Dynamic Resource Allocation

Algorithmic Design Principles and Spectrum of Achievable Performances

Akshit Kumar

joint with Omar Besbes and Yash Kanoria

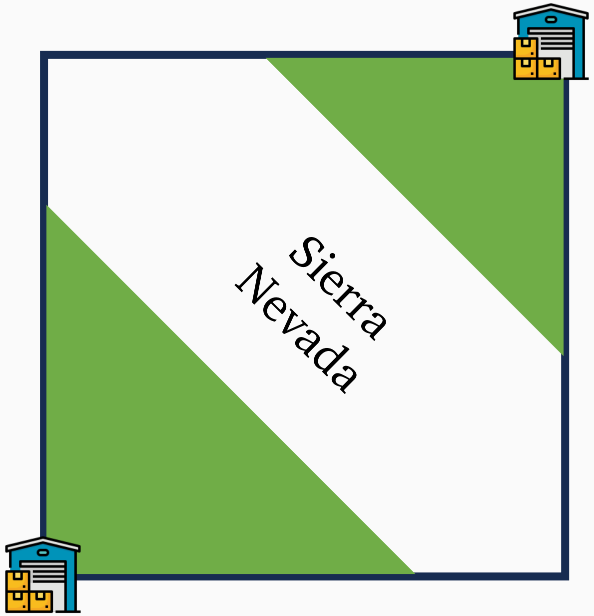
Columbia Business School

Dynamic Resource Allocation is ubiquitous

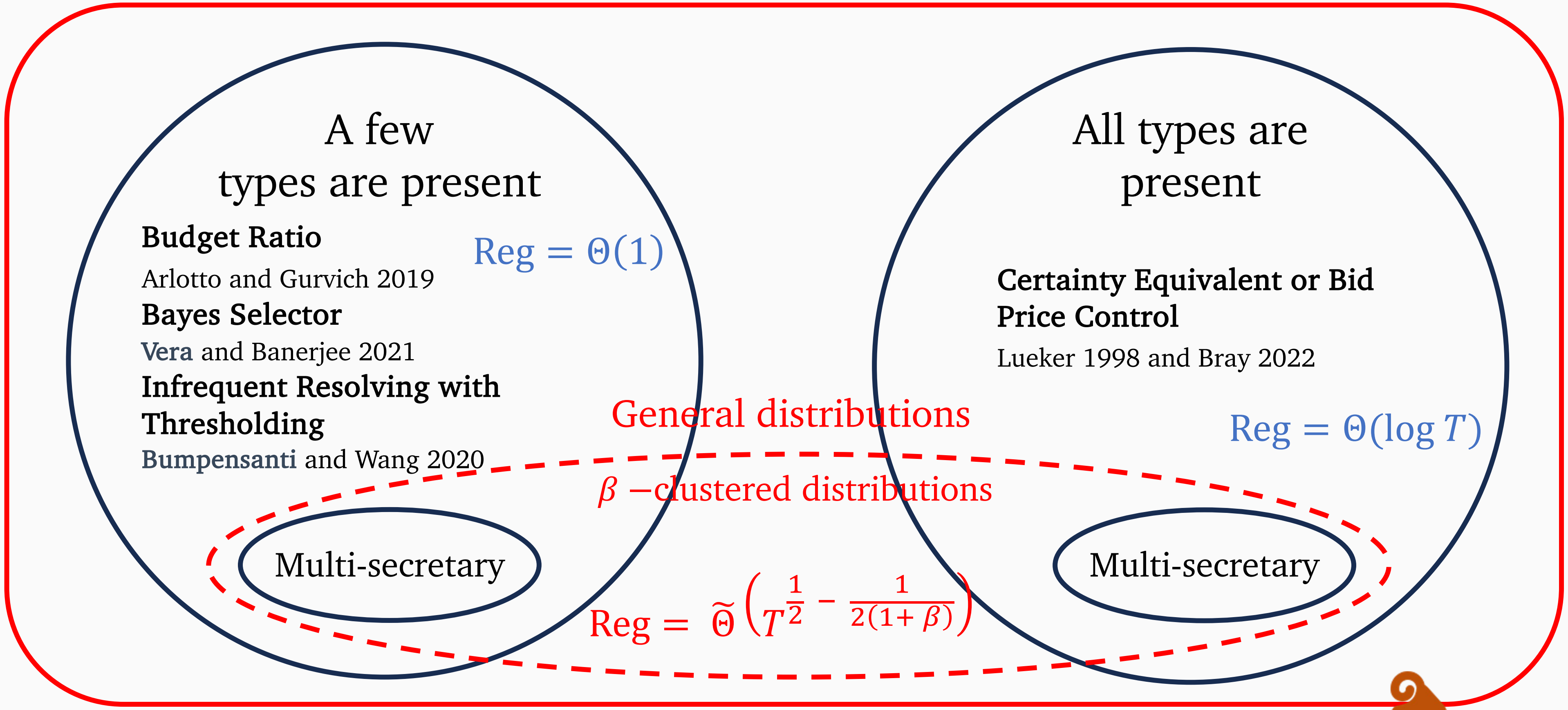
Network Revenue Management

Budget Management in Ad auctions

Order Fulfillment



In A Nutshell



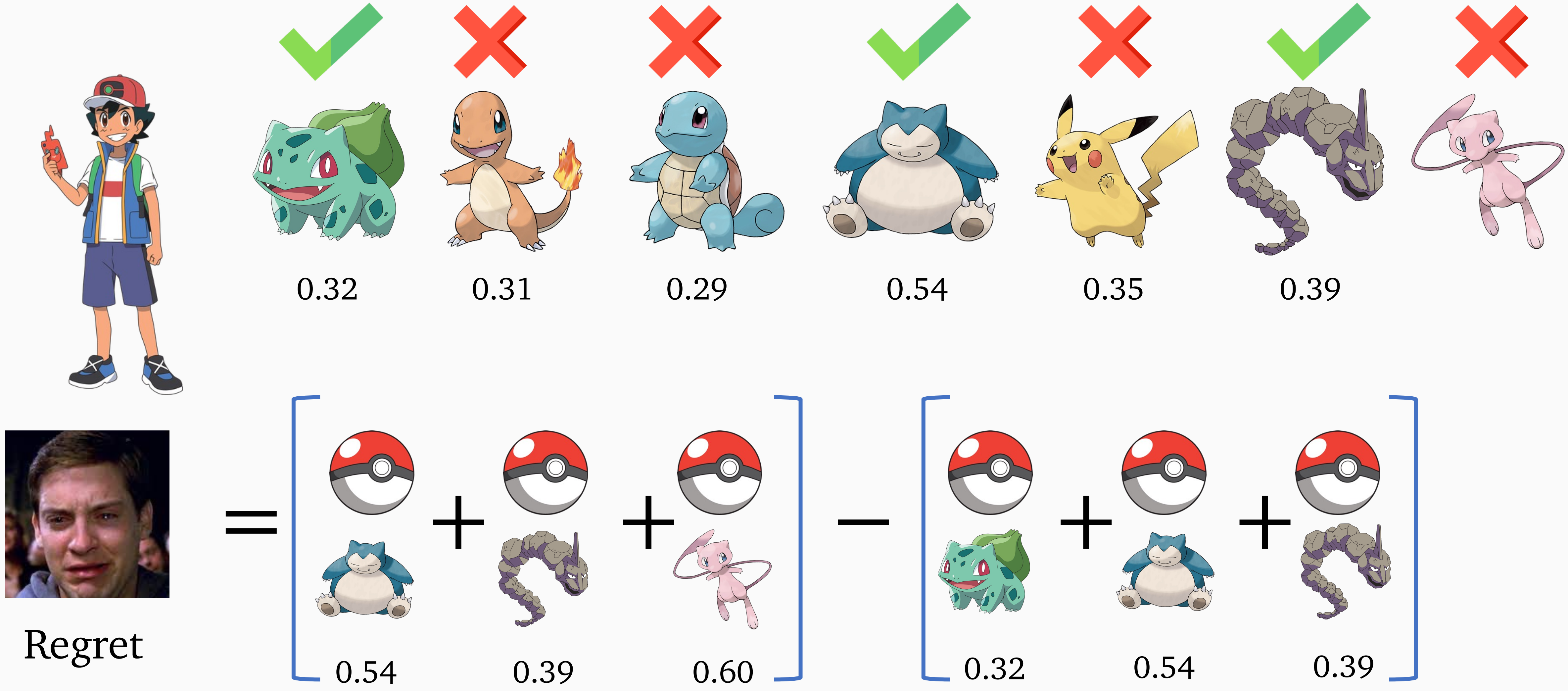
Repeatedly Act using Multiple Simulations

