Asymptotically Optimal and Computationally Efficient Average Treatment Effect Estimation in A/B testing

Vikas Deep [1] **Achal Bassamboo** [1] **Sandeep Juneja** [2]

Abstract

Motivated by practical applications in clinical trials and online platforms, we study A/B testing with the aim of estimating a confidence interval (CI) for the average treatment effect (ATE) using the minimum expected sample size. This CI should have a width at most ϵ while ensuring that the probability of the CI not containing the true ATE is at most δ. To answer this, we first establish a lower bound on the expected sample size needed for any adaptive policy which constructs a CI of ATE with desired properties. Specifically, we prove that the lower bound is based on the solution to a non-convex max-min optimization problem for small δ. Tailoring the "plug-in" approach for the ATE problem, we construct an adaptive policy that is asymptotically optimal, i.e., matches the lower bound on the expected sample size for small δ. Interestingly, we find that, for small ϵ and δ, the asymptotically optimal fraction of treatment assignment for A and B is proportional to the standard deviation of the outcome distributions of treatments A and B, respectively. However, as the proposed approach can be computationally intensive, we propose an alternative adaptive policy. This new policy, informed by insights from our lower bound analysis, is computationally efficient while remaining asymptotically optimal for small values of ϵ and δ. Numerical comparisons demonstrate that both policies perform similarly across practical values of ϵ and δ, offering efficient solutions for A/B testing.

[1]Kellogg School of Management, Northwestern University, Evanston, IL 60201 [2]Ashoka University, Sonipat, Haryana, India. Correspondence to: Vikas Deep <vikas.deep@kellogg.northwestern.edu>, Achal Bassamboo <achalb@kellogg.northwestern.edu>, Sandeep Juneja <sandeep.juneja2010@gmail.com>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

1. Introduction

The simplest controlled experiment where two variants are compared is referred to as an A/B test. In an A/B test, individuals arrive sequentially, the experiment designer assigns an arriving individual to either treatment A or treatment B and measures the response. There is a long history of use of A/B tests in clinical trials to assess the efficacy of a drug (treatment A) relative to another drug (treatment B). In recent years, A/B testing has gained widespread adoption among large-scale online platforms for assessing the performance of new product designs, web page layouts, or services (see Kohavi & Thomke (2017)). A typical objective of a sequential A/B test is to infer the better treatment. This inference problem is well-studied in the literature as the best arm identification or best treatment identification (BTI) problem (see Bubeck et al. (2011), Garivier & Kaufmann (2016)). A related but more informative metric of inference in an A/B test is the average treatment effect (ATE), which measures the difference in performance of treatment A vs treatment B. In this paper, we consider a sequential A/B test with the aim of estimating ATE. An experiment designer seeks to obtain a confidence interval (CI) with the width at most $\epsilon > 0$ such that the probability ATE does not lie in the CI is at most $\delta > 0$. We say that such CI has $(\epsilon, \delta)-$coverage guarantee. We aim to minimize the expected length of the experiment, i.e. expected sample size of the A/B test while delivering a CI with $(\epsilon, \delta)-$coverage guarantee. Both ϵ and δ are pre-specified at the beginning of the experiment.

A CI of ATE can be extremely useful in decision-making for online platforms in terms of deciding the design/treatment to pursue in future. ATE is especially useful to know how much value one is gaining via choosing the better treatment over the alternative when there is a deployment cost of designs/treatments. Further, a CI of ATE can be used in the design of future experiments as it precisely quantifies the value of one treatment over the other (see Johari et al. (2022)). Also, this CI is useful in clinical trials, as stated in Gardner & Altman (1986), "*In medical studies investigators are usually interested in determining the size of difference of a measured outcome between groups, rather than a simple indication of whether or not it is statistically significant.*"

Another application of estimating ATE, see Simchi-Levi & Wang (2023), is in situations where the best drug is not available due to shortage or some other factors. Then it is helpful to know the ATE of other drugs as compared to the best one.

Typically in A/B testing, the assignment rule used is the uniform randomized assignment rule, where incoming individuals are randomly assigned treatment A or B, with equal probabilities (see Tang et al. (2010) and Kohavi et al. (2013)). This is referred to as a randomized control trial (RCT). However, it is unclear whether this equal probability assignment for treatment A vs B is the best way to assign treatments to obtain the least sample size of the A/B test to estimate a CI of ATE with (ϵ, δ)-coverage guarantee. We aim to develop a policy that optimality assigns the treatments and uses the minimum sample size. To this end, we first develop an asymptotic lower bound on the expected sample size required for any adaptive experimental policy that provides a CI of the ATE with (ϵ, δ)-coverage guarantee as $\delta \to 0$. *It is important to note that the (ϵ, δ)-coverage guarantee for a CI of the ATE is valid for any $\epsilon > 0$ and any $\delta \in (0, 1)$.* We find that the lower bound on the expected sample size scales at the rate of $\log\left(\frac{1}{\delta}\right)$ as $\delta \to 0$, and the dependence of the lower bound on the parameters of the outcome distributions and ϵ can be expressed as the solution to a non-convex max-min optimization problem.

We then turn to develop a policy for A/B testing. Using the "plug-in" approach tailored for our ATE problem, we propose an adaptive experimental policy referred to as, $\mathfrak{P}_1$, which under mild assumptions is shown to be asymptotically optimal as $\delta \to 0$ and has $(\epsilon, \delta)-$coverage guarantee. The policy $\mathfrak{P}_1$ however is computationally expensive as it solves the non-convex max-min optimization problem before assigning treatments to each arriving individual. To improve the computation burden, we note that the max-min optimization problem that is defined in the lower bound provides the asymptotically optimal fraction of assignments for treatments A and B. This max-min optimization problem also provides insights into the asymptotically optimal CI of ATE.

We find that the asymptotically optimal fraction of treatment assignment for A and B is proportional to the standard deviation of their respective outcome distributions when $\epsilon \to 0$ and $\delta \to 0$. This assignment is in agreement with Neyman's allocation rule (see Neyman (1992) for the background) that aims to minimize the variance of the estimator and is extensively discussed in the existing literature. It is worth noting that, our objective function does not directly aim to minimize variance rather our goal is to minimize the expected sample size.

We note that the asymptotically optimal fraction of treatment assignment for A and B when $\epsilon > 0$, does not match with Neyman's allocation rule for general distributions. (It does for Gaussian distribution.) However, we prove theoretically that for small $\epsilon > 0$, the assignment of treatments suggested by the lower bound optimization problem is somewhat insensitive to the exact value of ϵ. This result is confirmed numerically for small (practical) values of ϵ. Using the above insights, we propose another adaptive policy $\mathfrak{P}_2$, which is computationally efficient compared to $\mathfrak{P}_1$ as it involves tracking the standard deviation of the outcome distributions, instead of solving the lower bound max-min optimization problem. We prove that $\mathfrak{P}_2$ is asymptotically optimal as $\epsilon \to 0$ and $\delta \to 0$ and has $(\epsilon, \delta)-$coverage guarantee. Further, we observe numerically that $\mathfrak{P}_2$'s performance is statistically indistinguishable from $\mathfrak{P}_1$ for finite practical values of ϵ and δ.

Our proposed policies, $\mathfrak{P}_1$ and $\mathfrak{P}_2$, exhibit a notable feature: they generate an anytime-valid confidence interval or sequence for the Average Treatment Effect (ATE) at any step, aligning with Darling's (1967) framework. These policies enable experimenters to conclude A/B tests prematurely, even when the (ϵ, δ)-coverage guarantee stopping rule is not met. Upon termination, while the resulting confidence interval (CI) width exceeds ϵ, it ensures that the ATE is captured in the CI with a minimum probability of $1 - \delta$. This feature is especially beneficial in environments like online platforms, where flexibility to stop experiments early without compromising validity is crucial. Termed as 'continuous monitoring,' or 'peeking' this approach allows for ongoing assessment of A/B tests, diverging from fixed end-point evaluations. This evolving methodology is supported by recent research, including works by (Waudby-Smith et al., 2021), (Ham et al., 2022), and (Lindon & Malek, 2022), highlighting its growing acknowledgement.

In Appendix F, we also quantify the value of the asymptotically optimal assignment rule in comparison to the uniform randomized assignment rule in terms of reducing the expected sample size. Additionally, unless specified otherwise, asymptotic optimality implies asymptotic optimality as $\delta \to 0$.

Organization of the paper. In the next section, we discuss the related literature. In Section 3, we present our main model. Section 4 provides the lower bound results on the expected sample size of the A/B test for constructing CI of ATE with $(\epsilon, \delta)-$coverage guarantee and our asymptotically optimal policy $\mathfrak{P}_1$ as $\delta \to 0$. In Section 5, we provide insights into the asymptotically optimal assignment rule for small δ. Section 6 leverages these insights from the lower bound to develop our asymptotically optimal policy $\mathfrak{P}_2$ when $\epsilon \to 0$ and $\delta \to 0$. In section 7, we provide the details of the numerical experiments. In Section 8, we provide the limitations of our work and future directions of the work. In Section 8 we provide the broader impact of our work.

2. Related literature

Our work is related to three streams of literature. The first stream is the work related to adaptive experiment design for ATE. The second stream is on sequential hypothesis. The third stream is concerned with bandit literature with pure exploration.

Adaptive Experiment Design for ATE: The field of experiment design, particularly for estimating the Average Treatment Effect (ATE), has been extensively studied. Recently adaptive experiments have been proposed by Hahn et al. (2011) and Kato et al. (2020) for minimizing asymptotic variance in ATE estimation. (Dai et al., 2024) aims to minimize the variance of an estimator of ATE within a fixed sample size in a design-based framework. Bhat et al. (2020) solved the allocation of treatments, using a dynamic optimization framework. Glynn et al. (2020) proposes a theoretical framework, to estimate ATE with asymptotic variance in the presence of temporal interference by reformulating it as a Markov decision problem. These methods primarily aim at minimizing asymptotic variance over a fixed experiment length, contrasting with our sequential approach focusing on non-asymptotic $(\epsilon, \delta)-$coverage guarantee. Simchi-Levi & Wang (2023) explores balancing efficiency and statistical power in adaptive experiments. Similar concepts in clinical trials were discussed by Hayre & Turnbull (1981) however the guarantees are only asymptotic.

Sequential Hypothesis Testing: The concept of sequential hypothesis testing, initiated in Wald (1945), involves concluding tests based on statistical criteria over time. Chernoff (1959) made seminal contributions in this area, see Naghshvar et al. (2013) and references there in for more recent advancements. Unlike the work in this stream that focuses on inferring hypotheses from finite sets related to outcome distributions, our objective is to minimize the length of the experiment to obtain an estimate of ATE with $(\epsilon, \delta)-$coverage guarantee.

Pure Exploration in Bandits: Substantial work has come up in recent times in the pure exploration problems in multi-armed bandit literature, see Bubeck et al. (2011) and Lattimore & Szepesvári (2020), where algorithm designer is interested in designing algorithms which at the end infers with statistical guarantees. One of the most studied problems in pure exploration is the best arm/treatment identification (BTI problem (see Mannor & Tsitsiklis (2004), Even-Dar et al. (2006), Audibert & Bubeck (2010), Garivier & Kaufmann (2016), Kaufmann et al. (2016), Russo (2016) see Juneja & Krishnasamy (2019) for generalizations). The BTI problem aims to identify with high probability the treatment that yields the highest mean outcome, with a minimum expected sample size. Our problem utilizes tools from this stream. However, our problem can not be solved by this framework directly (the BTI problem is in the same spirit as the sequential hypothesis testing problem), as we are interested in designing an experiment that estimates a CI of ATE with $(\epsilon, \delta)-$coverage guarantee.

3. Model

We consider a sequential framework with two treatments A and B. Our objective is to measure *Average Treatment Effect (ATE)* defined as the difference between the outcomes associated with treatment A and treatment B. To achieve this objective, we apply treatments to each incoming individual at a discrete time denoted by $n = 1, 2, 3, \ldots$. Specifically, either treatment A or B is assigned to the individual arriving at time n, represented as U_n. To define causal effects, we utilize the potential outcome notation introduced by Imbens & Rubin (2015). Let $(\mathfrak{X}_n(A), \mathfrak{X}_n(B))$ represent the random variable tuple whose first and second components capture the outcome that would have been observed if the n^{th} individual was assigned treatment $k \in \{A, B\}$. However, we observe only a single realized outcome, denoted as X_n, which corresponds to $\mathfrak{X}_n(U_n)$. We assume that $\{\mathfrak{X}_n(A); n = 1, 2, \ldots\}$ and $\{\mathfrak{X}_n(B); n = 1, 2, \ldots\}$ are independent and identically distributed according to distributions ν_A and ν_B, respectively. Let $\nu = \{\nu_A, \nu_B\}$ denote our true unknown underlying environment. Denoting the mean of a distribution by the function $m(\cdot)$, let $\mu_A = m(\nu_A)$ and $\mu_B = m(\nu_B)$. We denote the ATE of treatment A over treatment B is $\Delta = \mu_A - \mu_B$. Let $\{\mathcal{F}_n; n = 1, 2, \ldots\}$ denote the σ-algebra generated by $\{(U_j, X_j); j \leq n\}$. Our goal is to estimate a CI with the desired width $\epsilon > 0$, which contains the ATE, with a probability of at least $1 - \delta$, i.e., $(\epsilon, \delta)-$coverage guarantee. To achieve this goal, we choose an adaptive experimental policy consisting of the following three components:

1. Assignment rule: At each time $n = 1, 2, \ldots$, choose an assignment $U_n \in \{A, B\}$ adaptively and receives an independent draw X_n from ν_{U_n}.

2. Estimation rule: Let $[\hat{\Delta}_L(s), \hat{\Delta}_R(s)]$ represent the estimated confidence interval after observing $(U_1, X_1), (U_2, X_2), \ldots, (U_s, X_s)$ for $s = 1, 2, \ldots$.

3. Stopping rule: Let τ denote a stopping time with respect to $\{\mathcal{F}_n : n = 1, , 2, , 3...\}$.

Given the above three components, the adaptive policy yields the CI, $[\hat{\Delta}_L(\tau), \hat{\Delta}_R(\tau)]$. We next formally define the notion of $(\epsilon, \delta)-$coverage guarantee for an adaptive policy.

Definition 3.1. Given $\epsilon > 0$ and $\delta \in (0, 1)$, we say that an adaptive policy $\mathfrak{P}$, provides the $(\epsilon, \delta)-$coverage guarantee if the three components (assignment, estimation and stopping rule) above yield a confidence interval $[\hat{\Delta}_L(\tau), \hat{\Delta}_R(\tau)]$ such that $\mathbb{P}_\nu\{\Delta \notin [\hat{\Delta}_L(\tau), \hat{\Delta}_R(\tau)]\} \leq \delta$ and $\hat{\Delta}_R(\tau) - \hat{\Delta}_L(\tau) \leq \epsilon$, for all environment ν, where,

$\mathbb{P}_\nu(\cdot)$ denotes the probability measure induced by the environment ν.

For ease of exposition, we assume that ν_A and ν_B belong to the canonical single-parameter exponential family (SPEF) (see Cappé et al. (2013)), denoted as $\mathcal{S}$. In Appendix I, we extend our results to distributions of the outcome of treatments belonging to a non-parametric family with bounded support. The set $\mathcal{S}$ includes many commonly used distributions including Bernoulli, Poisson, Gaussian with known variance, and Gamma distribution with known shape parameter. Specifically $\nu_A, \nu_B \in \mathcal{S}$, where,

$$\mathcal{S} = \left\{ p_\theta : \theta \in \Theta \subset \mathbf{R}, \frac{dp_\theta}{d\xi} = \exp(\theta \cdot x - b(\theta)) \right\},$$

where ξ is some fixed reference measure on $\mathbf{R}$, and, $b(\cdot)$ is a fixed twice differentiable strictly convex function. The mean of the distribution p_θ is the derivative $b'(\theta)$ and the variance of the distribution is the double derivative $b''(\theta)$, for all $\theta \in \Theta$. Further, any distribution $p_\theta \in \mathcal{S}$ can be parameterized either by θ or by its mean. Let $KL(p_\theta, p_{\tilde{\theta}})$ represent the KL divergence of p_θ with respect to $p_{\tilde{\theta}}$. Since there is a one-to-one mapping between the mean of the distribution and the parameter θ, we define a divergence function, $d(\mu, \tilde{\mu}) \triangleq KL(p_{\theta(\mu)}, p_{\theta(\tilde{\mu})}) = b(\theta(\tilde{\mu})) - b(\theta(\mu)) - b'(\theta(\mu))(\theta(\tilde{\mu}) - \theta(\mu))$, such that $b'(\theta(\mu)) = \mu$, $b'(\theta(\tilde{\mu})) = \tilde{\mu}$ and $p_{\theta(\mu)}, p_{\theta(\tilde{\mu})} \in \mathcal{S}$. Let $\sigma(\mu)$ denote the standard deviation of the distribution with mean μ. For future use, let $\mathcal{I}$ denote the support of $d(\mu, \cdot)$ and define $\overline{\mu} = \sup \mathcal{I}$ and $\underline{\mu} = \inf \mathcal{I}$. We also assume that $\mathcal{S}$ is regular, that is, $\mathcal{I} = (\underline{\mu}, \overline{\mu})$. Note that $\overline{\mu}$ can be ∞ or $\underline{\mu}$ can be $-\infty$ as in the case of Gaussian with known variance. In Appendix D, we establish several key properties of the function $d(\mu, x)$, notably that $d(\mu, x)$ exhibits strict quasi-convexity in the second argument.

Remark 3.2. Since Δ can take values in between $(\underline{\mu} - \overline{\mu}, \overline{\mu} - \underline{\mu})$ when outcome distributions lie in $\mathcal{S}$, hence it follows that our ATE problem is well defined only if $\epsilon < 2(\overline{\mu} - \underline{\mu})$.

4. Lower bound and asymptotically optimal $(\epsilon, \delta)-$coverage guarantee policy

4.1. Lower bound

In this subsection, we first develop a lower bound on the expected sample size required for any adaptive policy which provides $(\epsilon, \delta)-$coverage guarantee. Recall that $\nu = \{\nu_A, \nu_B\}$ denotes our true underlying environment. In this environment, the distributions of the outcomes of treatments A and B are $\nu_A, \nu_B \in \mathcal{S}$ with means μ_A and μ_B, respectively.

Consider an adaptive policy denoted by $\mathfrak{P}$, which provides (ϵ, δ)-coverage guarantee. Let $N_A(n)$ and $N_B(n) = n - N_A(n)$ represent the number of times treatment A and B have been chosen for the first n assignments, respectively, by $\mathfrak{P}$. The estimation of the CI for ATE, i.e., Δ by $\mathfrak{P}$ is denoted by $[\hat{\Delta}_L(\tau), \hat{\Delta}_R(\tau)]$ at the stopping time τ, where $\hat{\Delta}_R(\tau) - \hat{\Delta}_L(\tau) \leq \epsilon$. Consider any alternate environment $\nu' = \{\nu'_A, \nu'_B\}$ in the set $\mathcal{K}$, where, $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2$, and $\mathcal{K}_1 = \{(\nu'_A, \nu'_B) : \nu'_A, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B < \Delta - \epsilon\}$ and $\mathcal{K}_2 = \{(\nu'_A, \nu'_B) : \nu'_A, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B > \Delta + \epsilon\}$. Using information theoretic arguments (see Section 33.2.1 in Lattimore & Szepesvári (2020) and Kaufmann (2020)), if τ is almost surely finite, we have,

$$\mathbb{E}_\nu[N_A(\tau)] \cdot d(\mu_A, \mu'_A) + \mathbb{E}_\nu[N_B(\tau)] \cdot d(\mu_B, \mu'_B) \geq \Psi(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})), \tag{1}$$

where, $\mathbb{E}_\nu[\cdot]$ denotes the expectation operator under our environment ν for any event $\mathcal{E} \in \mathcal{F}_\tau$, where $\Psi(p_1, p_2) \triangleq p_1 \log \frac{p_1}{p_2} + (1 - p_1) \log \left(\frac{1-p_1}{1-p_2} \right)$ for $p_1, p_2 \in (0, 1)$.

Set $\mathcal{E}$ to be $\{\Delta' \notin [\hat{\Delta}_L(\tau), \hat{\Delta}_R(\tau)]\}$, where $\Delta' = \mu'_A - \mu'_B$. Using the definition of $(\epsilon, \delta)-$coverage, we have $\mathbb{P}_{\nu'}(\mathcal{E}) \leq \delta$ and observing

$$\mathbb{P}_\nu(\mathcal{E}) \geq \mathbb{P}_\nu(\Delta \in [\hat{\Delta}_L(\tau), \hat{\Delta}_R(\tau)]) \geq 1 - \delta. \tag{2}$$

Thus we obtain that left hand side of (1) of bounded by $\Psi(1 - \delta, \delta)$, which in turn is greater than $\log(\frac{1}{4\delta})$. Further noting $\frac{\mathbb{E}_\nu[N_A(\tau)]}{\mathbb{E}_\nu[\tau]} = w \in [0, 1]$, $\frac{\mathbb{E}_\nu[N_B(\tau)]}{\mathbb{E}_\nu[\tau]} = 1 - w \in [0, 1]$, we have

$$\mathbb{E}_\nu[\tau] \geq \frac{\log(\frac{1}{4\delta})}{\sup\limits_{w \in [0,1]} \inf\limits_{\nu' \in \mathcal{K}} w d(\mu_A, \mu'_A) + (1 - w) d(\mu_B, \mu'_B)}.$$

We observe that the above result provides a lower bound on the expected sample size using the set of alternate environments $\mathcal{K}$. This bound turns out to be not tight and one can obtain a tighter bound by expanding the set $\mathcal{K}$. However, for any set larger than $\mathcal{K}$, the first inequality (left one) in (2) may not hold, thus we impose a stability condition to obtain a tighter bound which is asymptotic in the regime as $\delta \to 0$. From now on, we will index τ with subscript δ. Here is the formal definition of the stability condition.

Definition 4.1. Let τ_δ be the stopping time of a $(\epsilon, \delta)-$coverage adaptive experimental policy with the estimated CI of ATE, denoted by $[\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]$. The policy is said to be *stable with limiting CI* $[\Delta_L, \Delta_R]$, if $\hat{\Delta}_L(\tau_\delta) \xrightarrow{p} \Delta_L$ and $\hat{\Delta}_R(\tau_\delta) \xrightarrow{p} \Delta_R$ as $\delta \to 0$, where Δ_L and Δ_R are constants.

For stating the asymptotic lower bound, consider an adaptive policy denoted by $\mathfrak{P}$, which provides (ϵ, δ)-coverage guarantee with limiting CI $[\Delta_L, \Delta_R]$. Define the set of alternate environments $\mathcal{K}(\Delta_L, \Delta_R) = \mathcal{K}_1(\Delta_L) \cup \mathcal{K}_2(\Delta_R)$, and $\mathcal{K}_1(\Delta_L) = \{(\nu'_A, \nu'_B) : \nu'_A, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B < \Delta_L\}$ and $\mathcal{K}_2(\Delta_R) = \{(\nu'_A, \nu'_B) : \nu'_A, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B > \Delta_R\}$. Definition 4.1 and the $(\epsilon, \delta)-$coverage guarantee ensures that $\Delta_R - \Delta_L \leq \epsilon$ and $\Delta \in [\Delta_L, \Delta_R]$, which implies that $\mathcal{K} \subset \mathcal{K}(\Delta_L, \Delta_R)$. We use the same argument as before where we consider $\nu' \in \mathcal{K}(\Delta_L, \Delta_R)$ and $\mathcal{E} = \{\Delta' \notin [\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]\}$. We cannot use (2) as mentioned before, however the stability condition implies $\mathbb{P}_\nu(\mathcal{E}) \approx 1$ for small δ. The rest of the argument is similar to before and we obtain the asymptotic lower bound. We present a rigorous proof (see Appendix A.1) of the above argument. Before we state the formal result, we define some notation that will be useful for stating the theorem. For $w \in [0,1]$, let $\mathcal{C}(z) = \{(x,y) : x, y \in \mathcal{I}, x - y = z\}$ and,

$$T(\mu_A, \mu_B, w, z) \triangleq \min_{x,y \in \mathcal{C}(z)} wd(\mu_A, x) + (1-w)d(\mu_B, y). \tag{3}$$

Define $\Upsilon(\epsilon) = \{(\Delta_L, \Delta_R) : \Delta \in [\Delta_L, \Delta_R], \Delta_R = \Delta_L + \epsilon\}$. Now we state our lower bound result.

Theorem 4.2. *For given $\nu_A, \nu_B \in \mathcal{S}$ with mean μ_A and μ_B respectively and any $(\epsilon, \delta)-$coverage and stable adaptive experimental policy with an almost surely finite stopping time τ_δ, we have*

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\ell^*(\mu_A, \mu_B, \epsilon)}, \tag{4}$$

where $\ell^*(\mu_A, \mu_B, \epsilon)$ is the optimal value of the following optimization problem (denoted by $\mathfrak{L}$):

$$\ell^*(\mu_A, \mu_B, \epsilon) =$$

$$\sup_{\substack{w \in [0,1] \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \inf_{\nu' \in \mathcal{K}(\Delta_L, \Delta_R)} w\, d(\mu_A, \mu'_A) + (1-w)\, d(\mu_B, \mu'_B).$$

This further equals,

$$\sup_{\substack{w \in [0,1], \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_R)\}.$$

Let $w^*(\mu_A, \mu_B, \epsilon)$, $\Delta^*_L(\mu_A, \mu_B, \epsilon)$ and $\Delta^*_R(\mu_A, \mu_B, \epsilon)$ denote a solution to the optimization problem $\mathfrak{L}$ (later in Theorem 5.3, we demonstrate the existence of solution of $\mathfrak{L}$). It follows from the above theorem that, $\Delta^*_R(\mu_A, \mu_B, \epsilon) = \Delta^*_L(\mu_A, \mu_B, \epsilon) + \epsilon$. We provide a policy later in this section that chooses the assignment rule such that the fraction of assignment of treatment A tracks $w^*(\mu_A, \mu_B, \epsilon)$ and the limiting CI is $[\Delta^*_L(\mu_A, \mu_B, \epsilon), \Delta^*_R(\mu_A, \mu_B, \epsilon)]$. We prove that this proposed policy matches the lower bound for $\delta \to 0$, hence the lower bound provided above is asymptotically tight when $\delta \to 0$. Hence $w^*(\mu_A, \mu_B, \epsilon)$ and $[\Delta^*_L(\mu_A, \mu_B, \epsilon), \Delta^*_R(\mu_A, \mu_B, \epsilon)]$ can be interpreted as the asymptotic optimal fraction of assignment of treatment A and the limiting CI, respectively.

We now propose our policy $\mathfrak{P}_1$ that is shown to be asymptotically optimal and has $(\epsilon, \delta)-$coverage guarantee.

4.2. Asymptotically optimal policy $\mathfrak{P}_1$

Let $\hat{\mu}_A(n)$ and $\hat{\mu}_B(n)$ denote the sample average of outcomes of individuals who were assigned the treatment A and B by time n, respectively. Hence, $\hat{\mu}_A(n) = \frac{\sum_{t=1}^n X_t \mathbb{I}_{\{U_n = A\}}}{\sum_{t=1}^n \mathbb{I}_{\{U_n = A\}}}$, and $\hat{\mu}_B(n) = \frac{\sum_{t=1}^n X_t \mathbb{I}_{\{U_n = B\}}}{\sum_{t=1}^n \mathbb{I}_{\{U_n = B\}}}$.

To achieve asymptotic optimality, we choose the assignment rule to track $w^*(\mu_A, \mu_B, \epsilon)$. Further, our estimation and stopping rules aim to construct a CI of ATE $[\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]$ such that $\hat{\Delta}_L(\tau_\delta)$ converges to $\Delta^*_L(\mu_A, \mu_B, \epsilon)$ and $\hat{\Delta}_R(\tau_\delta)$ converges to $\Delta^*_R(\mu_A, \mu_B, \epsilon)$ in probability as $\delta \to 0$. All three components of $\mathfrak{P}_1$ is given by,

1. **Assignment rule:** Various tracking rules exist in bandit literature (see Garivier & Kaufmann (2016), Agrawal et al. (2020) and Degenne & Koolen (2019)). Here, we use the D-tracking rule introduced in Garivier & Kaufmann (2016). We define, $Q_n = \{k \in \{A, B\} : N_k(n) < \sqrt{n} - 1\}$. If $Q_n = \{\}$, then

   $$U_{n+1} = \arg\max_{k \in \{A,B\}} n \cdot w^*(\hat{\mu}_A(n), \hat{\mu}_B(n), \epsilon) - N_k(n),$$

   else $U_{n+1} = \arg\min_{k \in Q_n} N_k(n)$.

2. **Estimation rule:** We use the first-order optimality conditions for $\mathfrak{L}$ where we substitute $\hat{\mu}_A(n)$, $\hat{\mu}_B(n)$ in the place of μ_A and μ_B. Formally, we compute $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ that satisfies,

   $$T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_L(n)\right) \tag{5}$$
   $$= T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_R(n)\right) = \frac{\beta(n, \delta)}{n},$$

   (see (3) for the definition of $T(\mu_A, \mu_B, w, \Delta_L)$ function). Here, for a given $\alpha > 1$, we can choose $\beta(n\,\delta) = \log\left(\frac{c_1 n^\alpha}{\delta}\right)$, where $c_1 = c_1(\alpha)$ is an appropriately chosen positive constant. $\beta(n, \delta)$ is chosen to ensure that $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$ is wide enough such that ATE lies in the CI with probability $1 - \delta$ at each step. We show in Lemma E.6 in Appendix E that $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ satisfying (5) exist uniquely. Further, in Lemma E.3 in Appendix E, we show that $T(\mu_A, \mu_B, w, z)$ is increasing for $z \geq \Delta$ and decreasing $z \leq \Delta$, hence computing $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ requires simple binary search.

   Remark 4.3. The above threshold of $\beta(n, \delta)$ was used by Garivier & Kaufmann (2016). Tighter definitions of the threshold $\beta(n, \delta)$ have been proposed in various papers to get probabilistic guarantees in the context of the best treatment identification problem, which we can utilize in our ATE problem as well (see Agrawal et al. (2020), Kaufmann & Koolen (2021) and Jourdan et al. (2022)). Barrier (2023) gives another definition of $\beta(n, \delta)$ which works well empirically in practice although not theoretically supported.

3. **Stopping rule:** Set the stopping rule as

$$\tau_\delta = \inf\{n \in \mathbb{N} : \hat{\Delta}_R(n) - \hat{\Delta}_L(n) \leq \epsilon\}.$$

Next, we present results that state that $\mathfrak{P}_1$ has the (ϵ, δ)$-$coverage guarantee and is asymptotically optimal.

Theorem 4.4. (**(ϵ, δ)$-$coverage guarantee and stability of $\mathfrak{P}_1$**) *For $\mathfrak{P}_1$, there exists a $\epsilon_o > 0$ such that for $\epsilon \leq \epsilon_o$, we have:*

a) *For a given $\delta \in (0, 1)$, τ_δ is finite almost surely.*

b) *$\mathfrak{P}_1$ has the $(\epsilon, \delta)-$coverage guarantee.*

c) *$\mathfrak{P}_1$ is a stable policy.*

We assume the following technical property to prove the asymptotic optimality of $\mathfrak{P}_1$. In Section 6, we propose another policy $\mathfrak{P}_2$ based on the insights of the solution of lower bound $\mathfrak{L}$ which does not require this assumption.

