Best of both worlds: Stochastic & adversarial best-arm identification

Yasin Abbasi-Yadkori, Peter Bartlett, Victor Gabillon,
Alan Malek, Michal Valko

COLT - 9 July 2018

For $t = 1, 2, \ldots, n$,

- simultaneously, Learner picks arm $I_t \in [K]$, **(K arms)**
- Adversary /environment picks gain $\boldsymbol{g}_t \in [0,1]^K$.
- Then, the learner observes $\boldsymbol{g}_{t,I_t}$.

Recommend arm J_n hoping $J_n = k^\star$.

Objective: Minimizing the probability of misidentification of $k^\star$:

Adversarial	**Stochastic**
arbitrary $\boldsymbol{g}_{k,t}$	$\boldsymbol{g}_{k,t}$ sampled i.i.d. from ν_k
$k^\star_{\boldsymbol{g}} = \arg\max_{k\in[K]} G_k$ $G_k = \sum_{t=1}^{n} \boldsymbol{g}_{k,t}$	$k^\star_{\textbf{sto}} = \arg\max_{k\in[K]} \mu_k$
$e_{\textbf{adv}}(n) \triangleq \mathbb{P}\left(J_n \neq k^\star_{\boldsymbol{g}}\right)$	$e_{\textbf{sto}}(n) \triangleq \mathbb{P}\left(J_n \neq k^\star_{\textbf{sto}}\right)$
maximizes $e_{\textbf{adv}}(n)$	is indifferent to $e_{\textbf{sto}}(n)$

Worst-case adversarial analysis 😈

State of the art in : Successive Rejects (SR) **(Audibert et al, 2010)**

- SR can pull arm deterministically
- SR stops to pull some arms (eliminate/reject) during the game

SR can be tricked by an adversary 😈

- The learner needs to use internal randomization
- The learner should be careful about rejecting arm: no rejection!

Optimal uniform learner against 😈

RULE: I_t uniformly at random, returns the estimated best arm.

Theorem (Rule vs. 😈)

For all n, adversarial $\boldsymbol{g}$,

$$e_{\mathbf{adv}(\boldsymbol{g})}(n) = \mathcal{O}\left(\exp\left(-\frac{n}{H_{\mathrm{UNIF}(\boldsymbol{g})}}\right)\right).$$

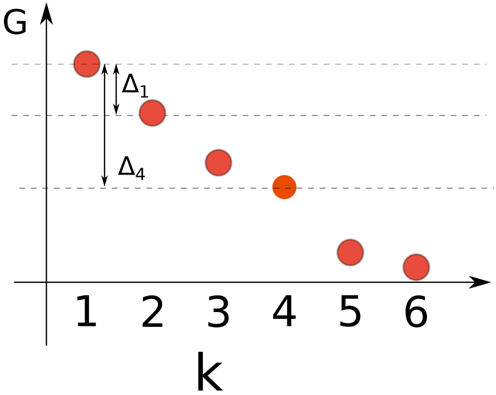
Theorem (😈 Lower bound)

For any learner, a $\boldsymbol{g}^1$ of complexity H_{UNIF},

$$e_{\boldsymbol{g}^1}(n) = \Omega\left(\exp\left(-\frac{n}{H_{\mathrm{UNIF}}}\right)\right).$$

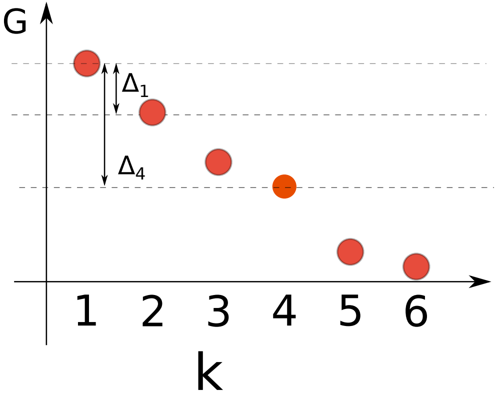
RULE: optimal gap-dependent rates against 😈.