Regret is the additive gap between the value of hindsight optimal problem and value of problem under some algorithm

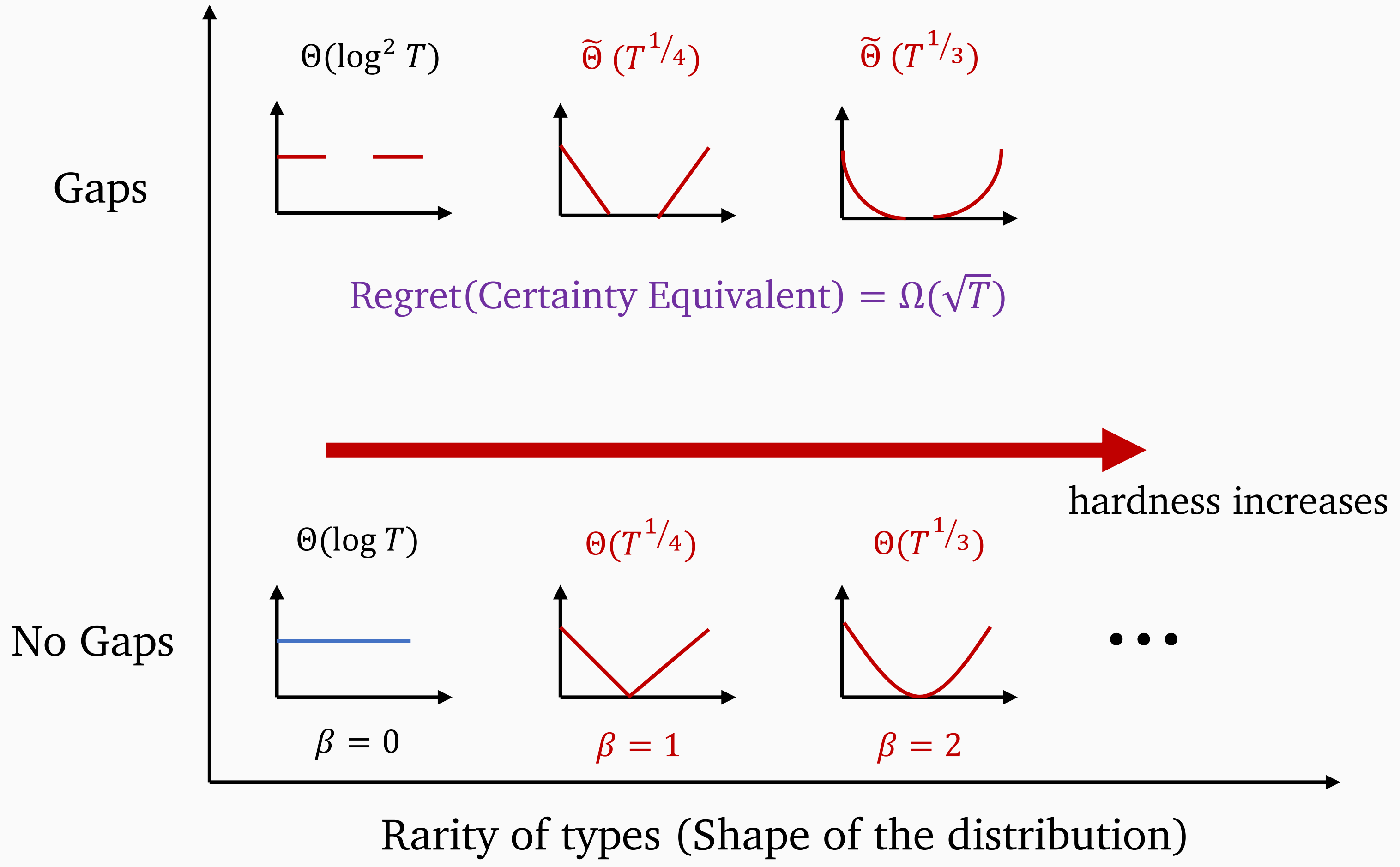
Multi-secretary problem

- Given a sequence of T secretaries and hiring budget B, a decision maker wants to hire the top B secretaries in terms of their ability.
- The secretaries arrive in an online fashion.
- The decision maker makes **irrevocable** hire or reject decisions.
- Assumption: The abilities (types) of the secretaries are drawn independently from a common and known distribution F.

Multi-Pokémon Problem



Spectrum of Achievable Performances



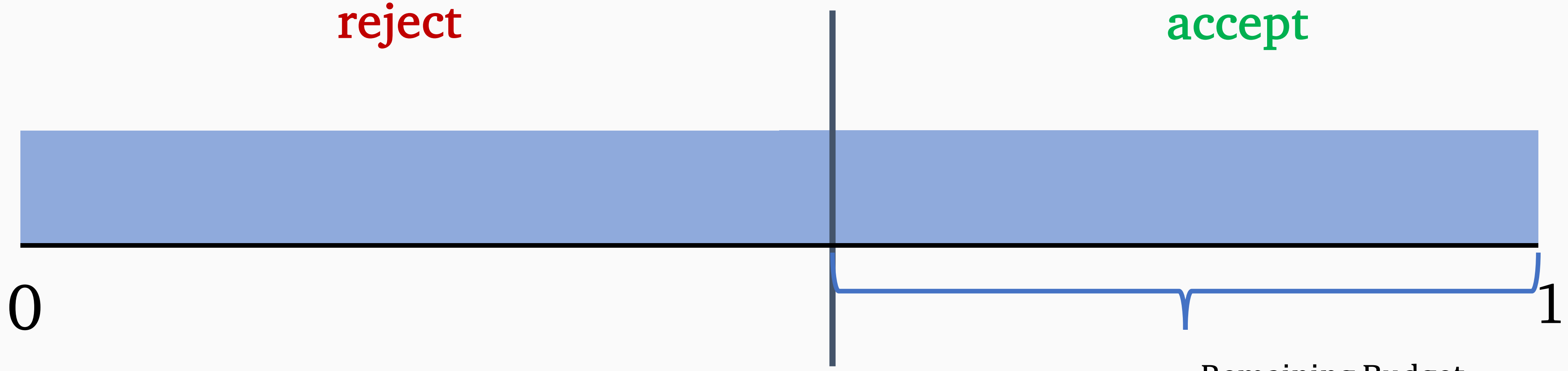
Distribution shape is fundamental driver of performance

Dealing with gaps is an algorithmic challenge

Conservativeness with respect to Gaps (CwG) principle enables near optimal performance