Assumption 4.5. For all $\mu_A, \mu_B \in \mathcal{I}$, $\min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}$ is a jointly strictly quasi-concave function in w and Δ_L for all $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$ (see (3) for the definition of $T(\mu_A, \mu_B, w, \Delta_L)$).

Remark 4.6. We prove in Appendix H.2, that the above technical property is satisfied by Gaussian distributions with known variance. For general outcome distributions in $\mathcal{S}$, we numerically show that the above technical property holds via plotting upper contour sets (See Appendix H.2). This technical assumption implies that the solution of $\mathfrak{L}$ is unique, i.e., $w^*(\mu_A, \mu_B, \epsilon)$, $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ and $\Delta_R^*(\mu_A, \mu_B, \epsilon)$ are unique for a given μ_A, μ_B and $\epsilon > 0$. (See statement and proof of Lemma E.7 in Appendix E).

Theorem 4.7. (Asymptotic optimality of $\mathfrak{P}_1$) *For $\mathfrak{P}_1$, there exists a $\epsilon_o > 0$ such that for $\epsilon \leq \epsilon_o$, we have:*

$$\mathbb{P}_\nu\left(\limsup_{\delta\to 0} \frac{\tau_\delta}{\log(1/\delta)} = \frac{1}{\ell^*(\mu_A, \mu_B, \epsilon)}\right) = 1 \textit{ and}$$

$$\lim_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} = \frac{1}{\ell^*(\mu_A, \mu_B, \epsilon)}.$$

Remark 4.8. Our proposed policy $\mathfrak{P}_1$ not only has (ϵ, δ)$-$coverage guarantee but its estimation rule constructs a confidence sequence/ anytime valid confidence interval (see Darling & Robbins (1967), Howard et al. (2021), and references within) for ATE.

Definition 4.9. The confidence interval generated by a policy's estimation rule $\{[\hat{\Delta}_L(n), \hat{\Delta}_R(n)], n = 1, \ldots\}$ is called a $(1-\delta)-$confidence sequence of Δ, if following holds,

$$\mathbb{P}_\nu(\forall n \in \mathbb{N} : \Delta \in [\hat{\Delta}_L(n), \hat{\Delta}_R(n)]) \geq 1 - \delta. \quad (6)$$

A policy generating a confidence sequence of ATE allows the experimenter to terminate the A/B test at any arbitrary point before the end of the experiment. Upon stopping, such a policy yields a confidence interval (CI) wider than ϵ (if it is stopped earlier than mandated by the stopping rule that ensures (ϵ, δ)$-$coverage guarantee). However, this CI contains the ATE with a probability of at least $1 - \delta$. Note that the stopping can occur at an arbitrary time, i.e., the experimenter at time n can use any of the CI that was generated at time $1, \ldots n-1$. This aspect is particularly valuable in practical scenarios, such as online platforms, where there is a demand for the flexibility to halt experiments at any juncture while still deriving valid conclusions. Our policy $\mathfrak{P}_1$ provide $(1-\delta)$-confidence sequence of Δ as it satisfies (6). See the formal statement in Theorem A.2 in Appendix A.2.

Observe that the assignment rule of $\mathfrak{P}_1$ requires the solution of $\mathfrak{L}$ at each time step, which is computationally expensive. In the remaining sections, we first develop structural insights about $w^*(\mu_A, \mu_B, \epsilon)$ from the lower bound, we then construct a policy, denoted as $\mathfrak{P}_2$, that utilizes the structure of $w^*(\mu_A, \mu_B, \epsilon)$ and has substantially lower computational burden compared to $\mathfrak{P}_1$. Further, we prove that $\mathfrak{P}_2$ is asymptotically optimal when $\epsilon \to 0$ and $\delta \to 0$.

5. Insights from lower bound

In this section, we first provide insights about any asymptotically optimal adaptive policy from the lower bound analysis, in particular about the assignment rule for our ATE problem. We now introduce *Fisher's information* for distributions in $\mathcal{S}$, this will aid us in stating the next result. Given $\nu \in \mathcal{S}$ with mean $m(\nu) = \mu$, the Fisher information $I(\mu)$ of distribution ν can be expressed as

$$I(\mu) = \left.\frac{\partial^2 d(\mu, x)}{\partial x^2}\right|_{x=\mu} = \frac{1}{\sigma^2(\mu)},$$

where $\sigma^2(\mu)$ is the variance of ν. For details, see Theorem 5.4, Chapter 2 in Lehmann & Casella (2006) and Section 15.3 in Agrawal (2022). It is worth noting that we prove in Appendix D.3 that $b(\theta) \in C^\infty$ as a function of θ. This implies that $d(\mu, x)$ is in C^∞ as a function of x (see Appendix D.1).

Our primary objective is to develop insights into the solution for $\mathfrak{L}$. To this end, we start with Gaussian distributions where the optimization problem $\mathfrak{L}$ has a unique solution and it can be explicitly characterized.

Proposition 5.1. *If underlying distributions of outcomes of both treatments are Gaussian with known variances, i.e.,* $\nu_A = N(\mu_A, \sigma_A^2)$ *and* $\nu_B = N(\mu_B, \sigma_B^2)$*, then* $w^*(\mu_A, \mu_B, \epsilon)$ *and* $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ *uniquely satisfy*

$$w^*(\mu_A, \mu_B, \epsilon) = \overline{w}(\mu_A, \mu_B) \text{ and } \Delta_L^*(\mu_A, \mu_B, \epsilon) = \Delta - \frac{\epsilon}{2},$$

where, $\overline{w}(\mu_A, \mu_B) \triangleq \frac{\sqrt{\frac{1}{I(\mu_A)}}}{\sqrt{\frac{1}{I(\mu_A)}}+\sqrt{\frac{1}{I(\mu_B)}}}$*. The above in turn implies that,* $w^*(\mu_A, \mu_B, \epsilon) = \frac{\sigma_A}{\sigma_A+\sigma_B}$.

Remark 5.2. It's noteworthy that the above proposition permits ν_A and ν_B to reside in distinct $\mathcal{S}$ sets. In the context of Gaussian distributions, ν_A and ν_B are confined to the same $\mathcal{S}$ solely if they share identical variances. As before, for the rest of the results, we operate under the assumption that ν_A and ν_B reside within the same $\mathcal{S}$.

The above proposition states that if the outcome distributions of treatments are Gaussian distributed with known variances, then the asymptotically optimal fraction of treatment assignment for A and B is unique and is proportional to standard deviation. This coincides with Neyman's allocation rule. Note that this assignment of treatments does not depend upon ϵ, however, this is not true in general. For a general distribution of $\nu_A, \nu_B \in \mathcal{S}$, $\mathfrak{L}$ is a non-convex optimization problem and the exact solution is not analytically tractable, we compute its behaviour near $\epsilon \approx 0$ using Taylor series expansion. The next result is our key finding.

Theorem 5.3. *For a given* $\nu_A, \nu_B \in \mathcal{S}$ *with mean* μ_A *and* μ_B *respectively, then following holds: a solution to the optimization problem* $\mathfrak{L}$ *exists, i.e.,* $w^*(\mu_A, \mu_B, \epsilon)$ *and* $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ *exists and any* $w^*(\mu_A, \mu_B, \epsilon)$ *and* $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ *that is a solution to the optimization problem* $\mathfrak{L}$ *satisfies,*

$$\lim_{\epsilon \to 0} \frac{\ell^*(\mu_A, \mu_B, \epsilon)}{\epsilon^2} = \frac{1}{8\left(\sqrt{\frac{1}{I(\mu_A)}} + \sqrt{\frac{1}{I(\mu_B)}}\right)^2}. \tag{7}$$

Further, $\lim_{\epsilon \to 0} \frac{\Delta - \Delta_L^*(\mu_A, \mu_B, \epsilon)}{\epsilon} = \frac{1}{2}$*, and*

$$\lim_{\epsilon \to 0} w^*(\mu_A, \mu_B, \epsilon) = \overline{w}(\mu_A, \mu_B).$$

The theorem states that as ϵ decreases, the value of $\ell^*(\mu_A, \mu_B, \epsilon)$ also decreases at a rate of ϵ^2. As a result, our ATE problem requires a larger expected sample size and the growth of $\mathbb{E}_\nu[\tau_\delta]$ is at the rate of $1/\epsilon^2$. Further, the above result states that the asymptotically optimal construction of CI around the estimator of ATE is symmetric when ϵ and δ are small. We find that the asymptotically optimal fraction of treatment assignment for A and B is inversely proportional to the square root of Fisher's information of the outcome distributions of treatments A and B, respectively, when ϵ and δ are small. This result aligns with Neyman's allocation rule, as for $\mathcal{S}$ the inverse of Fisher's information of distribution equals the variance of the distribution. Now we study how $w^*(\mu_A, \mu_B, \epsilon)$ changes with ϵ for general outcome distributions in $\mathcal{S}$.

Insensitivity of $w^*(\mu_A, \mu_B, \epsilon)$ to ϵ: Numerical observations and theoretical justification. We conducted a numerical study to understand how $w^*(\mu_A, \mu_B, \epsilon)$ behaves with changes in ϵ. Numerical exploration reveals that the asymptotically optimal fraction of treatments, as $\delta \to 0$, shows limited sensitivity to variations in ϵ, suggesting $w^*(\mu_A, \mu_B, \epsilon) \approx \overline{w}(\mu_A, \mu_B)$ across a range of ϵ. In our study, we examine outcome distributions for treatments A and B under two scenarios: exponential and Bernoulli distributions (refer to Figure 1). For both distributions, $w^*(\mu_A, \mu_B, \epsilon)$ remains relatively constant, closely approximating $\overline{w}(\mu_A, \mu_B)$, with a notable exception for Bernoulli distributions when $\mu_A = 0.5$ and $\mu_B = 0.08$, and $\epsilon > 0.2$.

Now we theoretically justify that $w^*(\mu_A, \mu_B, \epsilon) \approx \overline{w}(\mu_A, \mu_B)$ for reasonable values of ϵ by showing that the rate of change of $w^*(\mu_A, \mu_B, \epsilon)$ as a function of ϵ is 0 when $\epsilon \to 0$. We also characterize the lower order ϵ^2 term in the theorem below (see (8)).

Theorem 5.4. *For a given* $\nu_A, \nu_B \in \mathcal{S}$ *with mean* μ_A *and* μ_B *respectively, if* $d(\mu, x)$ *is four times continuously differentiable in* x *around a neighbourhood of* $x = \mu$*, then following holds:*

$$\lim_{\epsilon \to 0} \frac{w^*(\mu_A, \mu_B, \epsilon) - \overline{w}(\mu_A, \mu_B)}{\epsilon} = 0.$$

Further, we have,

$$\lim_{\epsilon \to 0} \frac{w^*(\mu_A, \mu_B, \epsilon) - \overline{w}(\mu_A, \mu_B)}{\epsilon^2} = \frac{v(\mu_A, \mu_B)}{96\sqrt{I(\mu_A) I(\mu_B)}}, \tag{8}$$

where, $v(\mu_A, \mu_B) \triangleq \left(\frac{\partial^4 d(\mu_B, x)}{\partial x^4}\Big|_{x=\mu_B} \right)(1 - \overline{w}(\mu_A, \mu_B))^4 - \left(\frac{\partial^4 d(\mu_A, x)}{\partial x^4}\Big|_{x=\mu_A} \right)\overline{w}^4(\mu_A, \mu_B)$.

Remark 5.5. It is worth noting that the ϵ^2 correction terms provided in (8) are zero for the Gaussian case and this aligns with our finding $w^*(\mu_A, \mu_B, \epsilon) = \overline{w}(\mu_A, \mu_B)$.

6. Computationally efficient asymptotically optimal policy $\mathfrak{P}_2$

We now present a policy $\mathfrak{P}_2$ which has less computational burden in comparison to $\mathfrak{P}_1$. As we study this policy in the asymptotic regime where $\epsilon \to 0$ and $\delta \to 0$, hence we index τ with both δ and ϵ and use the notation $\tau_{\delta,\epsilon}$. Further $\mathfrak{P}_2$ is asymptotically optimal as $\epsilon \to 0$ and $\delta \to 0$. We utilize the limiting behaviour of $w^*(\mu_A, \mu_B, \epsilon)$ when $\epsilon \to 0$ for $\mathfrak{P}_2$. Specifically, policy $\mathfrak{P}_2$ is identical to $\mathfrak{P}_1$

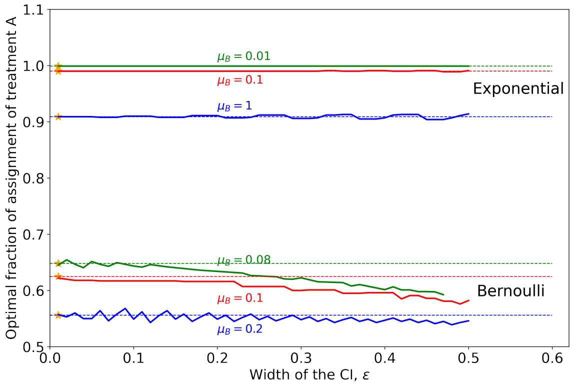

Figure 1: Plot of $w^*(\mu_A, \mu_B, \epsilon)$ with ϵ for two cases: first when outcome distributions are Bernoulli with $\mu_A = 0.5$ and three different values of μ_B, second when outcome distributions are exponential with $\mu_A = 10$ and three different values of μ_B. Horizontal lines starting from the star points represent the value of $\overline{w}(\mu_A, \mu_B)$. One can see that for the exponential cases, these horizontal lines completely overlap with the values of $w^*(\mu_A, \mu_B, \epsilon)$. A similar pattern follows for the Bernoulli cases as well, except when ϵ becomes larger than 0.2.

with the modification that in the assignment rule we replace $w^*(\hat{\mu}_A(n), \hat{\mu}_B(n), \epsilon)$ with $\overline{w}(\hat{\mu}_A(n), \hat{\mu}_B(n))$ and utilize the same stopping and estimation rule as in $\mathfrak{P}_1$.

Our policy $\mathfrak{P}_2$, similar to the policy $\mathfrak{P}_1$, is stable and has $(\epsilon, \delta)-$coverage guarantee. This implies, Theorem 4.4 holds for $\mathfrak{P}_2$ as well. Further, $\mathfrak{P}_2$ also constructs a $(1 - \delta)-$confidence sequence of ATE. Now we state the result which provides the asymptotic optimality of policy $\mathfrak{P}_2$.

Theorem 6.1. (Asymptotic optimality of policy $\mathfrak{P}_2$) *There exists a $\epsilon_o > 0$ such that the following holds for $\mathfrak{P}_2$ if $\epsilon \leq \epsilon_o$:*

$$\lim_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_{\delta,\epsilon}]}{\log(1/\delta)} = \frac{1}{\ell^{\mathfrak{P}_2}(\mu_A, \mu_B, \epsilon)},$$

where,

$$\ell^{\mathfrak{P}_2}(\mu_A, \mu_B, \epsilon) =$$

$$\sup_{(\Delta_L, \Delta_R) \in \Upsilon(\epsilon)} \min\{T(\mu_A, \mu_B, \overline{w}, \Delta_L), T(\mu_A, \mu_B, \overline{w}, \Delta_R)\}.$$

Further, we have $\lim_{\epsilon \to 0} \lim_{\delta \to 0} \frac{\epsilon^2 \mathbb{E}_\nu[\tau_{\delta,\epsilon}]}{\log(1/\delta)} =$

$$= \lim_{\epsilon \to 0} \frac{\epsilon^2}{\ell^{\mathfrak{P}_2}(\mu_A, \mu_B, \epsilon)} = \lim_{\epsilon \to 0} \frac{\epsilon^2}{\ell^*(\mu_A, \mu_B, \epsilon)} \tag{9}$$

$$= \frac{1}{8\left(\sqrt{\frac{1}{I(\mu_A)}} + \sqrt{\frac{1}{I(\mu_B)}}\right)^2}.$$

The lower bound developed in Theorem 4.2 and (9) above together imply the asymptotic optimality of the policy $\mathfrak{P}_2$ in the regime $\epsilon \to 0$ and $\delta \to 0$. Using the above theorem, we also obtain that for a given small $\epsilon > 0$, our policy $\mathfrak{P}_2$ is near optimal as $\delta \to 0$.

It is worth noting that $I(\mu)$ is a continuous function of μ for Gaussian with known variance, Gamma with known shape parameters, Bernoulli, Poisson and Geometric distribution. In general, it requires the $b(\theta)$ function to be twice continuously differentiable (see Appendix D for detailed discussion).

Comparison of $\mathfrak{P}_1$ and $\mathfrak{P}_2$. Even though theoretically $\mathfrak{P}_2$ policy is asymptotically optimal in $\epsilon \to 0$ and $\delta \to 0$ whereas $\mathfrak{P}_1$ policy is asymptotically optimal in $\delta \to 0$ for any small $\epsilon > 0$, we numerically verify that expected sample size taken by $\mathfrak{P}_2$ is statistically indistinguishable when compared to expected sample size taken by $\mathfrak{P}_1$ for reasonable values of $\epsilon > 0$ and $\delta \in (0, 1)$. This great performance of $\mathfrak{P}_2$ stems from the fact that $w^*(\mu_A, \mu_B, \epsilon) \approx \overline{w}(\mu_A, \mu_B)$ (see Theorem 5.4 and discussion around it). In the next section, we present the numerical performance of $\mathfrak{P}_2$ policy and compare it with a uniform randomized assignment policy where incoming individuals are randomly assigned treatment A or B, with equal probabilities.

7. Numerical experiments

In this section, we present a numerical analysis to demonstrate the performance of our policy, $\mathfrak{P}_2$, compared to a uniform randomized policy. According to Theorem 5.3, $\mathfrak{P}_2$ tends to sample in a manner close to uniform when the Fisher's information $I(\mu_A)$ and $I(\mu_B)$ for the two treatments are similar. Consequently, when $I(\mu_A)$ and $I(\mu_B)$ are nearly equal, the expected benefit of employing policy $\mathfrak{P}_2$ over a uniform randomized approach is relatively minor.

To highlight the advantages of our policy $\mathfrak{P}_2$ against the uniform randomized policy, we examine a scenario where $I(\mu_A)$ and $I(\mu_B)$ significantly diverge. Specifically, we model the outcomes for treatments A and B as exponentially distributed with means $\mu_A = 10$ and $\mu_B = 0.1$, respectively. We select $\epsilon = 0.5$ and explore different values of δ, including 10%, 5%, and 1%. For each δ setting, we generate 2000 sample paths and calculate the average outcomes. The efficacy of policy $\mathfrak{P}_2$ is then compared to that of a policy adhering to the randomized controlled trials (RCT) framework, which combines a uniform random assignment with our estimation and stopping criteria, referred to here as $\mathfrak{P}_{\text{RCT}}$. We present our findings in Table 1. Details of the numerical experiments are provided in Appendix G. We observe the approximately 50% reduction in the sample size by our policy $\mathfrak{P}_2$ over the performance of the policy $\mathfrak{P}_{\text{RCT}}$. We provide the theoretical support for this finding

Table 1: Performance of $\mathfrak{P}_2$ and $\mathfrak{P}_{\mathrm{RCT}}$ policy for finite δ values.The width of 95% CI for estimated $\mathbb{E}[\tau_\delta]$ for both policies is less than 150. We observe that the CI of ATE for both policies $\mathfrak{P}_2$ and $\mathfrak{P}_{\mathrm{RCT}}$ always contains ATE at the end, for all three cases of $\delta = 10, 5, 1\%$.

δ	Lower bound on $\mathbb{E}[\tau_\delta]$	$\mathbb{E}[\tau_\delta]$ for $\mathfrak{P}_2$	$\mathbb{E}[\tau_\delta]$ for $\mathfrak{P}_{\mathrm{RCT}}$
10%	7.52×10^3	3.74×10^4	7.85×10^4
5%	9.78×10^3	4.14×10^4	8.62×10^4
1%	1.50×10^4	4.98×10^4	1.02×10^5

Table 2: Performance of $\mathfrak{P}_2$, Clip-OGD and infeasible benchmark of Section 3 of Dai et al. (2024) for $\delta = 0.01$ and $\epsilon = 0.2$. The width of 95% CI for estimated $\mathbb{E}[\tau_\delta]$ for both policies is less than 100. We observe that the CI of ATE for all three policies always contains ATE at the end.

Policy	$\mathbb{E}[\tau_\delta]$ for the policy
$\mathfrak{P}_2$	2672
Clip-OGD, Dai et al. (2024)	4034
Neyman Benchmark, Dai et al. (2024)	4573

in Proposition F.2 in Appendix F. We quantify the value of any asymptotically optimal adaptive policy's assignment rule compared to the uniform randomized assignment rule in terms of reducing the sample size as a function of the ratio of Fisher's information of outcome distributions for small ϵ and δ.

We next present the comparison of policy $\mathfrak{P}_2$ with a policy that uses the adaptive sampling rule of clip-OGD described in Section 4 of Dai et al. (2024) along with the estimation and stopping rules of $\mathfrak{P}_2$. We refer to this policy as Clip-OGD. While our paper and Dai et al. (2024) refer to the asymptotically optimal assignment fraction of treatments as Neyman's allocation rule, they are not identical. The Neyman's allocation rule in Dai et al. (2024), is expressed as $\frac{S(A)}{S(A)+S(B)}$, where $S(A)$ and $S(B)$ are the square root of the second moment of treatment A and B, respectively. This represents the asymptotically optimal weight (assignment probability of treatment A) in their setting for a large sample size.

Using a numerical study, we show that this assignment rule disparity results in a notable difference in performance. We consider two well-separated treatments: the outcome distribution of treatment A is Bernoulli with mean $\mu_A = 0.98$ and of treatment B is Bernoulli with $\mu_B = 0.5$. We set $\epsilon = 0.2$ and $\delta = .01$, and then compare our policy $\mathfrak{P}_2$ with the Clip-OGD policy. The expected sample sizes are presented in the first two rows of Table 2. Our findings show that $\mathfrak{P}_2$ requires a significantly lower expected sample size than Clip-OGD. This may be due to differences in assignment rules. Our policy adaptively assigns close to 21.9% of samples to treatment A, which is the asymptotically optimal fraction of the assignment of treatment A. Here, we have $\sigma(\mu_A) = 0.14$ and $\sigma(\mu_B) = 0.5$, and $\frac{0.14}{0.14+0.5} \approx 21.9\%$. In contrast, the Clip-OGD adaptively assigns based on the second moments, $S(A) = 0.98$ and $S(B) = 0.5$, and thus allocates close to $\frac{0.98}{0.98+0.5} \approx 66.2\%$ to treatment A, which is significantly sub-optimal.

For completeness, we also include the benchmark non-adaptive infeasible Neyman's allocation rule of Section 3 of Dai et al. (2024), i.e., assigning the treatment A to each individual with fixed probability 66.2%, and uses the same estimation and stopping rule as $\mathfrak{P}_2$. The performance of this policy is reported in row 3 of the table. As the benchmark non-adaptive rule always allocates a fixed suboptimal fraction to treatment A, the performance deteriorates even further.

8. Limitations and future research directions

As mentioned earlier, we show that Assumption 4.5 holds when outcome distributions are Gaussian. For other outcome distributions in $\mathcal{S}$, we numerically show that Assumption 4.5 holds.

Often in practical scenarios, individuals come with their contexts. Hence it is worth exploring a setting where the decision whether to observe a sample from distribution A or B is conditioned on the context of an incoming individual. Hence extending our work to a contextual setting is an important direction for future research. In this work, outcome distributions are restricted to SPEF or have bounded support for the asymptotic optimality of the two policies. A more general class of outcome distributions should be explored as part of future work.

Impact Statement

This paper presents research aimed at enhancing A/B testing through statistical learning methods, a practice prevalent across various fields. While our work holds numerous potential societal implications, we believe that none require explicit emphasis within this context.

References

Agrawal, S. Bandits with heavy tails algorithms analysis and optimality. 2022.

Agrawal, S., Juneja, S., and Glynn, P. Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory*, pp. 61–110. PMLR, 2020.

Audibert, J.-Y. and Bubeck, S. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pp. 13–p, 2010.

Barrier, A. *Contributions à une théorie de l'exploration pure en statistique séquentielle*. PhD thesis, Lyon, École normale supérieure, 2023.

Bhat, N., Farias, V. F., Moallemi, C. C., and Sinha, D. Near-optimal ab testing. *Management Science*, 66(10): 4477–4495, 2020.

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

Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. Kullback-leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, pp. 1516–1541, 2013.

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

Dai, J., Gradu, P., and Harshaw, C. Clip-ogd: An experimental design for adaptive neyman allocation in sequential experiments. *Advances in Neural Information Processing Systems*, 36, 2024.

Darling, D. A. and Robbins, H. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences*, 58(1):66–68, 1967.

Degenne, R. and Koolen, W. M. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.

Dembo, A. and Zeitouni, O. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

Even-Dar, E., Mannor, S., and Mansour, Y. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006.

Gardner, M. J. and Altman, D. G. Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)*, 292(6522):746–750, 1986.

Garivier, A. and Kaufmann, E. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pp. 998–1027, 2016.

Glynn, P. W., Johari, R., and Rasouli, M. Adaptive experimental design with temporal interference: A maximum likelihood approach. *Advances in Neural Information Processing Systems*, 33:15054–15064, 2020.

Hahn, J., Hirano, K., and Karlan, D. Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1):96–108, 2011.

Ham, D. W., Bojinov, I., Lindon, M., and Tingley, M. Design-based confidence sequences for anytime-valid causal inference. *arXiv preprint arXiv:2210.08639*, 2022.

Hayre, L. S. and Turnbull, B. W. Estimation of the odds ratio in the two-armed bandit problem. *Biometrika*, 68 (3):661–668, 1981.

Honda, J. and Takemura, A. An asymptotically optimal bandit algorithm for bounded support models. In *COLT*, pp. 67–79. Citeseer, 2010.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. Time-uniform, nonparametric, nonasymptotic confidence sequences. 2021.

Imbens, G. W. and Rubin, D. B. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Johari, R., Koomen, P., Pekelis, L., and Walsh, D. Always valid inference: Continuous monitoring of a/b tests. *Operations Research*, 70(3):1806–1821, 2022.

Jourdan, M., Degenne, R., Baudry, D., de Heide, R., and Kaufmann, E. Top two algorithms revisited. *Advances in Neural Information Processing Systems*, 35:26791–26803, 2022.

Juneja, S. and Krishnasamy, S. Sample complexity of partition identification using multi-armed bandits. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1824–1852, Phoenix, USA, 25–28 Jun 2019. PMLR. URL `http://proceedings.mlr.press/v99/juneja19a.html`.

Kato, M., Ishihara, T., Honda, J., and Narita, Y. Efficient adaptive experimental design for average treatment effect estimation. *arXiv preprint arXiv:2002.05308*, 2020.

Kaufmann, E. *Contributions to the Optimal Solution of Several Bandit Problems*. PhD thesis, Université de Lille, 2020.

Kaufmann, E. and Koolen, W. M. Mixture martingales revisited with applications to sequential tests and confidence intervals. *The Journal of Machine Learning Research*, 22 (1):11140–11183, 2021.

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

Kohavi, R. and Thomke, S. The surprising power of online experiments. *Harvard business review*, 95(5):74–82, 2017.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., and Pohlmann, N. Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1168–1176, 2013.

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

Lehmann, E. L. and Casella, G. *Theory of point estimation*. Springer Science & Business Media, 2006.

Lindon, M. and Malek, A. Anytime-valid inference for multinomial count data. *Advances in Neural Information Processing Systems*, 35:2817–2831, 2022.

Mannor, S. and Tsitsiklis, J. N. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5(Jun):623–648, 2004.

Naghshvar, M., Javidi, T., et al. Active sequential hypothesis testing. *The Annals of Statistics*, 41(6):2703–2738, 2013.

Neyman, J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. In *Breakthroughs in Statistics: Methodology and Distribution*, pp. 123–150. Springer, 1992.

Russo, D. Simple bayesian algorithms for best arm identification. In *Conference on Learning Theory*, pp. 1417–1418, 2016.

Simchi-Levi, D. and Wang, C. Multi-armed bandit experimental design: Online decision-making and adaptive inference. In *International Conference on Artificial Intelligence and Statistics*, pp. 3086–3097. PMLR, 2023.

Sundaram, R. K. *A first course in optimization theory*. Cambridge university press, 1996.

Takayama, A. *Analytical methods in economics*. University of Michigan Press, 1993.

Tang, D., Agarwal, A., O'Brien, D., and Meyer, M. Overlapping experiment infrastructure: More, better, faster experimentation. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 17–26, 2010.

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

Waudby-Smith, I., Arbour, D., Sinha, R., Kennedy, E. H., and Ramdas, A. Time-uniform central limit theory with applications to anytime-valid causal inference. *arXiv preprint arXiv:2103.06476*, 2021.

A. Proofs of results in Section 4

A.1. Proofs of results in Section 4.1.

Proof of Theorem 4.2:

Recall Ξ is the set of adaptive policies which satisfy the $(\epsilon, \delta)-$coverage property and are stable. Consider $\mathfrak{P} \in \Xi$. As the policy $\mathfrak{P}$ is stable, it follows that

$$\hat{\Delta}_L(\tau_\delta) \xrightarrow{p} \Delta_L \text{ and } \hat{\Delta}_R(\tau_\delta) \xrightarrow{p} \Delta_R, \tag{10}$$

under the environment ν, where Δ_L and Δ_R are constants such that $\Delta_R - \Delta_L \leq \epsilon$, and Δ_L and Δ_R lies within the interval $[\Delta - \epsilon, \Delta]$ and $[\Delta, \Delta + \epsilon]$ respectively. Let $\nu' = \{\nu'_A, \nu'_B\}$ denote an alternate environment. In this alternate environment, the means of these distributions are μ'_A and μ'_B, respectively. We first choose $\nu' \in \mathcal{K}_1(\Delta_L)$ with ATE Δ', i.e., $\Delta' = (\mu'_A - \mu'_B)$. Let $\eta \triangleq \Delta_L - \Delta'$. It follows that η is a positive number.

Using equation (33.6) in Lattimore & Szepesvári (2020) / Lemma 1 in Kaufmann et al. (2016), for any $\mathcal{E} \in \mathcal{F}_{\tau_\delta}$, we have

$$\mathbb{E}_\nu[N_A(\tau_\delta)] \cdot d(\mu_A, \mu'_A) + \mathbb{E}_\nu[N_B(\tau_\delta)] \cdot d(\mu_B, \mu'_B) \geq \Psi(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})),$$

It follows that,

$$\mathbb{E}_\nu[\tau_\delta] \left(\frac{\mathbb{E}_\nu[N_A(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_A, \mu'_A) + \frac{\mathbb{E}_\nu[N_B(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_B, \mu'_B) \right) \geq \Psi(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E})). \tag{11}$$

We claim that the following holds:

$$\liminf_{\delta \to 0} \frac{\Psi(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\nu'}(\mathcal{E}))}{\log(1/\delta)} \geq 1, \tag{12}$$

where, $\mathcal{E} = \{\Delta' \notin [\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]\}$, where, $\hat{\Delta}_R(\tau_\delta) - \hat{\Delta}_L(\tau_\delta) \leq \epsilon$.

To prove the claim made in (12), using Definition 3.1, we get $\mathbb{P}_{\nu'}(\mathcal{E}) \leq \delta$.

Now observe that,

$$\mathbb{P}_\nu(\mathcal{E}) = \mathbb{P}\{\Delta' \notin [\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]\} \geq \mathbb{P}\{\Delta_L - \eta < \hat{\Delta}_L(\tau_\delta)\}.$$

Using (10), we get, $\lim_{\delta \to 0} \mathbb{P}_\nu(\mathcal{E}) = 1$. Using the definition of $\Psi(\cdot, \cdot)$, proof of the claim follows trivially. Now we come back to our original proof.

