The information complexity of sequential resource allocation

Emilie Kaufmann,

joint work with Olivier Cappé, Aurélien Garivier and Shivaram Kalyanakrishan

SMILE Seminar, ENS, June 8th, 2015

Sequential allocation : some examples

Clinical trial

- K possible treatments (with unknown effect)

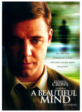




- Which treatment should be allocated to each patient based on their effect on previous patients ?

Movie recommendation

- K different movies



- Which movie should be recommended to each user, based on the ratings given by previous (similar) users ?

The "bandit" framework

One-armed bandit
= slot machine (or arm)

<u>Multi-armed bandit :</u> several arms.
Drawing arm $a \Leftrightarrow$ observing a sample from a distribution ν_a, with mean μ_a

Best arm $a^* = \text{argmax}_a \ \mu_a$

Which arm should be drawn based on the previous observed outcomes ?

Bandit model (more formal)

A **multi-armed bandit model** is a set of K arms where

- Each arm a is a probability distribution ν_a of mean μ_a
- Drawing arm a is observing a realization of ν_a
- Arms are assumed to be independent

At round t, an agent

- chooses arm A_t, and observes $X_t \sim \nu_{A_t}$
- (A_t) is his **strategy** or **bandit algorithm**, such that

$$A_{t+1} = F_t(A_1, X_1, \ldots, A_t, X_t)$$

Global objective : Learn which arm(s) have highest mean(s)

$$\mu^* = \max_a \ \mu_a \qquad a^* = \text{argmax}_a \ \mu_a$$

Objective 1 : Regret minimization

Samples are seen as **rewards**.

The agent ajusts (A_t) to

- maximize the (expected) sum of rewards accumulated,

$$\mathbb{E}\left[\sum_{t=1}^{T} X_t\right]$$

- or equivalently minimize his *regret* :

$$R_T = \mathbb{E}\left[T\mu^* - \sum_{t=1}^{T} X_t\right]$$

$\Rightarrow$ **exploration/exploitation tradeoff**

Objective 2 : Best arm identification

The agent has to **identify the best arm** a^*. (no loss when drawing "bad" arms)

To do so, he

- uses a sampling strategy (A_t)
- stops sampling the arms at some (random) time τ
- recommends an arm $\hat{a}_\tau$

His goal :

Fixed-budget setting	Fixed-confidence setting
$\tau = T$ minimize $\mathbb{P}(\hat{a}_\tau \neq a^*)$	minimize $\mathbb{E}[\tau]$ $\mathbb{P}(\hat{a}_\tau \neq a^*) \leq \delta$

$\Rightarrow$ **optimal exploration**

Comparison on the medical trials example

The doctor :

- chooses treatment A_t to give to patient t
- observes whether the patient is cured : $X_t \sim \mathcal{B}(\mu_{A_t})$

He can ajust his strategy (A_t) so as to

Regret minimization	Best arm identification
Maximize the number of patients cured among T patients	Identify the best treatment with probability at least $1-\delta$ (to always give this one later)

Outline

1. Regret minimization
2. m best arms identification
   - Lower bound on the sample complexity
   - An optimal algorithm ?
3. The complexity of A/B Testing
   - The Gaussian case
   - The Bernoulli case

A parametric assumption on the arms

$\nu_1, \ldots, \nu_K$ belong to a one-dimensional exponential family :

$\mathcal{P}_{\lambda,\Theta,b} = \{\nu_\theta, \theta \in \Theta : \nu_\theta$ has density $f_\theta(x) = \exp(\theta x - b(\theta))$ $w.r.t.$ $\lambda\}$

Example : Gaussian, Bernoulli, Poisson distributions...

- $\nu_k = \nu_{\theta_k}$ can also be parametrized by its mean $\mu_k = \dot{b}(\theta_k)$.

Notation : Kullback-Leibler divergence

$$\mathrm{KL}(p, q) = \mathbb{E}_{X \sim p}\left[\log \frac{dp}{dq}(X)\right]$$

For a given exponential family $\mathcal{P}$, we denote by

$$d_{\mathcal{P}}(\mu, \mu') := \mathrm{KL}(\nu_{\dot{b}^{-1}(\mu)}, \nu_{\dot{b}^{-1}(\mu')})$$

the KL divergence between the distributions of mean μ and μ'.

