Interactive Edge Orientation Identification: An Optimization Approach

Fabian Damken[1] Wouter M. Koolen[2,1] Rianne de Heide[1,2]

[1]University of Twente, Enschede [2]Centrum Wiskunde & Informatica, Amsterdam

Multi-Armed Bandit Notation

- $\nu = \big(\nu_i \mid i \in [K]\big)$ is a K-armed bandit
- reward distribution ν_i for arm $i \in [K]$
- reward mean $\mu_i = \mathbb{E}_{X \sim \nu_i}[X]$
- in each round t, agent selects an arm $I_t \in [K]$
- subsequently receives reward $X_t \sim \nu_{I_t}$

Pure Exploration

- find out certain details about the bandit
- e.g., identify best arm $\operatorname{argmax}_{i \in [K]} \mu_i$
- algorithm components: sampling rule $(\pi_t)_t$, stopping rule τ, decision rule $\hat{m}_\tau$
- fixed-budget: fix τ, max. $\mathbb{P}\big(\hat{m}_\tau = m(\nu)\big)$
- fixed-confidence: fix $\mathbb{P}\big(\hat{m}_\tau = m(\nu)\big)$, min. τ

Anytime-Valid CIs

- confidence intervals (CIs):
$$\forall t \in \mathbb{N} \colon \mathbb{P}\big(\mu \in \hat{\mu}(t) \pm \beta(t)\big) \geq 1 - \alpha$$
- only for *fixed* sample size t
- does *not* hold if t depends on CI itself
- anytime-valid CIs hold for all *stopping* times!
$$\forall \tau \colon \mathbb{P}\big(\mu \in \hat{\mu}(\tau) \pm \beta(\tau)\big) \geq 1 - \alpha$$
- naive: $\alpha = 6\tilde{\alpha}/(\pi^2 t^2)$, better versions exist [1]

References

[1] S. R. Howard et al. "Time-Uniform, Nonparametric, Nonasymptotic Confidence Sequences." In: *The Annals of Statistics* 49.2 (2021), pp. 1055–1080. ISSN: 0090-5364. JSTOR: 27169410.

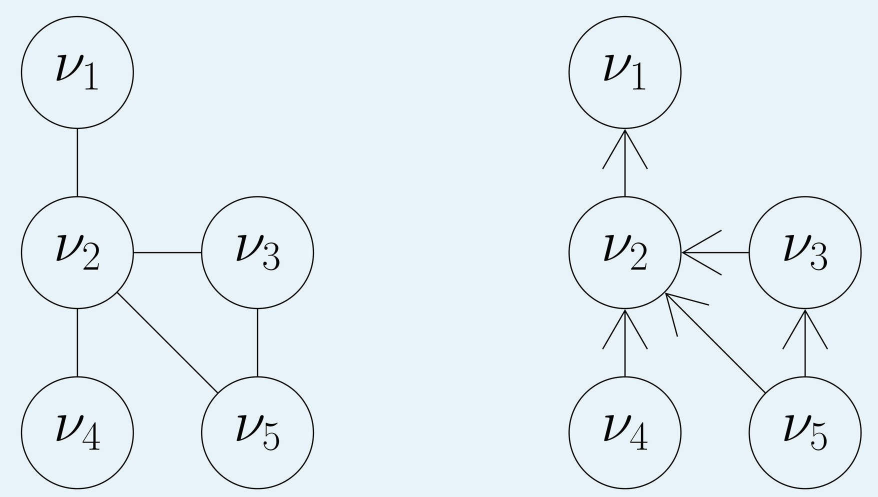
[2] A. Garivier and E. Kaufmann. "Optimal Best Arm Identification with Fixed Confidence." In: *Conference on Learning Theory*. Conference on Learning Theory. PMLR, June 6, 2016, pp. 998–1027.

[3] H. Zhuang, C. Wang, and Y. Wang. "Identifying Outlier Arms in Multi-Armed Bandit." In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.

Edge Orientation Problem

consider following 5-armed bandit and task:

- consider a 5-armed bandit with $\mu_1 < \mu_2 < \cdots < \mu_5$
- task definition: graph over arms



task: identify edge orientation

Formal Definition

- given a set of pairwise queries
$$\mathcal{C} \subseteq \{(i,j) \mid 1 \leq i < j \leq K\}$$
- we want to decide whether $\mu_i < \mu_j$
- correct answer:
$$m^*_{(i,j)}(\nu) = (-1)^{\mathbb{1}[\mu_i < \mu_j]}$$
i.e., $m_{(i,j)}(\nu) < 0$ iff $\mu_i < \mu_j$
- track-and-stop [2] lower bound

Separation Lemma [cf., 3]

