google... maximize cumulated payoff as fast as possible.

Difficulties: The expected revenue of an ad i is unknown; p_i cannot be estimated if ad i is not displayed.

Take risk and display new ads (to compute new and maybe high p_i) or be safe and display the best estimated ad ?

Static bandit – No queries

Structure of a specific instance

- **Decision set:** $\{1, \ldots, K\}$ (the set of "arms" ... ads).
- **Expected payoff** of arm k: $f^k \in [0, 1]$. Best ad $\star$, $f^\star$.
- **Problem difficulty:** $\Delta_k = f^\star - f^k$, $\Delta_{\min} = \min_{\Delta_k > 0} \Delta_k$

Repeated decision problem. At stage $t \in \mathbb{N}$,

- Choose $k_t \in \{1, \ldots, K\}$, receive $Y_t \in [0, 1]$ i.i.d. expectation f^{k_t}
- Observe only the payoff Y_t (and not f^{k_t}) and move to stage $t + 1$

Objectives: maximize cumulative expected payoff or

Minimize regret: $R_T = T.f^\star - \sum_{t=1}^T f^{k_t} = \sum_{t=1}^T \Delta_{k_t}$

Choose the quickest possible the best decision with noise.

Static Case: UCB

Lower bound for K=2: $R_T \geq \square \frac{\log\left(T\Delta_{\min}^2\right)}{\Delta_{\min}}$ with $\Delta_{\min} = \min f^\star - f^k$

Famous algo: **U**pper **C**onfidence **B**ound (and its variants)

Using UCB, $\mathbb{E}[R_T] \leq 8\sum_k \frac{\log(T)}{\Delta_k} \leq 8K\frac{\log(T)}{\Delta_{\min}}$

Static Case: UCB

Lower bound for K=2: $R_T \geq \square \frac{\log\left(T\Delta_{\min}^2\right)}{\Delta_{\min}}$ with $\Delta_{\min} = \min f^\star - f^k$

Famous algo: **U**pper **C**onfidence **B**ound (and its variants)

- Draw each arm $1, .., K$ once and observe $Y_1^1, .., Y_K^K$ (Round 1)
- After stage t, compute the following:
  - $t_k = \sharp\{\tau \leq t;\ k_\tau = k\}$ the number of times arm k was drawn;
  - $\bar{Y}_t^k = \frac{1}{t_k} \sum_{\tau \leq t;\ k_\tau = k} Y_\tau^k$ an estimate of f^k
- Draw the arm $k_{t+1} = \arg\max_k \bar{Y}_t^k + \sqrt{\frac{2\log(t)}{t_k}}$

Using UCB, $\mathbb{E}[R_T] \leq 8 \sum_k \frac{\log(T)}{\Delta_k} \leq 8K \frac{\log(T)}{\Delta_{\min}}$

Remarks on UCB

- **Lower bound for K=2:** $R_T \geq \square \frac{\log\left(T\Delta_{\min}^2\right)}{\Delta_{\min}}$, $\Delta_{\min} = \min_{\Delta_k>0} \Delta_k$
- **UCB algo:**
  - Draw each arm $1, .., K$ once and observe $Y_1^1, .., Y_K^K$ (Round 1)
  - Draw the arm $k_{t+1} = \arg\max_k \bar{Y}_t^k + \sqrt{\frac{2\log(t)}{t_k}}$
- **UCB Upper bound:** $\mathbb{E}[R_T] \leq 8\sum_k \frac{\log(T)}{\Delta_k} \leq 8K\frac{\log(T)}{\Delta_{\min}}$

Remarks:

- Proof based on Hoeffding inequality;
- Not intuitive: clearly suboptimal arms keep being drawn
- MOSS, a variant of UCB, achieves $\mathbb{E}[R_T] \leq \square K \frac{\log(T\Delta_{\min}^2/K)}{\Delta_{\min}}$
- Neither $\log(T)$ or $K\log(T\Delta_{\min}^2/K)$ sufficient with covariates.

Successive Elimination (SE)

Simple policy based on the intuition:

Determine the suboptimal arms, and do not play them.

Time is divided in **rounds** $n \in \mathbb{N}$:

- after round n: eliminate arms (with great proba.) suboptimal

  i.e., arm k s.t. $\bar{Y}_n^k + \sqrt{2\frac{\log(T/n)}{n}} \leq \bar{Y}_n^{k'} - \sqrt{2\frac{\log(T/n)}{n}}$
- at round $n+1$: draw each remaining arm once.

- Easy to describe, to understand (but not to analyse for $K > 2$...), intuitive.
- Simple confidence term (but requires knowledge of T).
- (SE) is a variant of Even-Dar et al. ('06) Auer and Ortner ('10)

Regret of successive elimination

Theorem [P. and Rigollet ('13)]

Played on K arms, the (SE) policy satisfies

$$\mathbb{E}[R_T] \leq \square \min \left\{ \sum_k \frac{\log(T\Delta_k^2)}{\Delta_k}, \sqrt{TK\log(K)} \right\}$$

- UCB: $\sum_k \frac{\log(T)}{\Delta_k}$, MOSS: $K\frac{\log(T\Delta_{\min}^2/K)}{\Delta_{\min}}$
- $\mathbb{E}[R_T] = \sum_k \Delta_k . \mathbb{E}[n_k]$ with n_k the number of draws of arm k
- Exact bound:

Regret of successive elimination

Theorem [P. and Rigollet ('13)]

Played on K arms, the (SE) policy satisfies

$$\mathbb{E}[R_T] \leq \square \min \left\{ \sum_k \frac{\log(T\Delta_k^2)}{\Delta_k}, \sqrt{TK\log(K)} \right\}$$

