Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling

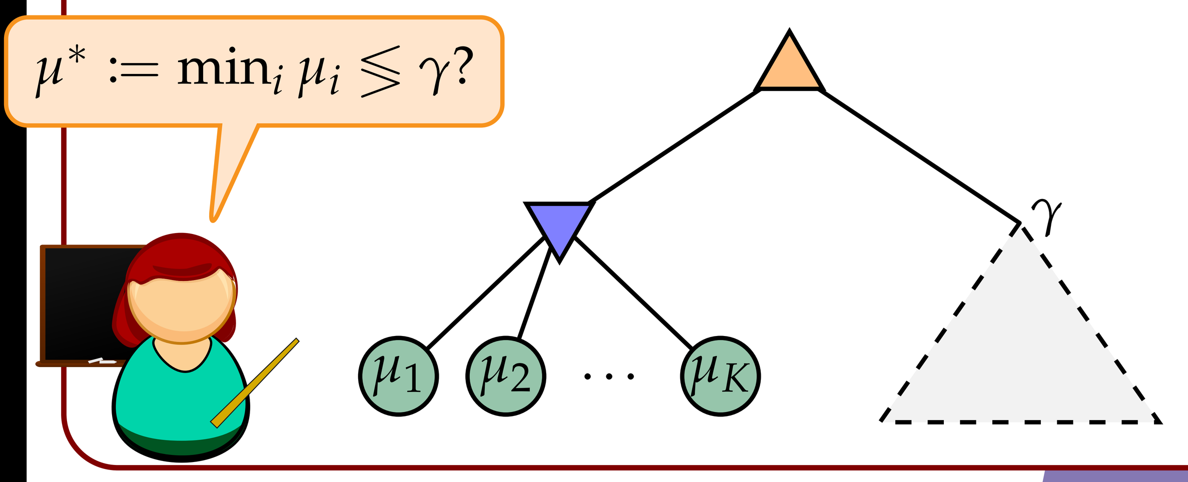
Emilie Kaufmann (CNRS) Wouter M. Koolen (CWI Centrum Wiskunde & Informatica) Aurélien Garivier (ENS de Lyon)

Motivating Questions

- Design of pure exploration **algorithms** for **complex queries**?
  - Monte Carlo Tree Search
- Valid anytime **confidence intervals** for **derived quantities**?
  - minimum

Stylised to Minimum Threshold Identification

Fix Bernoulli $\mu_1, \ldots, \mu_K$ and threshold γ. Learn δ-correct



For $t = 1, \ldots, \tau$
- Pick arm A_t
- See $X_t \sim \mu_{A_t}$

Say $\hat{m} \in \{<, >\}$

$\mathbb{P}_{\boldsymbol{\mu}}(\text{error}) < \delta$

Lower Bound

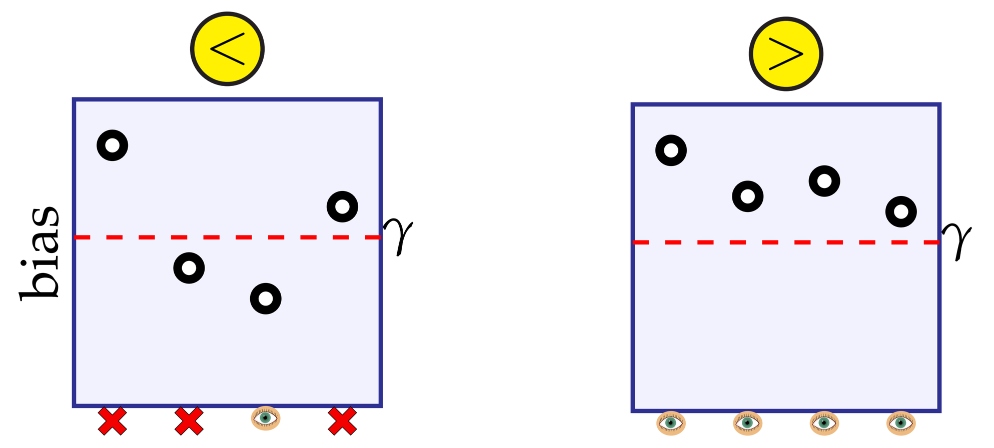
Generic lower bound [Castro, 2014, Garivier and Kaufmann, 2016] shows **sample complexity** for **any** δ-correct algorithm is at least

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \;\geq\; T^*(\boldsymbol{\mu}) \ln \frac{1}{\delta}.$$