Notice that (11) and (12) holds for any $\nu' \in \mathcal{K}_1(\Delta_L)$. Hence it follows that,

$$\liminf_{\delta \to 0} \left[\frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \right] \inf_{\nu' \in \mathcal{K}_1(\Delta_L)} \left(\frac{\mathbb{E}_\nu[N_A(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_A, \mu'_A) + \frac{\mathbb{E}_\nu[N_B(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_B, \mu'_B) \right) \geq 1. \tag{13}$$

Now we choose alternate environment $\nu' \in \mathcal{K}_2(\Delta_R)$ with ATE Δ', i.e., $\Delta' = (\mu'_A - \mu'_B)$. Let $\eta = \Delta' - \Delta_R$. Again it follows that η is a positive number. Using similar steps we get,

$$\liminf_{\delta \to 0} \left[\frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \right] \inf_{\nu' \in \mathcal{K}_2(\Delta_R)} \left(\frac{\mathbb{E}_\nu[N_A(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_A, \mu'_A) + \frac{\mathbb{E}_\nu[N_B(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_B, \mu'_B) \right) \geq 1. \tag{14}$$

Combining the (13) and (14), we get,

$$\liminf_{\delta \to 0} \left[\frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \right] \inf_{\nu' \in \mathcal{K}(\Delta_L, \Delta_R)} \left(\frac{\mathbb{E}_\nu[N_A(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_A, \mu'_A) + \frac{\mathbb{E}_\nu[N_B(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} d(\mu_B, \mu'_B) \right) \geq 1.$$

The above lower bound holds for a policy $\mathfrak{P}$, where $\frac{\mathbb{E}_\nu[N_A(\tau_\delta)]}{\mathbb{E}_\nu[\tau_\delta]} \in [0, 1]$. To get a lower bound for any adaptive stable policy with $(\epsilon, \delta)-$coverage guarantee, we have

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\sup_{\substack{w \in [0,1], \\ \Delta_L, \Delta_R : \Delta \in [\Delta_L, \Delta_R], \\ \Delta_R \leq \Delta_L + \epsilon}} \inf_{\nu' \in \mathcal{K}(\Delta_L, \Delta_R)} w d(\mu_A, \mu'_A) + (1 - w) d(\mu_B, \mu'_B)}.$$

To get the desired result, we need to show that, the above supremum will be achieved when $\Delta_R = \Delta_L + \epsilon$. To see that, $\mathcal{K}(\Delta_L, \Delta_L + \epsilon) \subseteq \mathcal{K}(\Delta_L, \Delta_R)$ for any $\Delta_R \leq \Delta_L + \epsilon$. Hence it follows that,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\ell^*(\mu_A, \mu_B, \epsilon)}.$$

This completes the first part of the proof.

Now we prove the second part of the theorem,

$$\ell^*(\mu_A, \mu_B, \epsilon) = \sup_{\substack{w \in [0,1], \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \inf_{\nu' \in \mathcal{K}(\Delta_L, \Delta_R)} wd(\mu_A, \mu'_A) + (1-w)d(\mu_B, \mu'_B).$$

where, $\mathcal{K}(\Delta_L, \Delta_R) = \mathcal{K}_1(\Delta_L) \cup \mathcal{K}_2(\Delta_R)$, and $\mathcal{K}_1(\Delta_L) = \{(\nu'_A, \nu'_B) : \nu'_A \in \mathcal{S}, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B < \Delta_L\}$ and $\mathcal{K}_2(\Delta_R) = \{(\nu'_A, \nu'_B) : \nu'_A \in \mathcal{S}, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B > \Delta_R\}$.

Now we define for $w \in [0, 1]$, $\Delta_L \in [\Delta - \epsilon, \Delta]$ and $\Delta_R = \Delta_L + \epsilon$,

$$\tilde{T}_1(\mu_A, \mu_B, w, \Delta_L) \triangleq \inf_{x, y \in \tilde{\mathcal{C}}_1(\Delta_L)} wd(\mu_A, x) + (1-w)d(\mu_B, y),$$

where, $\tilde{\mathcal{C}}_1(\Delta_L) = \{(x, y) : x, y \in \mathcal{I}, x - y < \Delta_L\}$.

$$\tilde{T}_2(\mu_A, \mu_B, w, \Delta_L) \triangleq \inf_{x, y \in \tilde{\mathcal{C}}_2(\Delta_R)} wd(\mu_A, x) + (1-w)d(\mu_B, y),$$

where, $\tilde{\mathcal{C}}_2(\Delta_R) = \{(x, y) : x, y \in \mathcal{I}, x - y > \Delta_R\}$.

Hence it follows that,

$$\ell^*(\mu_A, \mu_B, \epsilon) = \sup_{\substack{w \in [0,1], \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \min\{\tilde{T}_1(\mu_A, \mu_B, w, \Delta_L), \tilde{T}_2(\mu_A, \mu_B, w, \Delta_R)\}. \tag{15}$$

It is worth noting that for $\epsilon < 2(\overline{\mu} - \underline{\mu})$, for each $\Delta_L \in [\Delta - \epsilon, \Delta]$, at most one of the set $\tilde{\mathcal{C}}_1(\Delta_L)$ and $\tilde{\mathcal{C}}_1(\Delta_L + \epsilon)$ can be empty. In that case, we define inf over an empty set to be ∞.

Now we define for $w \in [0, 1]$, $\Delta_L \in [\Delta - \epsilon, \Delta]$ and $\Delta_R = \Delta_L + \epsilon$,

$$\hat{T}_1(\mu_A, \mu_B, w, \Delta_L) \triangleq \inf_{x, y \in \hat{\mathcal{C}}_1(\Delta_L)} wd(\mu_A, x) + (1-w)d(\mu_B, y), \tag{16}$$

where, $\hat{\mathcal{C}}_1(\Delta_L) = \{(x, y) : x, y \in \mathcal{I}, x - y \leq \Delta_L\}$.

$$\hat{T}_2(\mu_A, \mu_B, w, \Delta_R) \triangleq \inf_{x, y \in \hat{\mathcal{C}}_2(\Delta_R)} wd(\mu_A, x) + (1-w)d(\mu_B, y), \tag{17}$$

where, $\hat{\mathcal{C}}_2(\Delta_R) = \{(x, y) : x, y \in \mathcal{I}, x - y \geq \Delta_R\}$.

Now take any $x \in \mathcal{I}$ and $y \in \mathcal{I}$ such that $x - y = \Delta_L$ if such x, y exists. It follows that $\tilde{\mathcal{C}}_1(\Delta_L)$ does not contain $\{x, y\}$. But we one can construct a sequence of (x_n, y_n) for $n \geq 1$ such that $(x_n, y_n) \in \tilde{\mathcal{C}}_1(\Delta_L)$ and have the following property,

$$\lim_{n \to \infty} x_n = x, \ \lim_{n \to \infty} y_n = y.$$

Hence using continuity of $wd(\mu_A, x) + (1-w)d(\mu_B, y)$ in x and y, for $w \in [0,1]$, $\Delta_L \in [\Delta - \epsilon, \Delta]$ and $\Delta_R = \Delta_L + \epsilon$, we get,

$$\hat{T}_1(\mu_A, \mu_B, w, \Delta_L) = \tilde{T}_1(\mu_A, \mu_B, w, \Delta_L).$$

Similarly, one can get,

$$\hat{T}_2(\mu_A, \mu_B, w, \Delta_R) = \hat{T}_2(\mu_A, \mu_B, w, \Delta_R).$$

Hence it follows that,

$$\ell^*(\mu_A, \mu_B, \epsilon) = \sup_{\substack{w\in[0,1],\\ (\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \min\{\hat{T}_1(\mu_A, \mu_B, w, \Delta_L), \hat{T}_2(\mu_A, \mu_B, w, \Delta_R)\}. \tag{18}$$

Using strict quasi-convexity and uni-modality of $d(\mu, x)$ in x, we get that the solution of the optimization problem given in (16) and (17) exists. Hence it follows that for $w \in [0,1]$, $\Delta_L \in [\Delta - \epsilon, \Delta]$ and $\Delta_R = \Delta_L + \epsilon$,

$$\hat{T}_1(\mu_A, \mu_B, w, \Delta_L) \triangleq \min_{x,y\in\hat{\mathcal{C}}_1(\Delta_L)} wd(\mu_A, x) + (1-w)d(\mu_B, y).$$

$$\hat{T}_2(\mu_A, \mu_B, w, \Delta_R) \triangleq \min_{x,y\in\hat{\mathcal{C}}_2(\Delta_R)} wd(\mu_A, x) + (1-w)d(\mu_B, y).$$

Now again using strict quasi-convexity and uni-modality of $d(\mu, x)$ in x and (18), we get,

$$\ell^*(\mu_A, \mu_B, \epsilon) = \sup_{\substack{w\in[0,1],\\ (\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_R)\}. \tag{19}$$

This completes the proof.

□

Remark A.1. To get a non-asymptotic lower bound for any adaptive policy with $(\epsilon, \delta)-$coverage guarantee, for a given environment $\nu = \{\nu_A, \nu_B\}$, recall we have $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2$, where $\mathcal{K}_1 = \{(\nu'_A, \nu'_B) : \nu'_A, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B < \Delta - \epsilon\}$ and $\mathcal{K}_2 = \{(\nu'_A, \nu'_B) : \nu'_A, \nu'_B \in \mathcal{S}, \mu'_A - \mu'_B > \Delta + \epsilon\}$.

Now first fix an alternate environment $\nu' \in \mathcal{K}$, and then fix $\mathcal{E} = \{\Delta' \notin [\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]\}$, where $\Delta' = \mu'_A - \mu'_B$. Since for $\nu' \in \mathcal{K}$, $\Delta' < \Delta - \epsilon$ or $\Delta' > \Delta + \epsilon$, hence it follows using the fact that $\hat{\Delta}_R(\tau_\delta) - \hat{\Delta}_L(\tau_\delta) \leq \epsilon$,

$$\mathbb{P}_\nu(\mathcal{E}) \geq \mathbb{P}_\nu(\Delta \in [\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]) \geq 1 - \delta.$$

The last inequality in the above expression follows using $(\epsilon, \delta)-$coverage guarantee. Further, it follows trivially that $\mathbb{P}_{\nu'}(\mathcal{E}) \leq \delta$ using the $(\epsilon, \delta)-$coverage guarantee. Now using similar steps to the proof of this theorem and the fact that $\Psi(\delta, 1-\delta) \geq \log(1/4\delta)$ for $\delta \in (0,1)$, we get,

$$\frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/4\delta)} \geq \frac{1}{\sup_{w\in[0,1]} \inf_{\nu'\in\tilde{\mathcal{K}}} wd(\mu_A, \mu'_A) + (1-w)d(\mu_B, \mu'_B)}. \tag{20}$$

Since for any $\Delta_L \in [\Delta - \epsilon, \Delta]$ and $\Delta_R \in [\Delta, \Delta + \epsilon]$, it can be shown that $\tilde{\mathcal{K}} \subset \mathcal{K}(\Delta_L, \Delta_R)$, hence we have $\ell^*(\mu_A, \mu_B, \epsilon) < \sup_{w\in[0,1]} \inf_{\nu'\in\tilde{\mathcal{K}}} wd(\mu_A, \mu'_A) + (1-w)d(\mu_B, \mu'_B)$, which implies that this lower bound is not tight for adaptive stable policies with $(\epsilon, \delta)-$coverage guarantee as $\delta \to 0$.

A.2. Proofs of results in Section 4.2

Proof of Theorem 4.4(a) Given $\delta \in (0,1)$, we define event $\mathcal{E}_1 = \{\tau_\delta = \infty\}$. We need to show that, $\mathbb{P}(\mathcal{E}_1) = 0$. We prove it by contradiction, suppose $\mathbb{P}(\mathcal{E}_1) > 0$. Using the definition of τ_δ on any sample path in $\mathcal{E}_1$, it follows that $\forall\ n \in \mathbb{N}$, $\hat{\Delta}_R(n) - \hat{\Delta}_L(n) > \epsilon$. Since $\lim_{n\to\infty} \frac{\beta(n,\delta)}{n} = 0$, it follows from the definition that,

$$\lim_{n\to\infty} T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_L(n)\right) = T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_R(n)\right) = 0. \tag{21}$$

Using Lemma E.8, we get that, $\lim_{n\to\infty} \frac{N_A(n)}{n} = w^*(\mu_A, \mu_B, \epsilon)$, $\lim_{n\to\infty} \hat{\mu}_A(n) = \mu_A$ and $\lim_{n\to\infty} \hat{\mu}_B(n) = \mu_B$ almost surely.

Since we know that,

$$\underline{\mu} - \overline{\mu} < \hat{\Delta}_L(n) \le \hat{\mu}_A(n) - \hat{\mu}_B(n) \text{ and } \hat{\mu}_A(n) - \hat{\mu}_B(n) \le \hat{\Delta}_R(n) < \overline{\mu} - \underline{\mu}.$$

Using continuity of $d(\mu, x)$ in x, it follows that,

$$\underline{\mu} - \overline{\mu} \le \limsup_{n\to\infty} \hat{\Delta}_L(n) \le \mu_A - \mu_B \text{ and } \mu_A - \mu_B \le \liminf_{n\to\infty} \hat{\Delta}_R(n) \le \overline{\mu} - \underline{\mu}. \tag{22}$$

Now we claim that,

$$\lim_{n\to\infty} \hat{\Delta}_L(n) = \lim_{n\to\infty} \hat{\Delta}_R(n) = \Delta = \mu_A - \mu_B \text{ almost surely.}$$

This implies that $\lim_{n\to 0} \hat{\Delta}_R(n) - \hat{\Delta}_L(n) = 0$ which leads to contradiction as under the event $\mathcal{E}_1$, as we have $\hat{\Delta}_R(n) - \hat{\Delta}_L(n) \ge \epsilon$ for all $n \in \mathbb{N}$ under event $\mathcal{E}_1$. To complete the proof, we only need to prove the claim.

To prove the claim, we use the method of contradiction. Suppose, $\lim_{n\to\infty} \hat{\Delta}_L(n) \neq \mu_A - \mu_B$. A similar proof will follows for the case $\lim_{n\to\infty} \hat{\Delta}_R(n) \neq \mu_A - \mu_B$.

Using (22), we get that,

$$\limsup_{n\to\infty} \hat{\Delta}_L(n) = c \in [\underline{\mu} - \overline{\mu}, \Delta).$$

It is worth noting that $\underline{\mu} - \overline{\mu}$ can be ∞, depending the support of the family of distributions of outcome of treatments that we are considering within the $\mathcal{S}$. It implies that there exists a subsequence $\hat{\Delta}_L(n_k)$ of $\hat{\Delta}_L(n)$ such that $\lim_{k\to\infty} \hat{\Delta}_L(n_k) \to c$ and $\lim_{k\to\infty} n_k = \infty$.

Using Lemma E.2, we know that $T(\mu_A, \mu_B, w, \Delta_L)$ is jointly continuous function in $(w, \Delta_L, \mu_A, \mu_B)$ for $w \in [0,1]$, $\Delta_L \in (\underline{\mu} - \overline{\mu}, \overline{\mu} - \underline{\mu})$, $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$. Hence we have,

$$\lim_{k\to\infty} T\left(\hat{\mu}_A(n_k), \hat{\mu}_B(n_k), \frac{N_A(n_k)}{n_k}, \hat{\Delta}_L(n_k)\right) = T(\mu_A, \mu_B, w^*(\mu_A, \mu_B, \epsilon), c). \tag{23}$$

Since $c \neq \Delta$, hence using Lemma E.3, we get,

$$\lim_{k\to\infty} T\left(\hat{\mu}_A(n_k), \hat{\mu}_B(n_k), \frac{N_A(n_k)}{n_k}, \hat{\Delta}_L(n_k)\right) \neq 0.$$

This leads to a contradiction with (21). This completes the proof.

□

Proof of Theorem 4.4(b) We first state the following theorem, provide its proof then provide the proof of Theorem 4.4(b).

Theorem A.2. *The CI of ATE given by, $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$ by $\mathfrak{P}_1$ is a $(1-\delta)-$confidence sequence.*

Using the definition of $(1-\delta)-$confidence sequence, it suffices to show that,

$$\mathbb{P}\{\exists n \in \mathbb{Z}^+ : \Delta \notin [\hat{\Delta}_L(n), \hat{\Delta}_R(n)]\} \leq \delta.$$

Recall that,

$$T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_L(n)\right) = T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_R(n)\right) = \frac{\beta(n,\delta)}{n}.$$

Using the definition of $T(\mu_A, \mu_B, w, z)$ and Lemma E.3, we have,

$$\{\Delta \notin [\hat{\Delta}_L(n), \hat{\Delta}_R(n)]\} \subseteq \{N_A(n)d(\hat{\mu}_A(n), \mu_A) + N_B(n)d(\hat{\mu}_B(n), \mu_B) \geq \beta(n,\delta)\}.$$

Hence it suffices to show that,

$$\mathbb{P}\{\exists n \in \mathbb{Z}^+ : N_A(n)d(\hat{\mu}_A(n), \mu_A) + N_B(n)d(\hat{\mu}_B(n), \mu_B) \geq \beta(n,\delta)\} \leq \delta.$$

For $\beta(n\,\delta) = \log\left(\frac{c_1 n^\alpha}{\delta}\right)$, we can use Proposition 12 in Garivier & Kaufmann (2016) to get that $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$ is a $(1-\delta)-$confidence sequence. This completes the proof.

□

Now we prove Theorem 4.4(b). To get the $(\epsilon, \delta)-$coverage guarantee, we need to show that,

$$\mathbb{P}\{\Delta \notin [\hat{\Delta}_L(\tau_\delta), \hat{\Delta}_R(\tau_\delta)]\} \leq \delta.$$

Using part (a) of this theorem, it suffices to show that,

$$\mathbb{P}\{\exists n \in \mathbb{Z}^+ : \Delta \notin [\hat{\Delta}_L(n), \hat{\Delta}_R(n)]\} \leq \delta.$$

This follows from Theorem A.2. This completes the proof.

□

Proof of Theorem 4.4(c)

First we show that,

$$\lim_{\delta \to 0} \tau_\delta = \infty \text{ almost surely.} \tag{24}$$

To prove this, we use the method of contradiction. Hence, we define an event $\mathcal{E}_2 = \{\liminf_{\delta\to 0} \tau_\delta < \infty\}$ and assume $\mathbb{P}(\mathcal{E}_2) > 0$. Recall the definition of $\widehat{\Delta}_L(\tau_\delta)$ and $\widehat{\Delta}_R(\tau_\delta)$,

$$T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \widehat{\Delta}_L(\tau_\delta)\right) = T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \widehat{\Delta}_R(\tau_\delta)\right) = \frac{\beta(\tau_\delta, \delta)}{\tau_\delta}.$$

Let $f_1(w, \mu_A, \mu_B, x, y) = wd(\mu_A, x) + (1-w)d(\mu_B, y)$. Using the fact that $\lim_{\delta\to 0} \beta(\tau_\delta, \delta) = \infty$ for any sample path in $\mathcal{E}_2$. Hence we get, for any sample path in $\mathcal{E}_2$, there exists a subsequence τ_{δ_n} for $n \in \mathbb{Z}^+$ such that $\lim_{n\to\infty} \delta_n = 0$, $\lim_{n\to\infty} \tau_{\delta_n} < \infty$ and

$$\begin{aligned}
\lim_{n\to\infty} \min_{x,y \in \mathcal{C}(\widehat{\Delta}_L(\tau^\pi_{\delta_n}))} f_1\left(\frac{N_A(\tau_{\delta_n})}{\tau_{\delta_n}}, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) = \infty, \text{ and} \\
\lim_{n\to\infty} \min_{x,y \in \mathcal{C}(\widehat{\Delta}_R(\tau_{\delta_n}))} f_1\left(\frac{N_A(\tau_{\delta_n})}{\tau_{\delta_n}}, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) = \infty.
\end{aligned} \tag{25}$$

Using the definition of $\widehat{\Delta}_L(\tau_{\delta_n})$ and $\widehat{\Delta}_R(\tau_{\delta_n})$, we know that, $\widehat{\Delta}_R(\tau_{\delta_n}) - \widehat{\Delta}_L(\tau_{\delta_n}) \leq \epsilon$, hence it follows that,

$$\min_{x,y \in \mathcal{C}(\widehat{\Delta}_L(\tau_{\delta_n})) \cup \mathcal{C}(\widehat{\Delta}_R(\tau_{\delta_n}))} f_1\left(\frac{N_A(\tau_{\delta_n})}{\tau_{\delta_n}}, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) \leq$$
$$\sup_{\substack{w\in(0,1) \\ \Delta_L \in [\Delta-\epsilon, \Delta], \Delta_R \in [\Delta, \Delta+\epsilon]: \Delta_R - \Delta_L \leq \epsilon}} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right).$$

Using an argument similar to the proof of Theorem 4.2, we get,

$$\sup_{\substack{w\in(0,1) \\ \Delta_L \in [\Delta-\epsilon, \Delta], \Delta_R \in [\Delta, \Delta+\epsilon]: \Delta_R - \Delta_L \leq \epsilon}} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) =$$
$$\sup_{\substack{w\in(0,1) \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right).$$

Hence it follows that we have,

$$\min_{x,y \in \mathcal{C}(\widehat{\Delta}_L(\tau_{\delta_n})) \cup \mathcal{C}(\widehat{\Delta}_R(\tau_{\delta_n}))} f_1\left(\frac{N_A(\tau_{\delta_n})}{\tau_{\delta_n}}, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) \leq \tag{26}$$
$$\sup_{\substack{w\in(0,1) \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right).$$

Now using max-min inequality we get,

$$\sup_{\substack{w\in(0,1) \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) \leq \tag{27}$$
$$\sup_{\Delta_L \in [\Delta-\epsilon, \Delta]} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_L+\epsilon)} \sup_{w\in(0,1)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right).$$

Using some algebra and strict convexity of $d(\mu, x)$ in x, we get,

$$\sup_{\Delta_L \in [\Delta-\epsilon, \Delta]} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_L+\epsilon)} \sup_{w\in(0,1)} f_1\left(w, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) \leq$$
$$\max_{k\in\{A,B\}} \min\{d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) - \epsilon), d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) + \epsilon)\}.$$

Define $d(x,y) = \infty$ for $x \in \mathcal{I}$ and $y \notin \mathcal{I}$. Combining above, (26) and (27), we get for $\epsilon < \frac{(\overline{\mu} - \underline{\mu})}{2}$,

$$\min_{x,y \in \mathcal{C}(\widehat{\Delta}_L(\tau_{\delta_n})) \cup \mathcal{C}(\widehat{\Delta}_R(\tau_{\delta_n}))} f_1\left(\frac{N_A(\tau_{\delta_n})}{\tau_{\delta_n}}, \hat{\mu}_A(\tau_{\delta_n}), \hat{\mu}_B(\tau_{\delta_n}), x, y\right) \leq$$
$$\max_{k\in\{A,B\}} \min\{d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) - \epsilon), d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) + \epsilon)\}.$$

Now we claim that,

$$\lim_{n\to\infty} \max_{k\in\{A,B\}} \min\{d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) - \epsilon), d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) + \epsilon)\} < \infty. \tag{28}$$

It follows using the above claim that, we get a contradiction with (25), which completes the proof of (24). We now show that the above claim stated in (28) holds. Observe that for a given sample path in $\mathcal{E}_2$, we have $\lim_{n\to\infty} \tau_{\delta_n} < \infty$. Since τ_{δ_n} is a sequence on the space of positive integers, hence it follows that for a given sample path in $\mathcal{E}_2$, $\lim_{n\to\infty} \tau_{\delta_n} = \tau_{\delta_{n^*}}$ for some

$n^* \in \mathbb{Z}^+$. Recall that $\hat{\mu}_k(\tau_{\delta_n}) \in \mathcal{I}$, further for $\epsilon < \frac{(\overline{\mu}-\underline{\mu})}{2}$, either $\hat{\mu}_k(\tau_{\delta_n}) - \epsilon \in \mathcal{I}$ or $\hat{\mu}_k(\tau_{\delta_n}) + \epsilon \in \mathcal{I}$ for $k \in \{A, B\}$ for all $n \in \mathbb{Z}^+$. Hence, using the continuity of $d(x, y)$ in (x, y) for $x \in \mathcal{I}$ and $y \in \mathcal{I}$, we have,

$$\begin{aligned}\lim_{n\to\infty} \max_{k\in\{A,B\}} \min\{d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) - \epsilon), d(\hat{\mu}_k(\tau_{\delta_n}), \hat{\mu}_k(\tau_{\delta_n}) + \epsilon)\} = \\ \max_{k\in\{A,B\}} \min\{d(\hat{\mu}_k(\tau_{\delta_{n^*}}), \hat{\mu}_k(\tau_{\delta_{n^*}}) - \epsilon), d(\hat{\mu}_k(\tau_{\delta_{n^*}}), \hat{\mu}_k(\tau_{\delta_{n^*}}) + \epsilon)\}.\end{aligned} \tag{29}$$

Recall that $\hat{\mu}_k(\tau_{\delta_n}) \in \mathcal{I}$ and for $\epsilon < \frac{(\overline{\mu}-\underline{\mu})}{2}$, either $\hat{\mu}_k(\tau_{\delta_n}) - \epsilon \in \mathcal{I}$ or $\hat{\mu}_k(\tau_{\delta_n}) + \epsilon \in \mathcal{I}$ for $k \in \{A, B\}$ for all $n \in \mathbb{Z}^+$. Hence it follows that, $\max_{k\in\{A,B\}} \min\{d(\hat{\mu}_k(\tau_{\delta_{n^*}}), \hat{\mu}_k(\tau_{\delta_{n^*}}) - \epsilon), d(\hat{\mu}_k(\tau_{\delta_{n^*}}), \hat{\mu}_k(\tau_{\delta_{n^*}}) + \epsilon)\} < \infty$.

Combining the above with (29) we get that the claim stated in (28) holds.

Now we come back to our original proof. Using (24), Lemma E.8 we get,

$$\hat{\mu}_A(\tau_\delta) \to \mu_A \text{ and } \hat{\mu}_B(\tau_\delta) \to \mu_B \text{ almost surely as } \delta \to 0. \tag{30}$$

It follows that, $\hat{\Delta}_L(\tau_\delta)$ will satisfy the following: $\hat{\Delta}(\tau_\delta) - \epsilon \leq \hat{\Delta}_L(\tau_\delta) \leq \hat{\Delta}(\tau_\delta)$, where $\hat{\Delta}(\tau_\delta) = \hat{\mu}_A(\tau_\delta) - \hat{\mu}_B(\tau_\delta)$.

Using (30), it follows that $\hat{\Delta}_L(\tau_\delta)$ is a bounded sequence on each sample path. Similarly, $\hat{\Delta}_R(\tau_\delta)$ is a bounded sequence too on each sample path.

Using the definition of τ_δ, $\hat{\Delta}_L(\tau_\delta)$ and $\hat{\Delta}_R(\tau_\delta)$, it follows that,

$$\hat{\Delta}_R(\tau_\delta) \leq \hat{\Delta}_L(\tau_\delta) + \epsilon \text{ and } \hat{\Delta}_R(\tau_\delta - 1) > \hat{\Delta}_L(\tau_\delta - 1) + \epsilon. \tag{31}$$

$$T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_L(\tau_\delta)\right) = T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_R(\tau_\delta)\right). \tag{32}$$

Using Lemma E.3, (31) and (32), we get,

$$T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_L(\tau_\delta)\right) \leq T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_L(\tau_\delta) + \epsilon\right). \tag{33}$$

$$T\left(\hat{\mu}_A(\tau_\delta - 1), \hat{\mu}_B(\tau_\delta - 1), \frac{N_A(\tau_\delta - 1)}{\tau_\delta - 1}, \hat{\Delta}_L(\tau_\delta - 1)\right) \geq T\left(\hat{\mu}_A(\tau_\delta - 1), \hat{\mu}_B(\tau_\delta - 1), \frac{N_A(\tau_\delta - 1)}{\tau_\delta - 1}, \hat{\Delta}_L(\tau_\delta - 1) + \epsilon\right). \tag{34}$$

Now we show that $\lim_{\delta\to 0} \hat{\Delta}_L(\tau_\delta) = \Delta_L^*(\mu_A, \mu_B, \epsilon)$ almost surely. A similar proof will follow for $\lim_{\delta\to 0} \hat{\Delta}_R(\tau_\delta) = \Delta_R^*(\mu_A, \mu_B, \epsilon)$ on each sample path. This will imply the stability of the policy $\mathfrak{P}_1$.

We prove it by contradiction, suppose there exists a sequence $\hat{\Delta}_L(\tau_\delta)$ does not converge to $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ on a positive measure set $\mathcal{E}_3$, i.e., $\mathbb{P}(\mathcal{E}_3) > 0$. Fix any sample path in $\mathcal{E}_3$. Since $\hat{\Delta}_L(\tau_\delta)$ is a bounded sequence and hence we assume $\limsup_{\delta\to 0} \hat{\Delta}_L(\tau_\delta) = \overline{\Delta}$ and $\liminf_{\delta\to 0} = \underline{\Delta}$. Without loss of generality, we assume that $\overline{\Delta} \neq \Delta_L^*(\mu_A, \mu_B, \epsilon)$.

Since $\hat{\Delta}_L(\tau_\delta)$ is a bounded sequence, hence there will exist a sub-sequence $\{\delta_k,\ k \in \mathbb{Z}^+\}$ and $\delta_k \in (0, 1)$ such that $\lim_{k\to\infty} \delta_k = 0$ and $\lim_{k\to\infty} \hat{\Delta}_L(\tau_{\delta_k}) = \overline{\Delta}$. Using Lemma E.8, on this subsequence, $\lim_{\delta\to 0} \frac{N_A(\tau_{\delta_k})}{\tau_{\delta_k}} = w^*(\mu_A, \mu_B, \epsilon)$. Using (33) and (34) on the sub-sequence defined above and Lemma E.2, we get that,

$$\begin{aligned}\lim_{k\to\infty} T\left(\hat{\mu}_A(\tau_{\delta_k}), \hat{\mu}_B(\tau_{\delta_k}), \frac{N_A(\tau_{\delta_k})}{\tau_{\delta_k}}, \hat{\Delta}_L(\tau_{\delta_k})\right) &= \lim_{k\to\infty} T\left(\hat{\mu}_A(\tau_{\delta_k}), \hat{\mu}_B(\tau_{\delta_k}), \frac{N_A(\tau_{\delta_k})}{\tau_{\delta_k}}, \hat{\Delta}_L(\tau_{\delta_k}) + \epsilon\right) \\ &= T(w^*(\mu_A, \mu_B, \epsilon), \overline{\Delta}, \mu_A, \mu_B) = T(w^*(\mu_A, \mu_B, \epsilon), \overline{\Delta} + \epsilon, \mu_A, \mu_B).\end{aligned}$$

Using Lemma E.5, for a given μ_A, μ_B and $w^*(\mu_A, \mu_B, \epsilon)$, we know that there is a unique solution of the above equation and it equals $\Delta_L^*(\mu_A, \mu_B, \epsilon)$. Hence we get a contradiction. This completes the proof.