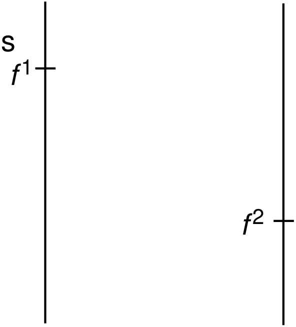
- UCB: $\sum_k \frac{\log(T)}{\Delta_k}$, MOSS: $K\frac{\log(T\Delta_{\min}^2/K)}{\Delta_{\min}}$
- $\mathbb{E}[R_T] = \sum_k \Delta_k.\mathbb{E}[n_k]$ with n_k the number of draws of arm k
- Exact bound:

$$\mathbb{E}[R_T] \leq \min \left\{ 646 \sum_k \frac{1}{\Delta_k} \log\left(\max\left[\frac{T\Delta_k^2}{18}, e\right]\right), 166\sqrt{TK\log(K)} \right\}$$

Successive Elimination: Example

Two arms

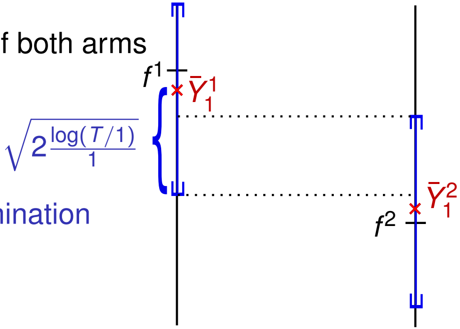
A round: a draw of both arms



Successive Elimination: Example

Two arms

A round: a draw of both arms

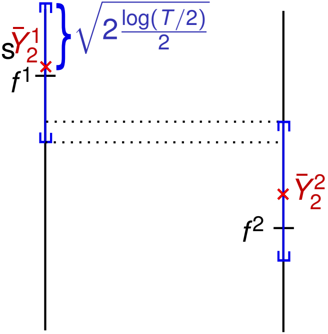


Round 1: no elimination

Successive Elimination: Example

Two arms

A round: a draw of both arms



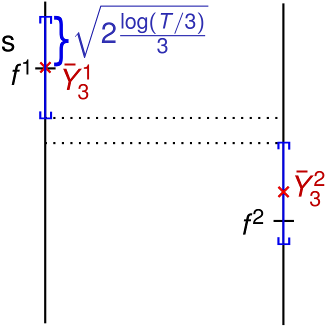
Round 1: no elimination

Round 2: no elimination

Successive Elimination: Example

Two arms

A round: a draw of both arms



Round 1: no elimination

Round 2: no elimination

Round 3: elimination

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$\bar{Y}_n^2 + \sqrt{2\frac{\log(T/n)}{n}} \leq \bar{Y}_n^1 - \sqrt{2\frac{\log(T/n)}{n}}$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^2 + \sqrt{2\frac{\log(T/n)}{n}} \leq f^1 - \sqrt{2\frac{\log(T/n)}{n}}$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}}$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 eliminated before round n_2

$$\mathbb{P}\left(\exists\, n \leq n_2,\ \bar{Y}_n^1 - \bar{Y}_n^2 \leq -2\sqrt{2\frac{\log(T/n)}{n}}\right) \leq \square \frac{n_2}{T}$$

Arm 2 **not eliminated** at round n_2.

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 eliminated before round n_2 (with proba. $\leq \square \frac{n_2}{T}$)

$$\mathbb{P}\left(\exists\, n \leq n_2,\ \bar{Y}_n^1 - \bar{Y}_n^2 \leq -2\sqrt{2\frac{\log(T/n)}{n}}\right) \leq \square \frac{n_2}{T}$$

Arm 2 **not eliminated** at round n_2.

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 **eliminated** before round n_2 (with **proba.** $\leq \square \frac{n_2}{T}$)

Arm 2 not eliminated at round n_2.

$$\mathbb{P}\left(\forall\, n \leq n_2, \bar{Y}_n^2 - \bar{Y}_n^1 \geq -2\sqrt{2\frac{\log(T/n)}{n}}\right)$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 **eliminated** before round n_2 (with **proba.** $\leq \square \frac{n_2}{T}$)

Arm 2 not eliminated at round n_2.

$$\mathbb{P}\left(\bar{Y}^2_{n_2} - \bar{Y}^1_{n_2} \geq -2\sqrt{2\frac{\log(T/n_2)}{n_2}}\right)$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 **eliminated** before round n_2 (with **proba.** $\leq \square \frac{n_2}{T}$)

Arm 2 not eliminated at round n_2.

$$\mathbb{P}\left(\bar{Y}^2_{n_2} - \bar{Y}^1_{n_2} \geq -2\Delta_2\right)$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 **eliminated** before round n_2 (with **proba.** $\leq \square \frac{n_2}{T}$)

Arm 2 not eliminated at round n_2. (with proba. $\leq \square \frac{n_2}{T}$)

$$\mathbb{P}\left([\bar{Y}^1_{n_2} - \bar{Y}^2_{n_2}] - \Delta_2 \leq -\Delta_2\right) \leq \exp\left(-\square n_2 \Delta_2^2\right) \leq \square \frac{n_2}{T}$$

Sketch of proof with $K = 2$

Basic idea: arm 2 (subopt.) eliminated at the first round n s.t.:

$$f^1 - f^2 = \Delta_2 \geq 2\sqrt{2\frac{\log(T/n)}{n}} \qquad n_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2^2}$$

What could go wrong:

Arm 1 **eliminated** before round n_2 (with **proba.** $\leq \square \frac{n_2}{T}$)

Arm 2 **not eliminated** at round n_2. (with **proba.** $\leq \square \frac{n_2}{T}$)

Number of draws of arm 2 (each incurs a regret of Δ_2):

T if something wrong (w.p. $\square \frac{n_2}{T}$), n_2 otherwise (w.p. ≤ 1):

$$\mathbb{E}[R_T] \leq \left[n_2 + \square \frac{n_2}{T} T\right] \Delta_2 \leq \square n_2 \Delta_2 \leq \square \frac{\log(T\Delta_2^2)}{\Delta_2}$$