Use RAMS (Repeatedly Act using Multiple Scenarios) to operationalize CwG

Certainty Equivalent



$$\text{Budget Ratio} = \frac{\text{Remaining Budget}}{\text{Remaining Time}}$$

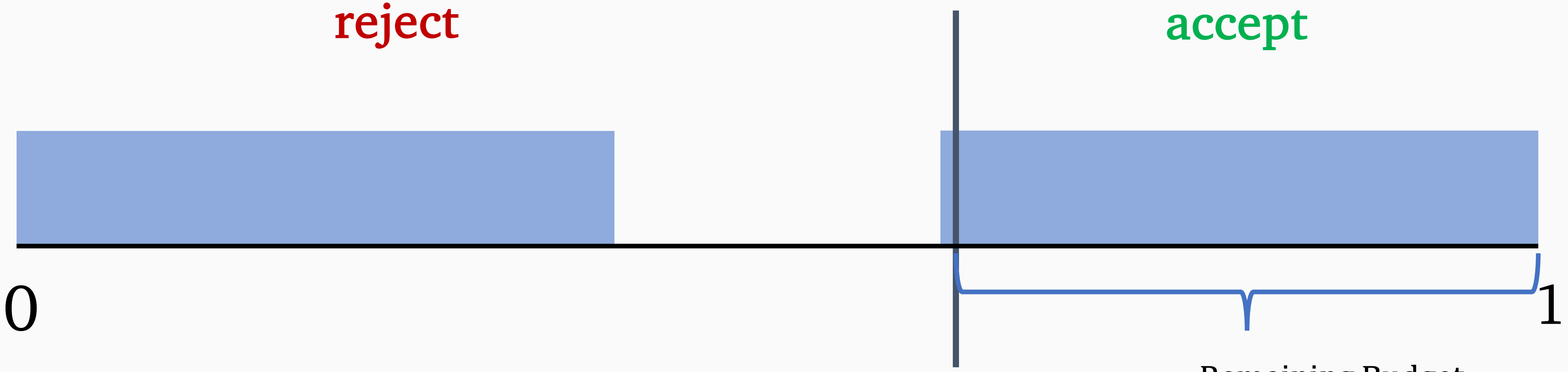
Uniform distribution (when all types are present)

$$\text{Regret(CE)} = \Theta(\log T)$$

(optimal regret scaling)

Lueker 1998 and Bray 2019

Certainty Equivalent



$$\text{Budget Ratio} = \frac{\text{Remaining Budget}}{\text{Remaining Time}}$$

Uniform distribution (when all types are present)

$$\text{Regret(CE)} = \Theta(\log T)$$

(optimal regret scaling)

Lueker 1998 and Bray 2019

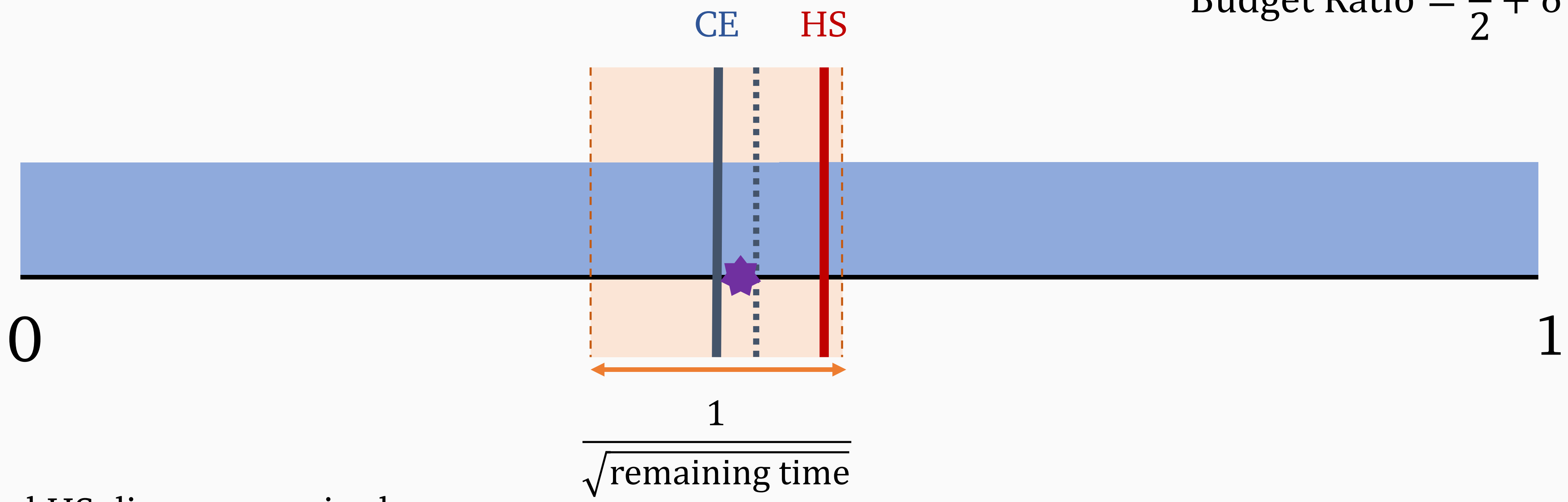
Bi-modal Uniform distribution (some types are absent)

$$\text{Regret(CE)} = \Omega(\sqrt{T})$$

(highly sub-optimal regret scaling)

Failure of Certainty Equivalent

$$\text{Budget Ratio} = \frac{1}{2} + \delta$$



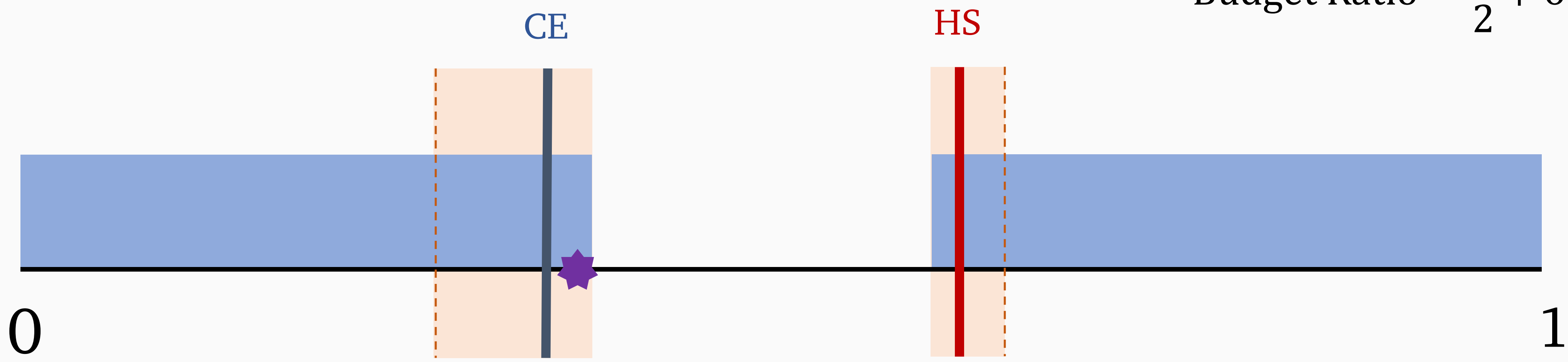
CE and HS disagree $\Rightarrow$ mistake

