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

Masahiro Kato [1] **Masaaki Imaizumi** [1] **Takuya Ishihara** [2] **Toru Kitagawa** [3]

Abstract

Experimental design is crucial in evidence-based decision-making with multiple treatment arms, such as online advertisements and medical treatments. This study investigates an experiment whose task is to identify the best treatment arm with the highest expected outcome. In our experiments, given a fixed sequence of sample-allocation rounds and multiple treatment arms, we allocate a sample to a treatment arm and observe a corresponding outcome at each round. At the end of the experiment, we recommend one of the treatment arms as the best based on the observations. We aim to design an experiment that minimizes the probability of misidentifying the *best treatment arm*. This problem has been explored under various names across numerous research fields, including *best arm identification* (BAI) and ordinal optimization. With this objective in mind, we initially derive lower bounds for the probability of misidentification through an information-theoretic approach, enabling discussions on the asymptotic optimality of experiments. In our analysis, we discover that the available information on the distribution of rewards for each treatment arm significantly influences the asymptotic optimality of experiments. Moreover, we find that the asymptotic optimality depends on a pre-specified set of *hypothetical best treatment arms* utilized for sample allocation. Existing experiments become asymptotically optimal when the true best treatment arm is in the set. The standard BAI is a special case in which all treatment arms are hypothetical best treatment arms. Based on the lower bounds, we design experiments whose probability of misidentification matches the lower bounds given the available information.

1. Introduction

Experimental design is integral to decision-making processes (Fisher, 1935; Robbins, 1952). This study explores scenarios involving multiple *treatment arms*, such as online advertisements, slot machine arms, diverse therapeutic strategies[1], or assorted unemployment assistance programs. The objective is to identify the treatment arm that provides the highest expected outcome at the end of an experiment. During the experiment, we allocate each sample to a treatment arm and recommend the treatment arm deemed the best at the end, with the aim of minimizing the probability of incorrectly identifying the best treatment arm. Both non-adaptive and adaptive experiments are considered. In non-adaptive experiments, we fix the sample allocation rule at the beginning of an experiment, while adaptive experiments allow us to optimize sample allocation throughout the experiment based on acquired data. This issue of designing such experiments has been examined in various research areas under a range of names, including best arm identification (BAI, Audibert et al., 2010), ordinal optimization (Ho et al., 1992), optimal budget allocation (Chen et al., 2000), and policy choice (Kasy & Sautmann, 2021).

Design of an optimal experiment depends on how much information on the distribution of treatment arms' rewards is available before the experiment. In scenarios where complete distributional information is available, we can discuss experimental designs that are *globally asymptotically optimal for any instances*, grounded on the large-deviation principle (Glynn & Juneja, 2004; Chen et al., 2000; Gärtner, 1977; Ellis, 1984). However, we often face situations where only partial or no distributional knowledge is available. Because optimal experiments are characterized by distributional information, incomplete information prevents attaining a globally asymptotically optimal experiments. Although we can obtain distributional information during an adaptive experiment, the estimation error from the missing information affects performance. We consider that such a information loss is a cause of non-existence of globally optimal experiments in this problem (Kaufmann, 2020).

The design of an optimal experiment depends on the ex-

[1]The term treatment arm is frequently used in clinical trials and economic contexts (Nair, 2019). Other literature refers to treatment arms by various names, including arms (Lattimore & Szepesvári, 2020), policies (Kasy & Sautmann, 2021), treatments (Hahn et al., 2011), designs (Chen et al., 2000), systems, populations (Glynn & Juneja, 2004), and alternatives (Shin et al., 2018).

tent of available information regarding the distribution of rewards of treatment arms prior to the experiment. In situations where complete distributional information is accessible, we can deliberate on experimental designs that are *globally asymptotically optimal for all instances*, based on the large-deviation principle (Glynn & Juneja, 2004; Chen et al., 2000; Gärtner, 1977; Ellis, 1984). However, it is common to encounter scenarios where only partial or no distributional knowledge is available. Since optimal experiments are characterized by distributional information, the lack of complete information hinders the attainment of globally asymptotically optimal experiments. Even though distributional information can be acquired during an adaptive experiment, the estimation error stemming from the absence of complete information impacts the experiment's performance. We propose that such information loss contributes to the nonexistence of globally optimal experiments in this context (Kaufmann, 2020; Degenne, 2023).

To explore the asymptotic optimality of experiments under information loss, we first establish lower bounds for the probability of misidentification based on the available information. From a theoretical perspective, information theory implies that the lower bounds for the probability of misidentification in experiments with complete information are characterized by the Kullback–Leibler (KL) divergence (Lai & Robbins, 1985; Kaufmann et al., 2016). However, when only incomplete information is available, we need to reflect this limitation in the lower bounds.

To address this issue, we perform a worst-case analysis, where the expected outcomes of both the optimal and suboptimal treatment arms converge to zero. We term this condition as the *small-gap regime*. In this regime, the information loss becomes relatively insignificant in comparison to the challenge of identifying the optimal treatment arm presented by the small gap. This scenario thus facilitates the evaluation of worst-case or local asymptotic optimality.

While the lower bounds with complete information are characterized by the KL divergence (Lai & Robbins, 1985; Kaufmann et al., 2016), those in the small-gap regime are characterized by variance, which arises from a second-order approximation of the KL divergence. Hence, knowledge of at least the variances is sufficient to design worst-case optimal experiments within the small-gap regime. Based on the lower bounds and available information, we design experiments that are either globally or locally optimal.

Furthermore, during our analysis, we find that experiments proposed in existing studies, such as Chen et al. (2000), Glynn & Juneja (2004), and Shin et al. (2018), achieve asymptotic optimality only when a specific treatment arm is presumed to be the best prior to an experiment. We refer to these treatment arms as the *hypothetical best treatment arms*. We demonstrate that such redesigned experiments are asymptotically optimal in that their misidentification probability matches the lower bounds when the hypothetical best treatment arm is, indeed, the best one. In this context, the lower bounds of experiments depend on the hypothetical best treatment arm and the knowledge about the distributional information.

Therefore, our analysis begins with the reformulation of the standard problem setting. We initially specify one or multiple hypothetical treatment arms and allocate samples to minimize the probability of misidentification when the hypothetical best treatment arms include the true optimal treatment arm. When all treatment arms are considered as hypothetical best treatment arms, this setting reduces to the standard setting of ordinal optimization and BAI.

Based on available information on distributions and hypothetical treatment arms, we classify experimental designs into four scenarios: *globally optimal non-adaptive experimental design* (GO-NonAED), *hypothetical locally optimal non-adaptive experimental design* (H-LO-NonAED), *hypothetical locally optimal adaptive experimental design* (H-LO-AED), and *locally optimal adaptive experimental design* (LO-AED). When complete information is available, we conduct non-adaptive experiments, referred to as the GO-NonAED. For instance, with Gaussian distributions and known means and variances, optimal experimental design can be computed (Chen et al., 2000). If only variances are known, we design locally optimal experiments under small gaps and refer to it as the H-LO-NonADE. As explained above, under the small-gap regime, optimal experiments depend only on variances because of the second-order approximation of the KL divergence. If we lack variances, we resort to adaptive experiments to estimate them, which allows in-experiment treatment allocation optimization based on past observations. When we have hypotheses, we refer to the problem as the H-LO-AED. When we do not have any hypotheses, we refer to it as the LO-AED. Experimental design without hypothetical treatment arms is the standard setting in BAI and ordinal optimization. In the LO-AED, we show the asymptotic optimality of experiments for the standard BAI and ordinal optimization under the small-gap. We also refer to experiments with known variances without a hypothetical best treatment arm as the *locally optimal non-adaptive experimental design* (LO-NonAED), which is just a special case of the LO-AED, and we omit the details. We summarize our categorization in Table 1.

In the GO-NonAED and the H-LO-NonAED, based on arguments by Glynn & Juneja (2004), we design non-adaptive experiments, referred to as the Non-Adaptive sampling (NA)-Empirical Best Arm recommendation (EBA) experiment. For the H-LO-ADE and the LO-ADE, we design adaptive experiments, referred to as the Two-Stage sampling (TS)-EBA experiment. We derive their upper bounds for the

Table 1. Comparison of experimental design when the number of treatment arms is more than three.

Complete information is available	Incomplete or no information is available		
		Variances are known (Non-adaptive experiments)	No prior information (Adaptive experiments)
GO-NonAED	Hypothetical best treatment arm		
	One arm is a candidate for the best.	H-LO-NonAED	H-LO-AED
	No information	LO-NonAED	LO-AED

misidentification probability and validate their asymptotic optimality through matching lower and upper bounds as the sample size goes to infinity, and gaps converge to zero.

Organization. In Section 2, we define our problem. In Section 3, we summarize our main contributions. In Section 4, we derive the lower bounds for an experiment based on the available information. In Section 5, we categorize problem settings based on the lower bounds and available information. In Appendix C and Section 6, for the established lower bounds, we design optimal experimental strategies. This study incorporates the findings from Kato et al. (2023b). To better highlight our contributions, we significantly revised the previous draft. Notably, we expanded the discussion on the asymptotic optimality of experiments. Moreover, while the previous draft tackled the aspect of contextual information, we omitted this section in this study, choosing to focus more on the problem of asymptotic optimal experiments in BAI. We intend to address the contextual information issue independently in a future publication.

2. Problem Setting

We consider the following setting. Given a fixed number of rounds T, referred to as a sample size or a budget, in each round $t \in [T] := \{1, 2, \dots, T\}$, an experimenter allocates a treatment arm $A_t \in [K] := \{1, 2, \dots, K\}$ to an experimental subject. Then, the experimenter immediately receives an outcome (or a reward) Y_t linked to the allocated treatment arm A_t. This setting is called bandit feedback. Our goal is to find a treatment arm with the highest expected outcome with a minimal probability of misidentification after observing the outcome in the round T.

Potential outcomes. To describe the data-generating process, we introduce potential outcomes following the Neyman-Rubin causal model (Neyman, 1923; Rubin, 1974). An outcome in round $t \in [T]$ is $Y_t = \sum_{a \in [K]} \mathbb{1}[A_t = a] Y_t^a$, where $Y_t^a \in \mathbb{R}$ is a potential independent outcome (random variable), and $Y_t^1, Y_t^2, \dots, Y_t^K$ are independent and identically distributed (i.i.d.) over $t \in [T] = \{1, 2, \dots, T\}$.

Let P be a joint distribution of K-potential outcomes, $(Y^1, Y^2, \dots, Y^K)$, and $(Y_{1,t}, Y_{2,t}, \dots, Y_{K,t})$ be an independent copy of $(Y^1, Y^2, \dots, Y^K)$ at round $t \in [T]$ under P. We refer to P as a statistical model[2]. For P, let $\mathbb{P}_P$, and $\mathbb{E}_P$ be the probability and expectation under P respectively and $\mu^a(P) = \mathbb{E}_P[Y^a]$ be the expected outcome. Let $\mathcal{P}$ be a set of all joint distributions P such that the the best treatment arm $a^*(P) = \arg\max_{a \in [K]} \mu^a(P)$ uniquely exists; that is, there exists $a^*(P) \in [K]$ such that $\mu^{a_0^*}(P) > \max_{b \in [K] \setminus a_0^*} \mu^b(P)$. Let $P_0 \in \mathcal{P}$ be the *true* statistical model that generates the potential outcomes.

Probability of misidentification. Our goal is to minimize the *probability of misidentification*, defined as $\mathbb{P}_{P_0}(\widehat{a}_T \neq a^*(P_0))$. It is known that for each fixed $P \in \mathcal{P}$, when $a^*(P_0)$ is unique, $\mathbb{P}_{P_0}(\widehat{a}_T \neq a^*(P_0))$ converges to zero with an exponential speed as $T \to \infty$. Therefore, to evaluate the exponential speed, we employ the following measure, called the *complexity*: $-\frac{1}{T} \log \mathbb{P}_{P_0}(\widehat{a}_T \neq a^*(P_0))$.

Experiment. We define an *experiment* as a combination of treatment allocation and best-treatment-arm recommendation rules; formally, with the sigma-algebras $\mathcal{F}_t = \sigma(A_1, Y_1, \dots, A_t, Y_t)$, an experiment is a pair $((A_t)_{t \in [T]}, \widehat{a}_T)$, where $((A_t)_{t \in [T]}$ is a sampling rule, which allocates a treatment arm $A_t \in [K]$ in each round t based on the past observations $\mathcal{F}_{t-1}$, and $\widehat{a}_T$ is a recommendation rule, which is an $\mathcal{F}_T$-measurable estimator of the best treatment arm $\widehat{a}^*(P)$ using observations up to round T. We denote an experiment by π and denote A_t and $\widehat{a}_T$ by A_t^π and $\widehat{a}_T^\pi$ when we emphasize that A_t and $\widehat{a}_T$ depend on π. In existing studies, an experiment is also referred to as different names, such as *strategy* and an algorithm. Our goal is equal to designing an experiment that minimizes its probability of misidentification when the null hypothesis false.

Hypothetical best treatment arm. In the following analysis, we also consider a situation where there is a candidate of the best treatment arm, denoted by $\widetilde{a} \in [K]$. We refer to it as a hypothetical best treatment arm and consider minimizing the probability of misidentification when $\widetilde{a}$ is equal to a_0^*; that is, $\mathbb{P}_{P_0}(\widehat{a}_T \neq a^*(P_0))$ for $P_0 \in \mathcal{P}$ such that $a^*(P_0) = \widetilde{a}$. We refer to the treatment arm $\widetilde{a}$ as the *hypothetical best treatment arm*. We raise the following examples for this problem.

Example (Online advertisement). Let $\widetilde{a} \in [K]$ be a treatment arm corresponding to a new advertisement. Our null hypothesis $a^* \neq \widetilde{a}$ implies that the existing advertisements $a \in [K] \setminus \{\widetilde{a}\}$ are superior to the new advertisement. Our goal is to reject the null hypothesis with a maximal probability when the null hypothesis is not correct; that is, the new hypothesis is better than the others.

Example (Clinical trial). Let $\widetilde{a} \in [K]$ be a new drug. Our null hypothesis $a^* \neq \widetilde{a}$ implies that the existing drug $a \in [K] \setminus \{\widetilde{a}\}$ is superior to the new drug (equivalently, the new drug is not good as the existing drugs). Our goal is to reject

[2] In particular, we refer to distributions of the potential outcomes $(Y^1, Y^2, \dots, Y^K)$ as full-data statistical models.

the null hypothesis with a maximal probability when the new drug is better than the others.

This setting corresponds to considering a null and alternative hypotheses H_0 and H_1 such that $H_0 : a^*(P_0) \neq \widetilde{a} \in [K]$ and $H_1 : a^*(P_0) = \widetilde{a}$; that is, the null hypothesis corresponds to a situation where the hypothetical best treatment arm is *not* the best, while the alternative hypothesis posits that the hypothetical best treatment arm is the best. Then, we consider minimizing the probability of misidentification when the alternative hypothesis is correct. This probability corresponds to power in hypothesis testing. We aim to minimize misidentification probability when the null hypothesis is false, corresponding to the *power of the test*.

We consider this setting not only because of its practical importance but also because the existence of the hypothetical best treatment arm is technically required in several existing methods. For instance, although not explicitly stated, the experiments in Glynn & Juneja (2004) require a hypothetical best treatment arm for asymptotic optimality.

Notation. For all $P \in \mathcal{P}$, and all $a \in [K]$, let $(\sigma^a(P))^2$ be the variance of Y^a. For the true statistical model $P_0 \in \mathcal{P}$, we denote $\mu^a(P_0) = \mu_0^a$, $(\sigma^a(P_0))^2 = (\sigma_0^a)^2$, and $a^*(P_0) = a_0^*$. Let $Y_t^{a_0^*} = Y_t^*$, and $\mu^{a_0^*}(P_0) = \mu_0^*$. Let $\Delta^a(P) = \mu^{a^*(P)}(P) - \mu^a(P)$ and $\Delta_0^a = \Delta^a(P_0) = \mu_0^* - \mu_0^a$. For $P \in \mathcal{P}$, let P^a be a distribution of a reward of treatment arm $a \in [K]$. For the two Bernoulli distributions with mean parameters $\mu, \mu' \in [0, 1]$, we denote the KL divergence by $d(\mu, \mu') = \mu \log(\mu/\mu') + (1 - \mu) \log((1 - \mu)/(1-\mu'))$ with the convention that $d(0, 0) = d(1, 1) = 0$.

3. Summary of Main Contributions: Information-Loss in Experimental Design

Experiments are designed based on information that is available prior to the experiments. We focus on hypothetical best treatment arm in hypothesis testing and distributional information. For example, when P^a follows a Gaussian distribution, we can consider the following information that can be used for experimental design: **hypothetical Best treatment arm:** a treatment arm $\widetilde{a} \in [K]$ that an experimenter expects that it is the best among treatment arms; **mean parameter:** a set $(\mu_0^a)_{a\in[K]}$; **variance parameter:** a set $((\sigma_0^a)^2)_{a\in[K]}$. Mean and variance parameters correspond to the distributional information. In general, statistical models can include more various parameters.

A hypothetical best treatment arm may not be the true best treatment arm. As well as statistical hypothesis testing, an experimenter has a null hypothesis that the hypothetical best treatment arm does not have the highest expected outcomes and an alternative hypothesis that the hypothetical best treatment is the best. As explained in Section 2, by regarding a hypothesis such that $a_0^* \neq \widetilde{a}$ as a null hypothesis, we discuss an experimental design that maximizes the probability of misidentification when the null is not correct; that is, the power of the test, $\mathbb{P}(\widehat{a}_T = a_0^*)$.

First, we consider an experimental design where the distributional information is completely known, and we have a hypothetical best treatment arm. This situation corresponds to a case such that we have conjectures on the distribution from pilot experiments conducted before the experiments that we aim to design. We refer to this experiment as an experiment with complete information or the *globally optimal non-adaptive experimental design* (GO-NonAED).

Following this, we analyze an experimental design when the only variances are known, and we have a hypothetical best treatment arm. Other information such as means is unknown. In such instances, we discuss the worst-case scenario of an experiment when the difference between $\mu^{a_0^*}$ and μ^a converges zero, which allows us to characterize optimal experiments by the variances. We refer to this experiment as the *hypothetical locally optimal non-adaptive experimental design* (H-LO-NonAED). We also refer to the evaluation under $\mu^{a_0^*} - \mu^a \to 0$ as the *small-gap regime*.

Our subsequent question is an optimal experiment even when the variance is unknown although we still have a hypothetical best treatment arm. In this case, we employ adaptive experiments, in which we can gather information during an experiment and update the sampling rule in the experiment. We refer to this experiment as the *hypothetical locally optimal adaptive experimental design* (H-LO-AED). We still evaluate the performance under the small-gap regime.

Lastly, we address the situation where we do not have a hypothetical best treatment arm (or all treatment arms are hypothetical best treatment arms). We refer to this setting as the *locally optimal adaptive experimental design* (LO-AED).

In the following sections, we first discuss information theoretic lower bounds that derive the experimental designs in Section 4. The lower bounds differ according to our amount of knowledge. Based on the lower bounds and the information loss, we introduce the above four scenarios, the GO-NonADE, the H-LO-NonAED, the H-LO-ADE, and the LO-ADE, in Section 5. Then, we introduce asymptotically optimal non-adaptive experiments in Appendix C and adaptive experiments for the H-LO-ADE, and the LO-ADE in Section 6. In Appendix E, we show experimental results.

4. Information Theoretic Lower Bounds and Information Loss

We derive lower bounds for $\mathbb{P}_P(\widehat{a}_T \neq a_0^*)$ (upper bound of $-\frac{1}{T} \log \mathbb{P}_P(\widehat{a}_T \neq a_0^*)$) under large and small gaps ($\Delta_0^a \to 0$

for all $a \in [K]$). We call an experiment asymptotically optimal if the asymptotic upper bound matches the lower bound under a large gap. We also call an experiment asymptotically locally (wost-case) optimal if the asymptotic upper bound matches the lower bound under a small gap.

4.1. Information Theoretic Lower Bounds

To derive the lower bound, we first restrict our experiment to a consistent experiment, which is also considered in Kaufmann et al. (2016).

Definition 4.1 (Consistent experiment)**.** A consistent experiment π is an experiment such that for each $P \in \mathcal{P}^*$, if a_0^* is unique, then $\mathbb{P}_P(\widehat{a}_T^\pi = a_0^*) \to 1$ as $T \to \infty$.

In large deviation efficiency of hypothesis testing, a similar consistency is assumed (van der Vaart, 1998).

Let us define an average sample allocation ratio $\kappa_{T,P}^\pi : [K] \to (0,1)$ such that $\sum_{a\in[K]} \kappa_{T,P}^\pi(a) = 1$ under a statistical model $P \in \mathcal{P}$ and an experiment $\pi \in \Pi$ as $\kappa_{T,P}^\pi(a) = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_P\left[\mathbb{1}[A_t^\pi = a]\right]$. This quantity represents the average sample allocation to each treatment arm a under an experiment. Then, Kaufmann et al. (2016) presents the following lower bounds for the probability of misidentification $\mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*)$.

Proposition 4.2 (From Lemma 1 in Kaufmann et al. (2016))**.** *For each $P_0 \in \mathcal{P}$, any consistent (Definition 4.1) experiment π satisfies* $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \limsup_{T\to\infty} \inf_{Q\in\mathcal{P}:a^*(Q)\neq a_0^*} \sum_{a\in[K]} \kappa_{T,Q}^\pi(a)\mathrm{KL}(Q^a, P_0^a)$.

Note that upper bounds for $-\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*)$ corresponds to lower bounds for $\mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*)$. Also see Kaufmann et al. (2016) and Remark 4.8. We use this measure to evaluate the tail probability of misidentification.

When $K = 2$, the lower bound can be simplified (Also see Appendix D and Theorem 12 in Kaufmann et al. (2016)). However, to the best of our knowledge, for the general case with $K \geq 3$, the existence of lower bounds is an open issue. One of the difficulties comes from the problem that the term $\kappa_{T,Q}^\pi(a)$ does not correspond to sample allocation under P (Kaufmann, 2020). To derive lower bounds, we connect $\kappa_{T,Q}^\pi(a)$ to $\kappa_{T,P}^\pi(a)$ by restricting experiments.

