Best Arm Identification with a Fixed Budget under a Small Gap

Masahiro Kato (CyberAgent, Inc. / University of Tokyo)

Kaito Ariu (CyberAgent, Inc.), Masaaki Imaizumi (University of Tokyo),

Masahiro Nomura (CyberAgent, Inc.).

Session: Adaptive Experimental Design for Policy Choice and Policy Learning

Experimental Design for Better Decision-Making

**Keywords:** Causal inference, decision-making, and experimental design.

**Treatment arm (arm / treatment / policy).** ex. drugs, advertisements, and economic policies.
- Each treatment arm has a potential outcome. By drawing an arm, we can observe the outcome.
- We are interested in decision-making on the choice of the treatment arm.
  → From treatment effect estimation to treatment choice (decision-making).

**Treatment (policy) choice:** Choose the best treatment arm (policy) using observations.

**Multi-armed bandit problem:** Optimize decision-making with adaptive experiments.
- Regret minimization: Choose the treatment arms to maximize the cumulative reward during the experiment.
  cf. Gittins (1979), and Lai and Robbins (1985). In-sample regret.
- Best arm identification (BAI): Choose the best treatment arm after the experiment.
Consider an adaptive experiment where we can draw a treatment arm in each round. Draw a treatment arm = allocate a treatment arm to an experimental unit and observe the realized outcome.

In this presentation, I consider BAI with a fixed budget.

- The number of rounds of an adaptive experiments (budget / sample size) is predetermined.
- Recommend the best treatment arm from multiple candidates after the experiment.
  ↔ BAI with fixed confidence: continue adaptive experiments until a certain criterion is satisfied. cf. sequential experiments.

Evaluation performance metrics:

- Probability of misidentifying the best treatment arm.
- Expected simple regret (difference between the expected outcomes of best and suboptimal arms).
  Also called expected relative welfare loss, out-sample regret, and policy regret (Kasy and Sautmann 2021)

Goal: recommend the best arm with smaller probability of misidentification or expected simple regret.
In this presentation, I discuss asymptotically optimal algorithms in BAI with a fixed budget. For simplicity, I focus on the following case:

- **Two** treatment arms are given. ex. treatment and control groups.
- Potential outcomes follow **Gaussian distributions**.
- Minimization of **the probability of misidentification**.

My presentation is based on the following our paper:

Kato, Ariu, Imaizumi, Nomura, and Qin (2022),

“Optimal Best Arm Identification in Two–Armed Bandits with a Fixed Budget under a Small Gap.” *

- We show that the Neyman allocation is the worst–case optimal in this setting.

Neyman allocation rule:

• Draw a treatment arm with the ratio of the standard deviations of the potential outcomes.

• When the standard deviations are known, the Neyman allocation (Neyman 1934) is optimal.
  
  cf. Chen et al. (2000), Glynn and Juneja (2004), and Kaufmann et al. (2016).

➢ Kato, Ariu, Imaizumi, Nomura, and Qin (2022).*

• The standard deviations are unknown and estimated in an adaptive experiment.

• The worst-case asymptotic optimality of the Neyman allocation rule. **

In addition to the above paper, I introduce several other findings in my project.

• (i) Beyond the Neyman allocation rule; (ii) minimization of the expected simple regret.

* https://arxiv.org/abs/2201.04469.  ** If we know the standard deviations, the Neyman allocation rule is globally optimal (Glynn and Juneja, 2004).
Optimal Best Arm Identification in Two-Armed Bandits with a Fixed Budget under a Small Gap
Kato, Ariu, Imaizumi, Nomura, and Qin (2022)
Problem Setting

- Adaptive experiment with $T$ rounds: $[T] = \{1, 2, ..., T\}$.

- Binary treatment arms: $\{1, 0\}$.
  
  • Each treatment arm $a \in \{1, 0\}$ has a potential outcome $Y_a \in \mathbb{R}$.
  
  The distributions of $(Y_1, Y_0)$ do not change across rounds, and $Y_1$ and $Y_0$ are independent.
  
  • At round $t$, by drawing a treatment arm $a \in \{1, 0\}$, we observe $Y_{a,t}$, which is an iid copy of $Y_a$.

- **Definition:** Two-armed Gaussian bandit models.