□

Proof of Theorem 4.7:

a) Almost sure sample size analysis. Observe that $\hat{\Delta}_L(\tau_\delta)$ and $\hat{\Delta}_R(\tau_\delta)$ satisfy the following:

$$T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_L(\tau_\delta)\right) = T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_R(\tau_\delta)\right) = \frac{\beta(\tau_\delta, \delta)}{\tau_\delta}. \tag{35}$$

Fix a sample path. Using (30), we get that $\lim_{\delta\to 0} \hat{\mu}_A(\tau_\delta) = \mu_A$ and $\lim_{\delta\to 0} \hat{\mu}_B(\tau_\delta) = \mu_B$. Using Lemma E.8 and the part (c) of Theorem 4.4, we get, $\lim_{\delta\to 0} \frac{N_A(\tau_\delta)}{\tau_\delta} = w^*(\mu_A, \mu_B, \epsilon)$, $\lim_{\delta\to 0} \hat{\Delta}_L(\tau_\delta) = \Delta_L^*(\mu_A, \mu_B, \epsilon)$ and $\lim_{\delta\to 0} \hat{\Delta}_R(\tau_\delta) = \Delta_L^*(\mu_A, \mu_B, \epsilon) + \epsilon$. Using Lemma E.2, we have,

$$\begin{aligned}\lim_{\delta\to 0} T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_L(\tau_\delta)\right) &= \lim_{\delta\to 0} T\left(\hat{\mu}_A(\tau_\delta), \hat{\mu}_B(\tau_\delta), \frac{N_A(\tau_\delta)}{\tau_\delta}, \hat{\Delta}_R(\tau_\delta)\right) \\ &= \ell^*(\mu_A, \mu_B, \epsilon).\end{aligned}$$

Hence we have,

$$\lim_{\delta\to 0} \frac{\beta(\tau_\delta, \delta)}{\tau_\delta} = \ell^*(\mu_A, \mu_B, \epsilon) \text{ almost surely.}$$

Since $\beta(n\,\delta) = \log\left(\frac{c_1 n^\alpha}{\delta}\right)$, hence we get the desired result. This completes the proof.

□

b) Expected sample size analysis.

To get the results of convergence in expectation from almost sure, we will show that collection of random variables indexed by δ, $\frac{\tau_\delta}{\log(1/\delta)}$ is uniformly integrable, which will complete the proof. Hence it suffices to show $\sup_{\delta\in(0,\delta_1)} \frac{\mathbb{E}[\tau(\delta)]^2}{(\log(1/\delta))^2} < \infty$, where, δ_1 is any small fixed number in $(0, 1)$. Observe that,

$$\mathbb{E}[\tau_\delta]^2 = \sum_{n=1}^{\infty} (2n-1)\mathbb{P}(\tau_\delta > n).$$

We start the analysis by the construction of the following set. Let $\mathcal{G}_n(\eta) = \cap_{i=h(n)}^n \max_{k\in\{A,B\}} |\hat{\mu}_k(i) - \mu_k| \le \zeta_1$, where ζ_1 is chosen to satisfy the following condition for a given small positive number $\eta > 0$. Here $h(n) = n^{1/4}$.

$$\max_{k\in\{A,B\}} |\mu_k' - \mu_k| \le \zeta_1 \implies |w^*(\mu_A', \mu_B', \epsilon) - w^*(\mu_A, \mu_B, \epsilon)| \le \eta.$$

Above holds true from Lemma E.7 and Lemma E.7. Hence it follows that,

$$\mathbb{E}[\tau_\delta]^2 \le \sum_{n=1}^{\infty} (2n-1)\mathbb{P}(\{\tau_\delta > n\} \cap \mathcal{G}_n(\eta)) + \sum_{n=1}^{\infty} (2n-1)\mathbb{P}(\mathcal{G}_n(\eta)^c). \tag{36}$$

We will handle the two series summations given in (36) separately, then we will come back to (36).

Upper bound on $\sum_{n=1}^{\infty} (2n-1)\mathbb{P}(\mathcal{G}_n(\eta)^c)$: Observe that,

$$\mathbb{P}(\mathcal{G}_n(\eta)^c) \le \sum_{i=h(n)}^{n} \sum_{k\in\{A,B\}} (\mathbb{P}(\hat{\mu}_k(i) \le \mu_k - \xi) + \mathbb{P}(\hat{\mu}_k(i) \ge \mu_k + \xi)).$$

Using Lemma E.10, we get that $N_k(i) \geq \sqrt{i} - 2$ for $i > 4$, we get,

$$\mathbb{P}(\hat{\mu}_k(i) \leq \mu_k - \xi) = \mathbb{P}(\hat{\mu}_k(i) \leq \mu_k - \xi \cap N_k(i) > \sqrt{i} - 2).$$

Using Chernoff's inequality for outcome distributions in $\mathcal{S}$,

$$\mathbb{P}(\hat{\mu}_k(i) \leq \mu_k - \xi) \leq \sum_{s=\sqrt{i}-1}^{i} e^{-sd(\mu_k - \xi, \mu_a)}.$$

Similarly,

$$\mathbb{P}(\hat{\mu}_k(i) \geq \mu_k + \xi) \leq \sum_{s=\sqrt{i}-1}^{i} e^{-sd(\mu_k + \xi, \mu_a)}.$$

Using some algebra, we get,

$$\mathbb{P}(\mathcal{G}_n(\eta)^c) \leq c_1 n e^{-c_2 n^{1/8}},$$

where c_1 and c_2 are well-chosen positive constants. Hence it follows that,

$$\sum_{n=1}^{\infty}(2n-1)\mathbb{P}(\mathcal{G}_n(\eta)^c) \leq \sum_{n=1}^{\infty} c_1 n(2n-1)e^{-c_2 n^{1/8}} \leq c_3, \tag{37}$$

where c_3 is a well-chosen positive constant which is independent of δ. It follows that the right most inequality in (37) holds as one can upper bound $\sum_{n=1}^{\infty} c_1 n(2n-1)e^{-c_2 n^{1/8}}$ using the definition of the Gamma function.

Upper bound on $\sum_{n=1}^{\infty}(2n-1)\mathbb{P}(\{\tau_\delta > n\} \cap \mathcal{G}_n(\eta))$: We claim that after $N(\delta) \triangleq O(\log(1/\delta))$ terms $\mathbb{P}(\{\tau_\delta > n\}$ will be 0 under the set $\mathcal{G}_n(\eta)$, i.e,

$$\mathbb{P}(\{\tau_\delta > n\} \cap \mathcal{G}_n(\eta)) = 0 \;\forall n \geq N(\delta). \tag{38}$$

Using (38), we get,

$$\sum_{n=1}^{\infty}(2n-1)\mathbb{P}(\{\tau_\delta > n\} \cap \mathcal{G}_n(\eta)) \leq \sum_{n=1}^{N(\delta)}(2n-1)\mathbb{P}(\tau_\delta > n) \leq O((\log(1/\delta))^2).$$

Combining the above inequality with (38), and using (37) and substituting them in (36), we get

$$\sup_{\delta \in (0,\delta_1)} \frac{\mathbb{E}[\tau(\delta)]^2}{(\log(1/\delta))^2} \leq \sup_{\delta \in (0,\delta_1)} \frac{O((\log(1/\delta))^2) + c_3}{(\log(1/\delta))^2} < \infty.$$

To complete the proof, all we need to show is that our claim (38) holds. To prove (38), observe that, we will show that, there exists a $N(\delta)$ which is $O(\log(1/\delta))$, for $n \geq N(\delta)$, one has $\mathcal{G}_n(\eta) \subseteq \{\tau_\delta \leq n\}$.

First, we define,

$$V(\mu_A, \mu_B, \epsilon) \triangleq \inf_{\substack{\mu'_A \in [\mu_A - \xi(\eta) + \mu_A + \xi(\eta)] \\ \mu'_B \in [\mu_B - \xi(\eta) + \mu_B + \xi(\eta)] \\ w' \in [w^*(\mu_A, \mu_B, \epsilon) - 3\eta + w^*(\mu_A, \mu_B, \epsilon) + 3\eta]}} J(\mu'_A, \mu'_B, w', \epsilon),$$

where,

$$J(\mu_A, \mu_B, w, \epsilon) = \max_{\Delta_L \in [\Delta - \epsilon, \Delta]} \min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}. \tag{39}$$

Using the the proof of Lemma E.5, we get,

$$J(\mu_A, \mu_B, w, \epsilon) = T(\mu_A, \mu_B, w, \tilde{\Delta}_L(\mu_A, \mu_B, w, \epsilon)),$$

where, $\tilde{\Delta}_L(\mu_A, \mu_B, w, \epsilon))$ uniquely satisfies,

$$T(\mu_A, \mu_B, w, \tilde{\Delta}_L(\mu_A, \mu_B, w, \epsilon)) = T(\mu_A, \mu_B, w, \tilde{\Delta}_L(\mu_A, \mu_B, w, \epsilon) + \epsilon).$$

First, we show that $J(\mu_A, \mu_B, w, \epsilon)$ is a jointly continuous function of (μ_A, μ_B, w). To see, we know from Lemma E.2 that $\min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}$ is a continuous function in $(\mu_A, \mu_B, w, \Delta_L)$ for $w \in (0, 1)$, $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$ and $\Delta_L \in (\underline{\mu} - \overline{\mu}, \overline{\mu} - \underline{\mu})$. Hence using (39) and Berge's Maximum Theorem, we get that $J(\mu_A, \mu_B, w, \epsilon)$ is a continuous function in in (μ_A, μ_B, w) for $w \in (0, 1)$, $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$. Since $\tilde{\Delta}_L(\mu_A, \mu_B, w, \epsilon)) \in (\Delta - \epsilon, \Delta)$ and $T(\mu_A, \mu_B, w, \tilde{\Delta}_L(\mu_A, \mu_B, w, \epsilon)) > 0$, hence $J(\mu_A, \mu_B, w, \epsilon) > 0$ for $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$ and $w \in (0, 1)$. Hence it follows that $V(\mu_A, \mu_B, \epsilon) > 0$.

Now we define $N(\delta)$ as follows,

$$N(\delta) \triangleq \inf\left\{n \in \mathbb{N} : n \geq n_\eta, \frac{\beta(n, \delta)}{V(\mu_A, \mu_B, \epsilon)} + \sqrt{n} \leq n\right\}.$$

Here n_η is defined in Lemma E.10. Using the definition of $\beta(n, \delta)$ and above, it follows that $N(\delta) = O(\log(1/\delta))$. Now we show that for $n \geq N(\delta)$, one has $\mathcal{G}_n(\eta) \subseteq \{\tau_\delta \leq n\}$. Using Lemma E.10, Lemma E.3 and the definition of set $\mathcal{G}_n(\eta)$, observe that for $n \geq N(\delta)$, we have under the set $\mathcal{G}_n(\eta)$,

$$\hat{\Delta}_R(n) - \hat{\Delta}_L(n) \leq \epsilon.$$

This implies that $\mathcal{G}_n(\eta) \subseteq \{\tau_\delta \leq n\}$. This completes the proof.

□

B. Proofs of results in Section 5.

For a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$, $w \in (0, 1)$ and $z \in [\Delta - \epsilon, \Delta + \epsilon]$, we denote any solution of (3) as $x^*(\mu_A, \mu_B, w, z)$ (in Lemma E.4, we show that it uniquely exists for small ϵ).

Proof of Preposition 5.1

If $\nu_A = N(\mu_A, \sigma_A^2)$ and $\nu_2 = N(\mu_B, \sigma_B^2)$, where σ_A^2 and σ_B^2 are known, then $\mathfrak{L}$ takes following form,

$$\max_{\substack{w\in[0,1],\\ (\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \min_{x,y\in\mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)} w\frac{(x-\mu_A)^2}{2\sigma_A^2} + (1-w)\frac{(y-\mu_B)^2}{2\sigma_B^2},$$

where $\mathcal{C}(z) = \{(x, y) : x, y \in \mathcal{I}, x - y = z\}$.

First, we solve the following optimization problem.

$$\min_{x,y\in\mathcal{C}(\Delta_L)} w\frac{(x-\mu_A)^2}{2\sigma_A^2} + (1-w)\frac{(y-\mu_B)^2}{2\sigma_B^2}.$$

It follows that the above optimization problem is a convex optimization problem for a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$, $w \in [0, 1]$, $\Delta_L \in [\Delta - \epsilon, \Delta]$. Hence for a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$, $w \in [0, 1]$ and $\Delta_L \in [\Delta - \epsilon, \Delta]$, $x^*(\mu_A, \mu_B, w, \Delta_L)$ uniquely satisfies,

$$w\frac{(x-\mu_A)}{\sigma_A^2}|_{x=x^*(\mu_A,\mu_B,w,\Delta_L)} + (1-w)\,\frac{(x-\Delta_L-\mu_B)}{\sigma_A^2}|_{x=x^*(\mu_A,\mu_B,w,\Delta_L)} = 0.$$

Also $y^*(\mu_A, \mu_B, w, \Delta_L) = x^*(\mu_A, \mu_B, w, \Delta_L) - \Delta_L$. Hence using some algebra, it follows that,

$$x^*(\mu_A, \mu_B, w, \Delta_L) = \frac{w\mu_A\sigma_B^2 + (1-w)\sigma_A^2\mu_B + \Delta_L\sigma_A^2(1-w)}{\sigma_{avg}^2(w)},$$

$$y^*(\mu_A, \mu_B, w, \Delta_L) = \frac{w\mu_A\sigma_B^2 + (1-w)\sigma_A^2\mu_B - \Delta_L\sigma_B^2 w}{\sigma_{avg}^2(w)},$$

where $\sigma^2_{avg}(w) = w\sigma_B^2 + (1-w)\sigma_A^2$.

It follows that,

$$\min_{x,y\in\mathcal{C}(\Delta_L)} w\frac{(x-\mu_A)^2}{2\sigma_A^2} + (1-w)\frac{(y-\mu_B)^2}{2\sigma_B^2} = \frac{w(1-w)}{2\sigma^2_{avg}(w)}(\mu_B-\mu_A+\Delta_L)^2. \tag{40}$$

One can similarly get,

$$\min_{x,y\in\mathcal{C}(\Delta_R)} w\frac{(x-\mu_A)^2}{2\sigma_A^2} + (1-w)\frac{(y-\mu_B)^2}{2\sigma_B^2} = \frac{w(1-w)}{2\sigma^2_{avg}(w)}(\mu_B-\mu_A+\Delta_R)^2. \tag{41}$$

Now using (40) and (41), we re-write $\mathfrak{L}$,

$$\max_{\substack{w\in[0,1],\\(\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \frac{w(1-w)}{2\sigma^2_{avg}(w)} \min\{(\mu_B-\mu_A+\Delta_R)^2, (\mu_B-\mu_A+\Delta_L)^2\}.$$

Using the fact that, $\Delta_R = \Delta_L + \epsilon$, we get,

$$\max_{\substack{w\in[0,1],\\ \Delta_L\in[\Delta-\epsilon,\Delta]}} \frac{w(1-w)}{2\sigma^2_{avg}(w)} \min\{(\mu_B-\mu_A+\Delta_L+\epsilon)^2, (\mu_B-\mu_A+\Delta_L)^2\}.$$

The above can be re-written as,

$$\max_{\Delta_L\in[\Delta-\epsilon,\Delta]} \min\{(\mu_B-\mu_A+\Delta_L+\epsilon)^2, (\mu_B-\mu_A+\Delta_L)^2\} \max_{w\in[0,1]} \frac{w(1-w)}{2\sigma^2_{avg}(w)}.$$

It follows that, $\Delta_L^*(\mu_A,\mu_B,\epsilon) = \Delta - \frac{\epsilon}{2}$ and $w^*(\mu_A,\mu_B,\epsilon)$ solves the following problem.

$$w^*(\mu_A,\mu_B,\epsilon) = \arg\max_{w\in[0,1]} \frac{w(1-w)}{2\sigma^2_{avg}(w)}.$$

Using the definition of $\sigma_{avg}(w)$, we get,

$$\frac{w(1-w)}{2\sigma^2_{avg}(w)} = \frac{1}{\frac{\sigma_A^2}{w} + \frac{\sigma_B^2}{(1-w)}}.$$

Hence,

$$w^*(\mu_A,\mu_B,\epsilon) = \arg\min_{w\in[0,1]} \frac{\sigma_A^2}{w} + \frac{\sigma_B^2}{(1-w)}.$$

Hence it, follows that, $w^*(\mu_A,\mu_B,\epsilon) = \frac{\sigma_A}{\sigma_A+\sigma_B}$. This completes the proof.

□

Proof of Theorem 5.3:

Using Lemma E.2 and Weierstrass existence theorem, we get that a solution of $\mathfrak{L}$ exists and the following holds,

$$\ell^*(\mu_A,\mu_B,\epsilon) = \max_{\substack{w\in[0,1],\\(\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \min\{T(\mu_A,\mu_B,w,\Delta_L), T(\mu_A,\mu_B,w,\Delta_R)\}. \tag{42}$$

Recall we denote any solution of $\mathfrak{L}$ as $w^*(\mu_A,\mu_B,\epsilon)$, $\Delta_L^*(\mu_A,\mu_B,\epsilon)$ and $\Delta_R^*(\mu_A,\mu_B,\epsilon)$. Now we claim that $w^*(\mu_A,\mu_B,\epsilon) \in (0,1)$ and $\Delta_L^*(\mu_A,\mu_B,\epsilon) \in (\Delta-\epsilon,\Delta)$. Notice that for any $w\in(0,1)$, $\Delta_L\in(\Delta-\epsilon,\Delta)$ and $\Delta_R = \Delta_L+\epsilon$, we have

$$T(\mu_A,\mu_B,w,\Delta_L) > 0 \text{ and } T(\mu_A,\mu_B,w,\Delta_R) > 0,$$

as $(\mu_A,\mu_B) \notin \mathcal{C}(\Delta_L) \cup \mathcal{C}(\Delta_R)$. Further, if $\Delta_L \in \{\Delta-\epsilon,\Delta\}$ and $\Delta_R = \Delta_L+\epsilon$, then we have $\min\{T(\mu_A,\mu_B,w,\Delta_L),T(\mu_A,\mu_B,w,\Delta_R)\} = 0$ for any $w \in [0,1]$ as $(\mu_A,\mu_B) \in \mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_R)$. Now for $w=0$ and $\Delta_L \in (\Delta-\epsilon,\Delta)$, there exists a $x \in \mathcal{I}$ and $y=\mu_B$ such that $(x,y) \in \mathcal{C}(\Delta_L)\cup\mathcal{C}(\Delta_L+\epsilon)$, hence $\min\{T(\mu_A,\mu_B,w,\Delta_L),T(\mu_A,\mu_B,w,\Delta_L+\epsilon)\} = 0$. Similarly for $w=1$ and $\Delta_L \in (\Delta-\epsilon,\Delta)$, we get that $\min\{T(\mu_A,\mu_B,w,\Delta_L),T(\mu_A,\mu_B,w,\Delta_L+\epsilon)\} = 0$. Hence we get that $w^*(\mu_A,\mu_B,\epsilon) \in (0,1)$, $\Delta_L^*(\mu_A,\mu_B,\epsilon) \in (\Delta-\epsilon,\Delta)$ and $\Delta_R^* = \Delta_L^*+\epsilon$.

For ease of readability, we suppress the notation we denote $w^*(\mu_A,\mu_B,\epsilon)$, $\Delta_L^*(\mu_A,\mu_B,\epsilon)$ and $\Delta_R^*(\mu_A,\mu_B,\epsilon)$ as w^*, Δ_L^* and Δ_R^* respectively. Using Lemma E.5, we get for small ϵ,

$$T(\mu_A,\mu_B,w^*,\Delta_L^*) = T(\mu_A,\mu_B,w^*,\Delta_R^*).$$

This completes the proof of the first two first-order conditions. Now we provide the proof of the last first-order condition. Re-writing the lower optimization problem as,

$$\begin{aligned}\ell^*(\mu_A,\mu_B,\epsilon) = \max_{\Delta_L\in(\Delta-\epsilon,\Delta)} \max_{w\in(0,1)} \max_{t\geq 0}\ & t,\\ & t \leq T(\mu_A,\mu_B,w,\Delta_L),\\ & t \leq T(\mu_A,\mu_B,w,\Delta_L+\epsilon).\end{aligned}$$

Writing the Lagrangian for above,

$$\max_{w\in(0,1)} \max_{\Delta_L\in(\Delta-\epsilon,\Delta)} \max_{t\geq 0} L(w,\Delta_L,t,\lambda_1,\lambda_2) = t + \lambda_1(T(\mu_A,\mu_B,w,\Delta_L)-t) + \lambda_2(T(\mu_A,\mu_B,w,\Delta_L+\epsilon)-t).$$

Using Lemma E.4, we get that $T(\mu_A,\mu_B,w,\Delta_L)$ and $T(\mu_A,\mu_B,w,\Delta_L+\epsilon)$ are continuously differentiable in (w,Δ_L) for small ϵ. Hence using the KKT conditions for small ϵ, we get that w^*, Δ_L^* and t^* satisfies,

$$\begin{aligned}\lambda_1+\lambda_2 &= 1.\\ \lambda_1 \geq 0,\ \lambda_2 &\geq 0.\end{aligned}$$

$$\begin{aligned}\lambda_1(T(\mu_A,\mu_B,w^*,\Delta_L^*)-t^*) = 0,\ T(\mu_A,\mu_B,w^*,\Delta_L^*) \geq t^*.\\ \lambda_2(T(\mu_A,\mu_B,w^*,\Delta_L^*+\epsilon)-t^*) = 0,\ T(\mu_A,\mu_B,w^*,\Delta_L^*+\epsilon) \geq t^*.\end{aligned}$$

$$\lambda_1 \frac{\partial T(\mu_A,\mu_B,w,\Delta_L)}{\partial \Delta_L}\Big|_{(w=w^*,\Delta_L=\Delta_L^*)} + \lambda_2 \frac{\partial T(\mu_A,\mu_B,w,\Delta_L+\epsilon)}{\partial \Delta_L}\Big|_{(w=w^*,\Delta_L=\Delta_L^*)} = 0. \tag{43}$$

$$\lambda_1 \frac{\partial T(\mu_A,\mu_B,w,\Delta_L)}{\partial w}\Big|_{(w=w^*,\Delta_L=\Delta_L^*)} + \lambda_2 \frac{\partial T(\mu_A,\mu_B,w,\Delta_L+\epsilon)}{\partial w}\Big|_{(w=w^*,\Delta_L=\Delta_L^*)} = 0.$$

Recall that,

$$T(\mu_A,\mu_B,w^*,\Delta_L^*) = T(\mu_A,\mu_B,w^*,\Delta_L^*+\epsilon).$$

It follows from above equations that $t^* = T(\mu_A,\mu_B,w^*,\Delta_L^*) = T(\mu_A,\mu_B,w^*,\Delta_L^*+\epsilon)$.

Let x_L^* represent $x^*(\mu_A,\mu_B,w^*,\Delta_L^*)$ and x_R^* represent $x^*(\mu_A,\mu_B,w^*,\Delta_R^*)$. Using Envelope theorem, the proof of Lemma E.4 and definition of x_L^* and x_R^*, we can re-write (43) and (B) as,

$$\lambda_1 \frac{\partial d(\mu_B,x_L^*-\Delta_L)}{\partial \Delta_L}\Big|_{\Delta_L=\Delta_L^*} + \lambda_2 \frac{\partial d(\mu_B,x_R^*-\Delta_L-\epsilon)}{\partial \Delta_L}\Big|_{\Delta_L=\Delta_L^*} = 0.$$

$$\lambda_1(d(\mu_A,x_L^*) - d(\mu_B,x_L^*-\Delta_L^*)) + \lambda_2(d(\mu_A,x_R^*) - d(\mu_B,x_R^*-\Delta_L^*-\epsilon)) = 0.$$

Using some algebra, we get that w^* and Δ_L^* satisfies

$$\begin{aligned}(d(\mu_A, x_L^*) - d(\mu_B, x_L^* - \Delta_L^*)) \left(\frac{\partial d(\mu_B, x_R^* - \Delta_L - \epsilon)}{\partial \Delta_L}|_{\Delta_L = \Delta_L^*}\right) = \\ (d(\mu_A, x_R^*) - d(\mu_B, x_R^* - \Delta_L^* - \epsilon)) \left(\frac{\partial d(\mu_B, x_L^* - \Delta_L)}{\partial \Delta_L}|_{\Delta_L = \Delta_L^*}\right).\end{aligned} \tag{44}$$

We now re-write the first-order conditions for $\mathfrak{L}$ and index everything by ϵ,

$$\Delta_{R,\epsilon}^* = \Delta_{L,\epsilon}^* + \epsilon$$

$$T(\mu_A, \mu_B, w_\epsilon^*, \Delta_{L,\epsilon}^*) = T(\mu_A, \mu_B, w_\epsilon^*, \Delta_{R,\epsilon}^*).$$

$$\begin{aligned}(d(\mu_A, x_{L,\epsilon}^*) - d(\mu_B, x_{L,\epsilon}^* - \Delta_{L,\epsilon}^*)) \left(\frac{\partial d(\mu_B, x_{R,\epsilon}^* - \Delta_L - \epsilon)}{\partial \Delta_L}|_{\Delta_L = \Delta_{L,\epsilon}^*}\right) = \\ (d(\mu_A, x_{R,\epsilon}^*) - d(\mu_B, x_{R,\epsilon}^* - \Delta_{R,\epsilon}^*)) \left(\frac{\partial d(\mu_B, x_{L,\epsilon}^* - \Delta_L)}{\partial \Delta_L}|_{\Delta_L = \Delta_{L,\epsilon}^*}\right).\end{aligned}$$

Using the definition of function T, we get,

$$w_\epsilon^* d(\mu_A, x_{L,\epsilon}^*) + (1 - w_\epsilon^*) d(\mu_B, x_{L,\epsilon}^* - \Delta_{L,\epsilon}^*) = w_\epsilon^* d(\mu_A, x_{R,\epsilon}^*) + (1 - w_\epsilon^*) d(\mu_B, x_{R,\epsilon}^* - \Delta_{R,\epsilon}^*). \tag{45}$$

We are interested in the limiting behaviour of $w^*(\mu_A, \mu_B, \epsilon)$ and $\ell^*(\mu_A, \mu_B, \epsilon)$ in the limiting regime of $\epsilon \to 0$. We will be using the twice continuous differentiability of $d(\mu, x)$ for x in a small neighbourhood around μ (see Appendix D.1). Let $\frac{\partial^2 d(\mu,x)}{\partial x^2}|_{x=c} = H(\mu, c)$. It follows that $H(\mu, \mu) = I(\mu)$.

Using Taylor series expansion of (45),

$$\begin{aligned}w_\epsilon^* \frac{(x_{L,\epsilon}^* - \mu_A)^2}{2} H(\mu_A, c_{1,\epsilon}) + (1 - w_\epsilon^*) \frac{(x_{L,\epsilon}^* - \Delta_{L,\epsilon}^* - \mu_B)^2}{2} H(\mu_B, c_{2,\epsilon}) = \\ w_\epsilon^* \frac{(x_{R,\epsilon}^* - \mu_A)^2}{2} H(\mu_A, c_{3,\epsilon}) + (1 - w_\epsilon^*) \frac{(x_{R,\epsilon}^* - \Delta_{R,\epsilon}^* - \mu_B)^2}{2} H(\mu_B, c_{4,\epsilon}).\end{aligned}$$

Here $c_{1,\epsilon} \in (x_{L,\epsilon}^*, \mu_A)$, $c_{2,\epsilon} \in (\mu_B, x_{L,\epsilon}^* - \Delta_{L,\epsilon}^*)$, $c_{3,\epsilon} \in (\mu_A, x_{R,\epsilon}^*)$ and $c_{4,\epsilon} \in (x_{R,\epsilon}^* - \Delta_{R,\epsilon}^*, \mu_B)$. Let

$$K_{1,\epsilon} \triangleq (\mu_A - x_{L,\epsilon}^*) K_{2,\epsilon} \triangleq (x_{R,\epsilon}^* - \mu_A), \text{ and } \alpha_\epsilon \triangleq \Delta - \Delta_{L,\epsilon}^*.$$

Using some algebra,

$$\begin{aligned}w_\epsilon^* H(\mu_A, c_{1,\epsilon})(K_{1,\epsilon}^2 - K_{2,\epsilon}^2) + (1 - w_\epsilon^*) H(\mu_B, c_{2,\epsilon})((-K_{1,\epsilon} + \alpha_\epsilon)^2 - (K_{2,\epsilon} + \alpha_\epsilon - \epsilon)^2) = \\ w_\epsilon^* K_{2,\epsilon}^2 (H(\mu_A, c_{3,\epsilon}) - H(\mu_A, c_{1,\epsilon})) + (1 - w_\epsilon^*)(K_{2,\epsilon} + \alpha_\epsilon - \epsilon)^2 (H(\mu_B, c_{4,\epsilon}) - H(\mu_B, c_{2,\epsilon})).\end{aligned} \tag{46}$$

Notice that we are not assuming in the proof that w_ϵ^* and $\Delta_{L,\epsilon}^*$ is unique for a given ϵ. Hence to prove this result, we select any arbitrary sequence of w_ϵ^* and $\Delta_{L,\epsilon}^*$, i.e., for each $\epsilon > 0$, we can choose any of the solutions of $\mathfrak{L}$ if it is not unique. We know that, $\lim_{\epsilon \to 0} \alpha_\epsilon = 0$, $\lim_{\epsilon \to 0} K_{1,\epsilon} = 0$ and $\lim_{\epsilon \to 0} K_{2,\epsilon} = 0$. It follows that using continuity of $H(\mu, x)$ in x, we get,

$$\lim_{\epsilon \to 0} H(\mu_A, c_{5,\epsilon}) = \lim_{\epsilon \to 0} H(\mu_A, c_{7,\epsilon}) = I(\mu_A).$$

$$\lim_{\epsilon \to 0} H(\mu_B, c_{6,\epsilon}) = \lim_{\epsilon \to 0} H(\mu_B, c_{8,\epsilon}) = I(\mu_B).$$

Since w_ϵ^* is a bounded sequence, it follows that the limit of RHS of (46) equals 0. This further implies,

$$\lim_{\epsilon \to 0} w_\epsilon^* H(\mu_A, c_{1,\epsilon})(K_{1,\epsilon}^2 - K_{2,\epsilon}^2) + (1 - w_\epsilon^*) H(\mu_B, c_{2,\epsilon})((-K_{1,\epsilon} + \alpha_\epsilon)^2 - (K_{2,\epsilon} + \alpha_\epsilon - \epsilon)^2) = 0. \tag{47}$$

It follows that $x^*_{L,\epsilon}$ and $x^*_{R,\epsilon}$ uniquely satisfy for small ϵ, we get,

$$w^*_\epsilon \frac{\partial d(\mu_A, x)}{\partial x}|_{x=x^*_{L,\epsilon}} + (1-w^*_\epsilon)\frac{\partial d(\mu_B, x-\Delta^*_{L,\epsilon})}{\partial x}|_{x=x^*_{L,\epsilon}} = 0. \tag{48}$$

$$w^*_\epsilon \frac{\partial d(\mu_A, x)}{\partial x}|_{x=x^*_{R,\epsilon}} + (1-w^*_\epsilon)\frac{\partial d(\mu_B, x-\Delta^*_{R,\epsilon})}{\partial x}|_{x=x^*_{R,\epsilon}} = 0. \tag{49}$$

Using the Taylor series expansion of the above equation, we get,

$$w^*_\epsilon H(\mu_A, c_{5,\epsilon})\frac{(x^*_{L,\epsilon}-\mu_A)}{\epsilon} + (1-w^*_\epsilon)\, H(\mu_B, c_{6,\epsilon})\frac{(x^*_{L,\epsilon}-\mu_B-\Delta^*_{L,\epsilon})}{\epsilon} = 0. \tag{50}$$

$$w^*_\epsilon H(\mu_A, c_{7,\epsilon})\frac{(x^*_{R,\epsilon}-\mu_A)}{\epsilon} + (1-w^*_\epsilon)\, H(\mu_B, c_{8,\epsilon})\frac{(x^*_{R,\epsilon}-\mu_B-\Delta^*_{R,\epsilon})}{\epsilon} = 0. \tag{51}$$

Here $c_{5,\epsilon} \in (x^*_{L,\epsilon}, \mu_A)$, $c_{6,\epsilon} \in (\mu_B, x^*_{L,\epsilon} - \Delta^*_{L,\epsilon})$, $c_{7,\epsilon} \in (\mu_A, x^*_{R,\epsilon})$ and $c_{8,\epsilon} \in (x^*_{R,\epsilon} - \Delta^*_{R,\epsilon}, \mu_B)$.