Gaps and complexities in hindsight



$$H_{\mathrm{UNIF}} \triangleq \frac{K}{\Delta_{(1)}^2} \quad .$$

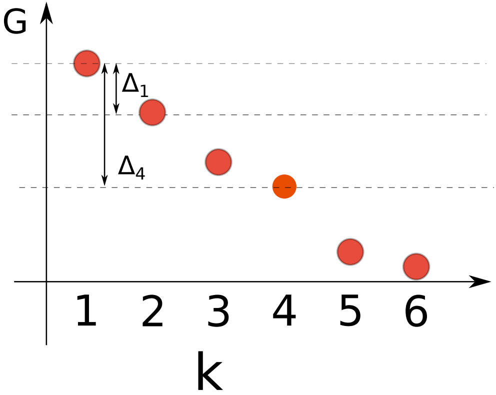
Gaps and complexities in hindsight



$$H_{\mathrm{UNIF}} \triangleq \frac{K}{\Delta_{(1)}^2} \qquad \& \qquad H_{\mathrm{SR}} \triangleq \max_{k\in[K]} \frac{k}{\Delta_{(k)}^2}.$$

Stochastic case

Gaps and complexities in hindsight



$$H_{\text{UNIF}} \triangleq \frac{K}{\Delta_{(1)}^2} \qquad \geq \qquad H_{\text{SR}} \triangleq \max_{k \in [K]} \frac{k}{\Delta_{(k)}^2}.$$

Stochastic case

¿Best of both worlds? (BOB)

Existing robust solutions?

		$e_{\text{sto}}(n)$		$e_{\text{adv}(g)}(n)$
SR	✔	$e^{\frac{-n}{H_{\mathrm{SR}} \log K}}$	✘	1
RULE	✘	$e^{\frac{-n}{H_{\text{UNIF}}}}$	✔	$e^{\frac{-n}{H_{\text{UNIF}}}}$

BOB question: *A learner performing optimally in* **both** *the stochastic and adversarial cases while not being aware of the nature of the rewards ?*

¿Best of both worlds? (BOB)

Existing robust solutions?

		$e_{\mathbf{sto}}(n)$		$e_{\mathbf{adv}(g)}(n)$
SR	✔	$e^{\frac{-n}{H_{\mathrm{SR}} \log K}}$	✘	1
RULE	✘	$e^{\frac{-n}{H_{\mathsf{UNIF}}}}$	✔	$e^{\frac{-n}{H_{\mathsf{UNIF}}}}$

BOB question: *A learner performing optimally in* **both** *the stochastic and adversarial cases while* <u>*not being aware*</u> *of the nature of the rewards ?*

¿Best of both worlds? (BOB)

Existing robust solutions?

		$e_{\mathbf{sto}}(n)$		$e_{\mathbf{adv}(g)}(n)$
SR	✔	$e^{\frac{-n}{H_{\mathrm{SR}} \log K}}$	✘	1
RULE	✘	$e^{\frac{-n}{H_{\mathrm{UNIF}}}}$	✔	$e^{\frac{-n}{H_{\mathrm{UNIF}}}}$

BOB question: *A learner performing optimally in* **both** *the stochastic and adversarial cases while not being aware of the nature of the rewards ?*

¡IMPOSSIBLE BOB!

New notion of complexity

$$H_{\mathrm{BOB}} \triangleq \frac{1}{\Delta_{(1)}} \max_{k\in[K]} \frac{k}{\Delta_{(k)}}.$$

Theorem (Lower bound for the BOB challenge)

For any learner, for any H_{BOB} there exists an stochastic problem with complexity H_{BOB} such that

if $$e_{\textbf{sto}}(n) \leq \frac{1}{64}\exp\left(-\frac{2048n}{H_{\mathrm{BOB}}}\right) \overset{\text{sometimes}}{=} \frac{1}{64}\exp\left(-\frac{2048n}{H_{\mathrm{SR}}\sqrt{K}}\right),$$

then *there exists an adversarial problem where*

$$e_{\textbf{adv}(g)}(n) \geq \frac{1}{16}.$$

¡IMPOSSIBLE BOB!

New notion of complexity

$$H_{\mathrm{BOB}} \triangleq \frac{1}{\Delta_{(1)}} \max_{k \in [K]} \frac{k}{\Delta_{(k)}}.$$

Theorem (Lower bound for the BOB challenge)

For any learner, for any H_{BOB} there exists an stochastic problem with complexity H_{BOB} such that

$$\textbf{if} \qquad e_{\textbf{sto}}(n) \leq \frac{1}{64} \exp\left(-\frac{2048n}{H_{\mathrm{BOB}}}\right) \overset{\textit{sometimes}}{=} \frac{1}{64} \exp\left(-\frac{2048n}{H_{\mathrm{SR}}\sqrt{K}}\right),$$

then *there exists an adversarial problem where*

$$e_{\textbf{adv}(g)}(n) \geq \frac{1}{16}.$$

Is there still a challenge?