General Model

Covariates: $X_t \in \mathcal{X} = [0,1]^d$, i.i.d., law μ (equivalent to) λ

- Examples: request received by Amazon or Google
- X_t observed before taking a decision at time $t \in \mathbb{N}$
- Equivalence: two unknown constants $\underline{c}\lambda(A) \leq \mu(A) \leq \overline{c}\lambda(A)$

Decisions: $k_t \in \mathcal{K} = \{1, .., K\}$; construction of a policy π

Payoff: $Y_t^k \in [0,1] \sim \nu^k(X_t)$, $\mathbb{E}[Y^k|X] = f^k(X)$

Objective: regret $R_T := \sum_{t=1}^T f^{\pi^\star(X_t)}(X_t) - f^{k_t}(X_t) \leq o(T)$

General Model

Covariates: $X_t \in \mathcal{X} = [0,1]^d$, i.i.d., law μ (equivalent to) λ

Decisions: $k_t \in \mathcal{K} = \{1, .., K\}$**; construction of a policy** π

- Examples: Choice of the ad to be displayed
- Decision k_t taken after the observation of X_t at time $t \in \mathbb{N}$
- Objectives: Find the best decision given the request

Payoff: $Y_t^k \in [0,1] \sim \nu^k(X_t)$, $\mathbb{E}[Y^k | X] = f^k(X)$

Objective: regret $R_T := \sum_{t=1}^T f^{\pi^\star(X_t)}(X_t) - f^{k_t}(X_t) \leq o(T)$

General Model

Covariates: $X_t \in \mathcal{X} = [0,1]^d$, i.i.d., law μ (equivalent to) λ

Decisions: $k_t \in \mathcal{K} = \{1, .., K\}$; construction of a policy π

Payoff: $Y_t^k \in [0,1] \sim \nu^k(X_t)$, $\mathbb{E}[Y^k|X] = f^k(X)$

- Examples: proba/reward of click on ad k function of the request
- Only $Y_t^{k_t}$ is observed before moving to stage $t+1$;
- Optimization: Find the decision k_t that maximizes $f^k(X_t)$

Objective: regret $R_T := \sum_{t=1}^T f^{\pi^\star(X_t)}(X_t) - f^{k_t}(X_t) \leq o(T)$

General Model

Covariates: $X_t \in \mathcal{X} = [0,1]^d$, i.i.d., law μ (equivalent to) λ

Decisions: $k_t \in \mathcal{K} = \{1, .., K\}$; construction of a policy π

Payoff: $Y_t^k \in [0,1] \sim \nu^k(X_t)$, $\mathbb{E}[Y^k|X] = f^k(X)$

Objective: regret $R_T := \sum_{t=1}^T f^{\pi^\star(X_t)}(X_t) - f^{k_t}(X_t) \leq o(T)$

- Optimal policy: $\pi^\star(X) = \arg\max f^k(X)$; and $f^{\pi^\star(X)}(X) = f^\star(X)$
- Maximize cumulated payoffs $\sum_{t=1}^T f^{k_t}(X_t)$ or minimize regret
- Find a policy π asymptotic. at least as well as $\pi^\star$ (in average)

Regularity assumptions

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

$$\exists\, L > 0,\ \forall\, x, y \in \mathcal{X},\ \|f(x) - f(y)\| \leq L\|x - y\|^{\beta}$$

2. **Complexity of the pb:** (α-margin condition) $\exists \delta_0 > 0$ and $C_0 > 0$

$$\mathbb{P}_X\left[0 < \left|f^1(x) - f^2(x)\right| < \delta\right] \leq C_0\delta^{\alpha}, \quad \forall \delta \in (0, \delta_0)$$

Regularity assumptions

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

$$\exists\, L > 0,\ \forall\, x, y \in \mathcal{X},\ \|f(x) - f(y)\| \leq L\|x - y\|^\beta$$

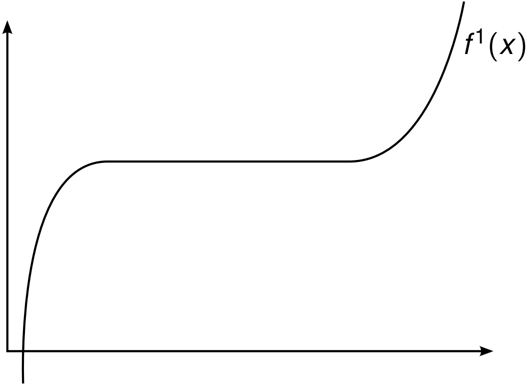
2. **Complexity of the pb:** (α-margin condition) $\exists \delta_0 > 0$ and $C_0 > 0$

$$\mathbb{P}_X\left[0 < \left|f^\star(x) - f^\sharp(x)\right| < \delta\right] \leq C_0 \delta^\alpha, \quad \forall \delta \in (0, \delta_0)$$

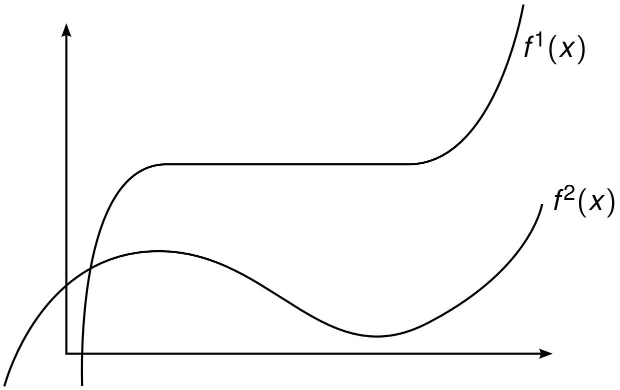
where $f^\star(x) = \max_k f^k(x)$ is the maximal f^k and
$f^\sharp(x) = \max\left\{f^k(x)\, s.t.\, f^k(x) < f^\star(x)\right\}$ is the second max.

