Logarithmic Neyman Regret for Adaptive Estimation of the Average Treatment Effect

Ojash Neopane
Machine Learning Department
Carnegie Mellon University
oneopane@andrew.cmu.edu

Aaditya Ramdas
Department of Statistics and Data Science
Machine Learning Department
Carnegie Mellon University
aramdas@cmu.edu

Aarti Singh
Machine Learning Department
Carnegie Mellon University
aartisingh@cmu.edu

Abstract

Estimation of the Average Treatment Effect (ATE) is a core problem in causal inference with strong connections to Off-Policy Evaluation in Reinforcement Learning. This paper considers the problem of adaptively selecting the treatment allocation probability in order to improve estimation of the ATE. The majority of prior work on adaptive ATE estimation focus on asymptotic guarantees, and in turn overlooks important practical considerations such as the difficulty of learning the optimal treatment allocation as well as hyper-parameter selection. Existing non-asymptotic methods are limited by poor empirical performance and exponential dependence on problem parameters. To address these gaps, we propose and analyze the Clipped Second Moment Tracking (ClipSMT) algorithm, a variant of an existing algorithm with strong asymptotic optimality guarantees, and provide finite sample bounds on its Neyman regret. Our analysis shows that ClipSMT achieves exponential improvements in Neyman regret over on two fronts: improving the dependence on T from $O(\sqrt{T})$ to $O(\log T)$, as well as reducing the exponential dependence on problem parameters to a polynomial dependence—although the setting we consider is slightly less general. We conclude with simulations demonstrating strong improvement of ClipSMT over existing approaches.

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

1 INTRODUCTION

Randomized Controlled Trials (RCTs) have long been considered the gold standard of evidence in a variety of disciplines, ranging from medicine (Hollis and Campbell, 1999), policy research (Wing et al., 2018), and economics (Banerjee et al., 2016). In their simplest form, RCTs involve a control arm and a treatment arm, and the objective is to determine if the treatment *causally* outperforms the control. This is typically achieved by fixing a treatment assignment probability (hereafter called an *allocation*), assigning experimental units to an arm, and using the resulting outcomes to estimate the Average Treatment Effect(ATE).

Despite the ubiquity of RCTs, many practitioners have noted that RCTs would benefit from the use of *adaptive* methods—methods in which practitioners vary some aspect of the experiment through the course of the experiment (Chow et al., 2005; Chow and Chang, 2008; US Food and Drug Administration, 2019). Although there are many reasons for desiring adaptivity, our primary focus is to adaptively select the treatment allocation probability in order to obtain the best possible estimate of the ATE. More concretely, our goal will be to minimize the MSE of our ATE estimate[1] This is the essence of the problem known as Adaptive Neyman Allocation (Dai et al., 2023) and is the primary focus of this work.

Despite the recent attention given to adaptive approaches, considerable work remains to ensure their success in practice. This is because a significant portion of prior work on this topic has focused on developing algorithms with strong asymptotic guarantees. In this asymptotic regime, much is known, such

[1] In general, one may wish to minimize the mean squared error of the ATE estimate. Since our work focuses on estimation using the unbiased Horvitz-Thompson estimator, this is equivalent to minimizing the variance.

as the semiparametric efficiency bound (Bickel et al., 1993; Kallus and Uehara, 2019) for non-adaptive approaches, as well as adaptive procedures which asymptotically match the performance of the *best possible* non-adaptive approach (Kato et al., 2020). While these results provide a solid foundation, their asymptotic nature overlooks many nuances crucial for practical application. At a high level, prior asymptotic approaches aim to identify the (unknown) variance-minimizing allocation and demonstrate that their allocation converges to this allocation. However, they do not adequately address the challenges of *efficiently learning* this allocation, which is often vital for practical implementation (Wagenmaker and Foster, 2023).

In order to address these subtleties, we believe a nonasymptotic analysis is required. Unfortunately, such analyses are currently scarce. The only work we are aware of which provides a nonasymptotic analysis is Dai et al. (2023) who propose the ClipOGD algorithm and show it attains $O(\sqrt{T})$ Neyman regret—a new measure of performance which we formally introduce in Section 3. Despite offering a promising starting point, this work has several limitations. As we further expand in Sections 2 and 4, ClipOGD can demonstrate poor empirical performance; this is explained by the exponential scaling of their bounds with respect to various problem parameters which they treat as constants.

In this paper, we advance the understanding of adaptive estimation procedures for the ATE by providing a finite sample analysis of the Clipped Second Moment Tracking algorithm, a variant of the procedure proposed in Cook et al. (2024), tailored for the Horvitz-Thompson estimator. Our analysis meticulously addresses various problem-specific parameters, demonstrating an exponential improvement with respect to problem parameters. We also establish a $O(\log T)$ bound on Neyman regret, representing another significant improvement over ClipOGD, although Dai et al. (2023) consider the more challenging fixed design setting, while we work in the superpopulation setting defined in Section 3. Additionally, our finite sample analysis also highlights some aspects of algorithm design that were previously unaddressed.

2 Related Works

Adaptive experimental design has a long and rich history, with work dating back as far as Neyman (1934), who introduced the notion of an optimal allocation in the context of experimental designs. Its significance has been emphasized by prominent researchers like Robbins (1952), who identified adaptive sampling as a crucial statistical problem. Building on this, Thompson (1933) proposed a Bayesian adaptive design, which later sparked significant interest in the Multi-Armed Bandit (MAB) problem. Thompson's Bayesian framework introduced the idea of sequentially updating beliefs about the arms (or treatments) based on observed outcomes, which became a foundational concept in MAB research (Lattimore and Szepesvári, 2020). However, the typical MAB formulations studied in the literature often focus on maximizing cumulative rewards across multiple rounds of exploration and exploitation, whereas our approach focuses on a different objective or setting, diverging from the standard MAB treatment.

Our work builds on a recent line of research focused on developing adaptive algorithms to improve the efficiency of ATE estimation. This line of work begins with Hahn et al. (2009), who propose a two-stage design similar to Explore-then-Commit style algorithms (Garivier et al., 2016) found in the MAB literature. Building on these ideas, Kato et al. (2020) propose a fully adaptive design which we term Regularized Variance Tracking. They demonstrate that this approach asymptotically attains the minimum-variance Semiparametric Efficiency (SPE) bound[2]. In a follow-up work, Cook et al. (2024) propose the Clipped Variance Tracking algorithm, which provides multiple benefits compared to the Regularized Variance Tracking algorithm. Specifically, Cook et al. (2024) show that their procedure attains the same minimum variance SPE bound under milder assumptions, can be used with modern uncertainty quantification techniques (Waudby-Smith et al., 2022), and has superior empirical performance. Finally, in an independent line of work, Li et al. (2024) demonstrates that a generalization of the two-stage approach from Hahn et al. (2009) can be applied to significantly more general problems including Non-Markov Decision Processes. They show that their approach also obtains the minimum-variance SPE bound.

The main drawback of these prior works is that they only give an asymptotic characterization of their respective procedures, thus painting an incomplete picture of the problem. For instance, the asymptotic analysis presented in Cook et al. (2024) fails to address the crucial aspect of tuning their algorithm's parameters, an issue that our finite sample analysis clarifies. These issues in conjunction with the broad asymptotic optimality of 'tracking' style algorithms jointly highlight the pressing need for a nonasymptotic analysis.

[2]By 'minimum-variance' SPE bound, we mean minimizing the asymptotic variance jointly over both the estimators and the treatment allocation. The standard SPE bound, in contrast, only considers the minimal asymptotic variance for a fixed treatment allocation.

When compared to asymptotic analyses, research on characterizing the nonasymptotic performance of adaptive algorithms is sparse; we are aware only of the recent work by Dai et al. (2023), who introduced the ClipOGD algorithm. Dai et al. (2023) analyzes the problem in the *fixed-design* setting and provides bounds on a scaled proxy to the variance of the resulting ATE estimate, termed the Neyman regret. While this work represents a significant first step, the performance of ClipOGD scales exponentially with respect to problem parameters such as the optimal allocation. Additionally, they consider the more challenging fixed-design setting, resulting in pessimistic bounds when assumptions about the underlying data-generating process, such as in the superpopulation setting, are reasonable. Our results address these scaling issues and highlight the significant improvements achievable in the superpopulation setting.

The problem of Off-Policy Evaluation (OPE), a generalization of ATE estimation[3], has also been thoroughly studied within the Reinforcement Learning literature (Dudík et al., 2011; Li et al., 2011; Jiang and Li, 2016). Here, the primary focus has been on offline estimation, and there has been extensive work culminating in precise characterizations of minimax lower bounds as well as matching upper bounds (Li et al., 2015; Wang et al., 2017; Duan et al., 2020; Ma et al., 2021). These ideas have been further generalized to estimating parameters other than the performance of a policy, such as the estimation of the Cumulative Distribution Function of the rewards (Huang et al., 2021, 2022). There is less work on studying adaptive versions of these methods; the only works we are aware of are Hanna et al. (2017) who focus on the problem of off-policy learning, and Konyushova et al. (2021) who combine offline OPE methods with an online data acquisition strategy to improve the sample efficiency of policy selection, though these works are primarily empirical.

Tangentially related to our work is a line of research on developing inference procedures using adaptively collected data. These works can be split into asymptotic and non-asymptotic approaches. On the asymptotic side, one line of work has focused on developing new estimators via re-weighting and demonstrating asymptotic normality (Hadad et al., 2021; Zhang et al., 2020, 2021). Another line of work eschews asymptotic results in favor of nonasymptotic results by utilizing modern martingale techniques to develop nonasymptotic confidence intervals and sequences for adaptively collected data, including quantities like the ATE (Howard et al., 2021; Waudby-Smith and Ramdas, 2023; Waudby-Smith et al., 2022).

To summarize, while significant progress has been made in the field of adaptive experimental design and related areas, there remain critical gaps, particularly in understanding the non-asymptotic performance of these methods. Our work aims to fill these gaps by providing a finite sample analysis that clarifies some aspects of algorithm design and serves as a starting point for analyzing the non-asymptotic behavior of more complicated algorithms.

3 PRELIMINARIES

Problem Setup. We consider the following interaction between an algorithm, Alg, and a problem instance, ν. At the start of each round t, Alg selects a treatment allocation, $\pi_t \in [0,1]$, based on the history of past observations $\mathcal{H}_{t-1} = \{(\pi_s, A_s, R_s)\}_{s=1}^{t-1}$. Then, the next experimental unit is assigned to either the control ($A_t = 0$) or the treatment ($A_t = 1$) arm by sampling $A_t \sim \text{Bernoulli}(\pi_t)$. Following this assignment, an outcome $R_t \in [0,1]$ is observed, marking the end of the round.

We formalize this interaction protocol as follows. First, we let $\mathcal{F}_t = \sigma(\mathcal{H}_t)$ denote the filtration generated by past observations. Then an algorithm $\text{Alg} = (\text{Alg}_t)$ is a sequence of $\mathcal{F}_{t-1}$ measurable mappings, $\text{Alg}_t : \mathcal{H}_{t-1} \to \mathcal{S}(\{0,1\})$, where $\mathcal{S}(\mathcal{X})$ is the set of distributions over $\mathcal{X}$. A problem instance $\nu : \{0,1\} \to \mathcal{S}([0,1])$ is a probability kernel which maps each arm to a distribution over outcomes which we assume to be bounded in the interval $[0,1]$. Finally, we let $R_t = \mathbb{I}[A_t = 0]R_t(0) + \mathbb{I}[A_t = 1]R_t(1)$, where $\mathbb{I}[\cdot]$ denotes the indicator function, and $R_t(A) \sim \nu(A)$ are called the *potential outcomes*. Within the causal inference literature, this framework is typically referred to as the superpopulation we potential outcomes framework (Neyman, 1923; Rubin, 1980; Imbens and Rubin, 2015).

Implicit in the above interaction protocol are the following assumptions:

1. *Bounded Observations:* We assume $R_t \in [0,1]$ almost surely.
2. *Stable Unit Treatment Value Assumption:* We assume that $R_t(A)$ is independent of $R_s(A)$.
3. *Unconfoundedness:* Given the history $\mathcal{H}_{t-1}$, we assume the treatment assignment A_t is independent of the potential outcomes $R_t(0)$ and $R_t(1)$. Formally, $R_t(A) \perp A_t \mid \mathcal{H}_{t-1}$ for $A \in \{0,1\}$.

[3] OPE is concerned with estimating the performance of a single policy, whereas ATE estimation can be thought of as estimating the difference in performance between two specific policies. However, the techniques developed in the OPE literature can be modified to estimate the difference in performance between two policies

While the second and third assumptions are commonplace in the causal inference literature and necessary for identification, the first assumption warrants a brief discussion. We make this assumption so that our methods are compatible with a recent line of work aimed at developing variance-adaptive sequential hypothesis tests (Karampatziakis et al., 2021; Waudby-Smith et al., 2022; Cook et al., 2024) where it is currently not known how to construct such tests without assuming bounded observations. However, our analysis and results can be easily modified to accommodate any class of distributions which guarantee concentration of the uncentered second moment. As we will discuss, this differs from existing work which assumes upper and lower bounds on the raw second moments. Indeed, our results don't treat any problem parameters as constant, and we take explicit care to understand scaling with respect to all problem parameters.

