NeurIPS2023 Workshop on Adaptive Experimental Design and Active Learning in the Real World

Semiparametric Efficient Inference in Adaptive Experiments

Thomas Cook[*] TJCOOK@UMASS.EDU
Department of Mathematics and Statistics
University of Massachusetts

Alan Mishler ALAN.MISHLER@JPMCHASE.COM
J.P. Morgan AI Research
J.P. Morgan Chase & Co.

Aaditya Ramdas ARAMDAS@CMU.EDU
Department of Statistics & Data Science and Machine Learning
Carnegie Mellon University

Abstract

We consider the problem of efficient inference of the *Average Treatment Effect* in a sequential experiment where the policy governing the assignment of subjects to treatment or control can change over time. We first provide a central limit theorem for the Adaptive Augmented Inverse-Probability Weighted estimator, which is semiparametric efficient, under weaker assumptions than those previously made in the literature. This central limit theorem enables efficient inference at fixed sample sizes. We then consider a sequential inference setting, deriving both asymptotic and nonasymptotic confidence sequences that are considerably tighter than previous methods. These *anytime-valid* methods enable inference under data-dependent stopping times (sample sizes). Additionally, we use propensity score truncation techniques from the recent off-policy estimation literature to reduce the finite sample variance of our estimator without affecting the asymptotic variance. Empirical results demonstrate that our methods yield narrower confidence sequences than those previously developed in the literature while maintaining time-uniform error control.

Keywords: Average Treatment Effect, Anytime-valid Inference, Confidence Sequences

1. Introduction

A/B tests, aka randomized experiments with two treatment arms, are widely used across many domains. Classical statistical tools (*fixed-time* methods) require the analyst to fix the experimental design and select the sample size in advance and only perform inference when this sample size is reached. However, modern A/B testing platforms enable continuous monitoring, which allows analysts to make repeated decisions about whether to stop or continue or modify an experiment based on the data observed so far. For example, an analyst might decide to run an experiment precisely until a test statistic becomes statistically significant, at which point they may stop and declare a treatment effective.

Statistical tools which enable valid inference in this setting are called *anytime-valid*. To illustrate the distinction, consider a confidence interval (CI) for a parameter of interest θ.

[*] Some of this work was performed while at J.P. Morgan Chase & Co.

A $(1-\alpha)$ CI for θ is an interval $[L_t, U_t]$ based on a sample of size t with the property that

$$\forall t \in \mathbb{N}^+, \mathbb{P}(\theta \in [L_t, U_t] \geq 1-\alpha. \tag{1}$$

The coverage guarantee in (1) only holds when the sample size (aka time) t is fixed in advance. By contrast, a *confidence sequence* (CS) for θ is a sequence of intervals such that

$$\mathbb{P}(\forall t \in \mathbb{N}^+, \theta \in [L_t, U_t]) \geq 1-\alpha. \tag{2}$$

The coverage guarantee in (2) is uniform in t, enabling valid inference at any t.

We consider inference for the Average Treatment Effect (ATE), which is the expected difference in outcomes between the two treatment arms, in the context of adaptive experiments. Kato et al. (2021) introduced the Adaptive Augmented IPW (A2IPW) estimator, which, when coupled with a particular adaptive design, yields asymptotically efficient CIs based on the central limit theorem (CLT) (Hahn et al., 2011). Furthermore, Kato et al. (2021) showed under certain conditions their adaptive design improves the regret bound compared to a non-adaptive design. They also provided a CS for the ATE using concentration inequalities based on nonasymptotic variants of the law of the iterated logarithm (LIL). In an independent line of research, Dai et al. (2023) proposed an experimental design such that the variance of an adaptive IPW estimator asymptotically achieves the variance under the optimal Neyman allocation, and provide an asymptotically-valid CI for the ATE.

Our contributions are both theoretical and empirical. Theoretically, we prove a CLT for the A2IPW estimator under weaker assumptions than those utilized by Kato et al. (2021), enabling approximately valid inference at fixed sample sizes. While these results are valid for arbitrary adaptive designs (with some mild restrictions), we propose a design which adaptively truncates the treatment assignment probabilities for finite sample stability (Waudby-Smith et al., 2024). We show that this estimator is semiparametric efficient when paired with the proposed design. Empirically, we couple the A2IPW estimator with anytime-valid methods based on test (super)martingales (Waudby-Smith and Ramdas, 2023) and asymptotic CSs (Waudby-Smith et al., 2023) which yield much tighter intervals (more powerful inference) than Kato's employed methods.

2. Problem Setting and Technical Preliminaries

2.1 Experimental Process

We follow the same problem setting and data generating process as described in Kato et al. (2021), with minor modifications to their notation. Subjects are indexed by $t \in \mathbb{N}$ and arrive sequentially. For each subject, the experimenter observes a context $X_t \in \mathcal{X}$, where $\mathcal{X}$ is the context domain, then assigns a treatment $A_t \in \{0, 1\}$, and then observes an outcome $Y_t \in \mathbb{R}$. We denote by $Y_t(a)$ the potential outcome corresponding to treatment a, for $a \in \{0, 1\}$, and we assume that $Y_t = \mathbb{1}[A_t = 0]Y_t(0) + \mathbb{1}[A_t = 1]Y_t(1)$, where $\mathbb{1}[\cdot]$ denotes the indicator function. That is, we assume that a given subject's outcome depends only on their own treatment assignment and not on the treatment assignments of other subjects (Rubin, 1980, 1986). The accumulated data after T subjects (equivalently, T time steps, where $T \in \{\mathbb{N} \cup \infty\}$) consists of a set $\{(X_t, A_t, Y_t)\}_{t=1}^{T}$, whose distribution is given by

$$(X_t, A_t, Y_t) \sim p(x)\pi_t(a|x, \Omega_{t-1})p(y|a, x),$$

where $\Omega_{t-1} = \{(X_s, A_s, Y_s) : s \leq t-1\}$ denotes the *history.* We denote the domain of Ω_{t-1} by $\mathcal{M}_{t-1}$. We assume that $\{X_t, Y_t(0), Y_t(1)\}_{t=1}^T$ are independent and identically distributed. However, our treatment assignments are not fixed over time, and depend on previous observations. We define the propensity score, $\pi_t(a|x, \Omega_{t-1})$ from the experimenter's *policy*, $\pi_t : \mathcal{A} \times \mathcal{X} \times \mathcal{M}_{t-1} \mapsto [0,1]$. By introducing dependence in the policy, the observed outcomes, $\{Y_t\}_{t=1}^T$, form a sequence of realizations of dependent random variables.

As data collection may be costly, time consuming, or high risk, the experimenter may not want to continue until some predetermined sample size. Conversely, an experimenter may reach this sample size and consider proceeding with further data collection. Such practice requires methods which can handle peeking (Ramdas et al., 2023), and is the focus of Section 3. Under the anytime-valid inference methods described in that section, the experimenter can choose to stop the experiment, continue under the current policy, or continue under a modified policy, without inflating the type-I error rate.

Additional notation: Our notation follows Kato et al. (2021) with minor modification. Let a be an action in $\mathcal{A}$. Let us denote $\mathbb{E}[Y_t(a) \mid x]$, $\mathbb{E}[Y_t^2(a) \mid x]$, $\mathrm{Var}(Y_t(a) \mid x)$, and $\mathbb{E}[Y_t(1) - Y_t(0) \mid x]$ as $f(a,x)$, $e(a,x)$, $v(a,x)$, and $\theta_0(x)$, respectively. Let $\hat{f}_t(a,x)$ and $\hat{e}_t(a,x)$ denote estimators of $f(a,x)$ and $e(a,x)$ constructed from Ω_t, respectively.[1] We denote the ℓ_2 norm of a function as $\|f\|_2^2 = \int \{f(x)\}^2 \, d\mathbb{P}(x)$.

