Simple regret for infinitely many armed bandit

Alexandra Carpentier* and Michal Valko**

* StatsLab, University of Cambridge
and
** Sequel team, INRIA Lille Nord-Europe

ICML 2015, July 7th 2015

The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

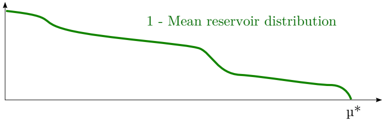
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 0$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

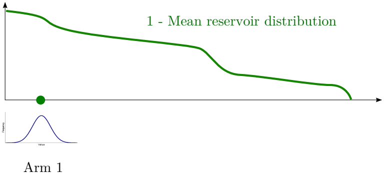
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 1$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

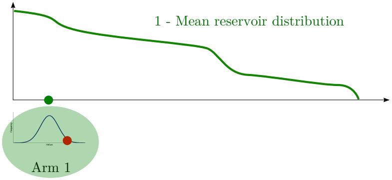
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 1$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \le n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \le K_{t-1}}$,

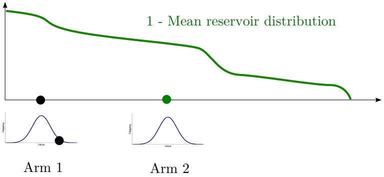
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 2$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

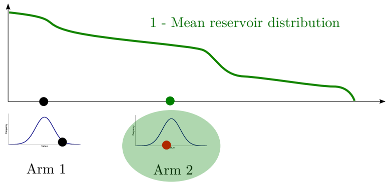
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 2$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \le n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \le K_{t-1}}$,

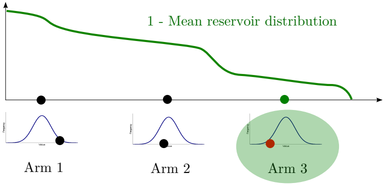
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 3$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

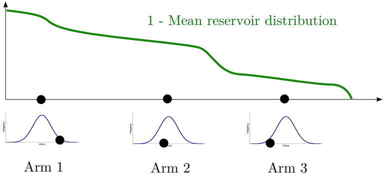
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 3$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

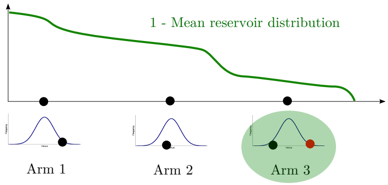
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 4$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

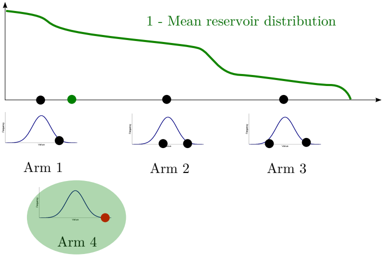
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 5$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

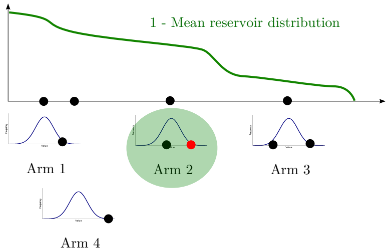
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 6$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

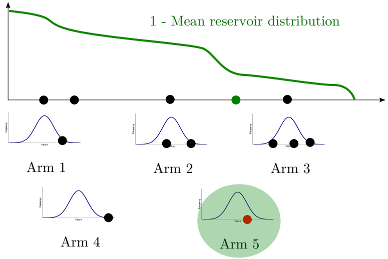
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 7$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

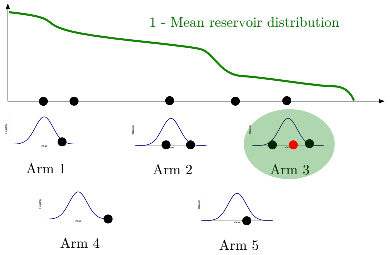
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 8$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