4.2. Asymptotically Invariant Experiment

Thus, further restrictions on a class of experiments are required to derive lower bounds when $K \geq 3$. In this study, we consider restricting strategies to ones such that its limit $\lim_{T\to\infty} \kappa_{T,P}^\pi(a)$ is invariant across $P \in \mathcal{P}^*$. We refer to the limit $\lim_{T\to\infty} \kappa_{T,P}^\pi(a)$ as the *target allocation ratio*.

Definition 4.3 (Asymptotically invariant experiment)**.** An experiment π is called asymptotically invariant if there exists a function $w^\pi : [K] \to (0,1)$ such that for any $Q \in \mathcal{P}$, and all $a \in [K]$, as $T \to \infty$,

$$\kappa_{T,Q}^\pi(a) = w^\pi(a) + o(1). \tag{1}$$

We raise examples for consistent and asymptotically invariant experiments in Appendix A, which also implies their existence.

Let $\mathcal{W}$ be a set of all functions $w^\pi : [K] \to (0,1)$ such that $\sum_{a\in[K]} w(a) = 1$, defined as $\mathcal{W} = \left\{w : [K] \to (0,1) \middle| \sum_{a\in[K]} w(a) = 1\right\}$. Given a class of asymptotic invariant experiments, there exists $w \in \mathcal{W}$ such that for all $P \in \mathcal{P}$, and $a \in [K]$, it holds that $\left|w(a) - \frac{1}{T}\sum_{t=1}^T \mathbb{E}_Q\left[\mathbb{1}[A_t = a]\right]\right| \to 0$. Therefore, we obtain the following theorem.

Theorem 4.4. *For each $P_0 \in \mathcal{P}$, any consistent (Definition 4.1) and asymptotically invariant (Definition 4.3) experiment π satisfies*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \max_{w\in\mathcal{W}} \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^*: \\ \mu^*(Q)-\mu^a(Q)<0}} w(a)\mathrm{KL}(Q^a, P_0^a) + o(1).$$

To prove this statement, from Lemma 4.6 and (1) (Definition 4.3), there exists w^π and bound the probability as $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^*: \\ \mu^*(Q)-\mu^a(Q)<0}} \sum_{a\in[K]} w^\pi(a)\mathrm{KL}(Q^a, P_0^a) + o(1) \leq \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^*: \\ \mu^*(Q)-\mu^a(Q)<0}} \max_{w\in\mathcal{W}} w(a)\mathrm{KL}(Q^a, P_0^a)$.

Thus, we can link the average sample allocation ratio to an actual experiment because we can compute w^*, independent from Q. For the asymptotically invariant experiment, we can show that the optimal allocation strategies proposed by Glynn & Juneja (2004) are asymptotically optimal for the information-theoretic lower bounds with asymptotically invariant experiments if we can compute $\mathrm{KL}(Q^a, P^a)$.

However, in many applications, $\mathrm{KL}(Q^a, P^a)$ (or complete information) is unavailable. We discuss lower bounds under such information loss in the following sections.

We discuss the relationship woth static proportions of Degenne (2023) in Appendix B

4.3. Localization and Fisher Information

We introduced the lower bound by Kaufmann et al. (2016). However, the lower bound depends on the full-distributional information, which may require the information in an experiment. Assuming the full-knowledge before experiments is unrealistic in many applications. For such cases, we consider localization (Huber, 1964; Shin et al., 2018).

The following arguments are intuitive and not rigorous. For the details, see the following sections. Suppose that there

is a statistical model $Q_{\boldsymbol{\varepsilon}}$ parameterized by $\boldsymbol{\varepsilon} = (\varepsilon^a)_{a\in[K]}$ such that $\mu^{a^*(Q_{\boldsymbol{\varepsilon}})}(Q_{\boldsymbol{\varepsilon}}) - \mu^a(Q_{\boldsymbol{\varepsilon}}) = \mu_0^* - \mu_0^a + \varepsilon^a$ and $P_0 = Q_{\mathbf{0}}$. Then, we consider lower bounds when $\varepsilon^a \to 0$ and $\Delta^a(P) \to 0$ for all $a \in [K]$ by expanding the KL divergence between P_0 and $Q_{\boldsymbol{\varepsilon}}$ around $\varepsilon^a = 0$. Because the second-order approximation of the KL divergence is the Fisher information of statistical models, and the inverse of the Fisher information corresponds to the variances of potential outcomes, we can characterize the lower bounds by the variances. We refer to the lower bounds as localized lower bounds under the small-gap regime, roughly given as $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \max_{w\in\mathcal{W}} \sum_{a\in[K]} w(a) 2I_0^a \left(\Delta_0^a\right)^2 + o\left(\left(\Delta_0^a\right)^2\right)$ as $\Delta_0^a \to 0$ for all $a \in \{a_0^*\}$, where I_0^a is some Fisher information for Δ_a under P_0. This argument is not rigorous because the definition of the Fisher information is unclear. In particular, when the distributions are nonparametric, they may not be unique. To deal with this problem, we consider the semiparametric analysis and derive the semiparametric efficiency bound, a lower bound.

A model is referred to as semiparametric if its distribution is characterized by finite-dimensional parameters of interest (gaps in expected outcomes) and other other (finite or infinite-dimensional) parameters. Different from parametric models, the expected value of a squared second-order approximated log-likelihood, which may not be unique in semiparametric models, equates to the Fisher information in parametric models. We then evaluate models where the information, measured as the squared second derivative of the log-likelihood on the gaps, is minimized (equivalently, variance is maximized). These models are called *least-favorable models*. In our study, we use the variance under the least-favorable models to derive lower bounds for the misidentification probability under the alternative.

Similar localization has been employed in existing studies, such as Dong & Zhu (2016) and Shin et al. (2018). However, their localization considers parametric models. Our result is a generalization of their approaches.

4.4. Local Location-Shift Models

To conduct localized analysis under the small-gap regime, we define a distribution at the limit of $\Delta_0^a \to 0$. We consider the following location-shift statistical models.

Definition 4.5 (Local location-shift models)**.** Statistical models $\mathcal{P}^*$ are local location-shift statistical models if the following conditions are satisfied:
(i) Absolute continuity. For all $P, Q \in \mathcal{P}^*$ and all $a \in [K]$, the distributions P^a and Q^a are mutually absolutely continuous and have density functions with respect to the Lebesgue measure.
(ii) Lipschitz continuity. For all $a \in [K]$, there exists a (unknown) constant $C > 0$ independent of T such that for all $P, P' \in \mathcal{P}^*$, $\left|\left(\sigma^a(P)\right)^2 - \left(\sigma^a(P')\right)^2\right| < C\left|\mu^a(P) - \mu^a(P')\right|$.
(iii) Boundedness of the moments. There exists a (unknown) constant $C > 0$ independent of T such that for all $P \in \mathcal{P}^*$, $a \in [K]$, and $\gamma \in \mathbb{N}$, $\mathbb{E}_P\left[|Y^a|^\gamma\right] < C$.
(iv) Lower bounds of variances. There exists a *known* constant $C_\sigma > 0$ such that for all $P \in \mathcal{P}^*$ and $a \in [K]$, $\left(\sigma^a(P)\right)^2 > C_\sigma$.
(v) Uniqueness of the best treatment arm. For all $P \in \mathcal{P}^*$, there exists an (unknown) unique best treatment arm $a^*(P)$; that is, there exits $a^*(P)$ such that $\mu^{a^*(P)}(P) > \max_{a\in[K]\setminus\{a^*(P)\}} \mu^a(P)$.

We refer to this class of statistical models as local location-shift bandit models. Our lower bounds are characterized by σ_0^a, a variance under the "true" statistical model $P_0 \in \mathcal{P}^*$.

We raise two examples for this class. The first class is a class of Gaussian models defined in Definition C.1. Another class is a class of Bernoulli bandits, which are statistical models whose potential outcomes follow Bernoulli distributions.

4.5. Localized Lower Bounds

Then, we show the following lower bounds. The proof is shown in Appendix J.

Lemma 4.6 (General lower bound for the local location-shift models)**.** *For any $P \in \mathcal{P}^*$ (Definition 4.5), any consistent (Definition 4.1) experiment π satisfies*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^*:\\ \mu^*(Q)-\mu^a(Q)<0}} \limsup_{T\to\infty} \frac{\left(\Delta_0^a\right)^2}{2\Omega_0^a(\kappa_{T,Q}^\pi)} + o\left(\left(\Delta_0^a\right)^2\right) \tag{2}$$

as $\Delta_0^a \to 0$ for all $a \in [K]\setminus\{a_0^*\}$, and $\Omega_0^a(\kappa_{T,Q}^\pi) = \frac{(\sigma_0^*)^2}{\kappa_{T,Q}^\pi(a_0^*)} + \frac{(\sigma_0^a)^2}{\kappa_{T,Q}^\pi(a)}$.

For $\overline{\Delta}_0 = \max_{a\in[K]\setminus\{a_0^*\}} \Delta_0^a$, the RHS in (2) is lower bounded as $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^*:\\ \mu^*(Q)-\mu^a(Q)<0}} \limsup_{T\to\infty} \frac{\left(\overline{\Delta}_0\right)^2}{2\Omega_0^a(\kappa_{T,Q}^\pi)} + o\left(\left(\overline{\Delta}_0\right)^2\right)$. Then, by restricting experiments to asymptotically invariant ones and maximizing the target allocation ratio, we can further lower bound the above lower bound as $\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \min_{a\in[K]\setminus\{a_0^*\}} \max_{w\in\mathcal{W}} \frac{\left(\overline{\Delta}_0\right)^2}{2\Omega_0^a(w)} + o\left(\left(\overline{\Delta}_0\right)^2\right)$. Note that by definition, $\min_{a\in[K]\setminus\{a_0^*\}} \max_{w\in\mathcal{W}} \frac{\left(\overline{\Delta}_0\right)^2}{2\Omega_0^a(w)} = \max_{w\in\mathcal{W}} \min_{a\in[K]\setminus\{a_0^*\}} \frac{\left(\overline{\Delta}_0\right)^2}{2\Omega_0^a(w)}$. Let $w^* \in$

$\arg\max_{w\in\mathcal{W}} \min_{a\in[K]\setminus\{a_0^*\}} \frac{1}{2\Omega_0^a(w)}$ be the target allocation ratio. By solving the maximization, under the small gap regime, the probability of misidentification is lower bounded as follows.

Theorem 4.7 (Localized lower bounds)**.** *For any $P_0 \in \mathcal{P}^*$, any consistent (Definition 4.1) and asymptotically invariant (Definition 4.3) experiment π satisfies*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \frac{\left(\overline{\Delta}_0\right)^2}{2\left(\sigma_0^* + \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}}\left(\sigma_0^a\right)^2}\right)^2} + o\left(\left(\overline{\Delta}_0\right)^2\right)$$

as $\overline{\Delta}_0 \to 0$, where $\overline{\Delta}_0 = \max_{a\in[K]\setminus\{a_0^*\}} \Delta_0^a$, and the target allocation ratio is given as

$$w^*(a_0^*) = \frac{\sigma_0^*}{\sigma_0^* + \sqrt{\sum_{b\in[K]\setminus\{a_0^*\}}\left(\sigma_0^b\right)^2}}, \tag{3}$$

$$w^*(a) = (1-w^*(a_0^*))\frac{\left(\sigma_0^a\right)^2}{\sum_{b\in[K]\setminus\{a_0^*\}}\left(\sigma_0^b\right)^2},\ \ \forall a\in[K]\setminus\{a_0^*\}.$$

Note that as shown in Section D, we can derive a lower bound that only holds for $K = 2$ without using additional restrictions on experiments.

Remark 4.8 (Complexity of Experiments and Bahadur Efficiency). The complexity $-\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*)$, has been widely adopted in the literature of ordinal optimization and BAI (Glynn & Juneja, 2004; Kaufmann et al., 2016). In the filed of hypothesis testing, Bahadur (1960) suggests a similar measure to asses a power of a test. The efficiency of a test under the criterion proposed by Bahadur (1960) is known as Bahadur efficiency. Although our problem is not hypothesis testing, it can be considered that our global asymptotic optimality parallels the global Bahadur efficiency, and our asymptotic optimality under the small-gap regime is analogous to the local Bahadur efficiency (Bahadur, 1960; Wieand, 1976; Akritas & Kourouklis, 1988). From a technical perspective, such localization has been utilized in evaluating various statistical procedures, such as estimation and hypothesis testing, since it enables us to approximate a broad range of distributions using Gaussian ones (Huber, 1964).

4.6. Target Allocation Ratio Independent from the True Best Treatment Arm

In the target allocation in (3), we use a_0^*, which is the true best treatment arm. Therefore, to design an optimal experiment, we need to know a_0^*, and under target sample allocation using a_0^*, we can show that the probability of misidentification is asymptotically optimal. This property requires us to introduce a proxy of a_0^*, and we refer to it as a hypothetical treatment arm, denoted by $\widetilde{a}$. We design experiments using $\widetilde{a}$, and the experiment is asymptotically optimal when $\widetilde{a} = a_0^*$.

While there are various applications where we can set such a hypothetical treatment arm, there are still many situations where we cannot specify it, as well as the standard setup of BAI. Therefore, we consider lower bounds under which the target allocation ratio is independent from a_0^*.

Consider experiments where an experimenter cannot specify a unique hypothetical best treatment arm. If an experimenter has null hypotheses that both of $\widetilde{a} \in [K]$ and $\widetilde{b} \in [K]$ ($\widetilde{a} \neq \widetilde{b}$) are *not* best, we consider minimize the probability of misidentification when the two null $H_0^{\widetilde{a}} : a_0^* \neq \widetilde{a}$ and $H_0^{\widetilde{b}} : a_0^* \neq \widetilde{a}$ are not true; that is, under P such that $a^*(P) = \widetilde{a}$ or $a^*(P) = \widetilde{b}$, we minimize $\mathbb{P}_P(\widehat{a}_T \neq a_0^*)$.

We consider restricted bandit models $\mathcal{P}^\dagger \subset \mathcal{P}^*$ such that there is an unique set of constants $\left\{\left(\sigma^a\right)^2\right\}_{a\in[K]}$ such that for all $P \in \mathcal{P}^\dagger$ and $a, b \in [K]$, it holds that $\left(\sigma^a\right)^2, \left(\sigma^b\right)^2 > C_\sigma$, and $\left(\sigma^a(P)\right)^2 \to \left(\sigma^a\right)^2$ and $\left(\sigma^b(P)\right)^2 \to \left(\sigma^b\right)^2$ as $\mu^a(P) - \mu^b(P) \to 0$. Then, by using Lemma 4.6, we can obtain the following theorem. For these models with multiple hypothetical treatment arms, we consider the following localized lower bounds for the choice of $\widetilde{a}, \widetilde{b}$: $\sup_{P\in\mathcal{P}^*:a^*(P)\in\{\widetilde{a},\widetilde{b}\}} \lim_{\overline{\Delta}(P)\to 0} \limsup_{T\to\infty} -\frac{1}{T\overline{\Delta}^2(P)}\log \mathbb{P}_P(\widehat{a}_T^\pi \neq a^*(P))$, where $\overline{\Delta}(P) = \max_{a\in[K]\setminus\{a^*(P)\}} \Delta^a(P)$.

Furthermore, by considering worst-cases for all possible hypothetical best treatment arms, we define the complexity as $\sup_{P\in\mathcal{P}^*} \lim_{\overline{\Delta}(P)\to 0} \limsup_{T\to\infty} -\frac{1}{T\overline{\Delta}^2(P)}\log \mathbb{P}_P(\widehat{a}_T^\pi \neq a^*(P))$, where $\overline{\Delta}(P) = \max_{a\in[K]\setminus\{a^*(P)\}} \Delta^a(P)$.

Theorem 4.9 (Localized lower bounds)**.** *When $K = 2$, for any $P \in \mathcal{P}^\dagger$, any consistent (Definition 4.1) and asymptotically invariant (Definition 4.3) experiment $\pi \in \Pi$ satisfies* $\sup_{P\in\mathcal{P}^\dagger} \lim_{\mu^1(P)-\mu^2(P)\to 0} \limsup_{T\to\infty} -\frac{1}{T(\mu^1(P)-\mu^2(P))^2}\log \mathbb{P}_P(\widehat{a}_T^\pi \neq a^*(P)) \leq \frac{1}{2(\sigma^1+\sigma^2)^2} + o(1)$, *and the target allocation ratio is given as $w^*(1) = \frac{\sigma_0^1}{\sigma_0^1+\sigma_0^2}$ and $w^*(2) = 1 - w^*(1)$. When $K \geq 3$, for any $P \in \mathcal{P}^\dagger$, any consistent (Definition 4.1) and asymptotically invariant (Definition 4.3) experiment $\pi \in \Pi$,*

$$\sup_{P\in\mathcal{P}^\dagger} \lim_{\overline{\Delta}(P)\to 0} \limsup_{T\to\infty} -\frac{1}{T\overline{\Delta}^2(P)}\log \mathbb{P}_P(\widehat{a}_T^\pi \neq a^*(P)) \leq \max_{w\in\mathcal{W}} \min_{b\in[K],a\in[K]\setminus\{b\}} \frac{1}{2\Omega^{b,a}(w)} + o(1),$$

where $\Omega^{b,a}(w) = \frac{\left(\sigma^b\right)^2}{w(b)} + \frac{\left(\sigma^a\right)^2}{w(a)}$. Here, the target allocation ratio is given as

$$w^*(a) = \arg\max_{w\in\mathcal{W}} \min_{b\in[K],a\in[K]\setminus\{b\}} \frac{1}{2\Omega^{b,a}(w)}. \tag{4}$$

The target allocation ratio is independent from a_0^*. The metric $\sup_{P\in\mathcal{P}^*}\lim_{\overline{\Delta}(P)\to 0}\limsup_{T\to\infty} -\frac{1}{T\overline{\Delta}^2(P)}\log \mathbb{P}_P(\widehat{a}_T^{\pi} \neq a^*(P))$ captures two worst-cases: one of the worst-cases is about $P \in \mathcal{P}^*$ and another is about the small gap.

5. Categories of Experimental Designs based on a Information Loss

In this section, we categorize experiments based on the perspective of information loss, building on the arguments on lower bound, established in Section 4.

(A) GO-NonAED. When complete distributional information is available, we design experiments in which the probability of misidentification matches the lower bound in Theorem 4.4. This setting has been tackled by Glynn & Juneja (2004), so we directly utilize their experiment, which we refer to as the non-adaptive (NA)-EBA experiment. For more details, see Appendix C. In this instance, since the distributional information is known, we consider that the hypothetical treatment arm can also be deduced from the available information.

(B) H-LO-NonADE. Requiring complete information can often be costly or even unfeasible. Therefore, we consider an experimental design that works with limited information. In particular, we explore optimal experiments when only variances are known. In this situation, the optimal experimental design is characterized by a worst-case scenario with regard to other parameters (localization). We specifically use the lower bounds in Theorem 4.7 under the small-gap regime. We employ the NA-EBA experiment with a target allocation ratio that differs from that of GO-NonADE.

(C) H-LO-ADE. The H-LO-NonADE still requires knowledge of the variances. We consider a situation where even the variances are unknown but can be estimated during an experiment. In this section, contrary to standard BAI, we consider a setting where we hypothesize that a treatment arm $a \in [K]$ is the best, as well as the GO-NonADE and the H-LO-NonADE. This setting is referred to as H-LO-ADE. Lower bounds for the probability of misidentification are given as in H-LO-NonADE; that is, the ones in Theorem 4.7. However, we need to design an adaptive experiment that efficiently allocates treatment arms and identifies the best one. Therefore, our focus is on the issue of whether there exists an optimal experiment, whose upper bound matches the lower bounds. In Section 6, we introduce the TS-EBA experiment, which is inspired from Kato et al. (2023a).

(D) LO-ADE. Lastly, we consider the LO-ADE, where neither the hypothetical best treatment nor distributional information is given. Interpreting the absence of a hypothetical best treatment as a scenario where all treatment arms are potential best treatment arms, we lower bound $\sup_{P\in\mathcal{P}^*}\lim_{\overline{\Delta}(P)\to 0}\limsup_{T\to\infty} -\frac{1}{T\overline{\Delta}^2(P)}\log \mathbb{P}_P(\widehat{a}_T^{\pi} \neq a^*(P))$ using Theorem 4.9. Then, we use the TS-EBA experiment with the target allocation ratio different from the one for the H-LO-ADE.

Algorithm 1 TS-EBA experiment.

Parameter: Hypothetical best treatment arm $\widetilde{a} \in [K]$. The sample splitting ratio $r \in (0,1)$.
Initialization: for $t = 1$ **do** Draw $A_t = t$. For each $a \in [K]$, set $\widehat{w}_t(a) = 1/K$. **end for**
Stage 1:
for $t = K+1$ to $\lceil rT \rceil$ **do**
  Draw a treatment arm a with probability $w^{(1)}$.
end for
Estimating the variances, for the H-LO-ADE, construct $\widehat{w}^{(2)}$ as (6); for the LO-ADE, construct $\widehat{w}^{(2)}$ as 8.
Stage 2:
for $t = \lceil rT \rceil + 1$ to T **do**
  $A_t = 1$ if $\gamma_t \leq \widehat{w}^{(2)}(1)$ and $A_t = a$ for $a \geq 2$ if $\gamma_t \in \left(\sum_{b=1}^{a-1}\widehat{w}^{(2)}(b), \sum_{b=1}^{a}\widehat{w}^{(2)}(b)\right]$.
end for
Construct $\widehat{\mu}_T^{\mathrm{SA},a}$ for each $a \in [K]$.
Recommend $\widehat{a}_T^{\mathrm{EBA}} = \arg\max_{a\in[K]}\widehat{\mu}_T^{\mathrm{SA},a}$.