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

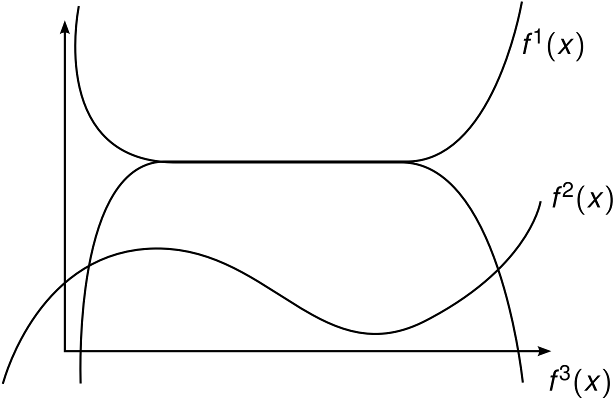
Regularity: an easy example (α big)



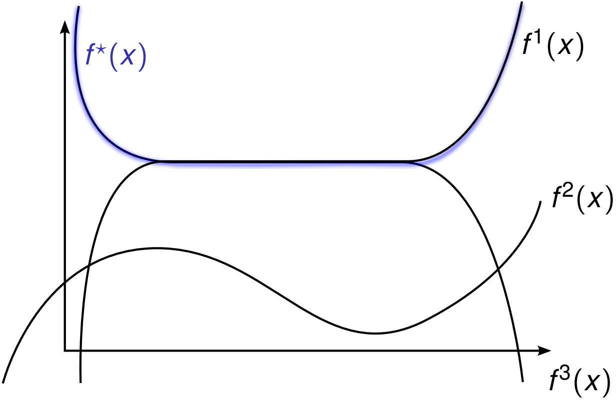
Regularity: an easy example (α big)



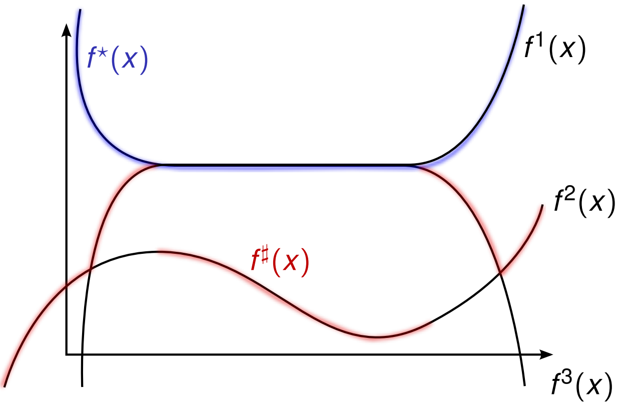
Regularity: an easy example (α big)



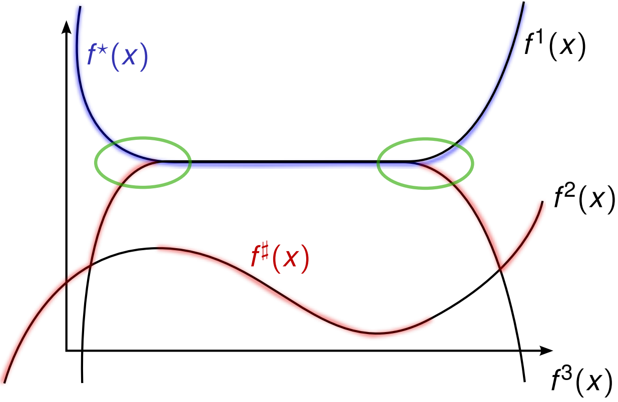
Regularity: an easy example (α big)



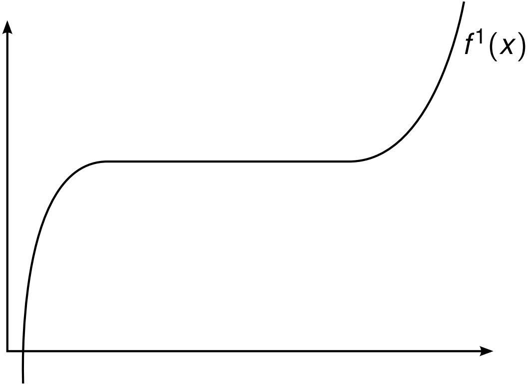
Regularity: an easy example (α big)



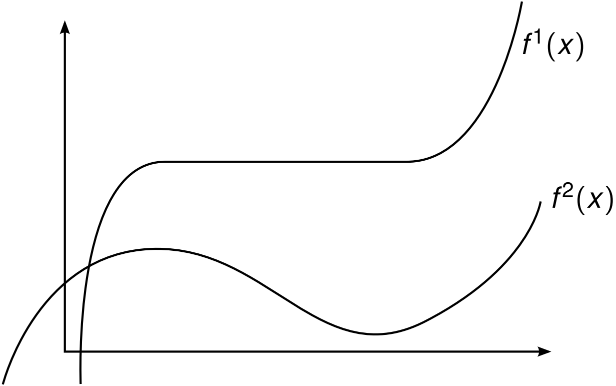
Regularity: an easy example (α big)



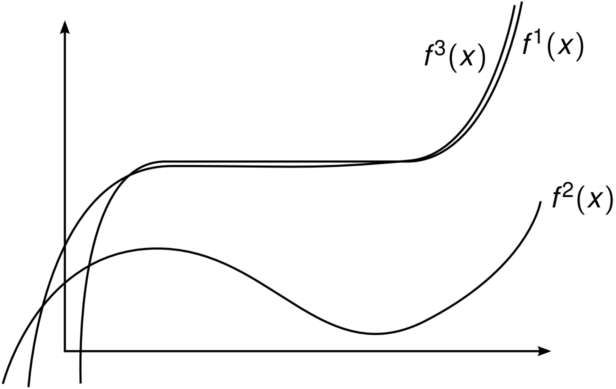
Regularity: a hard example (α small)