Efficient Estimation of the ATE. This work is concerned with designing algorithms to efficiently estimate the ATE, which we define as

$$\Delta = \mathbb{E}_{\nu}\left[R(1) - R(0)\right].$$

Efficient estimation of the ATE is roughly equivalent to minimizing the variance of an estimate of the ATE. This is because most standard estimates of the ATE are unbiased so the Mean Square Error (MSE) and widths of confidence intervals scale with the variance. We consider a variant of the standard Horvitz-Thompson (HT) estimator (Imbens and Rubin, 2015). For a fixed allocation, π, the HT estimator is defined as

$$\widehat{\Delta}_T^{(HT)}(\pi) = \frac{1}{T}\sum_{t=1}^{T} R_t \cdot \left(\frac{\mathbb{I}\left[A_t = 1\right]}{\pi} - \frac{\mathbb{I}\left[A_t = 0\right]}{1-\pi}\right). \tag{1}$$

It is well known that the HT estimator is unbiased, and a simple computation shows that its variance is

$$\mathbb{V}\left[\widehat{\Delta}_T^{(HT)}(\pi)\right] = \frac{1}{T}\left(\frac{m_1^2}{\pi} + \frac{m_0^2}{1-\pi} - \Delta^2\right), \tag{2}$$

where $m_A^2 = \mathbb{E}\left[R^2(A)\right]$ is the uncentered second moment of $\nu(A)$. The *Neyman allocation*, π_{Ney}, is the allocation that minimizes the variance of this estimate and is given by

$$\pi_{\text{Ney}} = \frac{m_1}{m_0 + m_1}. \tag{3}$$

Although π_{Ney} depends on unknown parameters, it can be estimated throughout an experiment, as we will do in this work.

Adaptive Estimation of the ATE. One issue with the standard HT estimator is that it requires knowledge of the marginal treatment assignment probability $\mathbb{P}(A_t = 1)$. For an adaptive algorithm, we do not know this probability *a priori* since it requires marginalizing over the interaction between Alg and the unknown problem instance ν. To remedy this issue, Kato et al. (2020) (and later Dai et al. (2023)) proposed the *adaptive Horvitz-Thompson* (aHT) estimator, defined as

$$\widehat{\Delta}_T = \frac{1}{T}\sum_{t=1}^{T} R_t \left(\frac{\mathbb{I}\left[A_t = 1\right]}{\pi_t} - \frac{\mathbb{I}\left[A_t = 0\right]}{1-\pi_t}\right). \tag{4}$$

The aHT is also an unbiased estimator of the ATE; this follows from the fact that π_t is $\mathcal{F}_{t-1}$-measurable in conjunction with the law of total expectation. In this work, we will aim to design Alg to minimize the variance of the aHT estimator.

The Neyman Regret. The preceding discussion has highlighted that our objective should be to minimize the variance of the aHT estimator. A straightforward calculation shows that

$$\mathbb{V}\left[\widehat{\Delta}_T\right] = \mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(f(\pi_t) - \Delta^2\right)\right], \tag{5}$$

where

$$f(\pi) = \frac{m_1^2}{\pi} + \frac{m_0^2}{1-\pi}$$

is the *Neyman loss*. One option to characterize the performance of Alg is to bound the variance of the resulting ATE estimate. However, such a quantity fails to normalize against the inherent difficulty of a problem instance — if the variance of an estimate using the Neyman allocation is large, then a bound on the variance of an adaptive procedure will paint a misleading picture of the performance of Alg. As such, in this work, we study the *Neyman regret*, $\mathfrak{N}_T$, recently introduced in Dai et al. (2023), and defined as

$$\mathfrak{N}_T = \sum_{t=1}^{T} f(\pi_t) - f(\pi_{\text{Ney}}). \tag{6}$$

Inspecting this quantity, we see that the Neyman regret is a scaled difference between the variance of the aHT estimator using an adaptive design and the variance of the HT estimator using the Neyman allocation. Therefore, demonstrating Alg obtains sublinear Neyman regret, we know that the variance of the resulting ATE estimate asymptotically approaches the minimum variance estimate.

4 ALGORITHM AND RESULTS

In this section, we introduce the Clipped Second Moment Tracking(ClipSMT) algorithm, state bounds on its Neyman regret, and compare its performance with existing algorithms. *To simplify our presentation and discussions, in this section, we will assume $\pi_{Ney} \leq \frac{1}{2}$.* However, we emphasize our results and analysis can be made to hold for all $\pi_{\text{Ney}} \in (0,1)$ by flipping the role of the two policies.

4.1 The ClipSMT Algorithm

We begin by describing the ClipSMT algorithm. The idea behind this approach is straightforward: since we do not know the Neyman allocation, we instead choose its empirical counterpart,

$$\tilde{\pi}_t = \frac{\widehat{m}_{t-1}(1)}{\widehat{m}_{t-1}(0) + \widehat{m}_{t-1}(1)}. \tag{7}$$

While this approach is appealing, it will not work without modification. This is because $\tilde{\pi}_t$ is overly sensitive to random fluctuations during the early rounds of interaction. As an extreme example, suppose that we select $A_1 = 1, A_2 = 0$ and observe $R_1 = 0, R_2 = 1$. Then, $\tilde{\pi}_t = 0$ for all the subsequent rounds, leading to infinite Neyman regret.

Therefore, we require some form of regularization to guarantee ClipSMT is robust to randomness early in the experiment. To regularize $\tilde{\pi}_t$, we follow Cook et al. (2024) and choose the allocation

$$\pi_t = \text{Clip}(\tilde{\pi}_t, c_t, 1 - c_t). \tag{8}$$

for some clipping sequence c_t. Our subsequent finite sample analysis will show that the setting $c_t = \frac{1}{2}t^{-\frac{1}{3}}$ is the correct choice in a worst-case sense. The complete algorithm can be found in Algorithm 1.

Algorithm 1 ClipSMT

Input: Clipping sequence (c_t)
for each round $t \in \mathbb{N}$ **do**
  Compute $\tilde{\pi}_t = \frac{\widehat{m}_{t-1}(1)}{\widehat{m}_{t-1}(0)+\widehat{m}_{t-1}(1)}$
  Set $\pi_t = \text{Clip}(\tilde{\pi}_t, c_t, 1 - c_t)$
  Play $A_t \sim \text{Bernoulli}(\pi_t)$ and observe R_t
end for

4.2 Understanding the Finite Sample Behavior of ClipSMT

We now present our finite sample analysis of ClipSMT. To begin, we will assume that the clipping sequence has polynomial decay so that $c_t = \frac{1}{2}t^{-\alpha}$ for some $\alpha \in (0,1)$. We discuss alternative choices for (c_t) in Appendix C.

Our analysis splits the behavior of ClipSMT into two phases — a clipping phase followed by a concentration phase. In the clipping phase, random fluctuations in R_t will induce large variations in $\tilde{\pi}_t$, leading our algorithm to clip $\tilde{\pi}_t$. The clipping phase ends once we can guarantee that our algorithm will no longer clip the plug-in allocation $\tilde{\pi}_t$, marking the start of the concentration phase, in which we can show that π_t converges to π_{Ney} at a $O\left(t^{-\frac{1}{2}}\right)$ rate.

Our first result characterizes the length of the clipping phase for various choices of α, demonstrating how to select α appropriately.

Lemma 4.1. *Assume for simplicity that $\pi_{Ney} \leq \frac{1}{2}$. Suppose we run* ClipSMT *with $c_t = \frac{1}{2}t^{-\alpha}$ for $\alpha \in (0,1)$. Let $p = \min\left(\alpha, \frac{1-\alpha}{2}\right)$ and define*

$$\tau = \tilde{O}\left(\left[\frac{1}{\pi_{Ney}} + \frac{1}{m_1}\left(\frac{1}{m_0} + \frac{1}{m_1}\right)^{\frac{1}{2}} \log\left(\frac{1}{\delta}\right)\right]^{\frac{1}{p}}\right). \tag{9}$$

Then with probability at least $1 - \delta$, for all $t \geq \tau$, we have that $\tilde{\pi}_t = \pi_t$.

Before proceeding we make a few remarks about this result. First, we can show that there exists a problem instance such that the above bound on the length of the clipping phase is tight (modulo some polylogarithmic factors). This implies that without additional knowledge on ν, setting $\alpha = \frac{1}{3}$ minimizes the length of the clipping phase in a worst-case sense. Furthermore, the proceeding results will show that in the concentration phase π_t converges to π_{Ney} at a rate that is independent of α, thus suggesting that $\alpha = \frac{1}{3}$ is in some sense the correct choice when we don't have additional information about the uncentered second moments.

The end of the clipping phase indicates sufficient data collection, mitigating the effects of random fluctuations on π_t, thus marking the start of the concentration phase. In this phase we can show that $\pi_t \in [\pi_{\min}, \pi_{\max}]$, so that $N_t(1) = \Omega(\pi_{\min} \cdot t)$. A simple computation shows that this implies that π_t converges to π_{Ney} at a $O\left((\pi_{\min} \cdot t)^{-\frac{1}{2}}\right)$ rate. While this leads to the correct dependence on t, the scaling with respect to $\pi_{\min}$ is suboptimal—we expect the scaling to be with respect to π_{Ney}. To see why, note that as the interaction progresses, we expect π_t to eventually converge to π_{Ney}. Consequently, we anticipate $N_t(1) = \Theta(\pi_{\text{Ney}} \cdot t)$ which further implies that π_t converges to π_{Ney} at a $O\left((\pi_{\text{Ney}} \cdot t)^{-\frac{1}{2}}\right)$ rate. To remedy this issue, we develop a 'double bounding' technique that uses these initial bounds on π_t and refines them to obtain the correct dependence on π_{Ney}. This gives us the following result which shows that π_t converges to π_{Ney} at the desired rate.

Lemma 4.2. *Assume for simplicity that $\pi_{Ney} \leq \frac{1}{2}$.*

Define

$$\tau = \tilde{O}\left(\left[\frac{1}{\pi_{Ney}} + \frac{1}{m_1}\left(\frac{1}{m_0} + \frac{1}{m_1}\right)^{\frac{1}{2}} \log\left(\frac{1}{\delta}\right)\right]^3\right). \tag{10}$$

Then with probability at least $1-\delta$, for all $t \geq \tau$, ClipSMT *guarantees that*

$$|\pi_{Ney} - \pi_{t+1}| \leq O\left(\sqrt{\frac{\ell(t,\delta)}{t}}\right) \tag{11}$$

where $\ell(t,\delta) = O\left(\log\log t + \log \frac{1}{\delta}\right)$.

The above result shows that following an additional burn-in period after the clipping phase, π_t will converge to π_{Ney} at the desired $O\left((\pi_{\text{Ney}} \cdot t)^{-\frac{1}{2}}\right)$ rate. We also make a remark about the $\sqrt{m_A}$ terms that appear in our bound. These terms appear because of the concentration inequalities we use for m_A. Unfortunately, we can show that this term is asymptotically unavoidable (see Remark B.7 in Appendix B.3).

4.3 Bounding the Neyman Regret

Before stating our bound on the Neyman regret of ClipSMT, we first give an alternative expression for the simple Neyman regret and provides insight into our Neyman regret bound.

Lemma 4.3. *Fix $\pi_t \in [0,1]$ and let $\epsilon_t = \pi_t - \pi_{Ney}$. Then we have that*

$$f(\pi_t) - f(\pi_{Ney}) = \Theta\left(\epsilon_t^2\right) \tag{12}$$

The proof of this result can be found in Appendix A. Surprisingly, this result shows that if π_t converges to π_{Ney} at a $O\left(t^{-\frac{1}{2}}\right)$ rate, then the simple Neyman regret will shrink at a $O\left(t^{-1}\right)$ rate. Our next result uses this fact in conjunction with the prior bounds on π_t to bound the Neyman regret.

Theorem 4.4. *Assume for simplicity that $\pi_{Ney} \leq \frac{1}{2}$. Suppose we run* ClipSMT *with $c_t = \frac{1}{2}t^{-\frac{1}{3}}$. Then probability at least $1-\delta$, the Neyman Regret is at most*

$$\tilde{O}\left(\pi_{Ney}^{-1} \cdot \log(T)\right). \tag{13}$$

The proof of this result can be found in Appendix A.2. We have just shown that ClipSMT obtains *logarithmic* Neyman regret, providing an exponential improvement from the $O\left(\sqrt{T}\right)$ Neyman regret obtained by prior works. As the proceeding discussion highlights, ClipOGD works in a more general "design-based" setup. However, it highlights the significant improvements that can be gained in the superpopluation setting considered in this papers.

4.4 Comparisons with Prior Work

We continue by comparing our results with past works.

Comparison with Dai et al. (2023). When comparing our Neyman regret bounds to ClipOGD, we observe exponential improvements in scaling with respect to π_{Ney} and T.