- consider a confidence intervals $k \in \{1,2\}$:
$$\mathbb{P}(\mu_k \in \hat{\mu}_k \pm \beta_k) \geq 1 - \alpha$$
- conditional on the CI holding:
$$2(\beta_1 + \beta_2) \leq \Delta \quad \Longrightarrow \quad \beta_1 + \beta_2 \leq \hat{\Delta}$$
- i.e., go from true to estimated mean gap

Core Idea

- place anytime-valid CIs on all arms:
$$\hat{\mu}_i(T) \pm \beta_i(T) \qquad \beta_i(T) = \frac{\gamma(T)}{\sqrt{N_i(T)}}$$
- decide for each $(i,j) \in \mathcal{C}$ whether $\mu_i < \mu_j$ based on these CIs
- directly optimize for stopping time:
$$\min_{N \in \mathbb{N}^K} \sum N$$
$$\text{s.t.} \quad \gamma\Big(\sum N\Big)\left(\frac{1}{\sqrt{N_i}} + \frac{1}{\sqrt{N_j}}\right) \leq \Delta_{ij}$$
where $\Delta_{ij} = |\mu_i - \mu_j|$

Reformulation

1. relax optimization problem into continuity
2. eliminate $\gamma\big(\sum N\big)$ from constraints and objective without changing the argmin
3. replace individual number of pulls $(N_i)_{i \in [K]}$ with total number of pulls T and sampling distribution $\omega \in \Delta_K$
4. massage constraints into the form LHS $\leq T$
5. replace objective with maximum over LHS's
6. take reciprocal of objective
$$\max_{\omega \in \Delta^K} \min_{(i,j) \in \mathcal{C}} \Delta_{ij}^2 \left(\frac{1}{\sqrt{\omega_i}} + \frac{1}{\sqrt{\omega_j}}\right)^{-2}$$

Algorithm Ingredients

1. *estimate* $\Delta_{ij}^2 \approx \hat{\Delta}_{ij}(t)^2 = \big|\hat{\mu}_i(t) - \hat{\mu}_j(t)\big|^2$
2. *optimize* above objective w.r.t. ω, yielding sequence $(\omega(t))_t$, once GD step per iteration
3. *track* cumulative distribution $\sum_{s=1}^t \omega(s)$
4. *stop* if confidence intervals are separated

Upper Bound

- from the above separation lemma
- and optimization problem reformulation
- algorithm with pull distribution ω stops with high probability after T pulls if
$$\frac{1}{T} \leq \frac{1}{4\gamma(T)^2} \underbrace{\min_{(i,j) \in \mathcal{C}} \Delta_{ij}^2 \left(\frac{1}{\sqrt{\omega_i}} + \frac{1}{\sqrt{\omega_j}}\right)^{-2}}_{f^*(\omega) = \quad / \quad f^{(t)} = (\Delta_{ij} \mapsto \hat{\Delta}_{ij}(t))}$$
- i.e., $4\gamma(T)^2 \leq T f^*\big(N(T)/T\big)$
- lower-bound RHS with reality:
$$\begin{aligned} &T f^*\big(N(T)/T\big) \\ &\geq T f^*\big(\textstyle\sum_{t=1}^T \omega(t)\big) - \mathcal{O}(1) \\ &\geq \textstyle\sum_{t=1}^T f^*\big(\omega(t)\big) - \mathcal{O}(1) \\ &\geq \textstyle\sum_{t=1}^T f^{(t)}\big(\omega(t)\big) - \mathcal{O}(\sqrt{T}) \\ &\geq \textstyle\sum_{t=1}^T f^{(t)}(\omega^*) - \mathcal{O}(\sqrt{T}) \\ &\geq T f^*(\omega^*) - \mathcal{O}(\sqrt{T}) \end{aligned}$$
where $\omega^* = \operatorname{argmax}_{\omega \in \Delta_K} f^*(\omega)$
- hence, our algorithm stops with high probability after T rounds if
$$4\gamma(T)^2 + \mathcal{O}(\sqrt{T}) \leq T f^*(\omega^*)$$

Proposed Algorithm

```
initialize μ̂_i(0) ~ ν_i, ω_i(0) = 1/K, N_i(0) = 1
foreach t = 1, 2, ..., τ do
    Δ̂_ij(t − 1) = |μ̂_i(t − 1) − μ̂_j(t − 1)|
    G^(t) = (∇_ω f^(t−1)(ω))|_{ω=ω(t−1)}
    ω̃(t) = ω(t − 1) + η_t G^(t)
    ω(t) = Π_{Δ_K^ε} ω̃(t)
    P(t) = Σ_{s=1}^t ω(s)
    pull i_t ∈ argmax_{i∈[K]} [P_i(t) − N_i(t − 1)]
    if γ(t)/√N_i(t) + γ(t)/√N_j(t) ≤ Δ̂_ij(t) then
        set τ = t
    end
end
return (m̂_τ)_(i,j) = (−1)^{1[μ̂_i(τ) < μ̂_j(τ)]}
```