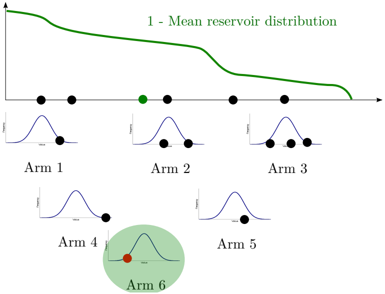
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = 9$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

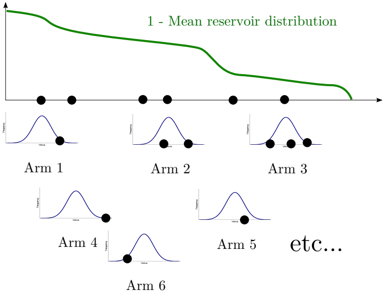
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time t...:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

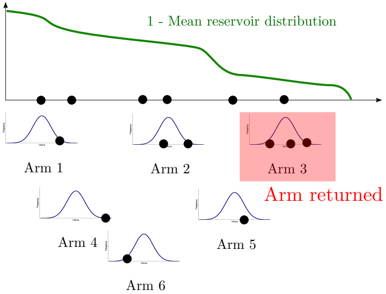
and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

At time $t = n$:



The bandit problem considered

Simple regret for infinitely many armed bandit

- Mean reservoir distr. F bounded by $\bar{\mu}^*$
- Limited sampling resources n

At time $t \leq n$ one can either

- sample a new arm ν_{K_t} from the reservoir distr. with mean $\mu_{K_t} \sim F$, and set $I_t = K_t$,
- or choose an arm I_t among the K_{t-1} observed arms $\{\nu_k\}_{k \leq K_{t-1}}$,

and then collect $X_t \sim \nu_{k_t}$

Objective: after n rounds, return an arm $\widehat{k}$ whose mean $\mu_{\widehat{k}}$ is as large as possible. Minimize the *simple regret*

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}},$$

where $\bar{\mu}^*$ is the right end point of $1 - F$.

Double exploration dilemma here: Allocation both to (i) learn the characteristics of the arm reservoir distr. (*meta-exploration*) and (ii) learn the characteristics of the arms (*exploration*).

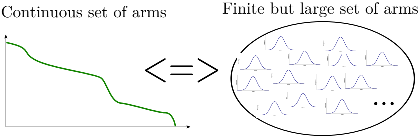
Main questions

How many arms should be sampled from the arm reservoir distribution? How aggressively should these arms be explored?

Applications

Simple-regret bandit problems with a *large number of arms* or with a *small budget*:

- Selection of a good biomarker
- Special case of *feature selection* where one wants to select a single feature [Hauskrecht et al., 2006]



Literature review

- **Simple regret bandits:** [Even-Dar et al., 2006], [Audibert et al., 2010], [Kalyanakrishnan et al., 2012], [Kaufmann et al., 2013], [Karnin et al., 2013], [Gabillon et al., 2012], and [Jamieson et al., 2014]
- **Infinitely many armed bandits with cumulative regret:** [Berry et al., 1997], [Wang et al., 2008], and [Bonald and Proutière, 2013].
- **Infinitely many armed settings with arm structure:** [Dani et al., 2008], [Kleinberg et al.,2008], [Munos, 2014], [Azar et al., 2014]

Literature review

- **Simple regret bandits:** [Even-Dar et al., 2006], [Audibert et al., 2010], [Kalyanakrishnan et al., 2012], [Kaufmann et al., 2013], [Karnin et al., 2013], [Gabillon et al., 2012], and [Jamieson et al., 2014]
- **Infinitely many armed bandits with cumulative regret:** [Berry et al., 1997], [Wang et al., 2008], and [Bonald and Proutière, 2013]
- **Infinitely many armed settings with arm structure:** [Dani et al., 2008], [Kleinberg et al.,2008], [Munos, 2014], [Azar et al., 2014]

Results:

- offer strategies that provide an optimal (or ϵ-optimal) arm with high probability.
- provide stopping rule based strategies that sample until they can provide an ϵ-optimal arm.

