Max K-armed bandit: On the ExtremeHunter algorithm and beyond

Mastane Achab, Stéphan Clémençon, Aurélien Garivier, Anne Sabourin, Claire Vernade

ECML PKDD 2017, Skopje

Outline

Outline

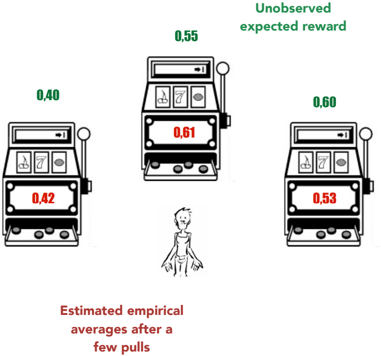
The classical Multi-Armed Bandit problem

At each time $t = 1, \dots, n$

- Play a slot machine ("pull an arm")
- Receive reward

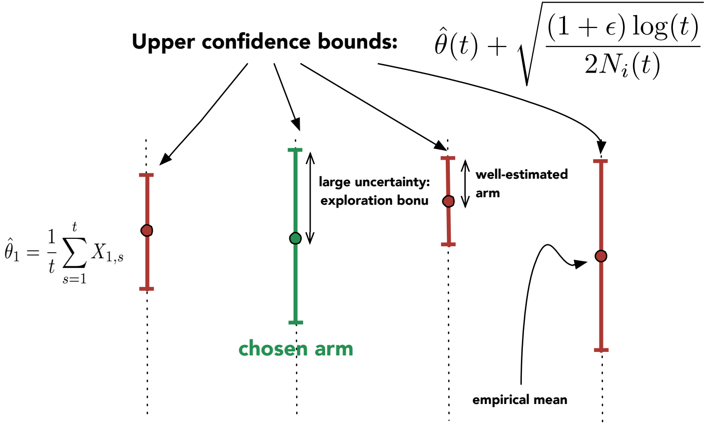
Goal: maximize cumulative reward!

Dilemma: exploration vs *exploitation*



A successful approach: UCB algorithm (Auer et al., 2002)

- *Initialization:* pull each arm once
- Then:



The Max K-Armed Bandit problem

At each time $t = 1, \ldots, n$

- Choose arm $k_t \in \{1, \ldots, K\}$
- Observe reward $X_{k_t,t}$

Multi-Armed Bandits

maximize $\mathbb{E}[\sum_{t=1}^{n} X_{k_t,t}]$

Max K-Armed Bandits (Cicirello and Smith, 2005)

maximize $\mathbb{E}[\max_{1 \leq t \leq n} X_{k_t,t}]$

Extreme Regret

- optimal arm

$$k^* = \underset{1 \leq k \leq K}{\arg\max} \, \mathbb{E}\left[\max_{1 \leq t \leq n} X_{k,t}\right]$$

- equivalent objective

Expected extreme regret

minimize $\mathbb{E}\left[R_n^\pi\right] = \mathbb{E}\left[\max_{1 \leq t \leq n} X_{k^*,t}\right] - \mathbb{E}\left[\max_{1 \leq t \leq n} X_{k_t,t}\right]$

2^{nd}-order Pareto

Definition

F is a 2^{nd}-order Pareto distribution if $\forall x \geq 0$

$$|1 - Cx^{-\alpha} - F(x)| \leq C' x^{-\alpha(1+\beta)} ,$$

with constants $\alpha, \beta, C, C' > 0$.

Some properties

- for $\beta = +\infty$, $F(x) = 1 - Cx^{-\alpha}$ (exact Pareto)
- finite moments of orders $r < \alpha$

Assumption: $\alpha > 1$ (finite mean).

Outline

Controlling $\mathbb{E}[\max_{1 \le t \le n} X_t]$

$X_{1:n} \sim^{\text{iid}} 2^{nd}$-order Pareto $(\alpha > 1, \beta, C, C')$

$\mathbb{E}[\max_{1 \le t \le n} X_t] \sim_{n \to \infty} (nC)^{1/\alpha} \Gamma(1 - 1/\alpha)$ (mean of a Fréchet distribution)

Theorem 1

$$\left| \mathbb{E}\left[\max_{1 \le t \le n} X_t\right] - (nC)^{1/\alpha} \Gamma\left(1 - 1/\alpha\right) \right| = \mathcal{O}\left(n^{-(\min(1,\beta) - 1/\alpha)}\right)$$

sharper than $\mathcal{O}\left(n^{\frac{1}{(1+\beta)\alpha}}\right)$ in C&V14.

Outline

ExtremeETC algorithm

- UCB idea:

$$B_k = \left(n(\widehat{C_k} + \Lambda_2)\right)^{\widehat{1/\alpha_k} + \Lambda_1} \Gamma(1 - \widehat{1/\alpha_k} - \Lambda_1) \geq \mathbb{E}\left[\max_{1 \leq t \leq T} X_{k,t}\right] \text{ with high probability} \tag{1}$$

- Initialization: pull each arm $N = A(\log n)^{\frac{2(2b+1)}{b}}$ times

ExtremeETC vs ExtremeHunter

Input: n: time horizon, K: number of arms, $b > 0$ such that $b \leq \min_k \beta_k$.
Initialize: Pull N times each arm k and compute index B_k (see Eq. (1)).