Starting out with the dependence on π_{Ney}, our bound scales like $O\left(\pi_{\text{Ney}}^{-1}\right)$ while ClipOGD scales like $O\left(\exp\left(\pi_{\text{Ney}}^{-1}\right)\right)$. We remark that it is not fully clear if the exponential scaling for ClipOGD is a product of the proof technique or is a fundamental drawback of ClipOGD. Inspecting the proof in Dai et al. (2023), this exponential dependence is introduced to tune the learning rate—if bounds on π_{Ney} are known, ClipOGD can be tuned to scale polynomially in π_{Ney}^{-1}. However, even then, not only is the exponent in their polynomial always worse than ours, but it also scales with $\sqrt{\log T}$, while ClipSMT does not. Finally, we empirically observe that ClipOGD is sensitive to parameter choices. The choices suggested by their analysis can often lead to poor performance (as we demonstrate in Section 5) indicating that the aforementioned exponential dependence is indeed a fundamental drawback.

Next, we see that our Neyman regret scales like $O(\log T)$ while ClipOGD scales like $O(\sqrt{T})$. While this is an exponential improvement, we believe this difference is primarily due to the differences in our problem settings—we consider the superpopulation setting where outcomes are stochastic whereas Dai et al. (2023) consider the fixed-design setting where the outcomes are a fixed sequence. In the fixed-design setting, the potential outcomes can be chosen adversarially, including with knowledge of the algorithm, thus increasing the problem's difficulty. The differences between these settings parallels the differences between stochastic and adversarial MABs where we observe similar gaps in regret bounds. In the stochastic bandit setting, the best one can obtain is $O(\log T)$ problem dependent regret(Auer et al., 2002); whereas in the adversarial bandit setting, the best one can hope to do is $O\left(\sqrt{T}\right)$ minimax regret (Auer et al., 1995).

Comparison with Cook et al. (2024). As we have mentioned, our algorithm is a variant of the algorithm proposed by Cook et al. (2024), tailored to the aHT estimator. The primary difference between our work and Cook et al. (2024) is that their focus is asymptotic while ours is nonasymptotic. The asymptotic perspective makes design choices such as the appropriate clipping sequence opaque. In their concluding remarks Cook et al. (2024) state that selection of the clipping sequence is an interesting question for future work – our finite sample analysis gives a concrete answer to this question. As an example of the difficulty in choosing the clipping sequence, Cook et al. (2024) uses a clipping sequence with exponential decay. Our finite sample analysis indicates that with constant probability, such a clipping sequence will result in an allocation that does not converge to π_{Ney}. Finally, we remark that using a clipping sequence with polynomial decay allows us to slightly generalize their asymptotic results by removing the requirement that bounds on π_{Ney} are known.

5 EXPERIMENTS

In this section, we experimentally[4] validate our algorithm. Our objective is to compare our algorithm to existing approaches as well as sensible baselines and to understand how well our theoretical characterization of ClipSMT aligns with its empirical behavior.

We start by comparing our algorithm to existing approaches and some non-adaptive baselines. In these experiments, we compare ClipSMT with ClipOGD, the infeasible Neyman Allocation, a balanced allocation with $\pi = \frac{1}{2}$, and a two-stage design we call Explore-then-Commit (EtC). For EtC, we select each treatment arm with equal probability for $T^{\frac{1}{3}}$ rounds, after which we compute the empirical Neyman allocation and use this allocation for the remaining rounds.

We evaluate each approach on nine problem instances, running them for T rounds, where T varies from 1000 to 20000 in increments of 1000. For each fixed value of T, we run ClipSMT, ClipOGD, and the two-stage design 5000 times to approximate the variance of the resulting ATE estimate. For the Neyman and balanced allocations, we compute their variance using equation (2). Our results show that ClipSMT outperforms ClipOGD and EtC, and adapts well to difficult problem instances (i.e., when the Neyman allocation deviates from $\frac{1}{2}$). The results of the experiments are displayed in Figure 1. Additionally, we perform a more comprehensive simulation of these algorithms in the small sample regime, where we observe similar behavior. The results of this experiment are shown in Figure 2 in Appendix D.

Next, we validate whether the length of the clipping phase predicted by our theory aligns with the empirical behavior of ClipSMT. To do this, we run ClipSMT using $c_t = \frac{1}{2}t^{-\alpha}$ for various values of $\alpha \in (0, 1)$. We run ClipSMT for each value of α and determine the 0.95 quantile of the clipping phase length based on 5000 simulations. Using these values, we compute the ratio between the theoretically predicted clipping time to the empirically computed clipping time. The results of this experiment are shown in Figure 3 (a) in Appendix D.

Inspecting these results, we find that the ratio peaks around $\alpha = \frac{1}{3}$. This behavior is due to a technical difficulty that arises in our proof which we take a brief moment to elucidate. Specifically, upper-bounding the length of the clipping phase involves bounding the quantity $\min\{t : \sum_i t^{p_i} \geq c_1 + c_2 \log\log t\}$, where $c_1 \approx \frac{1}{\pi_{\text{Ney}}^2}$ and $c_2 \approx \frac{1}{\pi_{\text{Ney}}}$. To accomplish this, we compute an upper bound on $\min\{t : t^{\max p_i} \geq c_1 + c_2 \log\log t\}$ which is also an upper bound on the initial quantity. Noting that $t^{\max p_i} \leq \sum_i t^{p_i}$, it is clear that this step introduces some looseness to our bound. However, we see that as $\pi_{\text{Ney}} \to 0$, we will have $\sum t^{p_i} = \Theta(t^{\max p_i})$ since the growth of $t^{\max p_i}$ will become the dominating term.

Instead, we bound $\min\{t : t^{\max p_i} \geq c_1 + c_2 \log\log t\}$, which provides the correct bound for large values of c_1 and c_2 but is loose for small and moderate values of c_1 and c_2. Therefore, while our bound is tight in the worst case, it has some looseness for specific problems — resolving this issue remains an interesting technical problem for future work.

In order to validate this worst-case optimality, we consider a sequence of problems with Bernoulli arms $\boldsymbol{\mu}^{(n)} = \left(0.5, \frac{0.5}{n}\right)$. These are chosen to guarantee π_{Ney} converges to 0 which captures the notion of increasing problem difficulty. We run ClipSMT with varying values of α on each problem instance n and compute the median length of the clipping phase over 5000 simulations. For each n, we then determined the value of α, which leads to the shortest clipping phase. The results of this experiment are shown in Figure 3 (b) in Appendix D, and confirm that setting $\alpha = \frac{1}{3}$ minimizes the length of the clipping phase in the worst case.

6 CONCLUSION

In this paper, we performed a finite sample analysis of the ClipSMT algorithm for adaptive estimation of the ATE. Our analysis clarified several aspects of algorithm design, including how to properly tune the clipping sequence. Furthermore, we demonstrated that our approach achieves exponential improvements in two distinct areas when compared to the only other method with a finite time analysis. Our comprehensive analysis meticulously addressed all problem pa-

[4] Code for replicating experiments can be found at the following GitHub repo: `https://github.com/oneopane/adaptive-ate-estimation`.

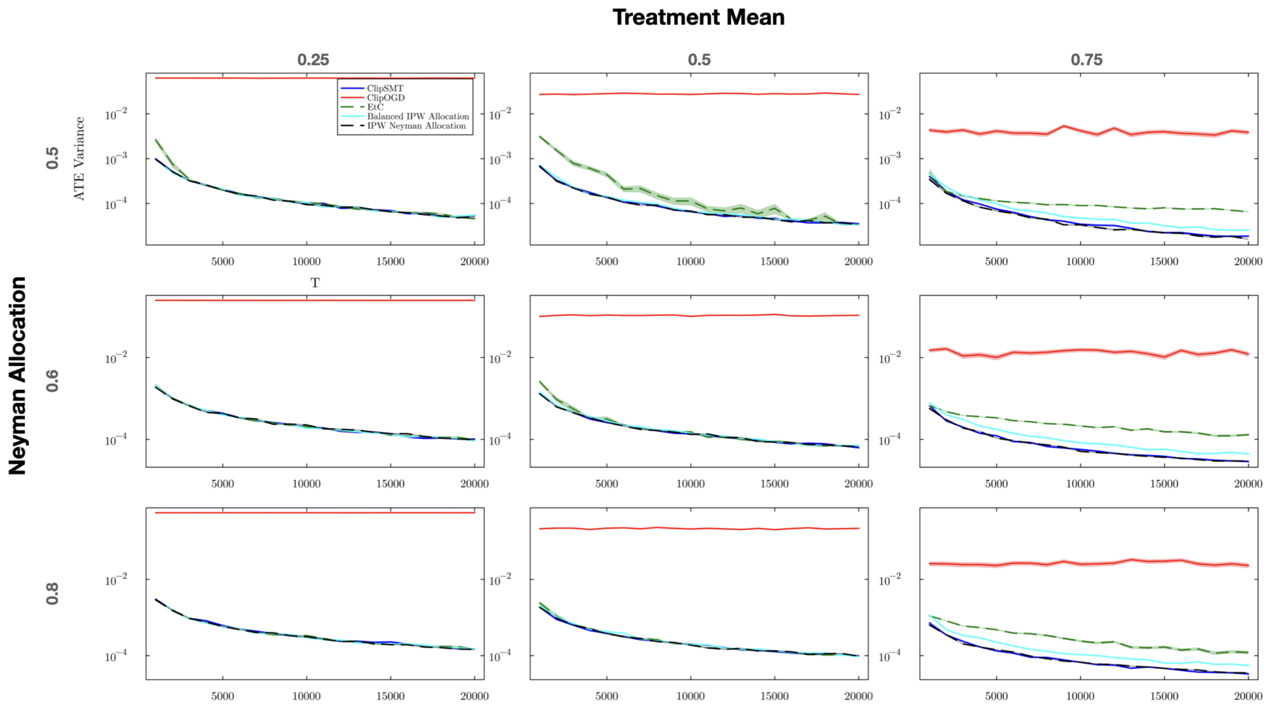


Figure 1: Comparison of the performance of ClipSMT, ClipOGD, Explore-then-Commit (EtC), Neyman allocation, and a balanced allocation with the treatment and control arms following Bernoulli distributions. Individual subplots plot the variance of each design against the number of samples for a fixed problem instance. Each column keeps the treatment mean fixed, and each row keeps the Neyman allocation fixed. Moving to the right increases the treatment mean and moving down increases the Neyman allocation. Overall the performance of ClipSMT is always competitive with the performance of the infeasible Neyman allocation and outperforms the other adaptive designs. Furthermore, as the Neyman allocation increases, we see that ClipSMT adapts to the increased difficulty while EtC and the balanced design do not. Note that error bars are plotted, however they are narrow due to the large number of simulations performed.

rameters, providing a clearer and more detailed understanding of the complexity of adaptive ATE estimation.

Several promising directions for future work emerge from our findings. One obvious direction is to extend our analysis to the Augmented Inverse Probability Weighted estimator, which has more desirable properties and is more appropriate when contextual information is available. Additionally, expanding these results to accommodate larger action spaces and stochastic context-dependent policies warrants further discussion.

Acknowledgments

This work is sponsored in part by AFRL Center of Excellence grant FA86501925209 and NSF AI Institute for Societal Decision Making under grant IIS2229881.

References

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multi-armed bandit problem. *Machine Learning*, 47:235–256.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (1995). Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 322–331. IEEE.

Banerjee, A. V., Duflo, E., and Kremer, M. (2016). The influence of randomized controlled trials on development economics research and on development policy. *The State of Economics, The State of the World*, pages 482–488.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer.

Chow, S.-C. and Chang, M. (2008). Adaptive design methods in clinical trials–a review. *Orphanet Journal of Rare Diseases*, 3:1–13.

Chow, S.-C., Chang, M., and Pong, A. (2005). Statistical consideration of adaptive methods in clinical development. *Journal of Biopharmaceutical Statistics*, 15(4):575–591.

Cook, T., Mishler, A., and Ramdas, A. (2024). Semiparametric efficient inference in adaptive experiments. In *Causal Learning and Reasoning*, pages 1033–1064. PMLR.

Dai, J., Gradu, P., and Harshaw, C. (2023). Clip-OGD: An experimental design for adaptive Neyman allocation in sequential experiments. *Advances in Neural Information Processing Systems*, 36.

Duan, Y., Jia, Z., and Wang, M. (2020). Minimax-optimal off-policy evaluation with linear function approximation. In *International Conference on Machine Learning*, pages 2701–2709. PMLR.

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, pages 1097–1104.

Garivier, A., Lattimore, T., and Kaufmann, E. (2016). On explore-then-commit strategies. *Advances in Neural Information Processing Systems*, 29.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences*, 118(15):e2014602118.

Hahn, J., Hirano, K., and Karlan, D. S. (2009). Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29:108 – 96.

Hanna, J. P., Thomas, P. S., Stone, P., and Niekum, S. (2017). Data-efficient policy evaluation through behavior policy search. In *International Conference on Machine Learning*.

Hollis, S. and Campbell, F. (1999). What is meant by intention to treat analysis? survey of published randomised controlled trials. *The BMJ*, 319(7211):670–674.

Howard, S. R., Ramdas, A., McAuliffe, J. D., and Sekhon, J. S. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*.

