Semiparametric Best Arm Identification with Contextual Information

Masahiro Kato, The University of Tokyo and CyberAgent, Inc.

Takuya Ishihara, Tohoku University

Masaaki Imaizumi, The University of Tokyo

Toru Kitagawa, Brown University

1. Abstract

➢ Best arm identification (BAI) with a fixed budget and contexts.

- BAI with a fixed budget: recommend the best arm from multiple arms in the final round of an adaptive experiment.
- Before drawing an arm, we can observe contexts (covariates).

■ Goal: recommend the best arm with less failure probability.

➢ Contributions:

1. **Asymptotically optimal algorithm under a small-gap regime.**
   Existence of an asymptotically optimal algorithm was unknown.
   → Propose an optimal algorithm under a small-gap regime.
2. **BAI with contextual information.**
   Few existing methods employ contextual information in BAI.
3. **Analytical solution of sample allocation ratio.**
   Existing studies require high computational costs to obtain a sample allocation ratio in an experiment.
   → We show an analytical solution of a sample allocation ratio.

2. Best Arm Identification with Contexts

■ **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.
By drawing a treatment arm, we observe a reward of the drawn arm.

■ Each arm a has a potential outcome $Y_t^a \in \mathbb{R}$.

The distributions of $Y_{a,t}$ do not change.
Denote the mean outcome of an arm a by $\mu^a = \mathbb{E}[Y_t^a]$.

■ Contexts: d-dimensional random variable $X_t \in \mathbb{R}^d$.

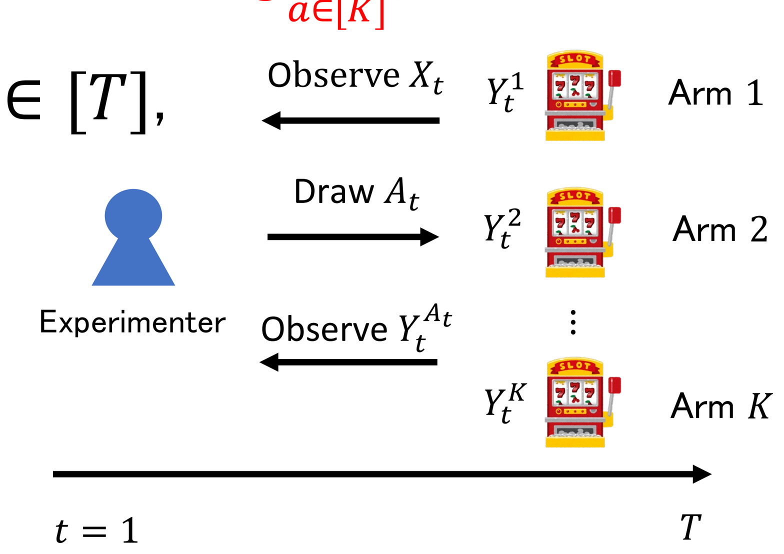
Side information such as a feature of arms.

■ Best treatment arm: an arm with the highest reward.

Denote the best treatment arm by $a^* = \arg\max_{a \in [K]} \mu^a$

■ **Bandit process**: In round $t \in [T]$,

- Observe contexts X_t.
- Pull an 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**: Minimizing the probability of misidentification: $\mathbb{P}[\hat{a}_T^* \neq a^*]$.

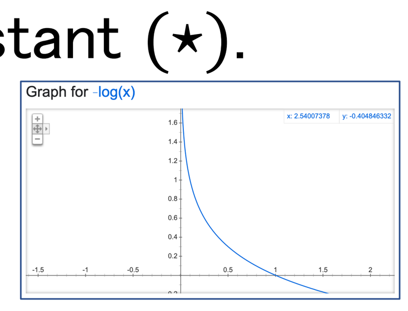
3. Evaluation

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

■ **Large deviation analysis**: tight evaluation of $\mathbb{P}[\hat{a}_T^* \neq a^*]$.

- Analysis under a "small-gap regime," where $\mu^{a^*} - \mu^a \to 0$.

Situation where it is difficult to identify the best arm.
Optimality under a large gap (constant $\mu^{a^*} - \mu^a$) is an open issue.

4. Lower Bound and Sample Allocation Ratio

■ Characterize the bound by the conditional variance.