$\mathbb{E}[\text{cost of mistake}] \sim \mathbb{P}(\bigstar \in [\text{CE}, \text{HS}]) \times |\bigstar - \text{HS}|$

compensation $\quad \sim 1/\text{remaining time}$

Failure of Certainty Equivalent

$$\text{Budget Ratio} = \frac{1}{2} + \delta$$



CE and HS disagree $\Rightarrow$ mistake

$\mathbb{E}[\text{cost of mistake}] \sim \mathbb{P}(\star \in [\text{CE}, \text{HS}]) \times |\star - \text{HS}|$

compensation $\sim 1/\text{remaining time}$

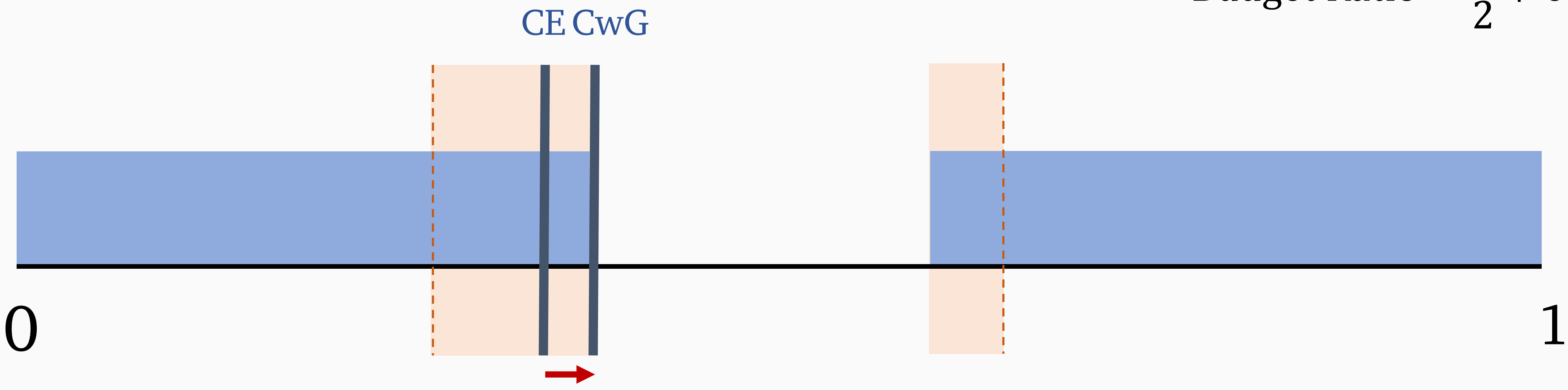
CE and HS disagree $\Rightarrow$ mistake

$\mathbb{E}[\text{cost of mistake}] \sim \mathbb{P}(\star \in [\text{CE}, \text{HS}]) \times |\star - \text{HS}|$

compensation $\sim 1/\sqrt{\text{remaining time}}$

CwG: Hedging against the risk of large regret in the future

Budget Ratio $= \frac{1}{2} + \delta$



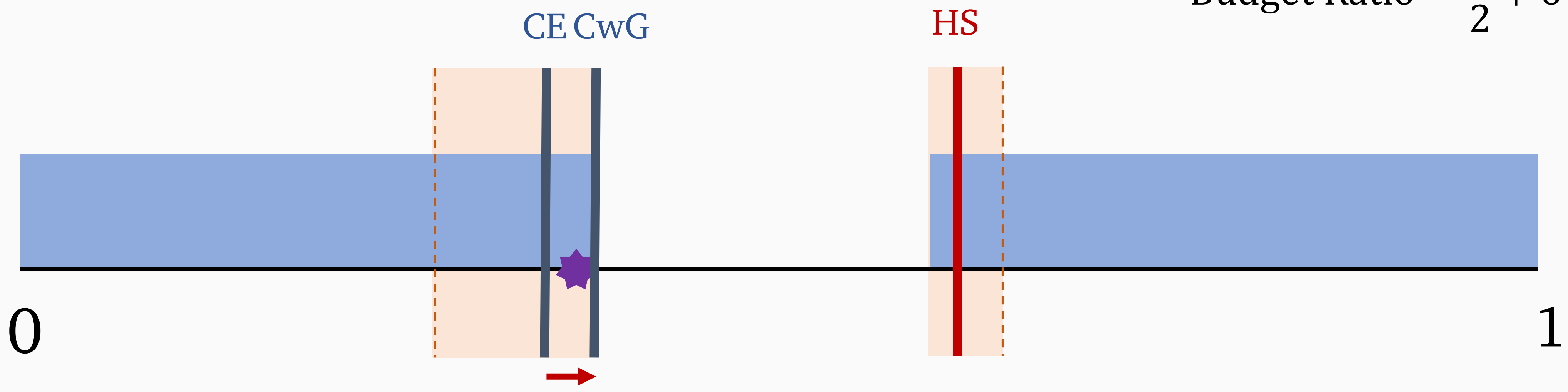
<u>Conservativeness with respect to gaps principle</u>

If the CE threshold is close to the gap, use the gap as the threshold.

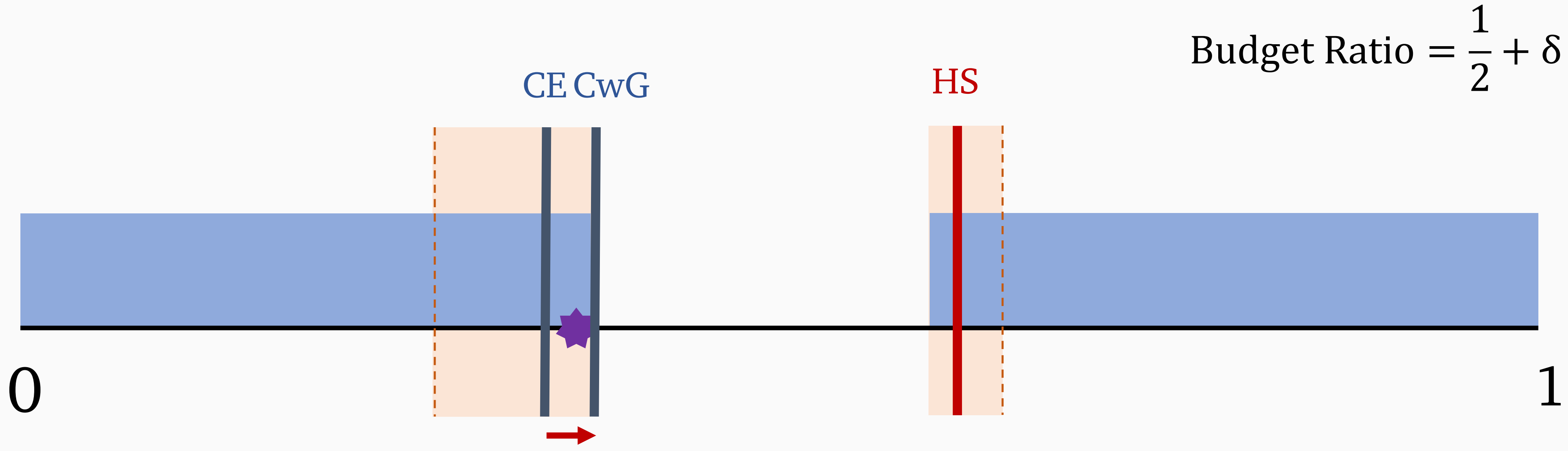
Otherwise use the CE threshold.