6. The TS-EBA Experiment

This section introduces our experiment, which is inspired by adaptive experiments using propensity score proposed by Hahn et al. (2011). Our experiment comprises the following sampling and recommendation rules: First, we divide the budget into two parts. In the first stage, we uniformly randomly draw a treatment arm. In the second stage, we draw treatment arms with the goal of achieving the target allocation ratio. After drawing treatment arms, we recommend the empirical best arm (EBA) using the sample average estimator. We refer to this experiment as the TS-EBA experiment.

6.1. The TS-EBA Experiment for the H-LO-ADE

First, we define a target allocation ratio, which is used to determine our sampling rule. Let $\sigma_0^{\widetilde{a}}$ be $\widetilde{\sigma}_0$. As shown in (3), we set the target allocation as

$$w^{\text{TS-EBA}}(\widetilde{a}) = \frac{\widetilde{\sigma}_0}{\widetilde{\sigma}_0 + \sqrt{\sum_{b\in[K]\setminus\{\widetilde{a}\}}\left(\sigma_0^b\right)^2}}, \tag{5}$$

$$w^{\text{TS-EBA}}(a) = \left(1 - w^{\text{TS-EBA}}(\widetilde{a})\right)\frac{\left(\sigma_0^a\right)^2}{\sum_{b\in[K]\setminus\{\widetilde{a}\}}\left(\sigma_0^b\right)^2} \quad \forall a \in [K]\setminus\{\widetilde{a}\}.$$

Here, we replaced the true best treatment arm a_0^* in (3) with the hypothetical best treatment arm $\widetilde{a}$. This because we do not know the true best treatment arm a_0^*, but our goal is to minimize the probability when the null hypothesis is not correct; that is, $a_0^* = \widetilde{a}$. When the variances are unknown, this target allocation ratio is also unknown. Therefore, we estimate it during the adaptive experiment and use the estimator to estimate w^*.

The TS-EBA experiment consists of the following two stage experiments. We first fix $r \in (0,1)$ independent from T[3] and $w^{(1)}(d)$ such that $w^{(1)}(d) > C$ and $\sum_{a\in[K]} w^{(1)}(a) = 1$, where C is independent from T. Let $t \in \{1, 2, \dots, \lceil rT \rceil\}$ be Stage 1 and $t \in \{\lceil rT \rceil + 1, \lceil rT \rceil + 2, \dots, T\}$ be Stage 2.

In Stage 1, after drawing each treatment arm $1, 2, \dots, K$ at each round $1, 2, \dots, K$, we draw treatment arm $A_t = a$ with probability $w^{(1)}(a)$ until $t = \lceil rT \rceil$. After Stage 1, we draw treatment arms with probability $w^{(2)}$, chosen to minimize empirical version of the lower bounds as follows:

$$\widehat{w}^{(2)}(\widetilde{a}) = \frac{\frac{\widehat{\widetilde{\sigma}}_{\lceil rT \rceil}}{\widehat{\widetilde{\sigma}}_{\lceil rT \rceil} + \sqrt{\sum_{b\in[K]\setminus\{a_0^*\}} \left(\widehat{\sigma}^b_{\lceil rT \rceil}\right)^2}} - r\widehat{w}^{(1)}(a_0^*)}{1-r}, \tag{6}$$

$$\widehat{w}^{(2)}(a) = \frac{\left(1 - \widehat{w}^{(2)}(\widetilde{a})\right) \frac{\left(\widehat{\sigma}^a_{\lceil rT \rceil}\right)^2}{\sum_{b\in[K]\setminus\{\widetilde{a}\}} \left(\widehat{\sigma}^b_{\lceil rT \rceil}\right)^2} - r\widehat{w}^{(1)}(a)}{1-r} \quad \forall a \in [K]\setminus\{\widetilde{a}\},$$

where $\widehat{\widetilde{\sigma}}_{\lceil rT \rceil}$ and $\widehat{\sigma}^a_{\lceil rT \rceil}$ are sample variances that are estimators of $\widetilde{\sigma}_0$ and σ_0^a using observations until $t = \lceil rT \rceil$, respectively.

After Stage 2 (after round T), for each $a \in [K]$, we estimate μ^a for each $a \in [K]$ and recommend the maximum. To estimate μ^a, we use the SA estimator. Finally, we recommend $\widehat{a}_T^{\mathrm{EBA}} = \arg\max_{a\in[K]} \widehat{\mu}_T^{\mathrm{SA},a}$ as the best treatment arm (estimator of a_0^*). We show the pseudo-code in Algorithm 1. In practice, instead of random sampling, we can allocate treatment arms by a way of a round-robin (Appendix N.2).

6.2. The TS-EBA Experiment for the LO-ADE

In the LO-ADE, the target sample allocation ratio is

$$w^{\text{TS-EBA}}(a) = \arg\max_{w\in\mathcal{W}} \min_{b\in[K],a\in[K]\setminus\{b\}} \frac{1}{2\Omega^{b,a}(w)}, \tag{7}$$

which is identical to that in (4). This is because it is independent from $\widetilde{a}$.

Therefore, in the second stage, we compute the target allocation ratio as

$$\widehat{w}^{(2)} = \arg\max_{w\in\mathcal{W}} \min_{b\in[K],a\in[K]\setminus\{b\}} \frac{1}{2\widehat{\Omega}^{b,a}(w)}, \tag{8}$$

where $\widehat{\Omega}^{b,a}(w) = \frac{\left(\widehat{\sigma}^b_{\lceil rT \rceil}\right)^2}{w(b)} + \frac{\left(\widehat{\sigma}^a_{\lceil rT \rceil}\right)^2}{w(a)}$. We replace $\widehat{w}^{(2)}$ in the previous section with this. Then, we conduct the same experiment.

6.3. Probability of Misidentification and Asymptotic Optimality of the TS-EBA Experiment

Next, we derive upper bounds for the probability of misidentification under the TS-EBA experiment.

Proposition 6.1 (Upper bound of the TS-EBA experiment)**.** *For each $P_0 \in \mathcal{P}^*$, $a \in [K]\setminus\{a_0^*\}$, and any $\varepsilon > 0$, there exists $T_0 > 0$ such that for all $T > T_0$, $\mathbb{P}_{P_0}\left(\widehat{\mu}_T^{\mathrm{SA},a_0^*} \le \widehat{\mu}_T^{\mathrm{SA},a}\right) \le \exp\left(- \frac{T(\Delta_0^a)^2}{2\Omega_0^a(w^{\text{TS-EBA}})} + \left\{ \frac{\sqrt{T}\Delta_0^a}{\sqrt{\Omega_0^a(w^{\text{TS-EBA}})}} + \frac{T(\Delta_0^a)^2}{2\Omega_0^a(w^{\text{TS-EBA}})} \right\} \varepsilon \right)$.*

For the proof, see Appendix N.1. This proposition immediately yields the following theorem.

Theorem 6.2 (Asymptotic optimality of the TS-EBA experiment)**.** *For each $P_0 \in \mathcal{P}^*$,*

$$\liminf_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}\left(\widehat{a}_T^{\mathrm{EBA}} \neq a_0^*\right) \ge \min_{a\neq a_0^*} \frac{(\underline{\Delta}_0)^2}{2\Omega_0^a(w^{\text{TS-EBA}})} - o\left((\underline{\Delta}_0)^2\right)$$

as $\underline{\Delta}_0 \to 0$, where $\underline{\Delta}_0 = \min_{a\in[K]\setminus\{a_0^*\}} \Delta_0^a$.

Recall that a_0^* is unique; that is, $\mu_0^* - \mu_0^a > 0$ for all $a \in [K]\setminus\{a_0^*\}$.

In the H-LO-ADE, when using (5) as the target allocation ratio, the probability of misidentification matches the lower bound in Theorem 4.7. In the LO-ADE, when using (7) as the target allocation ratio, the probability of misidentification matches the lower bound in Theorem 4.9.

7. Conclusion

We investigate the problem of experimental design. We found that depending on available information, different experiments become asymptotically optimal. Based on our findings, we categorized experimental design into the GO-NonADE, the H-LO-NonADE, the H-LO-ADE, and the LO-NonADE. For the H-LO-ADE and the LO-NonADE, we proposed an asymptotically optimal adaptive experimental design. We confirmed the soundness of the proposed methods via simulation studies.

[3] Although r is assumed to be independent from T, we use r such that $\lceil rT \rceil > K + 1$ in the following sections.

References

Akritas, M. G. and Kourouklis, S. Local bahadur efficiency of score tests. *Journal of Statistical Planning and Inference*, 19(2):187–199, 1988. ISSN 0378-3758.

Ariu, K., Kato, M., Komiyama, J., McAlinn, K., and Qin, C. Policy choice and best arm identification: Asymptotic analysis of exploration sampling, 2021.

Atsidakou, A., Katariya, S., Sanghavi, S., and Kveton, B. Bayesian fixed-budget best-arm identification, 2023.

Audibert, J.-Y., Bubeck, S., and Munos, R. Best arm identification in multi-armed bandits. In *Conference on Learning Theory*, pp. 41–53, 2010.

Bahadur, R. R. Stochastic Comparison of Tests. *The Annals of Mathematical Statistics*, 31(2):276 – 295, 1960.

Bechhofer, R., Kiefer, J., and Sobel, M. *Sequential Identification and Ranking Procedures: With Special Reference to Koopman-Darmois Populations*. University of Chicago Press, 1968.

Bechhofer, R. E. A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with known Variances. *The Annals of Mathematical Statistics*, 25(1):16 – 39, 1954.

Branke, J., Chick, S. E., and Schmidt, C. Selecting a selection procedure. *Management Science*, 53(12):1916–1932, 2007.

Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 2011.

Carpentier, A. and Locatelli, A. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *COLT*, 2016.

CDER, C. Adaptive designs for clinical trials of drugs and biologics guidance for industry draft guidance, 05 2018.

Chen, C.-H., Lin, J., Yücesan, E., and Chick, S. E. Simulation budget allocation for further enhancing theefficiency of ordinal optimization. *Discrete Event Dynamic Systems*, 10(3):251–270, 2000.

Chernoff, H. Sequential Design of Experiments. *The Annals of Mathematical Statistics*, 30(3):755 – 770, 1959.

Chow, S.-C. and Chang, M. *Adaptive Design Methods in Clinical Trials*. Chapman and Hall/CRC, 2 edition, 2011.

Cramér, H. Sur un nouveau théorème-limite de la théorie des probabilités. In *Colloque consacré à la théorie des probabilités*, volume 736, pp. 2–23. Hermann, 1938.

Degenne, R. On the existence of a complexity in fixed budget bandit identification, 2023.

Dembo, A. and Zeitouni, O. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2009.

Dong, J. and Zhu, Y. Three asymptotic regimes for ranking and selection with general sample distributions. In *2016 Winter Simulation Conference (WSC)*, pp. 277–288, 2016.

Ellis, R. S. Large Deviations for a General Class of Random Vectors. *The Annals of Probability*, 12(1):1 – 12, 1984.

Even-Dar, E., Mannor, S., Mansour, Y., and Mahadevan, S. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 2006.

Fisher, R. A. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.

Garivier, A. and Kaufmann, E. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, 2016.

Glynn, P. and Juneja, S. A large deviations perspective on ordinal optimization. In *Proceedings of the 2004 Winter Simulation Conference*, volume 1. IEEE, 2004.

Gupta, S. *On a Decision Rule for a Problem in Ranking Means*. University of North Carolina at Chapel Hill, 1956.

Gärtner, J. On large deviations from the invariant measure. *Theory of Probability & Its Applications*, 22(1):24–39, 1977.

Hahn, J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331, 1998.

Hahn, J., Hirano, K., and Karlan, D. Adaptive experimental design using the propensity score. *Journal of Business and Economic Statistics*, 2011.

Hansen, B. E. A modern gauss–markov theorem. *Econometrica*, 90(3):1283–1294, 2022.

Hayashi, F. *Econometrics*. Princeton Univ. Press, Princeton, NJ [u.a.], 2000.

Ho, Y., Sreenivas, R., and Vakili, P. Ordinal optimization of deds. *Discrete Event Dynamic Systems: Theory and Applications*, 2(1):61–88, July 1992.

Hong, L. J., Fan, W., and Luo, J. Review on ranking and selection: A new perspective. *Frontiers of Engineering Management*, 8(3):321–343, mar 2021.

Huber, P. J. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964.

Kasy, M. and Sautmann, A. Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1): 113–132, 2021.

Kato, M., Ishihara, T., Honda, J., and Narita, Y. Adaptive experimental design for efficient treatment effect estimation: Randomized allocation via contextual bandit algorithm, 2020.

Kato, M., Imaizumi, M., Ishihara, T., and Kitagawa, T. Semiparametric best arm identification with contextual information, 2022.

Kato, M., Imaizumi, M., Ishihara, T., and Kitagawa, T. Asymptotically minimax optimal fixed-budget best arm identification for expected simple regret minimization, 2023a.

Kato, M., Imaizumi, M., Ishihara, T., and Kitagawa, T. Best arm identification with contextual information under a small gap, 2023b.

Kaufmann, E. *Contributions to the Optimal Solution of Several Bandits Problems*. Habilitation á Diriger des Recherches, Université de Lille, 2020.

Kaufmann, E., Cappé, O., and Garivier, A. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1): 1–42, 2016.

Komiyama, J., Tsuchiya, T., and Honda, J. Minimax optimal algorithms for fixed-budget best arm identification. In *Advances in Neural Information Processing Systems*, 2022.

Lai, T. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 1985.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

Le Cam, L. Limits of experiments. In *Theory of Statistics*, pp. 245–282. University of California Press, 1972.

Manski, C. Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice. *Journal of Econometrics*, 95(2):415–442, 2000.

Manski, C. F. Treatment choice under ambiguity induced by inferential problems. *Journal of Statistical Planning and Inference*, 105(1):67–82, 2002.

Manski, C. F. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246, 2004.

Murphy, S. A. and van der Vaart, A. W. Semiparametric likelihood ratio inference. *The Annals of Statistics*, 25(4): 1471 – 1509, 1997.

Nair, B. Clinical trial designs. *Indian Dermatol. Online J.*, 10(2):193–201, March 2019.

Neyman, J. Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. *Statistical Science*, 5:463–472, 1923.

Paulson, E. A Sequential Procedure for Selecting the Population with the Largest Mean from k Normal Populations. *The Annals of Mathematical Statistics*, 1964.

Peirce, C. S. and de Waal, C. *Illustrations of the Logic of Science*. Chicago, Illinois: Open Court, 1887.

Peirce, C. S. and Jastrow, J. On small differences in sensation. *Memoirs of the National Academy of Sciences*, 3: 75–83, 1884.

Pong, A. and Chow, S.-C. *Handbook of adaptive designs in pharmaceutical and clinical development*. Chapman and Hall/CRC, 04 2016.

Qin, C. and Russo, D. Adaptivity and confounding in multi-armed bandit experiments, 2022.

Qin, C., Klabjan, D., and Russo, D. Improving the expected improvement algorithm. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Robbins, H. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 1952.

Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 1974.

Russo, D. Simple bayesian algorithms for best arm identification, 2016.

Sanov, I. N. On the probability of large deviations of random variables, 1958.

Shang, X., de Heide, R., Menard, P., Kaufmann, E., and Valko, M. Fixed-confidence guarantees for bayesian best-arm identification. In *International Conference on Artificial Intelligence and Statistics*, volume 108, pp. 1823–1832, 2020.

Shin, D., Broadie, M., and Zeevi, A. Tractable sampling strategies for ordinal optimization. *Operations Research*, 66(6):1693–1712, 2018.

Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 1933.

Tsiatis, A. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer New York, 2007.

van der Laan, M. J. The construction and analysis of adaptive group sequential designs, 2008.

van der Vaart, A. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

Wald, A. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

Wald, A. Statistical Decision Functions. *The Annals of Mathematical Statistics*, 20(2):165 – 205, 1949.

Wieand, H. S. A Condition Under Which the Pitman and Bahadur Approaches to Efficiency Coincide. *The Annals of Statistics*, 4(5):1003 – 1011, 1976.

A. Examples of Consistent and asymptotically invariant strategies

Example (Consistent and asymptotically invariant strategies). One of the instances of consistent strategies involves sampling all treatment arms with probabilities that are bounded below by a strictly positive value that is independent of T. Then, we proceed to consider strategies that satisfy the definition of asymptotically invariant strategies. The first example is uniform sampling, which allocates treatment arms with an equal allocation ratio, i.e., $w^\pi(1) = \cdots = w(K) = 1/K$. Additionally, we can consider w^π that depends solely on the fixed variance, fixed standard deviation, or fixed treatment arm, independent of P. Consider a set of statistical models $\mathcal{P}^\dagger \subset \mathcal{P}$ whose variances σ_0^a are the same among $P \in \mathcal{P}^\dagger$; that is, for all $P, Q \in \mathcal{P}^\dagger$ and $a \in [K]$, $\sigma^a(P) = \sigma_0^a(Q) = \sigma_0^a$. Then, for example, $w^\pi(a) = \frac{\sigma_0^a}{\sum_{b\in[K]}\sigma_0^b}$ and $w^\pi(a) = \frac{(\sigma_0^a)^2}{\sum_{b\in[K]}(\sigma_0^b)^2}$ for all $a \in [K]$ are members of the asymptotically invariant strategies because σ_0^a is fixed among $P \in \mathcal{P}^\dagger$. If we fix some $b \in [K]$, independent of $P \in \mathcal{P}^\dagger$, then w^π that depends on b also belongs to the asymptotically invariant strategies, such as $w^\pi(b) = 1/2$ and $w^\pi(a) = 1/2(K-1)$ for all $a \in [K]\backslash b$. We can also construct w^π that depends on a_0^* for all $P \in \mathcal{P}^\dagger$, such as $w^\pi(a_0^*) = 1/2$ and $w^\pi(a) = 1/2(K-1)$ for all $a \in [K]\backslash a_0^*$.

B. Static Proportions

Remark B.1 (Static proportions). While our previous version of this study at arXiv, Kato et al. (2023b), discusses asymptotically invariant experiments[4], Degenne (2023) also independently proposes static proportions, which is a almost same notion as our asymptotically invariant experiments. When they upload the draft with the finding that the experiment of Glynn & Juneja (2004) is asymptotically optimal under that class of experiments, we were in the process of including the same finding. However, the final conclusions between ours and Degenne (2023) are significantly different. While we show the existence of asymptotically optimal experiments, they show the non-existence of them. These results are not contradict because while we consider the small-gap regime, Degenne (2023) considers the fixed-gap regime (the gap is constant independent from T) as well as Carpentier & Locatelli (2016). Additionally, we further analyze the properties of the class from the viewpoint of hypothetical best treatment arm and derive the analytical solutions of the target sample allocation ratio.

C. The NA-EBA Experiment

In this section, based on the arguments in Section 4, we design asymptotically optimal non-adaptive experiments for the GO-NonADE and the H-LO-NonADE.

C.1. The NA-EBA Experiment for the GO-NonAED

We define an experiment, which consists of the non-adaptive (NA) sampling rule and the EBA recommendation rule. We refer to our experiment as the NA-EBA experiment for the GO-NonAED. The experiment for the GO-NonAED is non-adaptive; that is, we do not update the sampling rule during an experiment. Consider the following procedure of an experiment. At the beginning of an experiment, we compute the target allocation ratio as

$$w^{\text{NA-EBA}} = \arg\max_{w\in\mathcal{W}} \min_{Q\in\mathcal{P}:a^*(Q)\neq a_0^*} \sum_{a\in[K]} w(a)\mathrm{KL}(Q^a, P^a).$$

Then, we allocate samples $t \in \{1, 2, \ldots, \lfloor Tw^{\text{NA-EBA}}(1)\rfloor\}$ to arm 1, and for $a \geq 2$, samples $t \in \{\lfloor T\sum_{b=1}^{a-1} w^{\text{NA-EBA}}(b)\rfloor + 1, \lfloor T\sum_{b=1}^{a-1} w^{\text{NA-EBA}}(b)\rfloor + 2, \ldots, \lfloor T\sum_{b=1}^{a} w^{\text{NA-EBA}}(b)\rfloor\}$ to each treatment arm a to arm a. At the end of an experiment, we estimate the expected rewards and recommend a treatment arm with the highest expected reward as the best treatment arm. To estimate μ^a, we use the following Sample Average (SA) estimator: $\widehat{\mu}_T^{\text{SA},a} = \frac{1}{\sum_{s=1}^T \mathbb{1}[A_s=a]}\sum_{s=1}^T \mathbb{1}[A_s = a]Y_s^a$. Then, we recommend the following empirical best treatment arm:

$$\widehat{a}_T^{\text{EBA}} = \arg\max_{a\in[K]} \widehat{\mu}_T^{\text{SA},a}. \tag{9}$$

Computation of the Target Allocation Ratio. Based on the result in Glynn & Juneja (2004), they show that the target

[4]The initial version of Kato et al. (2023b) uploaded in Sept. 2022 lacks the condition of the asymptotically invariant experiments (Kato et al., 2022). After the initial upload, we found that the necessity of the condition and revised the draft in Jan. 2023.

allocation ratio $w^{\text{NA-EBA}}$ that minimizes the probability of misidentification satisfies

$$\sum_{a\in[K]\setminus\{a_0^*\}} \frac{\partial G^a(w^{\text{NA-EBA}}(a_0^*), w^{\text{NA-EBA}}(a))/\partial w^{\text{NA-EBA}}(a_0^*)}{\partial G^a(w^{\text{NA-EBA}}(a_0^*), w^{\text{NA-EBA}}(a))/\partial w^{\text{NA-EBA}}(a)} = 1, \tag{10}$$

$$G^a(w^{\text{NA-EBA}}(a_0^*), w^*(a)) = G^b(w^{\text{NA-EBA}}(a_0^*), w^*(a)) \qquad \forall (a,b) \in [K]\setminus\{a_0^*\}, \tag{11}$$

where $G^a(w^{\text{NA-EBA}}(a_0^*), w^{\text{NA-EBA}}(a)) = \inf_{z\in\mathbb{R}} \left\{ w^{\text{NA-EBA}}(a_0^*)\mathcal{I}^{a_0^*}(z) + w^{\text{NA-EBA}}(a)\mathcal{I}^a(z) \right\}$, and $\mathcal{I}^a(z)$ denote the Fenchel-Legendre transform of the log-moment generating function $\Lambda^a(\theta) = \log \mathbb{E}[\exp(\theta Y^a)]$ defined as $\mathcal{I}^a(z) = \sup_{\theta\in\mathbb{R}}\{\theta z - \Lambda^a(\theta)\}$.