The above two equations can be written as,

$$K_{1,\epsilon} = \frac{\alpha_\epsilon(1-w^*_\epsilon)H(\mu_B, c_{6,\epsilon})}{(1-w^*_\epsilon)H(\mu_B, c_{6,\epsilon}) + w^*_\epsilon H(\mu_A, c_{5,\epsilon})}.$$

$$K_{2,\epsilon} = \frac{(-\alpha_\epsilon+\epsilon)(1-w^*_\epsilon)H(\mu_B, c_{8,\epsilon})}{(1-w^*_\epsilon)H(\mu_B, c_{8,\epsilon}) + w^*_\epsilon H(\mu_A, c_{7,\epsilon})}.$$

Notice that $\alpha_\epsilon \in [0, \epsilon]$ for all $\epsilon > 0$, hence it follows that $\frac{\alpha_\epsilon}{\epsilon}$ is a bounded sequence in $[0, 1]$. Hence, we re-write the above two equations as,

$$\frac{K_{1,\epsilon}}{\epsilon} = \frac{\alpha_\epsilon}{\epsilon}\left(\frac{(1-w^*_\epsilon)H(\mu_B, c_{6,\epsilon})}{(1-w^*_\epsilon)H(\mu_B, c_{6,\epsilon}) + w^*_\epsilon H(\mu_A, c_{5,\epsilon})}\right). \tag{52}$$

$$\frac{K_{2,\epsilon}}{\epsilon} = \frac{(-\alpha_\epsilon+\epsilon)}{\epsilon}\left(\frac{(1-w^*_\epsilon)H(\mu_B, c_{8,\epsilon})}{(1-w^*_\epsilon)H(\mu_B, c_{8,\epsilon}) + w^*_\epsilon H(\mu_A, c_{7,\epsilon})}\right). \tag{53}$$

Since we know that w^*_ϵ and $\frac{\alpha_\epsilon}{\epsilon}$ are bounded sequences in $[0, 1]$, hence there exists a converging subsequence of w^*_ϵ, and $\frac{\alpha_\epsilon}{\epsilon}$ i.e., $\lim_{k\to\infty} w^*_{\epsilon_k} = \mathbb{W}$, and $\lim_{k\to\infty} \frac{\alpha_{\epsilon_k}}{\epsilon_k} = \eta$ such that $\lim_{k\to\infty} \epsilon_k = 0$, where $\mathbb{W} \in [0, 1]$ and $\eta \in [0, 1]$. It follows that $\mathbb{W}$ and η are functions of μ_A and μ_B.

Hence writing the (52) and (53) for the subsequence defined above, we get,

$$\gamma_1 \triangleq \lim_{k\to\infty} \frac{K_{1,\epsilon_k}}{\epsilon_k} = \eta\left(\frac{(1-\mathbb{W})I(\mu_B)}{(1-\mathbb{W})I(\mu_B) + \mathbb{W}I(\mu_A)}\right). \tag{54}$$

$$\gamma_2 \triangleq \lim_{k\to\infty} \frac{K_{2,\epsilon_k}}{\epsilon_k} = (1-\eta)\left(\frac{(1-\mathbb{W})I(\mu_B)}{(1-\mathbb{W})I(\mu_B) + \mathbb{W}I(\mu_A)}\right). \tag{55}$$

In the above, γ_1 and γ_2 are function of μ_A and μ_B. Using above two limiting results in (47) for the subsequence, we get,

$$\mathbb{W}I(\mu_A)(\gamma_1+\gamma_2)(\gamma_1-\gamma_2) + (1-\mathbb{W})I(\mu_B)(\gamma_1-\gamma_2-2\eta+1)(\gamma_1+\gamma_2-1) = 0. \tag{56}$$

Using the definition of γ_1 and γ_2, we get,

$$\gamma_1 + \gamma_2 = (1-\mathcal{W}).$$

$$\gamma_1 - \gamma_2 = (2\eta-1)(1-\mathcal{W}),$$

where, $\mathcal{W} \triangleq \left(\frac{\mathbb{W}I(\mu_A)}{(1-\mathbb{W})I(\mu_B)+\mathbb{W}I(\mu_A)}\right)$.

Substituting the values of γ_1 and γ_2 in (56), we get,

$$\mathcal{W}(1-\mathcal{W})^2(2\eta-1)+(1-\mathcal{W})(2\eta-1)(\mathcal{W})^2=0.$$

The above can be simplified as,

$$\mathcal{W}(1-\mathcal{W})(2\eta-1)=0.$$

There are three possibilities for the above equation to be satisfied. $\mathcal{W}=0$ or $\mathcal{W}=1$ or $\eta=1/2$.

Case 1: First we take $\eta=\frac{1}{2}$. We later show that the other two cases are not possible. Now we move to the second FOC of the optimization problem $\mathfrak{L}$.

$$(d(\mu_A, x^*_{L,\epsilon})-d(\mu_B, x^*_{L,\epsilon}-\Delta^*_{L,\epsilon}))\left(\frac{\partial d(\mu_B, x^*_{R,\epsilon}-\Delta_L-\epsilon)}{\partial \Delta_L}|_{\Delta_L=\Delta^*_{L,\epsilon}}\right)=$$
$$(d(\mu_A, x^*_{R,\epsilon})-d(\mu_B, x^*_{R,\epsilon}-\Delta^*_{R,\epsilon}))\left(\frac{\partial d(\mu_B, x^*_{L,\epsilon}-\Delta_L)}{\partial \Delta_L}|_{\Delta_L=\Delta^*_{L,\epsilon}}\right).$$

Using (48) and (49), we get,

$$(d(\mu_A, x^*_{L,\epsilon})-d(\mu_B, x^*_{L,\epsilon}-\Delta^*_{L,\epsilon}))\left(\frac{\partial d(\mu_A, x)}{\partial x}|_{x=x^*_{R,\epsilon}}\right)=$$
$$(d(\mu_A, x^*_{R,\epsilon})-d(\mu_B, x^*_{R,\epsilon}-\Delta^*_{R,\epsilon}))\left(\frac{\partial d(\mu_A, x)}{\partial x}|_{x=x^*_{L,\epsilon}}\right).$$

Using the Taylor series expansion of the above, we get,

$$H(\mu_A, c_{7,\epsilon})\left(\frac{K_{2,\epsilon}}{\epsilon}\right)\left[\left(\frac{K_{1,\epsilon}}{\epsilon}\right)^2 H(\mu_A, c_{1,\epsilon})-\left(\frac{K_{1,\epsilon}-\alpha_\epsilon}{\epsilon}\right)^2 H(\mu_B, c_{2,\epsilon})\right]=$$
$$H(\mu_A, c_{5,\epsilon})\left(\frac{-K_{1,\epsilon}}{\epsilon}\right)\left[\left(\frac{K_{2,\epsilon}}{\epsilon}\right)^2 H(\mu_A, c_{3,\epsilon})-\left(\frac{K_{2,\epsilon}+\alpha_\epsilon-\epsilon}{\epsilon}\right)^2 H(\mu_B, c_{4,\epsilon})\right]. \tag{57}$$

Coming back the converging subsequence $\lim_{k\to\infty} w^*_{\epsilon_k}=\mathbb{W}$, and $\lim_{k\to\infty}\frac{\alpha_{\epsilon_k}}{\epsilon_k}=\frac{1}{2}$ such that $\lim_{k\to\infty}\epsilon_k=0$, where $\mathbb{W}\in(0,1)$. It follows that,

$$\gamma_1=\gamma_2=\frac{1}{2}\left(\frac{(1-\mathbb{W})I(\mu_B)}{(1-\mathbb{W})I(\mu_B)+\mathbb{W}I(\mu_A)}\right). \tag{58}$$

Writing (57) for above mentioned subsequence, we get,

$$\gamma_2(\gamma_1^2 I(\mu_A)-(\gamma_1-1/2)^2 I(\mu_B))=-\gamma_1(\gamma_2^2-I(\mu_A)-(\gamma_2-1/2)^2 I(\mu_B)).$$

Since $\gamma_1=\gamma_2$ and using (58), we know that, $\gamma_1\in(0,1/2)$, hence it follows that,

$$\gamma_1=\gamma_2=\frac{\sqrt{I(\mu_B)}}{2(\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)})}.$$

Substituting the value of γ_1 from (58), we get,

$$\mathbb{W}=\frac{\sqrt{I(\mu_B)}}{\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)}}.$$

To summarise, we started with the solutions w^*_ϵ and $\Delta^*_{L,\epsilon}$ for each ϵ, which satisfy the FOCs of $\mathfrak{L}$. Since w^*_ϵ and $\Delta^*_{L,\epsilon}$ may not be unique, hence we selected any one of the solutions of $\mathfrak{L}$ for a given ϵ. Then we showed that there exists a converging subsequence $w^*_{\epsilon_k}$ and Δ^*_{L,ϵ_k} such that $\lim_{k\to\infty}\epsilon_k \to 0$, $\lim_{k\to\infty} w^*_{\epsilon_k} = \frac{\sqrt{I(\mu_B)}}{\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)}}$ and $\lim_{k\to\infty}\frac{\Delta-\Delta^*_{L,\epsilon_k}}{\epsilon_k} = \frac{1}{2}$. Since we can show that each of the subsequences of w^*_ϵ will have a converging subsequence going to the same limit. Hence it follows that,

$$\lim_{\epsilon\to 0} w^*(\mu_A,\mu_B,\epsilon) = \frac{\sqrt{I(\mu_B)}}{\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)}},$$

and

$$\lim_{\epsilon\to 0}\frac{\Delta-\Delta^*_L(\mu_A,\mu_B,\epsilon)}{\epsilon} = \frac{1}{2}.$$

Now we show that,

$$\lim_{\epsilon\to 0}\frac{\ell^*(\mu_A,\mu_B,\epsilon)}{\epsilon^2} = \frac{(I(\mu_A)I(\mu_B))}{8(\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)})^2}.$$

Recall from (42), we get,

$$\ell^*(\mu_A,\mu_B,\epsilon) = \max_{\substack{w\in[0,1],\\ (\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \min\{T(\mu_A,\mu_B,w,\Delta_L), T(\mu_A,\mu_B,w,\Delta_L+\epsilon)\}.$$

Using the FOCs, we know that for any $\epsilon > 0$,

$$\ell^*(\mu_A,\mu_B,\epsilon) = T(\mu_A,\mu_B,w^*,\Delta^*_L).$$

Now using the definition of $T(\mu_A,\mu_B,w,\Delta_L)$, we get,

$$\ell^*(\mu_A,\mu_B,\epsilon) = w^* d(\mu_A,x^*_L) + (1-w^*)d(\mu_B,x^*_L-\Delta^*_L).$$

Indexing w^*, x^*_L and Δ^*_L with ϵ, we get,

$$\ell^*(\mu_A,\mu_B,\epsilon) = w^*_\epsilon d(\mu_A,x^*_{L,\epsilon}) + (1-w^*_\epsilon)d(\mu_B,x^*_{L,\epsilon}-\Delta^*_{L,\epsilon}).$$

Using Taylor series expansion,

$$\ell^*(\mu_A,\mu_B,\epsilon) = w^*_\epsilon\frac{(x^*_{L,\epsilon}-\mu_A)^2}{2}H(\mu_A,c_{1,\epsilon}) + (1-w^*_\epsilon)\frac{(x^*_{L,\epsilon}-\Delta^*_{L,\epsilon}-\mu_B)^2}{2}H(\mu_B,c_{2,\epsilon}). \tag{59}$$

Above can be re-written as,

$$\frac{\ell^*(\mu_A,\mu_B,\epsilon)}{\epsilon^2} = w^*_\epsilon\frac{(x^*_{L,\epsilon}-\mu_A)^2}{2\epsilon^2}H(\mu_A,c_{1,\epsilon}) + (1-w^*_\epsilon)\frac{(x^*_{L,\epsilon}-\Delta^*_{L,\epsilon}-\mu_B)^2}{2\epsilon^2}H(\mu_B,c_{2,\epsilon}).$$

Taking the limit of $\epsilon\to 0$, we get,

$$\lim_{\epsilon\to 0}\frac{\ell^*(\mu_A,\mu_B,\epsilon)}{\epsilon^2} = \frac{\gamma_1^2}{2}I(\mu_A)\lim_{\epsilon\to 0} w^*_\epsilon + \frac{(1/2-\gamma_1)^2}{2}I^*(\mu_B)\lim_{\epsilon\to 0}(1-w^*_\epsilon).$$

Substituting the value of limit of w^*_ϵ and γ_1, we get,

$$\lim_{\epsilon\to 0}\frac{\ell^*(\mu_A,\mu_B,\epsilon)}{\epsilon^2} = \frac{\gamma_1^2}{2}I(\mu_A).$$

$$\lim_{\epsilon\to 0}\frac{\ell^*(\mu_A,\mu_B,\epsilon)}{\epsilon^2} = \frac{I(\mu_A)I(\mu_B)}{8(\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)})^2}.$$

To complete the proof, we need to show that the remaining two cases are not possible.

Case 2: Suppose $\mathcal{W} = 0$, then for the subsequence $w^*_{\epsilon_k}$, we have $\lim_{k\to\infty} w^*_{\epsilon_k} = \mathbb{W} = 0$. Using (54) and (55), we get, $\gamma_1 = \eta$ and $\gamma_2 = 1 - \eta$. Taking $k \to \infty$ in (57), we get $\eta = 0$ or 1. In both cases, using the definition of $\ell^*(\mu_A, \mu_B, \epsilon_k)$ (see (59)), it follows that, $\lim_{k\to\infty} \frac{\ell^*(\mu_A,\mu_B,\epsilon_k)}{\epsilon_k^2} = 0$. We now show the contradiction by choosing a feasible point for the optimization problem $\mathfrak{L}$, $w_{\epsilon_k} = \frac{\sqrt{I(\mu_B)}}{\sqrt{I(\mu_A)}+\sqrt{I(\mu_B)}}$ and $\Delta_{L,\epsilon_k} = \Delta - \epsilon_k/2$, we get that,

$$\lim_{k\to\infty} \frac{\min\{T(\mu_A, \mu_B, w_{\epsilon_k}, \Delta_{L,\epsilon_k}), T(\mu_A, \mu_B, w_{\epsilon_k}, \Delta_{L,\epsilon_k} + \epsilon_k)\}}{\epsilon_k^2} = \frac{I(\mu_A)\cdot I(\mu_B)}{8(\sqrt{I(\mu_A)} + \sqrt{I(\mu_B)})^2}.$$

This leads to contradiction as $\lim_{k\to\infty} \frac{\ell^*(\mu_A,\mu_B,\epsilon_k)}{\epsilon_k^2} < \lim_{k\to\infty} \frac{\min\{T(\mu_A,\mu_B,w_{\epsilon_k},\Delta_{L,\epsilon_k}),T(\mu_A,\mu_B,w_{\epsilon_k},\Delta_{L,\epsilon_k}+\epsilon_k)\}}{\epsilon_k^2}$ and $\ell^*(\mu_A, \mu_B, \epsilon_k)$ is the optimal value of the optimization problem $\mathfrak{L}$. We get a similar proof for the last case when $\mathcal{W} = 1$. This completes the proof. $\square$

Proof of Theorem 5.4: First we show that,

$$\lim_{\epsilon\to 0} \frac{w^*(\mu_A, \mu_B, \epsilon) - \overline{w}(\mu_A, \mu_B)}{\epsilon} = 0,$$

where $\overline{w}(\mu_A, \mu_B) = \frac{\sqrt{I(\mu_B)}}{\sqrt{I(\mu_B)}+\sqrt{I(\mu_A)}}$. To prove this, we utilize the first-order conditions of $\mathfrak{L}$. We index everything by ϵ in (44) and use the notations defined in the proof of Theorem 5.3. Using the third order Taylor series expansion around $\epsilon = 0$ of (44) and dividing both sides by ϵ^4,we get,

$$H(\mu_A, c_{7,\epsilon}) \left(\frac{K_{2,\epsilon}}{\epsilon}\right) B_{1,\epsilon} = H(\mu_A, c_{5,\epsilon}) \left(\frac{-K_{1,\epsilon}}{\epsilon}\right) B_{2,\epsilon}, \tag{60}$$

where,

$$B_{1,\epsilon} = \left(\frac{K_{1,\epsilon}^2}{\epsilon^3}\right) I(\mu_A) - \left(\frac{K_{1,\epsilon}^3}{3\epsilon^3}\right) O(\mu_A, c_{9,\epsilon}) - \frac{(\alpha_\epsilon - K_{1,\epsilon})^2}{\epsilon^3} I(\mu_B) - \frac{(\alpha_\epsilon - K_{1,\epsilon})^3}{3\epsilon^3} O(\mu_B, c_{10,\epsilon}).$$

$$B_{2,\epsilon} = \left(\frac{K_{2,\epsilon}^2}{\epsilon^3}\right) I(\mu_A) + \left(\frac{K_{2,\epsilon}^3}{3\epsilon^3}\right) O(\mu_A, c_{11,\epsilon}) - \frac{(K_{2,\epsilon} + \alpha_\epsilon - \epsilon)^2}{\epsilon^3} I(\mu_B) - \frac{(K_{2,\epsilon} + \alpha_\epsilon - \epsilon)^3}{3\epsilon^3} O(\mu_B, c_{12,\epsilon}).$$

Here $O(\mu_A, a) \triangleq \frac{\partial^3 d(\mu_A,x)}{\partial x^3}|_{x=a}$. Further $c_{9,\epsilon} \in (x^*_{L,\epsilon}, \mu_A)$, $c_{10,\epsilon} \in (\mu_B, x^*_{L,\epsilon} - \Delta^*_{L,\epsilon})$, $c_{11,\epsilon} \in (\mu_A, x^*_{R,\epsilon})$ and $c_{12,\epsilon} \in (x^*_{R,\epsilon} - \Delta^*_{R,\epsilon}, \mu_B)$. Using the (52) and (53) in the proof of Theorem 5.3, we get,

$$\lim_{\epsilon\to 0} \frac{K_{1,\epsilon}}{\epsilon} = \overline{w}(\mu_A, \mu_B)/2,\ \lim_{\epsilon\to 0} \frac{K_{2,\epsilon}}{\epsilon} = \overline{w}(\mu_A, \mu_B)/2.$$

$$\lim_{\epsilon\to 0} \frac{\alpha_{1,\epsilon}}{\epsilon} = 1/2.$$

It follows from the definition that, $\lim_{\epsilon\to 0} c_{5,\epsilon} = \lim_{\epsilon\to 0} c_{7,\epsilon} = \lim_{\epsilon\to 0} c_{9,\epsilon} = \lim_{\epsilon\to 0} c_{11,\epsilon} = \mu_A$ and $\lim_{\epsilon\to 0} c_{10,\epsilon} = \lim_{\epsilon\to 0} c_{12,\epsilon} = \mu_B$. Taking the limit of $\epsilon \to 0$ on both side in (60), we get,

$$\lim_{\epsilon\to 0} (B_{1,\epsilon} + B_{2,\epsilon}) = 0.$$

Using some algebra above can be re-arranged as,

$$\lim_{\epsilon\to 0} \left[I(\mu_A) \left(\frac{K_{1,\epsilon}^2 + K_{2,\epsilon}^2}{\epsilon^3}\right) - I(\mu_B) \left(\frac{(\alpha_\epsilon - K_{1,\epsilon})^2 + (K_{2,\epsilon} + \alpha_\epsilon - \epsilon)^2}{\epsilon^3}\right)\right] = 0.$$

Recall the definition of $K_{1,\epsilon}$ and $K_{2,\epsilon}$,

$$K_{1,\epsilon} = \frac{\alpha_\epsilon(1-w_\epsilon^*)H(\mu_B, c_{6,\epsilon})}{(1-w_\epsilon^*)H(\mu_B, c_{6,\epsilon}) + w_\epsilon^* H(\mu_A, c_{5,\epsilon})}.$$

$$K_{2,\epsilon} = \frac{(-\alpha_\epsilon+\epsilon)(1-w_\epsilon^*)H(\mu_B, c_{8,\epsilon})}{(1-w_\epsilon^*)H(\mu_B, c_{8,\epsilon}) + w_\epsilon^* H(\mu_A, c_{7,\epsilon})}.$$

Using the above definitions, we get,

$$\lim_{\epsilon\to 0}\left[I(\mu_A)\left(\frac{I^2(\mu_B)(1-w_\epsilon^*)^2}{\epsilon}\right) - I(\mu_B)\left(\frac{I^2(\mu_A)(w_\epsilon^*)^2}{\epsilon}\right)\right] = 0.$$

The above can be simplified to,

$$\lim_{\epsilon\to 0}\frac{w_\epsilon^* - \overline{w}(\mu_A,\mu_B)}{\epsilon} = 0.$$

This completes the proof of the first part. Now we show that,

$$\lim_{\epsilon\to 0}\frac{w^*(\mu_A,\mu_B,\epsilon) - \overline{w}(\mu_A,\mu_B)}{\epsilon^2} = \frac{v(\mu_A,\mu_B)}{96\sqrt{I(\mu_A)I(\mu_B)}},$$

where, $v(\mu_A,\mu_B) = \left(\frac{\partial^4 d(\mu_B,x)}{\partial x^4}\Big|_{x=\mu_B}\right)(1-\overline{w}(\mu_A,\mu_B))^4 - \left(\frac{\partial^4 d(\mu_A,x)}{\partial x^4}\Big|_{x=\mu_A}\right)\overline{w}^4(\mu_A,\mu_B)$.

To prove this, we again utilize the first-order conditions of $\mathfrak{L}$. Using the fourth order Taylor series expansion around $\epsilon = 0$ of (44) and dividing both sides by ϵ^5, and similar to the proof of the first part of this theorem, we get,

$$\lim_{\epsilon\to 0}(C_{1,\epsilon} + C_{2,\epsilon} + C_{3,\epsilon}) = 0. \tag{61}$$

$$C_{1,\epsilon} = \left(\frac{K_{1,\epsilon}^2 + K_{2,\epsilon}^2}{\epsilon^4}\right)I(\mu_A) - \left(\frac{(\alpha_\epsilon - K_{1,\epsilon})^2}{\epsilon^4} + \frac{(K_{2,\epsilon}+\alpha_\epsilon-\epsilon)^2}{\epsilon^4}\right)I(\mu_B).$$

$$C_{2,\epsilon} = \left(\frac{K_{2,\epsilon}^3 - K_{1,\epsilon}^3}{3\epsilon^4}\right)O(\mu_A,\mu_A) - \left(\frac{(\alpha_\epsilon - K_{1,\epsilon})^3}{3\epsilon^4} + \frac{(K_{2,\epsilon}+\alpha_\epsilon-\epsilon)^3}{3\epsilon^4}\right)O(\mu_B,\mu_B).$$

$$C_{3,\epsilon} = \left(\frac{K_{1,\epsilon}^4}{12\epsilon^4}\right)E(\mu_A, c_{13,\epsilon}) + \left(\frac{K_{2,\epsilon}^4}{12\epsilon^4}\right)E(\mu_A, c_{15,\epsilon}) - \frac{(\alpha_\epsilon - K_{1,\epsilon})^4}{12\epsilon^4}E(\mu_B, c_{14,\epsilon}) - \frac{(K_{2,\epsilon}+\alpha_\epsilon-\epsilon)^4}{12\epsilon^4}E(\mu_B, c_{16,\epsilon}),$$

where, $E(\mu,c) \triangleq \frac{\partial^4 d(\mu,x)}{\partial x^4}\Big|_{x=c}$.

First, using some algebra it follows that,

$$\lim_{\epsilon\to 0} C_{3,\epsilon} = \frac{v(\mu_A,\mu_B)}{96}. \tag{62}$$

Now we consider $C_{2,\epsilon}$. Using the fact, that $\lim_{\epsilon\to 0}\frac{K_{1,\epsilon}}{\epsilon w_\epsilon^*} = \lim_{\epsilon\to 0}\frac{K_{2,\epsilon}}{\epsilon w_\epsilon^*} = 1/2$, and the first part of this theorem, we get,

$$\lim_{\epsilon\to 0} C_{2,\epsilon} = 0. \tag{63}$$

Now we consider,

$$C_{1,\epsilon} = \left(\frac{K_{1,\epsilon}^2 + K_{2,\epsilon}^2}{\epsilon^4}\right)I(\mu_A) - \left(\frac{(\alpha_\epsilon - K_{1,\epsilon})^2}{\epsilon^4} + \frac{(K_{2,\epsilon}+\alpha_\epsilon-\epsilon)^2}{\epsilon^4}\right)I(\mu_B).$$

Using the fact, that $\lim_{\epsilon\to 0} \frac{K_{1,\epsilon}}{\epsilon w^*_\epsilon} = \lim_{\epsilon\to 0} \frac{K_{1,\epsilon}}{\epsilon w^*_\epsilon} = 1/2$, we get,

$$\lim_{\epsilon\to 0} C_{1,\epsilon} = \left(\frac{(w^*_\epsilon)^2}{2\epsilon^2}\right) I(\mu_A) - \left(\frac{(1-w^*_\epsilon)^2}{2\epsilon^2}\right) I(\mu_B).$$

It can be re-written as,

$$\lim_{\epsilon\to 0} C_{1,\epsilon} = \left(\frac{(w^*_\epsilon)\sqrt{I(\mu_A)} - (1-w^*_\epsilon)\sqrt{I(\mu_B)}}{2\epsilon^2}\right)(\overline{w}(\mu_A,\mu_B)\sqrt{I(\mu_A)} + (1-\overline{w}(\mu_A,\mu_B))\sqrt{I(\mu_B)}).$$

Hence,

$$\lim_{\epsilon\to 0} C_{1,\epsilon} = \left(\frac{w^*_\epsilon - \overline{w}(\mu_A,\mu_B)}{\epsilon^2}\right)(\sqrt{I(\mu_A)I(\mu_B)}). \tag{64}$$

Combining (62), (63), (64) and substituting in (61), we get the desired result. This completes the proof.

□

C. Proofs of results in Section 6.

Proof of Theorem 6.1: Proof of the corollary follows similar to the proof of Theorem 4.4. Since $\frac{\sqrt{\frac{1}{I(\mu_A)}}}{\sqrt{\frac{1}{I(\mu_A)}}+\sqrt{\frac{1}{I(\mu_B)}}}$ is jointly continuous function of μ_A and μ_B since $I(\mu)$ is a continuous function of μ for $\mu \in \mathcal{I}$ and the fact that $I(\mu) = \frac{1}{\sigma^2(\mu)} > 0$ (see Appendix D.2).

□

D. Properties of $d(\mu, x)$ and Fisher's information $I(\mu)$.

D.1. Properties of $d(\mu, x)$ as a function of x.

Recall from Section 3,

$$d(\mu,\tilde{\mu}) \triangleq KL(p_{\theta(\mu)}, p_{\theta(\tilde{\mu})}) = b(\theta(\tilde{\mu})) - b(\theta(\mu)) - b'(\theta(\mu))(\theta(\tilde{\mu}) - \theta(\mu)),$$

such that $b'(\theta_\mu) = \mu$, $b'(\theta(\tilde{\mu})) = \tilde{\mu}$ and $p_{\theta(\mu)}, p_{\theta(\tilde{\mu})} \in \mathcal{S}$.

Using Lemma D.1 and Section 3, we know that $b(\theta) \in C^\infty$ and a strictly convex function, hence it follows that $\theta(\mu) = b'^{-1}(\mu)$ is a continuous function in μ as well. Further we also know that, $\sigma^2(\mu) = b''(b'^{-1}(\mu)) > 0$.

$$d(\mu, x) = b(b'^{-1}(x)) - b(b'^{-1}(\mu)) + \mu(b'^{-1}(x) - b'^{-1}(\mu)).$$

Hence we get that $d(\mu, x)$ is a continuous function in (x, μ) for $\mu \in \mathcal{I}$ and $x \in \mathcal{I}$.

Now we move to the differentiability of $d(\mu, x)$ in x. Using Inverse mapping theorem, we know that $\theta(\mu) = b'^{-1}(\mu)$, is differentiable in μ and we have,

$$\frac{d\theta(x)}{dx} = \frac{1}{b''(b'^{-1}(x))}. \tag{65}$$

Hence it follows that, $d(\mu, x)$ is differentiable in x and we have,

$$\frac{\partial d(\mu,x)}{\partial x} = (x-\mu)\frac{d\theta(x)}{dx} = (x-\mu)\frac{1}{b''(b'^{-1}(x))}.$$

Since $b(\theta)$ is a twice differentiable strictly convex function in θ, hence it follows that $b''(\theta) > 0$. This further implies that, $\frac{\partial d(\mu,x)}{\partial x} > 0$ for $x > \mu$, $\frac{\partial d(\mu,x)}{\partial x} < 0$ for $x < \mu$ and $\frac{\partial d(\mu,x)}{\partial x} = 0$ at $x = \mu$. Hence we get that $d(\mu, x)$ is a strictly quasi-convex and unimodal function in x. Further $d(\mu, x) = 0$ if only if $x = \mu$.

Now we move to the twice differentiability of $d(\mu, x)$ in x. Recall,

$$\frac{\partial d(\mu, x)}{\partial x} = (x - \mu)\frac{d\theta(x)}{dx} = (x - \mu)\frac{1}{b''(b'^{-1}(x))}.$$

Now to get twice differentiability of $d(\mu, x)$ in x, $\theta(x)$ should be twice differentiable in x. Since $b(\theta) \in C^\infty$, we have,

$$\frac{d^2\theta(x)}{dx^2} = \frac{-b'''(b'^{-1}(x))}{(b''(b'^{-1}(x)))^3}. \tag{66}$$

Further, following holds as well,

$$\frac{\partial^2 d(\mu, x)}{\partial x^2} = \frac{d\theta(x)}{dx} + (x - \mu)\frac{d^2\theta(x)}{dx^2}. \tag{67}$$

It follows that for $d(\mu, x)$ to be twice continuous differentiable in x, we need $b(\theta)$ to be thrice continuous differentiable in θ. Since $b(\theta) \in C^\infty$, hence we get that $d(\mu, x) \in C^\infty$ as a function of x.

D.2. Fisher's information $I(\mu)$ and continuity of $I(\mu)$ in μ

For a given $\nu \in \mathcal{S}$ with mean μ, using Theorem 5.4 , Chapter 2 in Lehmann & Casella (2006), we know that,

$$I(\mu) = \frac{1}{\sigma^2(\mu)}.$$

Since $d(\mu, x)$ is twice differentiable in x, hence using (66), (66) and (67), we have,

$$I(\mu) = \frac{1}{\sigma^2(\mu)} = \frac{\partial^2 d(\mu, x)}{\partial x^2}\Big|_{x=\mu}.$$

Since $\sigma^2(\mu) = b''(b'^{-1}(\mu))$, hence it follows that $I(\mu)$ is continuous as $b(\theta) \in C^\infty$ (see Lemma D.1).