Example : Bernoulli distributions

$$d(\mu, \mu') = \mathrm{KL}(\mathcal{B}(\mu), \mathcal{B}(\mu')) = \mu \log \frac{\mu}{\mu'} + (1 - \mu) \log \frac{1 - \mu}{1 - \mu'}.$$

Outline

1. Regret minimization
2. m best arms identification
   - Lower bound on the sample complexity
   - An optimal algorithm ?
3. The complexity of A/B Testing
   - The Gaussian case
   - The Bernoulli case

Optimal algorithms for regret minimization

$$\nu = \nu_\theta = (\nu_{\theta_1}, \dots, \nu_{\theta_K}) \in \mathcal{M} = (\mathcal{P})^K.$$

$N_a(t)$: number of draws of arm a up to time t

$$R_T(\nu) = \sum_{a=1}^{K} (\mu^* - \mu_a) \mathbb{E}_\nu[N_a(T)]$$

- consistent algorithm : $\forall \nu \in \mathcal{M}, \forall \alpha \in]0,1[$, $R_T(\nu) = o(T^\alpha)$
- [Lai and Robbins 1985] : every consistent algorithm satisfies

$$\mu_a < \mu^* \Rightarrow \liminf_{T\to\infty} \frac{\mathbb{E}_\nu[N_a(T)]}{\log T} \geq \frac{1}{d(\mu_a, \mu^*)}$$

Definition

A bandit algorithm is **asymptotically optimal** if, for every $\nu \in \mathcal{M}$,

$$\mu_a < \mu^* \Rightarrow \limsup_{T\to\infty} \frac{\mathbb{E}_\nu[N_a(T)]}{\log T} \leq \frac{1}{d(\mu_a, \mu^*)}$$

Towards asymptotically optimal algorithms

- A UCB-type algorithm chooses at time $t+1$

$$A_{t+1} = \underset{a}{\arg\max}\ UCB_a(t)$$

where $UCB_a(t)$ is some upper confidence bound on μ_a.

Examples for binary bandits (Bernoulli distributions)

- UCB1 [Auer et al. 02] uses Hoeffding bounds :

$$UCB_a(t) = \frac{S_a(t)}{N_a(t)} + \sqrt{\frac{2\log(t)}{N_a(t)}}.$$

$S_a(t)$: sum of rewards from arm a up to time t

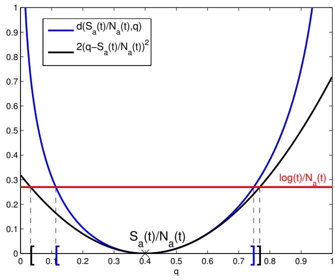
$$\mathbb{E}_\nu[N_a(T)] \leq \frac{K_1}{2(\mu^* - \mu_a)^2} \log T + K_2, \quad \text{with } K_1 > 1.$$

KL-UCB : an asymptotically optimal algorithm

- KL-UCB [Cappé et al. 2013] uses the index :

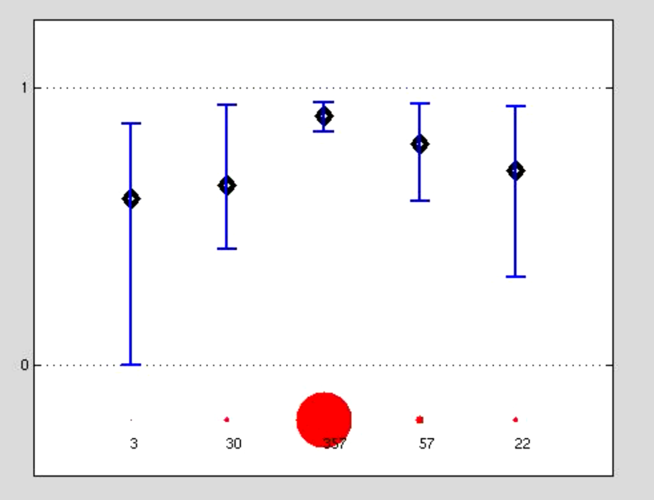
$$u_a(t) = \underset{x > \frac{S_a(t)}{N_a(t)}}{\operatorname{argmax}} \left\{ d\left(\frac{S_a(t)}{N_a(t)}, x\right) \leq \frac{\log(t) + c \log\log(t)}{N_a(t)} \right\},$$

where $d(p,q) = \mathrm{KL}\left(\mathcal{B}(p), \mathcal{B}(q)\right) = p \log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right)$.