For our problem the **characteristic time** and **oracle weights** are

$$T^*(\boldsymbol{\mu}) \;=\; \begin{cases} \dfrac{1}{d(\mu^*, \gamma)} & \mu^* < \gamma, \\ \displaystyle\sum_a \frac{1}{d(\mu_a, \gamma)} & \mu^* > \gamma, \end{cases} \qquad w_a^*(\boldsymbol{\mu}) \;=\; \begin{cases} \mathbf{1}_{a=a^*} & \mu^* < \gamma, \\ \dfrac{\frac{1}{d(\mu_a,\gamma)}}{\sum_j \frac{1}{d(\mu_j,\gamma)}} & \mu^* > \gamma. \end{cases}$$

Dichotomous Oracle Behaviour! Sampling Rule?



Real Algorithms Must Sample Every Arm

For **symmetric** algorithms we **boost** the lower bound on $\mu^* < \gamma$ to

$$\mathbb{E}_{\boldsymbol{\mu}}[\tau] \;\geq\; \frac{\ln \frac{1}{\delta}}{d(\mu^*, \gamma)} + \frac{C/K^2}{\max_a d(\mu_a, \gamma)} \sum_{a=1}^{K} \left(1 - \frac{d_+(\mu_a, \gamma)}{d(\mu^*, \gamma)}\right).$$

Sampling Rules

- **Lower Confidence Bounds**
  Play $A_t = \arg\min_a \ \mathrm{LCB}_a(t)$
- **Thompson Sampling** (Π_{t-1} is posterior after $t-1$ rounds)
  Sample $\boldsymbol{\theta} \sim \Pi_{t-1}$, then play $A_t = \arg\min_a \ \theta_a$.
- **Murphy Sampling** **condition on *low* minimum mean** (NEW)
  Sample $\boldsymbol{\theta} \sim \Pi_{t-1}\left(\cdot \middle| \min_a \theta_a < \gamma\right)$, then play $A_t = \arg\min_a \ \theta_a$.

Intuition for Murphy Sampling

- When $\mu^* < \gamma$ conditioning is immaterial: $\boldsymbol{\theta} \approx \boldsymbol{\mu}$ and MS $\equiv$ TS.
- When $\mu^* > \gamma$ conditioning results in $\boldsymbol{\theta} \approx (\mu_1, \ldots, \gamma, \ldots, \mu_K)$. Index a lowered to γ with probability $\propto \frac{1}{d(\mu_a, \gamma)}$ [Russo, 2016].

Main Result 1 : Murphy Sampling Rule

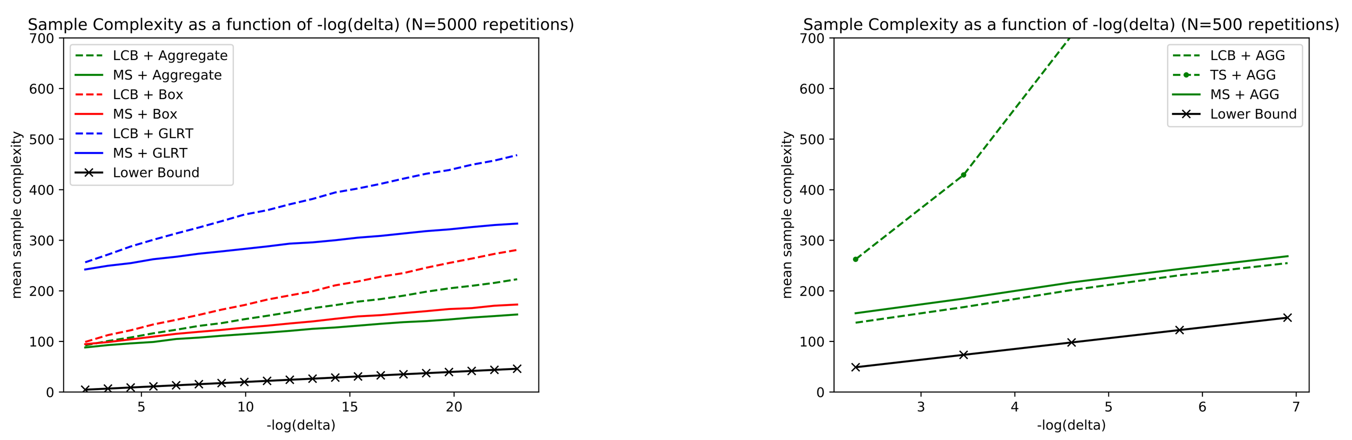
Theorem 1. *Asymptotic optimality: $N_a(t)/t \to w_a^*(\boldsymbol{\mu})$ for all $\boldsymbol{\mu}$*

