Adaptive Generalized Neyman Allocation: Local Asymptotic Minimax Optimal Best Arm Identification

arXiv Preprint. URL: https://arxiv.org/abs/2405.19317.

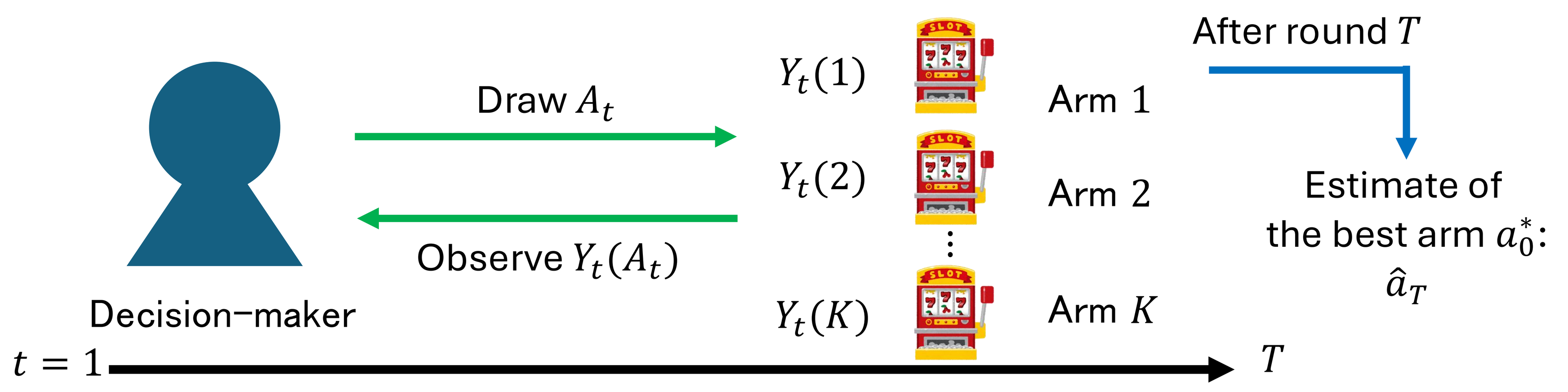
Masahiro Kato

Mizuho-DL Financial Technology

Best Arm Identification (BAI) with a Fixed Budget

- K arms: $[K] = \{1,2,\dots,K\}$.
  - Ex: slot machines, therapies, advertisements, marketing treatments
- Each arm a has the potential outcome $Y(a)$.
  - $(Y(1), Y(2), \dots, Y(K))$ is generated from P.
  - By drawing a treatment arm, we observe a reward of the drawn arm.
  - Denote the mean outcome of an arm a by $\mu_P(a) := \mathbb{E}_P[Y(a)]$.
  - Denote the variance by $\sigma_P^2(a)$.
- Best arm under P: $a^*(P) = \arg\max_{a\in[K]} \mu_P(a)$.
- **Adaptive experiment** with T rounds: $[T] = \{1,2,\dots,T\}$.
  - In each round $t \in [T]$, we draw arm $A_t \in [K]$.
  - Then, we observe the corresponding outcome $Y_t(A_t)$.

  At the end, we construct an estimator of the best arm, denoted by $\hat{a}_T$.
- Denote the true distribution that generates the data by P_0.
  - The best arm, mean, and variance under P_0: $a_0^* = a^*(P_0)$, $\mu_0(a) := \mu_P(a)$, and $\sigma_0^2(a)$.
- Goal: Minimize the probability of misidentifying the best arm: $\mathbb{P}_{P_0}(\hat{a}_T \neq a_0^*)$.



- Purpose of this study:
  - Proposal of asymptotically optimal strategy.

Neyman allocation

- The asymptotically optimal strategy in a case with $K = 2$.
- Draw arm $a \in \{1,2\}$ with the ratio of the standard deviation $\frac{\sigma_0(a)}{\sigma_0(1)+\sigma_0(2)}$.

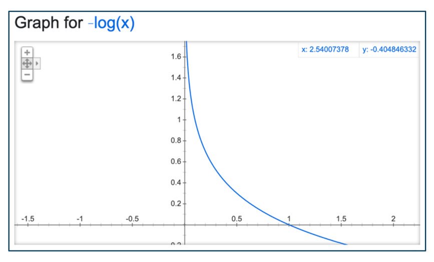
➢ **Q.** Generalize this strategy for a case with $K \geq 3$ and unknown variances.

- Contributions:
  - ✓ Proposal of the generalized Neyman allocation for a case with $K \geq 3$.
  - ✓ Prove that the strategy is asymptotically minimax optimal when the gap between the means of the best and the other arms approach zero.

Performance Measure and Lower Bound

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

    → $\mathbb{P}(\hat{a}_T^* \neq a^*) = \exp(-T(\star))$ for a constant $(\star)$.



- Performance measure: 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 $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}(\hat{a}_T^* \neq a^*)$.

➢ Lower bound for the probability of misidentifying the best arm.