CwG: Hedging against the risk of large regret in the future

$$\text{Budget Ratio} = \frac{1}{2} + \delta$$



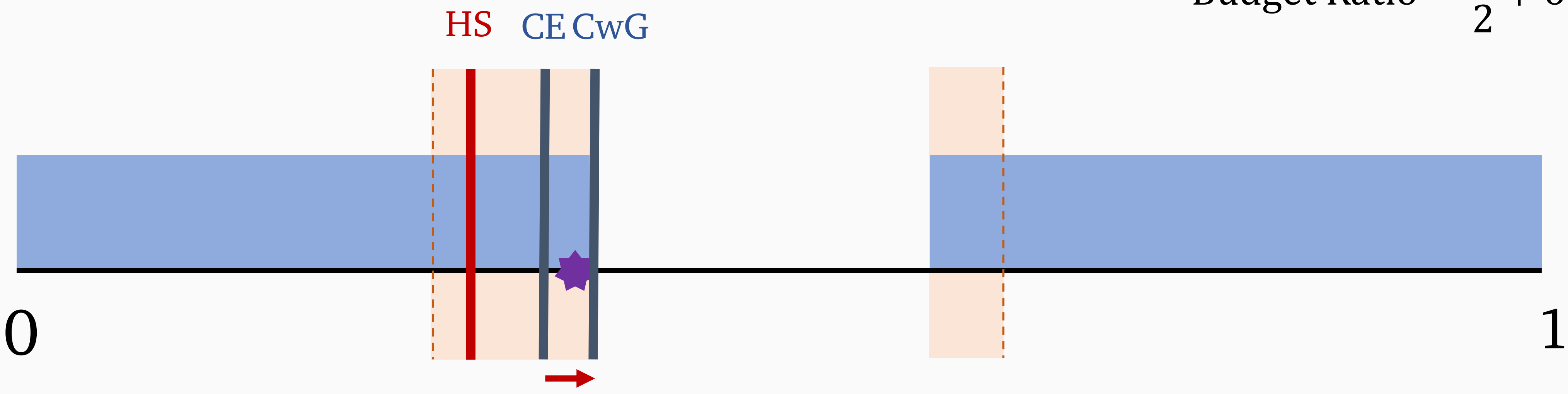
CwG: Hedging against the risk of large regret in the future



CwG and HS agree $\Rightarrow$ no compensation required

CwG: Hedging against the risk of large regret in the future

$$\text{Budget Ratio} = \frac{1}{2} + \delta$$



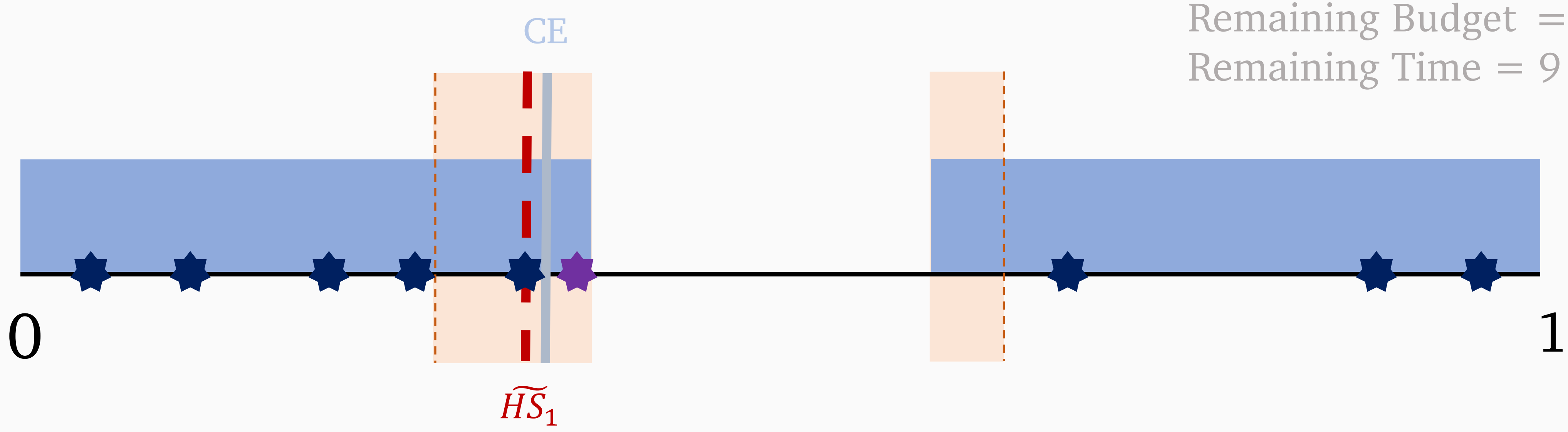
CwG and HS disagree, but expected cost of mistake $\sim \dfrac{1}{\text{remaining time}}$