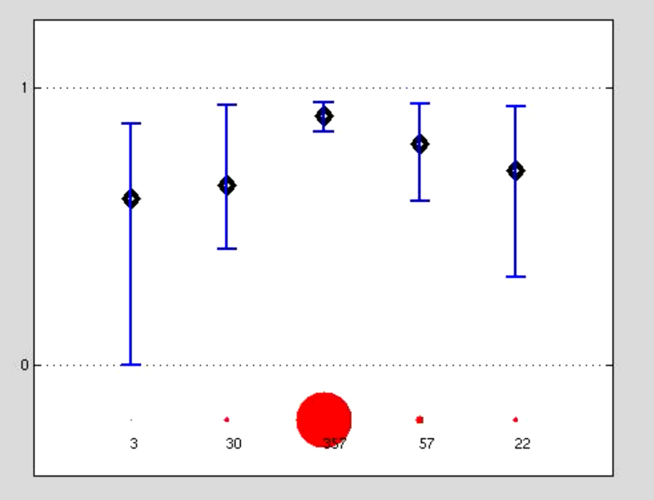


$$\mathbb{E}[N_a(T)] \leq \frac{1}{d(\mu_a, \mu^*)} \log T + O(\sqrt{\log(T)})$$

KL-UCB in action



KL-UCB in action



The information complexity of regret minimization

Letting

$$\kappa_R(\nu) := \inf_{\mathcal{A} \text{ consistent}} \liminf_{T\to\infty} \frac{R_T(\nu)}{\log(T)},$$

we showed that

$$\kappa_R(\nu) = \sum_{a=1}^{K} \frac{(\mu^* - \mu_a)}{d(\mu_a, \mu^*)}.$$

Remarks :

- an asymptotic notion of optimality
- still worth fighting for more efficient algorithms (e.g. Bayesian algorithms)

Outline

m best arms identification (fixed-confidence setting)

$\nu = (\nu_{\theta_1}, \dots, \nu_{\theta_K}) \in \mathcal{M} = (\mathcal{P})^K$ such that $\mu_{[m]} > \mu_{[m+1]}$.

Parameters and notation :

- m a fixed number of arms
- $\delta \in]0, 1[$ a risk parameter
- $\mathcal{S}_m^*$ the set of m arms with highest means

The agent's strategy : $\mathcal{A} = (A_t, \tau, \hat{S})$

- sampling rule : A_t arm chosen at time t
- stopping rule : at time τ he stops sampling the arms
- recommendation rule : a guess $\hat{\mathcal{S}}$ for the m best arms

His goal :

- $\forall \nu \in \mathcal{M} : \mu_{[m]} > \mu_{[m+1]}, \ \mathbb{P}_\nu(\hat{\mathcal{S}} = \mathcal{S}_m^*) \geq 1 - \delta$
  (the algorithm is δ-PAC on $\mathcal{M}$)
- The sample complexity $\mathbb{E}_\nu[\tau]$ is small

Towards the complexity of best-arm identification

The literature presents δ-PAC algorithm such that

$$\mathbb{E}_\nu[\tau] \leq C H(\nu) \log(1/\delta)$$

[Even-Dar et al. 06], [Kalyanakrishnan et al.12]
but no lower bound on $\mathbb{E}_\nu[\tau]$.

$\Rightarrow$ **No notion of optimal algorithm**

We propose

- a lower bound on $\mathbb{E}_\nu[\tau]$
- new algorithms (close to) reaching the lower bound

Outline

1. Regret minimization
2. *m* best arms identification
   - Lower bound on the sample complexity
   - An optimal algorithm ?
3. The complexity of A/B Testing
   - The Gaussian case
   - The Bernoulli case

A general lower bound

$\nu \in \mathcal{M}$ such that $\mu_1 \geq \cdots \geq \mu_m > \mu_{m+1} \geq \cdots \geq \mu_K$.

Theorem [K.,Cappé, Garivier 14]

Any algorithm that is δ-PAC on $\mathcal{M}$ satisfies, for all $\delta \in]0,1[$,

$$\mathbb{E}_\nu[\tau] \geq \left(\sum_{a=1}^{m} \frac{1}{d(\mu_a, \mu_{m+1})} + \sum_{a=m+1}^{K} \frac{1}{d(\mu_a, \mu_m)} \right) \log \left(\frac{1}{2.4\delta} \right).$$

- First lower bound for $m > 1$
- Involves information-theoretic quantities

A general lower bound

$\nu \in \mathcal{M}$ such that $\mu_1 \geq \cdots \geq \mu_m > \mu_{m+1} \geq \cdots \geq \mu_K$.