Gaussian models. As an example, we consider Gaussian statistical models defined as follows.

Definition C.1. Statistical models $\mathcal{P}^{\text{G}} \subset \mathcal{P}$ are Gaussian statistical models if for any $P \in \mathcal{P}$, Y^a is generated from $\mathcal{N}\left(\mu^a(P), (\sigma^a)^2\right)$ for all $a \in [K]$, where $\mu^a(P) \in \mathbb{R}$ and $(\sigma^a)^2 > C_\sigma$ are constants independent from T, and $\mathcal{N}(\mu^a(P), (\sigma^a)^2)$ is a Gaussian distribution with a mean $\mu^a(P)$ and a variance $(\sigma^a)^2$ (variance is fixed for any $P \in \mathcal{P}^{\text{G}}$).

Let $\sigma^{a_0^*} = \sigma^*$. For Gaussian models, Glynn & Juneja (2004) and Chen et al. (2000) show that

$$G^a(w(a_0^*), w(a)) = \frac{(\mu_0^a - \mu_0^*)^2}{2\left((\sigma^*)^2/w(a_0^*) + (\sigma^a)^2/w(a)\right)}.$$

Then, they derive that the target allocation ratio as w^*, satisfying

$$w^{\text{NA-EBA}}(a_0^*) = \sigma^* \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} (w^{\text{NA-EBA}}(a)/\sigma^a)^2},$$

$$\frac{(\mu_0^a - \mu_0^*)^2}{\frac{(\sigma^*)^2}{w^{\text{NA-EBA}}(a_0^*)} + \frac{(\sigma^a)^2}{w^{\text{NA-EBA}}(a)}} = \frac{(\mu_0^b - \mu_0^*)^2}{\frac{(\sigma^*)^2}{w^{\text{NA-EBA}}(a_0^*)} + \frac{(\sigma^b)^2}{w^{\text{NA-EBA}}(b)}} \qquad \forall (a,b) \in [K]\setminus\{a_0^*\},$$

$$\sum_{a\in[K]\setminus\{a_0^*\}} \frac{(\sigma^*)^2/w^{\text{NA-EBA}}(a_0^*)}{(\sigma^a)^2/w^{\text{NA-EBA}}(a)} = 1.$$

C.2. The NA-EBA Experiment for the H-LO-NonAED

We use an experiment that is the basically same as the NA-EBA experiment for the GO-NonAED except for the target allocation ratio. Instead of (10), we use the target allocation ratio defined in (3) by relacing a_0^* with a hypothetical treatment arm $\widetilde{a}$. Let $\sigma_0^{\widetilde{a}}$ be $\widetilde{\sigma}_0$. Therefore, we set the target allocation as

$$w^{\text{NA-EBA}}(\widetilde{a}) = \frac{\widetilde{\sigma}_0}{\widetilde{\sigma}_0 + \sqrt{\sum_{b\in[K]\setminus\{\widetilde{a}\}} \left(\sigma_0^b\right)^2}}, \tag{12}$$

$$w^{\text{NA-EBA}}(a) = \left(1 - w^{\text{TS-EBA}}(\widetilde{a})\right) \frac{(\sigma_0^a)^2}{\sum_{b\in[K]\setminus\{\widetilde{a}\}} \left(\sigma_0^b\right)^2} \qquad \forall a \in [K]\setminus\{\widetilde{a}\}. \tag{13}$$

Comparison with results in Glynn & Juneja (2004). This allocation ratio matches the results of Glynn & Juneja (2004) under the small-gap regime. First, consider a case with Gaussian models. By approximating the probability limit in (Chen et al., 2000) and Glynn & Juneja (2004) under the small-gap regime, as a solution of a non-linear optimization problem, the target allocation ratio satisfies

$$w^{\text{NA-EBA}}(\widetilde{a}) = \sigma^{\widetilde{a}} \sqrt{\sum_{a\in[K]\setminus\{\widetilde{a}\}} (w^{\text{NA-EBA}}(a)/\sigma^a)^2}, \tag{14}$$

$$\frac{1}{\frac{(\sigma^{\widetilde{a}})^2}{w^{\text{NA-EBA}}(a_0^*)} + \frac{(\sigma^a)^2}{w^{\text{NA-EBA}}(a)}} = \frac{1}{\frac{(\sigma^{\widetilde{a}})^2}{w^{\text{NA-EBA}}(a_0^*)} + \frac{(\sigma^b)^2}{w^{\text{NA-EBA}}(b)}} \qquad \forall (a,b) \in [K]\setminus\{\widetilde{a}\}, \tag{15}$$

$$\sum_{a\in[K]\setminus\{\widetilde{a}\}} \frac{\left(\sigma^{\widetilde{a}}\right)^2 / w^{\text{NA-EBA}}(\widetilde{a})}{\left(\sigma^{a}\right)^2 / w^{\text{NA-EBA}}(a)} = 1. \tag{16}$$

We can obtain the analytical solution for (14) as (3). Therefore, from the result of Glynn & Juneja (2004), we obtain an upper bound of the experiment that matches the lower bound in Theorem 4.7.

C.3. Probability of Misidentification and Asymptotic Optimality of the NA-EBA Experiment

This experiment is the same as that presented in Glynn & Juneja (2004), which is shown as globally asymptotically optimal. They find that the probability of misidentification is

$$\lim_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^{\text{EBA}} \neq a_0^*) = \min_{a\in[K]\setminus\{a_0^*\}} G^a(w^{\text{NA-EBA}}(a_0^*), w^{\text{NA-EBA}}(a)),$$

Note that from Sanov's theorem (Sanov, 1958), it holds that

$$\min_{a\in[K]\setminus\{a_0^*\}} G^a(w^{\text{NA-EBA}}(a_0^*), w^{\text{NA-EBA}}(a)) = \inf_{Q\in\mathcal{P}:a^*(Q)\neq a_0^*} \sum_{a\in[K]} w^{\text{NA-EBA}}(a)\text{KL}(Q^a, P^a).$$

In summary, we obtain the following proposition from Glynn & Juneja (2004).

Proposition C.2.

$$\lim_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^{\text{EBA}} \neq a_0^*) = \inf_{Q\in\mathcal{P}:a^*(Q)\neq a_0^*} \sum_{a\in[K]} w^{\text{NA-EBA}}(a)\text{KL}(Q^a, P^a). \tag{17}$$

Although Glynn & Juneja (2004) does not discuss lower bounds, we discover that the probability of misidentification under the alternative hypothesis in Glynn & Juneja (2004) matches our lower bound in Theorem 4.4 under the asymptotically invariant experiments. Note that Degenne (2023) independently gives the same explanation to experiments of Glynn & Juneja (2004). See page 5 and Theorem 12 in Degenne (2023).

D. Experiments with Two Treatment Arms

Interestingly, even without assuming asymptotically invariant experiments, we can derive the same lower bound for two-armed local location-shit statistical models only with assuming consistent experiments.

Theorem D.1 (Lower bound for two-armed local location-shift bandit models)**.** *When $K = 2$, for any $P_0 \in \mathcal{P}^{\text{G}}$ (Definition 4.5) and consistent (Definition 4.1) experiment π,*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^{\pi} \neq a_0^*) \leq \frac{\left(\mu_0^1 - \mu_0^2\right)^2}{2\left(\sigma^1 + \sigma^2\right)^2} + o\left(\left(\mu_0^1 - \mu_0^2\right)^2\right)$$

as $\mu_0^1 - \mu_0^2 \to 0$.

These results imply that even adding restrictions of Definitions 4.3, the lower bounds and the target allocation ratios do not change. This is because the lower bounds are characterized by the best and one suboptimal treatment arm, and the choice of the one suboptimal treatment arm affects the lower bound. However, when there are only two treatment arms, the pair of the best treatment arm and one suboptimal treatment arm is fixed (only includes $a = 1, 2$).

E. Simulation Studies

We investigate performances of our experiments in the settings of the GO-NonADE, the H-LO-NonADE, the H-LO-ADE, and the LO-ADE, and the existing Uniform-EBA experiment (Uniform, Bubeck et al., 2011) using simulation studies, which allocates treatment arms with the same allocation ratio ($1/K$). Let $K \in \{2, 5, 10\}$. Let r in the TS-EBA experiment be 0.5. The best treatment arm is arm 1 and $\mu_0^1 = 1$. The expected outcomes of suboptimal treatment arms are drawn from a uniform distribution with support $[0.75, 0.90]$ for $a \in [K]\setminus\{1, 2\}$, while $\mu_0^2 = 0.75$. The variances are drawn from a uniform distribution with support $[0.5, 5]$. We continue the experiments until $T = 5,000$ when $\widetilde{\mu} = 0.80$ and $T = 15,000$ when $\widetilde{\mu} = 0.90$. We conduct 100 independent trials for each setting. At each $T \in \{100, 500, 1000, \cdots, 14500, 15000\}$, we plot the empirical probability of misidentification in Figure 1. From the results, as the theory predicts, experiments using more information can achieve a lower probability of misidentification.

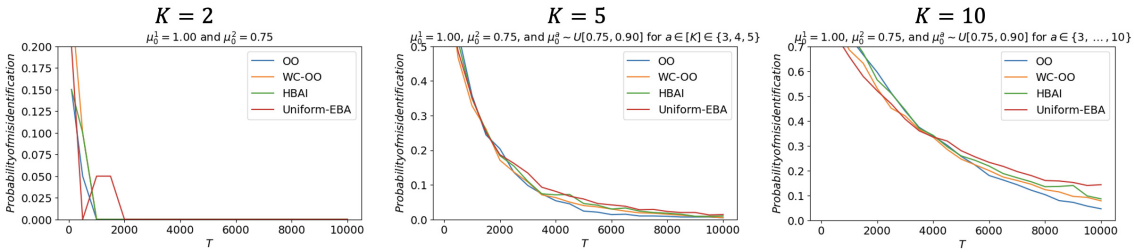

Figure 1. Experimental results. The y-axis and x-axis denote the probability of misidentification and T, respectively.

F. Related Work

Researchers have acknowledged the importance of statistical inference and experimental approaches as essential scientific tools (Peirce & Jastrow, 1884; Peirce & de Waal, 1887). With the advancement of these statistical methodologies, the experimental design also began attracting attention. Fisher (1935) develops the groundwork for the principles of experimental design. Wald (1949) establishes fundamental theories for statistical decision-making, bridging statistical inference and decision-making. These methodologies have been investigated across various disciplines, such as medicine, epidemiology, economics, operations research, and computer science, transcending their origins in statistics.

Ordinal optimization involves sample allocation to each treatment arm and selects a certain treatment arm based on a decision-making criterion; therefore, this problem is also known as the optimal computing budget allocation problem. The development of ordinal optimization is closely related to ranking and selection problems in simulation, originating from agricultural and clinical applications in the 1950s (Gupta, 1956; Bechhofer, 1954; Paulson, 1964; Branke et al., 2007; Hong et al., 2021). A modern formulation of ordinal optimization was established in the early 2000s (Chen et al., 2000; Glynn & Juneja, 2004). Existing research has found that the probability of misidentification converges at an exponential rate for a large set of problems. By employing large deviation principles (Cramér, 1938; Ellis, 1984; Gärtner, 1977; Dembo & Zeitouni, 2009), Glynn & Juneja (2004) proposes asymptotically optimal algorithms for ordinal optimization.

However, a promising idea for enhancing the efficiency of experiments is adaptive experimental design. In this approach, information from past experiments can be utilized to optimize the allocation of samples in subsequent trials. The concept of adaptive experimental design dates back to the 1970s (Pong & Chow, 2016). Presently, its significance is acknowledged (CDER, 2018; Chow & Chang, 2011). Adaptive experiments have also been studied within the domain of machine learning, and the multi-armed bandit (MAB) problem (Thompson, 1933; Robbins, 1952; Lai & Robbins, 1985) is an instance. The Best Arm Identification (BAI) is a paradigm of this problem (Even-Dar et al., 2006; Audibert et al., 2010; Bubeck et al., 2011), influenced by sequential testing, ranking, selection problems, and ordinal optimization (Bechhofer et al., 1968). There are two formulations in BAI: fixed-confidence (Garivier & Kaufmann, 2016) and fixed-budget BAI. In the former, the sample size (budget) is a random variable, and an experimenter stops an experiment when a certain criterion is satisfied, as well as sequential testing Wald (1945); Chernoff (1959). In contrast, the latter fixes the sample size (budget) and minimizes a certain criterion given the sample size. BAI in this study corresponds to fixed-budget BAI (Bubeck et al., 2011; Audibert et al., 2010; Bubeck et al., 2011). There is no strict distinction between the ordinal optimization and BAI [5].

A focal point of research interest has been to establish a tight lower bound on the probability of misidentification, representing a theoretical performance limit (Kaufmann, 2020). A BAI strategy (algorithm) is termed *asymptotically optimal* if, under this strategy, some criterion, such as the probability of misidentification and expected simple regret, matches the lower bound as the budget goes infinity. The existence of such an asymptotically optimal strategy has long been a perplexing issue in this field (Kaufmann, 2020; Qin & Russo, 2022). As an example, Glynn & Juneja (2004) proposes strategies that are asymptotically optimal based on optimally selected sample allocation ratios in a non-adaptive experiment. However, their approach requires complete distributional information. Kaufmann et al. (2016) derives lower bounds for the probability of misidentification for general settings, including adaptive and non-adaptive experiments. Despite this significant contribution, no optimal strategies corresponding to these lower bounds have been suggested. Indeed, Carpentier & Locatelli (2016) demonstrates that no strategy exists whereby the probability of misidentification matches the lower bounds deduced by Kaufmann et al. (2016). Also see (Kaufmann, 2020; Ariu et al., 2021). While the problem remains unresolved (Kaufmann, 2020; Qin & Russo, 2022), several approaches have been suggested (Komiyama et al., 2022; Degenne, 2023). The lower

[5]While ordinal optimization mainly addresses non-adaptive experiments, BAI mainly considers adaptive experiments. However, there are also studies about adaptive experiments in ordinal optimization; similarly, BAI also discuss non-adaptive experiments

bound of Carpentier & Locatelli (2016) is based on a minimax evaluation of the probability of misidentification under a large gap. From a Bayesian perspective, Russo (2016), Qin et al. (2017), and Shang et al. (2020) propose Bayesian BAI strategies that are optimal in terms of posterior convergence rate. However, it has been shown by Kasy & Sautmann (2021) and Ariu et al. (2021) that such optimality does not extend to asymptotic optimality for the probability of misidentification. From a different perspective, Atsidakou et al. (2023) proposes a Bayes optimal strategy for minimizing the probability of misidentification. However, it is still an open issue whether there exists an experiment whose probability of misidentification matches the lower bound conjectured by Kaufmann et al. (2016).

Independently of us, Degenne (2023) analyzes the BAI strategies under the class of static proportions, which correspond to our asymptotically invariant experiments. That study shows some impossibility theorems under the fixed-gap regime; that is, when the gap is fixed, there does not exist optimal experiments. In contrast, we show the existence of optimal experiments under the small-gap regime.

Our problem has close ties to theories of statistical decision-making (Wald, 1949; Manski, 2000; 2002; 2004), limits of experiments (Le Cam, 1972; van der Vaart, 1998), and semiparametric theory (Hahn, 1998). The semiparametric theory is particularly crucial as it enables the characterization of lower bounds through the semiparametric analog of Fisher information (van der Vaart, 1998).

G. Proof of General Lower Bound (Lemma 4.6)

In this section, we provide proof of Lemma 4.6. Our argument is based on a change-of-measure argument, which has been applied to BAI (Kaufmann et al., 2016). In this derivation, we relate the likelihood ratio to the lower bound. Inspired by Murphy & van der Vaart (1997), we expand the semiparametric likelihood ratio, where the gap parameter $\mu_0^* - \mu_0^a$ is regarded as a parameter of interest and the other parameters as nuisance parameters. By using a semiparametric efficient score function, we apply a series expansion to the likelihood ratio of the distribution-dependent lower bound around the gap parameter $\mu_0^* - \mu_0^a$ under a statistical model of an alternative hypothesis. Then, when the gap parameter goes to 0, the lower bound is characterized by the variance of the semiparametric influence function. Our proof is also inspired by van der Vaart (1998) and Hahn (1998).

Kato et al. (2023a) also presents worst-case lower bounds for the expected simple regret employing our proof techniques, but the proof procedure is different in some points. The difference mainly comes from the asymmetricity of the KL divergence. In our case, when evaluating the performance for each given statistical model, we face the notorious problem of reverse KL problem (Kaufmann, 2020). A technical issue with this problem is that the target allocation in theoretical analysis depends on the hypothetical statistical models rather than the true statistical model. This dependency is one of the reasons why we restrict our strategies to asymptotically invariant ones. However, in minimax evaluation, we do not suffer from the problem of reverse KL problem. For this property, we do not have to put the restriction on the asymptotically invariant strategies. Besides, this difference also affects the construction of parametric submodels. As a result, although the proofs might look similar for some readers, the details are different and cannot be applied to each other without modifications of the proof.

Precisely, our proof follows these steps. First, the goal is to express the lower bound of the probability of misidentification by using the gap parameter. In Proposition G.1 of Appendix G.1, we introduce a bound for some event based on a change-of-measure argument (Kaufmann et al., 2016). We apply this bound to derive lower bounds for the probability of misidentification in the final step of the proof. Next, we consider distributions of observations. Although we defined distributions of the potential random variables $(Y_{1,t}, Y_{2,t}, \dots, Y_{K,t})$ (full-data statistical models), we can only observe an outcome of a chosen treatment arm, $Y_t^{A_t}$, and cannot observe other outcomes $(Y_{a,t})_{a \in [K] \setminus \{A_t\}}$. Therefore, distributions of observations are different from the full-data statistical models. We induce the former from the latter in Appendix G.2 to discuss optimality. With these preparations, in Appendix G.3, we introduce a parameter into the true nonparametric full-data statistical models to differentiate the log-likelihood around the gap parameter; that is, the gap parameter is introduced so that it corresponds to $\mu_0^* - \mu_0^a$. This parameter is a technical device for the proof, and the parametrized models are called parametric submodels, which are subsets of $\mathcal{P}^*$. The derivative is then defined with respect to this parameter, and we consider applying the series expansion to the log-likelihood. However, the derivative (score function) is not uniquely defined because it includes nuisance parameters other than the parameter of interest. Therefore, to specify a score function with the tightest lower bound, it is necessary to consider information on the distribution of the observations. To perform these operations, we associate the full-data statistical models with the distribution of the observed data in Appendix G.4. Then, in Appendix G.5, we derive the parametric submodel of the distribution of observations from the parametric submodels of the full-data statistical models and define a score function for the parametric submodel of the distribution of observations. For

deriving lower bounds, an alternative hypothesis plays an important role, and we define a class of alternative hypotheses (alternative statistical models) in Appendix G.6. For the score function and alternative statistical models in Appendix G.6, we apply the series expansion to the log-likelihood in Appendix G.7 and characterize the bound in Proposition G.1 of Appendix G.1 with the gap parameter. Then, in Appendix G.8, we derive the information bound of the second moment of the score function; then, in Appendix G.9, we specify a score function whose second moment is equal to the information bound in Appendix G.8. Finally, combining them, we derive the lower bound for the probability of misidentification in Appendix G.10.

Throughout the proof, for simplicity, $\mathcal{P}^*$ is denoted by $\mathcal{P}$.

G.1. Transportation Lemma

Our lower bound derivation is based on change-of-measure arguments, which have been extensively used in the bandit literature (Lai & Robbins, 1985). (Kaufmann et al., 2016) derives the following result based on change-of-measure argument, which is the principal tool in our lower bound. Let us define a density of $(Y^1, Y^2, \dots, Y^K)$ under a statistical model $P \in \mathcal{P}$ as

$$p_P(y^1, y^2, \dots, y^K) = \prod_{a \in [K]} f_P^a(y^a)$$

Let $f_P^{a_0^*}$ be denoted by f_P^*.

Proposition G.1 (Lemma 1 in (Kaufmann et al., 2016))**.** *For any two statistical model $P, Q \in \mathcal{P}$ with K treatment arms such that for all $a \in [K]$, $f_P^a(y^a)$ and $f_Q^a(y^a)$ are mutually absolutely continuous,*

$$\mathbb{E}_Q\left[\sum_{t=1}^T \mathbb{1}[A_t = a] \log\left(\frac{f_Q^a(Y_t^a)}{f_P^a(Y_t^a)}\right)\right] \geq \sup_{\mathcal{E} \in \mathcal{F}_T} d(\mathbb{P}_Q(\mathcal{E}), \mathbb{P}_P(\mathcal{E})).$$

Recall that $d(p, q)$ indicates the KL divergence between two Bernoulli distributions with parameters $p, q \in (0, 1)$.

This "transportation" lemma provides the distribution-dependent characterization of events under a given statistical model P and corresponding perturbed statistical model P'.

Between the true statistical model $P \in \mathcal{P}$ and a statistical model $Q \in \mathcal{P}$, following the proof of Lemma 1 in Kaufmann et al. (2016), we define the log-likelihood ratio as

$$L_T = \sum_{t=1}^T \sum_{a \in [K]} \mathbb{1}[A_t = a] \log\left(\frac{f_Q^a(Y_t^a)}{f_P^a(Y_t^a)}\right).$$

For this log-likelihood ratio, from Lemma G.1, between the true model P, we have

$$\mathbb{E}_Q[L_T] \geq \sup_{\mathcal{E} \in \mathcal{F}_T} d(\mathbb{P}_Q(\mathcal{E}), \mathbb{P}_{P_0}(\mathcal{E})).$$

We consider an approximation of $\mathbb{E}_Q[L_T]$ under an appropriate alternative hypothesis $Q \in \mathcal{P}$ when the gaps between the expected outcomes of the best treatment arm and suboptimal treatment arms are small.