Huang, A., Leqi, L., Lipton, Z., and Azizzadenesheli, K. (2021). Off-policy risk assessment in contextual bandits. *Advances in Neural Information Processing Systems*, 34:23714–23726.

Huang, A., Leqi, L., Lipton, Z., and Azizzadenesheli, K. (2022). Off-policy risk assessment for markov decision processes. In *International Conference on Artificial Intelligence and Statistics*, pages 5022–5050. PMLR.

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

Jiang, N. and Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 652–661. PMLR.

Kallus, N. and Uehara, M. (2019). Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *Journal of Machine Learning Research*, 21:167:1–167:63.

Karampatziakis, N., Mineiro, P., and Ramdas, A. (2021). Off-policy confidence sequences. In *International Conference on Machine Learning*, pages 5301–5310. PMLR.

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

Konyushova, K., Chen, Y., Paine, T., Gulcehre, C., Paduraru, C., Mankowitz, D. J., Denil, M., and de Freitas, N. (2021). Active offline policy selection. *Advances in Neural Information Processing Systems*, 34:24631–24644.

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

Li, L., Chu, W., Langford, J., and Wang, X. (2011). Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 297–306.

Li, L., Munos, R., and Szepesvári, C. (2015). Toward minimax off-policy value estimation. In *Artificial Intelligence and Statistics*, pages 608–616. PMLR.

Li, T., Shi, C., Wang, J., Zhou, F., et al. (2024). Optimal treatment allocation for efficient policy evaluation in sequential decision making. *Advances in Neural Information Processing Systems*, 36.

Ma, C., Zhu, B., Jiao, J., and Wainwright, M. J. (2021). Minimax off-policy evaluation for multi-armed bandits. *IEEE Transactions on Information Theory*, 68:5314–5339.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. *Annals of Agricultural Sciences*, pages 1–51.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97:123–150.

Robbins, H. E. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58:527–535.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294.

US Food and Drug Administration (2019). Adaptive designs for clinical trials of drugs and biologics: guidance for industry. *Rockville: Food and Drug Administration*, page 2020.

Wagenmaker, A. J. and Foster, D. J. (2023). Instance-optimality in interactive decision making: Toward a non-asymptotic theory. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 1322–1472. PMLR.

Wang, Y.-X., Agarwal, A., and Dudık, M. (2017). Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning*, pages 3589–3597. PMLR.

Waudby-Smith, I. and Ramdas, A. (2023). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B Methodological*.

Waudby-Smith, I., Wu, L., Ramdas, A., Karampatziakis, N., and Mineiro, P. (2022). Anytime-valid off-policy inference for contextual bandits. *ACM/JMS Journal of Data Science*.

Wing, C., Simon, K., and Bello-Gomez, R. A. (2018). Designing difference in difference studies: Best practices for public health policy research. *Annual Review of Public Health*, 39:453–469.

Zhang, K., Janson, L., and Murphy, S. (2020). Inference for batched bandits. *Advances in Neural Information Processing Systems*, 33:9818–9829.

Zhang, K., Janson, L., and Murphy, S. (2021). Statistical inference with m-estimators on adaptively collected data. *Advances in Neural Information Processing Systems*, 34:7460–7471.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. Code for replicating experiments can be found at the following github repo: `https://github.com/oneopane/adaptive-ate-estimation`
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes
   (b) Complete proofs of all theoretical results. Yes
   (c) Clear explanations of any assumptions. Yes
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results. Code for replicating experiments can be found at the following github repo: `https://github.com/oneopane/adaptive-ate-estimation`
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Yes
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Yes
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Not applicable
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. Not Applicable
   (b) The license information of the assets, if applicable. Not Applicable
   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable
   (d) Information about consent from data providers/curators. Not Applicable
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable
5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. Not Applicable

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

A Analysis of ClipSMT

In this section, we will prove our bound on the Neyman regret of ClipSMT.

A.1 Preliminaries.

Before we proceed to the analysis, we first introduce some notation and define a 'good event' which we will assume to hold throughout the analysis. We define the following events

$$\mathcal{E}_1(\delta_1) = \bigcap_{t=1}^{\infty} \left\{ N_t(1) \in \left[\sum_{s=1}^{t} \pi_s - \beta_1(t,\delta_1), \sum_{s=1}^{t} \pi_s + \beta_1(t,\delta_1) \right] \right\} \tag{14}$$

$$\mathcal{E}_2(\delta_2) = \bigcap_{A\in\{0,1\}} \bigcap_{t=1}^{\infty} \left\{ m_A \in [\widehat{m}_t(A) - \beta_2(t,\delta), \widehat{m}_t(A) + \beta_2(t,\delta_2)] \right\}. \tag{15}$$

Applying Lemmas B.5 and B.6, using β_1 and β_2 respectively defined in equations (89) and (92) with $\delta_1 = \frac{\delta}{3}, \delta_2 = \frac{2\delta}{3}$, we see that the event $\mathcal{E} = \mathcal{E}_1(\delta_1) \bigcap \mathcal{E}_2(\delta_2)$ occurs with probability at least $1-\delta$. For the remainder of the section, we will assume that this event hold.

A.2 Bounding the Neyman Regret (Theorem 4.4)

We will bound the cumulative Neyman regret by bounding the simple Neyman regret and then summing over those terms. In order to do so, we will handle the clipping phase and concentration phases separately.

For the clipping phase, Lemma A.1 demonstrates that we can guarantee $\pi_t \in [\pi_{\min}, \pi_{\max}]$ where $\pi_{\min}, \pi_{\max}$ only depend on m_A. This implies that the instantaneous Neyman regret for each round in the clipping phase can be upper bounded by a constant $c(m_0, m_1) = \max_{\pi\in\{\pi_{\min},\pi_{\max}\}} f(\pi) - f(\pi_{\text{Ney}})$ which only depends on m_A. Furthermore, Lemma 4.1 shows that the length of the clipping phase is at most τ so that the cumulative Neyman regret from the clipping phase can be upper bounded as $c(m_0, m_1)\cdot\tau$ which is independent of T.

For the concentration phase, we apply Lemma 4.2 which shows that $\epsilon_t \le \tilde{O}\left(t^{-\frac{1}{2}}\right)$ so that Lemma 4.3 implies that the instantaneous Neyman regret for each round of the concentration phase is at most

$$16\left(\frac{1}{m_0+m_1}\right)^2 \left(\frac{1}{\sqrt{m_0\,(1-\pi_{\text{Ney}})}} + \frac{1}{\sqrt{m_1\pi_{\text{Ney}}}}\right)^2 \frac{\ell(t,\delta)}{t}. \tag{16}$$

Therefore, we can bound the cumulative Neyman regret during the clipping phase as

$$\begin{aligned}
\sum_{t=\tau+1}^{T} f(\pi_t) - f(\pi_{\text{Ney}}) &\le 16\left(\frac{1}{m_0+m_1}\right)^2 \left(\frac{1}{\sqrt{m_0\,(1-\pi_{\text{Ney}})}} + \frac{1}{\sqrt{m_1\pi_{\text{Ney}}}}\right)^2 \sum_{t=\tau+1}^{T} \frac{\ell(t,\delta)}{t} && (17)\\
&\le 16\left(\frac{1}{m_0+m_1}\right)^2 \left(\frac{1}{\sqrt{m_0\,(1-\pi_{\text{Ney}})}} + \frac{1}{\sqrt{m_1\pi_{\text{Ney}}}}\right)^2 \sum_{t=1}^{T} \frac{\ell(t,\delta)}{t} && (18)\\
&\le 16\left(\frac{1}{m_0+m_1}\right)^2 \left(\frac{1}{\sqrt{m_0\,(1-\pi_{\text{Ney}})}} + \frac{1}{\sqrt{m_1\pi_{\text{Ney}}}}\right)^2 \ell(T,\delta)\log(T). && (19)
\end{aligned}$$

Combining these bounds we see that the Neyman regret can be bounded as

$$c(m_0,m_1)\cdot\tau + 16\left(\frac{1}{m_0+m_1}\right)^2 \left(\frac{1}{\sqrt{m_0\,(1-\pi_{\text{Ney}})}} + \frac{1}{\sqrt{m_1\pi_{\text{Ney}}}}\right)^2 \ell(T,\delta)\log(T) = \tilde{O}(\log(T)), \tag{20}$$

which gives the desired result.

Lemma 4.3. *Fix $\pi_t \in [0,1]$ and let $\epsilon_t = \pi_t - \pi_{Ney}$. Then we have that*

$$f(\pi_t) - f(\pi_{Ney}) = \Theta\left(\epsilon_t^2\right) \tag{12}$$

Proof. The proof follows from the following series of algebraic manipulations:

$$
\begin{aligned}
f(\pi_{\text{Ney}} + \epsilon_t) - f(\pi_{\text{Ney}}) &= \frac{m_1^2}{\pi_{\text{Ney}} + \epsilon_t} + \frac{m_0^2}{(1 - \pi_{\text{Ney}} - \epsilon_t)} - \frac{m_1^2}{\pi_{\text{Ney}}} + \frac{m_0^2}{(1 - \pi_{\text{Ney}})} && (21)\\
&= \epsilon_t \left(\frac{m_0^2}{(1 - \pi_{\text{Ney}})(1 - \pi_{\text{Ney}} - \epsilon_t)} - \frac{m_1^2}{\pi_{\text{Ney}}(\pi_{\text{Ney}} + \epsilon_t)} \right) && (22)\\
&\overset{(a)}{=} \epsilon_t \left(\frac{m_0^2}{\left(\frac{m_0}{m_0+m_1}\right)\left(\frac{m_0}{m_0+m_1} - \epsilon_t\right)} - \frac{m_1^2}{\left(\frac{m_1}{m_0+m_1}\right)\left(\frac{m_1}{m_0+m_1} + \epsilon_t\right)} \right) && (23)\\
&= \epsilon_t \left(\frac{m_0^2 (m_0+m_1)^2}{m_0^2 - m_0 (m_0+m_1) \epsilon_t} - \frac{m_1^2 (m_0+m_1)^2}{m_1^2 - m_1 (m_0+m_1) \epsilon_t} \right) && (24)\\
&= \epsilon_t \left(\left[\frac{m_0^2 (m_0+m_1)^2}{m_0^2 - m_0 (m_0+m_1) \epsilon_t} - (m_0+m_1)^2 \right] \right.\\
&\qquad \left. + \left[(m_0+m_1)^2 - \frac{m_1^2 (m_0+m_1)^2}{m_1^2 - m_1 (m_0+m_1) \epsilon_t} \right] \right) && (25)\\
&= \epsilon_t^2 (m_0+m_1)^3 \left(\frac{1}{m_0 - (m_0+m_1) \epsilon_t} - \frac{1}{m_1 - (m_0+m_1) \epsilon_t} \right), && (26)
\end{aligned}
$$

where in (a) we have used the fact that $\pi_{\text{Ney}} = \frac{m_1}{m_0+m_1}$. ☐

A.3 Clipping Phase

We now cover various proofs related to the analysis of the clipping phase of our algorithm.

We begin by proving Lemma 4.1 which we restate for the reader's convenience.

Lemma 4.1. *Assume for simplicity that $\pi_{Ney} \leq \frac{1}{2}$. Suppose we run* ClipSMT *with $c_t = \frac{1}{2} t^{-\alpha}$ for $\alpha \in (0,1)$. Let $p = \min\left(\alpha, \frac{1-\alpha}{2}\right)$ and define*

$$
\tau = \tilde{O}\left(\left[\frac{1}{\pi_{Ney}} + \frac{1}{m_1} \left(\frac{1}{m_0} + \frac{1}{m_1} \right)^{\frac{1}{2}} \log\left(\frac{1}{\delta}\right) \right]^{\frac{1}{p}} \right). \quad (9)
$$

Then with probability at least $1-\delta$, for all $t \geq \tau$, we have that $\tilde{\pi}_t = \pi_t$.

Proof. To begin, we observe that since the function $x, y \mapsto \frac{x}{x+y}$ is monotonic increasing (resp. decreasing) in x (resp. y) we have (on the event $\mathcal{E}$) that

$$
\tilde{\pi}_{t+1} \in \left[\frac{m_1 - \beta_2(N_t(1), \delta_2)}{m_0 + \beta_2(N_t(0), \delta_2) + m_1 - \beta_2(N_t(1), \delta_2)}, \frac{m_1 + \beta_2(N_t(1), \delta_2)}{m_0 - \beta_2(N_t(0), \delta_2) + m_1 + \beta_2(N_t(1), \delta_2)} \right]. \quad (27)
$$

We note the above interval is random because $N_t(A)$ is random. In order to construct bounds on $N_t(A)$ we use the fact that $\pi_t \in [c_t, 1-c_t]$ so that an integral-sum argument demonstrates

$$
\sum_{s=1}^{t} \pi_s \in \left[\frac{1}{2} \cdot \frac{t^{1-\alpha} - 1}{1-\alpha}, \frac{1}{2} \cdot \frac{t^{1-\alpha}}{1-\alpha} \right]. \quad (28)
$$

Therefore, on the event $\mathcal{E}$, we obtain

$$
\begin{aligned}
N_t(1) \in \mathcal{N}(t, \delta_1) &= \left[\frac{1}{2} \cdot \frac{t^{1-\alpha} - 1}{1-\alpha} - \beta_1(t, \delta_1), t - \frac{1}{2} \cdot \frac{t^{1-\alpha}}{1-\alpha} + \beta_1(t, \delta_1) \right]\\
&= \left[\frac{1}{2} \cdot \frac{t^{1-\alpha} - 1}{1-\alpha} - \sqrt{t \cdot \ell(t, \delta_1)}, t - \frac{1}{2} \cdot \frac{t^{1-\alpha}}{1-\alpha} + \sqrt{t \cdot \ell(t, \delta_1)} \right], && (29)
\end{aligned}
$$

where we have set $\ell(t,\delta) = \sqrt{.7225\left(\log\log t + 0.72\log\frac{5.2}{\delta}\right)}$.

