Constant or Logarithmic Regret in Asynchronous Multiplayer Bandits with Limited Communication

Hugo Richard[1], Etienne Boursier[2], Vianney Perchet[1,3]

CIRM, December 15, 2023

[1]Criteo AI Lab, France [2]Inria, France [3]ENSAE, France

Introduction

Setting

Bandits

K arms

For $t \in \{1, \ldots, T\}$:

1. Choose arm $k_t \in [K]$
2. Receive reward $X_{k_t,t}$ subgaussian with mean μ_{k_t}
3. Observe $X_{k_t,t}$

Bandits

K arms

For $t \in \{1, \dots, T\}$:

1. Choose arm $k_t \in [K]$
2. Receive reward $X_{k_t,t}$ subgaussian with mean μ_{k_t}
3. Observe $X_{k_t,t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} X_{k_t,t}]$

Bandits

K arms

For $t \in \{1, \dots, T\}$:

1. Choose arm $k_t \in [K]$
2. Receive reward $X_{k_t,t}$ subgaussian with mean μ_{k_t}
3. Observe $X_{k_t,t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} X_{k_t,t}]$

Comparator: $\max_{k \in [K]} \mathbb{E}[\sum_{t=1}^{T} X_{k,t}] = T\mu_{(K)}$

Bandits

K arms

For $t \in \{1, \ldots, T\}$:

1. Choose arm $k_t \in [K]$
2. Receive reward $X_{k_t,t}$ subgaussian with mean μ_{k_t}
3. Observe $X_{k_t,t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} X_{k_t,t}]$

Comparator: $\max_{k\in[K]} \mathbb{E}[\sum_{t=1}^{T} X_{k,t}] = T\mu_{(K)}$

Regret: $R = T\mu_{(K)} - \mathbb{E}[\sum_{t=1}^{T} X_{k_t,t}]$

Bandits

K arms

For $t \in \{1, \ldots, T\}$:

1. Choose arm $k_t \in [K]$
2. Receive reward $X_{k_t,t}$ subgaussian with mean μ_{k_t}
3. Observe $X_{k_t,t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} X_{k_t,t}]$

Comparator: $\max_{k\in[K]} \mathbb{E}[\sum_{t=1}^{T} X_{k,t}] = T\mu_{(K)}$

Regret: $R = T\mu_{(K)} - \mathbb{E}[\sum_{t=1}^{T} X_{k_t,t}]$

Optimal algorithms achieve $R \approx \sum_{k=1}^{K-1} \frac{\log(T)}{\mu_{(K)}-\mu_{(k)}}$ (Auer, 2002)

Bandits with multiple plays

K arms, $M \leq K$ choices

For $t \in \{1, \ldots, T\}$:

1. Choose M arms $k_{1,t}, \ldots, k_{M,t} \in [K]$
2. Receive reward $\sum_{m=1}^{M} X_{k_{m,t},t}$ where $X_{k,t} \sim B(\mu_k)$
3. Observe $X_{k_{1,t},t}, \ldots, X_{k_{M,t},t}$

Bandits with multiple plays

K arms, $M \leq K$ choices

For $t \in \{1, \dots, T\}$:

1. Choose M arms $k_{1,t}, \dots, k_{M,t} \in [K]$
2. Receive reward $\sum_{m=1}^{M} X_{k_{m,t},t}$ where $X_{k,t} \sim B(\mu_k)$
3. Observe $X_{k_{1,t},t}, \dots, X_{k_{M,t},t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}]$

Bandits with multiple plays

K arms, $M \leq K$ choices

For $t \in \{1, \dots, T\}$:

1. Choose M arms $k_{1,t}, \dots, k_{M,t} \in [K]$
2. Receive reward $\sum_{m=1}^{M} X_{k_{m,t},t}$ where $X_{k,t} \sim B(\mu_k)$
3. Observe $X_{k_{1,t},t}, \dots, X_{k_{M,t},t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}]$

Comparator: $T \sum_{k=K-M+1}^{K} \mu_{(k)}$

Bandits with multiple plays

K arms, $M \leq K$ choices

For $t \in \{1, \dots, T\}$:

1. Choose M arms $k_{1,t}, \dots, k_{M,t} \in [K]$
2. Receive reward $\sum_{m=1}^{M} X_{k_{m,t},t}$ where $X_{k,t} \sim B(\mu_k)$
3. Observe $X_{k_{1,t},t}, \dots, X_{k_{M,t},t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}]$

Comparator: $T \sum_{k=K-M+1}^{K} \mu_{(k)}$

Regret: $R = T \sum_{k=K-M+1}^{K} \mu_{(k)} - \mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}]$

Bandits with multiple plays

K arms, $M \leq K$ choices

For $t \in \{1, \ldots, T\}$:

1. Choose M arms $k_{1,t}, \ldots, k_{M,t} \in [K]$
2. Receive reward $\sum_{m=1}^{M} X_{k_{m,t},t}$ where $X_{k,t} \sim B(\mu_k)$
3. Observe $X_{k_{1,t},t}, \ldots, X_{k_{M,t},t}$

Goal: Maximize $\mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}]$

Comparator: $T \sum_{k=K-M+1}^{K} \mu_{(k)}$

Regret: $R = T \sum_{k=K-M+1}^{K} \mu_{(k)} - \mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}]$

Optimal algorithms achieve $R \approx \sum_{k=1}^{K-M} \frac{\log(T)}{\mu_{(K)} - \mu_{(k)}}$ (Komiyama, 2015)

Multiplayer bandits

K arms, $M \leq K$ players

For $t \in \{1, \dots, T\}$:

