Fixed-Budget Hypothesis Best Arm Identification: On the Information Loss in Experimental Design

Masahiro Kato, The University of Tokyo

Masaaki Imaizumi, The University of Tokyo

Takuya Ishihara, Tohoku University

Toru Kitagawa, Brown University

1. ABSTRACT

- **Best arm identification (BAI) with a fixed budget.**
- Recommend the best arm at the end of an experiment.
- **Contribution: Optimal experiments for BAI with a fixed budget.**

  Existence of an asymptotically optimal algorithm was unknown.

→ **Optimal experiments depend on available information.**

2. PROBLEM SETTING

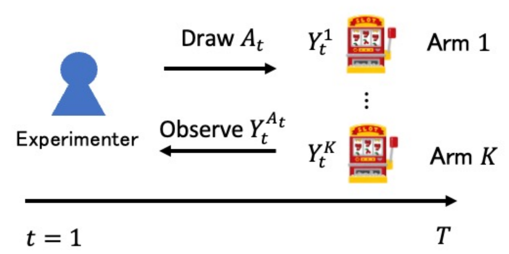
- **Adaptive experiment with T rounds**: $[T] = \{1,2,\dots,T\}$.
- **K treatment arms**: $[K] = \{1,2,\dots,K\}$.

  Treatment arms: alternatives of medicine, policy, and advertisements.
- Each arm a has a potential outcome $Y_t^a \in \mathbb{R}$.

  The distributions of Y_t^a do not change.

  Denote the mean outcome of an arm a by $\mu^a = \mathbb{E}[Y_t^a]$.
- Best treatment arm: an arm with the highest reward.

  Denote the best treatment arm by $a^* = \arg\max_{a\in[K]} \mu^a$
- $P = (P^1,\dots,P^K)$: distribution of $(Y_t^1,\dots,Y_t^K)$. $\mathcal{P}$: a set of P.
- $P_0 = (P_0^1,\dots,P_0^K)$: true distribution of the observations.
- **Sequence of an experiment**: at round $t \in [T]$,
- Draw a treatment arm $A_t \in [K]$.
- Observe a reward $Y_t^{A_t}$



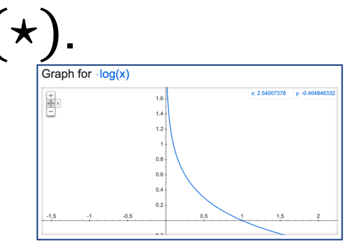
- After the final round T, an algorithm recommends an estimated best treatment arm $\hat{a}_T \in [K]$.
- **Goal**: Minimize the probability of misidentification: $\mathbb{P}[\hat{a}_T^* \neq a^*]$.

3. EVALUATION

- $\mathbb{P}[\hat{a}_T^* \neq a^*]$ converges to 0 with an exponential speed.

  → $\mathbb{P}[\hat{a}_T^* \neq a^*] = \exp(-T(\star))$ for a constant $(\star)$.
- Consider evaluating the term $(\star)$ by

  $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^* \neq a^*]$.



- A performance lower (upper) bound of $\mathbb{P}[\hat{a}_T^* \neq a^*]$ is an upper (lower) bound of $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^* \neq a^*]$.

4. INFORMATION LOSS IN EXPERIMENTS

- Different experiments are asymptotically optimal under different available information about outcomes' distribution .
- Consider lower bounds for $\mathbb{P}[\hat{a}_T \neq a^*]$ (Upper bounds for $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^* \neq a^*]$).
- **Case 1: complete distributional information is available.**
- Any experiment π satisfies

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^\pi \neq a^*] \leq \min_{a\in[K]} \inf_{\substack{Q\in\mathcal{P} \\ \mu^*(Q)-\mu^a(Q)<0}} \max_{w\in\mathcal{W}} \sum_{a\in[K]} w(a)\mathrm{KL}(Q^a, P_0^a).$$

- $\mathcal{W}$ is a set of functions $w: [K] \to [0,1]$ such that $\sum_{a\in[K]} w(a) = 1$.
- $w^* = \arg\max_{w\in\mathcal{W}} \sum_{a\in[K]} w(a)\mathrm{KL}(Q^a, P_0^a)$ is referred to as the target allocation ratio, which represents optimal ratio of arm draws.
- **Case 2: When only partial information is available, we approximate the KL divergence by the variances of outcomes.**
- $(\sigma^a)^2$: variance of an outcome Y_t^a.
- Assume that the variances are the same among $P \in \mathcal{P}$.
- As $\overline{\Delta} = \max_a \mu^{a^*} - \mu^a \to 0$, $\mathrm{KL}(Q^a, P_0^a)$ is approximated by the variance.
- Any experiment π satisfies