  • A class $\mathcal{M}$ of joint distributions $\nu$ (bandit models) of $(Y_1, Y_0)$.
  
  • $(Y_1, Y_0)$ under $\nu \in \mathcal{M}$ follow Gaussian distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_0, \sigma_0^2)$.
  
  • $\sigma_a^2$ is the variance of a potential outcome $Y_a$, which is fixed across bandit models $\nu \in \mathcal{M}$. 
**Problem Setting**

- **Best treatment arm**: an arm with the highest expected outcome, $a^* = \arg\max_{a \in \{1,0\}} \mu_a$.
  
  For simplicity, we assume that the best arm is unique.

- **Bandit process**: In each round $t \in \{1,2,\ldots,T\}$, under a bandit model $\nu \in \mathcal{M}$,
  
  - Draw a treatment arm $A_t \in \{1,0\}$.
  
  - Observe an outcome $Y_{A_t,t}$ of the chosen arm $A_t$.
  
  - Stop the trial at round $t = T$
  
  - After the final round $T$, an algorithm recommends an estimated best treatment arm $\hat{a}_T \in \{1,0\}$. 

![Diagram showing the bandit process]
Best Arm Identification (BAI) Strategy

- **Probability of misidentification** $\mathbb{P}_\nu[\hat{a}_T \neq a^*]$, where $\mathbb{P}_\nu$ is a probability law under $\nu \in \mathcal{M}$. = A probability of an event that we recommend a suboptimal arm instead of the best arm.

- **Goal**: Find the best treatment arm $a^*$ efficiently with smaller $\mathbb{P}_\nu[\hat{a}_T \neq a^*]$.

- **Our actions**: Using past observations, we can optimize $A_t$ during the bandit process. We recommend an estimated best treatment arm after the experiment.

- **These actions are components of algorithms for BAI, called a strategy.**

  - **Sampling rule** $(A_1, A_2, ...)$: How we draw a treatment arm in each round $t$.

  - **Recommendation rule** $\hat{a}_T \in \{1, 0\}$: Which treatment arm we recommend as the best arm.
Contributions

Main result of Kato, Ariu, Imaizumi, Nomura, and Qin (2022).

Optimal strategy for minimization of the probability of misidentification under a small gap.

- Consider a lower bound of $\mathbb{P}_v[\hat{a}_T \neq a^*]$ that any strategy cannot exceeds.

- Propose a strategy using the Neyman allocation rule and the AIPW estimator.
  
  In the strategy, we use the standard deviations during an experiment.
  
  Using estimated standard deviations, we draw a treatment arm in each round.

- The probability of misidentification matches the lower bound when $\mu_1 - \mu_0 \to 0$. 
Probability of Misidentification

- Assume that the best arm $a^*$ is unique.
- $\mathbb{P}_\nu[\hat{a}_T \neq a^*]$ converges to 0 with an exponential speed:
  $$\mathbb{P}_\nu[\hat{a}_T \neq a^*] = \exp(-T(\star))$$
  for a constant (\star).

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

  cf. Kaufmann et al. (2016).

- Large deviation analysis: tight evaluation of $\mathbb{P}_\nu[\hat{a}_T \neq a^*]$
Kaufmann et al. (2016) gives a lower bound for two-armed Gaussian bandit models.

• To derive a lower bound, we restrict a class of strategies.

Definition: consistent strategy.

• A strategy is called consistent for a class \( \mathcal{M} \) if for each \( \nu \in \mathcal{M} \), \( \mathbb{P}_\nu[\hat{a}_T \neq a^*] \to 1 \).

Lower bound (Theorem 12 in Kaufmann et al., 2016)

• For any bandit model \( \nu \in \mathcal{M} \), any consistent strategy satisfies

\[
\limsup_{T \to \infty} -\frac{1}{T} \log \mathbb{P}_\nu[\hat{a}_T \neq a^*] \leq \frac{\Delta^2}{2(\sigma_1 + \sigma_0)^2}.
\]

Any strategy cannot exceed this convergence rate of the probability of misidentification.

A lower bound of the probability of misidentification \( \mathbb{P}_\nu[\hat{a}_T \neq a^*] \) is an upper bound of \( \frac{1}{T} \log \mathbb{P}_\nu[\hat{a}_T \neq a^*] \).