Theorem [K.,Cappé, Garivier 14]

Any algorithm that is δ-PAC on $\mathcal{M}$ satisfies, for all $\delta \in]0,1[$,

$$\mathbb{E}_\nu[\tau] \geq \left(\sum_{a=1}^{m} \frac{1}{d(\mu_a, \mu_{m+1})} + \sum_{a=m+1}^{K} \frac{1}{d(\mu_a, \mu_m)} \right) \log\left(\frac{1}{2.4\delta}\right).$$

- First lower bound for $m > 1$
- Involves information-theoretic quantities

$$\mathbb{E}[\tau] = \sum_{a=1}^{K} \mathbb{E}[N_a(\tau)]$$

Behind the lower bound : changes of distribution

Lemma [K., Cappé, Garivier 2014]

$\nu = (\nu_1, \nu_2, \dots, \nu_K)$, $\nu' = (\nu'_1, \nu'_2, \dots, \nu'_K)$ two bandit models.

$$\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(\tau)]\mathrm{KL}(\nu_a, \nu'_a) \geq \sup_{\mathcal{E} \in \mathcal{F}_\tau} \mathrm{kl}(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})).$$

with $\mathrm{kl}(x, y) = x \log(x/y) + (1-x)\log((1-x)/(1-y))$.

Behind the lower bound : changes of distribution

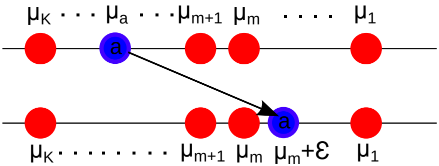
Lemma [K., Cappé, Garivier 2014]

$\nu = (\nu_1, \nu_2, \dots, \nu_K)$, $\nu' = (\nu'_1, \nu'_2, \dots, \nu'_K)$ two bandit models.

$$\sum_{a=1}^{K} \mathbb{E}_\nu[N_a(\tau)] \mathrm{KL}(\nu_a, \nu'_a) \geq \sup_{\mathcal{E} \in \mathcal{F}_\tau} \mathrm{kl}(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})).$$

with $\mathrm{kl}(x, y) = x \log(x/y) + (1-x) \log((1-x)/(1-y))$.

① choose ν' such that $\mathcal{S}^*_m(\nu') \neq \{1, \dots, m\}$:



$$\mathcal{S}^*_m(\nu) = \{1, \cdots, m-1, m\}$$

$$\mathcal{S}^*_m(\nu') = \{1, \cdots, m-1, a\}$$

② $\mathcal{E} = (\hat{S} = \mathcal{S}^*_m(\nu))$: $\mathbb{P}_\nu(\mathcal{E}) \geq 1 - \delta$ and $\mathbb{P}_{\nu'}(\mathcal{E}) \leq \delta$.

$$\Rightarrow \mathbb{E}_\nu[N_a(\tau)] d(\mu_a, \mu_m + \epsilon) \geq kl(\delta, 1-\delta) \geq \log(1/2.4\delta).$$

Outline

1. Regret minimization
2. *m* best arms identification
   - Lower bound on the sample complexity
   - An optimal algorithm ?
3. The complexity of A/B Testing
   - The Gaussian case
   - The Bernoulli case

The KL-LUCB algorithm

Generic notation :

- confidence interval (C.I.) on the mean of arm a at round t :

$$\mathcal{I}_a(t) = [L_a(t), U_a(t)]$$

- $J(t)$ the set of m arms with highest empirical means

Our contribution : Introduce KL-based confidence intervals

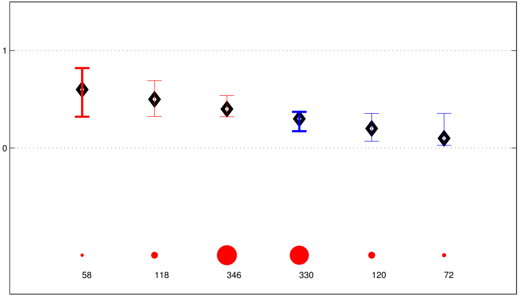
$$\begin{aligned} U_a(t) &= \max \{q \geq \hat{\mu}_a(t) : N_a(t) d(\hat{\mu}_a(t), q) \leq \beta(t,\delta)\} \\ L_a(t) &= \min \{q \leq \hat{\mu}_a(t) : N_a(t) d(\hat{\mu}_a(t), q) \leq \beta(t,\delta)\} \end{aligned}$$

for $\beta(t,\delta)$ some exploration rate.