Our strategy moving forward will be to use these bounds on $N_t(1)$ to construct a time τ such that for all $t \geq \tau$, we have $\tilde{\pi}_{t+1} \in [c_{t+1}, 1-c_{t+1}]$. We demonstrate how to do so in order to guarantee $\tilde{\pi}_{t+1} \geq c_{t+1}$ as the other case is entirely analogous. Observe that our initial (random) lower bound on $\tilde{\pi}_{t+1}$ together with our bounds on $N_t(1)$ imply that on the event $\mathcal{E}$, we have

$$\begin{aligned}
\tilde{\pi}_{t+1} &\geq \min_{n\in\mathcal{N}(t,\delta_1)} \frac{m_1 - \sqrt{\frac{\ell(n,\delta_2)}{m_1\cdot n}}}{m_0 + \sqrt{\frac{\ell(t-n,\delta_2)}{m_0\cdot(t-n)}} + m_1 - \sqrt{\frac{\ell(n,\delta_2)}{m_1\cdot n}}} \\
&\geq \min_{n\in\mathcal{N}(t,\delta_1)} \frac{m_1 - \sqrt{\frac{\ell(t,\delta_2)}{m_1\cdot n}}}{m_0 + \sqrt{\frac{\ell(t,\delta_2)}{m_0\cdot(t-n)}} + m_1 - \sqrt{\frac{\ell(t,\delta_2)}{m_1\cdot n}}}, \qquad (30)
\end{aligned}$$

where the final inequality follows from the monotonic properties of the map $x, y \mapsto \frac{x}{x+y}$. Therefore, our objective is to upper bound the quantity

$$\begin{aligned}
\underline{\tau} &= \min\left\{t : \min_{n\in\mathcal{N}(t,\delta_1)} \frac{m_1 - \sqrt{\frac{\ell(t,\delta_2)}{m_1\cdot n}}}{m_0 + \sqrt{\frac{\ell(t,\delta_2)}{m_0\cdot(t-n)}} + m_1 - \sqrt{\frac{\ell(t,\delta_2)}{m_1\cdot n}}} \geq \frac{1}{2}(t+1)^{-\alpha}\right\} \qquad (31)\\
&\leq \min\left\{t : \min_{n\in\mathcal{N}(t,\delta_1)} \frac{m_1 - \sqrt{\frac{\ell(t,\delta_2)}{m_1\cdot n}}}{m_0 + \sqrt{\frac{\ell(t,\delta_2)}{m_0\cdot(t-n)}} + m_1 - \sqrt{\frac{\ell(t,\delta_2)}{m_1\cdot n}}} \geq \frac{1}{2}t^{-\alpha}\right\}, \qquad (32)
\end{aligned}$$

where the inequality follows from the fact that the LHS in increasing in t and the RHS is decreasing in t. Letting $n^\star$ denote the minimizer of equation (30), by applying Lemma B.8 we observe that

$$n^\star \in \left\{\frac{1}{2}\cdot\frac{t^{1-\alpha}-1}{1-\alpha} - \sqrt{t\cdot\ell(t,\delta_1)},\, t - \frac{1}{2}\cdot\frac{t^{1-\alpha}}{1-\alpha} + \sqrt{t\cdot\ell(t,\delta_1)}\right\}. \qquad (33)$$

Therefore, we can compute an upper bound for each of the two cases so that taking the maximum of these bounds will result in an upper bound on equation (32).

We will demonstrate this for the case $n^\star = \frac{1}{2}\cdot\frac{t^{1-\alpha}}{1-\alpha} - \sqrt{t\cdot\ell(t,\delta_1)}$ since the other case is similar. After plugging this value of $n^\star$ into equation (32), rearranging terms shows that

$$\min\left\{t : m_1 \geq \frac{1}{2}t^{-\alpha}(m_0+m_1) + \frac{1}{2}t^{-\frac{2\alpha+1}{2}}\left(\frac{\ell(t,\delta_2)}{m_0 f(t,\delta_1,\alpha)}\right)^{\frac{1}{2}} + t^{\frac{\alpha-1}{2}}\left(1-t^{-\alpha}\right)\left(\frac{\ell(t,\delta_2)}{m_1 g(t,\delta_1,\alpha)}\right)^{\frac{1}{2}}\right\}, \qquad (34)$$

where

$$\begin{aligned}
f(t,\delta,\alpha) &= 1 + t^{-\frac{1}{2}}\sqrt{\ell(t,\delta)} + \frac{t^{-1}-t^{-\alpha}}{2(1-\alpha)}, \\
g(t,\delta,\alpha) &= \frac{1-t^{\alpha-1}}{2(1-\alpha)} - t^{\frac{2\alpha-1}{2}}\sqrt{\ell(t,\delta)}.
\end{aligned}$$

Defining $p = \min\left\{\alpha, \frac{1-\alpha}{2}\right\}$, we can upper bound the RHS of equation (34) with

$$t^{-p}\left((m_0+m_1) + \left(\frac{\ell(t,\delta_2)}{m_0 f(t,\delta_1,p)}\right)^{\frac{1}{2}} + \left(\frac{\ell(t,\delta_2)}{m_1 g(t,\delta_1,p)}\right)^{\frac{1}{2}}\right). \qquad (35)$$

Rearranging terms demonstrates that it is sufficient to bound

$$\min\left\{t : t^p \geq \frac{1}{\pi_{\text{Ney}}} + \frac{\sqrt{\ell(t,\delta_2)}}{m_1}\left[\left(\frac{1}{m_0 f(t,\delta_1,p)}\right)^{\frac{1}{2}} + \left(\frac{1}{m_1 g(t,\delta_1,p)}\right)^{\frac{1}{2}}\right]\right\}. \qquad (36)$$

Squaring both sides and applying the inequality $(a+b)^2 \leq a^2 + b^2$ twice shows that we can bound

$$\min \left\{ t : t^{2p} \geq \frac{2}{\pi_{\text{Ney}}^2} + \frac{4\ell(t,\delta_2)}{m_1^2} \left[\frac{1}{m_0 f\left(t,\delta_1,p\right)} + \frac{1}{m_1 g\left(t,\delta_1,p\right)} \right] \right\}. \tag{37}$$

Next, we apply Lemma B.11 and B.13 which show that when

$$t \geq O\left(\max\left\{ \left(\frac{1}{1-p}\right)^{\frac{1}{p}}, \left(\log\left(\frac{1}{\delta_1}\right)\right)^{\frac{1}{1-2p}} \right\} \right),$$

we have that $g(t,\delta_1,p), f(t,\delta_1,p) \geq \frac{1}{2}$. Applying Lemma B.10 to equation (37) using the above bounds on g, f demonstrates that

$$\underline{\tau} \leq \underline{\mathfrak{T}}^{\frac{1}{2p}} \tag{38}$$

where

$$\underline{\mathfrak{T}} = c_1(\pi_{\text{Ney}}) + c_2(\pi_{\text{Ney}}) \cdot c_3(\pi_{\text{Ney}}) \cdot \log\log c_1(\pi_{\text{Ney}}), \tag{39}$$

$$c_1(\pi_{\text{Ney}}) = \frac{2}{\pi_{\text{Ney}}^2} + \frac{4}{m_1^2}\left(\frac{1}{m_0} + \frac{1}{m_1}\right)\log\left(\frac{5.2}{\delta_2}\right), \tag{40}$$

$$c_2(\pi_{\text{Ney}}) = \frac{4}{m_1^2}\left(\frac{1}{m_0} + \frac{1}{m_1}\right), \tag{41}$$

$$c_3(\pi_{\text{Ney}}) = \frac{\log\log c_1(\pi_{\text{Ney}}) - \log(2p)}{\log\log c_1(\pi_{\text{Ney}})} \cdot \frac{c_1(\pi_{\text{Ney}})\log c_1(\pi_{\text{Ney}})}{c_1(\pi_{\text{Ney}})\log c_1(\pi_{\text{Ney}}) - c_2(\pi_{\text{Ney}})}. \tag{42}$$

Repeating the argument for the other choice of $n^\star$ yields the same result.

Finally, we can repeat the above argument for the upper bound on $\tilde{\pi}_{t+1}$ which shows that $\overline{\tau} \leq \overline{\mathfrak{T}}^{\frac{1}{2p}}$, where

$$\overline{\mathfrak{T}} = c_1(\pi_{\text{Ney}}) + c_2(\pi_{\text{Ney}}) \cdot c_3(\pi_{\text{Ney}}) \cdot \log\log c_1(\pi_{\text{Ney}}) \tag{43}$$

$$c_1(\pi_{\text{Ney}}) = \frac{2}{(1-\pi_{\text{Ney}})^2} + \frac{4}{m_0^2}\left(\frac{1}{m_0} + \frac{1}{m_1}\right)\log\left(\frac{5.2}{\delta_2}\right) \tag{44}$$

$$c_2(\pi_{\text{Ney}}) = \frac{4}{m_0^2}\left(\frac{1}{m_0} + \frac{1}{m_1}\right). \tag{45}$$

Letting $\tau = \max\{\underline{\tau}, \overline{\tau}\}$ gives the desired result. $\square$

A.4 Concentration Phase

In this section, we will prove Lemma 4.2 which we restate for the readers convenience below.

Lemma 4.2. *Assume for simplicity that $\pi_{Ney} \leq \frac{1}{2}$. Define*

$$\tau = \tilde{O}\left(\left[\frac{1}{\pi_{Ney}} + \frac{1}{m_1}\left(\frac{1}{m_0} + \frac{1}{m_1}\right)^{\frac{1}{2}} \log\left(\frac{1}{\delta}\right) \right]^3 \right). \tag{10}$$

Then with probability at least $1-\delta$, for all $t \geq \tau$, ClipSMT *guarantees that*

$$|\pi_{Ney} - \pi_{t+1}| \leq O\left(\sqrt{\frac{\ell(t,\delta)}{t}}\right) \tag{11}$$

where $\ell(t,\delta) = O\left(\log\log t + \log\frac{1}{\delta}\right)$.

Proof. To begin, we fix $t \geq \tau$ and let $s \in [\tau, t-1]$. Invoking Lemma A.1 implies that on the event $\mathcal{E}$ we have

$$N_s(1) \in \left[\pi_{\min} \cdot s - \sqrt{s\ell(s,\delta_1)}, s - \pi_{\min} \cdot s + \sqrt{s\ell(s,\delta_1)}\right]. \tag{46}$$

We will use this to construct a lower bound on π_{s+1} by solving the optimization problem in equation (30) using the interval defined above. Applying Lemma B.8, we can construct a lower bound by considering $N_s(1) \in \left\{\pi_{\min} \cdot s - \sqrt{s\ell(s,\delta_1)}, s - \pi_{\min} \cdot s - \sqrt{s\ell(s,\delta_1)}\right\}$. We demonstrate this for $N_s(1) = \pi_{\min} \cdot s - \sqrt{s\ell(s,\delta_1)}$. In this case, we have that

$$\begin{aligned}
\pi_{s+1} &\geq \frac{m_1 - \sqrt{\frac{\ell(s,\delta_2)}{m_1 N_s(1)}}}{m_0 + \sqrt{\frac{\ell(s,\delta_2)}{m_1(s-N_s(1))}} + m_1 - \sqrt{\frac{\ell(s,\delta_2)}{m_1 N_s(1)}}} && (47)\\
&= \pi_{\mathrm{Ney}} \cdot \frac{m_0 + m_1}{m_0 + m_1 + h(\pi_{\min}, s)\sqrt{\frac{\ell(s,\delta_2)}{s}}} - \frac{a_1(\pi_{\min}, s)\sqrt{\frac{\ell(s,\delta_1)}{s}}}{m_0 + m_1 + h(\pi_{\min}, s)\sqrt{\frac{\ell(s,\delta_2)}{s}}} && (48)\\
&= \pi_{\mathrm{Ney}} \cdot \frac{\sqrt{\frac{s}{\ell(s,\delta_2)}}\,(m_0 + m_1)}{\sqrt{\frac{s}{\ell(s,\delta_2)}}\,(m_0 + m_1) + h(\pi_{\min}, s)} - \frac{a_1(\pi_{\min}, s)}{\sqrt{\frac{s}{\ell(s,\delta_2)}}\,(m_0 + m_1) + h(\pi_{\min}, s)} && (49)\\
&= \underline{\pi}_{s+1}, && (50)
\end{aligned}$$

where we have defined

$$\begin{aligned}
a_1(\pi, s) &= \sqrt{\frac{1}{m_1\left(\pi - \sqrt{\frac{\ell(s,\delta_1)}{s}}\right)}}, && (51)\\
a_0(\pi, s) &= \sqrt{\frac{1}{m_0\left((1-\pi) - \sqrt{\frac{\ell(s,\delta_1)}{s}}\right)}}, && (52)\\
h(\pi, s) &= a_0(\pi, s) - a_1(\pi, s). && (53)
\end{aligned}$$

