**Simple regret for infinitely many armed bandits**

ALEXANDRA CARPENTIER and MICHAL VALKO

a.carpentier@statslab.cam.ac.uk and michal.valko@inria.fr

---

**Setting**

At time $t \leq n$

- pull a known arm $k$ or try a new one
- get the sample $X_t \sim \nu_{k,t}$

Reward only at the end $r_n = \bar{\mu}^* - \mu_k$

**arm selection tradeoff**
- take many to get at least one good
- take few to evaluate them well different from exploration/exploitation tradeoff

**Where is it useful?**

- When we are faced with many choices
  - but we can’t try them all even once.
- Applicable to finite but extremely large cases.
- Single feature selection (biomarkers).

**Other infinite bandits**

- X-armed bandits, bandits in metric spaces, ...
- Linear bandits, convex bandits, ...
- All require contextual information (embedding).

**Definitions**

$A_t = \{\nu_1, \ldots, \nu_{K_t}\}$

$\nu_{K_t + 1} \sim \bar{\nu}$

**Algorithm**

**SIRI - Simple Regret for Infinitely Many Arm Bandits**

**START:** Sample $T_k$ Arms and pull each once

Update $B$-values (estimates + confidence intervals)

$B_{k,t} \leftarrow \tilde{\mu}_{k,t} + 2 \sqrt{\frac{C}{T_k,t} \log \left( \frac{2V_3/\beta}{T_k,t} \right)}$

Pick arm $k$ with the highest $B$ value

Pull arm $k$ to double the samples from it

**END:** return the arm most pulled

**Upper bounds of SIRI**

- $\beta=2$: whp $r_n \leq E n^{-1/2}$
- $\beta=2$: whp $r_n \leq E (n \log n)^{-1/3} \text{polylog } n$

**Lower bounds**

- $\beta=2$: wp > 1/3 $\inf \sup r_n \geq \frac{E n^{-1/2}}{\xi} \text{ for } \xi \leq \xi_0$
- $\beta=2$: wp > 1/3 $\inf \sup r_n \geq \frac{E n^{-1/2}}{\xi} \text{ for } \xi \leq \xi_0$

**Proof sketch**

Based on 2 events that hold with high probability:

- $\xi_1$: controls the number of arm at a given distance from $\mu^*$
- $\xi_2$: controls the distance between empirical and true means $\mu$

Given $\xi_1$ and $\xi_2$ we show that:
- Given the suboptimality gap we can bound the number of suboptimal arms.
- Among $T_k$ arms pulled, there is at least one good enough.
- Empirical means are close to the true ones. (True means are random!)
- We can bound the number of suboptimal arms.
- We can upper bound the number of suboptimal pulls.
- There is a near-optimal arm pulled more than $n/2$ times.
- By definition, this near-optimal arm is selected by SIRI.

**References**

- Prior work that considered the cumulative regret case
  - Berry et al. 1997
  - Formalization and motivation
  - Asymptotic result
- Wang et al. 2008 – UCB-F
  - Finite time result
- Bonald and Proutière, 2013
  - Tight results for the uniform reservoir
  - Simple regret work that considered the finite arm case
- Jamieson et al. 2014 – HUCB
  - Best arm in the fixed confidence setting
- Audibert et al. 2010 – UCB-E
  - Best arm in the fixed budget setting

**Experiments**

- 2 options
  - 1) doubling trick
  - 2) UCB-AR method (Wang et al. 2008)

In both cases regret only worsened by $	ext{polylog } n$.

**Definitions**

- Cumulative regret $R_T = \sum_{t=1}^{T} r_t$
- Simple regret $r_T = \sum_{t=1}^{T} r_t$
- Minimax optimal rates
- UCB
- Beta(1,1) reservoir ~ 100 simulations
- $\beta$-SIRI algorithm
- Devote $n^{1/2}$ samples to estimate $\beta$.
- Get $n^{1/2}$ arms and sample them $n^{1/2}$ times each.
- Same guarantees as for SIRI (up to loglog $n$).

**Rewards in $[0,1]$ with $\mu^{\alpha}$-law?**

- The variance of the near-optimal arms is small.
  - Empirical Bernstein-modified SIRI (idea by Wang et al. 2008)
  - Improved minimax optimal rates (up to polylog $n$)
- $\beta=1$: whp $\tilde{\beta} \leq (2^{1/2}) \text{polylog } n$
- $\beta=2$: whp $\tilde{\beta} \leq (2^{1/2}) \text{polylog } n$

**Anytime algorithm?**

- 2 options
  - 1) doubling trick
  - 2) UCB-AR method (Wang et al. 2008)

In both cases regret only worsened by polylog $n$. 

---

**Comparisons**

- Cumulative regret $\sup_{k,t \in K_t} \frac{1}{n} \sum_{u=1}^{n} \nu_{k,u} = \sup_{k,t \in K_t} \frac{1}{n} \sum_{u=1}^{n} \nu_{k,u} = \sup_{k,t \in K_t} \frac{1}{n} \sum_{u=1}^{n} \nu_{k,u}$

- $\beta=2$: whp $\tilde{\beta} \leq (2^{1/2}) \text{polylog } n$
- $\beta=2$: whp $\tilde{\beta} \leq (2^{1/2}) \text{polylog } n$

**Algorithm**

**Upper bounds of SIRI**

- $\beta=2$: whp $r_n \leq E n^{-1/2}$
- $\beta=2$: whp $r_n \leq E (n \log n)^{-1/3} \text{polylog } n$

**Lower bounds**

- $\beta=2$: wp > 1/3 $\inf \sup r_n \geq \frac{E n^{-1/2}}{\xi} \text{ for } \xi \leq \xi_0$
- $\beta=2$: wp > 1/3 $\inf \sup r_n \geq \frac{E n^{-1/2}}{\xi} \text{ for } \xi \leq \xi_0$