■ $\left(\sigma^a(X_t)\right)^2$: conditional variance of Y_t^a given X_t.

Theorem 1 (Lower bound)

- For all $a \in [K]$, there exist $\Delta_0, C > 0$ such that $\mu^{a^*} - \mu^a < \Delta_0$ and $\mu^{a_0^*}(x) - \mu^a(x) = C(\mu^{a^*} - \mu^a)$.
- When $K = 2$, the lower bound of $\mathbb{P}[\hat{a}_T^* \neq a^*]$ is

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^* \neq a^*] \leq \frac{\Delta_0^2}{2\int \left(\sigma^1(x) + \sigma^2(x)\right)^2 \zeta(x)dx} + o(\Delta_0^2)$$

- When $K = 2$, the lower bound of $\mathbb{P}[\hat{a}_T^* \neq a^*]$ is

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^* \neq a^*] \leq \frac{\Delta_0^2}{2\sum_{a=1}^K \int \left(\sigma^a(x)\right)^2 \zeta(x)dx} + o(\Delta_0^2)$$

■ This lower bound suggests drawing an arm a with the following probability $w^*(a|X_t)$ (sample allocation ratio):

- When $K = 2$, $w^*(a|X_t) = \frac{\sigma^a(X_t)}{\sigma^1(X_t) + \sigma^2(X_t)}$,
- When $K \geq 3$, $w^*(a|X_t) = \frac{\left(\sigma^a(X_t)\right)^2}{\sum_{b\in[K]}\left(\sigma^b(X_t)\right)^2}$. $\forall a \in [K]$.

➢ **Lower bound under a small-gap regime.**

- $\Delta_0 \to 0$ means $\mu^{a^*} - \mu^a \to 0$.
- Draw the best arm with higher probability based on $\left(\sigma^a(X_t)\right)^2$.
- This bound gives the analytical solution of the sample allocation ratio.

5. Optimal Strategy and Upper Bound

■ Algorithm (strategy). Contextual RS-AIPW strategy.

- **RS**: random sampling of each treatment arm
- **AIPW**: recommendation using an augmented inverse probability weighting (AIPW) estimator.

An asymptotically efficient estimator of an expected reward.
This estimator is often used in causal inference literature.

➢ **Procedure of Contextual RS-AIPW strategy**:

1. In each round $t \in [T]$, estimate $\sigma^a(x)$ and a^*.
2. Using estimators of $\sigma^a(x)$ and a^*, estimate w^*.
3. Draw a treatment arm with the estimated probability $\widehat{w}_t$.
4. In round T, estimate μ^a using the AIPW estimator:

$$\hat{\mu}_T^{\mathrm{AIPW},a} = \frac{1}{T}\sum_{t=1}^{T} \frac{1[A_t = a]\left(Y_t^a - \hat{\mu}_t^a(X_t)\right)}{\widehat{w}_t(a|X_t)} + \hat{\mu}_t^a(X_t)$$

– $\hat{\mu}_t^a(X_t)$: an estimator of μ^a using samples until round t.
– This estimator consists a martingale difference sequence.

- Recommend $\hat{a}_T^{\mathrm{AIPW}} = \arg\max_{a\in[K]} \hat{\mu}_T^{\mathrm{AIPW},a}$

Theorem 2 (Upper bound)

If the estimator $\widehat{w}_t$ is consistent, when $K = 2$,

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^{\mathrm{AIPW}} \neq a^*] \geq \frac{\Delta_0^2}{2\int \left(\sigma^1(x) + \sigma^2(x)\right)^2 \zeta(x)dx} - o(\Delta_0^2);$$

when $K \geq 3$,

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}[\hat{a}_T^{\mathrm{AIPW}} \neq a^*] \geq \frac{\Delta_0^2}{2\sum_{a=1}^K \int \left(\sigma^a(x)\right)^2 \zeta(x)dx} - o(\Delta_0^2)$$

■ **Under a small-gap regime ($\mu^{a^*} - \mu^a \to 0$), the upper and lower bounds match = asymptotically optimal.**

- Estimation error of w^* is trivial under a small-gap regime.

References

M Kato, M Imaizumi, T Ishihara, T Kitagawa (2022), "Semiparametric Best Arm Identification with Contextual Information," Preprint on arXiv.