Optimal strategy: a strategy under which \( \mathbb{P}_\nu[\hat{a}_T \neq a^*] \) matches the lower bound.
Neyman Allocation Rule

- **Target allocation ratio.**

  - A ratio of the expected number of arm draws \( \left( \frac{1}{T} \mathbb{E}_v[\sum_{t=1}^T 1[A_t = a]] \right) \) under a sampling rule.
    
    \[ \frac{1}{T} \mathbb{E}_v[\sum_{t=1}^T 1[A_t = a]] / \sum_{b \in [K]} \frac{1}{T} \mathbb{E}_v[\sum_{t=1}^T 1[A_t = b]]. \]

    - \( \mathbb{E}_v \) is an expectation under a bandit model \( v \in \mathcal{M} \).

  - **Neyman allocation rule.**

    - Target allocation ratio is the ratio of the standard deviations.

    \[ = \text{Draw a treatment arm as } \frac{1}{T} \mathbb{E}_v[\sum_{t=1}^T 1[A_t = 1]]: \frac{1}{T} \mathbb{E}_v[\sum_{t=1}^T 1[A_t = 0]] = \sigma_1: \sigma_0. \]

  - **When the standard deviations \( \sigma_1 \) and \( \sigma_0 \) are known, the Neyman allocation is optimal.**


  - **An optimal strategy is unknown when the standard deviations are unknown.**

    - In our strategy, we estimate \((\sigma_1, \sigma_0)\) and draw an arm \( a \) with the probability \( \frac{\hat{\sigma}_a}{\hat{\sigma}_1 + \hat{\sigma}_0} \).
NA–AIPW Strategy

 Proposed strategy: **NA–AIPW strategy.**

- **NA**: sampling rule following the Neyman Allocation rule.
- **AIPW**: recommendation rule using an Augmented Inverse Probability Weighting (AIPW) estimator.

➢ Procedure of the NA–AIPW strategy:

1. In each round $t \in [T]$, estimate $\sigma_a^2$ using observations obtained until round $t$.

2. Draw a treatment arm $a \in \{1,0\}$ with a probability $\hat{w}_t(a) = \frac{\hat{\sigma}_{a,t}}{\hat{\sigma}_{1,t} + \hat{\sigma}_{0,t}}$ (Neyman allocation rule).

3. In round $T$, estimate $\mu^a$ using the AIPW estimator: $\hat{\mu}_a,T = \frac{1}{T} \sum_{t=1}^{T} \frac{1[A_t=a](Y_{a,t}-\hat{\mu}_{a,t})}{\hat{w}_t(a)} + \hat{\mu}_{a,t}$.  

\[
\hat{\mu}_{a,t} = \frac{1}{\sum_{s=1}^{t-1} 1[A_s=a]} \sum_{s=1}^{t} 1[A_s = a]Y_{a,t}
\]
is an estimator of $\mu_a$ using observations until round $t$.

4. Recommend $\hat{a}_T^{AIPW} = \arg \max_{a \in \{1,0\}} \hat{\mu}_a,T$ as an estimated best treatment arm.

We can apply this strategy to a case with batched updates (multiple waves)
Assume some regularity conditions. Suppose that the estimator $\hat{w}_t$ converges to $w^*$ almost surely (with a certain rate).

Then, for any $\nu \in \mathcal{M}$ such that $0 < \mu_1 - \mu_0 \leq C$ for some constant $C > 0$, the upper bound is

$$\limsup_{T \to \infty} - \frac{1}{T} \log P_\nu[\hat{a}_T^{AIPW} \neq a^*] \geq \frac{\Delta^2}{2(\sigma_1 + \sigma_0)^2} - \tilde{C}(\Delta^3 + \Delta^4),$$

where $\tilde{C}$ is some constant.

This result implies that

$$\lim_{\Delta \to 0} \limsup_{T \to \infty} - \frac{1}{\Delta^2 T} \log P_\nu[\hat{a}_T^{AIPW} \neq a^*] \geq \frac{1}{2(\sigma_1 + \sigma_0)^2} - o(1).$$

Under a small-gap regime ($\Delta = \mu_1 - \mu_0 \to 0$), the upper and lower bounds match.

Thus, the NA–AIPW strategy is asymptotically optimal under the small gap.
On the Optimality under the Small Gap

- Asymptotically optimal strategy under a small gap.