The KL-LUCB algorithm

At round t, the algorithm :

- draws two well-chosen arms : u_t and l_t (in bold)
- stops when C.I. for arms in $J(t)$ and $J(t)^c$ are separated



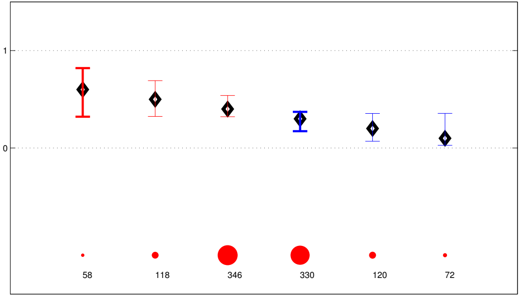
$m = 3, K = 6$

Set $J(t)$, arm l_t in bold Set $J(t)^c$, arm u_t in bold

The KL-LUCB algorithm

At round t, the algorithm :

- draws two well-chosen arms : u_t and l_t (in bold)
- stops when C.I. for arms in $J(t)$ and $J(t)^c$ are separated

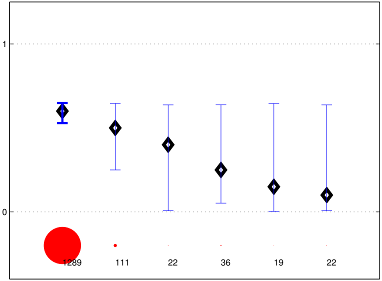


$$m = 3, K = 6$$

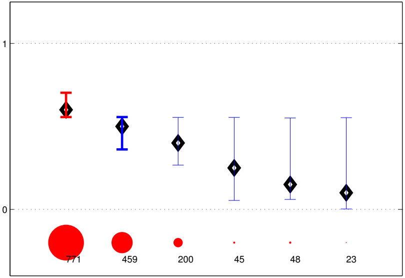
Set $J(t)$, arm l_t in bold Set $J(t)^c$, arm u_t in bold

Remark : KL-UCB versus KL-LUCB

Similar tools for a different behavior :



KL-UCB



KL-LUCB
$(m = 1)$

Theoretical guarantees

Theorem [K.,Kalyanakrishnan 2013]

KL-LUCB using the exploration rate

$$\beta(t,\delta) = \log\left(\frac{k_1 K t^{\alpha}}{\delta}\right),$$

with $\alpha > 1$ and $k_1 > 1 + \frac{1}{\alpha - 1}$ satisfies $\mathbb{P}_\nu(\hat{\mathcal{S}} = \mathcal{S}_m^*) \geq 1 - \delta$.
For $\alpha > 2$,

$$\mathbb{E}_\nu[\tau] \leq 4\alpha H^* \log\left(\frac{1}{\delta}\right) + o_{\delta \to 0}\left(\log \frac{1}{\delta}\right),$$

with

$$H^* = \min_{c \in [\mu_{m+1};\mu_m]} \sum_{a=1}^{K} \frac{1}{d^*(\mu_a, c)}.$$

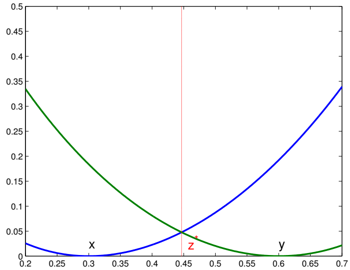
Theoretical guarantees

- **Another informational quantity : Chernoff information**

$$d^*(x, y) := d(z^*, x) = d(z^*, y),$$

where z^* is defined by the equality

$$d(z^*, x) = d(z^*, y).$$



Complexity of m best arm identification (FC)

Define the following complexity term :

$$\kappa_C(\nu) = \inf_{\substack{\text{PAC} \\ \text{algorithms}}} \limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)}$$

Lower bound

$$\kappa_C(\nu) \geq \sum_{t=1}^{m} \frac{1}{d(\mu_a, \mu_{m+1})} + \sum_{t=m+1}^{K} \frac{1}{d(\mu_a, \mu_m)}$$

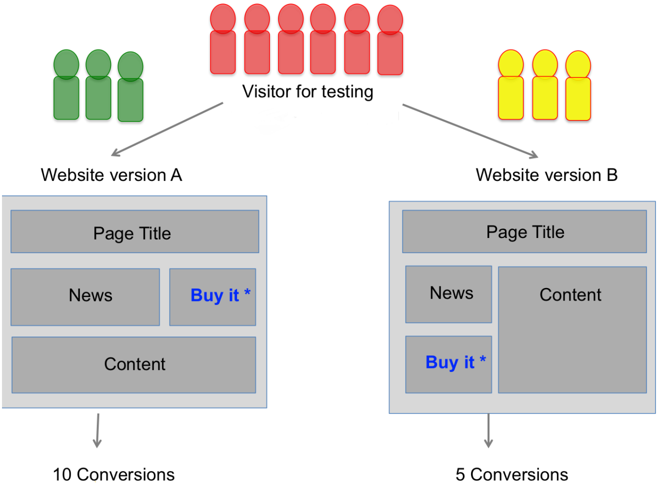
Upper bound (for KL-LUCB)

$$\kappa_C(\nu) \leq 8 \min_{c \in [\mu_{m+1}; \mu_m]} \sum_{a=1}^{K} \frac{1}{d^*(\mu_a, c)}$$