1. Each player m chooses an arm $k_{m,t} \in [K]$
2. Each player m receives
$$X_{k_{m,t},t} \underbrace{\mathbb{1}\{\text{Exactly one player pulls arm } k_{m,t}\}}_{\eta_{k_{m,t},t}}$$
3. Each player m observes $X_{k_{m,t},t}\eta_{k_{m,t},t}$, $\eta_{k_{m,t},t}$

Multiplayer bandits

K arms, $M \leq K$ players

For $t \in \{1, \dots, T\}$:

1. Each player m chooses an arm $k_{m,t} \in [K]$
2. Each player m receives
   $X_{k_{m,t},t} \underbrace{\mathbb{1}\{\text{Exactly one player pulls arm } k_{m,t}\}}_{\eta_{k_{m,t},t}}$
3. Each player m observes $X_{k_{m,t},t}\eta_{k_{m,t},t}$, $\eta_{k_{m,t},t}$

Regret: $R = T \sum_{k=K-M+1}^{K} \mu_{(k)} - \mathbb{E}[\sum_{t=1}^{T} \sum_{m=1}^{M} X_{k_{m,t},t}\eta_{k_{m,t},t}]$

Multiplayer bandits

K arms, $M \leq K$ players

For $t \in \{1, \ldots, T\}$:

1. Each player m chooses an arm $k_{m,t} \in [K]$
2. Each player m receives
   $X_{k_{m,t},t} \underbrace{\mathbb{1}\{\text{Exactly one player pulls arm } k_{m,t}\}}_{\eta_{k_{m,t},t}}$
3. Each player m observes $X_{k_{m,t},t}\eta_{k_{m,t},t}$, $\eta_{k_{m,t},t}$

Regret: $R = T\sum_{k=K-M+1}^{K} \mu_{(k)} - \mathbb{E}[\sum_{t=1}^{T}\sum_{m=1}^{M} X_{k_{m,t},t}\eta_{k_{m,t},t}]$

Optimal algorithms achieve $R \approx \sum_{k=1}^{K-M} \frac{\log(T)}{\mu_{(K)}-\mu_{(k)}}$ (Boursier, 2019) (Wang, 2020)

Asynchronous Multiplayer bandits

K arms, $M > K$ players, $(p_m)_{m=1}^M$ activation probabilities

For $t \in \{1, \dots, T\}$:

1. Each player m chooses an arm $k_{m,t} \in [K]$
2. Each player m is active with probability p_m
3. Each player m receives
   $$X_{k_{m,t},t} \underbrace{\mathbb{1}\{\text{Exactly one player is active and pulls arm } k_{m,t}\}}_{\eta_{k_{m,t},t}}$$
4. Each player m observe $X_{k_{m,t},t}\eta_{k_{m,t},t}$, $\eta_{k_{m,t},t}$

Asynchronous Multiplayer bandits

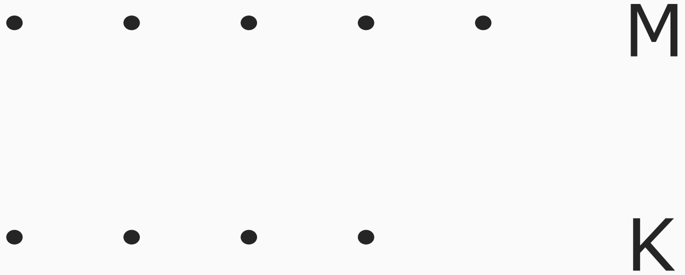
K arms, $M > K$ players, $(p_m)_{m=1}^M$ activation probabilities

For $t \in \{1, \dots, T\}$:

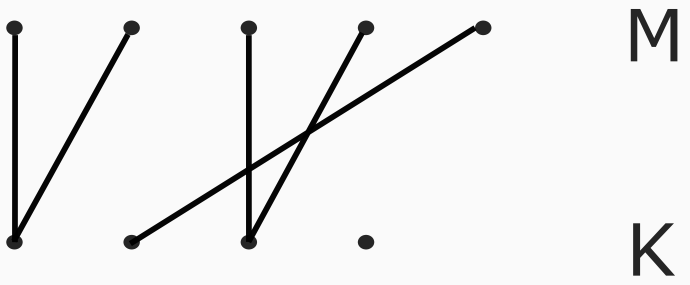
1. Each player m chooses an arm $k_{m,t} \in [K]$
2. Each player m is active with probability p_m
3. Each player m receives
   $X_{k_{m,t},t} \underbrace{\mathbb{1}\{\text{Exactly one player is active and pulls arm } k_{m,t}\}}_{\eta_{k_{m,t},t}}$
4. Each player m observe $X_{k_{m,t},t}\eta_{k_{m,t},t}$, $\eta_{k_{m,t},t}$

Regret: $R = \max_{k_1,\dots,k_M \in [K]} \sum_{t=1}^T \mathbb{E}[\sum_{m=1}^M X_{k_m,t}\eta_{k_m,t}] - \sum_{t=1}^T \mathbb{E}[\sum_{m=1}^M X_{k_{m,t},t}\eta_{k_{m,t},t}]$