But:

- Fixed number of arms that is *smaller* than the budget n (importance of *trying each arm*).

Literature review

- **Simple regret bandits:** [Even-Dar et al., 2006], [Audibert et al., 2010], [Kalyanakrishnan et al., 2012], [Kaufmann et al., 2013], [Karnin et al., 2013], [Gabillon et al., 2012], and [Jamieson et al., 2014]
- **Infinitely many armed bandits with cumulative regret:** [Berry et al., 1997], [Wang et al., 2008], and [Bonald and Proutière, 2013]
- **Infinitely many armed settings with arm structure:** [Dani et al., 2008], [Kleinberg et al.,2008], [Munos, 2014], [Azar et al., 2014]

Results:

- Provide optimal strategies under shape constraint on F and boundedness of the arm distributions.

But:

- Cumulative regret.

Note

We will discuss this in details soon...

Literature review

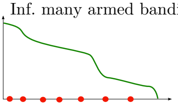
- **Simple regret bandits:** [Even-Dar et al., 2006], [Audibert et al., 2010], [Kalyanakrishnan et al., 2012], [Kaufmann et al., 2013], [Karnin et al., 2013], [Gabillon et al., 2012], and [Jamieson et al., 2014]
- **Infinitely many armed bandits with cumulative regret:** [Berry et al., 1997], [Wang et al., 2008], and [Bonald and Proutière, 2013]
- **Infinitely many armed settings with arm structure:** [Dani et al., 2008], [Kleinberg et al.,2008], [Munos, 2014], [Azar et al., 2014]

Results:

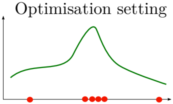
- Provide optimal strategies for specific structured bandits.

But:

- Structure or contextual information needed.



No control on the position of where one samples from the reservoir distr.



One *selects* where one samples based on proximity to good points

Back on infinitely many armed bandits literature

IMAB with cumulative regret: [Berry et al., 1997], [Wang et al., 2008], and [Bonald and Proutière, 2013].

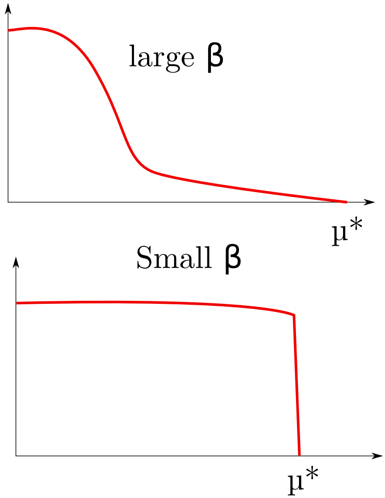
Cumulative regret:

$$R_n^C = n\bar{\mu}^* - \sum_{t \leq n} X_t.$$

Crucial assumption:

$$\mathbb{P}_{\mu \sim F}\left(\bar{\mu}^* - \mu \geq \varepsilon\right) \approx \varepsilon^{\beta},$$

i.e. $1 - F$ is β-regularly varying in $\bar{\mu}^*$.



Back on infinitely many armed bandits literature

IMAB with cumulative regret: [Berry et al., 1997], [Wang et al., 2008], and [Bonald and Proutière, 2013].

Cumulative regret:

$$R_n^C = n\bar{\mu}^* - \sum_{t \leq n} X_t.$$

Crucial assumption:

$$\mathbb{P}_{\mu \sim F}\left(\bar{\mu}^* - \mu \geq \varepsilon\right) \approx \varepsilon^{\beta},$$

i.e. $1 - F$ is β-regularly varying in $\bar{\mu}^*$.

Requirements: Bounded arm distributions and *knowledge of* β *for choosing the nb. of arms.*

Theorem (Regret bound)

Minimax bound on $\mathbb{E}(R_n^C)$

$$\mathcal{O}\left(\max\left(n^{\frac{\beta}{\beta+1}}, \sqrt{n}\right)\right) \text{ up to } \log(n).$$

Special case: If arm distr. bounded by $\bar{\mu}^*$, different rate.