Outline

1. Regret minimization
2. *m* best arms identification
   - Lower bound on the sample complexity
   - An optimal algorithm ?
3. The complexity of A/B Testing
   - The Gaussian case
   - The Bernoulli case

Motivation : A/B Testing



Two possible goals

The agent's goal is to design a strategy $\mathcal{A} = ((A_t), \tau, \hat{a}_\tau)$ satisfying

Fixed-confidence setting	Fixed-budget setting
$\mathbb{P}_\nu(\hat{a}_\tau \neq a^*) \leq \delta$	$\tau = t$
$\mathbb{E}_\nu[\tau]$ as small as possible	$p_t(\nu) := \mathbb{P}_\nu(\hat{a}_t \neq a^*)$ as small as possible

An algorithm using uniform sampling is

Fixed-confidence setting	Fixed-budget setting
a sequential test of $(\mu_1 > \mu_2)$ against $(\mu_1 < \mu_2)$ with probability of error uniformly bounded by δ	a test of $(\mu_1 > \mu_2)$ against $(\mu_1 < \mu_2)$ based on $(t/2)$ samples

[Siegmund 85] : sequential tests can save samples !

Two complexity terms

$\mathcal{M}$ a class of bandit models. $\mathcal{A} = ((A_t), \tau, \hat{a}_\tau)$ is...

Fixed-confidence setting	**Fixed-budget setting**
δ-PAC on $\mathcal{M}$ if $\forall \nu \in \mathcal{M}$, $\mathbb{P}_\nu(\hat{a}_\tau \neq a^*) \leq \delta$	consistent on $\mathcal{M}$ if $\forall \nu \in \mathcal{M}$, $p_t(\nu) := \mathbb{P}_\nu(\hat{a}_\tau \neq a^*_m) \underset{t\to\infty}{\longrightarrow} 0$

Two complexities :

$\kappa_{\mathrm{C}}(\nu) = \inf_{\mathcal{A}\ \mathrm{PAC}} \limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)}$	$\kappa_{\mathrm{B}}(\nu) = \inf_{\mathcal{A}\ \mathrm{cons.}} \left(\limsup_{t\to\infty} -\frac{1}{t} \log p_t(\nu) \right)^{-1}$
for a probability of error $\leq \delta$ $\mathbb{E}_\nu[\tau] \simeq \kappa_C(\nu) \log \frac{1}{\delta}$	for a probability of error $\leq \delta$, budget $t \simeq \kappa_B(\nu) \log \frac{1}{\delta}$

In all our examples, $\hat{a}_\tau = \mathrm{argmax}_a\ \hat{\mu}_a(\tau)$ (empirical best arm)

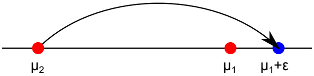
Lower bounds in the two-armed case

From the previous Lemma...

$\mathcal{A}$ is δ-PAC. $\nu = (\nu_1, \nu_2)$,$\nu' = (\nu'_1, \nu'_2)$: $\mu_1 > \mu_2$ and $\mu'_1 < \mu'_2$.

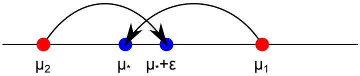
$$\mathbb{E}_\nu[N_1(\tau)]\mathrm{KL}(\nu_1, \nu'_1) + \mathbb{E}_\nu[N_2(\tau)]\mathrm{KL}(\nu_2, \nu'_2) \geq \log\left(\frac{1}{2.4\delta}\right).$$

previously,



$$\begin{aligned} \mu'_1 &= \mu_1 \\ \mu'_2 &= \mu_1 + \epsilon \end{aligned}$$

a new change of distribution :



$$\begin{aligned} \mu'_1 &= \mu_* \\ \mu'_2 &= \mu_* + \epsilon \end{aligned}$$

- choosing $\mu_* : d(\mu_1, \mu_*) = d(\mu_2, \mu_*) := d_*(\mu_1, \mu_2)$:

$$d_*(\mu_1, \mu_2)\mathbb{E}_\nu[\tau] \geq \log\left(\frac{1}{2.4\delta}\right).$$

Lower bounds in the two-armed case

- Exponential families bandit models :
  $\mathcal{M} = \{\nu \in (\mathcal{P})^2 : \mu_1 \neq \mu_2\}$

