Censored Semi-Bandits: A Framework for Resource Allocation with Censored Feedback

Arun Verma, IIT Bombay, India, v.arun@iitb.ac.in

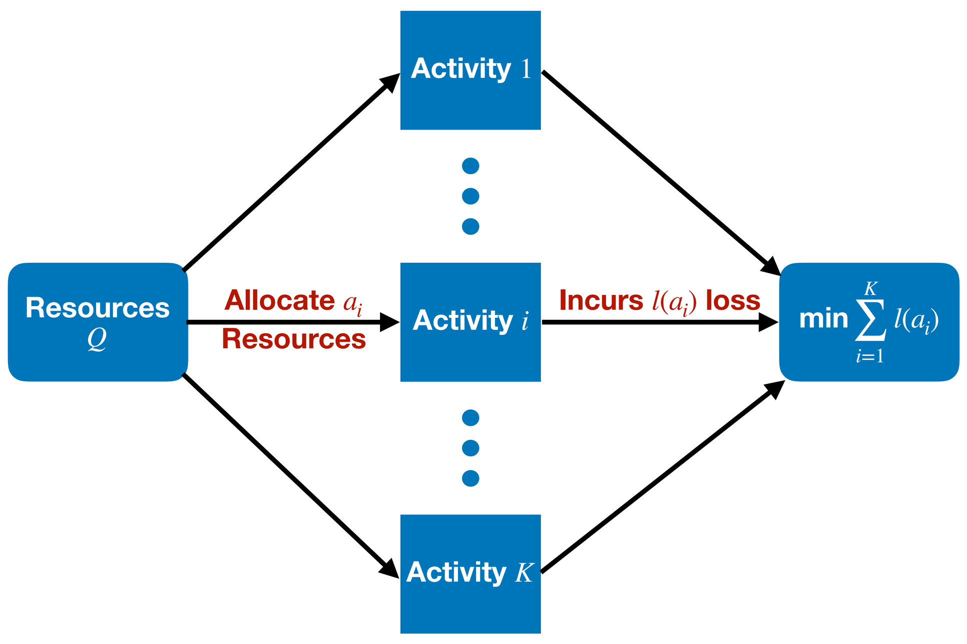
Manjesh K. Hanawal, IIT Bombay, India, mhanawal@iitb.ac.in

Arun Rajkumar, IIT Madras, India, arunr@cse.iitm.ac.in

Raman Sankaran, LinkedIn India, rsankara@linkedin.com

Resource Allocation Problem

- Fixed amount of resources need to be allocated among different activities such that the total mean loss is minimized.

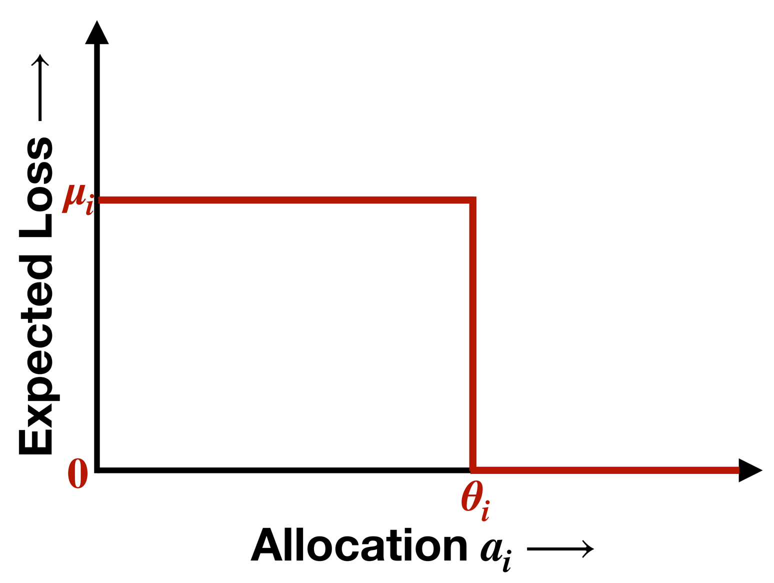


Resource Allocation Problem.

- Examples: advertisement budget allocation, police patrolling, supplier selection, etc.