Theorem (Special regret)

Minimax bound on $\mathbb{E}(R_n^C)$

$$\mathcal{O}\left(n^{\frac{\beta}{\beta+1}}\right) \text{ up to } \log(n).$$

The simple regret setting and assumptions

Objective:

Minimize the simple regret in the infinitely many armed setting

$$r_n = \bar{\mu}^* - \mu_{\widehat{k}}.$$

Same assumptions as for IMAB with cumulative regret:

- Regularly varying mean reservoir distr. :

$$\mathbb{P}_{\mu \sim F}\left(\bar{\mu}^* - \mu \geq \varepsilon\right) \approx \varepsilon^{\beta}$$

- Distributions from the arms are bounded/sub-Gaussian.

Lower bound

The following lower bound holds.

Theorem (CV15)

The expected simple regret $\mathbb{E}(r_n)$ can be lower bounded as

$$\max\left(n^{-\frac{1}{\beta}}, n^{-1/2}\right).$$

Remark: Different bottleneck as for the cumulative regret

$$\mathbb{E}[R_n^C] = \mathcal{O}\Big(\max\left(n^{\frac{\beta}{\beta+1}}, \sqrt{n}\right)\Big).$$

Strategy that attains this bound?

The SiRI strategy

Parameters: β, C, δ.
Pick $\bar{T}_\beta \approx n^{\min(\beta,2)/2}$ arms from the reservoir
Pull each of $\bar{T}_\beta$ arms once and set $t \leftarrow \bar{T}_\beta$.
while $t \leq n$ **do**
  For any $k \leq \bar{T}_\beta$, set

$$B_{k,t} \leftarrow \widehat{\mu}_{k,t} + 2\sqrt{\frac{C}{T_{k,t}} \log\left(\frac{n\delta}{T_{k,t}}\right)} + \frac{2C}{T_{k,t}} \log\left(\frac{n\delta}{T_{k,t}}\right)$$

  Pull $T_{k_t,t}$ times the arm k_t that maximizes the $B_{k,t}$
  Set $t \leftarrow t + T_{k_t,t}$.
end while
Output: Return the most pulled arm $\widehat{k}$.

The SiRI strategy

Parameters: β, C, δ.
Pick $\bar{T}_\beta \approx n^{\min(\beta,2)/2}$ arms from the reservoir
Pull each of $\bar{T}_\beta$ arms once and set $t \leftarrow \bar{T}_\beta$.
while $t \leq n$ **do**
  For any $k \leq \bar{T}_\beta$, set

$$B_{k,t} \leftarrow \widehat{\mu}_{k,t} + 2\sqrt{\frac{C}{T_{k,t}} \log\left(\frac{n\delta}{T_{k,t}}\right)} + \frac{2C}{T_{k,t}} \log\left(\frac{n\delta}{T_{k,t}}\right)$$

  Pull $T_{k_t,t}$ times the arm k_t that maximizes the $B_{k,t}$
  Set $t \leftarrow t + T_{k_t,t}$.
end while
Output: Return the most pulled arm $\widehat{k}$.

The SiRI strategy

Parameters: β, C, δ.
Pick $\bar{T}_\beta \approx n^{\min(\beta,2)/2}$ arms from the reservoir
Pull each of $\bar{T}_\beta$ arms once and set $t \leftarrow \bar{T}_\beta$.
while $t \leq n$ **do**
For any $k \leq \bar{T}_\beta$, set

$$B_{k,t} \leftarrow \widehat{\mu}_{k,t} + 2\sqrt{\frac{C}{T_{k,t}} \log\left(\frac{n\delta}{T_{k,t}}\right)} + \frac{2C}{T_{k,t}} \log\left(\frac{n\delta}{T_{k,t}}\right)$$

Pull $T_{k_t,t}$ times the arm k_t that maximizes the $B_{k,t}$
Set $t \leftarrow t + T_{k_t,t}$.
end while
Output: Return the most pulled arm $\widehat{k}$.