□

D.3. Properties of $b(\theta)$ and $\mu(\theta)$ as a function of θ for $\theta \in \Theta$.

Lemma D.1. *For $\theta \in \Theta$, $b(\theta) \in C^\infty$.*

Lemma D.2. *For $\theta \in \Theta$, $\mu(\theta)$ is a strictly increasing function of θ. Further $\mu(\theta) \in C^\infty$.*

Proof of Lemma D.1:

Fix a $\theta \in \Theta$. Since p_θ is a Radon–Nikodym derivative. Hence it follows that,

$$\int_{\mathbf{R}} \exp(\theta \cdot x - b(\theta))d\xi(x) = 1.$$

It can be re-written as,

$$b(\theta) = \int_{\mathbf{R}} \exp(\theta \cdot x)d\xi(x).$$

Define a function $M_1(\theta, \eta) = \mathbb{E}[exp(\eta \cdot X_\theta))]$, where X_θ is distributed according to p_θ and η is a small number such that $\theta + \eta \in \Theta$. Hence it follows that,

$$M_1(\theta, \eta) = \int_{\mathbf{R}} \exp((\theta + \eta) \cdot x - b(\theta))d\xi(x).$$

The above can be re-written as,

$$M_1(\theta, \eta) = exp(b(\theta + \eta) - b(\theta)) \int_{\mathbf{R}} \exp((\theta + \eta) \cdot x - b(\theta + \eta)) d\xi(x).$$

Since $\theta + \eta \in \Theta$, hence it follows that,

$$M_1(\theta, \eta) = exp(b(\theta + \eta) - b(\theta)).$$

The above can be re-written as,

$$b(\theta + \eta) = \log(M_1(\theta, \eta)) + b(\theta).$$

Since $\log(M_1(\theta, \eta))$ as a function of η is the log-moment generating function of X_θ) for a given $\theta \in \Theta$. Using e.g., 2.2.24 in Dembo & Zeitouni (2009), we know that $\log(M_1(\theta, \eta))$ is infinitely differentiable in η for small η such that $\theta + \eta \in \Theta$ as $b(\theta + \eta)$ is well defined for such values of η. This further implies that, $\frac{\partial^n b(\theta+\eta)}{\partial \eta^n}\big|_{\eta=0}$ exists for all $n \in \mathbb{Z}^+$. This completes the proof.

□

Proof of Lemma D.2:

Since we know that $\mu(\theta) = b'(\theta)$. Using Lemma D.1 and the fact that $b(\theta)$ is a strictly convex function, we get the desired result.

□

D.4. Specific examples in canonical SPEF $\mathcal{S}$.

Now we consider the special cases of canonical SPEF (see Cappé et al. (2013)).

Gaussian distribution with variance σ^2: Here $\mathcal{I} = \mathbb{R}$, $\Theta = \mathbb{R}$, $\theta(\mu) = \frac{\mu}{\sigma^2}$ and $b(\theta) = \frac{\sigma^2\theta^2}{2}$.

$$d(\mu, x) = \frac{(x - \mu)^2}{\sigma^2}.$$

Binomial distribution with n samples: Here $I = (0, n)$, $\Theta = \mathbb{R}$, $\theta(\mu) = \log(\mu/(n - \mu))$ and $b(\theta) = n(\log(1 + e^\theta))$.

Notice that the special case of $n = 1$ corresponds to the Bernoulli distribution.

$$d(\mu, x) = \mu \log\left(\frac{\mu}{x}\right) + (n - \mu) \log\left(\frac{n - \mu}{n - x}\right).$$

Poisson distribution: Here $I = (0, +\infty)$, $\Theta = \mathbb{R}$, $\theta(\mu) = \log(\mu)$ and $b(\theta) = e^\theta$.

$$d(\mu, x) = x - \mu + \mu \log\left(\frac{\mu}{x}\right).$$

Gamma distributions with known shape parameter $r > 0$: Here $I = (0, \infty)$, $\Theta = (-\infty, 0)$, $\theta(\mu) = -r/\mu$ and $b(\theta) = -r\log(-\theta)$.

Notice that the special case of $r = 1$ corresponds to the exponential distribution.

$$d(\mu, x) = r\left(\frac{\mu}{x} - 1 - \log\frac{\mu}{x}\right).$$

Negative binomial distributions with known shape parameter $r > 0$: Here $I = (0, \infty)$, $\Theta = (-\infty, 0)$, $\theta(\mu) = \log(\mu/(r + \mu))$ and $b(\theta) = -r\log(1 - e^\theta)$.

Notice that the case r = 1 corresponds to geometric distributions.

$$d(\mu, x) = r\log\left(\frac{r + x}{r + \mu}\right) + \mu \log\left(\frac{\mu(r + x)}{x(r + \mu)}\right).$$

□

E. Supporting Lemmas

Lemma E.1. *If ν_1 and $\nu_2 \in \mathcal{S}$ with mean μ and x respectively, then there exists a $\eta > 0$ such that $d(\mu, x)$ is strictly convex function in x for $x \in (\mu - \eta, \mu + \eta)$ and a given $\mu \in \mathcal{I}$.*

Proof of Lemma E.1: We know that $\frac{\partial^2 d(\mu,x)}{\partial x^2}\Big|_{x=\mu} = \frac{1}{\sigma^2(\mu)}$. Since $d(\mu, x) \in C^\infty$ as a function of x for a given μ and $\sigma^2(\mu) > 0$, hence it follows that there exists a $\eta > 0$ such that $\frac{\partial^2 d(\mu,x)}{\partial x^2}\Big|_{x=c} > 0$ for $c \in (\mu - \eta, \mu + \eta)$. This completes the proof. □

Lemma E.2. *$T(\mu_A, \mu_B, w, z)$ is a jointly continuous function in (μ_A, μ_B, w, z) for $w \in [0, 1]$, $z \in (\underline{\mu} - \overline{\mu}, \overline{\mu} - \underline{\mu})$, $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$.*

Proof of Lemma E.2:

Recall the definition of $T(\mu_A, \mu_B, w, z)$,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \mathcal{C}(z)} wd(\mu_A, x) + (1 - w)d(\mu_B, y), \text{ where,}$$

$\mathcal{C}(z) = \{(x, y) : x, y \in \mathcal{I}, x - y = z\}$.

First, we consider a case for $z \leq \mu_A - \mu_B$. Using strict quasi-convexity of $d(\mu, x)$ in x, we get,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \overline{\mathcal{C}}(z)} wd(\mu_A, x) + (1 - w)d(\mu_B, y), \text{ where,}$$

$\overline{\mathcal{C}}(z) = \{(x, y) : x, y \in \mathcal{I}, x \in [\mu_B + z, \mu_A], y \in [\mu_B, \mu_A - z]\}$.

Using Berge's Maximum theorem, we get the desired result. A similar proof will follow for the other case which is $\Delta_L \geq \mu_A - \mu_B$.

□

Lemma E.3. *For a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$ and $w \in (0, 1)$, following holds: $T(\mu_A, \mu_B, w, z)$ is a strictly decreasing function of z for $z \in (\underline{\mu} - \overline{\mu}, \Delta)$ and a strictly increasing function of z for $z \in (\Delta, \overline{\mu} - \underline{\mu})$. Further, $T(\mu_A, \mu_B, w, z) = 0$ iff $z = \Delta$, for a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$ and $w \in (0, 1)$.*

Proof of Lemma E.3: Recall,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \mathcal{C}(z)} wd(\mu_A, x) + (1 - w)d(\mu_B, y), \text{ where,}$$

$\mathcal{C}(z) = \{(x, y) : x, y \in \mathcal{I}, x - y = z\}$. Using the strict quasi convexity of $d(\mu_A, x)$ in x, we get, for $z \in (\underline{\mu} - \overline{\mu}, \Delta)$, we have,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \hat{\mathcal{C}}_1(z)} wd(\mu_A, x) + (1 - w)d(\mu_B, y), \text{ where,}$$

$\hat{\mathcal{C}}_1(z) = \{(x, y) : x, y \in \mathcal{I}, x - y \leq z\}$. Hence it follows trivially that for $z \in (\underline{\mu} - \overline{\mu}, \Delta)$, $T(\mu_A, \mu_B, w, z)$ is decreasing in z. To get that $T(\mu_A, \mu_B, w, z)$ is strictly decreasing in z, first recall from the proof of Lemma E.2 ,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \overline{\mathcal{C}}(z)} wd(\mu_A, x) + (1 - w)d(\mu_B, y), \text{ where,}$$

$\overline{\mathcal{C}}(z) = \{(x, y) : x, y \in \mathcal{I}, x \in [\mu_B + z, \mu_A], y \in [\mu_B, \mu_A - z]\}$.

observe that for any $z < z' < \Delta$, take any $(x, y) \in \overline{\mathcal{C}}(z)$, it follows using the fact that $d(\mu, x)$ is strictly increasing in x for $x > \mu$ and strictly decreasing in x for $x < \mu$, there exists a $(x', y') \in \overline{\mathcal{C}}(z')$, such that,

$$wd(\mu_A, x) + (1 - w)d(\mu_B, y) > wd(\mu_A, x') + (1 - w)d(\mu_B, y').$$

Using the Lemma E.2 and the compactness of the set $\overline{\mathcal{C}}(z)$, we get the desired result. A similar proof will follow for $z \in (\Delta, \overline{\mu} - \underline{\mu})$. This completes the proof of the lemma.

□

Lemma E.4. *There exists a $\epsilon_1 > 0$, such that $T(\mu_A, \mu_B, w, z)$ is a continuously differentiable in (w, Δ_L) for $w \in (0,1)$ and $z \in [\Delta - \epsilon, \Delta + \epsilon]$ for $\epsilon \leq \epsilon_1$, $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$.*

Proof of Lemma E.4: Recall,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \mathcal{C}(z)} wd(\mu_A, x) + (1-w)d(\mu_B, y), \text{ where,}$$

$\mathcal{C}(z) = \{(x,y) : x, y \in \mathcal{I}, x - y = z\}$.

It follows that for any $w \in (0,1)$, $z \in (\underline{\mu} - \overline{\mu}, \overline{\mu} - \underline{\mu})$, $\mathcal{C}(z)$ is a non-empty set. First, we consider a case for $z \leq \Delta$. Using strict quasi-convexity of $d(\mu, x)$ in x, we get,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \overline{\mathcal{C}}(z)} wd(\mu_A, x) + (1-w)d(\mu_B, y), \text{ where,} \tag{68}$$

$\overline{\mathcal{C}}(z) = \{(x,y) : x, y \in \mathcal{I}, x \in [\mu_B + z, \mu_A], y \in [\mu_B, \mu_A - z]\}$.

For the case, $z \geq \Delta$. Using strict quasi-convexity of $d(\mu, x)$ in x, we get,

$$T(\mu_A, \mu_B, w, z) = \min_{x,y \in \underline{\mathcal{C}}(z)} wd(\mu_A, x) + (1-w)d(\mu_B, y), \text{ where,} \tag{69}$$

$\underline{\mathcal{C}}(z) = \{(x,y) : x \in \mathcal{I}, y \in x \in \mathcal{I}, x \in [\mu_A, \mu_B + z], y \in [\mu_A - z, \mu_B]\}$.

Using Lemma E.1, we know that $d(\mu, x)$ is strictly convex for $x \in (\mu - \eta, \mu + \eta)$ where η is a small positive number. Hence it follows that there exists a $\epsilon_1 > 0$ such that for $\epsilon \leq \epsilon_1$, the optimization problem given in (68) is a convex optimization problem for a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$, $w \in (0,1)$, $z \in [\Delta - \epsilon, \Delta]$. Similarly, the optimization problem given in (69) is a convex optimization problem for a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$, $w \in (0,1)$, $z \in [\Delta, \Delta + \epsilon]$.

Hence for a given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$, $w \in (0,1)$ and $z \in [\Delta - \epsilon, \Delta + \epsilon]$, the solution of (3), denoted as $x^*(\mu_A, \mu_B, w, z)$, uniquely satisfies,

$$w \frac{\partial d(\mu_A, x)}{\partial x}\Big|_{x=x^*} + (1-w)\, \frac{\partial d(\mu_B, x - z)}{\partial x}\Big|_{x=x^*} = 0.$$

Also $y^*(\mu_A, \mu_B, w, z) = x^*(\mu_A, \mu_B, w, z) - z$.

Using the implicit function and strict convexity of $d(\mu, x)$ in x for $x \in (\mu - \eta, \mu + \eta)$ and uniqueness of $x^*(\mu_A, \mu_B, w, z)$, we get that $x^*(\mu_A, \mu_B, w, z)$ is twice differentiable in (w, z). It follows that,

$$T(\mu_A, \mu_B, w, z) = wd(\mu_A, x^*(\mu_A, \mu_B, w, z)) + (1-w)d(\mu_B, y^*(\mu_A, \mu_B, w, z)).$$

We get the desired result from the above. This completes the proof.

□

Lemma E.5. *There exists a $\epsilon_1 > 0$, for given $\mu_A \in \mathcal{I}$, $\mu_B \in \mathcal{I}$ and any $w \in (0,1)$, there exists a $\tilde{\Delta}_L(w, \mu_A, \mu_B, \epsilon) \in (\Delta - \epsilon, \Delta)$ such that following holds for $\epsilon \leq \epsilon_1$,*

$$T(\mu_A, \mu_B, w, \tilde{\Delta}_L(w, \mu_A, \mu_B, \epsilon)) = T(\mu_A, \mu_B, w, \tilde{\Delta}_L(w, \mu_A, \mu_B, \epsilon) + \epsilon).$$

Further, we have,

$$T(\mu_A, \mu_B, w^*, \Delta_L^*) = T(\mu_A, \mu_B, w^*, \Delta_R^*).$$

Proof of Lemma E.5:

For any given $w \in (0,1)$, using Lemma E.3, we get that $T(\mu_A, \mu_B, w, z)$ is a strictly increasing function in z for $z \in (\Delta, \overline{\mu} - \underline{\mu})$ and is a strictly decreasing function in z for $z \in (\underline{\mu} - \overline{\mu}, \Delta)$.

Using Lemma E.3, we also know that $T(\mu_A, \mu_B, w, \Delta) = 0$. Now using the above properties it follows that, for any $w \in (0,1)$, there exists a $\tilde{\Delta}_L(w, \mu_A, \mu_B, \epsilon) \in (\Delta - \epsilon, \Delta)$ such that following holds for small ϵ,

$$
\begin{aligned}
\max_{\Delta_L \in (\Delta-\epsilon, \Delta)} \min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\} &= T(\mu_A, \mu_B, w, \tilde{\Delta}_L(w, \mu_A, \mu_B, \epsilon)) \\
&= T(\mu_A, \mu_B, w, \tilde{\Delta}_L(w, \mu_A, \mu_B, \epsilon) + \epsilon).
\end{aligned}
$$

Since the above holds for any $w \in (0,1)$, it will hold for w^*. Hence,

$$
\begin{aligned}
\max_{\Delta_L \in (\Delta-\epsilon, \Delta)} \min\{T(\mu_A, \mu_B, w^*, \Delta_L), T(\mu_A, \mu_B, w^*, \Delta_L + \epsilon)\} &= T(\mu_A, \mu_B, w^*, \tilde{\Delta}_L(w^*, \mu_A, \mu_B, \epsilon)) \\
&= T(\mu_A, \mu_B, w^*, \tilde{\Delta}_L(w^*, \mu_A, \mu_B, \epsilon) + \epsilon).
\end{aligned}
$$

It follows that, $\Delta_L^* = \tilde{\Delta}_L(w^*, \mu_A, \mu_B, \epsilon)$, hence this completes the proof of the lemma.

□

Lemma E.6. *There exists an unique $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ which satisfies the (5).*

Recall that we find $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ such that following holds,

$$
T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_L(n)\right) = T\left(\hat{\mu}_A(n), \hat{\mu}_B(n), \frac{N_A(n)}{n}, \hat{\Delta}_R(n)\right) = \frac{\beta(n,\delta)}{n}.
$$

Using Lemma E.5, we get, for any given $\hat{\mu}_A(n)$, $\hat{\mu}_B(n)$, $\frac{N_A(n)}{n}$, there exists an unique $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ which satisfies the above equation for large n,. This completes the proof.

□

Lemma E.7. *For a given $w \in (0,1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, if $\min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}$ is a jointly strictly quasi-concave function in w and Δ_L, $w^*(\mu_A, \mu_B, \epsilon)$ and $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ are unique and are jointly continuous functions in (μ_A, μ_B) for $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$.*

Proof of Lemma E.7

Recall that, $w^*(\mu_A, \mu_B, \epsilon)$ and $\Delta_L^*(\mu_A, \mu_B, \epsilon)$ is the solution of following equation.

$$
\max_{\substack{w \in [0,1], \\ \Delta_L \in [\Delta_L - \epsilon, \Delta]}} \min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}.
$$

Using Lemma E.2 and Sundaram (1996) (Corollary 9.20), we get the desired result.

□

Lemma E.8. *Under the assumptions of Lemma E.7, for the assignment rule of the policy $\mathfrak{P}_1$, following holds*

$$
\lim_{n\to\infty} \frac{N_A(n)}{n} = w^*(\mu_A, \mu_B, \epsilon) \textit{ almost surely.}
$$

Further following holds as well,

$$
\lim_{n\to\infty} \hat{\mu}_A(n) = \mu_A \textit{ and } \lim_{n\to\infty} \hat{\mu}_B(n) = \mu_B \textit{ almost surely.}
$$

Proof of the above lemma follows trivially from the Proposition 9 of Garivier & Kaufmann (2016), if $w^*(\mu_A, \mu_B, \epsilon)$ is a continuous function in (μ_A, μ_B) for $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$. It follows that we have the continuity of $w^*(\mu_A, \mu_B, \epsilon)$ in (μ_A, μ_B) for $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$ using Lemma E.7.

Lemma E.9. *For the assignment rule of the policy $\mathfrak{P}_2$, following holds*

$$\lim_{n\to\infty} \frac{N_A(n)}{n} = \overline{w}(\mu_A, \mu_B) \text{ almost surely,}$$

where $\overline{w}(\mu_A, \mu_B) = \frac{\sqrt{\frac{1}{I(\mu_A)}}}{\sqrt{\frac{1}{I(\mu_A)}}+\sqrt{\frac{1}{I(\mu_B)}}}$. *Further following holds as well,*

$$\lim_{n\to\infty} \hat{\mu}_A(n) = \mu_A \text{ and } \lim_{n\to\infty} \hat{\mu}_B(n) = \mu_B \text{ almost surely.}$$

Proof of the above lemma follows trivially from Proposition 9 of Garivier & Kaufmann (2016) as long as $\overline{w}(\mu_A, \mu_B)$ is a continuous function in (μ_A, μ_B) for $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$. It follows that under the assumption that $I(\mu)$ is continuous in μ for $\mu \in \mathcal{I}$ and we know that $I(\mu) > 0$ using the fact that $\sigma^2(\mu) > 0$ and $I(\mu) = \frac{1}{\sigma^2(\mu)}$ for $\mu \in \mathcal{I}$.

□

Lemma E.10. *([Garivier & Kaufmann (2016)], Lemma 8 and Lemma 20) For the policy $\mathfrak{P}_1$, under the assumptions stated in Lemma E.7, D-Tracking rule ensures that $N_k(n) \geq \max\{\sqrt{n}-1, 0\} - 1$ for $k \in \{A, B\}$. Further, there exists a constant n_η such that for $n \geq n_\eta$, it holds for our assignment rule of policy $\mathfrak{P}_1$, under the set $\mathcal{G}_\eta(n)$,*

$$\forall i \geq \sqrt{n}, \max_{k\in A,B} \left| \frac{N_k(i)}{i} - w^*(\mu_A, \mu_B, \epsilon) \right| \leq 3\eta.$$

Also for the assignment rule of policy $\mathfrak{P}_2$, one needs to replace $w^*(\mu_A, \mu_B, \epsilon)$ with $\overline{w}(\mu_A, \mu_B)$ in the above.

Proof of Lemma E.10 From Lemma E.7, we get that $w^*(\mu_A, \mu_B, \epsilon)$ is unique and continuous in (μ_A, μ_B) for $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$. Once we have the uniqueness and continuity of $w^*(\mu_A, \mu_B, \epsilon)$, proof follows similarly as given in the Garivier & Kaufmann (2016). Similarly for $\mathfrak{P}_2$, using the proof of Lemma E.10, we get that $\overline{w}(\mu_A, \mu_B)$ is continuous in (μ_A, μ_B) for $\mu_A \in \mathcal{I}$ and $\mu_B \in \mathcal{I}$ under the assumption that $I(\mu)$ is a continuous function of μ for $\mu \in \mathcal{I}$. This completes the proof.

□

F. Value of asymptotically optimal adaptive policy over uniform randomized policy

Here, we aim to quantify the value of any asymptotically optimal adaptive policy's assignment rule relative to the practically used uniform randomized assignment rule in terms of reducing the expected sample size. In fact, we theoretically prove that to deliver on $(\epsilon, \delta)-$correct guarantee the expected sample size required by any uniform randomized policy can be at most twice as large by any asymptotically optimal adaptive policy for small δ.

To prove the above, first, we prove a result about the lower bound on the sample size required for any uniform randomized policy to estimate a CI of ATE with $(\epsilon, \delta)-$coverage guarantee in the limiting regime of $\delta \to 0$.

Proposition F.1. *For given $\nu_A, \nu_B \in \mathcal{S}$ with mean μ_A and μ_B respectively and any $(\epsilon, \delta)-$coverage, uniform randomized stable policy with an almost surely finite stopping time τ_δ, we have*

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)}, \tag{70}$$

where,

$$\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon) = \sup_{(\Delta_L, \Delta_R)\in\Upsilon(\epsilon)} \min\{T(\mu_A, \mu_B, 0.5, \Delta_L), T(\mu_A, \mu_B, 0.5, \Delta_R)\}.$$

It is worth noting that we can show that the lower bound for any uniform randomized policy is also tight via developing a uniform randomized policy which matches the lower bound via modifying the assignment rule of our policy $\mathfrak{P}_1$ to the uniform randomized assignment rule and keeping estimation and stopping rule same. Recall we denoted it as $\mathfrak{P}_{\text{RCT}}$. Hence we compare the lower bound on the expected sample size for any adaptive policy and the lower bound on the expected sample size for any uniform randomized policy. Note the above lower bound parallels Theorem 4.2 in terms of order, and the constant is $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$.

Now we denote any uniform randomized stable policy which matches the lower bound provided in (70) when $\delta \to 0$ and has $(\epsilon, \delta)-$coverage guarantee, as $\mathfrak{P}^{\text{unif}}$. It follows that $\mathfrak{P}_{\text{RCT}}$ belongs to the set $\mathfrak{P}^{\text{unif}}$. Hence to compare the performance of any a.o. adaptive policy with any $\mathfrak{P}^{\text{unif}}$, we must compare $\ell^*(\mu_A, \mu_B, \epsilon)$ with $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$.

Proposition F.2. *For given $\nu_A, \nu_B \in \mathcal{S}$ with mean μ_A and μ_B respectively, then following holds,*

$$\ell^*(\mu_A, \mu_B, \epsilon) \geq \ell^{\text{unif}}(\mu_A, \mu_B, \epsilon) \geq \frac{\ell^*(\mu_A, \mu_B, \epsilon)}{2}. \tag{71}$$

Further, if $d(\mu, x)$ is twice continuously differentiable in x around a neighbourhood of $x = \mu$, we have

$$\lim_{\epsilon \to 0} \frac{\ell^*(\mu_A, \mu_B, \epsilon)}{\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)} = R\left(\frac{I(\mu_A)}{I(\mu_B)}\right), \tag{72}$$

where,

$$R(x) = \frac{2\left(\sqrt{x} + \frac{1}{\sqrt{x}}\right)}{\left(\sqrt{x} + \frac{1}{\sqrt{x}} + 2\right)}.$$

Since by definition, we know that any uniform randomized policy will be a feasible policy under the set of adaptive policies, hence it follows that $\ell^*(\mu_A, \mu_B, \epsilon) \geq \ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$. In (71), we also show that $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon) \geq \frac{\ell^*(\mu_A, \mu_B, \epsilon)}{2}$ which implies that expected sample size to deliver on $(\epsilon, \delta)-$correct guarantee with any $\mathfrak{P}^{\text{unif}}$ can be at most twice as large as compared to any asymptotically optimal adaptive policy for small δ. Interestingly in (72), we characterize the ratio of $\frac{\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)}{\ell^*(\mu_A, \mu_B, \epsilon)}$ when $\epsilon \to 0$ which is given by $R\left(\frac{I(\mu_A)}{I(\mu_B)}\right)$. We now plot the function $R(x)$ with $x = \frac{\min\{I(\mu_A), I(\mu_B)\}}{\max\{I(\mu_A), I(\mu_B)\}}$ takes value in between $(0, 1]$. The plot is shown in Figure 2. This result helps us in characterizing the regime when there is a value of any asymptotically optimal adaptive policy over any $\mathfrak{P}^{\text{unif}}$ and when there is no gain of any asymptotically optimal adaptive policy over $\mathfrak{P}^{\text{unif}}$ for small ϵ and δ. We observe that when $\frac{\min\{I(\mu_A), I(\mu_B)\}}{\max\{I(\mu_A), I(\mu_B)\}}$ is 1, then any $\mathfrak{P}^{\text{unif}}$ is asymptotically optimal for small ϵ and δ. Hence, in this case, there is no value of any a.o policy over $\mathfrak{P}^{\text{unif}}$. But as $\frac{\min\{I(\mu_A), I(\mu_B)\}}{\max\{I(\mu_A), I(\mu_B)\}}$ ratio decreases, there is a value of any a.o policy over any $\mathfrak{P}^{\text{unif}}$ and as $\frac{\min\{I(\mu_A), I(\mu_B)\}}{\max\{I(\mu_A), I(\mu_B)\}}$ approaches 0, the value of any a.o policy over any $\mathfrak{P}^{\text{unif}}$ increases to 50% reduction in the sample size.

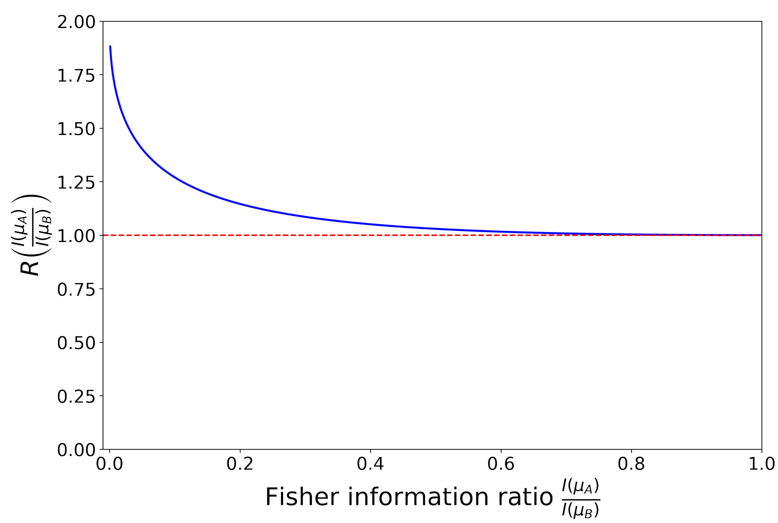

Figure 2: Value of adaptivity over uniform randomized policy

Proof of Proposition F.1: Proof of (70) follows similar to the proof of Theorem 4.2.

□

Proof of Proposition F.2: Now we show that (71) holds. From the definition of $\ell^*(\mu_A, \mu_B, \epsilon)$ and $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$, it follows trivially that, $\ell^*(\mu_A, \mu_B, \epsilon) \geq \ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$.

Now we show that, $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon) \geq \ell^*(\mu_A, \mu_B, \epsilon)/2$. Define a function for a given μ_A and μ_B for any $w_1 \geq 0$ and $w_2 \geq 0$,

$$\mathfrak{f}(w_1, w_2, \Delta_L) \triangleq \min_{x,y \in \mathcal{C}(\Delta_L) \cup \mathcal{C}(\Delta_L + \epsilon)} w_1 d(\mu_A, x) + w_2 d(\mu_B, y).$$

First, $\mathfrak{f}$ is a homogeneous function in (w_1, w_2) of degree one, i.e., $\mathfrak{f}(cw_1, cw_2, \Delta_L) = c\mathfrak{f}(w_1, w_2, \Delta_L)$ for any $c \geq 1$. second it follows that $\mathfrak{f}$ is non-decreasing in (w_1, w_2) for $w_1 \geq 0$ and $w_2 \geq 0$.

Let w^* and Δ_L^* denote any solution of $\mathfrak{L}$ (since we are not assuming that w^* and Δ_L^* is unique in the theorem, hence for the proof we choose any solution of $\mathfrak{L}$). It follows that,

$$r\mathfrak{f}(0.5, 0.5, \Delta_L) = \mathfrak{f}(0.5r, 0.5r, \Delta_L) \geq \mathfrak{f}(w^*, 1 - w^*, \Delta_L),$$

where $r = \max\{w^*/0.5, (1 - w^*)/0.5\}$. The above can be re-written as,

$$\mathfrak{f}(0.5, 0.5, \Delta_L) \geq 2\mathfrak{f}(w^*, 1 - w^*, \Delta_L).$$

It follows that,

$$\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon) = \max_{\Delta_L \in (\Delta - \epsilon, \Delta)} \mathfrak{f}(0.5, 0.5, \Delta_L) \text{ and } \ell^*(\mu_A, \mu_B, \epsilon) = \max_{\Delta_L \in (\Delta - \epsilon, \Delta)} \mathfrak{f}(w^*, 1 - w^*, \Delta_L).$$

Using the above, we complete the proof of (71).

Now we move to the proof of (72). In the definition of $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$, we have maximization over Δ_L and Δ_R and we have fixed $w = 0.5$. Using Lemma E.5, for small ϵ, we get that the solution of optimization problem in $\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)$, denoted as $\tilde{\Delta}_L$ and $\tilde{\Delta}_R$, is unique and satisfies,

$$T(\mu_A, \mu_B, 0.5, \tilde{\Delta}_L) = T(\mu_A, \mu_B, 0.5, \tilde{\Delta}_R), \text{ and } \tilde{\Delta}_R = \tilde{\Delta}_L + \epsilon.$$

Similar to the proof of Theorem 5.3, using the second-order Taylor series expansion of the above equation, we get,

$$\lim_{\epsilon \to 0} \frac{\ell^{\text{unif}}(\mu_A, \mu_B, \epsilon)}{\epsilon^2} = \frac{1}{16} \left(\frac{I(\mu_A) I(\mu_B)}{I(\mu_A) + I(\mu_B)} \right).$$

Recall that, from Theorem 5.3, we get,

$$\lim_{\epsilon \to 0} \frac{\ell^*(\mu_A, \mu_B, \epsilon)}{\epsilon^2} = \frac{1}{8 \left(\sqrt{\frac{1}{I(\mu_A)}} + \sqrt{\frac{1}{I(\mu_B)}} \right)^2}.$$

Using the above two equations, we get the desired result.

□

G. Details of numerical experiments mentioned in Section 7

We set the distributions of the outcome of treatment A and B to be exponential distributions with mean $\mu_A = 10$ and $\mu_B = 0.1$ respectively. We choose ϵ to be 0.5 and we set $\delta = 10\%$, 5% and 1%. For each choice of δ, we generate 2000 sample paths and report the average numbers. We present the detailed performance of our policy $\mathfrak{P}_2$ and $\mathfrak{P}_{\text{RCT}}$ in Table 3 (in Table 1, we presented a small version of this table).