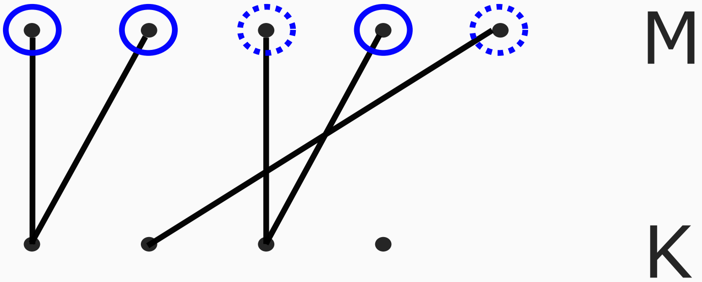
Asynchronous Multiplayer bandits



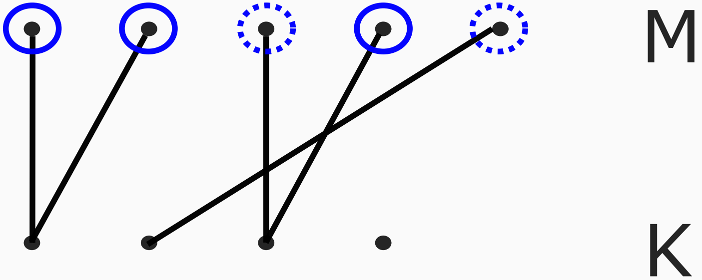
Asynchronous Multiplayer bandits



Asynchronous Multiplayer bandits



Asynchronous Multiplayer bandits



Player 4 receives $X_{3,t}$

Player $1, 2$ receive 0 but observes $\eta_{1,t} = 0$

Asynchronous Multiplayer bandits

- (Bonnefois, 2017) Selfish algorithms are promising
- (Dakdouk, 2022) $O(T^{\frac{2}{3}})$ regret with limited communication

Communication model

By (Dakdouk 2022):

Communication abilities
At each t, players can either

1. Attempt to send a message to a *gateway* (one player at a time)
2. Listen to messages from the gateway

Communication model

By (Dakdouk 2022):

Communication abilities
At each t, players can either

1. Attempt to send a message to a *gateway* (one player at a time)
2. Listen to messages from the gateway

- Succeeds with probability p_g
- Does not cost anything
- We keep count of the number of communication attempts

Communication model

- After τ attemps a communication fails with probability $(1-p_g)^{\tau}$

Communication model

- After τ attemps a communication fails with probability $(1 - p_g)^\tau$
- After $\tau = \frac{\log(T)}{-\log(1-p_g)}$ attemps a communication fails with probability $\frac{1}{T}$

Communication model

- After τ attemps a communication fails with probability $(1-p_g)^\tau$
- After $\tau = \frac{\log(T)}{-\log(1-p_g)}$ attemps a communication fails with probability $\frac{1}{T}$
- On average after $\frac{1}{p_g}$ attemps, the communication succeeds

Communication model

- After τ attemps a communication fails with probability $(1 - p_g)^\tau$
- After $\tau = \frac{\log(T)}{-\log(1-p_g)}$ attemps a communication fails with probability $\frac{1}{T}$
- On average after $\frac{1}{p_g}$ attemps, the communication succeeds

(Dakdouk 2022): $\tilde{O}(\frac{1}{p_g})$ expected communication attempts (at most $\tilde{O}(\frac{\log(T)}{-\log(1-p_g)})$)

Assumptions

Homogeneous activation probabilities

$$\forall m \in [M], p_m = p$$

Assumptions

Define $\mathbf{M}(t) = (M_1(t), \ldots, M_K(t))$

M_k = number of players choosing arm k at time t

Assumptions

Define $\mathbf{M}(t) = (M_1(t), \ldots, M_K(t))$
M_k = number of players choosing arm k at time t

Total expected reward at time t

$$= \mathbb{E}[\sum_{k=1}^{K} X_{k,t} \mathbb{1}\{\text{Exactly 1 player is active among } M_k(t)\}]$$

Assumptions

Define $\mathbf{M}(t) = (M_1(t), \ldots, M_K(t))$
M_k = number of players choosing arm k at time t

Total expected reward at time t

$$= \mathbb{E}[\sum_{k=1}^{K} X_{k,t} \mathbb{1}\{\text{Exactly 1 player is active among } M_k(t)\}]$$
$$= \mathbb{E}[\sum_{k=1}^{K} \mu_k \underbrace{pM_k(t)(1-p)^{M_k(t)-1}}_{g(M_k(t))}]$$

Assumptions

Define $\mathbf{M}(t) = (M_1(t), \ldots, M_K(t))$
$M_k =$ number of players choosing arm k at time t

Total expected reward at time t

$$= \mathbb{E}[\sum_{k=1}^{K} X_{k,t} \mathbb{1}\{\text{Exactly 1 player is active among } M_k(t)\}]$$

$$= \mathbb{E}[\sum_{k=1}^{K} \mu_k \underbrace{p M_k(t)(1-p)^{M_k(t)-1}}_{g(M_k(t))}]$$

$$= \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}(t)) \rangle]$$

Assumptions

Define $\mathbf{M}^* = \text{argmax}_{\mathbf{M}, \sum_{k=1}^{K} M_k = M} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle]$

Upper bound on optimal allocation

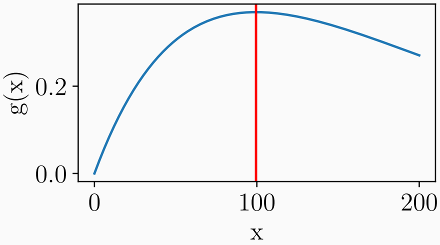
$\forall k \in [K], \mathbf{M}_k^* \leq -\frac{1}{\log(1-p)}$

Assumptions

Define $\mathbf{M}^* = \text{argmax}_{\mathbf{M}, \sum_{k=1}^K M_k = M} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle]$