G.2. Observed-Data statistical models

Next, we define a semiparametric model for observed data (Y_t, A_t), as we can only observe the triple (Y_t, A_t) and cannot observe the full-data $(Y_{1,t}, Y_{2,t}, \dots, Y_{K,t})$.

Then, we first show the following lemma. We show the proof in Appendix H.

Lemma G.2. *For $P, Q, P \in \mathcal{P}$,*

$$\frac{1}{T}\mathbb{E}_P[L_T] = \sum_{a \in [K]} \mathbb{E}_P\left[\log \frac{f_Q^a(Y_{a,t})}{f_P^a(Y_{a,t})}\right] \kappa_{T,P}(a).$$

Based on Lemma G.2, for some $\kappa \in \mathcal{W}$, we consider the following samples $\{(\overline{Y}_t, \overline{A}_t)\}_{t=1}^T$, instead of $\{(Y_t, A_t)\}_{t=1}^T$, generated as

$$\{(\overline{Y}_t, \overline{A}_t)\}_{t=1}^T \overset{\text{i.i.d}}{\sim} r(y,d) = \prod_{a\in[K]} \{f_P^a(y^a)\kappa(a)\}^{\mathbb{1}[d=a]},$$

where $\kappa(a)$ corresponds to the conditional expectation of $\mathbb{1}[\overline{A}_t = a]$. The expectation of L_T for $\{(\overline{Y}_t, \overline{A}_t)\}_{t=1}^T$ on P is identical to that for $\{(Y_t, A_t)\}_{t=1}^T$ from the result of Lemma G.2 when $\kappa = \kappa_{T,P}$. Therefore, to derive the lower bound for $\{(Y_t, A_t)\}_{t=1}^T$, we consider that for $\{(\overline{Y}_t, \overline{A}_t)\}_{t=1}^T$. Note that this data generating process is induced by a full-data statistical model $P \in \mathcal{P}$; therefore, we call it an observed-data statistical model.

Formally, for a statistical model $P \in \mathcal{P}$ and some $\kappa \in \mathcal{W}$, by using a density function of P, let $\overline{R}_P^\kappa$ be a distribution of an observed-data statistical model $\{(\overline{Y}_t, \overline{A}_t)\}_{t=1}^T$ with the density given as

$$\overline{r}_P^\kappa(y,d) = \prod_{a\in[K]} \{f_P^a(y)\kappa(a)\}^{\mathbb{1}[d=a]}.$$

We call it an observed-data distribution. To avoid the complexity of the notation, we will denote $\{(\overline{Y}_t, \overline{A}_t)\}_{t=1}^T$ as $\{(Y_t, A_t)\}_{t=1}^T$ in the following arguments. Let $\mathcal{R} = \left\{\overline{R}_P : P \in \mathcal{P}\right\}$ be a set of all observed-data statistical models $\overline{R}_P$. For $P \in \mathcal{P}$, let $\overline{R}_P^\kappa = \overline{R}_0^\kappa$, and $\overline{r}_P^\kappa = \overline{r}_0^\kappa$.

G.3. Parametric Submodels for the Full-Data statistical models

The purpose of this section is to introduce parametric submodels for the true full-data statistical model $P \in \mathcal{P}$, which is indexed by a real-valued parameter and a set of distributions contained in the larger set $\mathcal{P}$, and define the derivative of the parametric submodels.

In Section G.5, we define parametric submodels for observed-data statistical models under the true full-data statistical model, which is a set of distributions contained in the larger set $\mathcal{R}_0$, by using the parametric submodels for full-data statistical models. These definitions of parametric submodels are preparations for the series expansion of the log-likelihood; that is, we consider approximation of the log-likelihood $L_T = \sum_{t=1}^T \sum_{a\in[K]} \mathbb{1}[A_t = a] \log\left(\frac{f_Q^a(Y_t^a)}{f_P^a(Y_t^a)}\right)$ using $\mu_0^* - \mu_0^a$, where $Q \in \mathcal{P}$ is an alternative statistical model.

This section consists of the following two parts. In the first part, we define parametric submodels as (18) with condition (19). Then, in the following part, we confirm the differentiability (24) and define score functions.

Definition of parametric submodels for the observed-data distribution. First, we define parametric submodels for the true full-data statistical model P with the density function $p_P(y^1, \ldots, y^K)$ by introducing a parameter $\boldsymbol{\varepsilon} = (\varepsilon^a)_{a\in[K]\setminus\{a_0^*\}}$ $\varepsilon^a \in \Theta$ with some compact space Θ. We construct our parametric submodels so that the parameter can be interpreted as the gap parameter of a parametric submodel. For $P \in \mathcal{P}$, we define a set of parametric submodels $\left\{P_{\boldsymbol{\varepsilon}} : \boldsymbol{\varepsilon} \in \Theta^{K-1}\right\} \subset \mathcal{P}$ as follows: for a set of some functions $(g^a)_{a\in[K]\setminus\{a_0^*\}}$ such that $g^a : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, a parametric submodel $P_{\boldsymbol{\varepsilon}}$ has a density such that for each $a \in [K]\setminus\{a_0^*\}$, $g^a(\phi_\tau^*(y), \phi_\tau^a(y)) = 0$, and

$$p_{\boldsymbol{\varepsilon}}(y_*, y_a) = \left(1 + \varepsilon^a g^a\left(\phi_\tau^*(y), \phi_\tau^a(y)\right)\right) p_P(y_*, y_a), \tag{18}$$

where for a constant $\tau > 0$ and each $d \in [K]$, $\phi_\tau^d : \mathbb{R}\times\mathcal{X} \to (-\tau, \tau)$ is a truncation function such that for $\varepsilon^a < c(\tau)$,

$$\phi_\tau^d(y) = y\mathbb{1}[|y| < \tau] - \mathbb{E}_P[Y_t^d \mathbb{1}[|Y_t^d| < \tau]], \qquad |\varepsilon^a g^a\left(\phi_\tau^*(y), \phi_\tau^a(y)\right)| < 1,$$

and $c(\tau)$ is some decreasing scalar function with regard to τ such that for the inverse $c^{-1}(e) = \tau$, $\tau \to \infty$ as $e \to 0$. Let $\phi^{a_0^*}$ be denoted by ϕ^*. This is a standard construction of parametric submodels with unbounded random variables (Hansen, 2022). For $a \in [K]\setminus\{a_0^*\}$, this parametric submodel must satisfy $\mathbb{E}_P[g^a(\phi_\tau^*(Y_t), \phi_\tau^a(Y_t))] = 0$, $\mathbb{E}_P[(g^a(\phi_\tau^*(Y_t), \phi_\tau^a(Y_t)))^2] < \infty$, and

$$\int\int (y_* - y_a)\, p_{\boldsymbol{\varepsilon}}(y_*, y_a) \mathrm{d}y_* \mathrm{d}y_a \mathrm{d}x = \mu_* - \mu_a + \varepsilon_a. \tag{19}$$

In Section G.8, we specify functions $(g^a)_{a\in[K]\backslash\{a_0^*\}}$ and confirm that the specified g^a satisfies (19). Note that the parametric submodels are usually not unique. For each $a \in [K]\backslash\{a_0^*\}$, the parametric submodel $p_{\boldsymbol{\varepsilon}}(y_*, y_a)$ is equivalent to $p_P(y^*, y^a)$ when $\varepsilon^a = 0$ for any $(\varepsilon^e)_{e\in[K]\backslash\{a^*,a\}}$.

For each $a \in [K]\backslash\{a_0^*\}$ and a parametric submodel $P_{\boldsymbol{\varepsilon}}$, let $f^*_{\boldsymbol{\varepsilon}}(y)$ and $f^a_{\boldsymbol{\varepsilon}}(y) = f^a_{\varepsilon^a}(y)$ be the densities of Y_t^* and $Y_{a,t}$, which satisfies (18) and (19) as

$$p_{\boldsymbol{\varepsilon}}(y_*, y_a) = f^*_{\boldsymbol{\varepsilon}}(y) f^a_{\varepsilon^a}(y),$$
$$\int\int (y_* - y_a)\, f^*_{\boldsymbol{\varepsilon}}(y) f_{a,\varepsilon^a}(y) \mathrm{d}y_* \mathrm{d}y_a = \mu_* - \mu_a + \varepsilon_a.$$

According to the definition of the parametric submodels, $f^*_{\mathbf{0}}(y) = f^*_P(y)$, $f^a_{\mathbf{0}}(y) = f^a_0(y) = f^a_P(y)$.

Differentiablity and score functions of the parametric submodels for the full-data distribution. Next, we confirm the differentiablity of $p_{\boldsymbol{\varepsilon}}(y_*, y_a)$. Because $\sqrt{p_{\boldsymbol{\varepsilon}}(y_*, y_a)}$ is continuously differentiable for every (y_*, y_a), and $\int \left(\frac{\dot{p}_{\boldsymbol{\varepsilon}}(y_*,y_a)}{p_{\boldsymbol{\varepsilon}}(y_*,y_a)}\right)^2 p_{\boldsymbol{\varepsilon}}(y_*, y_a)\mathrm{d}m$ are well defined and continuous in $\boldsymbol{\varepsilon}$, where m is some reference measure on (y_*, y_a), from Lemma 7.6 of van der Vaart (1998), we see that the parametric submodel has the score function g_a in the L_2 sense; that is, the density $p_{\boldsymbol{\varepsilon}}(y_*, y_a)$ is differentiable in quadratic mean (DQM): for $a \in [K]\backslash\{a_*\}$, and any $(\varepsilon^b)_{b\in[K]\backslash\{a_*,a\}}$,

$$\int \left[p^{1/2}_{\boldsymbol{\varepsilon}}(y_*, y_a) - p^{1/2}_P(y_*, y_a) - \frac{1}{2}\varepsilon^a g_a(\phi^*_\tau(y), \phi_{a,\tau}(y)) p^{1/2}_P(y_*, y_a)\right]^2 \mathrm{d}m = o\left(\varepsilon_a\right). \tag{20}$$

This relationship is derived from

$$\frac{\partial}{\partial\varepsilon_a}\Big|_{\varepsilon^a=0} \log p_{\boldsymbol{\varepsilon}}(y_*, y_a) = \frac{g^a(\phi_{*,\tau}(y), \phi_{a,\tau}(y))}{1 + \varepsilon_a g_a(\phi_{*,\tau}(y), \phi_{a,\tau}(y))}\Big|_{\varepsilon^a=0} = g_a(\phi_{*,\tau}(y), \phi_{a,\tau}(y)),$$

for any $(\varepsilon_b)_{b\in[K]\backslash\{a^*,a\}}$.

To clarify the relationship between g_a and a score function, for each $a \in [K]\backslash\{a_0^*\}$, and any $(\varepsilon_b)_{b\in[K]\backslash\{a^*,a\}}$, we express the score function as

$$g^a(\phi_{*,\tau}(y), \phi^a_\tau(y)) = \frac{\partial}{\partial\varepsilon^a}\Big|_{\varepsilon_a=0} \log p_{\boldsymbol{\varepsilon}}(y_*, y_a) = S_f^{a,a_0^*}(y) + S_f^{a,a}(y),$$

where

$$S_f^{a,a_0^*}(y) = \frac{\partial}{\partial\varepsilon^a}\Big|_{\varepsilon^a=0} \log f^*_{\boldsymbol{\varepsilon}}(y), \quad S_f^{a,a}(y) = \frac{\partial}{\partial\varepsilon^a}\Big|_{\varepsilon^a=0} \log f^a_{\varepsilon^a}(y).$$

G.4. Mapping from Observed-Data to Full-Data statistical models

According to Section 7.2 of Tsiatis (2007), we define a mapping from full-data to observed-data as $y = \mathcal{T}^d(y^*, y^a)$, where $\mathcal{T}^d : \mathbb{R}^2 \to \mathbb{R}$ is a known many-to-one function, which maps the full-data (y^*, y^a) to observed-data statistical models (y^d).

We only consider a case where $(Y_t^*, Y_{a,t})$ is continuous and define a function $V^d : \mathbb{R}^2 \to \mathbb{R}$ as a counterfactual value of the observation; that is, $V^d(Y_t^*, Y_{a,t}) = ((Y_t^b)_{b\in\{a^*,a\}\backslash\{d\}})$. Then, the mapping

$$(Y_t^*, Y_{a,t}) \mapsto \{\mathcal{T}^d(Y_t^*, Y_{a,t}), V^d(Y_t^*, Y_{a,t})\}$$

is one-to-one for all $a \in [K]\backslash\{a_0^*\}$ and $d \in \{a^*, a\}$. For $a \in [K]\backslash\{a_0^*\}$, $d \in \{a^*, a\}$, $\tau^d = (y^d)$, and $v^d = ((y^b)_{b\in\{a^*,a\}\backslash\{d\}})$, which correspond to $\mathcal{T}^d$ and V^d respectively, we define the inverse transformation as

$$(y^*, y^a) = H^d(\tau^d, v^d),$$

Then, by the standard formula for change of variables, let us define the density of (τ^d, v^d) under $\mathcal{T}^d$ and V^d as

$$p_{\mathcal{T}^d,V^d}(\tau^d, v^d) = p_P(H^d(\tau^d, v^d))J(\tau^d, v^d), \tag{21}$$

where J is the Jacobian of H^d with respect to (τ^d, v^d). To find the density of the observed data $\overline{r}^{\kappa}_P(y,d)$, we can use

$$\overline{r}^{\kappa}_P(y,d) = \int \overline{r}^{\kappa}_{P,V^d}(\tau^d, d, v^d) dv^d, \tag{22}$$

where

$$\overline{r}^{\kappa}_{P,V^d}(\tau^d, d, v^d) = \kappa(d) p_{\mathcal{T}^d,V^d}(\tau^d, v^d). \tag{23}$$

Consequently, using (21) and (23), we can rewrite (22) as

$$\overline{r}^{\kappa}_P(y,d) = \int \kappa\big(d\big) p_P(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d.$$

G.5. Parametric Submodels for the Observed-Data statistical models and Tangent Space

This section consists of the following three parts. In the first part, we define parametric submodels as (18) with condition (19). Then, in the following part, we confirm the differentiability (24) and define score functions. Finally, we define a set of score functions, called a tangent set in the final paragraph.

By using the parametric submodels and tangent set, in Section G.7, we demonstrate the series expansion of the log-likelihood (Lemma G.5). In this section and Section G.7, we abstractly provide definitions and conditions for the parametric submodels and do not specify them. However, in Sections G.8 and G.9, we show a concrete form of the parametric submodel by finding score functions satisfying the conditions imposed in this section.

By using the parametric submodels for the true full-data statistical model $P \in \mathcal{P}$ in Section G.3, we define parametric submodels for observed-data statistical models under the true full-data statistical model $P \in \mathcal{P}$. Because we define the density functions of the parametric submodel of the true full-data statistical model, the parametric submodels for the observed-data statistical models are given as follows:

$$\overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,a) = f^a_{\varepsilon^a}(y)\kappa(a) \qquad \forall a \in [K]\backslash\{a^*_0\}, \qquad \overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,a^*_0) = f^*_{\boldsymbol{\varepsilon}}(y)\kappa(a^*_0).$$

Differentiablity and score functions of the parametric submodels for the observed-data distribution. Next, we confirm the differentiablity of $\overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d)$. Because $\sqrt{\overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d)}$ is continuously differentiable for every y given $d \in [K]$, and $\int \left(\frac{\dot{\overline{r}}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d)}{\overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d)}\right)^2 \overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d)\mathrm{d}m$ are well defined and continuous in $\boldsymbol{\varepsilon}$, where m is some reference measure on (y,d), from Lemma 7.6 of van der Vaart (1998), we see that the parametric submodel has the score function g^a in the L_2 sense; that is, the density $\overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d)$ is differentiable in quadratic mean (DQM): for $a \in [K]\backslash\{a^*_0\}$, $d \in \{a^*, a\}$, and any $(\varepsilon^b)_{b\in[K]\backslash\{a^*,a\}}$,

Then we show the differentiablity in quadratic mean at $\varepsilon^a = 0$ of $\overline{r}^{\kappa,1/2}_{\boldsymbol{\varepsilon}}$ in the following lemma. We show the proof in Appendix I.

Lemma G.3. *For $a \in [K]\backslash\{a^*_0\}$ and $d \in \{a^*, a\}$,*

$$\int \left[\overline{r}^{\kappa\ 1/2}_{\boldsymbol{\varepsilon}}(y,d) - \overline{r}^{\kappa\ 1/2}_0(y,d) - \frac{1}{2}\varepsilon^a S^a(y,d)\overline{r}^{\kappa\ 1/2}_0(y,d)\right]^2 \mathrm{d}m = o\left(\varepsilon^a\right). \tag{24}$$

where

$$S^a(y,d) = \mathbb{E}_P\left[g^a\left(\phi^*_\tau(Y^*_t), \phi^a_\tau(Y_{a,t})\right) | \mathcal{T}^d(Y^*_t, Y^*_t) = y\right]. \tag{25}$$

In the following section, we specify a measurable function S^a wigh g^a, satisfying the conditions (18) and (19), which corresponds to a score function of $\overline{r}^{\kappa}_0(y,a)$ and $\overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,a^*_0)$ for each $a \in [K]\backslash\{a^*_0\}$. To clarify the relationship between g^a and a score function, for each $a \in [K]\backslash\{a^*_0\}$, and any $(\varepsilon^b)_{b\in[K]\backslash\{a^*,a\}}$, we denote the score function as

$$S^a(y,d) = \frac{\partial}{\partial \varepsilon^a}\Big|_{\varepsilon^a=0} \log \overline{r}^{\kappa}_{\boldsymbol{\varepsilon}}(y,d) = \mathbb{1}[d = a^*_0] S^{a,a^*_0}_f(y) + \mathbb{1}[d=a] S^{a,a}_f(y) \quad \forall d \in \{a^*, a\},$$
$$S^a(y,d) = 0 \quad \forall d \in [K]\backslash\{a^*, a\}.$$

Note that $\frac{\partial}{\partial \varepsilon^a} \log \kappa(a) = 0$.

Definition of the tangent set. Recall that parametric submodels and corresponding score functions are not unique. Here, we consider a set of score functions. For a set of the parametric submodels $\left\{ \overline{R}^{\kappa}_{\boldsymbol{\varepsilon}} : \boldsymbol{\varepsilon} \in \Theta^{K-1} \right\}$, we obtain a corresponding set of score functions g^a in the Hilbert space $L_2(\overline{R}_Q)$, which we call a tangent set of $\mathcal{R}$ at $\overline{R}^{\kappa}_0$ and denote it by $\dot{\mathcal{R}}^a$. Because $\mathbb{E}_{\overline{R}^{\kappa}_0}[(g^a(\phi^{A_t}_{\tau}(Y_t), A_t))^2]$ is automatically finite, the tangent set can be identified with a subset of the Hilbert space $L_2(\overline{R}^{\kappa}_0)$, up to equivalence classes. For our parametric submodels, the tangent set at $\overline{R}^{\kappa}_0$ in $L_2(\overline{R}^{\kappa}_0)$ is given as

$$\dot{\mathcal{R}}^a = \left\{ \mathbb{1}[d = a^*_0] S^{a, a^*_0}_f(y) + \mathbb{1}[d = a] S^{a,a}_f(y) \right\}.$$

G.6. Alternative statistical model

Then, we define a class of alternative hypotheses. To derive a tight lower bound by applying the change-of-measure arguments, we use an appropriately defined alternative hypothesis. Our alternative hypothesis is defined using the parametric submodel of P as follows:

Definition G.4. Let $\mathrm{Alt}(P) \subset \mathcal{P}$ be alternative statistical models such that for all $Q \in \mathrm{Alt}(P)$, $a^*(Q) \neq a^*$, and $\overline{R}^{\kappa_T, Q}_{\boldsymbol{\varepsilon}} = \overline{R}^{\kappa_T, Q}_Q$, where $\boldsymbol{\varepsilon} = (\varepsilon^a)_{a \in [K] \setminus \{a^*_0\}}$, $\varepsilon^a = \left(\mu^{a^*_0}(Q) - \mu^a(Q)\right) - (\mu^*_0 - \mu^a_0)$.

This also implies that for all $Q \in \mathrm{Alt}(P)$, for all $a \in [K] \setminus \{a^*_0\}$, $\mu^*_0 - \mu^a_0 > 0$ and there exists $a \in [K] \setminus \{a^*_0\}$ such that $\mu^{a^*_0}(Q) - \mu^a(Q) < 0$. Let $\mu^{a^*_0}(Q)$ be denoted by $\mu^*(Q)$.

G.7. Semiparametric Likelihood Ratio

For $a \in [K] \setminus \{a^*_0\}$, let $\boldsymbol{\varepsilon}$ be $(0, \ldots, 0, \varepsilon^a, 0, \ldots, 0)$. Let us also define

$$L^a_T = \sum_{t=1}^{T} \left\{ \mathbb{1}[A_t = a^*_0] \log \left(\frac{f^*_{\boldsymbol{\varepsilon}}(Y^*_t)}{f^*_P(Y^*_t)} \right) + \mathbb{1}[A_t = a] \log \left(\frac{f^a_{\boldsymbol{\varepsilon}}(Y^a_t)}{f^a_P(Y^a_t)} \right) \right\}.$$

We consider series expansion of the log-likelihood L^a_T defined between $P \in \mathcal{P}$ and $Q \in \mathrm{Alt}(P)$, where $\mathbb{E}_Q[L_T]$ works as a lower bound for the probability of misidentification as shown in Appendix G.10. We consider an approximation of L^a_T under a small-gap regime (small $\mu^*_0 - \mu^a_0$), which is upper-bounded by the variance of the score function. Our argument is inspired by that in Murphy & van der Vaart (1997).