As mentioned in Section 7, for empirical studies, we use Kaufmann & Koolen (2021) for the choice of $\beta(n, \delta)$ for the estimation rule in $\mathfrak{P}_2$ and $\mathfrak{P}_{\text{RCT}}$. For exponential outcome distributions, using Theorem 10 of the same paper, we choose $\beta(n, \delta) = 2 \sum_{k \in \{A,B\}} \log[4 + \log(N_k(n))] + 2\mathcal{T}\left(\frac{\log(2/\delta)}{2}\right)$ accordingly as we have chosen exponential outcome

$\epsilon = 0.5$, $\mu_A = 10$ and $\mu_B = 0.1$					
δ value	Lower bound on $\mathbb{E}[\tau_\delta]$ given in (4)	Estimated $\mathbb{E}[\tau_\delta]$ for $\mathfrak{P}_2$	Estimated $\mathbb{E}[\tau_\delta]$ for $\mathfrak{P}_{\mathrm{RCT}}$	$\hat{P}_{\text{confidence sequence}}$ for $\mathfrak{P}_2$ in percentage(%)	$\hat{P}_{\text{confidence sequence}}$ for $\mathfrak{P}_{\mathrm{RCT}}$ in percentage(%)
10%	7.52×10^3	3.74×10^4	7.85×10^4	99.85 ± 0.1	99.60 ± 0.1
5%	9.78×10^3	4.14×10^4	8.62×10^4	100	100
1%	1.50×10^4	4.98×10^4	1.02×10^5	100	100

Table 3: Performance of $\mathfrak{P}_2$ and $\mathfrak{P}_{\mathrm{RCT}}$ policy for finite δ values. The width of 95% CI for estimated $\mathbb{E}[\tau_\delta]$ for both policies is less than 150.

distributions for the numerical experiment. To state the definition of $\mathcal{T}(x)$, we need to introduce two functions. First for $u \geq 1$ the function $z = \psi(u) = u - \ln u$ and its inverse $u = \psi^{-1}(z)$ for $z \geq 1$. And the other function is defined for any $y \in [1, e]$ and $x \geq 0$ and given by

$$\tilde{\psi}_y(x) = \begin{cases} e^{1/\psi^{-1}(x)}\psi^{-1}(x) & \text{if } x \geq \psi^{-1}(1/\ln y), \\ y(x - \ln\ln y) & \text{o.w.} \end{cases}$$

Now we define function $\mathcal{T}(x) : \mathbb{R}^+ \to \mathbb{R}^+$ as follows

$$\mathcal{T}(x) = 2\tilde{\psi}_{3/2}\left(\frac{\psi^{-1}(1+x) + \ln 2\zeta(2)}{2}\right)$$

where $\zeta(2) = \sum_{n=1}^{\infty} n^{-2}$.

Using the definition of $\mathcal{T}(x)$, it follows that $\mathcal{T}(x) = O(x)$, where $O(\cdot)$ denotes the Big O notation. Hence the above choice of $\beta(n, \delta) = O(\log(\log(n))/\delta))$.

We report the estimated number of the samples taken by each policy, i.e. estimated $\mathbb{E}[\tau_\delta]$. We report the asymptotic lower bound valid for small δ given in Theorem (4.2). We also report the estimation of the probability that ATE lies in the CI at the end of the experiment, we refer to it as $\hat{P}_{(\epsilon,\delta)-\text{coverage}}$. Last, we also report the estimation of probability which tells whether the confidence interval always contains the ATE or not during the entire sample path till the policy stops and we refer to it as $\hat{P}_{\text{confidence sequence}}$. The reason for estimating $\hat{P}_{\text{confidence sequence}}$ is to show that our estimation and stopping rule constructs the confidence sequence of ATE which allows "continuous monitoring" of A/B tests while central limit theorem-based confidence interval will fail miserably on this property. We have made one change in the estimation rule for $\mathfrak{P}_2$ and hence one change in the estimation rule for $\mathfrak{P}_{\mathrm{RCT}}$ for the numerical experiment. Recall in $\mathfrak{P}_2$ and $\mathfrak{P}_{\mathrm{RCT}}$, we estimate $(1-\delta)-$confidence sequence at each n, denoted as $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$. To stop early for a given δ, we estimate running intersection of $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$, which is defined as,

$$[\tilde{\Delta}_L(n), \tilde{\Delta}_R(n)] = \cap_{s=1}^{n}[\hat{\Delta}_L(s), \hat{\Delta}_R(s)],$$

and stop when $\tilde{\Delta}_R(n) - \tilde{\Delta}_L(n) \leq \epsilon$ for the first time. It follows that using $[\tilde{\Delta}_L(n), \tilde{\Delta}_R(n)]$, we will stop early and we will still have $(\epsilon, \delta)-$correct guarantee due to the definition of confidence sequence (see (6)). We observe that the CI of ATE of our policy $\mathfrak{P}_2$ and for $\mathfrak{P}_{\mathrm{RCT}}$ always contains ATE at the end of the experiment, i.e., $\hat{P}_{(\epsilon,\delta)-\text{coverage}} = 1$ and the ratio of actual estimated $\mathbb{E}[\tau_\delta]$ with the lower bound value is decreasing sharply as δ gets smaller.

Last, we present the performance of the $\mathfrak{P}_2$ with aggressive choice of $\beta(n, \delta) = \log\left(\frac{1+\log(n)}{\delta}\right)$ as mentioned in Barrier (2023) which is not theoretically supported but works well in practice. Results are shown in Table 4. In our set-up as well, we improve the performance of our policy $\mathfrak{P}_2$ as it stops much earlier and empirically $(\epsilon, \delta)-$correct guarantee holds as well. We also notice that, this aggressive choice of $\beta(n, \delta)$ loses out on the confidence sequence property.

□

$\epsilon = 0.5$, $\mu_A = 10$ and $\mu_B = 0.1$				
δ value	Lower bound on $\mathbb{E}[\tau_\delta]$ given in (4)	Estimated $\mathbb{E}[\tau_\delta]$	$\hat{P}_{(\epsilon,\delta)-\text{coverage}}$ in percentage(%)	$\hat{P}_{\text{confidence sequence}}$ in percentage(%)
10%	7.52×10^3	1.17×10^4	97.4 ± 0.5	84 ± 1
5%	9.78×10^3	1.40×10^4	99.6 ± 0.2	92.05 ± 1
1%	1.50×10^4	1.92×10^4	99.70 ± 0.1	97.50 ± 0.6

Table 4: Performance of $\mathfrak{P}_2$ policy with aggressive choice of $\beta(n, \delta)$ rule for finite δ values. The width of 95% CI for estimated $\mathbb{E}[\tau_\delta]$ for $\mathfrak{P}_2$ policy with aggressive choice of $\beta(n, \delta)$ is less than 150.

H. Discussion on stable policies and Remark 4.6

H.1. Discussion on stable policies

Recall a policy is called stable if $\hat{\Delta}_L(\tau_\delta) \xrightarrow{p} a$ and $\hat{\Delta}_R(\tau_\delta) \xrightarrow{p} b$ as $\delta \to 0$, where a and b are constants.

When a policy constructs a symmetric confidence interval at the conclusion of an A/B test and the estimator of the ATE is consistent, and as δ tends to zero, $\tau_\delta \to \infty$ almost surely, then, applying the Law of Large Numbers, it follows that the policy satisfies the stability assumption. For example, an A/B test with a symmetric confidence interval based on the central limit theorem (CLT) will meet the stability assumption for the reasons mentioned above. In this scenario, the values for a and b can be determined as $a = \Delta - \epsilon/2$ and $b = \Delta + \epsilon/2$. In general, if a policy produces a consistent estimator of the ATE, denoted as $\hat{\Delta}(n)$, and the boundaries of the confidence interval, $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$, are continuous functions of the ATE estimator, and as $\delta \to 0$, $\tau_\delta \to \infty$ almost surely, then by applying the Law of Large Numbers and considering the continuity of the confidence interval boundaries, we can establish the stability assumption. Our policy is stable for the same reason. This tells that stable assumption allows for a lot of tractable policies with very natural requirements such as consistency of estimator, length of the A/B approaching infinity when $\delta \to 0$ and continuity of the confidence interval boundaries with respect to the estimator of ATE.

H.2. Discussion on Assumption 4.5

Here we show that for given $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, $\min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}$ is jointly strictly quasi-concave in w and Δ_L. We will provide the proof when $\nu_A = N(\mu_A, \sigma^2)$ and $\nu_B = N(\mu_B, \sigma^2)$ with known variances. For other outcome distributions in $\mathcal{S}$, we do not have a formal proof but we provide a numerical study to show that the above assumption holds via plotting upper contour sets.

First, let $\nu_A = N(\mu_A, \sigma^2)$ and $\nu_B = N(\mu_B, \sigma^2)$ with known variances. It follows from the proof of Proposition 5.1, we have,

$$T(\mu_A, \mu_B, w, \Delta_L) = \frac{w(1-w)}{2\sigma^2}(\Delta_L - \Delta)^2.$$

We now show that $T(\mu_A, \mu_B, w, \Delta_L)$ is jointly strictly quasi-concave in w and Δ_L for $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$. For ease of readability, we suppress the notation and denote $T(\mu_A, \mu_B, w, \Delta_L)$ as $T(w, \Delta_L)$.

Now to show the strict quasi concavity of $T(\mu_A, \mu_B, w, \Delta_L)$ in (w, Δ_L), we use the sufficient conditions provided in Theorem 1.12 in Takayama (1993). We first compute the matrices $C_1(w, \Delta)$ and $C_2(w, \Delta_L)$. These matrices are defined as follows,

$$C_1(w, \Delta_L) \triangleq \begin{bmatrix} 0 & \frac{\partial T(w,\Delta_L)}{\partial \Delta_L} \\ \frac{\partial T(w,\Delta_L)}{\partial \Delta_L} & \frac{\partial^2 T(w,\Delta_L)}{\partial \Delta_L^2} \end{bmatrix},$$

$$C_2(w, \Delta_L) \triangleq \begin{bmatrix} 0 & \frac{\partial T(w,\Delta_L)}{\partial \Delta_L} & \frac{\partial T(w,\Delta_L)}{\partial w} \\ \frac{\partial T(w,\Delta_L)}{\partial \Delta_L} & \frac{\partial^2 T(w,\Delta_L)}{\partial \Delta_L^2} & \frac{\partial^2 T(w,\Delta_L)}{\partial \Delta_L \partial w} \\ \frac{\partial T(w,\Delta_L)}{\partial w} & \frac{\partial^2 T(w,\Delta_L)}{\partial w \partial \Delta_L} & \frac{\partial^2 T(w,\Delta_L)}{\partial w^2} \end{bmatrix}.$$

Using the definition of $T(w, \Delta_L)$, we get,

$$C_1(w, \Delta_L) = \begin{bmatrix} 0 & \frac{(\Delta_L-\Delta)w(1-w)}{\sigma^2} \\ \frac{(\Delta_L-\Delta)w(1-w)}{\sigma^2} & \frac{w(1-w)}{\sigma^2} \end{bmatrix},$$

$$C_2(w, \Delta_L) = \begin{bmatrix} 0 & \frac{(\Delta_L-\Delta)w(1-w)}{\sigma^2} & \frac{(1-2w)(\Delta_L-\Delta)^2}{2\sigma^2} \\ \frac{(\Delta_L-\Delta)w(1-w)}{\sigma^2} & \frac{w(1-w)}{\sigma^2} & \frac{(1-2w)(\Delta_L-\Delta)}{\sigma^2} \\ \frac{(1-2w)(\Delta_L-\Delta)^2}{2\sigma^2} & \frac{(1-2w)(\Delta_L-\Delta)}{\sigma^2} & \frac{-(\Delta_L-\Delta)^2}{\sigma^2} \end{bmatrix}.$$

To prove the quasi concavity of $T(w, \Delta_L)$ in (w, Δ_L), we need to show that $det(C_1(w, \Delta_L)) < 0$ and $det(C_2(w, \Delta_L)) > 0$ for all $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$.

Using some algebra, we get that for any $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$,

$$det(C_1(w, \Delta_L)) = -\left(\frac{(\Delta_L - \Delta)w(1-w)}{\sigma^2}\right)^2 < 0.$$

Using some algebra, we get that for any $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$,

$$det(C_2(w, \Delta_L)) = \frac{w(1-w)(\Delta_L - \Delta)^4}{\sigma^6}\left(\frac{3(1-2w)^2 + 4w(1-w)}{4}\right) > 0.$$

A similar proof will follow for the joint strictly quasi-concavity of $T(\mu_A, \mu_B, w, \Delta_L + \epsilon)$ in w and Δ_L for $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$. Since min of two strict quasi concave functions is also a strict quasi concave, hence it follows that $\min\{T(\mu_A, \mu_B, w, \Delta_L), T(\mu_A, \mu_B, w, \Delta_L + \epsilon)\}$ is jointly strictly quasi-concave in w and Δ_L. This completes the proof when outcome distributions are Gaussian with known variance. For the other relevant distributions in $\mathcal{S}$, we show upper contour plots of $T(\mu_A, \mu_B, w, \Delta_L)$ for $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$ in Figure 3, 4, 5 and 6. Since these upper contour sets are always strictly convex in all 4 figures. Using the definition of strict quasi concavity, it shows the joint strictly quasi-concavity of $T(\mu_A, \mu_B, w, \Delta_L + \epsilon)$ in w and Δ_L for $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$. numerically.

$\square$

I. Generalization to the non-parametric family with bounded support

In this section, we generalise our results to the setting where outcome distributions belong to the non-parametric family with bounded support in $[0, 1]$ denoted as $\mathcal{B}$. Recall that $m(\nu)$ denotes the mean of a distribution ν. For $\nu \in \mathcal{B}$, let $\sigma(\nu)$ denote the standard deviation of the distribution ν. Now we define the following functions which will help us to generalize the results (see Appendix F in Jourdan et al. (2022) and Honda & Takemura (2010)). For $\nu \in \mathcal{B}$, we define

$$D^U_{inf}(\nu, x) \triangleq \inf_{\kappa \in \mathcal{B}: m(\kappa) \geq x} KL(\nu, \kappa).$$

$$D^L_{inf}(\nu, x) \triangleq \inf_{\kappa \in \mathcal{B}: m(\kappa) \leq x} KL(\nu, \kappa).$$

Now we define,

$$D_{inf}(\nu, x) = \max\{D^L_{inf}(\nu, x), D^U_{inf}(\nu, x)\} \tag{73}$$

In the analysis, $D_{inf}(\nu, x)$ will replace the $d(\mu, x)$ function as KL divergence between two non-parametric distributions as it can not be defined via their means. Intuitively, $D^U_{inf}(\nu, x)(D^L_{inf}(\nu, x))$ represents the minimum KL divergence between ν and all the distributions in $\mathcal{B}$ which have a mean higher(less) or equal than x.

We will exclude the point mass distributions as our ATE problem is not well-defined for them. Hence it follows that for $\nu \in \mathcal{B}$ except point mass distribution will have mean in $(0, 1)$. We first state properties of $D^U_{inf}(\nu, x)$ and $D^L_{inf}(\nu, x)$.

Dual Representation of $D^U_{inf}(\nu, x)$ and $D^L_{inf}(\nu, x)$: It is well mentioned in the literature that dual representations of $D^U_{inf}(\nu, x)$ and $D^L_{inf}(\nu, x)$ are much more tractable. We now rewrite Theorem 3 of Jourdan et al. (2022). For $(\lambda, \nu, x) \in$

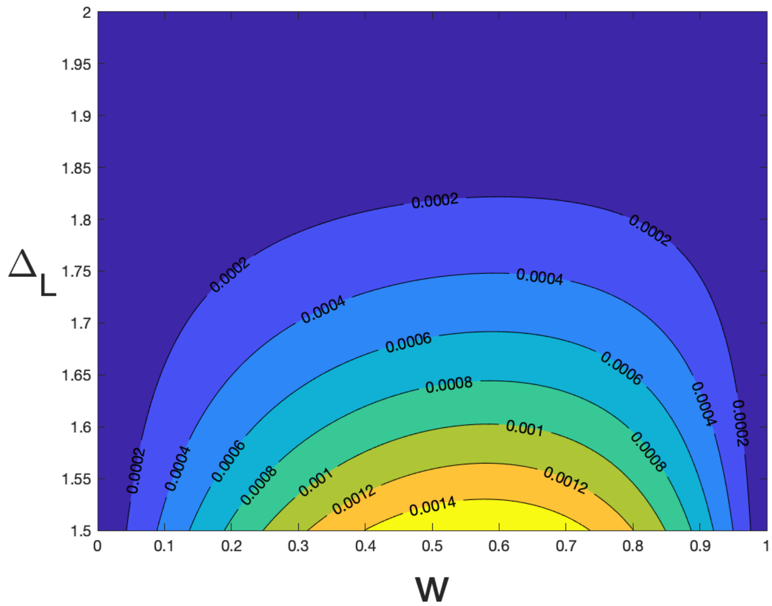

Figure 3: Upper contour plots for function $T(\mu_A, \mu_B, w, \Delta_L)$ in the space of (w, Δ_L) for a given $\mu_A = 5$ and $\mu_B = 3$ when outcome distributions are Geometric. Here we choose, $\epsilon = 0.5$, hence $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, i.e., $\Delta_L \in (1.5, 2)$. There are various lines in the above figure. The number on a given line in the above figure represents the value of $T(\mu_A, \mu_B, w, \Delta_L)$ for all values of (w, Δ_L) satisfying the trajectory of that given line for $\mu_A = 5$ and $\mu_B = 3$. One can observe that all the upper contour plots are strictly convex.

$\mathbb{N} \times \mathcal{B} \times [0, 1]$, let $H^+(\lambda, \nu, x) = \mathbb{E}_\nu[\log(1 - \lambda(X - x))]$, where we define $\log(x) = -\infty$ for $x \le 0$. Let $H^-(\lambda, \nu, x) = \mathbb{E}_\nu[\log(1 + \lambda(X - x))]$.

Theorem I.1. *For all $\nu \in \mathcal{B}$ and $x \in (0, 1)$, we have,*

$$D^U_{inf}(\nu, x) = \sup_{\lambda \in [0, 1/(1-x)]} H^+(\lambda, \nu, x).$$

$$D^L_{inf}(\nu, x) = \sup_{\lambda \in [0, 1/x]} H^-(\lambda, \nu, x).$$

Now we re-write some properties of $D^U_{inf}(\nu, x)$ and $D^L_{inf}(\nu, x)$ functions which are proven in Honda & Takemura (2010), Agrawal (2022) and Jourdan et al. (2022).

I.1. Properties of $D^U_{inf}(\nu, x)$ and $D^L_{inf}(\nu, x)$:

1. The function $D^U_{inf}(\nu, x)$ (resp. $D^L_{inf}(\nu, x)$) is continuous on $\mathcal{B} \times [0, 1)$ (resp. $\mathcal{B} \times (0, 1]$).
2. For all $(\nu, x) \in \mathcal{B} \times [0, 1)$, $D^U_{inf}(\nu, x) \le -\log(1 - x)$.
3. For all $(\nu, x) \in \mathcal{B} \times (0, 1]$, $D^L_{inf}(\nu, x) \le -\log(x)$.
4. The function $x \to D^U_{inf}(\nu, x)$ is strictly convex on $(m(\nu), 1]$. Further, the function $x \to D^L_{inf}(\nu, x)$ is strictly convex on $[0, m(\nu))$.
5. Let $\lambda^U(\nu, x) = \arg\max_{\lambda \in [0, 1/(1-x)]} H^+(\lambda, \nu, x)$ and $\lambda^L(\nu, x) = \arg\max_{\lambda \in [0, 1/x]} H^-(\lambda, \nu, x)$. $\lambda^U(\nu, x)$ is unique except for the case ν is a point mass distribution. For the case when ν is a point mass distribution, we define $\lambda^U(\nu, x) = \frac{1}{(1-x)}$. Similarly, $\lambda^L(\nu, x)$ is unique except for the case ν is a point mass distribution. For the case when ν is a point mass distribution, we define $\lambda^U(\nu, x) = \frac{1}{x}$.

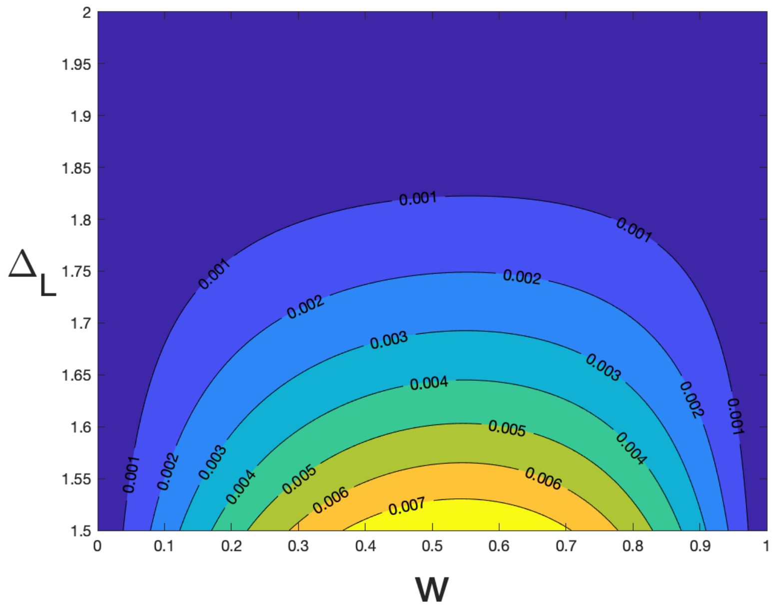

Figure 4: Upper contour plots for function $T(\mu_A, \mu_B, w, \Delta_L)$ in the space of (w, Δ_L) for a given $\mu_A = 5$ and $\mu_B = 3$ when outcome distributions are Poisson. Here we choose, $\epsilon = 0.5$, hence $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, i.e., $\Delta_L \in (1.5, 2)$. There are various lines in the above figure. The number on a given line in the above figure represents the value of $T(\mu_A, \mu_B, w, \Delta_L)$ for all values of (w, Δ_L) satisfying the trajectory of that given line for $\mu_A = 5$ and $\mu_B = 3$. One can observe that all the upper contour plots are strictly convex.

6. Let $\nu \in \mathcal{B}$ and $x^U(\nu) = 1 - \frac{1}{\mathbb{E}_{X\sim\nu}[1/(1-X)]} \geq m(\nu)$. We have,

$$\lambda^U(\nu, x) = 0 \iff x \leq m(\nu).$$

$$x \in (m(\nu), x^U(\nu)] \Longrightarrow \mathbb{E}_\nu \left[\frac{1}{1 - \lambda^U(\nu, x)(X - x)}\right] = 1.$$

$$\lambda^U(\nu, x) = \frac{1}{1 - x} \iff x \geq x^U(\nu).$$

7. For all $\nu \in \mathcal{B}$ and $x \in (m(\nu), 1]$, $x \to D^U_{inf}(\nu, x)$ is differentiable and

$$\frac{\partial D^U_{inf}(\nu, x)}{\partial x} = \lambda^U(\nu, x).$$

8. For all $\nu \in \mathcal{B}$ and $x \in [0, m(\nu))$, $x \to D^L_{inf}(\nu, x)$ is differentiable and

$$\frac{\partial D^L_{inf}(\nu, x)}{\partial x} = \lambda^L(\nu, x).$$

Remark I.2. Similar to the case of outcome distributions in $\mathcal{S}$ (see Remark 3.2), our problem is well-defined for $\epsilon < 2$ when outcome distributions have bounded support in [0,1].

In order to generalize Theorem 5.3 and Theorem 5.4, we need to do Taylor series expansion of the $D_{inf}(\nu, x)$ around $x = m(\nu)$. Hence we will need the four times continuous differentiability of $D_{inf}(\nu, x)$ for $x \in (m(\nu) - \eta, m(\nu) + \eta)$, where η is a small positive number. We next present a result which ensures the above.

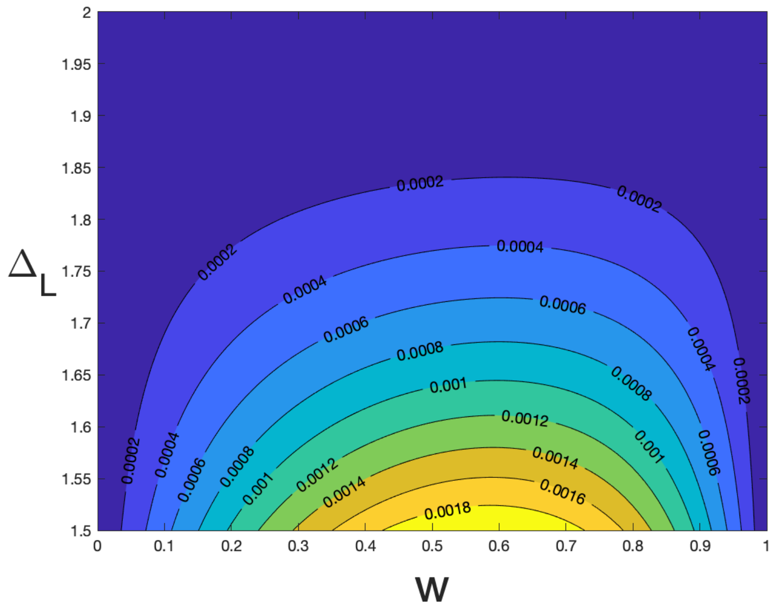

Figure 5: Upper contour plots for function $T(\mu_A, \mu_B, w, \Delta_L)$ in the space of (w, Δ_L) for a given $\mu_A = 5$ and $\mu_B = 3$ when outcome distributions are Exponential. Here we choose, $\epsilon = 0.5$, hence $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, i.e., $\Delta_L \in (1.5, 2)$. There are various lines in the above figure. The number on a given line in the above figure represents the value of $T(\mu_A, \mu_B, w, \Delta_L)$ for all values of (w, Δ_L) satisfying the trajectory of that given line for $\mu_A = 5$ and $\mu_B = 3$. One can observe that all the upper contour plots are strictly convex.

Lemma I.3. *For a given $\nu \in \mathcal{B}$, such that $m(\nu) \in (0, 1)$, there exists a $\eta > 0$ such that for $x \in (m(\nu) - \eta, m(\nu) + \eta)$, $D_{inf}(\nu, x)$ is four times continuously differentiable function in x. Further, we have,*

$$\frac{\partial^2 D_{inf}(\nu, x)}{\partial x^2}\Big|_{x=m(\nu)} = \frac{1}{\sigma^2(\nu)}.$$

$$\frac{\partial^3 D_{inf}(\nu, x)}{\partial x^3}\Big|_{x=m(\nu)} = \frac{2}{\sigma^6(\nu)\mathbb{E}_\nu[X - m(\nu)]^3}.$$

Now we are ready to generalize the main results for non-parametric bounded support distributions which were presented in the main sections of the paper for outcome distributions in $\mathcal{S}$.

I.2. ATE problem

Let

$$T(\nu_A, \nu_B, w, z) \triangleq \inf_{(x,y) \in \mathcal{C}^b(z)} w D_{inf}(\nu_A, x) + (1 - w) D_{inf}(\nu_A, y),$$

where $\mathcal{C}^b(z) = \{(x, y) : x \in (0, 1), y \in (0, 1), x - y = z\}$.

Theorem I.4. *For given $\nu_A, \nu_B \in \mathcal{B}$, any $(\epsilon, \delta)-$coverage and stable adaptive policy with an almost surely finite stopping time τ_δ, we have*

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\ell^*(\nu_A, \nu_B, \epsilon)},$$

where $\ell^*(\nu_A, \nu_B, \epsilon)$ is the solution of the following optimization problem (denoted by $\mathfrak{L}^b$),

$$\ell^*(\nu_A, \nu_B, \epsilon) = \sup_{\substack{w \in [0,1], \\ (\Delta_L, \Delta_R) \in \Upsilon(\epsilon)}} \min\{T(\nu_A, \nu_B, w, \Delta_L), T(\nu_A, \nu_B, w, \Delta_R)\}.$$

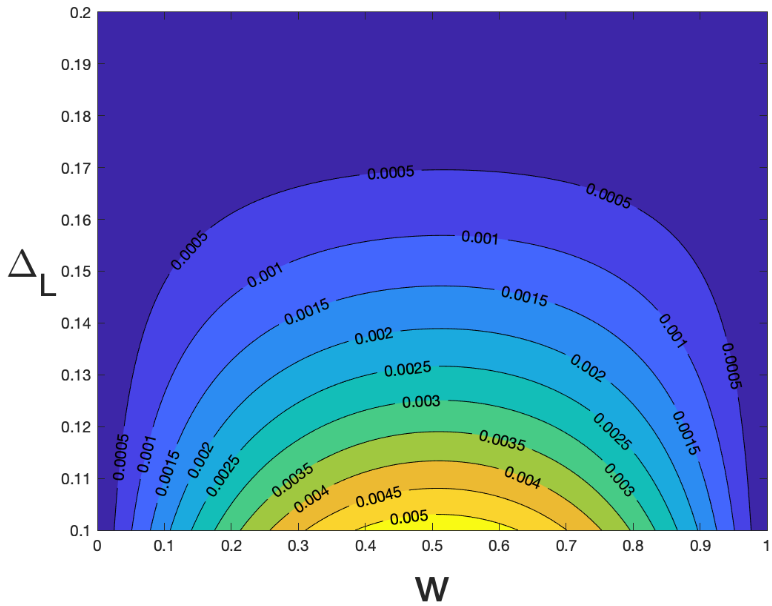

Figure 6: Upper contour plots for function $T(\mu_A, \mu_B, w, \Delta_L)$ in the space of (w, Δ_L) for a given $\mu_A = 0.5$ and $\mu_B = 0.3$ when outcome distributions are Bernoulli. Here we choose, $\epsilon = 0.1$, hence $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, i.e., $\Delta_L \in (1.5, 2)$. There are various lines in the above figure. The number on a given line in the above figure represents the value of $T(\mu_A, \mu_B, w, \Delta_L)$ for all values of (w, Δ_L) satisfying the trajectory of that given line for $\mu_A = 0.5$ and $\mu_B = 0.3$. One can observe that all the upper contour plots are strictly convex.