Upper bound on optimal allocation

$\forall k \in [K], \mathbf{M}_k^* \leq -\frac{1}{\log(1-p)}$

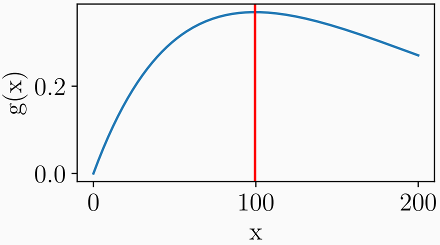


Assumptions

Define $\mathbf{M}^* = \text{argmax}_{\mathbf{M}, \sum_{k=1}^{K} M_k = M} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle]$

Upper bound on optimal allocation

$\forall k \in [K], \mathbf{M}_k^* \leq -\frac{1}{\log(1-p)}$



$\max_{\mathbf{M}, \sum_{k=1}^{K} M_k = M, M_k \leq -\frac{1}{\log(1-p)}} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$ easy to solve

(Bonnefois, 2017) (Dakdouk, 2022)

Cautious Greedy

Results

$M > K$, $p < 1$, number of communications $\tilde{O}(\log(T))$ in expectation and at most $\tilde{O}(\log^2(T))$

Results

$M > K$, $p < 1$, number of communications $\tilde{O}(\log(T))$ in expectation and at most $\tilde{O}(\log^2(T))$

Define $\nu^* = \|\mathbf{M}^*\|_0 = |\{k, M_k^* = 0\}|$

$$R = \tilde{O}(\frac{1}{r} + \sum_{\nu=1}^{\nu^*} \frac{\log(T)}{\mu_{(\nu^*+1)} - \mu_{(\nu)}})$$

r: data-dependent gap, dependency in K, M, p, p_g hidden.

Results

$M > K$, $p < 1$, number of communications $\tilde{O}(\log(T))$ in expectation and at most $\tilde{O}(\log^2(T))$

Define $\nu^* = \|\mathbf{M}^*\|_0 = |\{k, M_k^* = 0\}|$

$$R = \tilde{O}\left(\frac{1}{r} + \sum_{\nu=1}^{\nu^*} \frac{\log(T)}{\mu_{(\nu^*+1)} - \mu_{(\nu)}}\right)$$

r: data-dependent gap, dependency in K, M, p, p_g hidden.

Lower bounds

- $\nu^* = 0$: The dependency in r is optimal
- $\nu^* > 0$: The term $\sum_{\nu=1}^{\nu^*} \frac{\log(T)}{\mu_{(\nu^*+1)} - \mu_{(\nu)}}$ is optimal.

Quasi-centralized Cautious Greedy

K arms, $M > K$ players, p activation probability

For $t \in \{1, \ldots, T\}$:

1. Choose an assignment of player $\mathbf{M}(t)$
2. Each player m is active with probability p
3. Receive $\langle \mathbf{X}_t, \boldsymbol{\eta}(\mathbf{M}(t)) \rangle$, $X_{i,t} \sim B(\mu_i)$
4. Observe $\mathbf{X}_t \odot \boldsymbol{\eta}(\mathbf{M}(t))$ and $\boldsymbol{\eta}(\mathbf{M}(t))$

Quasi-centralized Cautious Greedy

K arms, $M > K$ players, p activation probability

For $t \in \{1, \ldots, T\}$:

1. Choose an assignment of player $\mathbf{M}(t)$
2. Each player m is active with probability p
3. Receive $\langle \mathbf{X}_t, \boldsymbol{\eta}(\mathbf{M}(t)) \rangle$, $X_{i,t} \sim B(\mu_i)$
4. Observe $\mathbf{X}_t \odot \boldsymbol{\eta}(\mathbf{M}(t))$ and $\boldsymbol{\eta}(\mathbf{M}(t))$

Regret: $R = \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]$

Regret of Cautious Greedy $R_{CG} = \tilde{O}(\frac{1}{r} + \sum_{\nu=1}^{\nu^*} \frac{\log(T)}{\mu_{(\nu^*+1)} - \mu_{(\nu)}})$

Why is constant regret possible ?

Assume $\nu^* = 0$ i.e. support$(\mathbf{M}^*) = [K]$

Why is constant regret possible ?

Assume $\nu^* = 0$ i.e. support$(\mathbf{M}^*) = [K]$

Greedy

- Compute $\hat{\mu}_i(t) = \frac{\sum_{\tau=1}^{t} X_{i,t}\eta_{i,t}}{\sum_{\tau=1}^{t} \eta_{i,t}}$
- Play $\mathrm{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k \le M, M_k \le \frac{-1}{\log(1-p)}\}} \langle \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}) \rangle$

Why is constant regret possible ?

$$R = \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]$$

Why is constant regret possible ?

$$
\begin{aligned}
R &= \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\
&\le \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]
\end{aligned}
$$

Why is constant regret possible ?

$$
\begin{aligned}
R &= \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\
&\leq \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\
&\lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}
\end{aligned}
$$

Why is constant regret possible ?

$$
\begin{aligned}
R &= \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\
&\leq \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\
&\lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}
\end{aligned}
$$

Why is constant regret possible ?

$$R = \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]$$
$$\leq \sum_{t=1}^{T} \mathbb{E}[\langle \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]$$
$$\lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}$$

Call $\mathbf{M}^{\hat{\mu}} = \mathrm{argmin}_{\mathbf{M}} \langle \hat{\boldsymbol{\mu}}, g(\mathbf{M}) \rangle$:

$r = \min_{\{\hat{\mu}, \mathbf{M}^{\hat{\mu}} \neq \mathbf{M}^*\}} \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_\infty$

Why is constant regret possible

$$R \lesssim \sum_{t=1}^{T} \mathbb{E}[\|\mu - \hat{\mu}(t)\|_\infty \mathbb{1}\{\mathsf{M}(t) \neq \mathsf{M}^*\}$$