Fixed-confidence setting	Fixed-budget setting
any δ-PAC algorithm satisfies	any consistent algorithm satisfies
$\mathbb{E}_\nu[\tau] \geq \frac{1}{d_*(\mu_1,\mu_2)} \log\left(\frac{1}{2\delta}\right)$	$\limsup_{t\to\infty} -\frac{1}{t}\log p_t(\nu) \leq d^*(\mu_1,\mu_2)$

- Gaussian bandit models, with σ_1, σ_2 known :

$$\mathcal{M} = \left\{\nu = \left(\mathcal{N}\left(\mu_1,\sigma_1^2\right), \mathcal{N}\left(\mu_2,\sigma_2^2\right)\right) : (\mu_1,\mu_2) \in \mathbb{R}^2, \mu_1 \neq \mu_2\right\},$$

$\mathbb{E}_\nu[\tau] \geq \frac{2(\sigma_1+\sigma_2)^2}{(\mu_1-\mu_2)^2} \log\left(\frac{1}{2\delta}\right)$	$\limsup_{t\to\infty} -\frac{1}{t}\log p_t(\nu) \leq \frac{(\mu_1-\mu_2)^2}{2(\sigma_1+\sigma_2)^2}$

Outline

Fixed-budget setting

$$\mathcal{M} = \left\{\nu = \left(\mathcal{N}\left(\mu_1, \sigma_1^2\right), \mathcal{N}\left(\mu_2, \sigma_2^2\right)\right) : (\mu_1, \mu_2) \in \mathbb{R}^2, \mu_1 \neq \mu_2\right\}$$

- From the lower bound :

$$\kappa_B(\nu) \geq \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2}$$

- A strategy allocating $t_1 = \left\lceil \frac{\sigma_1}{\sigma_1+\sigma_2} t \right\rceil$ samples to arm 1 and $t_2 = t - t_1$ samples to arm 1 satisfies

$$\liminf_{t\to\infty} -\frac{1}{t} \log p_t(\nu) \geq \frac{(\mu_1 - \mu_2)^2}{2(\sigma_1 + \sigma_2)^2}$$

$$\boxed{\kappa_B(\nu) = \frac{2(\sigma_1 + \sigma_2)^2}{(\mu_1 - \mu_2)^2}}$$

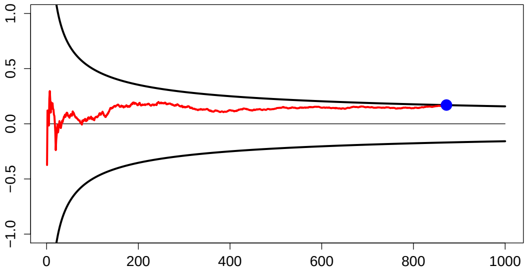
Fixed-confidence setting : algorithm

The α-Elimination algorithm with exploration rate $\beta(t,\delta)$

➜ chooses A_t in order to keep a proportion $N_1(t)/t \simeq \alpha$
i.e. $A_t = 2$ if and only if $\lceil \alpha t \rceil = \lceil \alpha(t+1) \rceil$

➜ if $\hat{\mu}_a(t)$ is the empirical mean of rewards obtained from a up to time t, $\sigma_t^2(\alpha) = \sigma_1^2/\lceil \alpha t \rceil + \sigma_2^2/(t - \lceil \alpha t \rceil)$,

$$\tau = \inf \left\{ t \in \mathbb{N} : |\hat{\mu}_1(t) - \hat{\mu}_2(t)| > \sqrt{2\sigma_t^2(\alpha)\beta(t,\delta)} \right\}$$



Fixed-confidence setting : results

- From the lower bound :

$$\mathbb{E}_\nu[\tau] \geq \frac{2(\sigma_1+\sigma_2)^2}{(\mu_1-\mu_2)^2}\log\left(\frac{1}{2\delta}\right)$$

Theorem

With $\alpha = \frac{\sigma_1}{\sigma_1+\sigma_2}$ and $\beta(t,\delta) = \log\frac{t}{\delta} + 2\log\log(6t)$,

α-Elimination is δ-PAC and

$$\forall \epsilon > 0, \quad \mathbb{E}_\nu[\tau] \leq (1+\epsilon)\frac{2(\sigma_1+\sigma_2)^2}{(\mu_1-\mu_2)^2}\log\left(\frac{1}{2\delta}\right) + \underset{\delta\to 0}{o_\epsilon}\left(\log\frac{1}{\delta}\right)$$

$$\kappa_C(\nu) = \frac{2(\sigma_1+\sigma_2)^2}{(\mu_1-\mu_2)^2}$$

Outline

Lower bounds for Bernoulli bandit models

$$\mathcal{M} = \{\nu = (\mathcal{B}(\mu_1), \mathcal{B}(\mu_2)) : (\mu_1, \mu_2) \in]0;1[^2, \mu_1 \neq \mu_2\},$$