YES! because

$$H_{\mathrm{SR}} \leq H_{\mathrm{BOB}} \leq H_{\mathrm{UNIF}}.$$

Why is the BOB question challenging?

- **Bias** of estimator $\widehat{G}_{k,t} \propto \sum_{t'=1}^{t} \mathbf{1}\{I_{t'} = k\} g_{k,t'}$ (simple average)
- **Variance** of $\widetilde{G}_{k,t} = \sum_{t'=1}^{t} \frac{g_{k,t'}}{p_{k,t'}} \mathbf{1}\{I_{t'} = k\}$ (importance weights)

Pull uniformly for too long and incur a large **variance** of order K in $\widetilde{G}_{k,t}$.

Objective: reduce the variance of the estimators of the best arms
$\approx$ find the best arm

The P1 algorithm

P1 pulls

- the $\widehat{best}$ arm with 'probability' $\mathbf{1}$
- the second $\widehat{best}$ arm with 'probability' $\frac{1}{2}$
- the third $\widehat{best}$ arm with 'probability' $\frac{1}{3}$
- and so on...
- the i-th $\widehat{best}$ arm with 'probability' $\frac{1}{i}$
- and the $\widehat{worst}$ arm with 'probability' $\frac{1}{K}$
- (and normalize)

The P1 algorithm

P1 pulls

- the $\widehat{best}$ arm with 'probability' $\mathbf{1}$
- the second $\widehat{best}$ arm with 'probability' $\mathbf{\frac{1}{2}}$
- the third $\widehat{best}$ arm with 'probability' $\frac{1}{3}$
- and so on...
- the i-th $\widehat{best}$ arm with 'probability' $\frac{1}{i}$
- and the $\widehat{worst}$ arm with 'probability' $\frac{1}{K}$
- (and normalize)

The P1 algorithm

P1 pulls

- the $\widehat{best}$ arm with 'probability' $\mathbf{1}$
- the second $\widehat{best}$ arm with 'probability' $\frac{1}{2}$
- the third $\widehat{best}$ arm with 'probability' $\frac{1}{3}$
- and so on...
- the i-th $\widehat{best}$ arm with 'probability' $\frac{1}{i}$
- and the $\widehat{worst}$ arm with 'probability' $\frac{1}{K}$
- (and normalize)

The P1 algorithm

P1 pulls
- the $\widehat{best}$ arm with 'probability' $\mathbf{1}$
- the second $\widehat{best}$ arm with 'probability' $\frac{1}{2}$
- the third $\widehat{best}$ arm with 'probability' $\frac{1}{3}$
- and so on...
- the i-th $\widehat{best}$ arm with 'probability' $\frac{1}{i}$
- and the $\widehat{worst}$ arm with 'probability' $\frac{1}{K}$
- (and normalize)

The P1 algorithm

P1 pulls

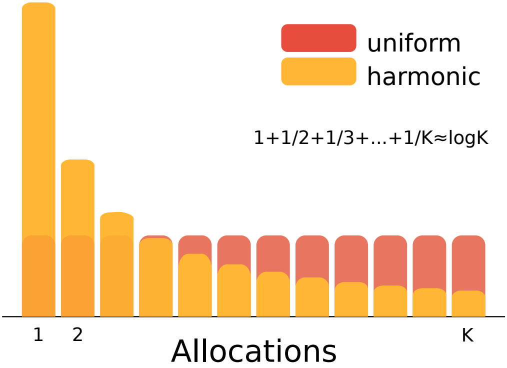
- the $\widehat{best}$ arm with 'probability' $\mathbf{1}$
- the second $\widehat{best}$ arm with 'probability' $\frac{1}{2}$
- the third $\widehat{best}$ arm with 'probability' $\frac{1}{3}$
- and so on...
- the i-th $\widehat{best}$ arm with 'probability' $\frac{1}{i}$
- and the $\widehat{worst}$ arm with 'probability' $\frac{1}{K}$
- (and normalize)

