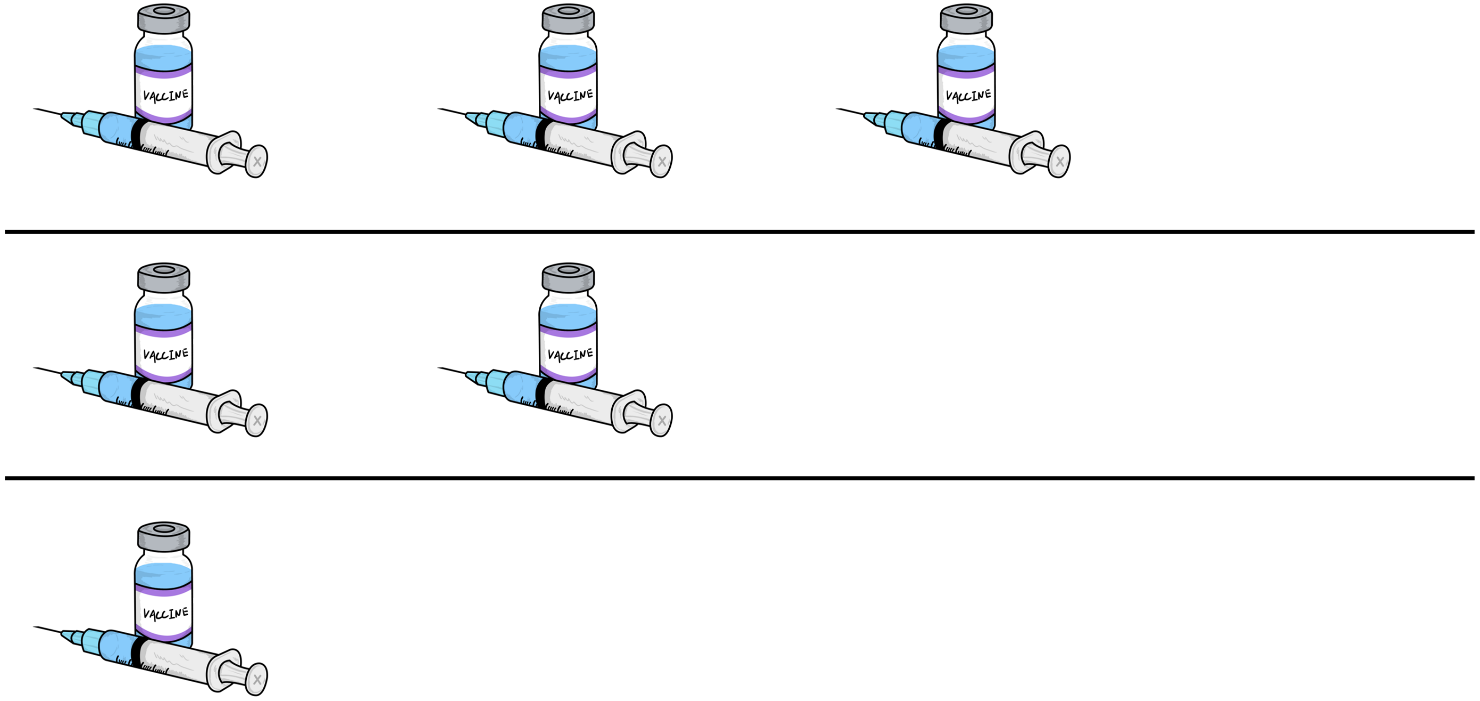
Multi-Fidelity Multi-Armed Bandits Revisited

Xuchuang Wang (CUHK), Qingyun Wu (PSU),

Wei Chen (MSRA), John C.S. Lui (CUHK)

NeurIPS' 23

Effectiveness of Multi-Dose Vaccines



Ground Truth Effectiveness

Approximate Effectiveness

Low Approximate Effectiveness

Multi-Fidelity Multi-Armed Bandits

- $K \in \mathbb{N}^+$ arms and $M \in \mathbb{N}^+$ fidelities
- In each time t, pull an arm k at fidelity m, and
  - Pay cost $\lambda^{(m)}$
  - Observes a stochastic reward with *unknown* mean $\mu_k^{(m)}$
- Accuracy: $\left|\mu_k^{(m)} - \mu_k^{(M)}\right| \leq \zeta^{(m)}$ for all fidelity m
- Objective: find the arm with highest $\mu_k^{(M)}$ (ground true mean reward)

Optimal Fidelity: most effective choice

- Assume $\mu_1^{(M)} > \mu_2^{(M)} > \cdots > \mu_K^{(M)}$
- Denote $\Delta_k^{(m)} := \begin{cases} \mu_1^{(M)} - \left(\mu_k^{(m)} + \zeta^{(m)}\right) & k \neq 1 \\ \left(\mu_1^{(m)} - \zeta^{(m)}\right) - \mu_2^{(m)} & k = 1 \end{cases}$
- Each arm k has an optimal fidelity (Theorem 3.1)

$$m_k^* := \underset{m}{\operatorname{argmax}} \frac{\Delta_k^{(m)}}{\sqrt{\lambda^{(m)}}}$$

Best Arm Identification: Algorithm

- Lower-Upper Confidence Bound (LUCB) (Algorithm 1)
  - Stop when best arm’s **LCB** exceeds the **UCB** of second-best
- Fidelity selection procedures (Algorithm 2)
  - Explore-A: find the optimal fidelity
    - Design f-UCB indices for each fidelity
    - Choose the fidelity with highest f-UCB
  - Explore-B: find a good fidelity
    - Uniformly choose fidelity at beginning
    - Commit to one good fidelity with condition

Best Arm Identification: Analysis

- Cost complexity:
  - the total cost for identifying the best arm with confidence $1 - \delta$
- Explore-A: find the optimal fidelity

$$\mathbb{E}[\Lambda] = O\left(H\log\frac{1}{\delta} + G\log\left(\log\frac{1}{\delta}\right)\right)$$

- Explore-B: find a good fidelity

$$\mathbb{E}[\Lambda] = O\left(H\sum_{m}\frac{\lambda^{(m)}}{\lambda^{(1)}}\log\frac{1}{\delta}\right)$$

- Where $H = \sum_m \frac{\lambda^{(m_k^*)}}{\left(\Delta_k^{(m_k^*)}\right)^2}$ and $G = \sum_k \sum_m \left(\frac{\Delta_k^{(m_k^*)}}{\sqrt{\lambda^{(m_k^*)}}} - \frac{\Delta_k^{(m)}}{\sqrt{\lambda^{(m)}}}\right)^{-2}$

Thank you for Listening!

More new results about regret minimization

in full paper at [*https://openreview.net/pdf?id=oi45JlpSOT*](https://openreview.net/pdf?id=oi45JlpSOT)