Sampling rule	<	>
Thompson Sampling	✓	✗
Lower Confidence Bounds	✗	✓
Murphy Sampling	✓	✓

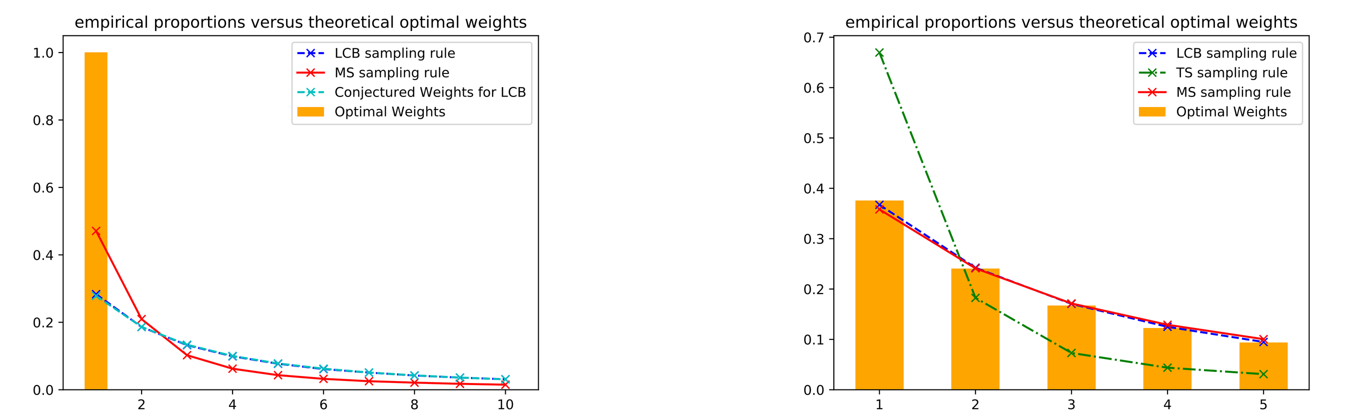
Lemma 2. *Any anytime sampling strategy (A_t) ensuring $\frac{\boldsymbol{N}_t}{t} \to \boldsymbol{w}^*(\boldsymbol{\mu})$ and **good** stopping rule τ_δ guarantee* $\limsup_{\delta \to 0} \frac{\tau_\delta}{\ln \frac{1}{\delta}} \leq T^*(\boldsymbol{\mu})$.

Numerical Results

$\boldsymbol{\mu} = \text{linspace}(-1, 1, 10) \in \mathcal{H}_<$ $\qquad$ $\boldsymbol{\mu} = \text{linspace}(1/2, 1, 5) \in \mathcal{H}_>$



$\mathbb{E}[\tau_\delta]$ as a function of $\ln(1/\delta)$. Throughout $\gamma = 0$.



Sampling proportions vs oracle, $\delta = e^{-23}$ (l) and $\delta = e^{-7}$ (r).

(Non-Asymptotic) Adaptivity

Multiple low arms (identical or similar) $\Rightarrow$ $\begin{cases} \text{conclude } \mu^* < \gamma \textbf{ faster} \\ \textbf{tighter} \text{ confidence interval for } \mu^* \end{cases}$?

Confidence Interval for Minimum

For **LCB** we adopt the obvious $\mathrm{LCB}_{\min}(t) = \min_a \mathrm{LCB}_a(t)$.

For **UCB** we investigate three approaches:
- **Box**: Straightforward idea: $\mathrm{UCB}_{\min}(t) = \min_a \mathrm{UCB}_a(t)$.
- **GLRT**: New sum-of-deviations confidence bound.
- **Agg**: Pool samples from multiple arms. Upper bound on **any average** is upper bound on **minimum**. **Biased** but **narrower**.