Then, we prove the following lemma:

Lemma G.5. *For $P \in \mathcal{P}$, $Q \in \mathrm{Alt}(P)$, and each $a \in [K] \setminus \{a^*_0\}$,*

$$\frac{1}{T} \mathbb{E}_Q[L^a_T] = \frac{(\varepsilon^a)^2}{2} \mathbb{E}_P\left[(S^a(Y_t, A_t))^2 \right] + o\left((\varepsilon^a)^2 \right).$$

To prove this lemma, for $a \in [K] \setminus \{a^*_0\}$ and $d \in [K]$, we define

$$\ell^a_{\boldsymbol{\varepsilon}}(y, d) = \mathbb{1}[d = a^*_0] \log f^*_{\boldsymbol{\varepsilon}}(y) + \mathbb{1}[d = a] \log f^a_{\varepsilon^a}(y).$$

Note that if $\varepsilon^a = 0$, then

$$\ell^a_{\boldsymbol{\varepsilon}}(y, d) = \mathbb{1}[d = a^*_0] \log f^*_P(y) + \mathbb{1}[d = a] \log f^a_P(y).$$

Proof of Lemma G.5. By using the parametric submodel defined in the previous section, from the series expansion,

$$\begin{aligned} L^a_T &= \sum_{t=1}^{T} \left\{ \mathbb{1}[A_t = a^*_0] \log \left(\frac{f^*_{\boldsymbol{\varepsilon}}(Y^*_t)}{f^*_P(Y^*_t)} \right) + \mathbb{1}[A_t = a] \log \left(\frac{f^a_{\boldsymbol{\varepsilon}}(Y^a_t)}{f^a_P(Y^a_t)} \right) \right\} \\ &= \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial \varepsilon^a}\Big|_{\varepsilon^a = 0} \ell^a_{\boldsymbol{\varepsilon}}(Y_t, A_t) \varepsilon^a + \frac{\partial^2}{\partial (\varepsilon^a)^2}\Big|_{\varepsilon^a = 0} \ell^a_{\boldsymbol{\varepsilon}}(Y_t, A_t) \frac{(\varepsilon^a)^2}{2} + O\left((\varepsilon^a)^3 \right) \right\}, \end{aligned}$$

Here, we fix $(\varepsilon^b)_{b\in[K]\setminus\{a^*,a\}}$, where $\varepsilon^b = 0$. Note that

$$\frac{\partial}{\partial\varepsilon^a}\Big|_{\varepsilon^a=0}\ell^a_{\boldsymbol{\varepsilon}}(y,d) = S^a(y,d), \qquad \frac{\partial}{\partial(\varepsilon^a)^2}\Big|_{\varepsilon^a=0}\ell^a_{\boldsymbol{\varepsilon}}(y,d) = -\left(S^a(y,d)\right)^2.$$

Let $\overline{R}^{\kappa_T,Q}_{\boldsymbol{\varepsilon}} = \overline{R}_{\boldsymbol{\varepsilon}}$, $\overline{r}^{\kappa_T,Q}_{\boldsymbol{\varepsilon}}(y,d) = \overline{r}_{\boldsymbol{\varepsilon}}(y,d)$, and $\overline{r}^{\kappa_T,Q}_0(y,d) = \overline{r}_0(y,d)$. Then,

$$\begin{aligned}
&\mathbb{E}_Q\left[S^a(Y_t,A_t)\right] = \mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}}\left[S^a(Y_t,A_t)\right] \\
&= \mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}}\left[S^a(Y_t,A_t)\right] - \sum_{d\in[K]}\int S^a(y,d)\left(1+\frac{1}{2}\varepsilon^a S^a(y,d)\right)^2\overline{r}_0(y,d)\mathrm{d}y \\
&\quad + \sum_{d\in[K]}\int S^a(y,d)\left(1+\frac{1}{2}\varepsilon^a S^a(y,d)\right)^2\overline{r}_0(y,d)\mathrm{d}y \\
&= \sum_{d\in[K]}\int S^a(y,d)\left\{\overline{r}_{\boldsymbol{\varepsilon}}(y,d) - \left(1+\frac{1}{2}\varepsilon^a S^a(y,d)\right)^2\overline{r}_0(y,d)\right\}\mathrm{d}y\mathrm{d}x \\
&\quad + \sum_{d\in[K]}\int S^a(y,d)\left(1+\frac{1}{2}\varepsilon^a S^a(y,d)\right)^2\overline{r}_0(y,d)\mathrm{d}y \\
&= \sum_{d\in[K]}\int S^a(y,d)\left\{\overline{r}_{\boldsymbol{\varepsilon}}(y,d) - \left(1+\frac{1}{2}\varepsilon^a S^a(y,d)\right)^2\overline{r}_0(y,d)\right\}\mathrm{d}y \\
&\quad + \mathbb{E}_P\left[S^a(Y_t,A_t)\right] + \varepsilon^a\mathbb{E}_P\left[\left(S^a(Y_t,A_t)\right)^2\right] + \frac{1}{4}(\varepsilon^a)^2\mathbb{E}_P\left[\left(S^a(Y_t,A_t)\right)^2\right],
\end{aligned}$$

where we used

$$\begin{aligned}
&\sum_{d\in[K]}\int S^a(y,d)\overline{r}_0(y,d)\mathrm{d}y \\
&= \sum_{d\in[K]}\int\left\{\mathbb{1}[d=a_0^*]S_f^{a,a_0^*}(y) + \mathbb{1}[d=a]S_f^{a,a}(y)\right\}\overline{r}_0(y,d)\mathrm{d}y.
\end{aligned}$$

Then, because the density $\overline{r}_{\boldsymbol{\varepsilon}}(y,d)$ is DQM (24), as $\varepsilon^a \to 0$,

$$\mathbb{E}_Q\left[S^a(Y_t,A_t)\right] - \mathbb{E}_P\left[S^a(Y_t,A_t)\right] - \varepsilon^a\mathbb{E}_P\left[\left(S^a(Y_t,A_t)\right)^2\right] = o(\varepsilon^a).$$

Similarly,

$$-\mathbb{E}_Q\left[\left(S^a(Y_t,A_t)\right)^2\right] + \mathbb{E}_P\left[\left(S^a(Y_t,A_t)\right)^2\right] - \varepsilon^a\mathbb{E}_P\left[\left(S^a(Y_t,A_t)\right)^3\right] = o(\varepsilon^a).$$

By using these expansions, we approximate $\mathbb{E}_Q\left[L_T\right]$. Here, by definition, $\mathbb{E}_P\left[S^a(Y_t,A_t)\right] = 0$. Then, we approximate the likelihood ratio as follows:

$$\frac{1}{T}\mathbb{E}_Q[L_T^a] = \frac{\left(\varepsilon^a\right)^2}{2}\mathbb{E}_P\left[\left(S^a(Y_t,A_t)\right)^2\right] + O\left(\left(\varepsilon^a\right)^3\right).$$

□

G.8. Observed-Data Semiparametric Efficient Influence Function

Our remaining task is to specify the score function S^a. Because there can be several score functions for our parametric submodel due to directions of the derivative, we find a parametric submodel that has a score function with the largest variance, called a least-favorable parametric submodel (van der Vaart, 1998).

In this section, instead of the original observed-data statistical model $\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}}$, we consider an alternative observed-data statistical model $\overline{R}_0^{\kappa_{T,Q}\,\dagger}$, which is a distribution of $\{(\phi_\tau^{A_t}(Y_t), A_t)\}_{t=1}^T$. Let $\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}$ be parametric submodel defined as well as Section G.5, $\mathcal{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}$ be a set of all $\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}$, and $\overline{r}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}(y, d = f_{\varepsilon^d}^{d,\dagger}(y)\kappa_{T,Q}(d)_{\boldsymbol{\varepsilon}}$. For each $a \in [K]\backslash\{a_0^*\}$, let $S^{a\,\dagger}(y, d)$ and $\dot{\mathcal{R}}^{a\,\dagger}$ be a corresponding score function and tangent space, respectively.

As a preparation, we define a parameter $\mu^*(Q) - \mu^a(Q)$ as a function $\psi^a : \mathcal{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger} \to \mathbb{R}$ such that $\psi^a(\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}) = \mu_0^* - \mu_0^a + \varepsilon^a$. The information bound for $\psi^a\left(\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}\right)$ of interest is called semiparametric efficiency bound. Let $\overline{\text{lin}}\dot{\mathcal{R}}^{a\,\dagger}$ be the closure of the tangent space. Then, $\psi^a\left(\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}\right) = \mu_0^* - \mu_0^a + \varepsilon^a$ is pathwise differentiable relative to the tangent space $\dot{\mathcal{R}}^{a\,\dagger}$ if and only if there exists a function $\widetilde{\psi}^a \in \overline{\text{lin}}\dot{\mathcal{R}}^{a\,\dagger}$ such that

$$\frac{\partial}{\partial \varepsilon^a}\Big|_{\varepsilon^a=0} \psi^a\left(\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}\right)\left(= \frac{\partial}{\partial \varepsilon^a}\Big|_{\varepsilon^a=0}\left\{\mu_0^* - \mu_0^a + \varepsilon^a\right\} = 1\right) = \mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}}\left[\widetilde{\psi}^a(Y_t, A_t)S^{a\,\dagger}(Y_t, A_t)\right].$$

This function $\widetilde{\psi}^a$ is called the *semiparametric influence function*.

Then, we prove the following lemma on the lower bound for $\mathbb{E}_P\left[\left(S^a(Y_t, A_t)\right)^2\right]$, which is called the semiparametric efficiency bound:

Lemma G.6. *Any score function $S^{a\,\dagger} \in \dot{\mathcal{R}}^{a\,\dagger}$ satisfies*

$$\mathbb{E}_P\left[\left(S^{a\,\dagger}(Y_t, A_t)\right)^2\right] \geq \frac{1}{\mathbb{E}_P\left[\left(\widetilde{\psi}^a(Y_t, A_t)\right)^2\right]}.$$

Proof. From the Cauchy-Schwartz inequality, we have

$$1 = \mathbb{E}_P\left[\widetilde{\psi}^a(Y_t, A_t)S^{a\,\dagger}(Y_t, A_t)\right] \leq \sqrt{\mathbb{E}_P\left[\left(\widetilde{\psi}^a(Y_t, A_t)\right)^2\right]}\sqrt{\mathbb{E}_P\left[\left(S^{a\,\dagger}(Y_t, A_t)\right)^2\right]}.$$

Therefore,

$$\sup_{S^{a\,\dagger}\in\dot{\mathcal{R}}^{a\,\dagger}} \frac{1}{\mathbb{E}_P\left[\left(S^{a\,\dagger}(Y_t, A_t)\right)^2\right]} \leq \mathbb{E}_P\left[\left(\widetilde{\psi}^a(Y_t, A_t)\right)^2\right].$$

□

For $a \in [K]\backslash\{a_0^*\}$ and $d \in [K]\backslash\{a^*, a\}$, let us define a *semiparametric efficient score function* $S_{\text{eff}}^a(y, d) \in \overline{\text{lin}}\dot{\mathcal{R}}^{a\,\dagger}$ as

$$S_{\text{eff}}^a(y, d) = \frac{\widetilde{\psi}^a(y, d)}{\mathbb{E}_P\left[\left(\widetilde{\psi}^a(Y_t, A_t)\right)^2\right]}.$$

Next, we consider finding $\widetilde{\psi}^a \in \overline{\text{lin}}\dot{\mathcal{R}}^{a\,\dagger}$. We can use the result of Hahn (1998). Let us guess that for each $a \in [K]\backslash\{a_0^*\}$ and $d \in \{a^*, a\}$, $\widetilde{\psi}^a(y, d)$ is given as follows:

$$\widetilde{\psi}^a(y, d) = \frac{\mathbb{1}[d = a](\phi_\tau^*(y) - \mu_0^*)}{\kappa_{T,Q}(a_0^*)} - \frac{\mathbb{1}[d = a](\phi_\tau^a(y) - \mu_0^a)}{\kappa_{T,Q}(a)}. \tag{26}$$

Then, as shown by Hahn (1998), the condition $1 = \mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}^{\kappa_{T,Q}\,\dagger}}\left[\widetilde{\psi}^a(Y_t, A_t)S^a(Y_t, A_t)\right]$ holds under (26) when for each $a \in [K]\backslash\{a_0^*\}$ and $d \in \{a^*, a\}$, the semiparametric efficient score functions are given as

$$S_{\text{eff}}^a(y, d) = \mathbb{1}[d = a_0^*]S_{f,\text{eff}}^{a,a_0^*}(y) + \mathbb{1}[d = a]S_{f,\text{eff}}^{a,a}(y),$$

$$S^{a,a_0^*}_{f,\mathrm{eff}}(y) = \frac{(\phi^*_\tau(y) - \mu^*_0)}{\kappa_{T,Q}(a^*_0)} / \widetilde{V}^a_0(\kappa_{T,Q};\tau),$$
$$S^{a,a}_{f,\mathrm{eff}}(y) = \frac{(\phi^a_\tau(y) - \mu^a_0)}{\kappa_{T,Q}(a)} / \widetilde{V}^a_0(\kappa_{T,Q};\tau).$$

where

$$\widetilde{V}^a_0(\kappa_{T,Q};\tau) = \frac{(\sigma^*_0(\tau))^2}{\kappa_{T,Q}(a^*_0)} + \frac{(\sigma^a_0(\tau))^2}{\kappa_{T,Q}(a)},$$
$$(\sigma^*_0(\tau))^2 := (\phi^*_\tau(Y_t;\tau) - \mu^*_0)^2,$$
$$(\sigma^a_0(\tau))^2 := (\phi^a_\tau(Y_t;\tau) - \mu^a_0)^2.$$

Here, note that for each $d \in [K]$,

$$\begin{aligned}
&\mathbb{E}_P\left[\left(\phi^d_\tau(Y_t) - \mu^d_0\right)^2\right] \\
&= \mathbb{E}_P\left[\left(Y^d_t \mathbb{1}[|Y^d_t| < \tau] - Y^d_t \mathbb{1}[|Y^d_t| < \tau]\right)^2\right] \\
&= \mathbb{E}_P\left[\left(Y^d_t\right)^2 \mathbb{1}[|Y^d_t| < \tau]\right] - \left(\mathbb{E}_P[Y^d_t \mathbb{1}[|Y^d_t| < \tau]]\right)^2.
\end{aligned}$$

We also note that $\mathbb{E}_{\overline{R}^{\kappa_{T,Q}\,\dagger}_{\boldsymbol{\varepsilon}}}[S^a_{\mathrm{eff}}(Y_t, A_t)] = 0$ and

$$\mathbb{E}_{\overline{R}^{\kappa_{T,Q}\,\dagger}_{\boldsymbol{\varepsilon}}}\left[\left(S^a_{\mathrm{eff}}(Y_t, A_t)\right)^2\right] = \widetilde{V}^a_0(\kappa_{T,Q};\tau) = \left(\mathbb{E}_{\overline{R}^{\kappa_{T,Q}\,\dagger}_{\boldsymbol{\varepsilon}}}\left[\left(\widetilde{\psi}^a(Y_t, A_t)\right)^2\right]\right)^{-1}.$$

Summarizing the above arguments, we obtain the following lemma.

Lemma G.7. *For $a \in [K]\backslash\{a^*_0\}$ and $d \in [K]\backslash\{a^*, a\}$, the semiparametric efficient influence function is*

$$\widetilde{\psi}^a(y, d) = \frac{\mathbb{1}[d = a^*_0](\phi^*_\tau(y) - \mu^*_0)}{\kappa_{T,Q}(a^*_0)} - \frac{\mathbb{1}[d = a](\phi^a_\tau(y) - \mu^a_0)}{\kappa_{T,Q}(a)}.$$

We also define the limit of the semiparametric efficient influence function when $\tau \to \infty$ and the variance as

$$\widetilde{\psi}^a_\infty(y, d) = \frac{\mathbb{1}[d = a^*_0](Y^*_t - \mu^*_0)}{\kappa_{T,Q}(a^*_0)} - \frac{\mathbb{1}[d = a](Y_{a,t} - \mu^a_0)}{\kappa_{T,Q}(a)},$$
$$\widetilde{V}^a_0(\kappa_{T,Q}) = \mathbb{E}_P\left[\left(\widetilde{\psi}^a_\infty(Y_t, A_t)\right)^2\right] = \mathbb{E}_P\left[\frac{(\sigma^*_0)^2}{\kappa_{T,Q}(a^*_0)} + \frac{(\sigma^a_0)^2}{\kappa_{T,Q}(a)}\right] \geq \Omega^a_0(\kappa_{T,Q}),$$

where $C > 0$ is a constant.

G.9. Specification of the Observed-Data Score Function

According to Lemma G.6, we can conjecture that if we use the semiparametric efficient score function for our score function, we can obtain a tight upper bound for $\mathbb{E}_P[L^a_T]$, which is related to a lower bound for the probability of misidentification. Note that the variance of the semiparametric efficient score function is equivalent to the lower bound in Lemma G.6. However, we cannot use the semiparametric efficient score function because it is derived for $\overline{R}^{\kappa_{T,Q}\,\dagger}_{\boldsymbol{\varepsilon}}$, rather than $\overline{R}^{\kappa_{T,Q}}_{\boldsymbol{\varepsilon}}$. Furthermore, if we use the semiparametric efficient score function for our score function, the constant (19) is not satisfied. Therefore, based on our obtained result, we specify our score function, which differs from the semiparametric efficient score function, but they match when $\tau \to \infty$.

We specify our score function $S^a(y, d) = \mathbb{1}[d = a^*_0] S^{a,a^*_0}_f(y) + \mathbb{1}[d = a] S^{a,a}_f(y)$ as follows:

$$S^{a,a^*_0}_f(y) = \frac{\phi^*_\tau(y) - \mu^*_0}{\kappa_{T,Q}(a^*_0)} / V^a_0(\kappa_{T,Q};\tau) = S^{a,a^*_0)}_{f,\mathrm{eff}}(y) \widetilde{V}^a_0(\kappa_{T,Q};\tau) / V^a_0(\kappa_{T,Q};\tau),$$

$$S_f^{a,a}(y) = \frac{\phi_\tau^a(y) - \mu_0^a}{\kappa_{T,Q}(a)} / V_0^a(\kappa_{T,Q};\tau) = S_{f,\mathrm{eff}}^{a,a}(y)\widetilde{V}_0^a(\kappa_{T,Q};\tau)/V_0^a(\kappa_{T,Q};\tau),$$

where

$$\begin{aligned} V_0^a(\kappa_{T,Q};\tau) &= \widetilde{V}_0^a(\kappa_{T,Q};\tau) + \sum_{d\in\{a^*,a\}} \mathbb{E}_P\left[\frac{\mu_0^d \mathbb{E}_P[Y_t^d \mathbb{1}[|Y_t^d| < \tau]] - \left(\mathbb{E}_P[Y_t^d \mathbb{1}[|Y_t^d| < \tau]]\right)^2}{\kappa_{T,Q}(d)}\right] \\ &= \mathbb{E}_P\left[\frac{Y_t^* \left(\phi_\tau^*(Y_t) - \mu_0^*\right)}{\kappa_{T,Q}(a_0^*)} + \frac{Y_{a,t}\left(\phi_\tau^a(Y_t) - \mu_0^a\right)}{\kappa_{T,Q}(a)} + \left(\left(\mu_0^* - \mu_0^a\right) - \left(\mu_0^* - \mu_0^a\right)\right)^2\right]. \end{aligned} \tag{27}$$

Here, note that for $d \in [K]$,

$$\begin{aligned} \mathbb{E}_P\left[Y_t^d\left(\phi_\tau^d(Y_t) - \mu_0^d\right)\right] &= \mathbb{E}_P\left[\left(\left(Y_t^d\right)^2 \mathbb{1}[|Y_t^d| < \tau] - Y_t^d \mathbb{E}_P[Y_t^d \mathbb{1}[|Y_t^d| < \tau]]\right)\right] \\ &= \mathbb{E}_P\left[\mathbb{E}_P\left[\left(Y_t^d\right)^2 \mathbb{1}[|Y_t^d| < \tau]\right] - \mu_0^d \mathbb{E}_P[Y_t^d \mathbb{1}[|Y_t^d| < \tau]]\right]. \end{aligned}$$

We note that $V_0^a(\kappa_{T,Q};\tau) \to \widetilde{V}_0^a(\kappa_{T,Q})$ as $\varepsilon^a \to 0$ and $\tau \to \infty$,.

From the definition of the parametric submodel, we have

$$g^a(\phi_\tau^*(y), \phi_\tau^a(y)) = S_f^{a,a_0^*}(y) + S_f^{a,a}(y) = \left\{\frac{(\phi_\tau^*(y) - \mu_0^*)}{\kappa_{T,Q}(a_0^*)} - \frac{(\phi_\tau^a(y) - \mu_0^a)}{\kappa_{T,Q}(a)})\right\} / V_0^a(\kappa_{T,Q};\tau).$$

Then, we can also confirm that condition (19) holds for our specified g^a:

$$\begin{aligned} &\int\int (y^* - y^a)\left(1 + \varepsilon^a g^a(\phi_\tau^*(y), \phi_\tau^a(y))\right) p_P(a^*, a)\mathrm{d}y^*\mathrm{d}y^a\mathrm{d}x \\ &= \mu_0^* - \mu_0^a + \varepsilon^a\left\{\int\int (y^* - y^a)\, g^a(\phi_\tau^*(y), \phi_\tau^a(y))\overline{r}_0(y, a_0^*)\mathrm{d}y^*\mathrm{d}y^a\mathrm{d}x\right\} \\ &= \mu_0^* - \mu_0^a + \varepsilon^a, \end{aligned}$$

where we used the definition of the variance (27).