CwG via multiple simulations

$\text{Budget Ratio} = \frac{1}{2} + \delta$

$\text{Remaining Budget} = 5$

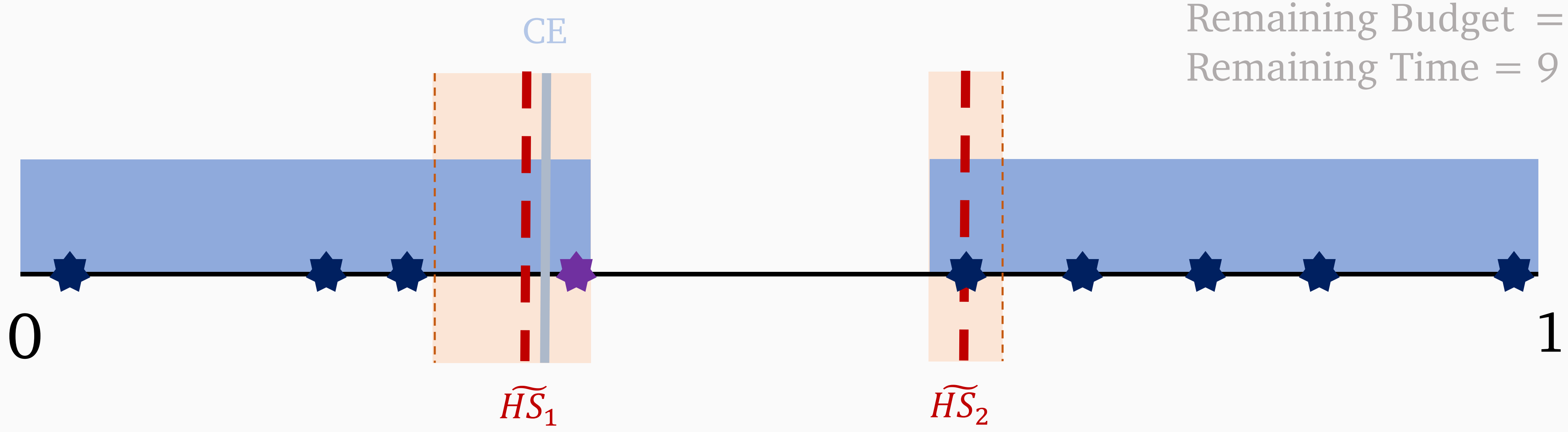
$\text{Remaining Time} = 9$



CwG via multiple simulations

$$\text{Budget Ratio} = \frac{1}{2} + \delta$$

Remaining Budget $= 5$
Remaining Time $= 9$

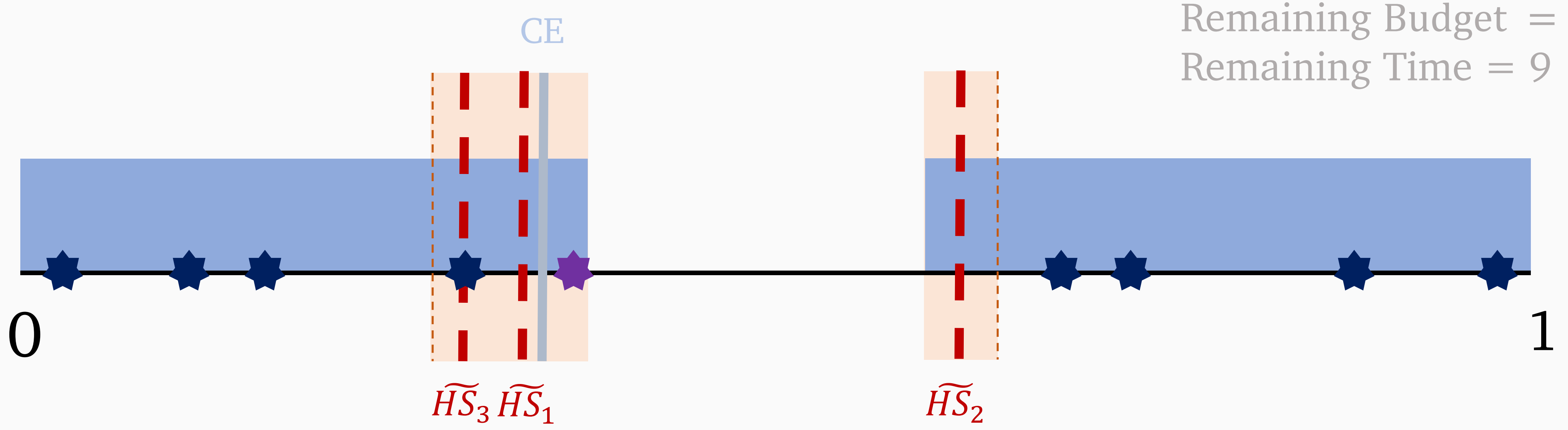


CwG via multiple simulations

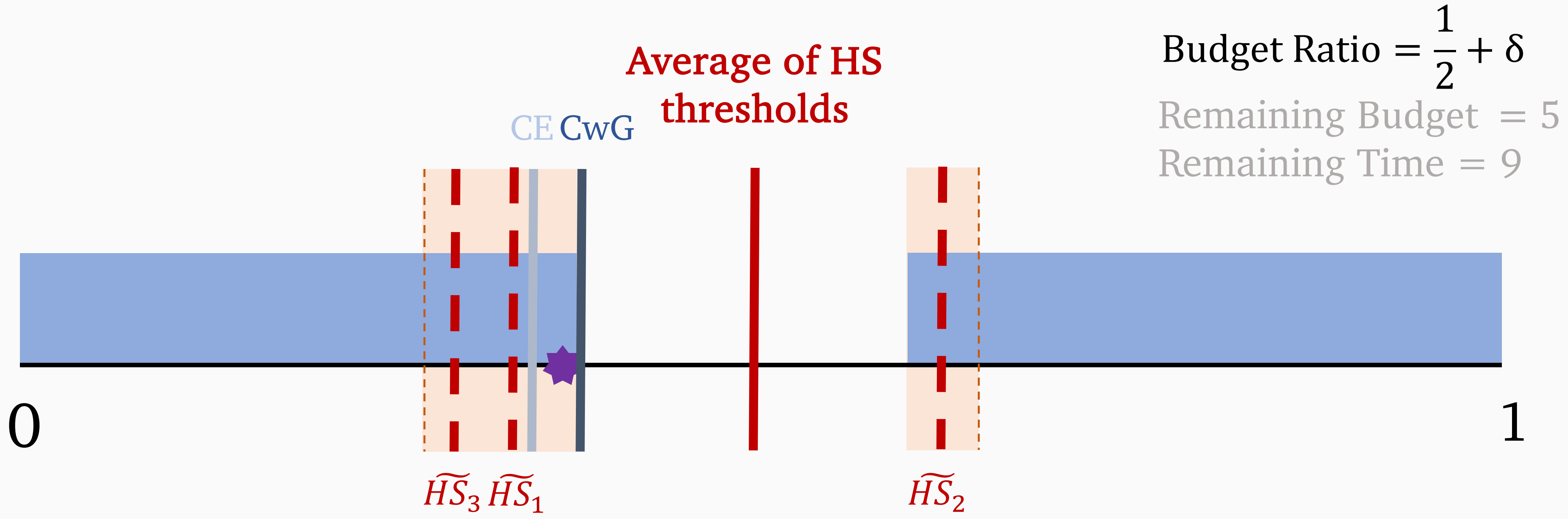
Budget Ratio $= \frac{1}{2} + \delta$

Remaining Budget $= 5$

Remaining Time $= 9$



CwG via multiple simulations



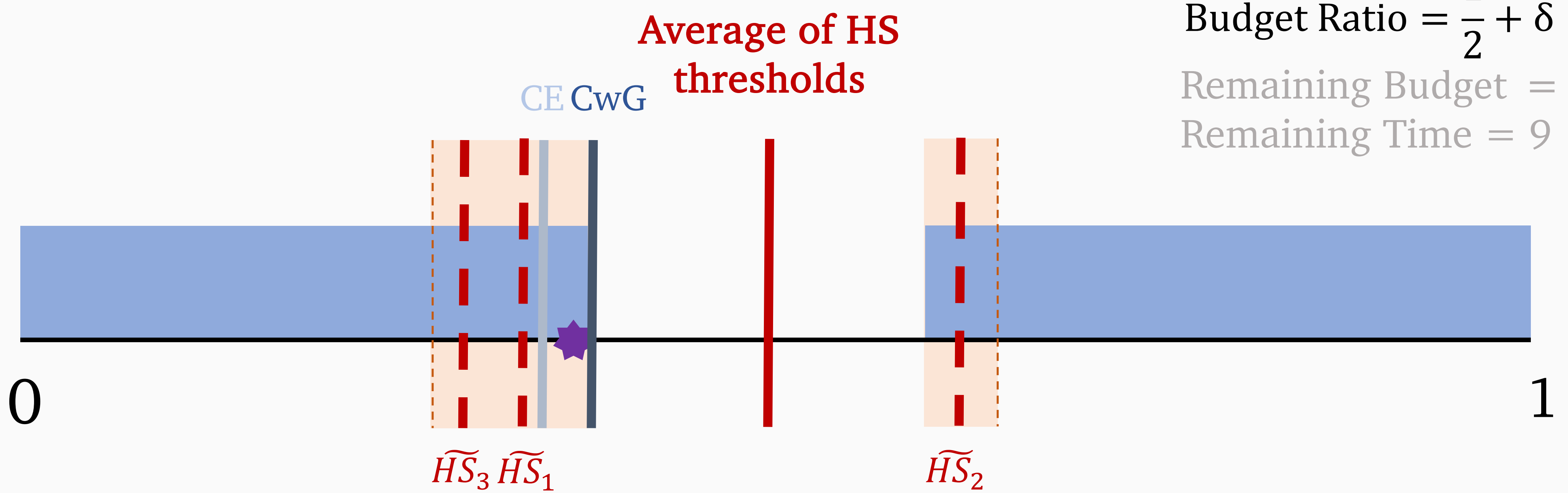
same action under the CwG policy and using the average of the simulated HS thresholds

CwG via multiple simulations

Budget Ratio $= \frac{1}{2} + \delta$

Remaining Budget $= 5$

Remaining Time $= 9$



Connection to "dual averaging" policy of Talluri and van Ryzin (1998)

think of the different HS thresholds as the shadow prices of the budget for different scenarios, the bid price is computed by averaging the HS thresholds