Adaptive Estimator: We denote the causal parameter of interest, the ATE, as $\theta_0 = \mathbb{E}(Y(1) - Y(0))$, where the subscript t is dropped to emphasize time invariance. In an experimental setting the treatment probabilities are known and the Inverse-Probability Weighted (IPW) estimator produces an unbiased estimate of θ_0. The Augmented IPW (AIPW) extends the IPW estimator to include regression estimates, which can reduce the variance of the estimator, while maintaining unbiasedness (Robins et al., 1994; Chernozhukov et al., 2018). Kato et al. (2021) extended the AIPW estimator to the setting of an adaptive experiment by defining the *Adaptive* AIPW estimator (A2IPW). The key difference between the two estimators is the use of data-dependent propensity scores. The A2IPW estimator, given that T subjects have been observed, is defined as $\hat{\theta}_T^{\mathrm{A2IPW}} = \frac{1}{T}\sum_{t=1}^T h_t$, where

$$h_t = \left(\frac{\mathbb{1}[A_t = 1](Y_t - \hat{f}_{t-1}(1, X_t))}{\pi_t(1|X_t, \Omega_{t-1})} - \frac{\mathbb{1}[A_t = 0](Y_t - \hat{f}_{t-1}(0, X_t))}{\pi_t(0|X_t, \Omega_{t-1})} + \hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) \right).$$

Hahn et al. (2011) showed that the policy π^{AIPW} minimizes the semiparametric lower bound of the asymptotic variance for regular estimators of the ATE, where

$$\pi^{\mathrm{AIPW}}(1|X_t) = \frac{\sqrt{v(1, X_t)}}{\sqrt{v(1, X_t)} + \sqrt{v(0, X_t)}}.$$

This policy depends on unknown quantities of the underlying data generating process. Kato et al. (2021) suggested estimating the unknown quantities. We define this policy as

$$\pi_t^{\mathrm{A2IPW}}(1|X_t, \Omega_{t-1}) = \left(\frac{\sqrt{\hat{v}(1, X_t)}}{\sqrt{\hat{v}(1, X_t)} + \sqrt{\hat{v}(0, X_t)}} \vee \frac{1}{k_t} \right) \wedge \left(1 - \frac{1}{k_t} \right), \tag{3}$$

[1] In general, $\hat{f}$ can be any arbitrary estimator. In Theorem 1, we simply require $\hat{f}$ to be consistent for f.

where $k_t \in [2, \infty)$ is a user-chosen[2] truncation parameter. Note that setting $k_t = 2$ results in $\pi_t(A_t = 0|X_t, \Omega_{t-1}) = \pi_t(A_t = 1|X_t, \Omega_{t-1}) = 0.5$, and $k_t \to \infty$ results in the policy given in Kato et al. (2021). With pointwise convergence assumptions on $\hat{f}$ and π_t, Kato et al. (2021) show that the A2IPW estimator achieves the semiparametric lower bound of the asymptotic variance. The policy truncation that we utilize is inspired by Waudby-Smith et al. (2024) where truncation circumvents required knowledge of the maximal importance weight in off-policy evaluation. Empirical results show truncation can improve finite-sample performance for well-chosen k_t.

2.2 Fixed-Time Confidence Intervals

We now turn to constructing CIs with asymptotic coverage guarantees. Kato et al. (2021) defined $z_t = h_t - \theta_0$ and showed that $\{z_t\}_{t=1}^T$ forms a *martingale difference sequence* (MDS). They then utilized a MDS CLT to show $\hat{\theta}^{\text{A2IPW}}$ is asymptotically Gaussian. They further showed that $\hat{\theta}^{\text{A2IPW}}$ is semiparametric efficient under the asymptotic policy. We provide the same results under weaker assumptions, as elaborated after the theorem.

Theorem 1 (Asymptotic Distribution of $\hat{\theta}_T^{\text{A2IPW}}$) *Assume $\{(X_t, A_t, Y_t)\}_{t=1}^T$ follow the data generating process described in Section 2.1. Let $\pi_t : \mathcal{A} \times \mathcal{X} \mapsto (0, 1)$ be an arbitrary sequence of truncated policies. Assume $1/\pi(a \mid x) < \infty$ and $v(a, x) < \infty$ for all $x \in \mathcal{X}$ and $a \in \{0, 1\}$. Further, assume $\text{Var}(Y_t) < \infty$, $k_t\|\hat{f}_t - f\|_2 = o_{\mathbb{P}}(1)$, and $k_t\|\pi_t - \pi\|_2 = o_{\mathbb{P}}(1)$ for some policy π. Under these assumptions we have*

$$\sqrt{T}(\hat{\theta}_T^{\text{A2IPW}} - \theta_0) \xrightarrow{d} N(0, \sigma^2),$$

where σ^2 is the semiparametric lower bound of the asymptotic variance for regular estimators of θ_0 under the policy π. If we have $\pi = \pi^{\text{AIPW}}$, then $\hat{\theta}_T^{\text{A2IPW}}$ is semiparametric efficient.

Details and proof are provided in Appendix A. Note that Kato et al. (2021) assumed that Y_t and $\hat{f}_t$ are uniformly bounded, that π_t is uniformly bounded away from 0, and that $\hat{f}_t$ and π_t converge pointwise. By contrast, we only assume that Y_t has finite (conditional) variance; we only assume that $\hat{f}_t$ and π_t converge in ℓ_2 norm; and, by utilizing truncation, we avoid the assumption that the policies π_t are uniformly bounded away from 0. Our proof uses a MDS CLT given by Dvoretzky (1972), which is used in a similar fashion by Zhang et al. (2021). This form of a MDS CLT allows weaker assumptions on $\hat{f}$ and π_t.

A t-statistic, along with an explicit CI, are defined in Appendix F.2. Although our interval is the same as the one given in Kato et al. (2021), our relaxed assumptions make its use applicable in more general settings, such as when outcomes have unbounded support.

3. Anytime-Valid Inference in Adaptive Experiments

We now construct CSs for the ATE. Kato et al. (2021) leveraged anytime-valid inference via concentration inequalities based on the law of the iterated logarithm (LIL) which are derived in Balsubramani (2015) and Balsubramani and Ramdas (2016). The concentration

[2] For our CLT-based CI (Theorem 1), we require k_t not to grow too quickly. For details see Appendix A.

inequality derived for $\hat{\theta}^{\text{A2IPW}}$ (Kato et al., 2021, Thm. 4) depends on the unknown treatment effect θ_0, although we believe it is probably a trivial extension to replace this with an estimate of θ_0. Indeed, though their derivation uses the true value θ_0, their experiments use a running estimate for θ_0 based on $\{h_t\}_{t=1}^{T-1}$. The theorems that follow derive CSs for the ATE based on recent, state-of-the-art methods for inference of means of random variables in sequential settings. All sequences are *fully empirical*, meaning they do not depend on unknown parameters. We will see that these these methods empirically yield much tighter intervals than methods based on the LIL.

3.1 Betting Confidence Sequences

We first derive a CS using results from Waudby-Smith and Ramdas (2023) and Waudby-Smith et al. (2024). Since these CSs do not require independence between observations, their use in the setting of adaptive experimentation is natural. The approach is based on a set of *capital processes*, each of which can be understood as the wealth a gambler would accumulate in a game against nature. More precisely, we construct one capital process for each $\theta' \in \Theta$, the parameter space. At each time, t, the confidence set corresponds to the set of $\theta' \in \Theta$ such that our capital process has not exceeded an improbable level of wealth for a fair game. Continuing with the betting analogy, the analyst must choose a predictable betting strategy for each game, $\lambda_t(\theta')$, which is typically chosen to be quasi-convex in θ' so that the confidence set forms an interval.

Theorem 2 (Hedged CS [Hedged]) *Assume we observe data following the data generating process of Section 2.1. Assume $Y_t \in [0,1]$ and $\pi_t(1 \mid X_t, \Omega_{t-1}) \in [k_t, 1-k_t]$ for all $t \in 1, \ldots, T$. If we define*

$$\mathcal{K}_T^+(\theta') := \prod_{t=1}^{T}(1 + \lambda_t(\theta')(h_t - \theta')), \quad \mathcal{K}_T^-(\theta') := \prod_{t=1}^{T}(1 - \lambda_t(\theta')(h_t - \theta')),$$

$$\mathcal{M}_T(\theta') := \frac{\mathcal{K}_T^+(\theta') + \mathcal{K}_T^-(\theta')}{2},$$

then

$$C_T^{Hedged} := \bigcap_{t \leq T} \left\{ \theta' \in [-1,1] : \mathcal{M}_T(\theta') < \frac{1}{\alpha} \right\},$$

forms a $(1-\alpha)$-CS for θ_0, where $(\lambda_t(\theta'))_{t=1}^{T} \in \left(\frac{-1}{k_t - \theta'}, \frac{1}{k_t + \theta'}\right)$ is a predictable sequence that may be interpreted as an analyst's betting strategy.

Proof, intuition, and other details can be found in Appendix C. We note that our result holds for any bounded Y_t by rescaling. The CS produced by Theorem 2 can be computationally expensive, as a grid search is performed over $\theta' \in [-1,1]$. Theorem 10 in Appendix E states a practical, closed-form CS with only small degradation in performance.

3.2 Asymptotic Confidence Sequences

Due to their time-uniform guarantees, the CSs defined so far produce wider intervals than their CI counterparts. In the fixed-time setting, coverage is guaranteed asymptotically.

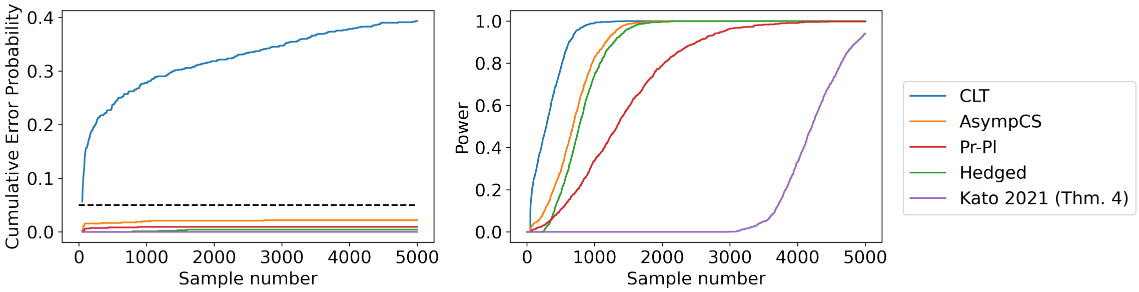

Figure 1: Cumulative error probability and power of experiment from Appendix F.3. Intervals based on the CLT (Theorem 1), AsympCS (Theorem 3), Pr-Pl (Theorem 10), Hedged (Theorem 2), and (Kato et al., 2021, Theorem 4) begin at $t = 50$.