Remark: SiRI is the combination of a choice of the number of arms and a UCB algorithm for *cumulative* regret.

Upper bound

The following upper bound holds.

Theorem (CV15)

The expected simple regret $\mathbb{E}(r_n)$ *of SiRI can be upper bounded up to* $\log(n)$ *factors as*

$$\max\left(n^{-1/2}, n^{-\frac{1}{\beta}}\right).$$

Lower and upper bound match up to $\log(n)$ factors (not present in all cases).

Extensions

In the paper we present three main extensions:

- **anytime SiRI**.
- **Distributions bounded by** μ^*: a Bernstein modification of SiRI has Minimax optimal simple regret

$$\max\left(n^{-1}, n^{-1/\beta}\right).$$

- **Unknown** β: possible to *estimate* β using arguments from extreme value theory. Simple regret rate is the same up to $\log(n)$ factors. Same could apply to cumulative regret.

Recap on the rates (up to $\log(n)$)

	Minimax optimal rates
Cumulative regret	$\max\left(n^{\frac{\beta}{\beta+1}}, \sqrt{n}\right)$
Cum. regret with arm bound $\bar{\mu}^*$	$n^{\frac{\beta}{\beta+1}}$
Simple regret	$\max\left(n^{-1/\beta}, n^{-1/2}\right)$
Simple regret with arm bound $\bar{\mu}^*$	$\max\left(n^{-1/\beta}, n^{-1}\right)$

Remark: Different bottleneck as for the cumulative regret.

Simulations

Comparison on synthetic data of SiRI with:

- lil'UCB [Jamieson et. al, 2014], to which the optimal oracle number of arms is given (algorithm for simple regret with finitely many arms)
- UCB-F of [Wang et. al, 2008] (algorithm for cumulative regret and infinitely many arms)

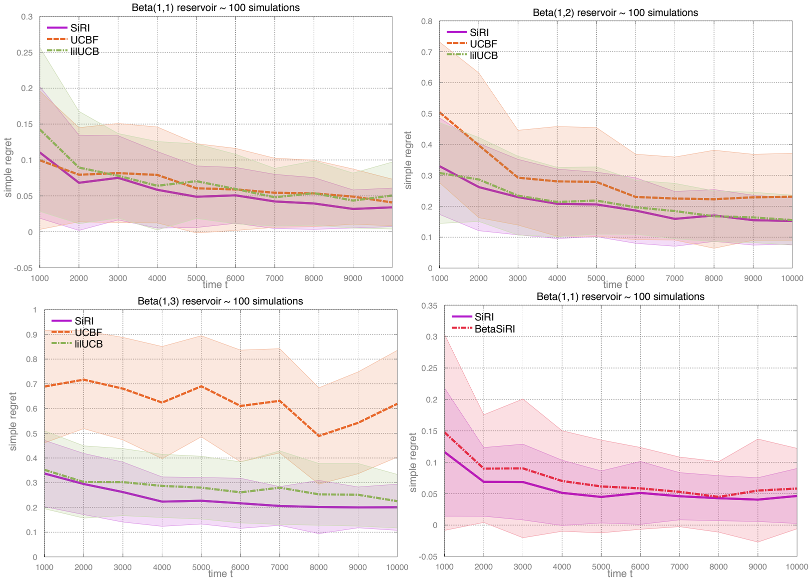


Figure: Comparison on B(1, 1) (UL), B(1, 2) (UR), and B(1, 3) (DL), and unknown β on B(1, 1) (DR)

Conclusion

Minimax optimal solution up to $\log(n)$ factors for the simple regret problem with infinitely many arms. Extensions:

- Unknown β
- Bernstein SiRI with minimax optimal performance when arm distributions are bounded by $\bar{\mu}^*$

Open problems:

- Closing the log gaps (some of them are already closed)?
- Heavy tailed mean reservoir distribution?

THANK YOU!

Acknowledgements This work was supported by the French Ministry of Higher Education and Research and the French National Research Agency (ANR) under project ExTra-Learn n.ANR-14-CE24-0010-01.