In summary, from Lemmas G.5, under our specified score function, we obtain the following lemma:

Lemma G.8. *For $P \in \mathcal{P}$ and $Q \in \mathrm{Alt}(P)$,*

$$\frac{1}{T}\mathbb{E}_Q[L_T^a] = \frac{\left(\varepsilon^a\right)^2}{2V_0^a(\kappa_{T,Q};\tau)} + O\left(\left(\varepsilon^a\right)^3\right).$$

G.10. Proof of Lemma 4.6: Derivation of a Lower Bound of the Probability of Misidentification

Here, we derive a lower bound for the probability of misidentification as follows, which is refined later:

Lemma G.9. *For any $P \in \mathcal{P}$, any consistent experiment satisfies*

$$\begin{aligned} &\lim_{\Delta_0 \to 0} \limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T \neq a_0^*) \\ &\leq \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q) < -(\mu_0^* - \mu_0^a)}} \limsup_{T\to\infty} \frac{1}{2\Omega_0^a(\kappa_{T,Q})} + o(1). \end{aligned}$$

Proof of Lemma G.9. For each $Q \in \mathrm{Alt}(P)$, $\mathbb{E}_Q[L_T] \geq \sup_{\mathcal{E}\in\mathcal{F}_T} d(\mathbb{P}_Q(\mathcal{E}), \mathbb{P}_P(\mathcal{E}))$ holds from Proposition G.1. Let $\mathcal{E} = \{\widehat{a}_T = a^*\}$. Because we assume that the experiment is consistent and asymptotically invariant for both models and from the definition of $\mathrm{Alt}(P)$, for each $\epsilon_1 \in (0,1)$ and $\epsilon_2 > 0$, there exists $t_0(\epsilon_1, \epsilon_2)$ such that for all $T \geq t_0(\epsilon_1)$, $\mathbb{P}_Q(\mathcal{E}) \leq \epsilon_1 \leq \mathbb{P}_{P_0}(\mathcal{E})$, and $\kappa_{T,Q}(a) \leq \kappa_{T,P}(a) + \epsilon_2$. Then, for all $T \geq t_0(\epsilon_1, \epsilon_2)$, $\mathbb{E}_Q[L_T] \geq d(\epsilon_1, 1 - \mathbb{P}_P(\widehat{a}_T \neq a_0^*)) = \epsilon \log \frac{\epsilon}{1-\mathbb{P}_{P_0}(\widehat{a}_T \neq a_0^*)} + (1-\epsilon_1)\log\frac{1-\epsilon_1}{\mathbb{P}_P(\widehat{a}_T \neq a_0^*)}$. Then, taking the limsup and letting $\epsilon_1, \epsilon_2 \to 0$,

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T \neq a_0^*) \leq \inf_{Q\in\mathrm{Alt}(P)} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_Q[L_T]$$

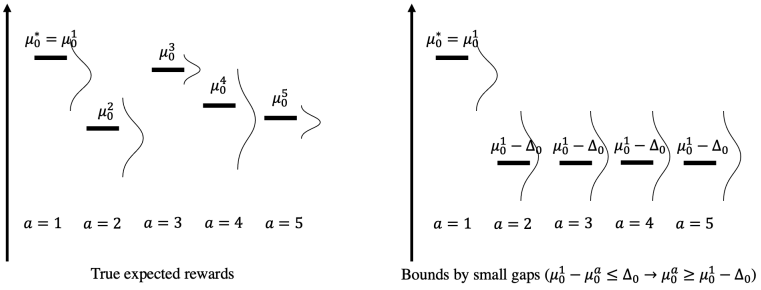

Figure 2. An idea in the derivation of the lower bounds. To lower bound the probability of misidentification, or equivalently upper bound $-\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T \neq a_0^*)$, it is sufficient to consider a case in the fight figure.

$$
\begin{aligned}
&\le \inf_{Q \in \mathrm{Alt}(P)} \limsup_{T\to\infty} \sum_{a\in[K]} \mathbb{E}_Q\left[\mathbb{E}_Q\left[\log \frac{f_Q^a(Y^a)\zeta_Q}{f_P^a(Y^a)\zeta_P}\right]\kappa_{T,Q}(a)\right] \\
&= \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P} \\ \mu^*(Q)-\mu^a(Q)<0}} \limsup_{T\to\infty} \mathbb{E}_Q\left[\mathbb{E}_Q\left[\log \frac{f_Q^a(Y^a)\zeta_Q}{f_P^a(Y^a)\zeta_P}\right]\kappa_{T,Q}(a)\right].
\end{aligned}
$$

For $\varepsilon^a(Q) = (\mu^*(Q) - \mu^a(Q)) - (\mu_0^* - \mu_0^a) < 0$, we have $\varepsilon^a(Q) < -(\mu_0^* - \mu_0^a) \Leftrightarrow \mu^*(Q) - \mu^a(Q) < 0$, where $\mu^*(Q) - \mu^a(Q) < 0$. Therefore,

$$
\begin{aligned}
&\min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P} \\ \mu^*(Q)-\mu^a(Q)<0}} \limsup_{T\to\infty} \sum_{a\in[K]} \mathbb{E}_Q\left[\mathbb{E}_Q\left[\log \frac{f_Q^a(Y_{a,t})\zeta_Q}{f_P^a(Y_{a,t})\zeta_P}\right]\kappa_{T,Q}(a)\right] \\
&= \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q)<-(\mu_0^*-\mu_0^a) \\ \forall b\in[K]\setminus\{a^*,a\}\ \varepsilon^b=0}} \limsup_{T\to\infty} \sum_{a\in[K]} \mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}}\left[\mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}}\left[\log \frac{f_{\boldsymbol{\varepsilon}}^a(Y_{a,t})\zeta_{\boldsymbol{\varepsilon}}}{f_P^a(Y_{a,t})\zeta_P}\right]\kappa_{T,Q}(a)\right] \\
&= \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q)<-(\mu_0^*-\mu_0^a)}} \limsup_{T\to\infty} \sum_{a\in\{a^*,a\}} \mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}}\left[\mathbb{E}_{\overline{R}_{\boldsymbol{\varepsilon}}}\left[\log \frac{f_{\boldsymbol{\varepsilon}}^a(Y_{a,t})\zeta_{\boldsymbol{\varepsilon}}}{f_P^a(Y_{a,t})\zeta_P}\right]\kappa_{T,Q}(a)\right] \\
&= \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q)<-(\mu_0^*-\mu_0^a)}} \limsup_{T\to\infty} \left\{\frac{(\varepsilon^a)^2}{2V_0^a(\kappa_{T,Q};\tau)} + O\left((\varepsilon^a)^3\right)\right\} \\
&\le \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q)<-(\mu_0^*-\mu_0^a)}} \limsup_{T\to\infty} \left\{\frac{(\mu_0^*-\mu_0^a)^2}{2V_0^a(\kappa_{T,Q};\tau)} - O\left((\mu_0^*-\mu_0^a)^3\right)\right\} \\
&\le \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q)<-(\mu_0^*-\mu_0^a)}} \limsup_{T\to\infty} \frac{(\mu_0^*-\mu_0^a)^2}{2V_0^a(\kappa_{T,Q};\tau)}.
\end{aligned}
$$

Then, as $\mu_0^* - \mu_0^a \to 0$, we obtain $V_0^a(w;\tau) \to \widetilde{V}_0^a(\kappa_{T,Q})$ by letting $\tau \to \infty$, which is the semiparametric efficiency bound in Lemmas G.6 and G.7. This also implies $1/V_0^a(w;\tau) = 1/\widetilde{V}_0^a(\kappa_{T,Q}) + o(1)$ as $\mu_0^* - \mu_0^a \to 0$.

Because all gaps $\mu_0^* - \mu_0^a$ are assumed to be upper bounded by Δ_a, we consider a situation where the expected outcomes of all suboptimal treatment arms are in $[\mu_0^* - \Delta_0, \mu_0^*)$. To obtain lower bounds, it is sufficient to consider a case where $\mu^b = \mu_0^* = \Delta_0$, under which the largest lower bounds are given (Figure 2).

Therefore, for $\Delta_0 > 0$ such that $\mu_0^* - \mu_0^a < \Delta_0$ for all $a \in [K]$,

$$
\begin{aligned}
&\lim_{\Delta_0\to 0} \limsup_{T\to\infty} -\frac{1}{\Delta_0^2 T}\log \mathbb{P}_{P_0}(\widehat{a}_T \neq a_0^*) \\
&\le \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \varepsilon^a(Q)<-(\mu_0^*-\mu_0^a)}} \limsup_{T\to\infty} \frac{1}{2\widetilde{V}_0^a(\kappa_{T,Q})} + o(1).
\end{aligned}
$$

From $\widetilde{V}_0^a(\kappa_{T,Q}) \geq \Omega_0^a(\kappa_{T,Q})$, the proof is complete. □

H. Proof of Lemma G.2

Proof.

$$\begin{aligned}\mathbb{E}_Q[L_T] &= \sum_{t=1}^{T} \mathbb{E}_Q \left[\sum_{a\in[K]} \mathbb{1}\{A_t = a\} \log \frac{f_Q^a(Y_t^a)}{f_P^a(Y_t^a)} \right] \\ &= \sum_{t=1}^{T} \mathbb{E}_Q^{\mathcal{F}_{t-1}} \left[\sum_{a\in[K]} \mathbb{E}_Q^{Y_{a,t},A_t} \left[\mathbb{1}[A_t = a] \log \frac{f_Q^a(Y_t^a)}{f_P^a(Y_t^a)} | \mathcal{F}_{t-1} \right] \right] \\ &= \sum_{a\in[K]} \mathbb{E}_Q^{Y^a} \left[\log \frac{f_Q^a(Y^a)}{f_P^a(Y^a)} \right] \sum_{t=1}^{T} \mathbb{E}_Q^{\mathcal{F}_t} \left[\mathbb{E}_Q \left[\mathbb{1}[A_t = a] | \mathcal{F}_{t-1} \right] \right]\end{aligned}$$

where $\mathbb{E}_Q^Z$ denotes an expectation of random variable Z over the distribution Q. We used that the observations $(Y_{1,t}, \ldots, Y_{K,t})$ are i.i.d. across $t \in \mathcal{T}$. □

I. Proof of Lemma G.3

Proof. For the parametric submodel of the observed-data statistical models, the log-likelihood for the observed data is

$$\log \overline{r}_{\boldsymbol{\varepsilon}}^{\kappa}(y, d) = \log \int \kappa(d) p_{\boldsymbol{\varepsilon}}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d,$$

where note that $p_{\boldsymbol{\varepsilon}}(H^d(\tau^d, v^d)) = p_{\boldsymbol{\varepsilon}}(y^*, y^a)$. Then, for $d \in \{a^*, a\}$,

$$\begin{aligned}S^a(y, d) &= \frac{\partial}{\partial \varepsilon^a} \left[\log \int \kappa(d) p_{\boldsymbol{\varepsilon}}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d \right] \Bigg|_{\varepsilon^a = 0} \\ &= \frac{\int \frac{\partial}{\partial \varepsilon^a} \kappa(d) p_{\boldsymbol{\varepsilon}}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d}{\int \kappa(d) p_{\boldsymbol{\varepsilon}}(y^*, y^a) p_{\boldsymbol{\varepsilon}}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d}. \end{aligned} \tag{28}$$

Dividing and multiplying by $p_{\boldsymbol{\varepsilon}}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d$ in the integral of the numerator of (28) yields

$$\begin{aligned}&\frac{\int \frac{\partial}{\partial \varepsilon^a} \kappa(d) p_{\boldsymbol{\varepsilon}:\varepsilon^a=0}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d}{\int \kappa(d) p_{\boldsymbol{\varepsilon}:\varepsilon^a=0}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d} \\ &= \frac{\int g^a(\phi_\tau^*(y), \phi_\tau^a(y)) p_{\boldsymbol{\varepsilon}:\varepsilon^a=0}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d}{\int \kappa(d) p_{\boldsymbol{\varepsilon}:\varepsilon^a=0}(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d} \\ &= \frac{\int g^a(\phi_\tau^*(y), \phi_\tau^a(y)) p_P(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d}{\int \kappa(d) p_P(H^d(\tau^d, v^d)) J(\tau^d, v^d) \mathrm{d}v^d}.\end{aligned}$$

Hence,

$$S^a(y, d) = \mathbb{E}_P \left[g^a(\phi_\tau^*(Y_t^*), \phi_\tau^a(Y_{a,t})) | \mathcal{T}^d(Y_t^*, Y_{a,t}) = (y) \right]$$

This concludes the proof. □

J. Proofs of Theorem 4.7

Proof. From Lemma 4.6, if there exists $\Delta_0 > 0$ such that $\mu_0^* - \mu_0^a \leq \Delta_0$ for all $a \in [K]$, the lower bounds are characterized by

$$\max_{w\in\mathcal{W}} \min_{a\neq a_0^*} \frac{1}{\frac{\left(\sigma_0^*\right)^2}{w(a_0^*)} + \frac{\left(\sigma_0^a\right)^2}{w(a)}}.$$

We consider maximising $R > 0$ such that $R \leq 1/2\frac{(\sigma_0^*)^2}{w(a_0^*)} + \frac{(\sigma_0^a)^2}{w(a)}$ for all $a \in [K]\backslash\{a_0^*\}$ by optimizing $w \in \mathcal{W}$. That is, we consider the following non-linear programming:

$$\begin{aligned}
&\max_{R>0, \boldsymbol{w}=\{w(1),w(2)\ldots,w(K)\}\in(0,1)^K} \quad R \\
&\text{s.t.} \quad R\left(\frac{(\sigma_0^*)^2}{w(a_0^*)} + \frac{(\sigma_0^a)^2}{w(a)}\right)\zeta - 1 \leq 0 \qquad \forall a \in [K]\backslash\{a_0^*\}, \\
&\quad \sum_{a\in[K]} w(a) - 1 = 0, \\
&\quad w(a) > 0 \qquad \forall a \in [K].
\end{aligned}$$

The maximum of R in the constraint optimization is equal to $\max_{w\in\mathcal{W}} \min_{a\neq a_0^*} \frac{1}{\frac{(\sigma_0^*)^2}{w(a_0^*)} + \frac{(\sigma_0^a)^2}{w(a)}}$.

Then, for $(K-1)$ Lagrangian multiplies $\boldsymbol{\lambda} = \{\lambda^a\}_{a\in[K]\backslash\{a_0^*\}}$ and γ such that $\lambda^a \leq 0$ and $\gamma \in \mathbb{R}$, we define the following Lagrangian function:

$$\begin{aligned}
&L(\boldsymbol{\lambda}, \boldsymbol{\gamma}; R, \boldsymbol{w}) \\
&= R + \sum_{a\in[K]\backslash\{a_0^*\}} \lambda^a \left\{ R\left(\frac{(\sigma_0^*)^2}{w(a_0^*)} + \frac{(\sigma_0^a)^2}{w(a)}\right) - 1 \right\} - \gamma \left\{ \sum_{a\in[K]} w(a) - 1 \right\} \\
&= R + \sum_{a\in[K]\backslash\{a_0^*\}} \lambda^a \left\{ R\left(\frac{(\sigma_0^*)^2}{w(a_0^*)} + \frac{(\sigma_0^a)^2}{w(a)}\right) - 1 \right\} - \gamma \left\{ \sum_{a\in[K]} w(a) - 1 \right\}.
\end{aligned}$$

Note that the objective (R) and constraints ($R\left(\frac{(\sigma_0^*)^2}{w(a_0^*)} + \frac{(\sigma_0^a)^2}{w(a)}\right) - 1 \leq 0$ and $\sum_{a\in[K]} w(a) - 1 = 0$) are differentiable convex functions for R and $\boldsymbol{w}$. Therefore, the global optimizer R^* and $\boldsymbol{w}^* = \{w^*(a)\} \in (0,1)^{KN}$ satisfies the KKT condition; that is, there are Lagrangian multipliers λ^{a*}, γ^*, and R^* such that

$$1 + \sum_{a\in[K]\backslash\{a_0^*\}} \lambda^{a*}\left(\frac{(\sigma_0^*)^2}{w^*(a_0^*)} + \frac{(\sigma_0^a)^2}{w^*(a)}\right) = 0 \tag{29}$$

$$-2\sum_{a\in[K]\backslash\{a_0^*\}} \lambda^{a*} R^* \frac{(\sigma_0^*)^2}{(w^*(a_0^*))^2} = \gamma^* \tag{30}$$

$$-2\lambda^{a*} R^* \frac{(\sigma_0^a)^2}{(w^*(a))^2} = \gamma^* \qquad \forall a \in [K]\backslash\{a_0^*\} \tag{31}$$

$$\lambda^{a*}\left\{ R^*\left(\frac{(\sigma_0^*)^2}{w^*(a_0^*)} + \frac{(\sigma_0^a)^2}{w^*(a)}\right) - 1 \right\} = 0 \qquad \forall a \in [K]\backslash\{a_0^*\} \tag{32}$$

$$\gamma^*\left\{ \sum_{c\in[K]} w^*(c) - 1 \right\} = 0$$

$$\lambda^{a*} \leq 0 \qquad \forall a \in [K]\backslash\{a_0^*\}. \tag{33}$$

Here, (29) implies $\lambda^{a*} < 0$ for some $a \in [K]\backslash\{a_0^*\}$. This is because if $\lambda^{a*} = 0$ for all $a \in [K]\backslash\{a_0^*\}$, $1 + 0 = 1 \neq 0$.

With $\lambda^{a*} < 0$, since $-\lambda^{a*} R^* \frac{(\sigma_0^a)^2}{(w^*(a))^2} > 0$ for all $a \in [K]$, it follows that $\gamma^* > 0$. This also implies that $\sum_{c\in[K]} w^{c*} - 1 = 0$. Then, (32) implies that

$$R^*\left(\frac{(\sigma_0^*)^2}{w^*(a_0^*)} + \frac{(\sigma_0^a)^2}{w^*(a)}\right) = 1 \qquad \forall a \in [K]\backslash\{a_0^*\}.$$

Therefore, we have

$$\frac{\left(\sigma_0^a\right)^2}{w^*(a)} = \frac{\left(\sigma_0^b\right)^2}{w^*(b)} \qquad \forall a, b \in [K]\backslash\{a_0^*\}. \tag{34}$$

Let $\frac{\left(\sigma_0^a\right)^2}{w^*(a)} = \frac{\left(\sigma_0^b\right)^2}{w^*(b)} = \frac{1}{R^*} - \frac{\left(\sigma_0^*\right)^2}{w^*(a_0^*)} = U$. From (34) and (29),

$$\sum_{b\in[K]\backslash\{a_0^*\}} \lambda^{b*} = -\frac{1}{\frac{\left(\sigma_0^*\right)^2}{w^*(a_0^*)} + U} \tag{35}$$

From (30) and (31),

$$\frac{\left(\sigma_0^*\right)^2}{(w^*(a_0^*))^2} \sum_{b\in[K]\backslash\{a_0^*\}} \lambda^{b*} = \lambda^{a*} \frac{\left(\sigma_0^a\right)^2}{(w^*(a))^2} \qquad \forall a \in [K]\backslash\{a_0^*\}. \tag{36}$$

From (35) and (36),

$$-\frac{\left(\sigma_0^*\right)^2}{(w^*(a_0^*))^2} = \lambda^{a*} \frac{\left(\sigma_0^a\right)^2}{(w^*(a))^2} \left(\frac{\left(\sigma_0^*\right)^2}{w^*(a_0^*)} + U \right) \qquad \forall a \in [K]\backslash\{a_0^*\}. \tag{37}$$

From (29) and (37),

$$w^*(a_0^*) = \sqrt{\left(\sigma_0^*\right)^2 \sum_{a\in[K]\backslash\{a_0^*\}} \frac{(w^*(a))^2}{\left(\sigma_0^a\right)^2}}.$$

In summary, we have the following KKT conditions:

$$\begin{aligned}
& w^*(a_0^*) = \sqrt{\left(\sigma_0^*\right)^2 \sum_{a\in[K]\backslash\{a_0^*\}} \frac{(w^*(a)^2}{\left(\sigma_0^a\right)^2}} \\
& \frac{\left(\sigma_0^*\right)^2}{(w^*(a_0^*))^2} = -\lambda^{a*} \frac{\left(\sigma_0^a\right)^2}{(w^*(a))^2} \left(\left(\frac{\left(\sigma_0^*\right)^2}{w^*(a_0^*)} + \frac{\left(\sigma_0^a\right)^2}{w^*(a)} \right) \right) \qquad \forall a \in [K]\backslash\{a_0^*\} \\
& -\lambda^{a*} \frac{\left(\sigma_0^a\right)^2}{(w^*(a))^2} = \widetilde{\gamma}^* \qquad \forall a \in [K]\backslash\{a_0^*\} \\
& \frac{\left(\sigma_0^a\right)^2}{w^*(a)} = \frac{1}{R^*} - \frac{\left(\sigma_0^*\right)^2}{w^*(a_0^*)} \qquad \forall a \in [K]\backslash\{a_0^*\} \\
& \sum_{a\in[K]} w^*(a) = 1 \\
& \lambda^{a*} \le 0 \qquad \forall a \in [K]\backslash\{a_0^*\},
\end{aligned}$$

where $\widetilde{\gamma}^* = \gamma^*/2R^*$. From $w^*(a_0^*) = \sqrt{\left(\sigma_0^*\right)^2 \sum_{a\in[K]\backslash\{a_0^*\}} \frac{(w^*(a))^2}{\left(\sigma_0^a\right)^2}}$ and $-\lambda^{a*} \frac{\left(\sigma_0^a\right)^2}{(w^*(a))^2} = \widetilde{\gamma}^*$, we have

$$\begin{aligned}
w^*(a_0^*) &= \sigma_0^* \sqrt{\sum_{a\in[K]\backslash\{a_0^*\}} -\lambda^{a*}} / \sqrt{\widetilde{\gamma}^*} \\
w(a) &= \sqrt{-\lambda^{a*}/\widetilde{\gamma}^*}\sigma_0^a.
\end{aligned}$$

From $\sum_{a\in[K]} w^*(a) = 1$,

$$\sigma_0^* \sqrt{\sum_{a\in[K]\backslash\{a_0^*\}} -\lambda^{a*}} / \sqrt{\widetilde{\gamma}^*} + \sum_{a\in[K]\backslash\{a_0^*\}} \sqrt{-\lambda^{a*}/\widetilde{\gamma}^*}\sigma_0^a = 1.$$