Theorem I.5. *For a given $\nu_A, \nu_B \in \mathcal{B}$, then following holds: a solution to the optimization problem $\mathfrak{L}^b$ exists, i.e., $w^*(\nu_A, \nu_B, \epsilon)$, $\Delta_L^*(\nu_A, \nu_B, \epsilon)$ and $\Delta_R^*(\nu_A, \nu_B, \epsilon)$ exists and any $\Delta_L^*(\nu_A, \nu_B, \epsilon)$ and $\Delta_R^*(\nu_A, \nu_B, \epsilon)$ that is a solution to the optimization problem $\mathfrak{L}^b$ satisfies,*

$$\Delta_R^*(\nu_A, \nu_B, \epsilon) = \Delta_L^*(\nu_A, \nu_B, \epsilon) + \epsilon \text{ and } \lim_{\epsilon \to 0} \frac{\ell^*(\nu_A, \nu_B, \epsilon)}{\epsilon^2} = \frac{1}{8\left(\sigma(\nu_A) + \sigma(\nu_B)\right)^2}.$$

$$\lim_{\epsilon \to 0} w^*(\nu_A, \nu_B, \epsilon) = \overline{w}^b(\nu_A, \nu_B) \text{ and } \lim_{\epsilon \to 0} \frac{\Delta_L^*(\nu_A, \nu_B, \epsilon) - \Delta}{\epsilon} = -\frac{1}{2}.$$

$$\lim_{\epsilon \to 0} \frac{w^*(\nu_A, \nu_B, \epsilon) - \overline{w}^b(\nu_A, \nu_B)}{\epsilon} = 0.$$

Further, we have,

$$\lim_{\epsilon \to 0} \frac{w^*(\nu_A, \nu_B, \epsilon) - \overline{w}(\nu_A, \nu_B)}{\epsilon^2} = \frac{v(\nu_A, \nu_B)}{96\sqrt{I(\mu_A) I(\mu_B)}},$$

where, $v(\nu_A, \nu_B) \triangleq \left(\frac{\partial^4 D_{inf}(\nu_B, x)}{\partial x^4}\Big|_{x=m(\nu_B)} \right) (1 - \overline{w}(\mu_A, \mu_B))^4 - \left(\frac{\partial^4 D_{inf}(\nu_A, x)}{\partial x^4}\Big|_{x=m(\nu_A)} \right) \overline{w}^4(\mu_A, \mu_B).$

Here, $\Delta = m(\nu_A) - m(\nu_B)$ and

$$\overline{w}^b(\nu_A, \nu_B) \triangleq \frac{\sigma(\nu_A)}{\sigma(\nu_A) + \sigma(\nu_B)}.$$

Remark I.6. It is worth noting that, for $\nu \in \mathcal{B}$ as well, using the above result, we get that the asymptotically optimal assignment rule of treatments is Neyman's allocation rule for small ϵ and δ. Using Lemma I.3, it also follows that $\frac{\partial^2 D_{inf}(\nu, x)}{\partial x^2}|_{x=m(\nu)}$ is the generalized notion of Fisher's information that we get in our non-parametric framework.

Proposition I.7. *For given $\nu_A, \nu_B \in \mathcal{B}$, any $(\epsilon, \delta)-$coverage, uniform randomized stable policy with an almost surely finite stopping time τ_δ, we have*

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\ell^{\text{unif}}(\nu_A, \nu_B, \epsilon)},$$

where,

$$\ell^{\text{unif}}(\nu_A, \nu_B, \epsilon) = \sup_{(\Delta_L, \Delta_R) \in \Upsilon(\epsilon)} \min\{T(\nu_A, \nu_B, 0.5, \Delta_L), T(\nu_A, \nu_B, 0.5, \Delta_R)\}.$$

Further following holds,

$$\ell^*(\nu_A, \nu_B, \epsilon) \geq \ell^{\text{unif}}(\nu_A, \nu_B, \epsilon) \geq \frac{\ell^*(\nu_A, \nu_B, \epsilon)}{2}.$$
$$\lim_{\epsilon \to 0} \frac{\ell^*(\nu_A, \nu_B, \epsilon)}{\ell^{\text{unif}}(\nu_A, \nu_B, \epsilon)} = R\left(\frac{\sigma^2(\nu_A)}{\sigma^2(\nu_B)}\right),$$

where,

$$R(x) = \frac{2\left(\sqrt{x} + \frac{1}{\sqrt{x}}\right)}{\left(\sqrt{x} + \frac{1}{\sqrt{x}} + 2\right)}.$$

I.3. Asymptotically optimal $(\epsilon, \delta)-$coverage guarantee policies $\mathfrak{P}_1^{\text{bound}}$ and $\mathfrak{P}_2^{\text{bound}}$

Let $N_A(n)$ and $N_B(n) = n - N_A(n)$ represent the number of times treatment A and B have been chosen for the first n assignments respectively under the policy $\mathfrak{P}_1^{\text{bound}}$. Recall $U_n \in \{A, B\}$ denotes the assignment of treatment A or B for the individual arriving at time n and X_n denotes the outcome of the individual arriving at time n once, treatment U_n was assigned. Let $\hat{\nu}_A(n)$ denote the empirical distribution corresponding to $N_A(n)$ samples from the outcomes of the treatment A by time n. Similar for the treatment B, we define $\hat{\nu}_B(n)$. It is worth noting that for ν_A and $\nu_B \in \mathcal{B}$, empirical distributions $\hat{\nu}_A(n)$ and $\hat{\nu}_B(n)$ will also in $\mathcal{B}$.

The estimation of the CI for ATE under the policy $\mathfrak{P}_1^{\text{bound}}$ is denoted by $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$, and the stopping rule of policy $\mathfrak{P}_1^{\text{bound}}$ is denoted by τ_δ for a given δ. All three components of $\mathfrak{P}_1^{\text{bound}}$ is given by,

1. **Assignment Rule:** For the assignment rule, we use the randomized tracking rule stated in Chapter 5 in Agrawal et al. (2020). Similar to the assignment rule of $\mathfrak{P}_1$, the key idea is to track the solution of the lower bound optimization problem to estimate the asymptotically optimal fraction of treatments and some forced exploration. We write it here for completeness as given below.
   (a) Initialize by assigning $m > 1$ samples in a round-robin way to generate at least $\lfloor (m/2) \rfloor$ samples from each treatment. Set $l = 1$ and let lm denote the total number of samples generated.
   (b) Check if the stopping criteria (discussed later in the stopping rule are met). If not, compute $w^*(\hat{\nu}_A(lm), \hat{\nu}_B(lm), \epsilon)$.
   (c) Compute starvation s_k for each treatment as $s_k = (\sqrt{(l+1)m} - N_k(lm))^+$.
   (d) if $m \geq \sum_{k \in A,B} s_k$, generate s_k samples from each treatment k. Specifically, first, generate s_A samples from treatment A, then s_B samples from treatment B. In addition, toss a coin, with head probability $w^*(\hat{\nu}_A(lm), \hat{\nu}_B(lm), \epsilon)$, $\max\{m - \sum_{k \in A,B} s_k, 0\}$ times independently. For each toss of the coin, generate a sample from treatment A if head comes up, otherwise from treatment B.
   (e) Else, if $\sum_{k \in A,B} s_k > m$ generate $\hat{s}_k$ samples from treatment $k \in \{A, B\}$, where $(\hat{s}_A, \hat{s}_B)$ are a solution to the load balancing problem: $\min(\max_k\{s_k - \hat{s}_k\})$ s.t. $s_k \geq \hat{s}_k \geq 0$ for $k \in \{A, B\}$, and $\sum_{k \in \{A,B\}} \hat{s}_k = m$. Again, first, generate $\hat{s}_A$ samples from treatment A, then $\hat{s}_B$ samples from treatment B.
   (f) Increment l by 1 and return to step (b).

2. **Estimation Rule:** Formally, we find $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ after n samples such that following holds,

$$\min_{x,y \in \mathcal{C}(\hat{\Delta}_L(n))} N_A(n) D_{inf}(\hat{\nu}_A(n), x) + N_B(n) D_{inf}(\hat{\nu}_B(n), y) \tag{74}$$
$$= \min_{x,y \in \mathcal{C}(\hat{\Delta}_R(n))} N_A(n) D_{inf}(\hat{\nu}_A(n), x) + N_B(n) D_{inf}(\hat{\nu}_B(n), y) = \beta(n, \delta).$$

Here $\beta(n,\delta) = \log(1/\delta) + 2\log(1+n) + 2$ (see Theorem 5.18 in Agrawal (2022)) is chosen to ensure that $[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$ is a $(1-\delta)-$confidence sequence of Δ. Similar to the policy $\mathfrak{P}_1$, one can show here as well that $\hat{\Delta}_L(n)$ and $\hat{\Delta}_R(n)$ uniquely solves the above equation.

3. **Stopping Rule:** Given that we have maintained the $(1-\delta)-$confidence sequence of Δ for ATE, we need to stop once CI width becomes less than ϵ, hence we define the stopping rule as follows:

$$\tau_\delta = \inf\{n \in \mathbb{N} : \hat{\Delta}_R(n) - \hat{\Delta}_L(n) \le \epsilon\}.$$

Theorem I.8. **($(\epsilon, \delta)-$coverage guarantee and stability of $\mathfrak{P}_1^{\text{bound}}$)** *For $\mathfrak{P}_1^{\text{bound}}$, there exists a $\epsilon_o > 0$ such that for $\epsilon \le \epsilon_o$, we have:*

a) *For a given $\delta \in (0,1)$, τ_δ is finite almost surely.*

b) *$[\hat{\Delta}_L(n), \hat{\Delta}_R(n)]$ is a $(1-\delta)-$confidence sequence of Δ. This in turn implies that $\mathfrak{P}_1^{\text{bound}}$ has the $(\epsilon,\delta)-$coverage guarantee.*

c) *$\mathfrak{P}_1^{\text{bound}}$ is a stable policy.*

Remark I.9. To get asymptotic optimality of $\mathfrak{P}_1^{\text{bound}}$ we assume that for a given $w \in (0,1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, $\min\{T(\nu_A,\nu_B,w,\Delta_L), T(\nu_A,\nu_B,w,\Delta_L+\epsilon)\}$ is jointly strictly quasi-concave function in w and Δ_L. Later in Remark J.1, we numerically verify this assumption. This assumption implies that for a given $\nu_A \in \mathcal{B}$, $\nu_B \in \mathcal{B}$ and $\epsilon > 0$, $w^*(\nu_A,\nu_B,\epsilon)$, $\Delta_L^*(\nu_A,\nu_B,\epsilon)$ and $\Delta_R^*(\nu_A,\nu_B,\epsilon)$ are unique.

Theorem I.10. **(Asymptotic optimality of $\mathfrak{P}_1^{\text{bound}}$)** *For $\mathfrak{P}_1^{\text{bound}}$, there exists a $\epsilon_o > 0$ such that for $\epsilon \le \epsilon_o$, we have:*

$$\mathbb{P}\left(\limsup_{\delta\to 0} \frac{\tau_\delta}{\log(1/\delta)} = \frac{1}{\ell^*(\nu_A,\nu_B,\epsilon)}\right) = 1 \textit{ and } \lim_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} = \frac{1}{\ell^*(\nu_A,\nu_B,\epsilon)}.$$

In the assignment rule of $\mathfrak{P}_1^{bounded}$, replace $w^*(\hat{\nu}_A(n),\hat{\nu}_B(n),\epsilon)$ with $\overline{w}^b(\sigma(\hat{\nu}_A(n)),\sigma(\hat{\nu}_B(n))) = \frac{\sigma(\hat{\nu}_A(n))}{\sigma(\hat{\nu}_A(n))+\sigma(\hat{\nu}_B(n))}$ and keep stopping and estimation rule same. We refer this policy as $\mathfrak{P}_2^{bounded}$. Here $\sigma(\hat{\nu}_A(n))$ denotes the standard deviation of the empirical distribution of outcomes generated by treatment A by time n.

Corollary I.11. **(Asymptotic optimality of policy $\mathfrak{P}_2^{\text{bound}}$)** *For $\mathfrak{P}_1^{\text{bound}}$, there exists a $\epsilon_o > 0$ such that for $\epsilon \le \epsilon_o$, we have:*

$$\lim_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} = \frac{1}{\ell^{\mathfrak{P}_2}(\nu_A,\nu_B,\epsilon)},$$

where,

$$\ell^{\mathfrak{P}_2}(\nu_A,\nu_B,\epsilon) = \sup_{(\Delta_L,\Delta_R)\in\Upsilon(\epsilon)} \min\{T(\nu_A,\nu_B,\overline{w}(\nu_A,\nu_B),\Delta_L), T(\nu_A,\nu_B,\overline{w}^b(\nu_A,\nu_B),\Delta_R)\},$$

where,

$$\overline{w}(\nu_A,\nu_B) = \frac{\sigma(\nu_A)}{\sigma(\nu_A)+\sigma(\nu_B)}.$$

Further following holds as well,

$$\lim_{\epsilon\to 0}\lim_{\delta\to 0} \frac{\epsilon^2\mathbb{E}_\nu[\tau_{\delta,\epsilon}]}{\log(1/\delta)} = \lim_{\epsilon\to 0} \frac{\epsilon^2}{\ell^{\mathfrak{P}_2}(\nu_A,\nu_B,\epsilon)} = \lim_{\epsilon\to 0} \frac{\epsilon^2}{\ell^*(\nu_A,\nu_B,\epsilon)}. \tag{75}$$

Since we are taking double limit of $\epsilon \to 0$ then $\delta \to 0$ in (75), hence we have indexed τ with both δ and ϵ.

Remark I.12. Our policy $\mathfrak{P}_2^{\text{bound}}$, similar to the policy $\mathfrak{P}_1^{\text{bound}}$, is a stable policy and constructs a $(1-\delta)-$confidence sequence of ATE. Hence it also follows that $\mathfrak{P}_2^{\text{bound}}$ has $(\epsilon,\delta)-$coverage guarantee. This implies,Theorem I.10 holds for $\mathfrak{P}_2^{\text{bound}}$ as well.

J. Proofs and supporting material related to Section I.

Remark J.1. Here we numerically show that for given $w \in (0,1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, $\min\{T(\nu_A, \nu_B, w, \Delta_L), T(\nu_A, \nu_B, w, \Delta_L + \epsilon)\}$ is jointly strictly quasi-concave in w and Δ_L via plotting upper contour plots when outcome distributions are beta and Bernoulli. For Bernoulli outcome distributions, one can show that $D_{\text{inf}}(\nu, x) = d(m(\nu), m(\nu'))$ such that ν' is a Bernoulli distribution with mean $m(\nu') = x$. Using Figure 6, we observe that $T(\nu_A, \nu_B, w, \Delta_L)$ is a strictly quasi concave function in (w, Δ_L).

Now we numerically show that $T(\nu_A, \nu_B, w, \Delta_L)$ is a strictly quasi concave function in (w, Δ_L) when outcome distributions are beta in Figure 7. Similarly, one can verify numerically that $w \in (0,1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, $T(\nu_A, \nu_B, w, \Delta_L + \epsilon)$ is a jointly strictly quasi-concave in w and Δ_L. This further implies that, $\min\{T(\nu_A, \nu_B, w, \Delta_L), T(\nu_A, \nu_B, w, \Delta_L + \epsilon)\}$ is a jointly strictly quasi-concave in w and Δ_L.

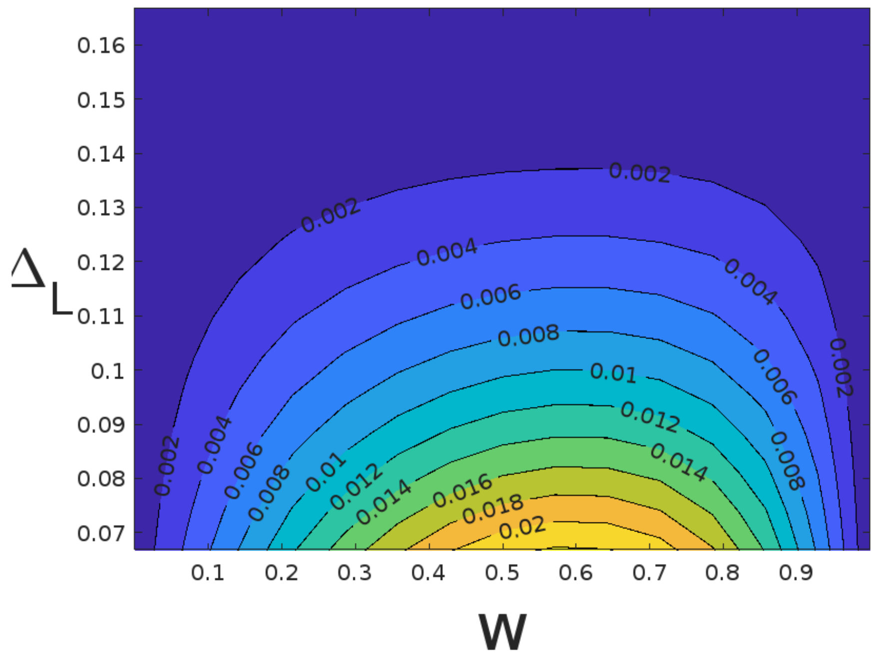

Figure 7: Upper contour plots for function $T(\nu_A, \nu_B, w, \Delta_L)$ in the space of (w, Δ_L) when outcomes follow beta distributions. ν_A is a beta distribution with shape parameters $(1, 1)$ and ν_B is a beta distribution with shape parameters $(2, 4)$. It follows that $m(\nu_A) = 0.5$, $m(\nu_B) = 1/3 = 0.333$ and hence $\Delta = 0.167$. Here we choose, $\epsilon = 0.1$, hence $w \in (0, 1)$ and $\Delta_L \in (\Delta - \epsilon, \Delta)$, i.e., $\Delta_L \in (0.067, 0.167)$. There are various lines in the above figure. The number on a given line in the above figure represents the value of $T(\nu_A, \nu_B, w, \Delta_L)$ for all values of (w, Δ_L) satisfying the trajectory of that given line for given ν_A and ν_B. One can observe that all the upper contour plots are strictly convex.

Proof of Lemma I.3:

First, observe that from the definition of $D_{inf}(\nu, x)$ for $x \in (m(\nu), 1)$ and $\nu \in \mathcal{B}$,

$$D_{inf}(\nu, x) = D^U_{inf}(\nu, x).$$

Similarly, for $x \in (0, m(\nu))$ and $\nu \in \mathcal{B}$,

$$D_{inf}(\nu, x) = D^L_{inf}(\nu, x).$$

Using the property number 6 mentioned above, we get that at $x = m(\nu)$, $\lambda^U(\nu, x) = 0$. Similarly, we will have at $x = m(\nu)$, $\lambda^L(\nu, x) = 0$. Hence using properties number 7 and 8, we get that $D_{inf}(\nu, x)$ is differentiable in x for $x \in (0, 1)$. Further,

$$\frac{\partial D_{inf}(\nu, x)}{\partial x} = \lambda^U(\nu, x), \tag{76}$$

for $x \in [m(\nu), 1)$ and

$$\frac{\partial D_{inf}(\nu, x)}{\partial x} = -\lambda^L(\nu, x), \tag{77}$$

for $x \in (0, m(\nu)]$.

Since we are interested in proving thrice differentiability of $D_{inf}(\nu, x)$ with respect to x for x near $m(\nu)$. Hence we are interested in differentiability of $\lambda^U(\nu, x)$ and $\lambda^L(\nu, x)$ for x near $m(\nu)$. Using the strict convexity of function $\frac{1}{1-x}$ for $x \in (0, 1)$, we get for $\nu \in \mathcal{B}$,

$$\mathbb{E}_{X\sim\nu}\left[\frac{1-x}{1-X}\right] > 1,$$

for $x = m(\nu)$. Using the continuity of $1 - x$ in x, we get that there exists a $\xi_1 > 0$ such that

$$\mathbb{E}_{X\sim\nu}\left[\frac{1-x}{1-X}\right] > 1, \tag{78}$$

for $x \in [m(\nu), m(\nu) + \xi_1]$. It is worth noticing that if $\mathbb{P}_{X\sim\nu}(X = 1) > 0$, then we define $\mathbb{E}_{X\sim\nu}\left[\frac{1}{1-X}\right] = \infty$. Since we have established the fact that (78) holds for $x \in [m(\nu), m(\nu) + \xi_1]$, hence we get from property number 6, that $\lambda^U(\nu, x)$ uniquely satisfy the following equation for $x \in [m(\nu), m(\nu) + \xi_1]$ since we are not choosing ν to be a point mass distribution.

$$\mathbb{E}_{X\sim\nu}\left[\frac{(X-x)}{1-(X-x)\lambda^U(\nu, x)}\right] = 0. \tag{79}$$

Above equation also implies that for $x \in [m(\nu), m(\nu) + \xi_1]$, $\lambda^U(\nu, x)$ satisfies $\mathbb{E}_{X\sim\nu}\left[\frac{1}{1-(X-x)\lambda^U(\nu,x)}\right] = 1$. Using (78), we get that for $\lambda^U(\nu, x) \neq 1/(1-x)$ for $x \in [m(\nu), m(\nu) + \xi_1]$. Hence using property number 9, we get $\lambda^U(\nu, x)$ is continuous in x for $x \in [m(\nu), m(\nu) + \xi_1]$. Let $\lambda^m = \max_{x\in[m(\nu),m(\nu)+\xi_1]} \lambda^U(\nu, x) > 0$.

To find the derivative of $\lambda^U(\nu, x)$ with respect to x for $x \in (m(\nu), m(\nu) + \xi_1)$, we will use implicit function. Observe that for a given ν, (79) can be re-written as,

$$F(\lambda^U(\nu, x), x) = 0,$$

where, $F(\lambda^U, x) = \mathbb{E}_{X\sim\nu}\left[\frac{(X-x)}{1-(X-x)\lambda^U}\right]$.

Let $B_1 = \{(\lambda, x) : x \in [m(\nu), m(\nu) + \xi_1], \lambda \in [0, \lambda^m], \lambda \in [0, 1/(1-x)\}$. It follows for $x \in [m(\nu), m(\nu) + \xi_1]$, $(x, \lambda^U(\nu, x)) \in B_1$. Since ξ_1 can be chosen small enough such that $\lambda^m < 1/(1 - m(\nu))$, this implies that for all $x \in [m(\nu), m(\nu) + \xi_1]$, $\lambda^m < 1/(1 - x)$. Further this implies that B_1 set is a closed interval.

Let $G(X, \lambda, x) \triangleq \frac{(X-x)}{1-(X-x)\lambda}$. Hence we have,

$$\frac{\partial G(X, \lambda, x)}{\partial x} = \frac{(X-x)^2}{(1-\lambda(X-x))^2}, \frac{\partial G(X, \lambda, x)}{\partial \lambda} = \frac{-1}{(1-\lambda(X-x))^2}.$$

For $X \in [0, 1]$, $(x, \lambda) \in B_1$, since $1-(X-x)\lambda > 0$. This implies that $G(X, \lambda, x)$, $\frac{\partial G(X,\lambda,x)}{\partial x}$ and $\frac{\partial G(X,\lambda,x)}{\partial \lambda}$ are continuous functions of (X, λ, x) for $X \in [0, 1]$ and $(\lambda, x) \in B_1$. It also follows that $G(X, \lambda, x)$ is jointly continuous differentiable in λ and x for $X \in [0, 1]$, $(x, \lambda) \in B_1$.

Using Leibniz rule, we get the differentiability of $F(\lambda^U, x)$ in λ^U and x for $(\lambda^U, x) \in B_1$. Using the Bounded convergence theorem, we get that $\frac{\partial F(\lambda^U,x)}{\partial x}$ and $\frac{\partial F(\lambda^U,x)}{\partial \lambda}$ are continuous in (λ^U, x) for $(\lambda^U, x) \in B_1$. Since continuous partial derivative implies the continuous differentiability in (λ, x) jointly, we get that, $F(\lambda^U, x)$ in continuous differentiable in (λ^U, x) for $(\lambda^U, x) \in B_1$. Hence using the implicit function theorem and the fact that $\lambda^U(\nu, x)$ uniquely satisfies $F(\lambda^U(\nu, x), x) = 0$ for $x \in [m(\nu), m(\nu) + \xi]$, we get,

$$\frac{\partial \lambda^U(\nu, x)}{\partial x} = \frac{\mathbb{E}_{X\sim\nu}\left(\frac{1}{(1-\lambda^U(\nu,x)(X-x))^2}\right)}{\mathbb{E}_{X\sim\nu}\left(\frac{(X-x)^2}{(1-\lambda^U(\nu,x)(X-x))^2}\right)}. \tag{80}$$

Since we know that for $(\lambda^U, x) \in B_1$, we have $1 - \lambda^U(X - x) > 0$ for all $X \in [0,1]$. Hence it follows that $\mathbb{E}_{X\sim\nu}\left(\frac{(X-x)^2}{(1-\lambda^U(\nu,x)(X-x))^2}\right) > 0$. This also gives us the differentiability of $\lambda^U(\nu, x)$ for $x \in (m(\nu), m(\nu) + \xi_2)$ and right hand derivative at $x = m(\nu)$.

It also follows that, at $x = m(\nu)$, since $\lambda^U(\nu, m(\nu)) = 0$, right hand derivative of $\lambda^U(\nu, x)$ with respect to x is $\frac{1}{\sigma^2(\nu)}$. Similar analysis will follow for the derivative of $-\lambda^L(\nu, x)$ for $x \in [m(\nu)\xi_2, m(\nu)]$, and we get the left hand derivative of $-\lambda^L(\nu, x)$ with respect to x is $\frac{1}{\sigma^2(\nu)}$ at $x = m(\nu)$. Here ξ_2 is a well-chosen small positive number. Choosing $\eta = \min\{\xi_1, \xi_2\}$ and using (76) and (77), we get that, for $x \in (m(\nu) - \eta), m(\nu) + \eta)$, $D_{inf}(\nu, x)$ is twice differentiable and $\frac{\partial^2 D_{inf}(\nu,x)}{\partial x^2}\Big|_{x=m(\nu)} = \frac{1}{\sigma^2(\nu)}$.

Again applying the Bounded convergence theorem on (80) and similarly for $\lambda^L(\nu, x)$, we get that $x \in (m(\nu)-\eta), m(\nu)+\eta)$, $D_{inf}(\nu, x)$ is twice continuous differentiable in x. One can similarly get the higher order differentiability of $D_{inf}(\nu, x)$ in x for $x \in (m(\nu) - \eta), m(\nu) + \eta)$ as well.

A similar proof will follow for $\lambda^L(\nu, x)$). This completes the proof.

□

Proof of Theorem I.4: Using the proof similar to the Theorem 4.2, we get,

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} \geq \frac{1}{\ell^*(\nu_A, \nu_B, \epsilon)},$$

where,

$$\ell^*(\nu_A, \nu_B, \epsilon) \triangleq \sup_{\substack{w\in[0,1],\\ (\Delta_L,\Delta_R)\in\Upsilon(\epsilon)}} \inf_{\nu'\in\mathcal{K}(\Delta_L,\Delta_R)} wKL(\nu_A, \nu'_A) + (1-w)KL(\nu_B, \nu'_B).$$

where, $\mathcal{K}^b(\Delta_L, \Delta_R) = \mathcal{K}^b_1(\Delta_L) \cup \mathcal{K}^b_2(\Delta_R)$, and $\mathcal{K}^b_1(\Delta_L) = \{(\nu'_A, \nu'_B) : \nu'_A \in \mathcal{B}, \nu'_B \in \mathcal{B}, m(\nu'_A) - m(\nu'_B) < \Delta_L\}$ and $\mathcal{K}^b_2(\Delta_R) = \{(\nu'_A, \nu'_B) : \nu'_A \in \mathcal{B}, \nu'_B \in \mathcal{B}, m(\nu'_A) - m(\nu'_B) > \Delta_R\}$.

Using definition of $D^L_{inf}(\nu, x)$ and $D^U_{inf}(\nu, x)$ and some algebra, we get that for $w \in [0,1]$ and $(\Delta_L, \Delta_R) \in \Upsilon(\epsilon)$, we have,

$$\inf_{\nu'\in\mathcal{K}^b_1(\Delta_L)} wKL(\nu_A, \nu'_A) + (1-w)KL(\nu_B, \nu'_B) = \min_{x,y\in\mathcal{C}^b(\Delta_L)} wD^L_{inf}(\nu_A, x) + (1-w)D^U_{inf}(\nu_B, y).$$

Using the definition of $D_{inf}(\nu, x)$, we get,

$$\min_{x,y\in\mathcal{C}^b(\Delta_L)} wD^L_{inf}(\nu_A, x) + (1-w)D^U_{inf}(\nu_B, y) = \min_{x,y\in\mathcal{C}^b(\Delta_L)} wD_{inf}(\nu_A, x) + (1-w)D_{inf}(\nu_B, y).$$

Similarly, we get, for $w \in [0,1]$ and $(\Delta_L, \Delta_R) \in \Upsilon(\epsilon)$, we have,

$$\inf_{\nu'\in\mathcal{K}^b_2(\Delta_R)} wKL(\nu_A, \nu'_A) + (1-w)KL(\nu_B, \nu'_B) = \min_{x,y\in\mathcal{C}^b(\Delta_R)} wD_{inf}(\nu_A, x) + (1-w)D_{inf}(\nu_B, y).$$

This completes the proof.

□

Remark J.2. Theorem I.5, Proposition I.7, Theorem I.8, Theorem I.10 and Corollary I.11 are generalizations of result present in main sections of this paper where we assume that outcome distributions lie in canonical SPEF. In all of these results, we have replaced $D_{inf}(\nu, x)$ instead of $d(\mu, x)$. It follows from Appendix D, Lemma E.1, Appendix I.1 and Lemma I.3, $D_{inf}(\nu, x)$ inherits most of the properties of $d(\mu, x)$ such as continuity in (ν, x), strict convexity in x with $D_{inf}(\nu, m(\nu)) = 0$. Now we mention few points which helps us in understanding that proofs of this Appendix can be generalized from the main sections of this paper.

1. Firstly, it is worth noting that, we have continuity of $D_{inf}(\nu, x)$ in the space of probability distributions. Hence we need to study the properties of the functions defined on the space of probability measures of bounded support. Hence we endow $\mathcal{B}$, with the topology of weak convergence. We denote the weak convergence of sequence ν_n in $\mathcal{B}$ to ν by $\nu_n \Rightarrow \nu$. This convergence is equivalent to convergence in the Lévy metric on $\mathcal{B}$, denoted by d_L, defined below.

Definition J.3. For $\nu_1, \nu_2 \in \mathcal{B}$,

$$d_L(\nu_1, \nu_2) = \inf\{\eta > 0 : F_{\nu_1}(x - \eta) \leq F_{\nu_2}(x) \leq F_{\nu_1}(x + \eta) + \eta, \forall x \in \mathfrak{R}\},$$

where $F_\nu(x)$ denotes the CDF of the distribution ν.

2. It is worth noting that, we have $D_{inf}(\nu, x)$ is four times continuously differentiable in x for a neighbourhood around $x = m(\nu)$, which is what we need to extend our results in this Appendix (see Lemma I.3).

   To show the above, one can use the continuity of $D_{inf}(\nu, x)$ in (ν, x) and the property numbers 2 and 3 in Appendix I.1.

3. It is worth noting that we have used the randomized tracking rule introduced in Chapter 5 in Agrawal et al. (2020), in our policies $\mathfrak{P}_1^{\text{bound}}$ and $\mathfrak{P}_2^{\text{bound}}$. Hence to prove the Theorem I.10 and Corollary I.11, we need to replace Lemma E.8, Lemma E.9 and Lemma E.10. We instead use Lemma 5.13, Lemma 5.14 and the proof of Theorem 5.15 of Chapter 5 in Agrawal et al. (2020), to define the set $\mathcal{G}_n(\eta)$ and to bound the probability $\mathbb{P}(\mathcal{G}_n(\eta)^c)$ for the policies $\mathfrak{P}_1^{\text{bound}}$ and $\mathfrak{P}_2^{\text{bound}}$.

4. Recall that Theorem 6.1 requires continuity of $I(\mu)$ or $\sigma^2(\mu)$. To get the result for Corollary I.11, let's take a sequence of outcome distributions of treatment A denoted as $\nu_{A,n} \in \mathcal{B}$ such that $\nu_{A,n} \Rightarrow \nu_A^*$ where $\nu_A^* \in \mathcal{B}$. Using Skorohod's Theorem and bounded convergence theorem, we can show that,

$$\lim_{n \to \infty} m(\nu_{A,n}) = m(\nu_A^*) \text{ and } \sigma(\nu_{A,n}) = \sigma(\nu_A^*).$$

   Hence it follows that we have continuity of $\sigma(\nu)$ for $\nu \in \mathcal{B}$ as well which is needed in the Corollary I.11.

$\square$