Main Result 2: Deviation Inequalities

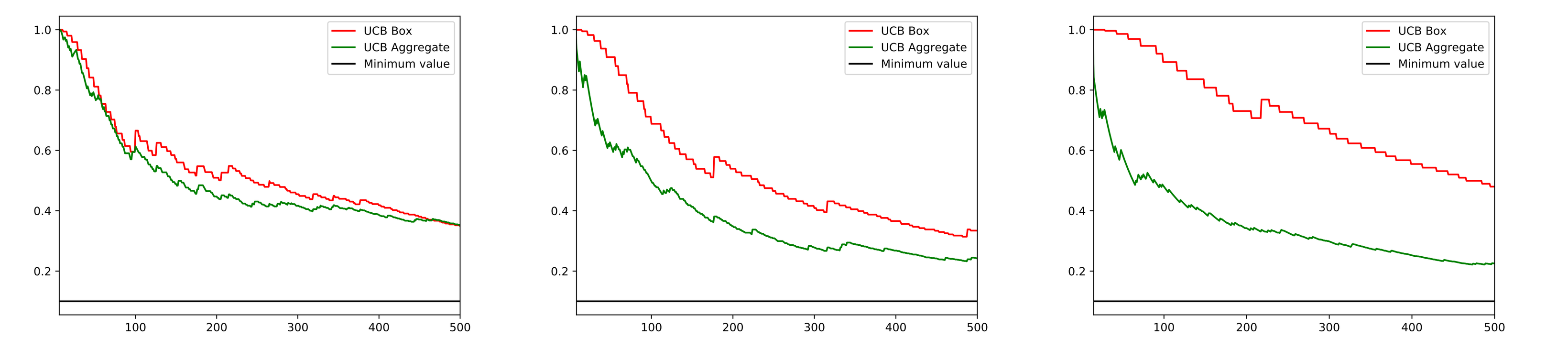
We identify **threshold function** $T(x) = x + o(x)$ such that for every fixed subset $\mathcal{S} \subseteq [K]$, w.h.p. $\geq 1 - \delta$,

$$\forall t : \left[N_{\mathcal{S}}(t) d^+\left(\hat{\mu}_{\mathcal{S}}(t), \min_{a \in \mathcal{S}} \mu_a\right) - \ln \ln N_{\mathcal{S}}(t)\right]^+ \leq T\left(\ln \tfrac{1}{\delta}\right),$$

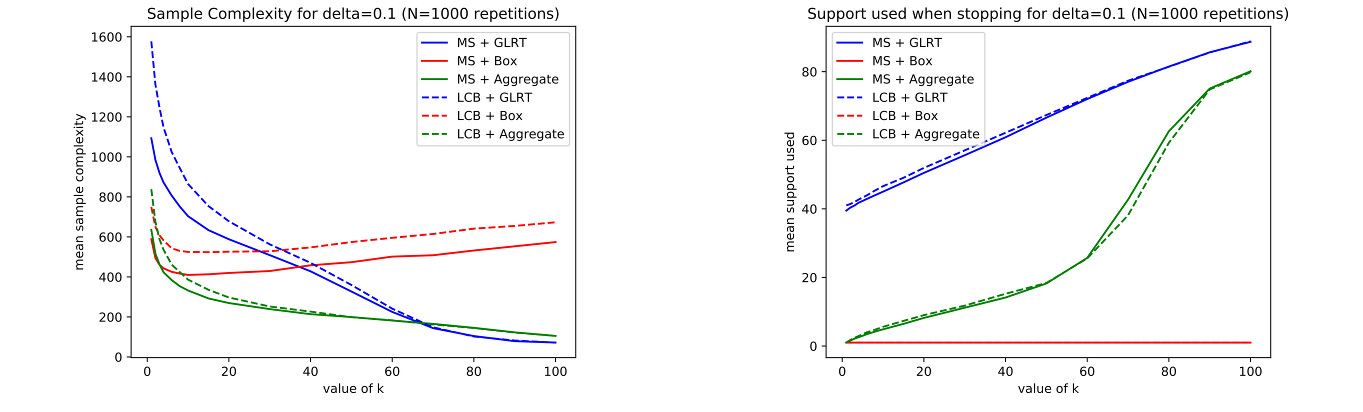
$$\forall t : \sum_{a \in \mathcal{S}} \left[N_a(t) d^+\left(\hat{\mu}_a(t), \min_{a \in \mathcal{S}} \mu_a\right) - \ln \ln N_a(t)\right]^+ \leq |\mathcal{S}| T\left(\frac{\ln \frac{1}{\delta}}{|\mathcal{S}|}\right).$$

Weighted union bound over subsets **learns** useful low-mean arms.

Numerical Results



UCB for minimum: Agg **dominates** Box with 1, 3 and 10 low arms.



Agg **beats** Box and GLRT in adapting to the number k of low arms. Here $\mu_a \in \{-1, 0\}$ and $\gamma = 0$.

What's Next

Deep trees. Adaptive tree expansion. Foundation for MCTS and RL.