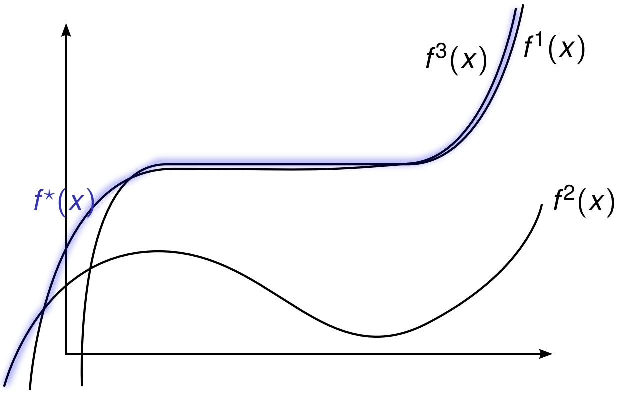
Regularity: a hard example (α small)



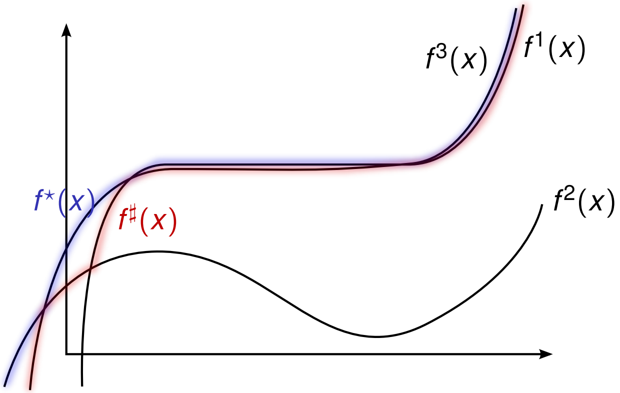
Regularity: a hard example (α small)



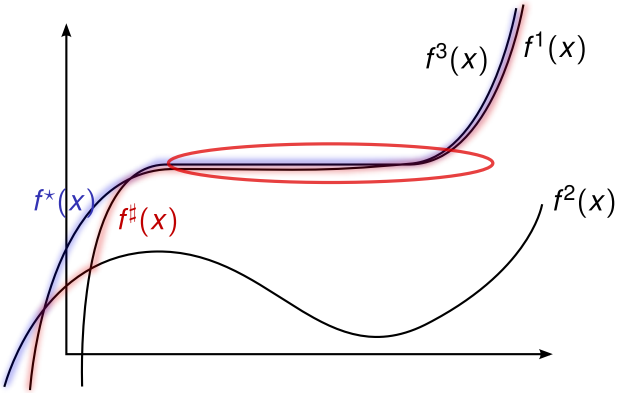
Regularity: a hard example (α small)



Regularity: a hard example (α small)



Regularity: a hard example (α small)



Conflict between α and β

$$\exists \delta_0,\ C_0,\ \mathbb{P}_X\left[0 < f^\star(x) - f^\sharp(x) < \delta\right] \le C_0 \delta^\alpha, \quad \forall \delta \in (0, \delta_0)$$

- First used by Goldenshluger and Zeevi ('08) – case $f^1 = 0$;

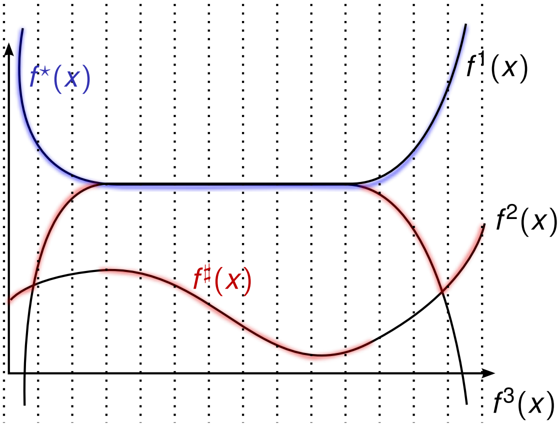
  It was an assumption on the distribution of X only
- Here: fixed marginal (uniform), measures closeness of functions.

Proposition: Conflict α vs. β

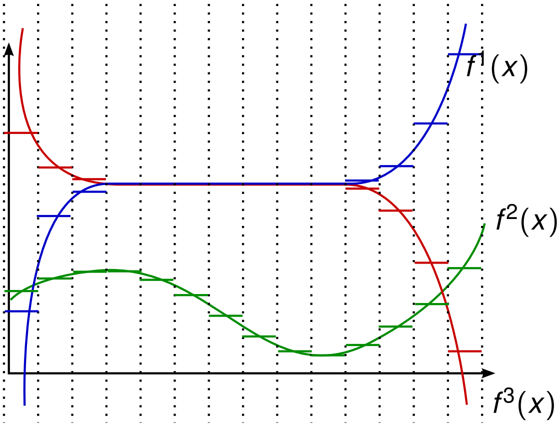
$\alpha\beta > d \implies$ all arms are either always or never optimal

Smoothness β is known, but complexity α is **not** known.

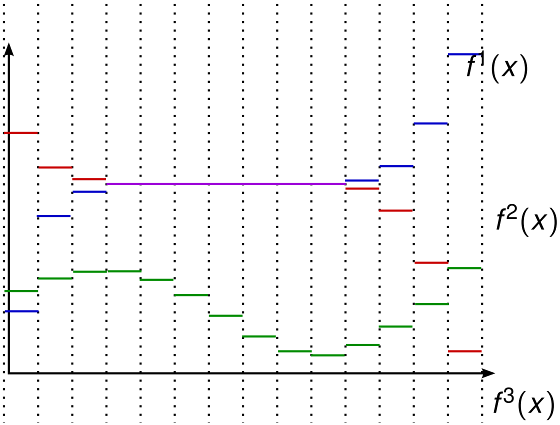
Binned policy



Binned policy



Binned policy



Binned policy

- Consider the uniform partition of $[0,1]^d$ into $1/M^d$ bins

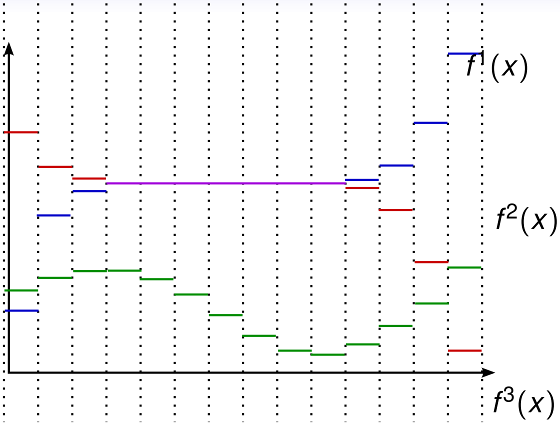
  Bins: hypercube B with side length $|B|$ equal to M.
- Each bin is an independent problem; exact value of X_t discarded
- Average reward of bin B: $\bar{f}_B^k = \frac{\int_B f^k(x)d\mathbb{P}(x)}{\mathbb{P}(B)} \quad (\mathbb{P}(B) \simeq M^d)$

Follow on each bin your favorite static policy.