Repeatedly Act using Multiple Simulations (RAMS)

current budget B_t
feasible set of actions A_t

observe request $\theta_t = (r_t, c_t)$

Repeatedly Act using Multiple Simulations (RAMS)

current budget B_t

feasible set of actions A_t

observe request $\theta_t = (r_t, c_t)$

Scenario 1

Scenario m

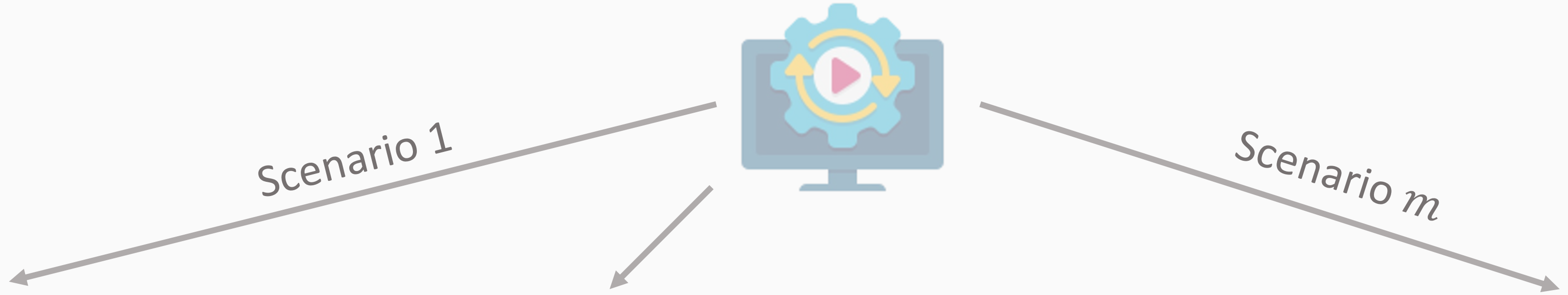
$\theta_t \ \tilde{\theta}^1_{t+1} \ \cdots \ \tilde{\theta}^1_T$ $\quad$ $\theta_t \ \tilde{\theta}^2_{t+1} \ \cdots \ \tilde{\theta}^2_T$ $\quad$ .... $\quad$ $\theta_t \ \tilde{\theta}^m_{t+1} \ \cdots \ \tilde{\theta}^m_T$

Repeatedly Act using Multiple Simulations (RAMS)

current budget B_t

feasible set of actions A_t

observe request $\theta_t = (r_t, c_t)$



$\theta_t \;\; \tilde{\theta}^1_{t+1} \;\; \cdots \;\; \tilde{\theta}^1_T$ $\qquad$ $\theta_t \;\; \tilde{\theta}^2_{t+1} \;\; \cdots \;\; \tilde{\theta}^2_T$ $\qquad \cdots \qquad$ $\theta_t \;\; \tilde{\theta}^m_{t+1} \;\; \cdots \;\; \tilde{\theta}^m_T$

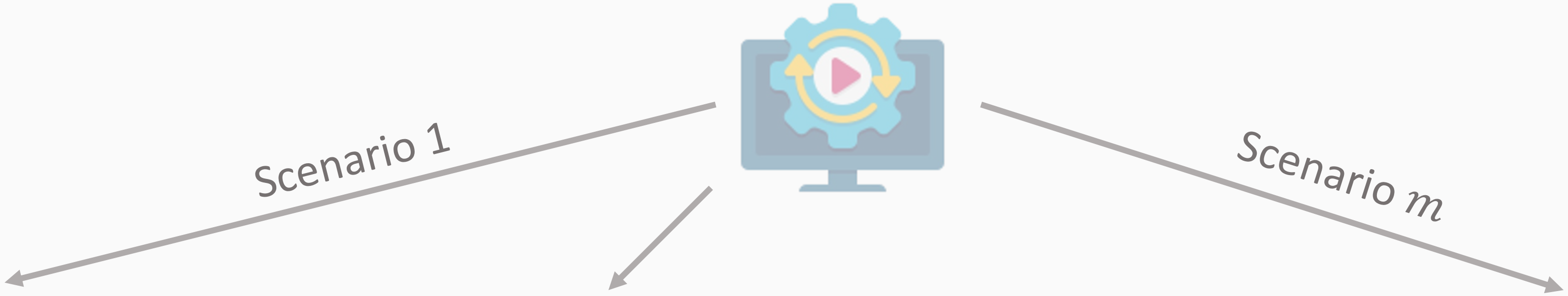
For each simulated scenario, compute the maximum total reward for each feasible current action at time *t*

Repeatedly Act using Multiple Simulations (RAMS)

current budget B_t

feasible set of actions A_t

observe request $\theta_t = (r_t, c_t)$



$\theta_t \;\; \tilde{\theta}^1_{t+1} \; \cdots \; \tilde{\theta}^1_T \qquad \theta_t \;\; \tilde{\theta}^2_{t+1} \; \cdots \; \tilde{\theta}^2_T \qquad \cdots \qquad \theta_t \;\; \tilde{\theta}^m_{t+1} \; \cdots \; \tilde{\theta}^m_T$

For each simulated scenario, compute the maximum total reward for each feasible current action at time t

Take the action which maximizes the average total reward over the multiple simulated scenarios

Repeatedly Act using Multiple Simulations (RAMS)

current budget B_{t+1}

feasible set of actions A_{t+1}

observe request $\theta_{t+1} = (r_{t+1}, c_{t+1})$

Scenario 1

Scenario m

$\theta_{t+1}\ \tilde{\theta}^1_{t+2}\ \cdots\ \tilde{\theta}^1_T$ $\quad$ $\theta_{t+1}\ \tilde{\theta}^2_{t+2}\ \cdots\ \tilde{\theta}^2_T$ $\quad \cdots \quad$ $\theta_{t+1}\ \tilde{\theta}^m_{t+2}\ \cdots\ \tilde{\theta}^m_T$

For each simulated scenario, compute the maximum total reward for each feasible current action at time t

Take the action which maximizes the average total reward over the multiple simulated scenarios

RAMS minimizes hindsight-based regret

Informal Meta Theorem.

Given an dynamic resource allocation setting, if there exists an algorithm **ALG** satisfying certain technical conditions, then

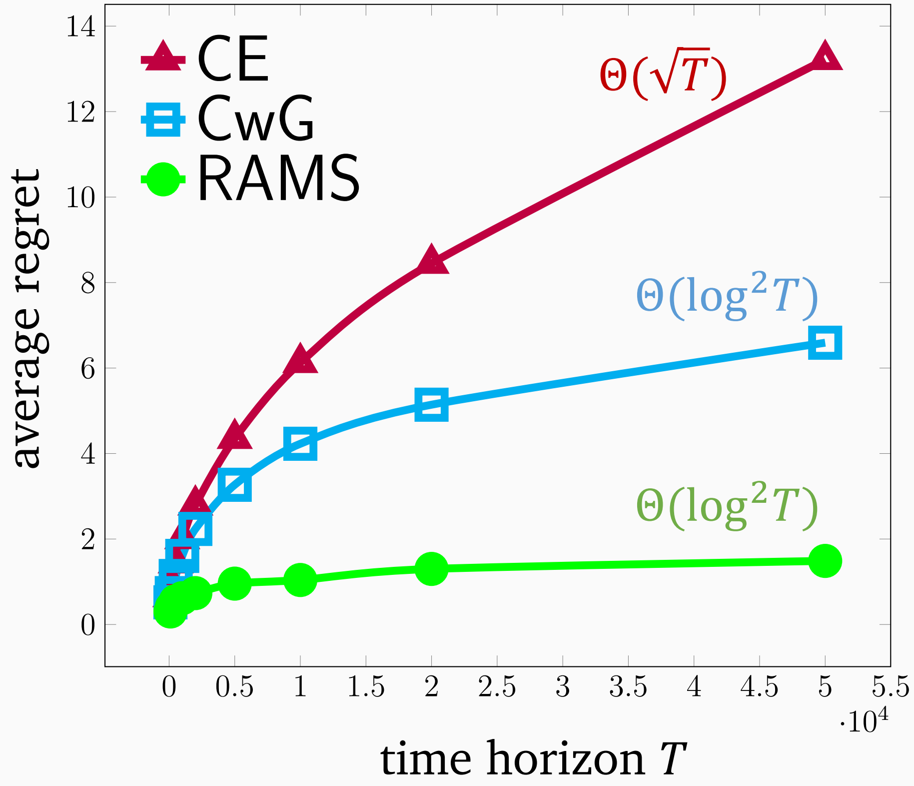
$$\text{Regret}(\mathsf{RAMS}) \leq \text{Regret Upper Bound of } \mathsf{ALG} + \text{Sampling Error}$$