  - This result implies the worst-case optimality of the proposed algorithm.

- A technical reason for the small gap.

  - There is no optimal strategy when the gap is fixed, and the standard deviations are unknown.

    ↔ When the standard deviations are known, the Neyman allocation is known to be optimal.

      cf. Chen et al. (2000), Glynn and Juneja (2004), and Kaufmann et al. (2016).

- When the gap is small, we can ignore the estimation error of the standard deviations.

  ↑ The estimation error is trivial compared with the difficulty of identifying the best arm under the small gap.

  ✓ Optimality under a large gap (constant $\mu_1 - \mu_0$) is an open issue.

Simulation Studies

Empirical performance of the NA–AIPW (NA) strategy.

- Compare the NA strategy with the $\alpha$-elimination (Alpha) and Uniform sampling (Uniform).
  
  The $\alpha$-elimination is a strategy using the Neyman allocation when the standard deviations are known (Kaufmann et al., 2016).
  
  The uniform sampling draw each treatment arm with equal probability. A randomized controlled trial without adaptation.

  - Setting 1: $\mu_1 = 0.05$, $\mu_0 = 0.01$, $\sigma_1^2 = 1$, $\sigma_0^2 = 0.2$.
  - Setting 2: $\mu_1 = 0.05$, $\mu_0 = 0.01$, $\sigma_1^2 = 1$, $\sigma_0^2 = 0.1$.

  We draw treatment arm 1 in Setting 2 more often than in Setting 1.

Strategies using the Neyman allocation outperform the RCT.

- Under the NA–AIPW strategy, we can identify the best arm with a lower probability of misidentification than the RCT (uniform sampling).
Beyond the Neyman Allocation Rule (ongoing)
Limitations of the Neyman Allocation Rule

- I briefly introduce my ongoing other work.
  - Several contents are still conjectures and not published.
- The Neyman allocation rule.
  - Consider a case where there are two treatment arms.
  - Not consider covariates (contextual information).
- Extensions of the NA–AIPW strategy with multiple treatment arms and contextual information.
  - \(K\) treatment arms: \([K] = \{1, 2, \ldots, K\}\).
  - Covariate (context): \(d\)-dimensional random variable \(X \in \mathcal{X} \subseteq \mathbb{R}^d\). Side information such as a feature of arms.
Problem Setting

- Let $\nu$ be a joint distribution of $(Y_1, \ldots, Y_K, X)$, called a bandit model.
  - $\mu_a(\nu) = \mathbb{E}_\nu[Y_{a,t}], \mu_a(\nu)(x) = \mathbb{E}_\nu[Y_{a,t}|X_t = x]$.

- **Best treatment arm**: an arm with the highest expected outcome, $a^*(\nu) = \arg\max_{a \in [K]} \mu_a(\nu)$.

- In each round $t \in \{1, 2, \ldots, T\}$, under a bandit model $\nu$,
  - Observe a covariate (context) $X_t \in \mathcal{X}$.
  - Draw a treatment arm $A_t \in [K]$.
  - Observe an outcome $Y_{A_t,t}$ of chosen arm $A_t$.
  - An algorithm recommends an estimated best treatment arm $\hat{a}_T \in [K]$. 

---

![Diagram](image)

**Observation Process**

- **Experiment**
  - Draw $A_t$ from $[K]$.
  - Observe $Y_{A_t,t}$.
  - Recommend $\hat{a}_T$. 

- **Bandit**
  - Observe $X_t$.
  - Observe $Y_{1:t}$.

- **Time**
  - $t = 1, 2, \ldots, T$
Bandit Models and Strategy Class

To derive lower bound, consider other restrictions on bandit models and strategies.

Definition: Location-shift bandit class $\mathcal{P}$.

- For all $\nu \in \mathcal{P}$ and $x \in \mathcal{X}$, the conditional variance of $Y_{\alpha,t}$ is constant.
- For all $a \in [K]$ and any $x \in \mathcal{X}$, there exists a constant $\sigma^2_a(x)$ such that $\text{Var}_\nu(Y_{\alpha,t}|X_t = x) = \sigma^2_a(x)$ for all $\nu \in \mathcal{P}$.
- For all $\nu \in \mathcal{P}$, $X$ has the same distribution and denote the density by $\zeta(x)$.
  
  ex. Gaussian distributions with fixed variances. An extension of Gaussian distributions.