Reduction to $1/M^d$ static bandits pb. with expected reward $(\bar{f}_B^1, .., \bar{f}_B^K)$.

see Rigollet and Zeevi ('10)

Binned Successive Elimination (BSE)



Binned Successive Elimination (BSE)

Theorem [P. and Rigollet ('11)]

If $0 < \alpha < 1$, $\mathbb{E}[R_T(\text{BSE})] \leq \square T \left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$ with the choice of parameter $M \simeq \left(\frac{K\log(K)}{T}\right)^{\frac{1}{2\beta+d}}$

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

- Same bound can be obtained in the full info. setting (Audibert and Tsybakov, '07)
- No $\log(T)$: difficulty of nonparametric estimation washes away the effects of exploration/exploitation.
- $\alpha < 1$: cannot attain fast rates

Sketch for $K = 2$

Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins ($\Delta_B < M^\beta$)**:**

$$R_{\mathrm{H}} \le M^\beta.\mathbb{P}\,(\text{Hard})\,T \le M^\beta.\mathbb{P}\left(0 < f^\star - f^\sharp < M^\beta\right) T \le TM^{\beta(1+\alpha)}$$

Easy bins ($\Delta_B \not< M^\beta$):

$$\text{with } \Delta_B = \sup_{x \in B} f^\star(x) - f^\sharp(x) \simeq \frac{\int_B f^\star - f^\sharp d\mathbb{P}}{\mathbb{P}(B)}$$

Sketch for $K = 2$

Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins $(\Delta_B < M^\beta)$**:** $R_{\mathrm{H}} \leq TM^{\beta(1+\alpha)} \leq T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

$$R_{\mathrm{H}} \leq M^\beta.\mathbb{P}\,(\text{Hard})\,T \leq M^\beta.\mathbb{P}\left(0 < f^\star - f^\sharp < M^\beta\right)T \leq TM^{\beta(1+\alpha)}$$

Easy bins $(\Delta_B \not< M^\beta)$:

$$\text{with } \Delta_B = \sup_{x \in B} f^\star(x) - f^\sharp(x) \simeq \frac{\int_B f^\star - f^\sharp d\mathbb{P}}{\mathbb{P}(B)}$$

Sketch for $K = 2$

Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins $(\Delta_B < M^\beta)$: $R_{\mathrm{H}} \leq TM^{\beta(1+\alpha)} \leq T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

Easy bins ($\Delta_B \not< M^\beta$**):**

$$R_{\mathrm{E}} \leq \square \sum_{\text{easy}} \frac{\log\left((TM^d)\Delta_B^2\right)}{\Delta_B}$$

$$\text{with } \Delta_B = \sup_{x \in B} f^\star(x) - f^\sharp(x) \simeq \frac{\int_B f^\star - f^\sharp d\mathbb{P}}{\mathbb{P}(B)}$$

Sketch for $K = 2$

Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins $(\Delta_B < M^\beta)$: $R_{\mathrm{H}} \leq TM^{\beta(1+\alpha)} \leq T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

Easy bins ($\Delta_B \geq M^\beta$):

$$R_{\mathrm{E}} \leq \square \sum_{\text{easy}} \frac{\log\left((TM^d)\Delta_B^2\right)}{\Delta_B}$$

Order the Δ_B as $\Delta_1 \leq \Delta_2 \leq \ldots \leq \Delta_{M^{-d}}$ then

$$\forall \ell \in \{1, .., M^{-d}\}, \;\; \ell M^d \leq \mathbb{P}\left(0 < f^\star - f^\sharp < \Delta_\ell\right) \leq \Delta_\ell^\alpha$$

Sketch for $K = 2$

Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins $(\Delta_B < M^\beta)$: $R_{\mathrm{H}} \leq TM^{\beta(1+\alpha)} \leq T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

Easy bins ($\Delta_B \geq M^\beta$):

$$R_{\mathrm{E}} \leq \square \sum_{\ell=M^{\alpha\beta-d}}^{M^{-d}} \frac{\log\left((TM^d)(\ell M^d)^{2/\alpha}\right)}{(\ell M^d)^{1/\alpha}}$$

Order the Δ_B as $\Delta_1 \leq \Delta_2 \leq \ldots \leq \Delta_{M^{-d}}$ then

$$\forall \ell \in \{1, .., M^{-d}\}, \quad \ell M^d \leq \mathbb{P}\left(0 < f^\star - f^\sharp < \Delta_\ell\right) \leq \Delta_\ell^\alpha$$

Sketch for $K = 2$

Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins $(\Delta_B < M^\beta)$: $R_{\mathrm{H}} \le TM^{\beta(1+\alpha)} \le T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

Easy bins ($\Delta_B \ge M^\beta$**):** $R_{\mathrm{E}} \le \square\, TM^{\beta(1+\alpha)} \le \square\, T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

$$R_{\mathrm{E}} \le \square \sum_{\ell=M^{\alpha\beta-d}}^{M^{-d}} \frac{\log\left((TM^d)(\ell M^d)^{2/\alpha}\right)}{(\ell M^d)^{1/\alpha}} \le TM^{\beta(1+\alpha)}$$

because (for $\alpha < 1$):

$$\sum_{\ell=M^{\alpha\beta-d}}^{M^{-d}} \frac{\log\left((TM^d)(\ell M^d)^{2/\alpha}\right)}{(\ell M^d)^{1/\alpha}} \le \frac{\log\left(TM^{2\beta+d}\right)}{M^{d+\beta(1-\alpha)}} \le TM^{\beta(1+\alpha)}$$