Therefore,

$$\sqrt{\widetilde{\gamma}^*} = \sigma_0^* \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} -\lambda^{a*}} + \sum_{a\in[K]\setminus\{a_0^*\}} \sqrt{-\lambda^{a*}}\sigma_0^a.$$

Hence,

$$w^*(a_0^*) = \frac{\sigma_0^* \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} -\lambda^{a*}}}{\sigma_0^* \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} -\lambda^{a*}} + \sum_{a\in[K]\setminus\{a_0^*\}} \sqrt{-\lambda^{a*}}\sigma_0^a}$$
$$w(a) = \frac{\sqrt{-\lambda^{a*}}\sigma_0^a}{\sigma_0^* \sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} -\lambda^{a*}} + \sum_{a\in[K]\setminus\{a_0^*\}} \sqrt{-\lambda^{a*}}\sigma_0^a},$$

where from $\frac{(\sigma_0^*)^2}{(w^*(a_0^*))^2} = -\lambda^{a*}\frac{(\sigma_0^a)^2}{(w^*(a))^2}\left(\frac{(\sigma_0^*)^2}{w^*(a_0^*)} + \frac{(\sigma_0^a)^2}{w^*(a)}\right)$, $(\lambda^{a_0^*})_{a\in[K]\setminus\{a_0^*\}}$ satisfies,

$$\begin{aligned}\frac{1}{\sum_{a\in[K]\setminus\{a_0^*\}} -\lambda^{a*}} &= \left(\frac{\sigma_0^*}{\sqrt{\sum_{a\in[K]\setminus\{a_0^*\}} -\lambda^{a*}}} + \frac{\sigma_0^a}{\sqrt{-\lambda^{a*}}}\right)\left(\sigma_0^* \sqrt{\sum_{c\in[K]\setminus\{a_0^*\}} -\lambda^{c*}} + \sum_{c\in[K]\setminus\{a_0^*\}} \sqrt{-\lambda^{c*}}\sigma_0^c\right) \\ &= \left(\sigma_0^* + \frac{\sigma_0^a}{\sqrt{-\lambda^{a*}}}\sqrt{\sum_{c\in[K]\setminus\{a_0^*\}} -\lambda^{c*}}\right)\left(\sigma_0^* + \frac{\sum_{c\in[K]\setminus\{a_0^*\}} \sqrt{-\lambda^{c*}}\sigma_0^c}{\sum_{c\in[K]\setminus\{a_0^*\}} -\lambda^{c*}}\sqrt{\sum_{c\in[K]\setminus\{a_0^*\}} -\lambda^{c*}}\right).\end{aligned}$$

Then, the following solutions satisfy the above KKT conditions:

$$\begin{aligned} R^*\left(\sigma_0^* + \sqrt{\sum_{b\in[K]\setminus\{a_0^*\}} \left(\sigma_0^b\right)^2}\right)^2 &= 1 \\ w^*(a_0^*) &= \frac{\sigma_0^* \sqrt{\sum_{b\in[K]\setminus\{a_0^*\}} \left(\sigma_0^b\right)^2}}{\sigma_0^* \sqrt{\sum_{b\in[K]\setminus\{a_0^*\}} \left(\sigma_0^b\right)^2} + \sum_{b\in[K]\setminus\{a_0^*\}} \left(\sigma_0^b\right)^2} \\ w^*(a) &= \frac{\left(\sigma_0^a\right)^2}{\sigma_0^* \sqrt{\sum_{b\in[K]\setminus\{a_0^*\}} \left(\sigma_0^b\right)^2} + \sum_{b\in[K]\setminus\{a_0^*\}} \left(\sigma_0^b\right)^2} \\ \lambda^{a*} &= -\left(\sigma_0^a\right)^2 \\ \gamma^* &= 2\left(\sigma_0^a\right)^2. \end{aligned}$$

□

Note that a target allocation ratio w in the maximum corresponds to a limit of an expectation of sampling rule $\frac{1}{T}\sum_{t=1}^T \mathbb{1}[A_t = a]$ from the definition of asymptotically invariant strategies.

K. Asymptotic Optimality for BAI

K.1. Proof of Theorem 4.9

We can prove Corollary 4.9 by using a proof procedure similar to Theorem 4.7.

Proof. We consider solving

$$\max_{a\in[K], w\in\mathcal{W}} \min_{b\in[K], b\neq a} \frac{1}{2\left(\frac{\left(\sigma_0^a\right)^2}{w(a)} + \frac{\left(\sigma_0^b\right)^2}{w(b)}\right)}.$$

We solved $\max_{w\in\mathcal{W}}\min_{a\neq a_0^*}\frac{1}{2\left(\frac{(\sigma_0^*)^2}{w(a_0^*)}+\frac{(\sigma_0^a)^2}{w(a)}\right)}$ in the proof of Theorem 4.7. Similarly, we solve this $\max_{w\in\mathcal{W}}\min_{a\neq b}\frac{1}{2\left(\frac{(\sigma_0^a)^2}{w(a)}+\frac{\left(\sigma_0^b\right)^2}{w(b)}\right)}$.

Therefore, we consider the following non-linear programming: To solve this problem, we consider maximising $R>0$ by solving

$$\begin{aligned}
&\max_{R>0,\boldsymbol{w}=\{w(a)\}\in(0,1)^K} R\\
\text{s.t.}\quad & R\left(\frac{\left(\sigma_0^a\right)^2}{w(a)}+\frac{\left(\sigma_0^b\right)^2}{w(b)}\right)-1\leq 0 \qquad \forall a\neq b\in[K],\\
&\sum_{a\in[K]} w(a)-1=0,\\
&w(a)>0 \qquad \forall a\in[K].
\end{aligned}$$

Then, for ${}_KC_2$ Lagrangian multiplies $\boldsymbol{\lambda}=\{\lambda_a^b\}_{a\in[K],b\in[K]:b>a}$, and γ such that $\lambda_b^a\leq 0$ and $\gamma\in\mathbb{R}$, we define the following Lagrangian function:

$$L(\boldsymbol{\lambda},\gamma;R,\boldsymbol{w})=R+\sum_{a\in[K]}\sum_{b\in[K]:b>a}\lambda_a^b\left\{R\left(\frac{\left(\sigma_0^a\right)^2}{w(a)}+\frac{\left(\sigma_0^b\right)^2}{w(b)}\right)-1\right\}-\gamma\left\{\sum_{a\in[K]}w(a)-1\right\}.$$

Note that the objective (R) and constraints are differentiable convex functions for R and $\boldsymbol{w}$. Therefore, the global optimizer R^* and $\boldsymbol{w}^*=\{w^*(a)\}\in(0,1)^K$ satisfies the KKT condition; that is, there are Lagrangian multipliers λ_a^{b*}, γ^*, and R^* such that

$$\begin{aligned}
&1+\sum_{a\in[K]}\sum_{b\in[K]:b>a}\lambda_a^{b*}\left(\frac{\left(\sigma_0^a\right)^2}{w^*(a)}+\frac{\left(\sigma_0^b\right)^2}{w^*(b)}\right)=0\\
&-2\sum_{b\in[K]\setminus\{c\}}\lambda_c^{b*}R^*\frac{\left(\sigma_0^c\right)^2}{(w^*(c))^2}-2\sum_{a\in[K]\setminus\{c\}}\lambda_a^{c*}R^*\frac{\left(\sigma_0^c\right)^2}{(w^*(c))^2}=\gamma^* \qquad \forall c\in[K]\\
&\lambda_a^{b*}\left\{R^*\left(\frac{\left(\sigma_0^a\right)^2}{w^*(a)}+\frac{\left(\sigma_0^b\right)^2}{w^*(b)}\right)-1\right\}=0 \qquad \forall a\in[K],\ \forall b\in[K]:b>a,\\
&\gamma^*\left\{\sum_{a\in[K]}w^*(a)-1\right\}=0\\
&\lambda_b^{a*}\leq 0 \qquad \forall a\in[K],\ \forall b\in[K]:b>a.
\end{aligned}$$

The solution differs according to the number of the treatment arms K. When $K=2$, $w^*(a)=\frac{\sigma_0^a}{\sigma_0^1+\sigma_0^2}$. When $K\geq 3$, we could not obtain a closed-form solution except for the following specific case.

Lower bounds for multi-armed equal-variance statistical models. Here, we show the second statement of Corollary M.2, where $\left(\sigma_0^1\right)^2=\cdots=\left(\sigma^K\right)^2=(\sigma)^2$. In this case, solutions that satisfy the KKT conditions are given as

$$\begin{aligned}
w^*(a)&=\frac{1}{K} \qquad \forall a\in[K],\\
R^*&=\frac{1}{2K\left(\sigma\right)^2},\\
\lambda_a^{b*}&=-\frac{R^*}{K-1} \qquad \forall a\neq b\in[K],\\
\gamma^*&=2K^2(R^*)^2\left(\sigma\right)^2.
\end{aligned}$$

□

L. Proof of Theorem D.1

The proof follows those in van der Laan (2008), Hahn et al. (2011), and Kato et al. (2020). We specifically follow the proof procedure of Proposition 1 of Kato et al. (2020).

Proof. Let $w \in (0,1)$. From Lemma 4.6,

$$\begin{aligned}
&\min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \text{s.t. } \mu^*(Q)-\mu^a(Q)<0}} \limsup_{T\to\infty} \frac{1}{2\Omega_0^a(\kappa_{T,Q}^\pi)} \\
&= \min_{a\in[K]\setminus\{a_0^*\}} \inf_{\substack{Q\in\mathcal{P}^* \\ \text{s.t. } \mu^*(Q)-\mu^a(Q)<0}} \sup_{w\in(0,1)} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)} \\
&= \sup_{w\in(0,1)} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)},
\end{aligned}$$

If there exists $\max_{w\in(0,1)} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)}$, we have

$$\sup_{w\in(0,1)} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)} = \max_{w\in(0,1)} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)}.$$

Here, let w^* be an point of maximum. Then, it holds that

$$\max_{w\in(0,1)} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)} = \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w^*}+\frac{\left(\sigma_0^2\right)^2}{1-w^*}\right)}.$$

Let us consider finding minimum of $2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)$. Obviously, as $1/z$ is strictly decreasing for $z>0$, then minimum will be at the point of maximum of z. Therefore,

$$\arg\max_{w\in\mathcal{W}} \frac{1}{2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)} = \arg\min_{w\in\mathcal{W}} 2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right).$$

Then, instead of the maximization problem, we consider

$$\min_{w\in\mathcal{W}} 2\left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right)$$

Therefore, let us define the following function $b:\mathcal{W}\to\mathbb{R}$:

$$b(w) = \left(\frac{\left(\sigma_0^1\right)^2}{w}+\frac{\left(\sigma_0^2\right)^2}{1-w}\right).$$

We consider minimizing $b(w)$ by minimizing $\tilde{b}(q) = \frac{\left(\sigma_0^1\right)^2}{q}+\frac{\left(\sigma_0^2\right)^2}{1-q}$ for $q\in(0,1)$. The first derivative of $\tilde{b}(q)$ with respect to q is given as follows:

$$\tilde{b}'(q) = -\frac{\left(\sigma_0^1\right)^2}{q^2}+\frac{\left(\sigma_0^2\right)^2}{(1-q)^2}.$$

The second derivative of $\widetilde{b}(q)$ is given as

$$\tilde{b}''(q) = 2\frac{\left(\sigma_0^1\right)^2}{q^3} + 2\frac{\left(\sigma_0^2\right)^2}{(1-q)^3}.$$

For $q \in (0,1)$, because $\tilde{b}''(q) > 0$, the minimizer q^* of $\tilde{b}$ satisfies the following equation:

$$-\frac{\left(\sigma_0^1\right)^2}{(q^*)^2} + \frac{\left(\sigma_0^2\right)^2}{(1-q^*)^2} = 0.$$

This equation is equivalent to

$$\begin{aligned}
&-(q^*)^2 e(0) + (1-q^*)^2 e(1) = 0 \\
\Leftrightarrow\ & q^* \sqrt{\left(\sigma_0^2\right)^2} = (1-q^*)\sqrt{\left(\sigma_0^1\right)^2} \\
\Leftrightarrow\ & q^* = \frac{\sqrt{\left(\sigma_0^1\right)^2}}{\sqrt{\left(\sigma_0^1\right)^2} + \sqrt{\left(\sigma_0^2\right)^2}}.
\end{aligned}$$

Therefore,

$$q^* = \frac{\sigma_0^1}{\sigma_0^1 + \sigma_0^2}.$$

□

M. Lower Bounds for Multi-Armed Equal-Variance Statistical Models

As a generalization of statistical models with potential outcomes adhering to one-parameter distributions, such as Bernoulli, Binomial, and Gamma distributions, we define the equal-variance statistical models.

Definition M.1 (Equal-variance statistical models)**.** statistical models $\mathcal{P}^{\mathrm{E}}$ are equal-variance statistical models if for local location-shift statistical model $\mathcal{P}^*$, $\sigma^1 = \sigma^2 = \cdots = \sigma^K = \sigma$ for any $x \in \mathcal{X}$, where σ is a constant.

When outcomes follow Bernoulli distributions, the statistical model belongs to the equal-variance statistical models because the variances are the same when the expected outcomes are the same. For this class, the lower bound are given as follows. We omit the proof because we just substitute $\sigma = \sigma_0^1 = \cdots = \sigma^K$ for Theorems D.1–4.9.

Corollary M.2 (Lower bounds for multi-armed equal-variance statistical models)**.** *Let $\Delta_0 > 0$ be a constant independent from T such that $\mu_0^* - \mu_0^a \leq \Delta_0$.*

- *Let Π be a class of consistent strategies (Definition 4.1). For $K = 2$, any $P \in \mathcal{P}^*$ and $\pi \in \Pi$,,*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \frac{1}{8\mathbb{E}_P\left[\left(\sigma\right)^2\right]} + o(1),$$

  where the target allocation ratio is given as

$$w^*(1) = w^*(2) = 1/2 \qquad \forall x \in \mathcal{X}.$$

- *Let Π be a class of consistent (Definition 4.1) and asymptotically invariant (Definition 4.3) strategies. For $K \geq 3$, any $P \in \mathcal{P}^*$ and $\pi \in \Pi$,*

$$\begin{aligned}
&\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \\
&\leq \frac{1}{2\left(1+\sqrt{K-1}\right)^2 \mathbb{E}_P\left[\left(\sigma\right)^2\right]} + o(1),
\end{aligned}$$

where the target allocation ratio is given as

$$w^*(a_0^*) = \frac{1}{1+\sqrt{K-1}},$$
$$w^*(a) = \frac{1}{(1+\sqrt{K-1})\sqrt{K-1}}, \qquad \forall a \in [K].$$

- *For $K \geq 3$, any $P \in \mathcal{P}^*$, any consistent (Definition 4.1) experiment π satisfies*

$$\limsup_{T\to\infty} -\frac{1}{T}\log \mathbb{P}_{P_0}(\widehat{a}_T^\pi \neq a_0^*) \leq \frac{1}{2K\mathbb{E}_P\left[(\sigma)^2\right]} + o(1),$$

where the target allocation ratio is given as

$$w^*(a) = 1/K \qquad \forall a \in [K], \qquad \forall x \in \mathcal{X}.$$

Because the variances are equal across treatment arms, the target allocation ratio is also equal across treatment arms. This lower bound and the target allocation ratio implies that the uniform-EBA experiment is optimal, where we choose each treatment arm with the same probability (the uniform sampling rule) and recommend a treatment arm with the highest sample average of observed outcomes (the empirical best arm (EBA) recommendation rule). The fact that the uniform-EBA experiment is approximately optimal for two-armed Bernoulli bandits is also reported by Kaufmann et al. (2016).

N. TS-EBA experiment

The proof of the upper bound is detailed in Section N.1. A modified version of the TS-EBA experiment is discussed in Section N.2.

N.1. Proof of the Upper Bound (Theorem 6.1)

Let us define $\widehat{\Delta}_T^{\mathrm{HIR},a,b} = \widehat{\mu}_T^{\mathrm{EBA},a_0^*} - \widehat{\mu}_T^{\mathrm{EBA},a}$ for all $a \in]K]$ The proof of the upper bound is based on that of the TS-HIR strategy discussed in Kato et al. (2023a). For the sake of completeness, we provide the proof herein. To prove the upper bound, we employ the following result from Hahn et al. (2011).

Proposition N.1 (Asymptotic normality of the SA estimator. From Theorem 1 of Hahn et al. (2011))**.** *Assume that w^* smoothly depends on $(\sigma_0^a)_{a\in[K]}$. Then,*

$$\sqrt{T}\left(\widehat{\Delta}_T^{\mathrm{HIR},a} - \Delta_0^a\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \Omega_0^a(w^*)\right),$$

where

$$\Omega_0^a(w^*) = \frac{(\sigma_0^*)^2}{w^*(a_0^*)} + \frac{(\sigma_0^a)^2}{w^*(a)}.$$

We also use the following result from Hayashi (2000).

Proposition N.2 (Convergence in distribution and in moments. Lemma 2.1 of Hayashi (2000))**.** *Let $\alpha_{s,n}$ be the s-th moment of z_n, and $\lim_{n\to\infty}\alpha_{s,n} = \alpha_s$, where α_s is finite. Suppose that for some $\delta > 0$, $\mathbb{E}\left[|z_n|^{s+\delta}\right] < M < \infty$ for all n and a constant $M > 0$ independent of n. If $z_n \xrightarrow{\mathrm{d}} z$, then α_s is the s-th moment of z.*

Then, we show Theorem 6.1 as follows.

Proof. Let us define

$$\xi_T^a = \frac{\sqrt{T}\left(\widehat{\Delta}_T^{\mathrm{HIR},a} - \Delta_0^a\right)}{\Omega_0^a(w^*)}.$$

By applying the Chernoff bound, for any $v \geq 0$ and any $\lambda < 0$,

$$\mathbb{P}_P\left(\widehat{\Delta}_T^{\mathrm{HIR},a} - \Delta_0^a \leq v\right) \leq \mathbb{E}_P\left[\exp\left(\lambda\sqrt{T}\xi_T^a(P)\right)\right]\exp\left(-\lambda T v\right).$$

Algorithm 2 TS-EBA experiment with round-robin.

Parameter: Hypothetical best treatment arm $\widetilde{a} \in [K]$. The sample splitting ratio $r \in (0,1)$.
Initialization: for $t = 1$ **do** Draw $A_t = t$. For each $a \in [K]$, set $\widehat{w}_t(a) = 1/K$. **end for**
Stage 1:
for $t = K+1$ to $\lceil rT \rceil$ **do**
  Draw a treatment arm a with probability $w^{(1)}$.
end for
Construct $\widehat{w}$ as (6) by estimating the variances.
Stage 2:
for $t = \lceil rT \rceil + 1$ to T **do**
  $A_t \in \arg\min_{a \in [K]} \left\{ \frac{1}{t-1} \sum_{s=1}^{t-1} \mathbb{1}[A_s = a] - \widehat{w}(a) \right\}$.
end for
Construct $\widehat{\mu}_T^{\mathrm{SA},a}$ for each $a \in [K]$.
Recommend $\widehat{a}_T^{\mathrm{SA}} = \arg\max_{a \in [K]} \widehat{\mu}_T^{\mathrm{SA},a}$.

By applying the Taylor series expansion for $\log \mathbb{E}_P \left[\exp\left(\lambda \sqrt{T} \xi_T^a \right) \right]$ around $\frac{\lambda}{\sqrt{T}} = 0$,

$$\log \mathbb{E}_P \left[\exp\left(\lambda \sqrt{T} \xi_T^a \right) \right] = \sqrt{T} \lambda \mathbb{E}_P \left[\xi_T^a \right] + \frac{T\lambda^2}{2} \mathbb{E}_P \left[\left(\xi_T^a \right)^2 \right] + \sum_{n=3}^{\infty} \frac{(\sqrt{T}\lambda)^n}{n!} c_{n,T},$$

where $c_{n,T}$ is the n-th cumulant of ξ_T^a. From Lemma 2.1 of Hayashi (2000) (Proposition N.2 in Appendix), Proposition N.1, we have $\lim_{T \to \infty} \mathbb{E}\left[\xi_T^a \right] = 0$, $\lim_{T \to \infty} \mathbb{E}\left[\left(\xi_T^a \right)^2 \right] = 1$, and $\lim_{T \to \infty} \mathbb{E}\left[\left(\xi_T^a \right)^n \right] = m_n$ for all $n \geq 3$, where m_n is the n-th moments. Then, we have $\lim_{T \to \infty} c_{n,T} = 0$ because cumulants of centered normal distributions are zero except for the second-order cumulant. Here, note that $\lim_{T \to \infty} \sum_{n=3}^{\infty} \frac{(\sqrt{T}\lambda)^{n-2}}{n!} = -\frac{1}{2}$. Therefore, for any $v, \varepsilon > 0$, there exist $T_0 > 0$ such that for all $T > T_0$,

$$\mathbb{P}_P \left(\sum_{t=1}^{T} \xi_t^a \leq v \right) \leq \exp\left(\frac{T\lambda^2}{2} - T\lambda v - \left\{ \sqrt{T}\lambda + T\lambda^2/2 \right\} \varepsilon \right).$$

By substituting $\lambda = v = -\frac{\Delta_0^a}{\sqrt{\Omega_0^a(w^*)}} < 0$, the claim follows. □

N.2. TS-EBA experiments with Round-Robin

We introduce the TS-EBA experiment that utilizes a round-robin-based sampling rule; in other words, we draw treatment arms that have been selected the least number of times, represented by $A_t \in \arg\min_{a \in [K]} \left\{ \frac{1}{t-1} \sum_{s=1}^{t-1} \mathbb{1}[A_s = a] - \widehat{w}(a) \right\}$. This approach is often used in existing studies, such as the α-Elimination strategy found in Kaufman et al. (2016). The pseudo-code for this method is depicted in Algorithm 2.