Using these bounds, on π_{s+1} we observe that on the event $\mathcal{E}$ we have

$$\begin{aligned}
N_t(1) \geq \sum_{s=1}^{t} \pi_s - \sqrt{t\ell(t,\delta_2)} &= \sum_{s=1}^{\tau} \pi_s + \sum_{s=\tau+1}^{t} \pi_s - \sqrt{t\ell(t,\delta_2)} && (54)\\
&\geq \frac{\tau^{1-\alpha} - 1}{2(1-\alpha)} + \sum_{s=\tau+1}^{t} \underline{\pi}_s - \sqrt{t\ell(t,\delta_2)} && (55)
\end{aligned}$$

We bound $\sum_{s=\tau+1}^{t} \underline{\pi}_s$ using Lemma B.1 so that

$$\begin{aligned}
N_t(1) \geq\ & \pi_{\mathrm{Ney}} \cdot t + \frac{\tau^{1-\alpha} - 1}{2(1-\alpha)} - \pi_{\mathrm{Ney}}(\tau - 1) \\
& - \sqrt{t\ell(t,\delta_2)}\left(2\frac{h(\pi_{\min}) + a_1(\pi_{\min}, \tau)}{m_0 + m_1} + 1\right) \\
& - \frac{a_1(\pi_{\min}, \tau)}{\sqrt{\frac{\tau}{\ell(\tau,\delta_2)}}\,(m_0 + m_1) + h(\pi_{\min}, \tau)} \\
=\ & \pi_{\mathrm{Ney}} \cdot t - c,
\end{aligned} \tag{56}$$

where we have defined $h(\pi_{\min}) = \lim_{t\to\infty} h(\pi_{\min}, t)$. By plugging this value of $N_t(1)$ into equation (30), we obtain

$$\pi_{t+1} \geq \underline{\pi}_{t+1} = \pi_{\mathrm{Ney}} \cdot \frac{\sqrt{\frac{t}{\ell(t,\delta_2)}}\,(m_0 + m_1)}{\sqrt{\frac{t}{\ell(t,\delta_2)}}\,(m_0 + m_1) + \tilde{h}(\pi_{\mathrm{Ney}}, t)} - \frac{\tilde{a}_1(\pi_{\mathrm{Ney}}, t)}{\sqrt{\frac{t}{\ell(t,\delta_2)}}\,(m_0 + m_1) + \tilde{h}(\pi_{\mathrm{Ney}}, t)} \tag{57}$$

where we have defined

$$\tilde{a}_0(\pi, t) = \sqrt{\frac{1}{m_0\left((1-\pi)+\frac{c}{t}\right)}} \tag{58}$$

$$\tilde{a}_1(\pi, t) = \sqrt{\frac{1}{m_1\left(\pi-\frac{c}{t}\right)}}, \tag{59}$$

$$\tilde{h}(\pi, t) = \tilde{a}_0(\pi, t) - \tilde{a}_1(\pi, t) \tag{60}$$

Therefore, we have

$$\begin{aligned}
\pi_{\text{Ney}} - \pi_{t+1} &\le \pi_{\text{Ney}} - \underline{\pi}_{t+1} && (61)\\
&= \sqrt{\frac{\ell(t,\delta_2)}{t}}\left(\frac{\pi_{\text{Ney}}\tilde{a}_0 + (1-\pi_{\text{Ney}})\tilde{a}_1}{m_0 + m_1 + \sqrt{\frac{\ell(t,\delta_2)}{t}}(\tilde{a}_0 - \tilde{a}_1)}\right) && (62)\\
&= \sqrt{\frac{\ell(t,\delta_2)}{t}}\left(\frac{\pi_{\text{Ney}}}{\sqrt{m_0\left((1-\pi_{\text{Ney}})+\frac{c}{t}\right)}} + \frac{1-\pi_{\text{Ney}}}{\sqrt{m_1\left(\pi_{\text{Ney}}-\frac{c}{t}\right)}}\right)\left(\frac{1}{m_0 + m_1 + \sqrt{\frac{\ell(t,\delta_2)}{t}}(\tilde{a}_0 - \tilde{a}_1)}\right) && (63)\\
&\le 4\sqrt{\frac{\ell(t,\delta)}{t}}\left(\frac{1}{\sqrt{m_0\left(1-\pi_{\text{Ney}}\right)}} + \frac{1}{\sqrt{m_1\pi_{\text{Ney}}}}\right)\left(\frac{1}{m_0+m_1}\right) && (64)
\end{aligned}$$

where the final inequality follows from the application of Lemmas B.2 and B.3 which shows that when $t \ge O(\tau)$ we have that $\frac{c}{t} \le \frac{1}{2}\pi_{\text{Ney}}$. □

Lemma A.1. *Suppose we run* ClipSMT *with* $c_t = \frac{1}{2}t^{-\alpha}$ *for some* $\alpha \in (0,1)$ *and let* $p = \min\left(\alpha, \frac{1-\alpha}{2}\right)$. *Then, on the event* $\mathcal{E}$*, for all* $t \ge 1$*, we have that* $\pi_t \in [\pi_{\min}, \pi_{\max}] = \left[\frac{1}{2}\underline{\mathfrak{T}}^{-\frac{\alpha}{2p}}, 1 - \frac{1}{2}\overline{\mathfrak{T}}^{-\frac{\alpha}{2p}}\right]$ *where* $\underline{\mathfrak{T}}, \overline{\mathfrak{T}}$ *are respectively defined in equations* (39) *and* (43).

Proof. During the clipping phase, we know that $\pi_t \in [c_t, 1-c_t]$. Additionally, once the clipping phase ends, we know that $\tilde{\pi}_t = \pi_t$ so that

$$\pi_{t+1} \in \left[\frac{m_1 - \beta_2(N_t(1),\delta_2)}{m_0 + \beta_2(N_t(0),\delta_2) + m_1 - \beta_2(N_t(1),\delta_2)}, \frac{m_1 + \beta_2(N_t(1),\delta_2)}{m_0 - \beta_2(N_t(0),\delta_2) + m_1 + \beta_2(N_t(1),\delta_2)}\right]. \tag{65}$$

It is easy to see that the above bonunds are monotonic in t—the lower bound is monotonically increasing and the upper bound is monotonically decreasing—which implies that π_t takes its minimum and maximum values at the end of the clipping phase. Therefore, we see that for all $t \ge 1$, we have that

$$1 - \frac{1}{2}\overline{\mathfrak{T}}^{-\frac{\alpha}{2p}} \ge 1 - \frac{1}{2}\tau^{-\alpha} = 1 - c_\tau \ge \pi_t \ge c_\tau = \frac{1}{2}\tau^{-\alpha} \ge \frac{1}{2}\underline{\mathfrak{T}}^{-\frac{\alpha}{2p}}, \tag{66}$$

where the first and last inequality follows from applying Lemma 4.1 which shows that $\tau \le \mathfrak{T}^{\frac{1}{2p}}$. □

B Supporting Lemmas

B.1 Intermediate Steps

Lemma B.1. *Define*

$$\underline{\pi}_{s+1} = \pi_{Ney} \cdot \frac{\sqrt{\frac{s}{\ell(s,\delta_2)}}(m_0+m_1)}{\sqrt{\frac{s}{\ell(s,\delta_2)}}(m_0+m_1) + h(\pi_{\min}, s)} - \frac{a_1(\pi_{\min}, s)}{\sqrt{\frac{s}{\ell(s,\delta_2)}}(m_0+m_1) + h(\pi_{\min}, s)}. \tag{67}$$

Then we have that

$$\sum_{s=\tau+1}^{t} \underline{\pi}_s \ge \pi_{Ney}(t-\tau-1) - 2\sqrt{t\ell(t,\delta_2)}\left(\frac{h(\pi_{\min}) + a_1(\pi_{\min},\tau)}{m_0+m_1}\right) - \frac{h(\pi_{\min},\tau)}{\sqrt{\frac{\tau}{\ell(\tau,\delta_2)}}(m_0+m_1) + h(\pi_{\min},\tau)}, \tag{68}$$

Proof. We begin by observing

$$\sum_{s=\tau+1}^{t} \underline{\pi}_s = \sum_{s=\tau}^{t-1} \underline{\pi}_{s+1} \tag{69}$$

$$= \sum_{s=\tau}^{t-1} \pi_{\mathrm{Ney}} \cdot \frac{\sqrt{\frac{s}{\ell(s,\delta_2)}}\,(m_0+m_1)}{\sqrt{\frac{s}{\ell(s,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min},s)} - \frac{a_1(\pi_{\min},s)}{\sqrt{\frac{s}{\ell(s,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min},s)} \tag{70}$$

$$\geq \pi_{\mathrm{Ney}} \cdot \underbrace{\sum_{s=\tau}^{t-1} \frac{\sqrt{\frac{s}{\ell(t,\delta_2)}}\,(m_0+m_1)}{\sqrt{\frac{s}{\ell(t,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min})}}_{\text{Term 1}} - \underbrace{\sum_{s=\tau}^{t-1} \frac{a_1(\pi_{\min},\tau)}{\sqrt{\frac{s}{\ell(t,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min},\tau)}}_{\text{Term 2}}, \tag{71}$$

where we have set

$$h(\pi) = \sqrt{\frac{1}{m_0\,(1-\pi)}} - \sqrt{\frac{1}{m_1 \pi}}$$

and the inequality follows from the monotonic properties of the map $x, y \mapsto \frac{x}{x+y}$ combined with the fact that $h(\pi, t)$ is increasing in t and $a_1(\pi, t)$ is decreasing in t. From here, we lower bound Term 1 and upper bound Term 2.

To lower bound Term 1, let $c_1 = \frac{m_0+m_1}{\sqrt{\ell(t,\delta_2)}}$ and $c_2 = h(\pi_{\min})$. Then we have

$$\text{Term 1} = \sum_{s=\tau}^{t-1} \frac{\sqrt{s}c_1}{\sqrt{s}c_1 + c_2} \tag{72}$$

$$= (t-\tau-1) - \sum_{s=\tau}^{t} \frac{c_2}{\sqrt{s}c_1 + c_2} \tag{73}$$

$$\geq (t-\tau-1) - 2\frac{c_2}{c_1}\sqrt{t} \tag{74}$$

$$= (t-\tau-1) - 2\frac{h(\pi_{\min})}{m_0+m_1}\sqrt{t\ell(t,\delta_2)} \tag{75}$$

where the inequality follows Lemma B.15.

To upper bound Term 2 we similarly apply Lemma B.15 so that

$$\text{Term 2} \leq \frac{a_1(\pi_{\min},\tau)}{\sqrt{\frac{\tau}{\ell(\tau,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min},\tau)} + 2\frac{a_1(\pi_{\min},\tau)\sqrt{t\ell(t,\delta_2)}}{m_0+m_1} \tag{76}$$

Combining these bounds shows that

$$\sum_{s=\tau+1}^{t} \underline{\pi}_s \geq \pi_{\mathrm{Ney}}(t-\tau-1) - 2\sqrt{t\ell(t,\delta_2)}\left(\frac{h(\pi_{\min})+a_1(\pi_{\min},\tau)}{m_0+m_1}\right) - \frac{h(\pi_{\min},\tau)}{\sqrt{\frac{\tau}{\ell(\tau,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min},\tau)}, \tag{77}$$

thus proving the desired result. $\square$

Lemma B.2. *Define*

$$c = \sqrt{t\ell(t,\delta)}\left(2\frac{h(\pi_{\min})+a_1(\pi_{\min},\tau)}{m_0+m_1}+1\right) + \frac{a_1(\pi_{\min},\tau)}{\sqrt{\frac{\tau}{\ell(\tau,\delta_2)}}\,(m_0+m_1)+h(\pi_{\min},\tau)} + \pi_{Ney}\,(\tau-1) - \frac{\tau^{1-\alpha}-1}{2\,(1-\alpha)} \tag{78}$$

If $t \geq 6\tau$, then $\frac{c}{t} \leq \frac{1}{2}\pi_{Ney}$.

Proof. Note that it is sufficient to bound the first three terms since we are subtracting the fourth term.