Why is constant regret possible

$$R \lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}$$

$$= \sum_{t=1}^{T} r\mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > r) + \int_r^\infty \mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > u)du$$

Why is constant regret possible

$$
\begin{aligned}
R &\lesssim \sum_{t=1}^{T} \mathbb{E}[\|\mu - \hat{\mu}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}] \\
&= \sum_{t=1}^{T} r\mathbb{P}(\|\mu - \hat{\mu}(t)\|_\infty > r) + \int_r^\infty \mathbb{P}(\|\mu - \hat{\mu}(t)\|_\infty > u) du
\end{aligned}
$$

Why is constant regret possible

$$R \lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}$$
$$= \sum_{t=1}^{T} r\mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > r) + \int_r^\infty \mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > u)du$$

$$\text{Number of samples on arm } k = \sum_{\tau=1}^{t} \eta(M_k(t))$$

Why is constant regret possible

$$R \lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}$$
$$= \sum_{t=1}^{T} r\mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > r) + \int_r^\infty \mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > u)du$$

$$\text{Number of samples on arm } k = \sum_{\tau=1}^{t} \eta(M_k(t))$$

$$\text{Expected number of samples on arm } k = \sum_{\tau=1}^{t} \mathbb{E}[g(M_k(t))]$$

Why is constant regret possible

$$R \lesssim \sum_{t=1}^{T} \mathbb{E}[\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty \mathbb{1}\{\mathbf{M}(t) \neq \mathbf{M}^*\}$$
$$= \sum_{t=1}^{T} r\mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > r) + \int_r^\infty \mathbb{P}(\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(t)\|_\infty > u)du$$

$$\text{Number of samples on arm } k = \sum_{\tau=1}^{t} \eta(M_k(t))$$
$$\text{Expected number of samples on arm } k = \sum_{\tau=1}^{t} \mathbb{E}[g(M_k(t))]$$
$$\geq p\mathbb{E}[g(1)] = pt$$

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?
- Maintain a lower bound ν of ν^*

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?
- Maintain a lower bound ν of ν^*

Building a lower bound on ν^*
Step 0: $\nu = 0$

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?
- Maintain a lower bound ν of ν^*

Building a lower bound on ν^*
Step 0: $\nu = 0$

Step 1: $\hat{\boldsymbol{\mu}}^L$, $\hat{\boldsymbol{\mu}}^U$ such that whp: $\hat{\boldsymbol{\mu}}^L \leq \boldsymbol{\mu} \leq \hat{\boldsymbol{\mu}}^U$

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?
- Maintain a lower bound ν of ν^*

Building a lower bound on ν^*

Step 0: $\nu = 0$

Step 1: $\hat{\boldsymbol{\mu}}^L$, $\hat{\boldsymbol{\mu}}^U$ such that whp: $\hat{\boldsymbol{\mu}}^L \leq \boldsymbol{\mu} \leq \hat{\boldsymbol{\mu}}^U$

Step 2: $r^L = \max_{\{\mathbf{M}, \sum_{k=1}^K M_k = K, M_k \leq \frac{-1}{\log(1-p)}\}} \langle \hat{\boldsymbol{\mu}}^L, g(\mathbf{M}) \rangle$

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?
- Maintain a lower bound ν of ν^*

Building a lower bound on ν^*

Step 0: $\nu = 0$

Step 1: $\hat{\boldsymbol{\mu}}^L$, $\hat{\boldsymbol{\mu}}^U$ such that whp: $\hat{\boldsymbol{\mu}}^L \leq \boldsymbol{\mu} \leq \hat{\boldsymbol{\mu}}^U$

Step 2: $r^L = \max_{\{\mathbf{M}, \sum_{k=1}^K M_k = K, M_k \leq \frac{-1}{\log(1-p)}\}} \langle \hat{\boldsymbol{\mu}}^L, g(\mathbf{M}) \rangle$

Step 3: $r^U_\nu = \max_{\{\mathbf{M}, \sum_{k=1}^K M_k = K, M_k \leq \frac{-1}{\log(1-p)}, \|\mathbf{M}^*\| = \nu\}} \langle \hat{\boldsymbol{\mu}}^U, g(\mathbf{M}) \rangle$

What could go wrong ?

When greedy fails

- Play Greedy assuming $\nu^* = 0$ (full support)
  What if $\nu^* > 0$?
- Maintain a lower bound ν of ν^*

Building a lower bound on ν^*
Step 0: $\nu = 0$

Step 1: $\hat{\boldsymbol{\mu}}^L$, $\hat{\boldsymbol{\mu}}^U$ such that whp: $\hat{\boldsymbol{\mu}}^L \leq \boldsymbol{\mu} \leq \hat{\boldsymbol{\mu}}^U$

Step 2: $r^L = \max_{\{\mathbf{M}, \sum_{k=1}^K M_k = K, M_k \leq \frac{-1}{\log(1-p)}\}} \langle \hat{\boldsymbol{\mu}}^L, g(\mathbf{M}) \rangle$

Step 3: $r_\nu^U = \max_{\{\mathbf{M}, \sum_{k=1}^K M_k = K, M_k \leq \frac{-1}{\log(1-p)}, \|\mathbf{M}^*\| = \nu\}} \langle \hat{\boldsymbol{\mu}}^U, g(\mathbf{M}) \rangle$

Step 4: If $r_\nu^H < r^L$: $\nu^* > \nu \rightarrow$ Set $\nu = \nu + 1$ and go to Step 3

Estimating the support

Wrong support

- Estimate ν, $\nu \leq \nu^*$

Estimating the support

Wrong support

- Estimate ν, $\nu \leq \nu^*$
- If $\nu = 0$ play Greedy
  $\operatorname{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k = M, M_k \leq -\frac{1}{\log(1-p)}, M_k > 0\}} \langle \hat{\mu}(t), g(\mathbf{M}) \rangle$

Estimating the support

Wrong support

- Estimate ν, $\nu \leq \nu^*$
- If $\nu = 0$ play Greedy
  $\operatorname{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k = M, M_k \leq -\frac{1}{\log(1-p)}, M_k > 0\}} \langle \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}) \rangle$
- If $\nu > 0$, many supports are possible, how to choose ?