Problem Setup

- Amount of resources: Q
- Number of arms (activities): K
- Resource allocated to arm i: a_i where $a_i \in \mathbb{R}_+$
- Allocation vector: $\boldsymbol{a} := \{a_i : i \in \{1, \ldots, K\}\}$
- Allocation $\boldsymbol{a}$ is feasible if $\sum_{i=1}^{K} a_i \leq Q$. The set of all feasible allocations is denoted by $\mathcal{A}_Q$.
- Expected loss observed from arm i is:



$$\mathbb{E}\left[l(a_i)\right] = \mu_i \mathbb{1}_{\{a_i < \theta_i\}}$$

where μ_i is the mean loss of arm i and θ_i is the associated threshold with arm i.

- **Note that both $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ are unknown.**

Goal: Find an resource allocation that minimizes the total mean loss.

Performance Measure: Regret

- Expected (pseudo) regret over a period of T:

$$\mathbb{E}\left[\mathcal{R}_T\right] = \sum_{t=1}^{T} \sum_{i=1}^{K} \mu_i \left(\mathbb{1}_{\{a_{t,i} < \theta_i\}} - \mathbb{1}_{\{a_i^\star < \theta_i\}} \right)$$

where $\boldsymbol{a}^\star \in \arg\min_{\boldsymbol{a} \in \mathcal{A}_Q} \sum_{i=1}^{K} \mu_i \mathbb{1}_{\{a_i < \theta_i\}}$.

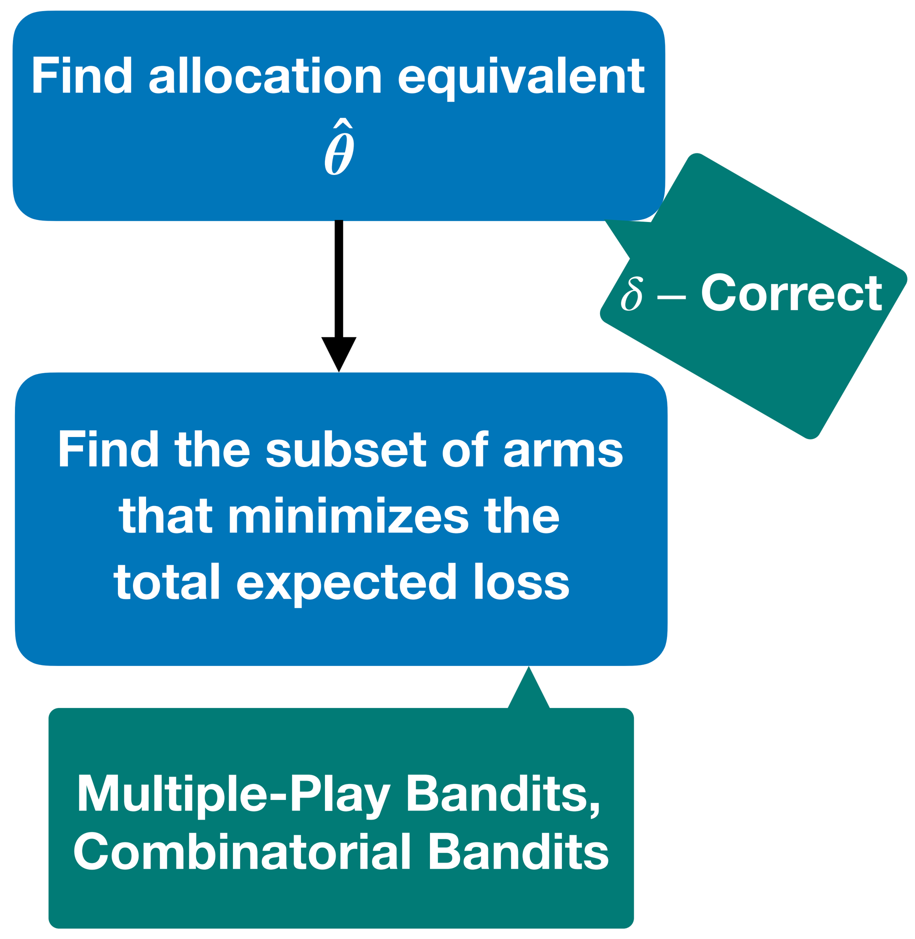
- A good policy should have sub-linear expected regret, i.e., $\mathbb{E}\left[\mathcal{R}_T\right]/T \to 0$ as $T \to \infty$.