Waudby-Smith et al. (2023) introduced a sequential analogue of asymptotic CIs, asymptotic CSs (AsympCS), by defining a CS which converges to some (unknown) CS. We now define our AsympCS for θ_0.

Theorem 3 (Asymptotic CS [AsympCS]) *Assume $\{(X_t, A_t, Y_t)\}_{t=1}^{T}$ follow the data generating process described in section 2.1. Furthermore, assume $\mathbb{E}(Y_t^{2+\delta}) < \infty$ for some $\delta > 0$. Let $\hat{\sigma}^2$ be an estimator of $Var(h_t)$, and $\rho > 0$ be a user-specified parameter, with a valid default being 0.5. For all $t \in 1, \ldots, T$, we have that*

$$C_T^{AsympCS} := \left(\frac{1}{T} \sum_{t=1}^{T} h_t \pm \sqrt{\frac{2(T\hat{\sigma}_T^2\rho^2 + 1)}{T^2\rho^2} \log\left(\frac{\sqrt{T\hat{\sigma}_T^2\rho^2 + 1}}{\alpha} \right)} \right),$$

forms a $(1 - \alpha)$-AsympCS for θ_0.

Proof and further details can be found in Appendix D. Although this interval does not yield exact coverage, empirically most errors occur quite early during the experiment. Its applicability for *reasonable* sample sizes provides a noticeable gain in power in comparison to the exact CSs. We also note that the theorem we make use of from Waudby-Smith et al. (2023) allows for time-varying conditional means. This suggests that the results of Theorem 3 can be extended to time-varying effects, which we leave for future work.

4. Empirical Results

We compare our methods to Kato et al. (2021). We run two simulations: one with Bernoulli outcomes, and one with continuous, bounded outcomes. We collect 5000 total samples for each iteration, constructing intervals after each sample. Details are given in Appendix F.

Our methods provide significantly narrower intervals, as is seen in the increased power in Figure 1. This is due to leveraging tighter concentration inequalities, as well as using time-varying truncation. Performance is inline with expectations from the CS literature. The effects of truncation on inference is further studied in Appendix F.6, and is a continued focus of this research.

References

Akshay Balsubramani. Sharp finite-time iterated-logarithm martingale concentration. *arXiv preprint 1405.2639*, 2015.

Akshay Balsubramani and Aaditya Ramdas. Sequential nonparametric testing with the law of the iterated logarithm. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'16, 2016.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Jessica Dai, Paula Gradu, and Christopher Harshaw. CLIP-OGD: An experimental design for adaptive neyman allocation in sequential experiments. In *Advances in Neural Information Processing Systems*, volume 37, 2023.

Aryeh Dvoretzky. Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, 1972.

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

Masahiro Kato, Takuya Ishihara, Junya Honda, and Yusuke Narita. Efficient adaptive experimental design for average treatment effect estimation. *arXiv preprint 2002.05308*, 2021.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 2023.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

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

Donald B. Rubin. Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81:961–962, 1986.

Jean Ville. *Étude critique de la notion de collectif.* 1939.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2023.

Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H. Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *arXiv preprint 2103.06476*, 2023.

Ian Waudby-Smith, Lili Wu, Aaditya Ramdas, Nikos Karampatziakis, and Paul Mineiro. Anytime-valid off-policy inference for contextual bandits. *ACM/IMS Journal of Data Science (to appear)*, 2024.

Kelly Zhang, Lucas Janson, and Susan Murphy. Statistical inference with M-estimators on adaptively collected data. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Disclaimer This paper was prepared for informational purposes in part by the Artificial Intelligence Research group of JPMorgan Chase & Co and its affiliates ("J.P. Morgan") and is not a product of the Research Department of J.P. Morgan. J.P. Morgan makes no representation and warranty whatsoever and disclaims all liability, for the completeness, accuracy or reliability of the information contained herein. This document is not intended as investment research or investment advice, or a recommendation, offer or solicitation for the purchase or sale of any security, financial instrument, financial product or service, or to be used in any way for evaluating the merits of participating in any transaction, and shall not constitute a solicitation under any jurisdiction or to any person, if such solicitation under such jurisdiction or to such person would be unlawful.

Acknowledgements AR acknowledges support from NSF grants IIS-2229881 and DMS-2310718.

Appendix A. Details and Proof of Theorem 1

A.1 On the Role of Policy Truncation

As mentioned in Section 2.2, Hahn et al. (2011) showed that the policy that minimizes the semiparametric lower bound of the asymptotic variance, σ^2, for regular estimators of the ATE is π^{AIPW}. For an arbitrary policy $\pi(a \mid x)$, we have that

$$\sigma^2 = \mathbb{E}\left[\sum_{a=0}^{1} \frac{v(a, X_t)}{\pi(a \mid X_t)} + (f(1, X_t) - f(0, X_t) - \theta_0)^2\right].$$

The policy π^{AIPW} depends on unknown quantities of the underlying data generating process. Kato et al. (2021) proposed sequentially estimating the unknown quantities from the observed data and and defined their suggested policy as

$$\pi_t^{\text{A2IPW,Kato}}(1 \mid X_t, \Omega_{t-1}) = \frac{\sqrt{\hat{v}_{t-1}(1, X_t)}}{\sqrt{\hat{v}_{t-1}(1, X_t)} + \sqrt{\hat{v}_{t-1}(0, X_t)}},$$

where $\hat{v}_{t-1}$ denotes an estimate of v using the first $t-1$ samples. For numerical stability, Kato et al. (2021) mixed this policy with a non-adaptive policy that assigns treatment with probability half. As the sample size grows, the mixing gradually assigns a greater weight to the estimated optimal policy. This mixing scheme prevents noisy estimates of v from inducing high variance in the observed $(h_t)_{t=1}^T$ early in the experiment and does not affect the asymptotic properties of the estimator. In a similar spirit, we explicitly define a truncation schedule for the propensity scores generated by our policy. However, our truncation schedule is not only useful for improving finite sample stability in practice; it is also a technical device that allows us to relax the assumptions needed for our results below. We can rewrite our policy from equation (3) as

$$\pi_t^{\text{A2IPW}}(1 \mid X_t, \Omega_{t-1}) = \left(\pi_t^{\text{A2IPW,Kato}}(1 \mid X_t, \Omega_{t-1}) \vee \frac{1}{k_t}\right) \wedge \left(1 - \frac{1}{k_t}\right). \tag{4}$$

Since our Theorem 1 holds in a more general setting, we can apply this truncation to arbitrary policies $\tilde{\pi}_t$, denoting

$$\pi_t = \left(\tilde{\pi}_t \vee \frac{1}{k_t}\right) \wedge \left(1 - \frac{1}{k_t}\right). \tag{5}$$

Note that setting $k_t \to \infty$ results in the non-truncated policy $\tilde{\pi}_t$.

In contrast to Kato et al. (2021), truncation plays a key role in our derivation of the asymptotic distribution of $\hat{\theta}_T^{\text{A2IPW}}$ and, in turn, the conditions required of k_t are of particular interest. If k_t increases to infinity, then as long as the non-truncated policy $\tilde{\pi}_t$ converges to some non-truncated policy $\tilde{\pi}$, the truncated policy π_t will also converge to $\tilde{\pi}$ (and the theorem would apply as long as k_t increased slowly enough that $k_t \max(\|\hat{f}_t - f\|_2, \|\pi_t - \pi\|_2) = o_{\mathbb{P}}(1)$). Instead, if k_t remains constant or increases to a finite bound, then the truncated policy π_t will converge to an appropriate truncation of $\tilde{\pi}$, and the theorem would still apply as long as $\max(\|\hat{f}_t - f\|_2, \|\pi_t - \pi\|_2) = o_{\mathbb{P}}(1)$. When we have $\pi = \pi^{\text{AIPW}}$, we are implicitly assuming that we have independently selected k_t such that truncation becomes asymptotically inactive. Note that if $\tilde{\pi}_t$ were uniformly bounded away from 0, as assumed in Kato et al. (2021), then we could simply set $k_t = 1/\min(\tilde{\pi}_t(x), 1-\tilde{\pi}_t(x))$ so that $\tilde{\pi}_t = \pi_t$, meaning we never actively truncate $\tilde{\pi}_t$. In that case, the conditions in (Kato et al., 2021, Theorem 1) would imply the conditions in our theorem. The following remark alludes to how selecting k_t can lead us to semiparametric efficient inference with the our proposed policy, π_t^{A2IPW}.