Estimating the support

Wrong support

- Estimate ν, $\nu \leq \nu^*$
- If $\nu = 0$ play Greedy
  $\operatorname{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k = M, M_k \leq -\frac{1}{\log(1-p)}, M_k > 0\}} \langle \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}) \rangle$
- If $\nu > 0$, many supports are possible, how to choose ?

Successive accept and reject

(Bubeck, 2012)

Estimating the support

Wrong support

- Estimate ν, $\nu \leq \nu^*$
- If $\nu = 0$ play Greedy
  $\operatorname{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k = M, M_k \leq -\frac{1}{\log(1-p)}, M_k > 0\}} \langle \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}) \rangle$
- If $\nu > 0$, many supports are possible, how to choose ?

Successive accept and reject

(Bubeck, 2012)

- If $\hat{\mu}_k^U < \hat{\mu}_{(\nu+1)}^L \rightarrow$ Reject

Estimating the support

Wrong support

- Estimate ν, $\nu \leq \nu^*$
- If $\nu = 0$ play Greedy
  $\text{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k = M, M_k \leq -\frac{1}{\log(1-p)}, M_k > 0\}} \langle \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}) \rangle$
- If $\nu > 0$, many supports are possible, how to choose ?

Successive accept and reject
(Bubeck, 2012)

- If $\hat{\mu}_k^U < \hat{\mu}_{(\nu+1)}^L \rightarrow$ Reject
- If $\hat{\mu}_k^L > \hat{\mu}_{(\nu)}^U \rightarrow$ Accept

Estimating the support

Successive accept and reject
(Bubeck, 2012)

- If $\hat{\mu}_k^U < \hat{\mu}_{(\nu+1)}^L \rightarrow$ Reject
- If $\hat{\mu}_k^L > \hat{\mu}_{(\nu)}^U \rightarrow$ Accept

Estimating the support

Successive accept and reject
(Bubeck, 2012)

- If $\hat{\mu}_k^U < \hat{\mu}_{(\nu+1)}^L \rightarrow$ Reject
- If $\hat{\mu}_k^L > \hat{\mu}_{(\nu)}^U \rightarrow$ Accept

Rejects and accepts

- What does it mean to reject ? A rejected arm is never assigned players again

Estimating the support

Successive accept and reject
(Bubeck, 2012)

- If $\hat{\mu}_k^U < \hat{\mu}_{(\nu+1)}^L \rightarrow$ Reject
- If $\hat{\mu}_k^L > \hat{\mu}_{(\nu)}^U \rightarrow$ Accept

Rejects and accepts

- What does it mean to reject ? A rejected arm is never assigned players again
- What does it mean to accept ? An accepted arm is played at every round until ν increases

Estimating the support

Successive accept and reject
(Bubeck, 2012)

- If $\hat{\mu}_k^U < \hat{\mu}_{(\nu+1)}^L \to$ Reject
- If $\hat{\mu}_k^L > \hat{\mu}_{(\nu)}^U \to$ Accept

Rejects and accepts

- What does it mean to reject ? A rejected arm is never assigned players again
- What does it mean to accept ? An accepted arm is played at every round until ν increases
- Rotate among other arms in a Round Robin fashion

Illustration

Play

$$\operatorname{argmax}_{\{\mathbf{M}, \sum_{k=1}^{K} M_k = K, M_k \leq -\frac{1}{\log(1-p)}, M_2 > 0, M_3 > 0, M_4 > 0\}} \langle \hat{\boldsymbol{\mu}}(t), g(\mathbf{M}) \rangle$$

Illustration

 K

$\nu^* = 0$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\emptyset$

Rejected arms: $\emptyset$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\emptyset$

Rejected arms: $\emptyset$

Illustration

K

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

K

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 1$

Accepted arms: $\{2\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\emptyset$

Rejected arms: $\emptyset$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

K

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

K

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\emptyset$

Rejected arms: $\emptyset$

Illustration

$\nu^* = 2$

Accepted arms: $\varnothing$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

K

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\varnothing$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Illustration

$\nu^* = 2$

Accepted arms: $\{1\}$

Rejected arms: $\{3\}$

Analysis

Regret decomposition
Call

$$\mathbf{M}_\nu = \operatorname{argmax}_{\mathbf{M}, \|\mathbf{M}\|_0 = \nu} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

Analysis

Regret decomposition
Call

$$\mathbf{M}_\nu = \mathrm{argmax}_{\mathbf{M}, \|\mathbf{M}\|_0 = \nu} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\mathbf{M}_\mathcal{E} = \mathrm{argmax}_{\mathbf{M}, \forall k \in \mathcal{E}, M_k > 0} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

Analysis

Regret decomposition

Call

$$\mathbf{M}_\nu = \mathrm{argmax}_{\mathbf{M}, \|\mathbf{M}\|_0 = \nu} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\mathbf{M}_\mathcal{E} = \mathrm{argmax}_{\mathbf{M}, \forall k \in \mathcal{E}, M_k > 0} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$R = \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]$$