Sketch for $K = 2$

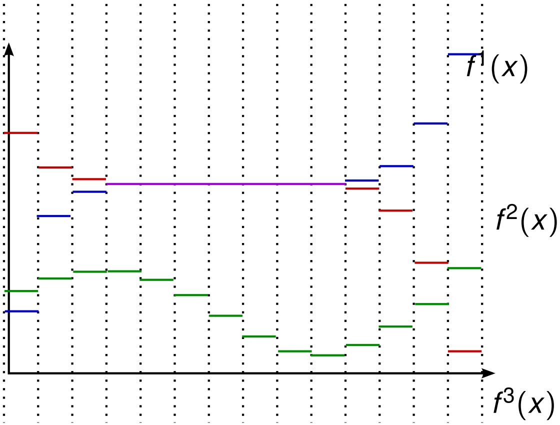
Decomposition of regret: $\mathbb{E}[R_T(\mathrm{BSE})] = R_{\mathrm{H}} + R_{\mathrm{E}}$

Hard bins $(\Delta_B < M^\beta)$: $R_{\mathrm{H}} \leq TM^{\beta(1+\alpha)} \leq T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

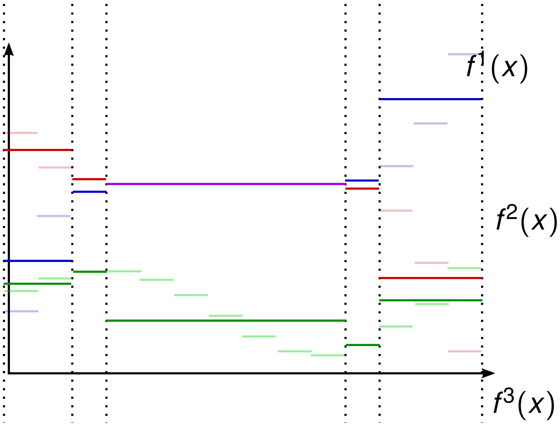
Easy bins $(\Delta_B \not< M^\beta)$: $R_{\mathrm{E}} \leq TM^{\beta(1+\alpha)} \leq T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$

- For $\alpha \geq 1$ additional terms: $\mathbb{E}[R_T]$ multiplied by $\log(T)$.
- We always pay the number of bins (that should be large enough for non-smooth functions)
- Problem is: too many bins. Solution: Online/adaptive construction of the bins.

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



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



Adaptative BSE (ABSE)

Basic idea: Given a bin of size $|B|$ (for $K = 2$):

$$\text{If } \bar{f}^1_B - \bar{f}^2_B \geq \Box |B|^\beta \text{ then } f^1 \geq f^2 \text{ on } B.$$

Adaptively Binned Successive Elimination

Start with $B = [0, 1]$ and $|B|_0 \simeq \left(\frac{K \log(K)}{T}\right)^{\frac{1}{2\beta+d}}$

- Draw samples (in rounds) of arms when covariates are in B;
- If $\bar{Y}^k_n - \bar{Y}^{k'}_n \geq \Box \sqrt{\frac{\log(T|B|^d/n)}{n}} + \Box |B|^\beta$ then eliminate arm k';
- Stop after n_B rounds and split B in two halves (of size $|B|/2$) with

$$\sqrt{\frac{\log(T|B|^d/n_B)}{n_B}} = |B|^\beta \quad \text{and} \quad n_B \simeq \frac{\log(T|B|^{2\beta+d})}{|B|^{2\beta}}$$

- Repeat the procedure on two halves (until $|B| \leq |B|_0$).

Regret of (ABSE)

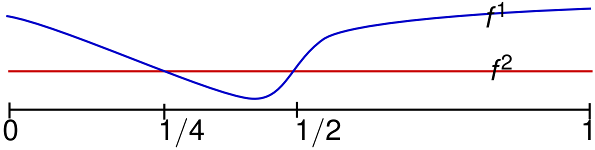
Theorem [P. and Rigollet ('11)]

Fix $\alpha > 0$ and $0 < \beta \leq 1$ then (ABSE) has a regret bounded as

$$\mathbb{E}[R_T(\mathrm{ABSE})] \leq \square T \left(\frac{K \log(K)}{T} \right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$$

- Minimax optimal (Rigollet and Zeevi, 2010. See also Audibert and Tsybakov, 2007)
- Slivkins (2011, COLT): Zooming (abstract setup, complicated algorithm); no real purpose nor measure to adaptive policy.

(ABSE) illustrated



$[0,1]$ $(1 \sim 2)$
keep 1 and 2

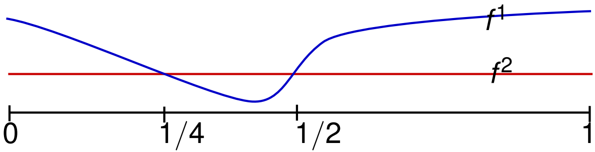
$[0,\frac{1}{2}]$ $(1 \sim 2)$
keep 1 and 2

$[\frac{1}{2},1]$ $(1 \geq 2)$
eliminate 2

$[0,\frac{1}{4}]$ $(1 \geq 2)$
eliminate 2

$[0,\frac{1}{4}]$ $(2 \geq 1)$
eliminate 1

(ABSE) illustrated



eliminate 1 or 2

$[0, 1]$ $(1 \sim 2)$

keep 1 and 2

eliminate 1 or 2

$[0, \frac{1}{2}]$ $(1 \sim 2)$

keep 1 and 2

eliminate 1 or not eliminate 2

$[\frac{1}{2}, 1]$ $(1 \geq 2)$

eliminate 2

eliminate 1 or not eliminate 2

$[0, \frac{1}{4}]$ $(1 \geq 2)$

eliminate 2

$[0, \frac{1}{4}]$ $(2 \geq 1)$

eliminate 1

eliminate 2 or not eliminate 1

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^{\beta}$ (so regret $\leq n_B|B|^{\beta}$).