Remark 4 (Semiparametric Efficiency) *Assume that we set k_t such that $\lim_{t\to\infty} k_t > \sup \frac{1}{\pi^{\text{AIPW}}}$. Assume that the estimated conditional variance function $\hat{v}_t$ is consistent for v such that $\|\pi_t^{\text{A2IPW}} - \pi^{\text{AIPW}}\|_2 = o_{\mathbb{P}}(1)$. If k_t grows at a rate such that $k_t\|\pi_t^{\text{A2IPW}} - \pi^{\text{AIPW}}\|_2 = o_{\mathbb{P}}(1)$ and all other assumptions of Theorem 1 hold, then $\hat{\theta}_T^{\text{A2IPW}}$ is semiparametric efficient.*

In the final sentence of Theorem 1, we state that if π_t converges to π^{AIPW}, then the semiparametric lower bound is minimized with respect to π (Hahn et al., 2011). In order to make use of this result, we require an adaptive policy that converges to π^{AIPW}. Remark 4 states that π_t^{A2IPW} is such a policy as long as our estimates of v are consistent and our truncation does not vanish too quickly. The rate at which k_t is allowed to increase as per the conditions in Remark 4 depends on the rate that $\hat{v}_t$ converges to v. In practice this rate is unobservable, and it is worth acknowledging this limitation. Overcoming this limitation is an interesting direction for future research.

A.2 High-Level Roadmap of Proof

Our proof follows a similar style to the proof of Kato et al. (2021). We consider a martingale difference sequence (MDS) and apply a central limit theorem to find the asymptotic distribution of the sample mean of the MDS. The main departure of our proof from their proof is the statement of the central limit theorem which is amenable to making assumptions standard in causal inference.

To outline the proof, first we state our assumptions. Next we establish that $\{z_t\}_{t=1}^T$, where $z_t = h_t - \theta_0$, is a MDS. We then state the MDS central limit theorem by Dvoretzky (1972) and show that $\{z_t\}_{t=1}^T$ satisfies the necessary conditions. For the sake of brevity, we defer much of the tedious algebra to Appendix B. Since $\bar{z}_T = T^{-1}\sum_{t=1}^T z_t = T^{-1}\sum_{t=1}^T (h_t - \theta_0) = \hat{\theta}_T^{\text{A2IPW}} - \theta_0$, this result allows us to characterize the asymptotic distribution of $\hat{\theta}_T^{\text{A2IPW}}$.

A.3 Assumptions

- *IID Contexts and Potential Outcomes* : $\{X_t, Y_t(0), Y_t(1)\}_{t=1}^T$ are independent and identically distributed.
- *Finite Variance* : $\text{Var}(Y) < \infty$.
- *Finite Conditional Variance* : $\text{Var}(Y(a) \mid x) < \infty$ for $a \in \{0,1\}$ and $x \in \mathcal{X}$.
- *Convergence of Regression* : $k_t\|\hat{f}_t - f\|_2 = o_{\mathbb{P}}(1)$
- *Convergence of Policy* : $k_t\|\pi_t - \pi\|_2 = o_{\mathbb{P}}(1)$.
- *π Bounded Away from 0* : $\frac{1}{\pi} < C_1$, for some $C_1 < \infty$.

- *Finite Variance of Predictions* : $\mathrm{Var}(\hat{f}_{t-1}(a, X_t)) < \infty$ for $a \in \{0, 1\}$.

A.4 z_t is a MDS

Kato et al. (2021) show the first necessary condition, $\mathbb{E}(z_t \mid \Omega_{t-1}) = 0$. For completeness we present this step here.

$$
\begin{aligned}
&\mathbb{E}\big[z_t \mid \Omega_{t-1}\big] \\
&= \mathbb{E}\left[\frac{\mathbb{1}[A_t = 1]\big(Y_t - \hat{f}_{t-1}(1, X_t)\big)}{\pi_t(1 \mid X_t, \Omega_{t-1})} - \frac{\mathbb{1}[A_t = k]\big(Y_t - \hat{f}_{t-1}(0, X_t)\big)}{\pi_t(0 \mid X_t, \Omega_{t-1})} + \hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0 \;\middle|\; \Omega_{t-1}\right] \\
&= \mathbb{E}\Bigg[\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0 \\
&\qquad + \mathbb{E}\left[\frac{\mathbb{1}[A_t = 1]\big(Y_t - \hat{f}_{t-1}(1, X_t)\big)}{\pi_t(1 \mid X_t, \Omega_{t-1})} - \frac{\mathbb{1}[A_t = 0]\big(Y_t - \hat{f}_{t-1}(0, X_t)\big)}{\pi_t(0 \mid X_t, \Omega_{t-1})} \;\middle|\; X_t, \Omega_{t-1}\right] \;\Bigg|\; \Omega_{t-1}\Bigg] \\
&= \mathbb{E}\left[\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0 + f(1, X_t) - f(0, X_t) - \hat{f}_{t-1}(1, X_t) + \hat{f}_{t-1}(0, X_t) \;\middle|\; \Omega_{t-1}\right] = 0.
\end{aligned}
$$

The second required condition is $\mathbb{E}|z_t| < \infty$. By Chebyshev's inequality,

$$\mathbb{P}(|z_t - \mathbb{E}(z_t)| \geq 1) = \mathbb{P}(|z_t| \geq 1) \leq \mathrm{Var}(z_t) = \mathbb{E}(z_t^2) < \infty.$$

The final inequality holds due to finite variance of the outcome and regression prediction, and truncated propensity scores.

A.5 MDS Central Limit Theorem

Kato et al. (2021) used a MDS CLT which requires (condition b) a finite $2 + \delta$ moment ($\delta > 0$) for $|z_t|$. Instead we use the MDS CLT as stated by Dvoretzky (1972). This statement contains a Lindeberg type condition where we must only consider the second moment of $|z_t|$. Since we do not assume boundedness, we opt for this Lindeberg-type statement. For completeness, we present this theorem as it is stated in Zhang et al. (2021, Theorem 2).

Theorem 5 (MDS Central Limit Theorem) *Let $Z_T(\mathcal{P})_{T \geq 1}$ be a sequence of random variables whose distributions are defined by some $\mathcal{P} \in \mathbb{P}$ and some nuisance component η. Moreover, let $Z_T(\mathcal{P})_{T \geq 1}$ be a martingale difference sequence with respect to Ω_t, meaning $\mathbb{E}_{\mathcal{P},\eta}[Z_t(\mathcal{P}) \mid \Omega_{t-1}] = 0$ for all $t \geq 1$ and $\mathcal{P} \in \mathbb{P}$. If we assume that,*

1. *$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathcal{P},\eta}\left[z_t^2 \mid \Omega_{t-1}\right] \xrightarrow{p} \sigma^2$ uniformly over $\mathcal{P} \in \mathbb{P}$, where σ^2 is a constant $0 < \sigma^2 < \infty$, and that,*

2. *for any $\epsilon > 0$, $\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathcal{P},\eta}\left[z_t(\mathcal{P})^2 \mathbb{I}\left[|z_t(\mathcal{P})| > \epsilon\right] \mid \Omega_{t-1}\right] \xrightarrow{p} 0$ uniformly over $\mathcal{P} \in \mathbb{P}$,*

then $\sqrt{T}(\bar{z}_t) \xrightarrow{d} N(0, \sigma^2)$ uniformly over $\mathcal{P} \in \mathbb{P}$.

Dropping the requirement of the conditions holding uniformly over $\mathcal{P} \in \mathbb{P}$ recovers the original result by Dvoretzky (1972). Below we show that these two conditions are satisfied. It follows that

$$\sqrt{T}(\bar{z}_t) = \frac{\hat{\theta}^{\text{A2IPW}} - \theta_0}{\sqrt{T}} \xrightarrow{d} N(0, \sigma^2),$$

where

$$\sigma^2 = \mathbb{E}\left[\sum_{a=0}^{1} \frac{v(a, X_t)}{\pi(a \mid X_t)} + (f(1, X_t) - f(0, X_t) - \theta_0)^2\right].$$

A.5.1 Condition 1 (Conditional Variance)

We wish to show that

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[z_t^2 \mid \Omega_{t-1}\right] \xrightarrow{p} \sigma^2 = \mathbb{E}\left[\sum_{a=0}^{1} \frac{\nu(a, X_t)}{\pi(a \mid X_t)} + \Big(f(1, X_t) - f(0, X_t) - \theta_0\Big)^2\right].$$

This is equivalent to showing

$$\frac{1}{T}\sum_{t=1}^{T} \left(\mathbb{E}\left[z_t^2 \mid \Omega_{t-1}\right] - \sigma^2\right) \xrightarrow{p} 0.$$