Analysis

Regret decomposition
Call

$$\mathbf{M}_\nu = \mathrm{argmax}_{\mathbf{M}, \|\mathbf{M}\|_0 = \nu} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\mathbf{M}_\mathcal{E} = \mathrm{argmax}_{\mathbf{M}, \forall k \in \mathcal{E}, M_k > 0} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$R = \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle]$$
$$= \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}_{\nu(t)}) \rangle]$$

Analysis

Regret decomposition
Call

$$\mathbf{M}_\nu = \text{argmax}_{\mathbf{M}, \|\mathbf{M}\|_0 = \nu} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\mathbf{M}_\mathcal{E} = \text{argmax}_{\mathbf{M}, \forall k \in \mathcal{E}, M_k > 0} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\begin{aligned} R &= \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\ &= \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}_{\nu(t)}) \rangle] + \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}_{\nu(t)}) - g(\mathbf{M}_{\mathcal{E}(t)}) \rangle] \end{aligned}$$

Analysis

Regret decomposition
Call

$$\mathbf{M}_\nu = \operatorname{argmax}_{\mathbf{M}, \|\mathbf{M}\|_0 = \nu} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\mathbf{M}_\mathcal{E} = \operatorname{argmax}_{\mathbf{M}, \forall k \in \mathcal{E}, M_k > 0} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$$

$$\begin{aligned} R &= \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}(t)) \rangle] \\ &= \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}^*) - g(\mathbf{M}_{\nu(t)}) \rangle] + \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}_{\nu(t)}) - g(\mathbf{M}_{\mathcal{E}(t)}) \rangle] \\ &+ \mathbb{E}[\sum_{t=1}^{T} \langle \boldsymbol{\mu}, g(\mathbf{M}_{\mathcal{E}(t)}) - g(\mathbf{M}(t)) \rangle] \end{aligned}$$

Regret due to the mismatch between ν and ν^*

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*) \rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*) \rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$

How long it takes to increase ν

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*) \rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M}) \rangle$

How long it takes to increase ν

ν increases when:

$\langle \hat{\boldsymbol{\mu}}^L(t), g(\mathbf{M}^*) \rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \hat{\boldsymbol{\mu}}^H(t), g(\mathbf{M}) \rangle \iff$

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M},\|\mathbf{M}\|_0=\nu\}}\langle \boldsymbol{\mu}, g(\mathbf{M})\rangle$

How long it takes to increase ν

ν increases when:

$\langle \hat{\boldsymbol{\mu}}^L(t), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M},\|\mathbf{M}\|_0=\nu\}}\langle \hat{\boldsymbol{\mu}}^H(t), g(\mathbf{M})\rangle \iff$

$$\langle \boldsymbol{\mu} - O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M},\|\mathbf{M}\|_0=\nu\}}\langle \boldsymbol{\mu} + O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M})\rangle$$

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle$

How long it takes to increase ν

ν increases when:

$\langle \hat{\boldsymbol{\mu}}^L(t), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \hat{\boldsymbol{\mu}}^H(t), g(\mathbf{M})\rangle \iff$

$$\langle \boldsymbol{\mu} - O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu} + O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M})\rangle$$

$$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle > O(\sqrt{\frac{\log(T)}{t}})$$

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle$

How long it takes to increase ν

ν increases when:

$\langle \hat{\boldsymbol{\mu}}^L(t), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \hat{\boldsymbol{\mu}}^H(t), g(\mathbf{M})\rangle \iff$

$$\langle \boldsymbol{\mu} - O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu} + O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M})\rangle$$

$$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0 = \nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle > O(\sqrt{\frac{\log(T)}{t}})$$

Regret due to the mismatch between ν and ν^*

Cost of using ν instead of ν^*

$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0=\nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle$

How long it takes to increase ν

ν increases when:

$\langle \hat{\boldsymbol{\mu}}^L(t), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0=\nu\}} \langle \hat{\boldsymbol{\mu}}^H(t), g(\mathbf{M})\rangle \iff$

$$\langle \boldsymbol{\mu} - O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M}^*)\rangle > \max_{\{\mathbf{M}, \|\mathbf{M}\|_0=\nu\}} \langle \boldsymbol{\mu} + O(\sqrt{\frac{\log(T)}{t}}), g(\mathbf{M})\rangle$$

$$\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0=\nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle > O(\sqrt{\frac{\log(T)}{t}})$$

It takes $t = O(\frac{\log(T)}{(\langle \boldsymbol{\mu}, g(\mathbf{M}^*)\rangle - \max_{\{\mathbf{M}, \|\mathbf{M}\|_0=\nu\}} \langle \boldsymbol{\mu}, g(\mathbf{M})\rangle)^2})$ steps

Communication

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$
- If $2^s < 16M\frac{\log(2TM)}{\log(1-p_g)}$ play greedy with $\nu = 0$

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$
- If $2^s < 16M\frac{\log(2TM)}{\log(1-p_g)}$ play greedy with $\nu = 0$
- Otherwise, only make updates ($\hat{\boldsymbol{\mu}}$, accepted arms, rejected arms, ν) at the end of a phase

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$
- If $2^s < 16M\frac{\log(2TM)}{\log(1-p_g)}$ play greedy with $\nu = 0$
- Otherwise, only make updates ($\hat{\boldsymbol{\mu}}$, accepted arms, rejected arms, ν) at the end of a phase
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to communicate statistics

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$
- If $2^s < 16M\frac{\log(2TM)}{\log(1-p_g)}$ play greedy with $\nu = 0$
- Otherwise, only make updates ($\hat{\boldsymbol{\mu}}$, accepted arms, rejected arms, ν) at the end of a phase
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to communicate statistics
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to receive statistics