Corollary of the Meta Theorem

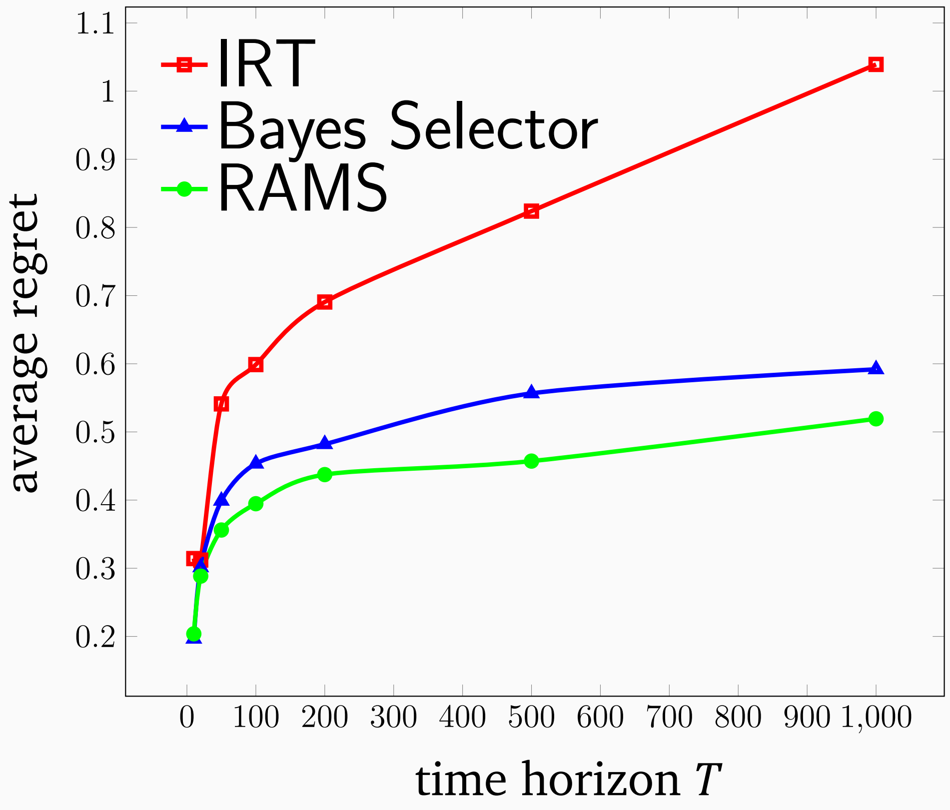
- bounded regret for NRM and online matching for a few types (Vera and Banerjee '21)
- logarithmic regret for NRM for many types with non-degeneracy assumption (Bray '22)
- log-squared regret for NRM for many types without non-degeneracy assumption (Jiang et. al '22)

Numerical Performance of RAMS

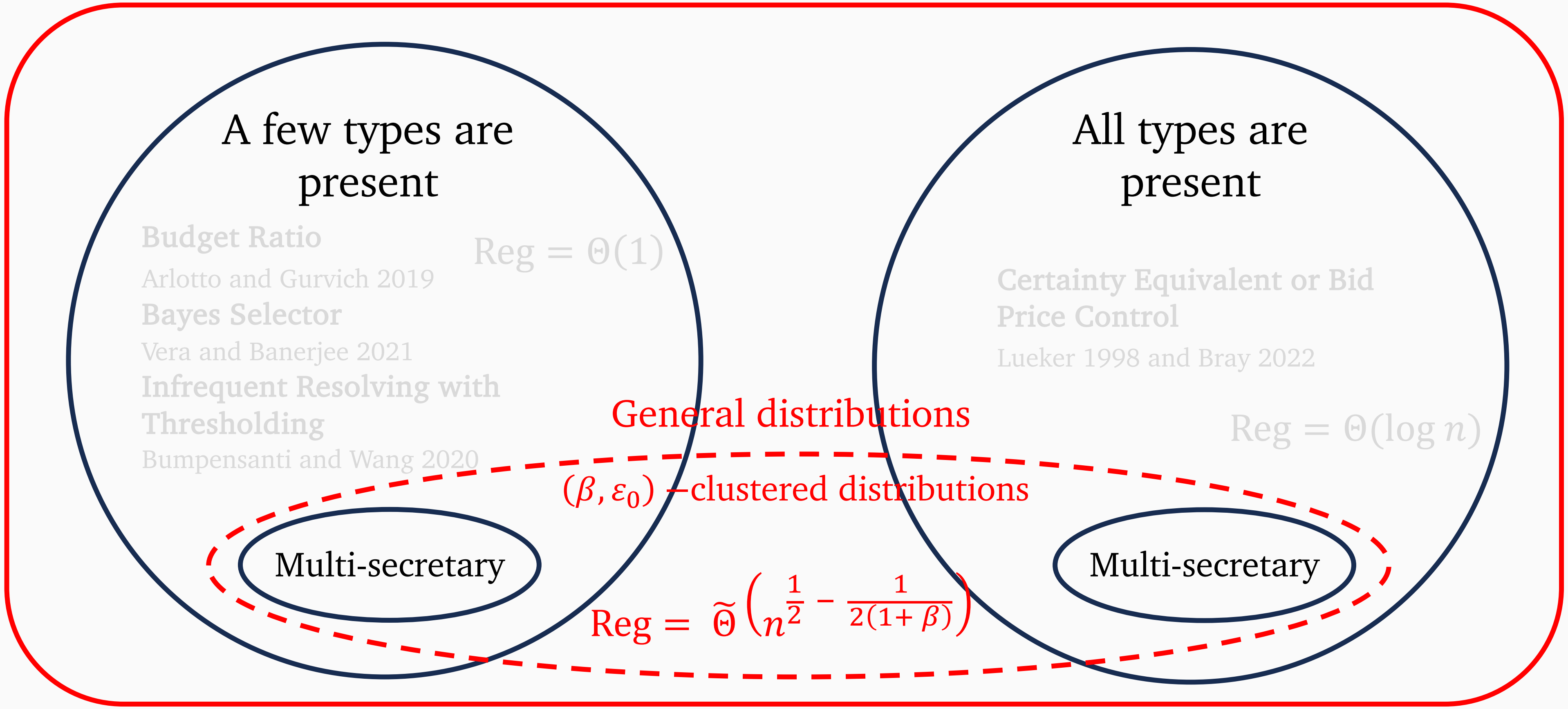
Multi-secretary Problem for the bimodal uniform distribution



NRM problem with two resources and a few types present



One policy to solve them all and with a simulator we bind them



Repeatedly Act using Multiple Simulations

Regret is with respect to the hindsight optimal problem

Thanks!

arxiv.org/abs/2205.09078