For the first term, we observe that when $\tau^{1-2\alpha} \geq 16\ell(\tau, \delta)$, which is satisfied by our definition of τ, we have that $a_1(\pi_{\min}, \tau) \leq \sqrt{\frac{2}{m_1 \pi_{\min}}}$. Therefore, some algebra shows that

$$2\frac{h(\pi_{\min}) + a_1(\pi_{\min}, \tau)}{m_0 + m_1} + 1 \leq 2\tau^{\frac{\alpha}{2}} \left(\frac{1}{m_0} + \frac{1}{m_1}\right)\left(\frac{1}{\sqrt{m_0}} + \frac{1}{\sqrt{m_1}}\right) + 1 = c_1 \tag{79}$$

Therefore if we want bound this term by $b_1 \pi_{\mathrm{Ney}}$, we require $t \geq \frac{c_1^2}{b_1^2}\ell(t, \delta)$. We can apply Lemma B.10 to bound this.

Next, some algebra shows that when $\tau^{1-\alpha} \geq \frac{3}{m_1(m_0+m_1)^2}\ell(\tau, \delta)$, which is satisfied by our definition of τ, we have that

$$\frac{a_1(\pi_{\min}, \tau)}{\sqrt{\frac{\tau}{\ell(\tau, \delta_2)}}\,(m_0 + m_1) + h(\pi_{\min}, \tau)} \leq 2. \tag{80}$$

As such, if we want to bound this term by b_2, we require $t \geq \frac{2}{b_2 \pi_{\mathrm{Ney}}}$

Finally, to bound the third term by $b_3 \pi_{\mathrm{Ney}}$, we observe that we require $t \geq \frac{\tau - 1}{b_2}$.

Setting $b_1 = b_2 = b_3 = \frac{1}{6}$ and $c_1 = \frac{1}{2}$, and using the above results, see that when

$$t \geq \max\left\{144\ell(t, \delta), 6\left(\tau - 1\right), \frac{12}{\pi_{\mathrm{Ney}}}\right\} \tag{81}$$

we have that $\frac{c}{t} \leq \frac{1}{2}\pi_{\mathrm{Ney}}$. □

Lemma B.3. *Suppose $t \geq 6\tau$. Then we have that*

$$\sqrt{\frac{\ell(t, \delta)}{t}}\left(\tilde{a}_1 - \tilde{a}_0\right) \leq \frac{1}{2}\left(m_0 + m_1\right). \tag{82}$$

Proof. To being, we see that it is sufficient to find compute an upper bound on the smallest t such that

$$t \geq \ell(t, \delta)\frac{\tilde{a}_1^2}{m_0 + m_1}.$$

Next, we apply Lemma B.2 which shows that when $t \geq 6\tau$, we have that $\tilde{a}_1^2 \leq \frac{2}{m_1 \pi_{\mathrm{Ney}}}$. Plugging this in and applying Lemma B.10 gives the desired result. □

B.2 Useful Tools

Lemma B.4. *We have that*

$$\frac{m_1 - \sqrt{\frac{\ell(t,\delta)}{m_1 N_t(1)}}}{m_0 + \sqrt{\frac{\ell(t,\delta)}{m_0(t - N_t(1))}} + m_1 - \sqrt{\frac{\ell(t,\delta)}{m_1 N_t(1)}}} \tag{83}$$

$$= \pi_{Ney}\frac{\sqrt{\frac{t}{\ell(t,\delta)}}\,(m_0 + m_1)}{\sqrt{\frac{t}{\ell(t,\delta)}}\,(m_0 + m_1) + a_0(t, N_t(1)) - a_1(t, N_t(1))} \tag{84}$$

$$- \frac{a_1(t, N_t(1))}{\sqrt{\frac{t}{\ell(t,\delta)}}\,(m_0 + m_1) + a_0(t, N_t(1)) - a_1(t, N_t(1))} \tag{85}$$

where

$$a_0(t, n) = \sqrt{\frac{1}{m_0\left(1 - \frac{n}{t}\right)}}, \tag{86}$$

$$a_0(t, n) = \sqrt{\frac{1}{m_1 \cdot \frac{n}{t}}}. \tag{87}$$

B.3 Concentration Results

Lemma B.5. *Let $X_1, X_2, \ldots$ be a sequence of random variables such that $X_t \sim \text{Bernoulli}(\pi_t)$ where π_t is $\mathcal{F}_{t-1}$ measurable and define $N_t = \sum_{s=1}^t X_s$. Then, with probability at least $1-\delta$, the following holds for all $t \in \mathbb{N}$*

$$\left| N_t - \sum_{s=1}^t \pi_t \right| \leq \beta_1(t, \delta), \tag{88}$$

where

$$\beta_1(t, \delta) = 0.85 \sqrt{t \left(\log \log t + 0.72 \log \left(\frac{5.2}{\delta} \right) \right)}. \tag{89}$$

Proof. Define $M_t^\lambda = \exp\left(\lambda(X - p_t) - \frac{\lambda^2}{8}\right)$. Note that by definition, $X_t \in [0,1]$ almost surely with $\mathbb{E}\left[X_t \mid \mathcal{F}_{t-1}\right] = p_t$ which implies that the following holds for every $\lambda \in \mathbb{R}$

$$\mathbb{E}\left[M_t^\lambda \mid \mathcal{F}_{t-1}\right] \leq 1. \tag{90}$$

Therefore, $D_t^\lambda = \prod_{s=1}^t M_s^\lambda$ is a test supermartingale and we can apply Theorem 1 from Howard et al. (2021) (see equation (11)) to obtain the desired result. ☐

Lemma B.6. *Let $X_1, X_2, \ldots$ be a sequence of random variables such that $X_t \in [0,1]$, $\mu = \mathbb{E}\left[X_t \mid \mathcal{F}_{t-1}\right]$, and $m^2 = \mathbb{E}\left[X_t^2 \mid \mathcal{F}_{t-1}\right]$. Define the empirical second moment as $\widehat{m}_t^2 = \frac{1}{t} \sum_{s=1}^t X_s^2$. Then, with probability at least $1-\delta$, the following holds for all $t \in \mathbb{N}$*

$$|\widehat{m}_t - m| \leq \beta_2(t, \delta) \tag{91}$$

where

$$\beta_2(t, \delta) = 0.85 \sqrt{\frac{\left(\log \log t + 0.72 \log\left(\frac{5.2}{\delta}\right)\right)}{m^2 \cdot t}}. \tag{92}$$

Proof. To see this, we first observe that

$$|\widehat{m}_t - m| = \frac{\left|\widehat{m}_t^2 - m^2\right|}{|\widehat{m}_t + m|} \leq \frac{\left|\widehat{m}_t^2 - m^2\right|}{\left|\sqrt{m^2}\right|}.$$

The result then follows by bounding $\left|\widehat{m}_t^2 - m^2\right|$ by applying Theorem 1 from Howard et al. (2021) (see equation (11)). ☐

Remark B.7. *Note that in our above result, the width of the confidence sequences scale like $O\left(\frac{1}{\sqrt{m^2 \cdot t}}\right)$. An application of the CLT along with the Delta Method shows that, asymptotically, the scaling with respect to $\frac{1}{\sqrt{m^2}}$ is unavoidable.*

B.4 Technical Results

Lemma B.8. *Let $t, \alpha_0, \alpha_1, \gamma_0, \gamma_1 > 0$ be fixed, and define the function $f : (0, t) \to \mathbb{R}$ by*

$$f(x) = \frac{\alpha_1 - \frac{\gamma_1}{\sqrt{x}}}{\alpha_0 + \frac{\gamma_0}{\sqrt{t-x}} + \alpha_1 - \frac{\gamma_1}{\sqrt{x}}}. \tag{93}$$

Given an interval $[s, r] \subseteq [1, t]$, any solution $x^\star$ to the optimization problem

$$\min_{x \in [s,r]} f(x), \tag{94}$$

must satisfy $x^\star \in \{s, r\}$.

Proof. Our proof will proceed by demonstrating that one of the preconditions of Lemma B.9 is satisfied, from which the desired result naturally follows. To begin, we let $f'(x) = \frac{d}{dx}f(x)$ denote the derivative of $f(x)$. We compute $f'(x)$ and perform some simplifications to show that

$$\begin{aligned} f'(x) &= -\left(\frac{\left(\frac{\gamma_0}{2(t-x)^{3/2}} + \frac{\gamma_1}{2x^{3/2}}\right)(\alpha_1 - \frac{\gamma_1}{\sqrt{x}})}{(\alpha_0 + \alpha_1 + \frac{\gamma_0}{\sqrt{t-x}} - \frac{\gamma_1}{\sqrt{x}})^2}\right) + \frac{\gamma_1}{2(\alpha_0 + \alpha_1 + \frac{\gamma_0}{\sqrt{t-x}} - \frac{\gamma_1}{\sqrt{x}})x^{3/2}} \\ &= \frac{(\gamma_0\gamma_1 t + \alpha_0\gamma_1 t\sqrt{t-x} - \alpha_0\gamma_1\sqrt{t-x}x - \alpha_1\gamma_0 x^{3/2})}{2(-\gamma_1\sqrt{t-x} + \gamma_0\sqrt{x} + \alpha_0\sqrt{t-x}\sqrt{x} + \alpha_1\sqrt{t-x}\sqrt{x})^2\sqrt{t-x}\sqrt{x}}. \end{aligned} \tag{95}$$

Observe that the denominator in (95) is always greater than zero. Therefore, $\text{sign}(f'(x))$ is determined by the numerator which we will now show to be strictly decreasing. The derivative of the numerator in (95) is

$$-\left(\frac{3(\alpha_0\gamma_1 t + \alpha_1\gamma_0\sqrt{t-x}\sqrt{x} - \alpha_0\gamma_1 x)}{2\sqrt{t-x}}\right).$$

From here, we have that by assumption $\alpha_0, \alpha_1, \gamma_0, \gamma_1 > 0$ and $x < t$ imply that the above quantity is strictly negative. Since the derivative of the numerator is strictly negative, we know that the numerator is strictly decreasing. Therefore, our earlier observation, in conjunction with this fact implies that one of the preconditions of Lemma B.9 must hold, thus enabling its application, which in turn implies the desired result. □

The next lemma essentially shows that the minimum of a concave-unimodal function over a closed interval must occur at one of the boundaries of the interval.

Lemma B.9. *Let $f : \mathcal{D} \to \mathbb{R}$ be any differential function such that its derivative, f', satisfies one of the following conditions:*

1. *$f'(x) > 0$ for all $x \in \mathcal{D}$*
2. *$f'(x) < 0$ for all $x \in \mathcal{D}$*
3. *There exists c such that for all $x < c$, $f'(x) > 0$ and for all $x > c$, $f'(x) < c$.*

Then for any $[a, b] \subset \mathcal{D}$, any solution $x^\star$ to optimization problem,

$$\min_{x\in[a,b]} f(x), \tag{96}$$

must satisfy $x^\star \in \{a, b\}$.

Proof. If $f'(x) > 0$ for all $x \in \mathcal{D}$, the function is monotonically increasing and the minimum will occur at $x^\star = a$. If $f'(x) < 0$ for all $x \in \mathcal{D}$, the function is monotonically decreasing and the minimum will occur at $x^\star = b$. For the final case, let c be as defined in the condition and let $\tilde{x}$ denote the minimum of f. If $a < \tilde{x} < c$ then $f(\tilde{x}) - f(a) = \int_a^{\tilde{x}} f'(t)dt > 0$ which is a contradiction. Similarly if $b > \tilde{x} > c$, then $f(b) - f(\tilde{x}) = \int_{\tilde{x}}^b f'(c)dc < 0$ which is also a contradiction. Therefore, for each of the cases, $x^\star$ must satisfy $x^\star \in \{a, b\}$. □

Lemma B.10. *Let $c_1, c_2, p > 0$ such that $\log c_1 > p$ and $c_1 \log c_1 > c_2$ and define*

$$\tau = \min\left\{t : t^p \geq c_1 + c_2 \log\log(t)\right\}. \tag{97}$$

We have that

$$\tau \leq \left(c_1 + c_2 \log\left(\log c_1\right) \frac{\log\log c_1 - \log(p)}{\log\log c_1} \cdot \frac{c_1 \log c_1}{c_1 \log c_1 - c_2}\right)^{\frac{1}{p}} \tag{98}$$

Proof. To prove this, we set

$$t = (c_1 + ac_2 \log\log c_1)^{\frac{1}{p}},$$

for some a to be chosen later. Our objective is to show that

$$\log\log\left[(c_1 + ac_2 \log\log c_1)^{\frac{1}{p}}\right] \leq a \log\log c_1.$$

To do so, we observe that

$$\begin{aligned}
&\log\left(\log\left((c_1 + ac_2 \log\log c_1)^{\frac{1}{p}}\right)\right) \\
&= \log\left(\frac{1}{p}\log\left(c_1 + ac_2 \log\log c_1\right)\right) \\
&= \log\left(\frac{1}{p}\left(\log(c_1) + \log\left(1 + \frac{ac_2}{c_1}\log\log c_1\right)\right)\right) \\
&\le \log\left(\frac{1}{p}\left(\log(c_1) + \frac{ac_2}{c_1}\log\log c_1\right)\right) \\
&= \log\left(\frac{1}{p}\log c_1\right) + \log\left(1 + \frac{ac_2}{c_1\log c_1}\log\log c_1\right) \\
&\le \log\left(\frac{1}{p}\log c_1\right) + \frac{ac_2}{c_1 \log c_1}\log\log c_1,
\end{aligned}$$

where the inequalities follow from applying the inequality $\log(1+x) \le x$. From here, we set a so the final line above equals $a \log\log c_1$. In particular, by setting

$$a = \frac{\log\log c_1 - \log(p)}{\log\log c_1} \cdot \frac{c_1 \log c_1}{c_1 \log c_1 - c_2},$$

the above series of inequalities proves that

$$\log\log\left[(c_1 + ac_2\log\log c_1)^{\frac{1}{p}}\right] \le a \log\log c_1,$$

as desired. □

Lemma B.11. *Fix $\alpha, \delta \in (0,1)$ and consider the function*

$$f(t,\delta,\alpha) = 1 + t^{-\frac{1}{2}}\sqrt{\ell(t,\delta)} + \frac{t^{-1} - t^{-\alpha}}{1-\alpha}.$$

For all $t \ge \left(\frac{2}{1-\alpha}\right)^{\frac{1}{\alpha}}$, we have that $g(t,\delta,\alpha) \ge \frac{1}{2}$.