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$
- If $2^s < 16M\frac{\log(2TM)}{\log(1-p_g)}$ play greedy with $\nu = 0$
- Otherwise, only make updates ($\hat{\boldsymbol{\mu}}$, accepted arms, rejected arms, ν) at the end of a phase
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to communicate statistics
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to receive statistics

Phases costs a factor 2

Communication

- Split the algorithm in phases of size $(2^s)_{s=1}^{\log(T)}$
- If $2^s < 16M\frac{\log(2TM)}{\log(1-p_g)}$ play greedy with $\nu = 0$
- Otherwise, only make updates ($\hat{\boldsymbol{\mu}}$, accepted arms, rejected arms, ν) at the end of a phase
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to communicate statistics
- At each phase, reserve $\frac{1}{4M}$ rounds for player m to receive statistics

Phases costs a factor 2

The low probability of miscommunication compensates errors

Lower bounds

$\nu^* = 0$
$K = 2$ arms, $M = 2N + 1$ players $p \leq \frac{1}{M+1}$, $r_0 < \frac{p}{12}$,
$T \geq \frac{1}{16g(M)r_0^2}$. For any algorithm A, there exists rewards $\boldsymbol{\mu}$ s.t
$r(\boldsymbol{\mu}) = r_0$ and

$$\mathbb{E}[R_A] \geq \frac{1}{256Mr_0}$$

Lower bounds

$\nu^* = 1$

For any $M \geq 5, \nu^* > 0, p \leq \frac{1}{M+1}$, any gaps $\Delta_{(1)}, \ldots, \Delta_{(\nu^*)} \leq \frac{p}{8(M-4)}$, and for any consistent algorithm A, there exists $(\mu_1, \ldots, \mu_{\nu^*+2})$ s.t $\mu_{(\nu^*+1)} - \mu_{(\nu)} = \Delta_{(\nu)}$ for all $\nu \in [\nu^*]$ and for some c:

$$\liminf_{T\to\infty} \frac{\mathbb{E}R_A}{\log(T)} \geq \sum_{\nu=1}^{\nu^*} \frac{c}{\Delta_{(\nu)}} .$$

Lower bounds

$\nu^* = 0$

Either $\boldsymbol{\mu} = (1/2, 1/2 + \Delta)$, $\boldsymbol{\mu} = (1/2 + \Delta, 1/2)$ with $\Delta \leq \frac{p}{2}$

1. $r = \Delta/2$.
2. Best solutions are $\mathbf{M}^* = (N, N+1)$ or $\mathbf{M}^* = (N+1, N)$
3. Similar to a 2-arm bandits with full info: $\mathbb{E}[R_A] \geq \frac{\exp(-1)}{128\Delta}$

Lower bounds

$\nu^* = 0$

Either $\boldsymbol{\mu} = (1/2, 1/2 + \Delta)$, $\boldsymbol{\mu} = (1/2 + \Delta, 1/2)$ with $\Delta \leq \frac{p}{2}$

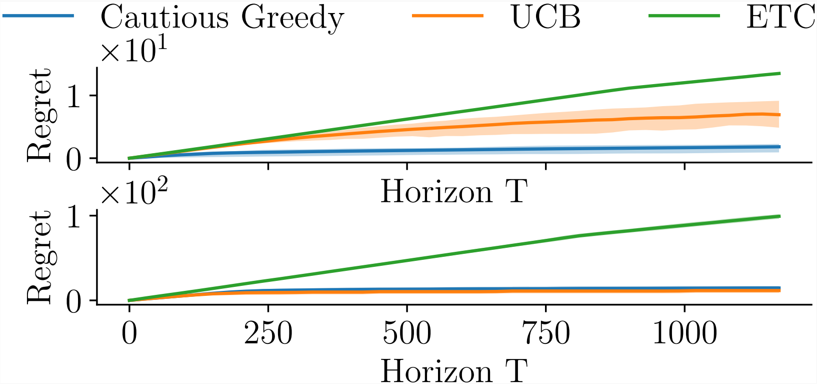
1. $r = \Delta/2$.
2. Best solutions are $\mathbf{M}^* = (N, N+1)$ or $\mathbf{M}^* = (N+1, N)$
3. Similar to a 2-arm bandits with full info: $\mathbb{E}[R_A] \geq \frac{\exp(-1)}{128\Delta}$

$\nu^* = 1$

Either $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_1 + \Delta)1/2 + \Delta)$ or $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_1 - \Delta$ with $\Delta \leq \frac{p}{2}$

1. Best solutions are $\mathbf{M}^* = (M-1, 1, 0)$ or $(M-1, 0, 1)$
2. $\mathbb{E}[R_A] \geq \frac{\log(T)}{\Delta}$

Simulations



$\nu^* = 0$ (top) $\nu^* = 1$ (bottom)

Conclusion

Conclusion

Contribution

- Cautious Greedy: optimal dependency in T, r and $(\mu_{\nu^*+1} - \mu_\nu)_{\nu=1}^{\nu^*}$
- Average $\log(T)$ communication steps

Conclusion

Contribution

- Cautious Greedy: optimal dependency in T, r and $(\mu_{\nu^*+1} - \mu_\nu)_{\nu=1}^{\nu^*}$
- Average $\log(T)$ communication steps

Future work

- No communications (Selfish algorithms)
- Better dependency in K, M, p, p_g
- Anytime version

Conclusion

Contribution

- Cautious Greedy: optimal dependency in T, r and $(\mu_{\nu^*+1} - \mu_\nu)_{\nu=1}^{\nu^*}$
- Average $\log(T)$ communication steps

Future work

- No communications (Selfish algorithms)
- Better dependency in K, M, p, p_g
- Anytime version

Thank you !