ExtremeETC	ExtremeHunter
$k_0 = \arg\max_k B_k$	**for** $t > KN$ **do**
for $t > KN$ **do**	Pull $k_t = \arg\max_k B_k$.
Pull arm k_0.	Update index B_{k_t}.
end for	**end for**

Complexity	Ex.ETC	Ex.Hunt.
Time	$\mathcal{O}\left(K(\log n)^{\frac{2(2b+1)}{b}}\right)$	$\mathcal{O}(n^2)$
Memory	$\mathcal{O}\left(K(\log n)^{\frac{2(2b+1)}{b}}\right)$	$\mathcal{O}(n)$

Tight regret bounds

Theorem

(i) *Upper bound for* ExtremeETC *and* ExtremeHunter

$$\mathbb{E}[R_n] = \mathcal{O}\Big((\log n)^{\frac{2(2b+1)}{b}} n^{-(1-1/\alpha_{k^*})} + n^{-(b-1/\alpha_{k^*})}\Big),$$

sharper than $\mathcal{O}\left(n^{\frac{1}{(1+b)\alpha_{k^*}}}\right)$ *in C&V(14).*

(ii) *Lower bound for any algorithm pulling each arm at least* N *times*

$$\mathbb{E}[R_n] = \Omega\left((\log n)^{\frac{2(2b+1)}{b}} n^{-(1-1/\alpha_{k^*})}\right) .$$

When $b \geq 1$, (i) and (ii) are tight !

Regret bounds - idea of proof

- favorable event $(\mathcal{A})$: $1/\alpha_k, C_k \in$ confidence intervals
- Lemma: under $(\mathcal{A})$, k^* always pulled
- use Theorem 1 to control $\mathbb{E}[\max \ldots]$

Outline

Reduction to Multi-Armed Bandits

- Idea: peak over threshold
- Truncated rewards: $Y_{k,t} = X_{k,t} \mathbb{1}_{\{X_{k,t} > u\}}$.
- $\mathbb{E}[Y_{k,1}] \sim_{u \to \infty} C_k \left(1 + \frac{1}{\alpha_k - 1}\right) u^{-\alpha_k + 1}$.
- For u and n large

$$\operatorname*{arg\,max}_{1 \le k \le K} \mathbb{E}[Y_{k,1}] = \operatorname*{arg\,min}_{1 \le k \le K} \alpha_k = k^* \ .$$

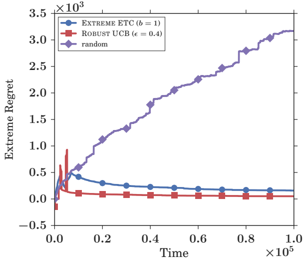
- MAB objective: maximize $\mathbb{E}\left[\sum_{t=1}^{n} Y_{k_t,t}\right]$.
- We use $\textsc{Robust UCB}$ (Bubeck et al., 2013)
  - parameters: $\epsilon < \min_{1 \le k \le K} \alpha_k - 1$, $v \ge \max_{k \in [K]} \mathbb{E}\left[\left|Y_{k,1}\right|^{1+\epsilon}\right]$
  - $\mathbb{E}[\#\text{ pulls arm } k \neq k^*] = \mathcal{O}(\log n) < N$.

Outline

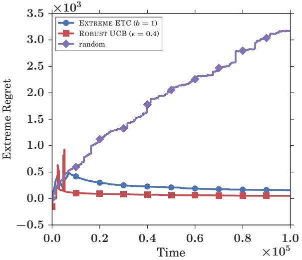
Experiments

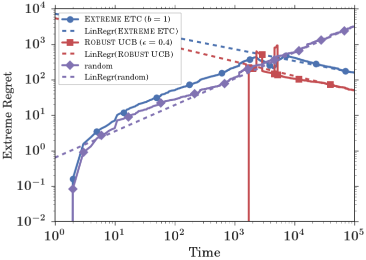
- time horizon $n = 10^5$
- $K = 3$ exact Pareto distributions ($\beta = +\infty$)

Arm	1	$k^* = 2$	3
α	15	1.5	10
C	10^8	1	10^5
$\mathbb{E}[X]$	3.7	3	3.5
$\mathbb{E}[\max_{1 \le t \le n} X_t]$	7.7	$5.8 \cdot 10^3$	11



Experiments





Linear regression for $\textsc{ExtremeETC}$ over $t = 5 \cdot 10^4, \dots, 10^5$ has slope ≈ -0.333 (with $R^2 \approx 0.97$)

$\rightarrow$ **validation of bounds !**