Proof. First note that

$$1 + t^{-\frac{1}{2}}\sqrt{\ell(t,\delta)} + \frac{t^{-1} - t^{-\alpha}}{1-\alpha} \ge 1 - \frac{t^{-\alpha}}{1-\alpha}.$$

Solving the inequality

$$1 - \frac{t^{-\alpha}}{1-\alpha} \ge \frac{1}{2},$$

for t gives the desired result. □

Corollary B.12. *For $\delta \in (0,1)$, $t \ge 27$ implies that $f(t,\delta,\frac{1}{3}) \ge \frac{1}{2}$.*

Lemma B.13. *Fix $\alpha \in \left(0, \frac{1}{2}\right)$, $\delta \in (0,1)$, and let*

$$g(t,\delta,\alpha) = \frac{1 - t^{\alpha-1}}{2\,(1-\alpha)} - t^{\frac{2\alpha-1}{2}}\sqrt{\ell(t,\delta)}.$$

We have that $g(t,\delta,\alpha) \ge \frac{1}{2}$ whenever

$$t \ge \left(c_1 + c_2 \log\left(\log c_1\right) \frac{\log\log c_1 - \log(1-2\alpha)}{\log\log c_1} \cdot \frac{c_1 \log c_1}{c_1 \log c_1 - c_2}\right)^{\frac{1}{1-2\alpha}},$$

where $c_1 = \frac{2}{\alpha^2} + \frac{8(1-\alpha)^2}{\alpha^2}\log\left(\frac{5.2}{\delta}\right)$ and $c_2 = \frac{8(1-\alpha)^2}{\alpha^2}$.

Proof. To begin, observe that

$$\frac{1-t^{\alpha-1}}{2\,(1-\alpha)} - t^{\frac{2\alpha-1}{2}}\sqrt{\ell(t,\delta)} \geq \frac{1}{2\,(1-\alpha)} - t^{\frac{2\alpha-1}{2}}\sqrt{\ell(t,\delta)} - \frac{t^{\frac{2\alpha-1}{2}}}{2\,(1-\alpha)}, \tag{99}$$

therefore it is sufficient to bound the quantity

$$\min\left\{t : \frac{1}{2\,(1-\alpha)} - t^{\frac{2\alpha-1}{2}}\left(\sqrt{\ell(t,\delta)} + \frac{1}{2\,(1-\alpha)}\right) \geq \frac{1}{2}\right\}. \tag{100}$$

Rearranging, we see that this is equivalent to bounding the quantity

$$\min\left\{t : t^{\frac{1}{2}-\alpha} \geq \frac{2\,(1-\alpha)}{\alpha}\sqrt{\ell(t,\delta)} + \frac{1}{\alpha}\right\}. \tag{101}$$

By squaring both sides and applying the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ we see that it is sufficient to bound

$$\min\left\{t : t^{1-2\alpha} \geq \frac{2}{\alpha^2} + \frac{8\,(1-\alpha)^2}{\alpha^2}\ell(t,\delta)\right\}. \tag{102}$$

Setting $c_1 = \frac{2}{\alpha^2} + \frac{8(1-\alpha)^2}{\alpha^2}\log\left(\frac{5.2}{\delta}\right)$ and $c_2 = \frac{8(1-\alpha)^2}{\alpha^2}$ we can apply Lemma B.10 to see that whenever

$$t \geq \left(c_1 + c_2\log\left(\log c_1\right)\frac{\log\log c_1 - \log(1-2\alpha)}{\log\log c_1}\cdot\frac{c_1\log c_1}{c_1\log c_1 - c_2}\right)^{\frac{1}{1-2\alpha}},$$

we have that $g(t,\delta,\alpha) \geq \frac{1}{2}$, as desired. □

Corollary B.14. *For $\delta \in (0,1)$, $t \geq O\left(\log\left(\frac{1}{\delta}\right)^3\right)$ implies that $g\left(t,\delta,\frac{1}{3}\right) \geq \frac{1}{2}$.*

Lemma B.15. *Fix c_1, c_2, c_3, τ, t such that $c_1, c_2 > 0$, $\tau < t$, and $c_2\sqrt{\tau} + c_3 > 0$. Then, we have that*

$$\sum_{s=\tau}^{t}\frac{c_1}{c_2\sqrt{s}+c_3} \leq \frac{c_1}{c_2\sqrt{\tau}+c_3} + \frac{2c_1}{c_2}\left(\sqrt{t}-\sqrt{\tau}\right) \tag{103}$$

Proof. Observe that under the stated conditions, we have that $\frac{c_1}{c_2\sqrt{s}+c_3}$ is monotonically decreasing in s. Therefore we can bound

$$\begin{aligned}
\sum_{s=\tau}^{t}\frac{c_1}{c_2\sqrt{s}+c_3} &\leq \frac{c_1}{c_2\sqrt{\tau}+c_3} + \int_{s=\tau}^{t}\frac{c_1}{c_2\sqrt{s}+c_3}ds \\
&\leq \frac{c_1}{c_2\sqrt{\tau}+c_3} + \left(\frac{2c_1}{c_2}\sqrt{s} - \frac{2c_1c_3}{c_2^2}\log\left(c_3 + c_2\sqrt{s}\right)\right)\Bigg|_{s=\tau}^{t} \\
&= \frac{c_1}{c_2\sqrt{\tau}+c_3} + \left(\frac{2c_1}{c_2}\sqrt{t} - \frac{2c_1c_3}{c_2^2}\log\left(c_3 + c_2\sqrt{t}\right)\right) - \left(\frac{2c_1}{c_2}\sqrt{\tau} - \frac{2c_1c_3}{c_2^2}\log\left(c_3 + c_2\sqrt{\tau}\right)\right) \\
&\leq \frac{c_1}{c_2\sqrt{\tau}+c_3} + \frac{2c_1}{c_2}\left(\sqrt{t}-\sqrt{\tau}\right)
\end{aligned}$$

□

C Discussion on Clipping Sequences

Recall that our proposed ClipSMT algorithm utilizes clipping sequence with polynomial decay so that $c_t = \frac{1}{2}t^{-\alpha}$ for $\alpha \in (0,1)$. It is natural to wonder if there are other valid choices for the clipping sequence. While there are, the choices of clipping sequences that will work depend on the assumptions that we make.

On one hand, if we do not assume a lower bound on m_A^2, then we must require that $\sum_t c_t$ diverges as $t \to \infty$. To see why, suppose the sum converges, i.e $\lim_{T\to\infty}\sum_{t=1}^{T} c_t = c$. Then, if we choose m_1^2, m_2^2 so that the length of the clipping phase is larger than c, this will ensure that π_t never converges to π_{Ney}. As a concrete example, this implies that in this most general setting, we should not use clipping sequence with exponential decay. However, if we are willing to assume a lower bound on m_0^2, m_1^2, then we can use a similar argument in order to select the rate of decay for a clipping sequence whose sum converges.

D Additional Figures from Experiments

This section contains additional plots not included in the main paper due to space constrains. We refer the reader to Section 5 for a more in-depth discussion on the experiments.

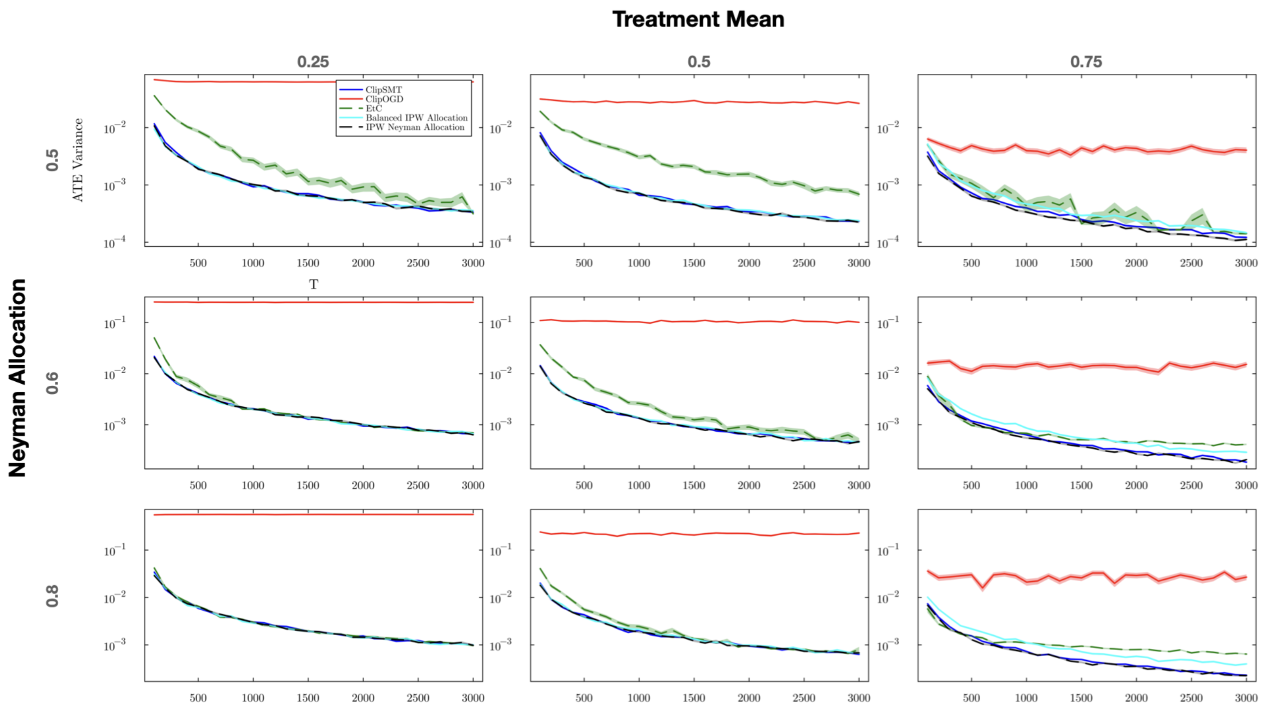
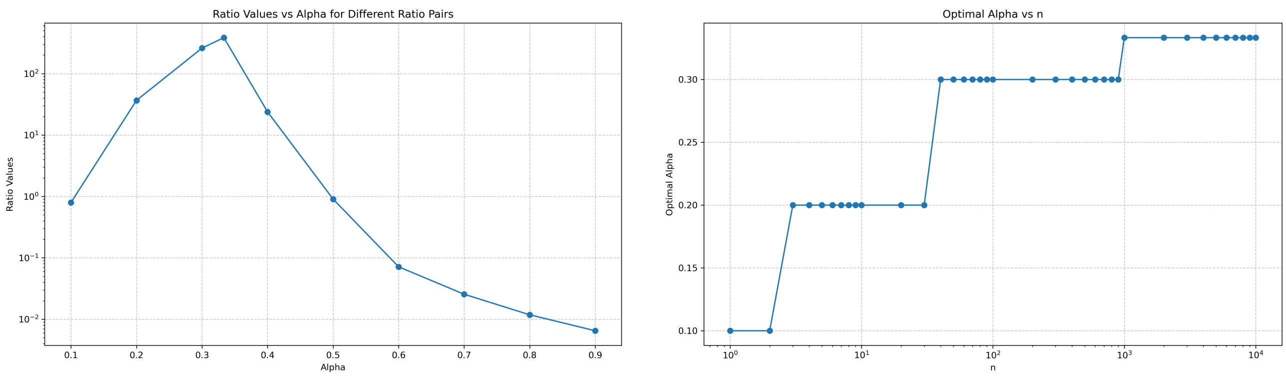


Figure 2: Comparison of the performance of ClipSMT, ClipOGD, Explore-then-Commit (EtC), Neyman allocation, and a balanced allocation with the treatment and control arms following Bernoulli distributions in the small sample regime. Notably, ClipSMT is competitive with the Oracle Neyman Allocation even for small sample sizes, indicating its practical utility.



(a) (b)

Figure 3: The figure on the left plots the optimal ratio for each problem instance, where the problems get harder as n increases. The figure on the left plots the ratio of the predicted versus empirically computed clipping times. Note that a smaller value implies our theory underestimates the empirical clipping time, implying that the true clipping times are larger.