The P1 algorithm

$\textsc{P1}$ pulls

- the $\widehat{best}$ arm with 'probability' $\mathbf{1/\log K}$
- the second $\widehat{best}$ arm with 'probability' $\mathbf{\frac{1}{2\log K}}$
- the third $\widehat{best}$ arm with 'probability' $\mathbf{\frac{1}{3\log K}}$
- and so on...
- the i-th $\widehat{best}$ arm with 'probability' $\mathbf{\frac{1}{i\log K}}$
- and the $\widehat{worst}$ arm with 'probability' $\mathbf{\frac{1}{K\log K}}$
- (and normalize)

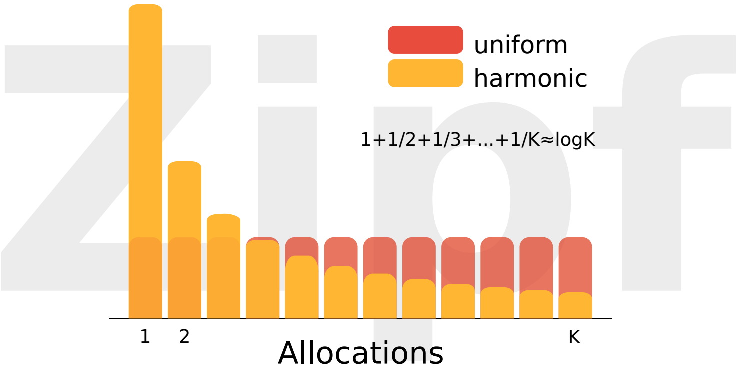
The P1 algorithm



W.r.t. Rule, p1 early bets are almost costless!

P1 follows the allocation proportions of SR

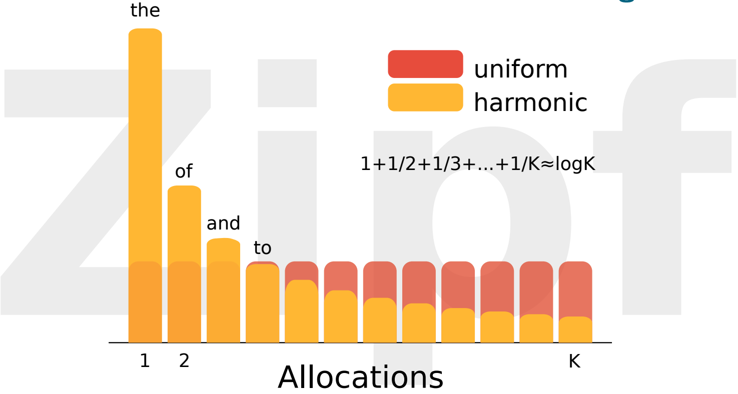
The P1 algorithm



W.r.t. Rule, p1 early bets are almost costless!

P1 follows the allocation proportions of SR

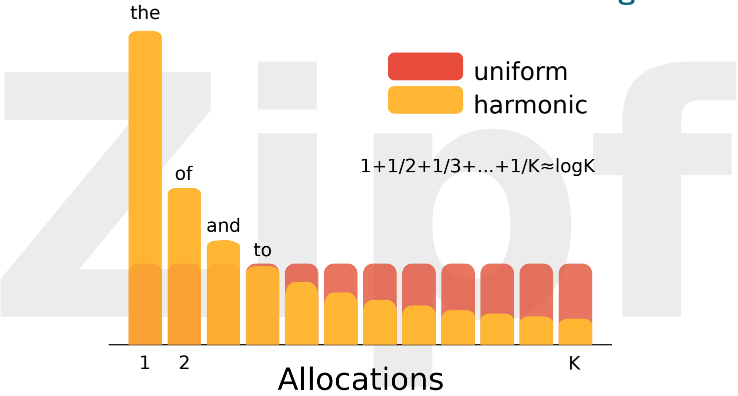
The P1 algorithm



W.r.t. Rule, p1 early bets are almost costless!

P1 follows the allocation proportions of SR

The P1 algorithm



W.r.t. Rule, p1 early bets are almost costless!

P1 follows the allocation proportions of SR

P1 achieves the '*best you can wish for*' (up to log factor) + we have some experiments

Thank you!