To reduce notational clutter, let $\mathbb{E}(X_t \mid \Omega_{t-1})$ be denoted as $\mathbb{E}^{t-1}(X_t)$. Kato et al. (2021, Appendix B) show

$$\begin{aligned}
\mathbb{E}\left[z_t^2 \mid \Omega_{t-1}\right] - \sigma^2 &= \mathbb{E}^{t-1}\Bigg[\frac{(Y_t(1) - \hat{f}_{t-1}(1, X_t))^2}{\pi_t(1 \mid X_t, \Omega_{t-1})} + \frac{(Y_t(0) - \hat{f}_{t-1}(0, X_t))^2}{\pi_t(0 \mid X_t, \Omega_{t-1})} \\
&+ \Big(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0\Big)^2 \\
&+ 2(f(1, X_t) - f(0, X_t) - \hat{f}_{t-1}(1, X_t) + \hat{f}_{t-1}(0, X_t))(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0)\Bigg] \\
&- \mathbb{E}^{t-1}\left[\frac{(Y_t(1) - f(1, X_t))^2}{\pi(1 \mid X_t)} + \frac{(Y_t(0) - f(0, X_t))^2}{\pi(0 \mid X_t)} + (f(1, X_t) - f(0, X_t) - \theta_0)^2\right] \\
&= \sum_{a=0}^{1} \mathbb{E}^{t-1}\left[\frac{\Big(Y_t(a) - \hat{f}_{t-1}(a, X_t)\Big)^2}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{(Y_t(a) - f(a, X_t))^2}{\pi(a \mid X_t)}\right] \qquad (6) \\
&+ 2\mathbb{E}^{t-1}\left[\Big(f(1, X_t) - f(0, X_t) - \hat{f}_{t-1}(1, X_t) + \hat{f}_{t-1}(0, X_t)\Big)\Big(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0\Big)\right] \qquad (7) \\
&+ \mathbb{E}^{t-1}\left[\Big(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0\Big)^2 - (f(1, X_t) - f(0, X_t) - \theta_0)^2\right]. \qquad (8)
\end{aligned}$$

We now consider terms (6), (7) and (8) individually. We make use of auxiliary lemmas and defer proofs to Appendix B. In all of the lemmas below, we keep all assumptions from Appendix A.3.

Lemma 6 (Convergence of (6)**)** *Under the assumptions of Theorem 1, we have*

$$\sum_{a=0}^{1} \mathbb{E}^{t-1}\left[\frac{\left(Y_t(a) - \hat{f}_{t-1}(a, X_t)\right)^2}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{(Y_t(a) - f(a, X_t))^2}{\pi(a \mid X_t)}\right] = o_{\mathbb{P}}(1).$$

Lemma 7 (Convergence of (7)**)** *Under the assumptions of Theorem 1, we have*

$$2\mathbb{E}^{t-1}\left[\left(f(1, X_t) - f(0, X_t) - \hat{f}_{t-1}(1, X_t) + \hat{f}_{t-1}(0, X_t)\right)\left(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0\right)\right] = o_{\mathbb{P}}(1).$$

Lemma 8 (Convergence of (8)**)** *Under the assumptions of Theorem 1, we have*

$$\mathbb{E}^{t-1}\left[\left(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0\right)^2 - (f(1, X_t) - f(0, X_t) - \theta_0)^2\right] = o_{\mathbb{P}}(1).$$

Given Lemmas 6, 7, and 8, convergence in probability to zero for each term is established, and therefore so is the convergence of the sum. Convergence of the conditional variance of the MDS is then established. We now focus on the convergence of the sample average of the conditional variances.

Following the same argument from Kato et al. (2021), for any $\epsilon > 0$, there exists a $\tilde{t} > 0$ such that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}^{t-1}\left[z_t^2\right] - \sigma^2\right) \leq \tilde{t}/T + \epsilon.$$

Since σ^2 does not depend on t, $\tilde{t}/T \to 0$ as $T \to \infty$, and so $\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}^{t-1}\left[z_t^2\right] - \sigma^2\right) \xrightarrow{p} 0$. Hence, condition 1 is satisfied.

A.5.2 Condition 2 (Conditional Lindeberg)

We seek to show that for any $\delta > 0$,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(z_t^2 \mathbb{1}\left[|z_t| > \delta\sqrt{T}\right] \;\middle|\; \Omega_{t-1}\right) \xrightarrow{p} 0.$$

Define $b_t = z_t^2 \mathbb{1}(|z_t| > \delta\sqrt{T})$. Then $b_t = z_t^2$ w.p. $\mathbb{P}(|z_t| > \delta\sqrt{T})$ and 0 otherwise. By Chebyshev's inequality,

$$\mathbb{P}(|z_t| > \delta\sqrt{T}) \leq \frac{\mathrm{Var}(z_t)}{\delta^2 T}.$$

We note that $\mathrm{Var}(z_t) = \mathbb{E}(z_t^2) < \infty$. This gives

$$\lim_{T\to\infty}\frac{\mathrm{Var}(z_t)}{\delta^2 T} = 0,$$

which implies that $b_t \xrightarrow{p} 0$, and $b_t \xrightarrow{d} 0$.

Note that $|z_t| \leq z_t^2$, and $\mathbb{E}(z_t^2) < \infty$. By the dominated convergence theorem, $\lim_{T\to\infty}\mathbb{E}(b_t) = \mathbb{E}(\lim_{T\to\infty} b_t) = 0$. Hence we have

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(z_t^2 \mathbb{1}\left[|z_t| > \delta\sqrt{T}\right] \;\middle|\; \Omega_{t-1}\right) \xrightarrow{p} 0.$$

Appendix B. Auxiliary Lemmas and Proofs

This appendix shows proofs for auxiliary lemmas used in Appendix A. The proofs involve tedious algebra and are included in full detail for completeness.

B.1 Proof of Lemma 6

The term considered in Lemma 6, specifically term (6), involves a summation over the potential treatments, we choose to focus on a single arbitrary treatment, a, and show that the term for an individual treatment converges to zero in probability, and hence, so does the sum.

$$\mathbb{E}^{t-1}\left[\frac{\left(Y_t(a)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}-\frac{(Y_t(a)-f(a,X_t))^2}{\pi(a\mid X_t)}\right]$$

$$=\mathbb{E}^{t-1}\left[\frac{\left(Y_t(a)-\hat{f}_{t-1}(a,X_t)+f(a,X_t)-f(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}-\frac{(Y_t(a)-f(a,X_t))^2}{\pi(a\mid X_t)}\right] \tag{9}$$

$$=\mathbb{E}^{t-1}\left[\frac{\left((Y_t(a)-f(a,X_t))+\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}-\frac{(Y_t(a)-f(a,X_t))^2}{\pi(a\mid X_t)}\right] \tag{10}$$

$$=\mathbb{E}^{t-1}\left[\frac{(Y_t(a)-f(a,X_t))^2}{\pi_t(a\mid X_t,\Omega_{t-1})}+\frac{2\,(Y_t(a)-f(a,X_t))\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}\right. \tag{11}$$

$$\left.+\frac{\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}-\frac{(Y_t(a)-f(a,X_t))^2}{\pi(a\mid X_t)}\right]$$

$$=\mathbb{E}^{t-1}\left[(Y_t(a)-f(a,X_t))^2\left(\frac{1}{\pi_t(a\mid X_t,\Omega_{t-1})}-\frac{1}{\pi(a\mid X_t)}\right)\right] \tag{12}$$

$$+\,2\mathbb{E}^{t-1}\left[\frac{(Y_t(a)-f(a,X_t))\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}+\frac{\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}\right].$$

Above, (9) simultaneously adds and subtracts $f(a,X_t)$, while (10) and (11) square the binomial term. In equation (12), we factor $(Y_t(a)-f(a,X_t))^2$ from the first and final terms, and utilize linearity of expectation. Continuing,

$$\mathbb{E}^{t-1}\left[(Y_t(a)-f(a,X_t))^2\left(\frac{1}{\pi_t(a\mid X_t,\Omega_{t-1})}-\frac{1}{\pi(a\mid X_t)}\right)\right]$$

$$+\,2\mathbb{E}^{t-1}\left[\frac{(Y_t(a)-f(a,X_t))\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}+\frac{\left(f(a,X_t)-\hat{f}_{t-1}(a,X_t)\right)^2}{\pi_t(a\mid X_t,\Omega_{t-1})}\right]$$

$$\leq \mathbb{E}^{t-1}\left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{1}{\pi(a \mid X_t)}\right)\right] \tag{13}$$

$$+ 2\mathbb{E}^{t-1}\left[\frac{(Y_t(a) - f(a, X_t))\left(f(a, X_t) - \hat{f}_{t-1}(a, X_t)\right)^2}{\pi_t(a \mid X_t, \Omega_{t-1})}\right] + \mathbb{E}^{t-1}\left[k_t\left(f(a, X_t) - \hat{f}_{t-1}(a, X_t)\right)^2\right]$$

$$= \mathbb{E}^{t-1}\left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{1}{\pi(a \mid X_t)}\right)\right] \tag{14}$$

$$+ 2\mathbb{E}^{t-1}\left[\frac{(Y_t(a) - f(a, X_t))\left(f(a, X_t) - \hat{f}_{t-1}(a, X_t)\right)^2}{\pi_t(a \mid X_t, \Omega_{t-1})}\right] + k_t\left(\|\hat{f}(a, X_t) - f(a, X_t))\|_2\right)^2.$$