Definition: Asymptotically invariant strategy.

- A strategy is asymptotically invariant for $\mathcal{P}$ if for any $\nu, \nu' \in \mathcal{P}$, for all $a \in [K]$ and any $x \in \mathcal{X}$,
  
  $\mathbb{E}_\nu \left[ \sum_{t=1}^T 1[A_t = a] | X_t = x \right] = \mathbb{E}_{\nu'} \left[ \sum_{t=1}^T 1[A_t = a] | X_t = x \right].$

  The sampling rule does not chance across $\nu \in \mathcal{P}$.

I conjecture that if potential results follow particular distributions, such as Bernoulli, such restrictions may not be necessary, and an RCT is optimal.
Lower Bound

Theorem (Lower bound)

• Consider a location–shift bandit class $\mathcal{P}$ and $\nu \in \mathcal{P}$.

• Assume that there is a unique best treatment arm $a^*(\nu)$.

• Assume that for all $a \in [K]$, there exists a constant $\Delta > 0$ such that $\mu_{a^*(\nu)}(\nu) - \mu_a(\nu) < \Delta$.

• Then, for any $\nu$ in a location–shift class, any consistent and asymptotically invariant strategy satisfies

  if $K = 2$: $\limsup_{T \to \infty} -\frac{1}{T} \log \mathbb{P}_\nu[\hat{a}_T^* \neq a^*(\nu)] \leq \frac{\Delta^2}{2 \int (\sigma_1(x) + \sigma_2(x))^2 \zeta(x) dx} + C_1 \Delta^3$;

  if $K \geq 3$ and strategy is invariant: $\limsup_{T \to \infty} -\frac{1}{T} \log \mathbb{P}_\nu[\hat{a}_T^* \neq a^*(\nu)] \leq \frac{\Delta^2}{2 \sum_{b \in [K]} \int \sigma_b^2(x) \zeta(x) dx} + C_2 \Delta^3$,

where $C_1, C_2 > 0$ are some constant.
Target Allocation Ratio and Optimal Strategy

The lower bound suggests drawing an arm $a$ with the following probability $w^*(a|X_t)$:

- if $K = 2$, $w^*(a|X_t) = \frac{\sigma_a(X_t)}{\sigma_1(X_t) + \sigma_2(X_t)}$ for $a \in [2]$; if $K \geq 3$, $w^*(a|X_t) = \frac{\sigma_a^2(X_t)}{\sum_{b \in [K]} \sigma_b^2(X_t)}$ for $a \in [K]$.

Beyond the Neyman allocation rule: when $K \geq 3$, draw arms with the ratio of the variances.

Replace the Neyman allocation rule in the NA–AIPW strategy with $w^*(a|x)$ defined here.

- In $t \in [T]$, estimate $\sigma_a(X_t)$ using samples until round $t$ and draw an arm with an estimated $\hat{\sigma}_t$.

- In round $T$, estimate $\mu_a(\nu)$ using the AIPW estimator: $\hat{\mu}_{a,T} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{[A_t = a]} (r_{a,t} - \hat{\mu}_{a,t}(X_t)) + \hat{\mu}_{a,t}(X_t)$. $\hat{\mu}_{a,t}(X_t)$: an estimator of $\mu_a(\nu)(x)$ using samples until round $t$.

- Recommend $\hat{a}_T^{AIPW} = \arg \max_{a \in [K]} \hat{\mu}_{a,T}^{AIPW}$ as an estimated best treatment arm.

This strategy is asymptotically optimal under the small gap as well as the NA–AIPW strategy.

When $K = 2$, the target allocation ratio is identical to that in average treatment effect estimation, such as Hahn Hirano, and Karlan (2011).
Expected Simple Regret

- Relationship between the probability of misidentification and expected simple regret.

- **Simple regret:** $r_T(\nu) = \mu_{a^*(\nu)}(\nu) - \mu_{\hat{a}_T}(\nu)$ under a bandit model $\nu$ (there is a randomness of $\hat{a}_T(\nu)$).