- From the lower bounds,

$$\kappa_C(\nu) \geq \frac{1}{d_*(\mu_1, \mu_2)} \quad \text{and} \quad \kappa_B(\nu) \geq \frac{1}{d^*(\mu_1, \mu_2)}.$$

$d_*(x,y) = d(x,z_*) = d(y,z_*)$ with z_* defined by $d(x,z_*) = d(y,z_*)$	$d^*(x,y) = d(z^*,x) = d(z^*,y)$ with z^* defined by $d(z^*,x) = d(z^*,y)$ (Chernoff information)

For Bernoulli distributions,

$$\boxed{d^*(\mu_1, \mu_2) > d_*(\mu_1, \mu_2)}$$

Fixed-budget setting

There exists $\alpha(\nu)$ such that a strategy allocating $t1 = \lceil \alpha(\nu) t \rceil$ samples to arm 1 and $t2 = t - t1$ samples to arm 2 satisfies

$$p_t(\nu) \leq \exp(-t d^*(\mu_1, \mu_2)).$$

$$\boxed{\kappa_B(\nu) = \frac{1}{d^*(\mu_1, \mu_2)}}$$

Remarks :

- the optimal strategy not implementable in practice
- using uniform sampling is very close to optimal

Consequence :

$$\boxed{\kappa_C(\nu) > \kappa_B(\nu)}$$

Fixed-confidence setting

Another lower bound

A δ-PAC algorithm using uniform sampling satisfy

$$\mathbb{E}_\nu[\tau] \geq \frac{1}{I_*(\mu_1, \mu_2)} \log\left(\frac{1}{2.4\delta}\right)$$

with

$$I_*(\mu_1, \mu_2) = \frac{d\left(\mu_1, \frac{\mu_1+\mu_2}{2}\right) + d\left(\mu_2, \frac{\mu_1+\mu_2}{2}\right)}{2}.$$

Remark : $I_*(\mu_1, \mu_2)$ is very close to $d_*(\mu_1, \mu_2)$!

$\Rightarrow$ in practice, use uniform sampling ?

The algorithm using uniform sampling and

$$\tau = \inf\left\{t \in \mathbb{N}^* : |\hat{\mu}_1(t) - \hat{\mu}_2(t)| > \log\left(\frac{t}{\delta}\right)\right\}$$

is δ-PAC but not optimal : $\frac{\mathbb{E}[\tau]}{\log(1/\delta)} \simeq \frac{2}{(\mu_1-\mu_2)^2} > \frac{1}{I_*(\mu_1,\mu_2)}$.

The SGLRT algorithm

The stopping rule

$$\tau = \inf \left\{ t \in \mathbb{N}^* : tI_*(\hat{\mu}_1(t), \hat{\mu}_2(t)) > \log\left(\frac{t}{\delta}\right) \right\}$$

corresponds to a Sequential Generalized Likelihood Ratio Test.

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau]}{\log(1/\delta)} \leq \frac{1}{I^*(\mu_1, \mu_2)}$$

SGLRT : optimal among strategies using uniform sampling
(hence, close to optimal)

Conclusion

- the complexity of regret minimization is well-understood
  $\Rightarrow$ complexity term involving Kullback-Leibler divergence
- Chernoff information appears as a relevant complexity measure for best arm identification among two arms
- complexity terms for the fixed-budget and fixed-confidence settings can be different !

Remaining questions

- A/B Testing : for which classes of distributions is uniform sampling a good idea ?
- the complexity of m best arm identification, $m > 1$

References

- O. Cappé, A. Garivier, O-A. Maillard, R. Munos, and G. Stoltz. Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Annals of Statistics*, 2013.
- E. Kaufmann, S. Kalyanakrishnan, Information Complexity in Bandit Subset Selection. In *Proceedings of the 26th Conference On Learning Theory* (COLT), 2013.
- E. Kaufmann, O. Cappé, A. Garivier. On the Complexity of A/B Testing. In *Proceedings of the 27th Conference On Learning Theory* (COLT), 2014.
- E. Kaufmann, O. Cappé, A. Garivier. On the Complexity of Best Arm Identification in Multi-Armed Bandit Models. to appear in the *Journal of Machine Learning Research*, 2015
- T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 1985.