Inequality (13) follows since our policy is truncated within $[\frac{1}{k_t}, 1 - \frac{1}{k_t}]$. Equation (14) uses norm notation in the final term so that we may reference our assumptions further in the proof. We now turn our focus to simplifying the middle term of equation (14),

$$\mathbb{E}^{t-1}\left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{1}{\pi(a \mid X_t)}\right)\right]$$

$$+ 2\mathbb{E}^{t-1}\left[\frac{(Y_t(a) - f(a, X_t))\left(f(a, X_t) - \hat{f}_{t-1}(a, X_t)\right)^2}{\pi_t(a \mid X_t, \Omega_{t-1})}\right] + k_t\left(\|\hat{f}(a, X_t) - f(a, X_t))\|_2\right)^2$$

$$= \mathbb{E}^{t-1}\left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{1}{\pi(a \mid X_t)}\right)\right] \tag{15}$$

$$+ 2\left(\mathbb{E}^{t-1}\left[\frac{Y_t(a)}{\pi_t(a \mid X_t, \Omega_{t-1})}\left(f(a, X_t) - \hat{f}(a, X_t)\right)^2\right] - \mathbb{E}^{t-1}\left[\frac{f(a, X_t)}{\pi_t(a \mid X_t, \Omega_{t-1})}\left(\hat{f}(a, X_t) - f(a, X_t)\right)^2\right]\right)$$

$$+ k_t\left(\|\hat{f}(a) - f(a))\|_2\right)^2$$

$$= \mathbb{E}^{t-1}\left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{1}{\pi(a \mid X_t)}\right)\right]$$

$$+ 2\Bigg(\mathbb{E}^{t-1}\left[\mathbb{E}\left[\frac{Y_t(a)}{\pi_t(a \mid X_t, \Omega_{t-1})}\left(f(a, X_t) - \hat{f}(a, X_t)\right)^2\right] \;\Big|\; X_t\right] \tag{16}$$

$$- 2\mathbb{E}^{t-1}\left[\mathbb{E}\left[\frac{f(a, X_t)}{\pi_t(a \mid X_t, \Omega_{t-1})}\left(\hat{f}(a, X_t) - f(a, X_t)\right)^2\right] \;\Big|\; X_t\right]\Bigg) \tag{17}$$

$$+ k_t\left(\|\hat{f}(a, X_t) - f(a, X_t))\|_2\right)^2,$$

where equation (15) expands the term of interest, and terms (16) and (17) apply the law of iterated expectation. Conditioning on X_t, the only non-constant term in terms (16) and (17) is $Y_t(a)$, whose conditional expectation on X_t is $f(a, X_t)$. Therefore, the terms (16) and (17) reduce to 0. Simplifying, we have

$$\mathbb{E}^{t-1}\left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{1}{\pi(a \mid X_t)}\right)\right]$$

$$
\begin{aligned}
&+2\Bigg(\mathbb{E}^{t-1}\left[\mathbb{E}\left[\frac{Y_t(a)}{\pi_t(a \mid X_t, \Omega_{t-1})}\left(f(a,X_t)-\hat{f}(a,X_t)\right)^2\right] \;\Big|\; X_t\right] \\
&-2\mathbb{E}^{t-1}\left[\mathbb{E}\left[\frac{f(a,X_t)}{\pi_t(a \mid X_t, \Omega_{t-1})}\left(\hat{f}(a,X_t)-f(a,X_t)\right)^2\right] \;\Big|\; X_t\right]\Bigg) \\
&+k_t\left(\|\hat{f}(a,X_t)-f(a,X_t))\|_2\right)^2 \\
&=\mathbb{E}^{t-1}\left[(Y_t(a)-f(a,X_t))^2\left(\frac{1}{\pi_t(a \mid X_t,\Omega_{t-1})}-\frac{1}{\pi(a \mid X_t)}\right)\right] \qquad (18) \\
&+k_t\left(\|\hat{f}(a,X_t)-f(a,X_t))\|_2\right)^2.
\end{aligned}
$$

By assumption, $k_t\|\hat{f}(a,X_t)-f(a,X_t)\|_2 = o_{\mathbb{P}}(1)$. It follows then that $k_t\left(\|\hat{f}(a,X_t)-f(a,X_t))\|_2\right)^2 = o_{\mathbb{P}}(1)$. Equation (18) can be further simplified to

$$
\begin{aligned}
&\mathbb{E}^{t-1}\left[(Y_t(a)-f(a,X_t))^2\left(\frac{1}{\pi_t(a \mid X_t,\Omega_{t-1})}-\frac{1}{\pi(a \mid X_t)}\right)\right] \\
&+k_t\left(\|\hat{f}(a,X_t)-f(a,X_t))\|_2\right)^2 \\
&=\mathbb{E}^{t-1}\left[(Y_t(a)-f(a,X_t))^2\left(\frac{1}{\pi_t(a \mid X_t,\Omega_{t-1})}-\frac{1}{\pi(a \mid X_t)}\right)\right]+o_{\mathbb{P}}(1) \\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{\pi(a \mid X_t)-\pi_t(a \mid X_t,\Omega_{t-1})}{\pi(a \mid X_t)\pi_t(a \mid X_t,\Omega_{t-1})}(Y_t(a)-f(a,X_t))^2 \;\Big|\; X_t,\Omega_{t-1}\right] \;\Big|\; \Omega_{t-1}\right]+o_{\mathbb{P}}(1) \qquad (19) \\
&=\mathbb{E}\left[\frac{\pi(a \mid X_t)-\pi_t(a \mid X_t,\Omega_{t-1})}{\pi(a \mid X_t)\pi_t(a \mid X_t,\Omega_{t-1})}\mathbb{E}\left[(Y_t(a)-f(a,X_t))^2 \;\Big|\; X_t,\Omega_{t-1}\right] \;\Big|\; \Omega_{t-1}\right]+o_{\mathbb{P}}(1) \qquad (20) \\
&\leq C_1 k_t \mathrm{Var}(Y_t)\mathbb{E}^{t-1}\left[\pi(a \mid X_t)-\pi_t(a \mid X_t,\Omega_{t-1})\right]+o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1). \qquad (21)
\end{aligned}
$$

Equation (19) follows from the law of total expectation. In equation (20) $\pi_t(a \mid X_t, \Omega_{t-1})$ given X_t and Ω_{t-1} is constant, and can be moved out of the inner expectation, away from $(Y_t(a) - f(a, X_t))^2$. The bound (21) then utilizes our policy truncation and our assumption that $\frac{1}{\pi}$ is bounded. We denote this bound as $C_1 < \infty$. We are able to bound the denominator with a constant, and move this constant outside of the expectation. Simultaneously, we note that the inner expectation is by definition the conditional variance of $Y_t(a)$ given X_t. We apply the law of total variance to bound this term by $\mathrm{Var}(Y_t(a))$, and move this constant out of the outer expectation. The bound 21 reduces to $o_{\mathbb{P}}(1)$, since converge in ℓ_2 implies convergence in ℓ_1, and the Lemma is proved.

B.2 Proof of Lemma 7

We look to prove that

$$
2\mathbb{E}^{t-1}\left[\left(f(1,X_t)-f(0,X_t)-\hat{f}_{t-1}(1,X_t)+\hat{f}_{t-1}(0,X_t)\right)\left(\hat{f}_{t-1}(1,X_t)-\hat{f}_{t-1}(0,X_t)-\theta_0\right)\right]=o_{\mathbb{P}}(1).
$$

For simplicity, we temporarily ignore the constant. Continuing,

$$
\mathbb{E}^{t-1}\left[\left(f(1,X_t)-f(0,X_t)-\hat{f}_{t-1}(1,X_t)+\hat{f}_{t-1}(0,X_t)\right)(f(1,X_t)-f(0,X_t)-\theta_0)\right]
$$

$$
\begin{aligned}
&= \mathbb{E}^{t-1}\left[\left(f(1,X_t)-\hat{f}_{t-1}(1,X_t)\right)(f(1,X_t)-f(0,X_t)-\theta_0)\right] && (22)\\
&\quad + \mathbb{E}^{t-1}\left[\left(f(0,X_t)-\hat{f}(0,X_t)\right)(f(1,X_t)-f(0,X_t)-\theta_0)\right]\\
&\leq \sqrt{\mathbb{E}^{t-1}\left[\left(f(1,X_t)-\hat{f}_{t-1}(1,X_t)\right)^2\right]\mathbb{E}^{t-1}\left[(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]} && (23)\\
&\quad + \sqrt{\mathbb{E}^{t-1}\left[\left(f(0,X_t)-\hat{f}(0,X_t)\right)^2\right]\mathbb{E}^{t-1}\left[(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]},
\end{aligned}
$$

where equation (22) separates terms from different treatments and utilizes the linearity of expectation. Bound (23) then follows from applying the Cauchy-Schwarz inequality. We conclude by showing

$$
\begin{aligned}
&\sqrt{\mathbb{E}^{t-1}\left[\left(f(1,X_t)-\hat{f}_{t-1}(1,X_t)\right)^2\right]\mathbb{E}^{t-1}\left[(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]}\\
&\quad + \sqrt{\mathbb{E}^{t-1}\left[\left(f(0,X_t)-\hat{f}(0,X_t)\right)^2\right]\mathbb{E}^{t-1}\left[(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]}\\
&= \|\hat{f}-f\|_2\sqrt{\mathbb{E}^{t-1}\left[(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]} + \|\hat{f}-f\|_2\sqrt{\mathbb{E}^{t-1}\left[(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]}\\
&= o_{\mathbb{P}}(1).
\end{aligned}
$$

Using norm notation and applying the assumption of convergence of regression in ℓ_2-norm, convergence is established, and the lemma is proved.