Allocation Equivalent

- For fix $\boldsymbol{\mu}$ and Q, $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ are `allocation equivalent` iff:

$$\min_{\boldsymbol{a} \in \mathcal{A}_Q} \sum_{i=1}^{K} \mu_i \mathbb{1}_{\{a_i < \theta_i\}} = \min_{\boldsymbol{a} \in \mathcal{A}_Q} \sum_{i=1}^{K} \mu_i \mathbb{1}_{\{a_i < \hat{\theta}_i\}}.$$

Algorithm Idea



- **$\boldsymbol{\delta}$−correct:** $\hat{\boldsymbol{\theta}}$ is an allocation equivalent to $\boldsymbol{\theta}$ with probability at least $1 - \delta$.
- Once $\hat{\theta}$ is known, a subset of arms is selected (using MP-TS [1] for the same threshold case and CTS [2] for the different threshold case) such that the total mean loss is minimized.

Algorithms

CSB-ST for Same Threshold Case

- $\forall i \in [K] : \theta_i = \theta_c$ where $\theta_c \in \mathbb{R}^+$ and $Q \geq \theta_c$.
- Let $M = \min\{\lfloor Q/\theta_c \rfloor, K\}$. Then the optimal allocation allocates the θ_c fraction of resources to top M arms with highest mean loss.
- Let $\Theta = \{Q/K, Q/(K-1), \cdots, Q\}$. Then the allocation equivalent of θ_c is $\hat{\theta}_c$ where $\hat{\theta}_c \in \Theta$.

CSB-DT for Different Threshold Case

- Threshold may not be the same for all arms.
- Optimal allocation is the solution of 0-1 knapsack having capacity Q and K items where item i has weight θ_i and value μ_i.
- Define $\gamma := \left(Q - \sum_{a_i^\star \geq \theta_i} \theta_i\right)/K > 0$ and $\forall i \in [K] : \hat{\theta}_i \in [\theta_i, \theta_i + \gamma]$ where $\theta_i \in [0, 1]$. Then $\hat{\boldsymbol{\theta}}$ is allocation equivalent of $\boldsymbol{\theta}$.
- All $\hat{\boldsymbol{\theta}}_i$ are estimated by using binary search in $[0, 1]$ interval.

Regret Bounds

- Let $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{K-M} < \mu_{K-M+1} \leq \ldots \leq \mu_K$, $\min_{i \in \{1,\ldots,K\}} \mu_i \geq \epsilon > 0$, $\Delta_a = \sum_{i=1}^{K} \mu_i (\mathbb{1}_{\{a_i < \theta_i\}} - \mathbb{1}_{\{a_i^\star < \theta_i\}})$, $\Delta_m = \max_{\boldsymbol{a} \in \mathcal{A}_Q} \Delta_a$, and $\delta = 1/T$. Then the expected regret of CSB-ST over a period of T is given by

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq \frac{\log(T \log_2(|\Theta|)) \log_2(|\Theta|) \Delta_m}{\max\{1, \lfloor Q \rfloor\} \log\left(1/(1-\epsilon)\right)} + O\left(\sum_{i \in [K] \setminus [K-M]} \frac{(\mu_i - \mu_{K-M}) \log T}{d(\mu_{K-M}, \mu_i)} \right).$$