- **Expected simple regret:** $\mathbb{E}_\nu[r_T(\nu)] = \mathbb{E}_\nu[\mu_{a^*(P)}(\nu) - \mu_{\hat{a}_T}(\nu)]$. ($\mathbb{E}_\nu$ is the expectation over $\hat{a}_T(\nu)$).
  - The expected simple regret represents an expected relative welfare loss.
  - In economics, the expected simple regret is often more meaningful than the probability of misidentification.

- A gap between the expected outcomes of arms $a, b \in [K]$: $\Delta^{a,b}(\nu) = \mu_a(\nu) - \mu_b(\nu)$.

- By using the gap $\Delta^{a,b}(\nu) = \mu_a(\nu) - \mu_b(\nu)$, the expected regret can be decomposed as
  \[
  \mathbb{E}_\nu[r_T(\nu)] = \mathbb{E}_\nu[\mu_{a^*(\nu)}(\nu) - \mu_{\hat{a}_T}(\nu)] = \sum_{b \notin \mathcal{A}^*_\nu} \Delta^{a^*(\nu),b}(\nu) \mathbb{P}_\nu[\hat{a}_T = b].
  \]
  - The probability of misidentification.
  - A set of the best treatment arms.

- For some constant $C > 0$, $\mathbb{E}_\nu[r_T(\nu)] = \sum_{b \notin \mathcal{A}^*_\nu} \Delta^{a^*(P),b}(\nu) \exp\left(-CT(\Delta^{a^*(P),b}(\nu))^2\right)$. 

Expected Simple Regret

The speed of convergence to zero of $\Delta^{a^*(P),b}(\nu)$ affects the of $\mathbb{E}_\nu[r_t(P)]$ regarding $T$.

1. $\Delta^{a^*(\nu),b}(\nu)$ is slower than $1/\sqrt{T} \rightarrow$ For some increasing function $g(T)$, $\mathbb{E}_\nu[r_t(\nu)] \approx \exp(-g(T))$.

2. $\Delta^{a^*(\nu),b}(\nu) = C_1/\sqrt{T}$ for some constant $C_1 \rightarrow$ For some constant $C_2 > 0$, $\mathbb{E}_\nu[r_t(\nu)] \approx \frac{C_2}{\sqrt{T}}$.

3. $\Delta^{a^*(\nu),b}(\nu)$ is faster than $1/\sqrt{T} \rightarrow \mathbb{E}_\nu[r_t(\nu)] \approx o(1/\sqrt{T})$

→ In the worst case, $\Delta^{a^*(\nu),b}$ converges to zero with $C_1/\sqrt{T}$ (Bubeck et al., 2011). cf. Limit of experiment framework.

✓ When $\Delta^{a,b}(\nu)$ is independent from $T$, evaluation of $\mathbb{E}_\nu[r_t(\nu)]$ is equivalent to that of $\mathbb{P}_\nu[\hat{a}^*_T = b]$.

• $\mathbb{P}_\nu[\hat{a}^*_T = b]$ converges to zero with an exponential speed if $\Delta^{a,b}(\nu)$ is independent from $T$.

• $\Delta^{a^*(\nu),b}$ does not affect the rate.

→ For some constant ($\star$), if $\mathbb{P}_\nu[\hat{a}^*_T = b] \approx \exp(-T(\star))$ for $b \notin \mathcal{A}^*(\nu)$, then $\mathbb{E}_\nu[r_t(\nu)] \approx \exp(-T(\star))$.

• Our result on the small gap optimality of $\mathbb{P}_\nu[\hat{a}^*_T = b]$ is directly applicable to the optimality of $\mathbb{E}_\nu[r_t(\nu)]$. 
Summary
Summary

- **Asymptotically optimal strategy** in two-armed Gaussian BAI with a fixed budget.

- Evaluating the performance of BAI strategies by the probability of misidentification.
  - The Neyman allocation rule is globally optimal when the standard deviations are known.
    - The Neyman allocation is known to be asymptotically optimal when potential outcomes of two treatment arms follow Gaussian distributions with any mean parameters and fixed variances.
    - The standard deviations are unknown and estimated during an experiment.
    - Under the NA–AIPW strategy, the probability of misidentification matches the lower bound when the gap between expected outcomes goes to zero.
      - The strategy based on the Neyman allocation is the worst-case optimal (small-gap optimal).
Reference


• Neyman, J (1934). “On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection.” JRSSB