B.3 Proof of Lemma 8

We wish to prove that

$$
\mathbb{E}^{t-1}\left[\left(\hat{f}_{t-1}(1,X_t)-\hat{f}_{t-1}(0,X_t)-\theta_0\right)^2-(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]=o_{\mathbb{P}}(1).
$$

We begin by expanding this term,

$$
\begin{aligned}
&\mathbb{E}^{t-1}\left[\left(\hat{f}_{t-1}(1,X_t)-\hat{f}_{t-1}(0,X_t)-\theta_0\right)^2-(f(1,X_t)-f(0,X_t)-\theta_0)^2\right]\\
&\quad = \mathbb{E}^{t-1}\Bigg[\left(\left(\hat{f}(1,X_t)-\hat{f}(0,X_t)-\theta_0\right)+(f(1,X_t)-f(0,X_t)-\theta_0)\right) && (24)\\
&\qquad\qquad \times\left(\left(\hat{f}(1,X_t)-\hat{f}(0,X_t)-\theta_0\right)-(f(1,X_t)-f(0,X_t)-\theta_0)\right)\Bigg]\\
&\quad = \mathbb{E}^{t-1}\Bigg[\left((\hat{f}(1,X_t)+f(1,X_t))-(f(0,X_t)+\hat{f}(0,X_t))-2\theta_0\right) && (25)\\
&\qquad\qquad \times\left(\left(\hat{f}(1,X_t)-f(1,X_t)\right)+\left(f(0,X_t)-\hat{f}(0,X_t)\right)\right)\Bigg].
\end{aligned}
$$

Equation (24) arises from the fact that $a^2 - b^2 = (a+b)(a-b)$ for real numbers a, b. Equation (25) then collapses θ_0 to a single term. We now add and subtract $f(1, X_t)$ and $f(0, X_t)$ to the first term of equation (25), giving

$$\begin{aligned}
&\mathbb{E}^{t-1}\Bigg[\Big((\hat{f}(1, X_t) + f(1, X_t)) - (f(0, X_t) + \hat{f}(0, X_t)) - 2\theta_0 \Big) \\
&\qquad\qquad \times \Big(\Big(\hat{f}(1, X_t) - f(1, X_t) \Big) + \Big(f(0, X_t) - \hat{f}(0, X_t) \Big) \Big) \Bigg] \\
&= \mathbb{E}^{t-1}\Bigg[\Big((\hat{f}(1, X_t) - f(1, X_t)) + (f(0, X_t) - \hat{f}(0, X_t)) + 2\left(f(1, X_t) - f(0, X_t) - \theta_0 \right) \Big) \\
&\qquad\qquad \times \Big(\Big(\hat{f}(1, X_t) - f(1, X_t) \Big) + \Big(f(0, X_t) - \hat{f}(0, X_t) \Big) \Big) \Bigg]. \qquad (26)
\end{aligned}$$

Equation (26) completes this step, and rearranges terms so that we may use the assumption of convergence of regression.

Next, distributing the second term in equation (26), along with the use of the linearity of expectation gives

$$\begin{aligned}
&\mathbb{E}^{t-1}\Bigg[\Big((\hat{f}(1, X_t) - f(1, X_t)) + (f(0, X_t) - \hat{f}(0, X_t)) + 2\left(f(1, X_t) - f(0, X_t) - \theta_0 \right) \Big) \\
&\qquad\qquad \times \Big(\Big(\hat{f}(1, X_t) - f(1, X_t) \Big) + \Big(f(0, X_t) - \hat{f}(0, X_t) \Big) \Big) \Bigg] \\
&= \mathbb{E}^{t-1}\Bigg[\Big((\hat{f}(1, X_t) - f(1, X_t)) + (f(0, X_t) - \hat{f}(0, X_t)) \Big)^2 \Bigg] \\
&+ 2\mathbb{E}^{t-1}\Bigg[(f(1, X_t) - f(0, X_t) - \theta_0) \Big(\Big(\hat{f}(1, X_t) - f(1, X_t) \Big) + \Big(f(0, X_t) - \hat{f}(0, X_t) \Big) \Big) \Bigg]. \qquad (27)
\end{aligned}$$

The first term in equation (27) converges in probability by assumption. For the second term, we distribute $f(1, X_t) - f(0, X_t) - \theta_0)$ yielding

$$\begin{aligned}
&\mathbb{E}^{t-1}\Bigg[\Big((\hat{f}(1, X_t) - f(1, X_t)) + (f(0, X_t) - \hat{f}(0, X_t)) \Big)^2 \Bigg] \qquad (28) \\
&+ 2\mathbb{E}^{t-1}\Bigg[(f(1, X_t) - f(0, X_t) - \theta_0) \Big(\Big(\hat{f}(1, X_t) - f(1, X_t) \Big) + \Big(f(0, X_t) - \hat{f}(0, X_t) \Big) \Big) \\
&= \mathbb{E}^{t-1}\Big[\left(2\left(f(1, X_t) - f(0, X_t) - \theta_0 \right) \right) \Big(f(1, X_t) - \hat{f}(1, X_t) \Big) \Big] \qquad (29) \\
&\quad + \mathbb{E}^{t-1}\Big[\left(2\left(f(1, X_t) - f(0, X_t) - \theta_0 \right) \right) \Big(f(0, X_t) - \hat{f}(0, X_t) \Big) \Big] + o_{\mathbb{P}}(1). \qquad (30)
\end{aligned}$$

Applying the Cauchy-Schwarz inequality to each term in equation (30) gives

$$\begin{aligned}
&\mathbb{E}^{t-1}\left[\left(2\left(f(1,X_t)-f(0,X_t)-\theta_0\right)\right)\left(f(1,X_t)-\hat{f}(1,X_t)\right)\right] && (31)\\
&\qquad+\mathbb{E}^{t-1}\left[\left(2\left(f(1,X_t)-f(0,X_t)-\theta_0\right)\right)\left(f(0,X_t)-\hat{f}(0,X_t)\right)\right]+o_{\mathbb{P}}(1)\\
&\leq\sqrt{\mathbb{E}^{t-1}\left[\left(2\left(f(1,X_t)-f(0,X_t)-\theta_0\right)\right)^2\right]\mathbb{E}^{t-1}\left[\left(f(1,X_t)-\hat{f}(1,X_t)\right)^2\right]} && (32)\\
&\qquad+\sqrt{\mathbb{E}^{t-1}\left[\left(2\left(f(1,X_t)-f(0,X_t)-\theta_0\right)\right)^2\right]\mathbb{E}^{t-1}\left[\left(f(0,X_t)-\hat{f}(0,X_t)\right)^2\right]}+o_{\mathbb{P}}(1)\\
&=\sqrt{\mathbb{E}^{t-1}\left[\left(2\left(f(1,X_t)-f(0,X_t)-\theta_0\right)\right)^2\right]}\|\hat{f}-f\|_2 && (33)\\
&\qquad+\sqrt{\mathbb{E}^{t-1}\left[\left(2\left(f(1,X_t)-f(0,X_t)-\theta_0\right)\right)^2\right]}\|\hat{f}-f\|_2+o_{\mathbb{P}}(1).
\end{aligned}$$

Equation (33) follows from the bound (32) by definition. Since $\|\hat{f}-f\|_2 = o_{\mathbb{P}}(1)$, Equation (33) reduces to $o_{\mathbb{P}}(1)$, and the lemma is proved.

Appendix C. Details and Proof of Theorem 2

C.1 Details and Intuition of Hedged CS

$\mathcal{K}_T^+(\theta')$ and $\mathcal{K}_T^-(\theta')$ can be interpreted as capital processes for a gambler who is betting in favor of $\theta_0 > \theta'$ and $\theta_0 < \theta'$ respectively in two separate games against nature. Since we wish to produce two-sided intervals, we take the mean of these two process to form $\mathcal{M}_T(\theta')$. This is equivalent to a gambler partitioning their wealth equally between two games. The analyst must choose a predictable betting strategy for each game, $\lambda_t(\theta')$. In theory, this betting strategy could be different at each possible value of θ'. Moreso, apart from a bounded range, the only restriction on $\lambda_t(\theta')$ is that it is predictable, meaning that it cannot depend on the current or any future observations. However, since our parameter space is continuous, an exhaustive search over an infinite set of θ' is not feasible. Waudby-Smith and Ramdas (2023) propose a method to set λ_t to be quasi-convex in θ' so that the confidence set forms an interval. With quasi-convexity, it is sufficient to partition the parameter space and perform a grid-search; see their paper for further details and a variety of settings of λ_t. We provide the explicit strategy that we use in Appendix C.4.