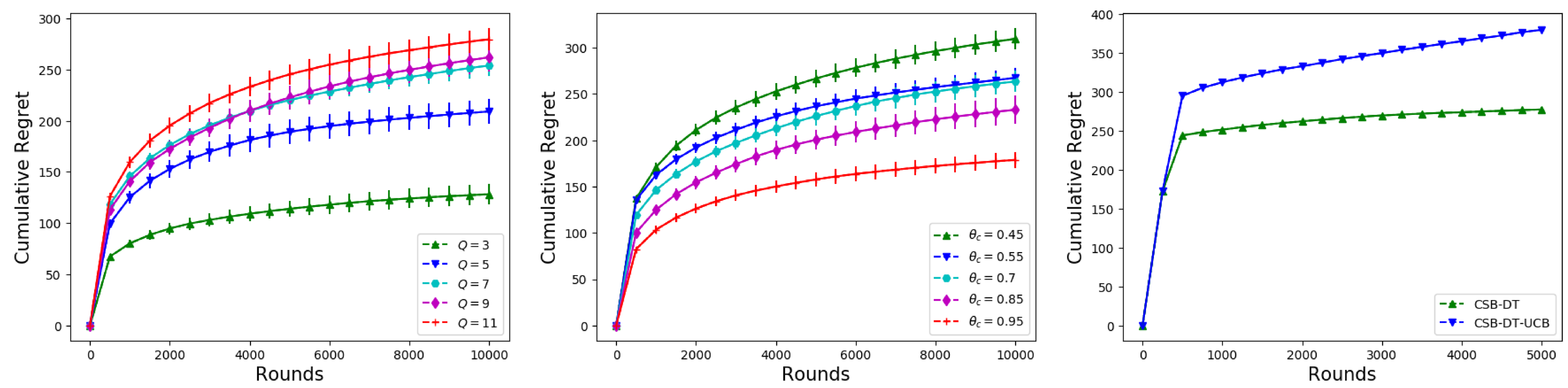
- Let $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_K$, $\min_{i \in \{1,\ldots,K\}} \mu_i \geq \epsilon > 0$, $\gamma > 0$, $S_a = \{i : a_i < \theta_i\}$ for any feasible allocation a, and $k^\star = |S_{a^\star}|$. Then for any η such that $\forall \boldsymbol{a} \in \mathcal{A}_Q, \Delta_a > 2(k^{\star 2} + 2)\eta$, the expected regret of CSB-DT over a period of T is given by

$$\mathbb{E}\left[\mathcal{R}_T\right] \leq \frac{K \log\left(KT \log_2\left(\lceil 1 + 1/\gamma \rceil\right)\right) \log_2\left(\lceil 1 + 1/\gamma \rceil\right) \Delta_m}{\max\{1, \lfloor Q \rfloor\} \log\left(1/(1-\epsilon)\right)} + O\left(\sum_{i \in [K]} \max_{S_a : i \in S_a} \frac{8|S_a| \log T}{\Delta_a - 2(k^{\star 2} + 2)\eta} \right).$$

- The first term in the regret bounds corresponds to the regret incurred during threshold estimation.

Experiment Results

- **Same Threshold Problem Instance:** $K = 20, Q = 7, \theta_c = 0.7, \delta = 0.1, \epsilon = 0.1$ and $T = 10000$. The loss of arm i is Bernoulli distribution with parameter $0.25 + (i-1)/50$.
- **Different Thresholds Problem Instance:** $K = 5, Q = 2, \delta = 0.1, \epsilon = 0.1, \gamma = 10^{-3}$ and $T = 5000$. The mean loss vector is $\boldsymbol{\mu} = [0.9, 0.89, 0.87, 0.58, 0.3]$ and corresponding threshold vector is $\boldsymbol{\theta} = [0.7, 0.7, 0.7, 0.6, 0.35]$. The loss of arm i is Bernoulli distributed with parameter μ_i.



Cumulative Regret of CSB-ST v/s Amount of Resource (Leftmost Fig.) and Different Values of Same Threshold (Middle Fig.). Comparing CSB-DT with UCB based Algorithm CDB-DT-UCB (Rightmost Fig.).

Future Directions

- We decoupled the problem of threshold and mean loss estimation. It can be done jointly, leading to better performance guarantees.
- Another extension of our work is to relax the assumptions that mean losses are strictly positive, and time horizon T is known.

References

[1] Junpei Komiyama, Junya Honda, and Hiroshi Nakagawa. Optimal regret analysis of thompson sampling in stochastic multi-armed bandit problem with multiple plays. In *International Conference on Machine Learning*, pages 1152–1161, 2015.

[2] Siwei Wang and Wei Chen. Thompson sampling for combinatorial semi-bandits. In *International Conference on Machine Learning*, pages 5101–5109, 2018.

[3] Arun Verma, Manjesh K Hanawal, Arun Rajkumar, and Raman Sankaran. Censored semi-bandits: A framework for resource allocation with censored feedback. In *Neural Information Processing Systems (NeurIPS)*, 2019.