- **What could go wrong**

Terminal node:

- Eliminate arm 1 or not eliminate arm 2: Same analysis for (SE)
- Happens with proba. less than $\square \frac{n_B}{T|B|^d}$
- Number of times covariates in B less than $\square T|B|^d$
- Regret each time less than $\Delta_B \leq |B|^{\beta}$

Non-terminal node:

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^\beta$ (so regret $\leq n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \leq \Box n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

- Eliminate arm 1 or not eliminate arm 2: Same analysis for (SE)
- Happens with proba. less than $\Box \frac{n_B}{T|B|^d}$
- Number of times covariates in B less than $\Box T|B|^d$
- Regret each time less than $\Delta_B \leq |B|^\beta$

Non-terminal node:

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^\beta$ (so regret $\leq n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \leq \square n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

Non-terminal node:

- Eliminate arm 1 or eliminate arm 2 ($\bar{f}_B^2 \leq \bar{f}_B^1 \leq \bar{f}_B^2 + |B|^\beta$)
- For arm 1, same analysis. For arm 2:

$$\exists\, n \leq n_B,\ \bar{Y}_n^1 - \sqrt{\frac{\log(T|B|^d/n)}{n}} \geq \bar{Y}_n^2 + \sqrt{\frac{\log(T|B|^d/n)}{n}} + |B|^\beta$$

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^\beta$ (so regret $\leq n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \leq \Box n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

Non-terminal node:

- Eliminate arm 1 or eliminate arm 2 ($\bar{f}_B^2 \leq \bar{f}_B^1 \leq \bar{f}_B^2 + |B|^\beta$)
- For arm 1, same analysis. For arm 2:

$$\exists\, n \leq n_B,\ \bar{Y}_n^1 - \bar{Y}_n^2 - \Delta_B \geq 2\sqrt{\frac{\log(T|B|^d/n)}{n}} + |B|^\beta - \Delta_B$$

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^\beta$ (so regret $\leq n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \leq \square n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

Non-terminal node:

- Eliminate arm 1 or eliminate arm 2 ($\bar{f}_B^2 \leq \bar{f}_B^1 \leq \bar{f}_B^2 + |B|^\beta$)
- For arm 1, same analysis. For arm 2:

$$\mathbb{P}\left(\exists\, n \leq n_B,\ \bar{Y}_n^1 - \bar{Y}_n^2 - \Delta_B \geq 2\sqrt{\frac{\log(T|B|^d/n)}{n}}\right) \leq \frac{n_B}{T|B|^d}$$

(ABSE) Sketch of proof

- **If everything goes right:**
When a bin B is reach, one has $\Delta_B \le |B|^\beta$ (so regret $\le n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \le \square n_B|B|^\beta \le \log(T|B|^{2\beta+d})|B|^{-\beta}$

Non-terminal node: $R_B \le \square n_B|B|^\beta \le \log(T|B|^{2\beta+d})|B|^{-\beta}$

- Eliminate arm 1 or eliminate arm 2 ($\bar{f}_B^2 \le \bar{f}_B^1 \le \bar{f}_B^2 + |B|^\beta$)
- For arm 1, same analysis. For arm 2:

$$\mathbb{P}\left(\exists\, n \le n_B,\ \bar{Y}_n^1 - \bar{Y}_n^2 - \Delta_B \ge 2\sqrt{\frac{\log(T|B|^d/n)}{n}}\right) \le \frac{n_B}{T|B|^d}$$

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^\beta$ (so regret $\leq n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \leq \square n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

Non-terminal node: $R_B \leq \square n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

$N_\ell =$number of bins of size $|B| = 2^{-\ell}$ (and $2^{\ell_0} = |B|_0$):

$$N_\ell.2^{-\ell d} \leq \mathbb{P}\left(0 < f^\star - f^\sharp < 2^{-\ell\beta}\right) \leq 2^{-\ell\alpha\beta} \quad \text{and}$$

$$\mathbb{E}[R_T] \leq \sum_B n_B|B|^\beta \leq \sum_{\ell=0}^{\ell_0} 2^{\ell(d-\alpha\beta)} \log\left(T2^{-\ell(2\beta+d)}\right) 2^{\ell\beta}$$

(ABSE) Sketch of proof

- **If everything goes right:**

When a bin B is reach, one has $\Delta_B \leq |B|^\beta$ (so regret $\leq n_B|B|^\beta$).

- **What could go wrong**

Terminal node: $R_B \leq \square n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

Non-terminal node: $R_B \leq \square n_B|B|^\beta \leq \log(T|B|^{2\beta+d})|B|^{-\beta}$

N_ℓ =number of bins of size $|B| = 2^{-\ell}$ (and $2^{\ell_0} = |B|_0$):

$$N_\ell.2^{-\ell d} \leq \mathbb{P}\left(0 < f^\star - f^\sharp < 2^{-\ell\beta}\right) \leq 2^{-\ell\alpha\beta} \quad \text{and}$$

$$\mathbb{E}[R_T] \leq \square T\left(\frac{K\log(K)}{T}\right)^{\frac{\beta(1+\alpha)}{2\beta+d}}$$

Conclusion

We introduced and analyzed new policies:

- Sequential Elimination: an intuitive policy with great potential for the static case;
- Binned SE: its generalization for **hard** dynamic pb;
- Adaptively BSE: again generalized for both **easy** and hard pb.

– There are all minimax optimal in T;

– Conjecture: also in K up to the term $\log(K)$.

– They require the knowledge of T (OK) and β (more arguable)

– Analysis more intricate when $K > 2$: optimal arm can be eliminated more easily, $f^\sharp$ non continuous

– **Future work:** adaptive policy w.r.t. β