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T \neq a^*] \leq \min_{a\in[K]\backslash\{a^*\}} \max_{w\in\mathcal{W}} \frac{\overline{\Delta}^2}{2\left(\frac{\sigma^{a^*}}{w(a^*)} + \frac{\sigma^a}{w(a)}\right)} + o(\overline{\Delta}^2).$$

- $w^* = \arg\max_{w\in\mathcal{W}} \sum_{a\in[K]} \frac{1}{\frac{\sigma^{a^*}}{w(a^*)} + \frac{\sigma^a}{w(a)}}$ has a closed form.
- $w^*(a^*) = \frac{\sigma^{a^*}}{\sigma^{a^*} + \sqrt{\sum_{b\in[K]\backslash\{a^*\}} (\sigma^b)^2}}$, $w^*(a) = (1 - w^*(a^*)) \frac{(\sigma^b)^2}{\sum_{b\in[K]\backslash\{a^*\}} (\sigma^b)^2}$.
- w^* depends on a^*.

→ Use some proxy (conjecture) for a^* in experiments

= We refer to this proxy as the hypothetical best treatment arm.

- **Case 3: If hypothetical best arm is also unknown, we consider the following worst-case lower bound.**
- $\overline{\Delta}^2(P) = \max_a \mu^{a^*}(P) - \mu^a(P)$, where $\mu^a(P)$ is the mean under P.
- Any experiment π satisfies

$$\sup_{P\in\mathcal{P}} \lim_{\overline{\Delta}(P)\to 0} \limsup_{T\to\infty} -\frac{1}{T\overline{\Delta}^2(P)}\log \mathbb{P}[\hat{a}_T^* \neq a^*(P)] \leq \max_{w\in\mathcal{W}} \min_{\substack{b\in[K] \\ a\in[K]\backslash\{b\}}} \frac{1}{2\left(\frac{\sigma^b}{w(b)} + \frac{\sigma^a}{w(a)}\right)} + o(1).$$

5. ASYMPTOTICALLY OPTIMAL EXPERIMENTS

- Based on available information, we design optimal experiments.
- A sample average estimator for μ^a: $\hat{\mu}_T^a = \frac{1}{\sum_{t=1}^T 1[A_t = a]} \sum_{t=1}^T 1[A_t = a] Y_t^a$.
- In all experiment, we recommend $\hat{a}_T = \arg\max_{a\in[K]} \hat{\mu}_T^a$.
- We optimize a sampling rule that generates a sequence of $A_1, A_2, \dots, A_T$.
- **Case 1: complete information is available.**
- Non-adaptive experiment (samples are i.i.d.).
- Compute $w^* = \arg\max_{w\in\mathcal{W}} \sum_{a\in[K]} w(a)\mathrm{KL}(Q^a, P_0^a)$.
- Draw treatment arms (A_t) following the ratio w^*.
- **Case 2: Only variance and hypothetical best arm are available.**
- Non-adaptive experiment (samples are i.i.d.).
- Compute $w^*(a^*) = \frac{\sigma^{a^*}}{\sigma^{a^*} + \sqrt{\sum_{b\in[K]\backslash\{a^*\}} (\sigma^b)^2}}$, $w^*(a) = (1 - w^*(a^*)) \frac{(\sigma^b)^2}{\sum_{b\in[K]\backslash\{a^*\}} (\sigma^b)^2}$.
- Draw treatment arms (A_t) following the ratio w^*.
- **Case 2': Only hypothetical best arm is available.**
- Adaptive experiment (samples are non-i.i.d.).
- Estimate w^* in Case 2 by replacing the variance $(\sigma^a)^2$ with its estimator $(\hat{\sigma}^a)^2$ (a sample variance).
- Procedure of experiments.
- **Split T rounds into two stages.**
- At the first stage, we uniformly randomly draw arms.
- At the second stage, we draw arms (A_t) following an estimated ratio $\widehat{w}$.
- **Case 3: No information is available.**
- Adaptive experiment (samples are non-i.i.d.).
- Compute $\widehat{w} = \max_{w\in\mathcal{W}} \min_{\substack{b\in[K] \\ a\in[K]\backslash\{b\}}} \frac{1}{2\left(\frac{\hat{\sigma}^b}{w(b)} + \frac{\hat{\sigma}^a}{w(a)}\right)}$.
- Under these experiments, the probabilities of misidentification match the lower bounds.

References

Kaufmann, E., Cappe, O., and Garivier, A. (2016), "On the Complexity of Best-Arm Identification in Multi-Armed ´ Bandit Models," Journal of Machine Learning Research, 17, 1–42