- Distribution class $\mathcal{P}(P_0, \overline{\Delta})$: Set of possible P of $(Y_1, Y_2, \dots, Y_K)$.
  - For any $P \in \mathcal{P}(P_0, \overline{\Delta})$, the best arm $a^*(P)$ is not a_0^*.
  - For any $P \in \mathcal{P}(P_0, \overline{\Delta})$, $\mu_0(a_0^*) - \mu_P(a) \leq \overline{\Delta}$ holds for all a.
  - For any $P \in \mathcal{P}(P_0, \overline{\Delta})$, the variance $\sigma_P^2(a)$ of $Y(a)$ is continuous function of mean $\mu_P(a) \in \mathbb{R}$ and denoted by $\sigma^*(a; \mu_P(a))$.

$\mathcal{P}(P_0, \overline{\Delta})$ is a set of P such that for any $P \in \mathcal{P}(P_0, \overline{\Delta})$, the best arm is not a_0^* and the means are bounded by $\overline{\Delta}$.

Lower bound

- For any consistent strategy, the following holds:

$$\lim_{\overline{\Delta}\to 0} \inf_{P\in\mathcal{P}(P_0,\overline{\Delta})} \limsup_{T\to\infty} -\frac{1}{\overline{\Delta}^2 T}\log \mathbb{P}_P\left(\hat{a}_T \neq a^*(P)\right) \leq V^*(P_0),$$

where $$V^*(P_0) := \frac{1}{2\left(\sigma(a_0^*; \theta^*) + \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} \sigma^2(a; \theta^*)}\right)^2}$$ and $\theta^* = \mu_0(a_0^*)$.

Adaptive Generalized Neyman Allocation

- Algorithm: Adaptive Generalized Neyman Allocation (AGNA) strategy.
- The lower bound suggests that optimal algorithms draw an arm $a \in [K]$ with the following ratio $w^*(a)$:
  - $w^*(a_0^*) = \frac{\sigma_P(a_0^*)}{\sigma_0(a_0^*) + \sqrt{\sum_{c\in[K]\setminus\{a_0^*\}} \sigma_0^2(c)}}$,
  - $w^*(a) = \left(1 - w^*(a_0^*)\right)\frac{\sigma(a;\theta^*)}{\sum_{c\in[K]\setminus\{a_0^*\}} \sigma_0^2(c)}$ for all $a \in [K] \setminus \{a_0^*\}$.

  If we draw each arm a with the ratio $w^*(a)$, we can estimate the best arm with the probability of misidentification matching the lower bound.
- **Issue:** The best arm a_0^* and the variance are unknown.
  - We need to estimate them during the experiment.

    → Estimation error of them may affect the performance.

AGNA strategy

- In each round $t \in [T]$, we estimate $\sigma_0^2(a)$ and a_0^*.
  1. Using estimators of $\sigma_0^2(a)$ and a_0^*, we estimate the optimal ratio w^*.
  2. Draw a treatment arm with the estimated ratio $\widehat{w}_t$.
- After round T, we estimate μ^a using the augmented inverse probability weighting (AIPW) estimator:

  $$\hat{\mu}_T^{\mathrm{AIPW}}(a) = \frac{1}{T}\sum_{t=1}^{T} \frac{1[A_t = a]\left(Y_t - \hat{\mu}_t(a)\right)}{\widehat{w}_t(a)} + \hat{\mu}_t(a).$$
  - $\hat{\mu}_t(a)$: an estimator of $\mu_0(a)$ using samples until round t.
  - This estimator has the smallest asymptotic variance.
  - Make the estimator insensitive to the estimation error of a_0^* and σ_0^2.
- Estimate the best arm as $\hat{a}_T^{\mathrm{AIPW}} = \arg\max_{a\in[K]} \hat{\mu}_T^{\mathrm{AIPW}}(a)$.

Probability of misidentification of the AGNA strategy

$$\lim_{\underline{\Delta}\to 0} \limsup_{T\to\infty} -\frac{1}{\underline{\Delta}^2 T}\log \mathbb{P}_{P_0}(\hat{a}_T \neq a_0^*) \geq V^*(P_0),$$

where $\underline{\Delta} := \min_{a\in[K]} \mu_0(a_0^*) - \mu_0(a)$.

- For any $\varepsilon > 0$, there exists $\underline{\delta} > 0$ such that for all P_0 such that $\underline{\Delta} \leq \underline{\delta}$, it holds that $\limsup_{T\to\infty} -\frac{1}{\underline{\Delta}^2 T}\log \mathbb{P}_{P_0}(\hat{a}_T \neq a_0^*) \leq V^*(P_0) + \varepsilon$.

Local Asymptotic Optimality

$$\lim_{\overline{\Delta}\to 0} \inf_{P\in\mathcal{P}(P_0,\overline{\Delta})} \limsup_{T\to\infty} -\frac{1}{\overline{\Delta}^2 T}\log \mathbb{P}_P\left(\hat{a}_T \neq a^*(P)\right) \leq V^*(P_0)$$

$$\leq \lim_{\underline{\Delta}\to 0} \limsup_{T\to\infty} -\frac{1}{\underline{\Delta}^2 T}\log \mathbb{P}_{P_0}(\hat{a}_T \neq a_0^*) \geq V^*(P_0).$$

- Locality: $\overline{\Delta} \to 0$ and $\underline{\Delta} \to 0$.

- Why the "local" and "minimax" are required for evaluation?
  - Lacked knowledge about the best arm, the mean, and the variance.
  - Locality considers the worst case about the mean and allows us to ignore the estimation error of the unknown parameters.
  - Minimax evaluation considers the worst case about the best arm.