C.2 Proof Outline

The proof *adapts* the proof of Waudby-Smith et al. (2024, Theorem 1) to our problem setting. The only departure in our proof is that our parameter space and $(\lambda_t)_{t=1}^T$ are not strictly non-negative. We include this proof to demonstrate how our bounds on λ_t originate as well as showing how our proof does not make use of the mirroring technique to form a $(1-\alpha)$-upper CS. Although these adjustments are immediate and obvious to those familiar with the anytime-valid inference literature, we include this proof for completeness. We begin by stating and proving a lemma that demonstrates how to construct an arbitrary $(1-\alpha)$ Betting-CS for our problem setting. We then construct a Hedged CS, where we specify the capital process, the convex combination, relevant user-specified parameters and invoke our adapted lemma.

C.3 Constructing a $(1-\alpha)$ Betting-CS

Lemma 9 *Assume we observe data following the data generating process of Section 2.1. Assume that $Y_t \in [0,1]$ $\forall t \in 1,\ldots,T$. Suppose that $\pi_t(1 \mid X_t, \Omega_{t-1}) \in [k_t, 1-k_t]$ for all $t \in 1,\ldots,T$, then*

$$C_T^{Betting} := \bigcap_{t \le T} \left\{ \theta' \in [-1,1] : \prod_{t=1}^{T} \left(1 + \lambda_t(\theta')(h_t - \theta')\right) < \frac{1}{\alpha} \right\},$$

forms a $(1-\alpha)$ CS for θ_0, where λ_t is a predictable sequence.

Proof Note that $\pi_t(a \mid X_t, \Omega_{t-1}) \in [\frac{1}{k_t}, 1-\frac{1}{k_t}]$ and consequently $h_t \in [-k_t, k_t]$. Inspired by the truncation technique used by (Waudby-Smith et al., 2024, Theorem 1), we show that $M_T(\theta_0)$ in Equation (34) is a test martingale,

$$M_T(\theta_0) := \prod_{t=1}^{T} \left(1 + \lambda_t(\theta_0)(h_t - \theta_0)\right). \tag{34}$$

For $M_T(\theta_0)$ to be a test martingale, we must show $M_0(\theta_0) = 1$, $\{M_T(\theta_0)\}_{t=1}^{T}$ is non-negative, and that $\mathbb{E}^{T-1}\left(M_T(\theta_0)\right) = M_{T-1}(\theta_0)$.

$M_T(\theta_0)$ is non-negative if $(1 + \lambda_t(\theta_0)(h_t - \theta_0)) > 0$ $\forall t \in 1,\ldots,T$. Waudby-Smith and Ramdas (2023) state this condition in their Proposition 3 as requiring $\lambda_t(\theta_0)\,(h_t - \theta_0) > -1$. Consider the case when $(h_t - \theta_0) < 0$. We have that

$$1 + \lambda_t(\theta_0)(h_t - \theta_0) \ge 1 + \lambda_t(\theta_0)(-k_t - \theta_0).$$

In this case, $\lambda_t(\theta_0) \in (-\infty, \frac{1}{k_t+\theta_0})$ will give

$$1 + \lambda_t(\theta_0)(-k_t - \theta_0) > 1 + \frac{-k_t - \theta_0}{k_t + \theta_0} = 0.$$

Next consider when $(h_t - \theta_0) > 0$, then setting $\lambda_t(\theta_0) \in \left(\frac{-1}{k_t+\theta_0}\right)$ guarantees $\lambda_t(\theta_0)(h_t - \theta_0) > -1$. Taking the union of these sets gives $\lambda_t(\theta_0) \in \left(\frac{-1}{k_t-\theta_0}, \frac{1}{k_t+\theta_0}\right)$, and we conclude that $M_T(\theta_0)$ is non-negative.

Next, we check the condition on the conditional expectation,

$$\begin{aligned}
\mathbb{E}^{T-1}\left(M_T(\theta_0)\right) &= \mathbb{E}^{T-1}\left(M_{T-1}(\theta_0) \times (1 + \lambda_T(\theta_0)(h_T - \theta_0))\right) \\
&= M_{T-1}(\theta_0)(1 + \lambda_T(\theta_0)\mathbb{E}^{T-1}(h_T - \theta_0) \\
&= M_{T-1}(\theta_0)(1 + \lambda_T(\theta_0) \times 0) = M_{T-1}(\theta_0).
\end{aligned}$$

$M_T(\theta_0)$ is therefore a test martingale. By the inequality for non-negative supermartingales due to Ville (1939), we have that

$$\mathbb{P}\left(\exists T \in \mathbb{N}, M_T(\theta_0) \ge \frac{1}{\alpha}\right) \le \alpha.$$

It follows that the set

$$C_T^{\text{Betting}} := \left\{ \theta' \in [-1, 1] : \prod_{t=1}^{T} \Big(1 + \lambda_t(\theta')(h_t - \theta')\Big) < \frac{1}{\alpha} \right\},$$

forms a $(1 - \alpha)$ confidence set. ■

C.4 Hedged CS

Following suggested values from Waudby-Smith and Ramdas (2023), we set

$$\lambda_t = \sqrt{\frac{2\log(2/\alpha)}{\hat{\sigma}_{t-1}^2 t \log(1+t)}} \wedge c, \text{ where } c = 0.5, \tag{35}$$

$$\hat{\theta}_t = \frac{\frac{1}{2} + \sum_{i=1}^{t-1} h_i}{t},$$

$$\hat{\sigma}_t^2 = \frac{\frac{1}{4} + \sum_{i=1}^{t} (h_i - \hat{\theta})^2}{t}.$$

We define

$$\mathcal{K}_T^+(\theta') := \prod_{t=1}^{T} (1 + \lambda_t(\theta')(h_t - \theta')), \quad \mathcal{K}_T^-(\theta') := \prod_{t=1}^{T} (1 - \lambda_t(\theta')(h_t - \theta')),$$

$$\mathcal{M}_T(\theta') := m\mathcal{K}_T^+(\theta') + (1 - m)\mathcal{K}_T^-(\theta'),$$

where $m = 0.5$ (in general, $m \in [0, 1]$). Letting $\lambda_t(\theta') = \lambda_t$ as defined in Equation (35), and truncated to fall within $\left(\frac{-1}{k_t - \theta'}, \frac{1}{k_t + \theta'}\right)$, both $\mathcal{K}_T^+(\theta')$ and $\mathcal{K}_T^-(\theta')$ are test martingales when $\theta' = \theta_0$. It follows that $\mathcal{M}_T(\theta')$ is also a test martingale when $\theta' = \theta_0$ (Waudby-Smith and Ramdas, 2023, Theorem 3). By Lemma 9,

$$C_T^{\text{Hedged}} := \bigcap_{t \le T} \left\{ \theta' \in [-1, 1] : \mathcal{M}_T(\theta') < \frac{1}{\alpha} \right\},$$

forms a valid $(1 - \alpha)$-CS. We now focus computing C_T^{Hedged}.

If λ_t does not depend on θ' (apart from truncating the domain), Waudby-Smith and Ramdas (2023) show that, empirically, C_T^{Hedged} forms an interval at each time T. We can then search over a grid of possible values of $\theta' \in [-1, 1]$, and set lower and upper bounds as

$$L_T^{\text{Hedged}} = \sup_{t \in \{1,\ldots,T\}} \inf_T \left\{ \theta' \in [-1, 1] : \mathcal{M}_T(\theta') < \frac{1}{\alpha} \right\},$$

$$U_T^{\text{Hedged}} = \inf_{t \in \{1,\ldots,T\}} \sup_T \left\{ \theta' \in [-1, 1] : \mathcal{M}_T(\theta') < \frac{1}{\alpha} \right\}.$$

As a result, $[L_T^{\text{Hedged}}, U_T^{\text{Hedged}}]$ forms a $(1 - \alpha)$-CS for θ_0.

Appendix D. Proof of Theorem 3

Proof Outline $(h_t)_{t=1}^T$ is recognized to be a sequence of random variables with conditional mean θ_0 and conditional variance σ^2. This allows us to invoke Theorem 2.5 from Waudby-Smith et al. (2023). In order to do so, we must verify three assumptions.

Assumption 1 (Cumulative variance diverges almost surely) This assumption is satisfied in Appendix A where we establish that the average conditional variance of z_t (which equals the average conditional variance of h_t) does not vanish. It follows that an infinite sum of a non-zero constant diverges.

Assumption 2 (Lindeberg-type uniform integrability) We mush show that there exists some $0 < \kappa < 1$ such that

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}\left[(h_t - \theta_0)^2 \mathbb{1}\left((h_t - \theta_0)^2 > V_t^{\kappa}\right) \mid \Omega_{t-1}\right]}{V_t^{\kappa}} < \infty \text{ almost surely,}$$

where $V_t = \sum_{i=1}^{t} \sigma_i^2$.

As is noted in Waudby-Smith et al. (2023), this equation is satisfied if $1/K \leq \mathbb{E} \mid h_t - \theta_0 \mid^q < K$ a.s. for all $t \geq 1$ and for some constant $K > 0$. Without loss of generality, assume $q = 2 + \delta$. We have that

$$\mathbb{E}|h_t - \theta_0|^q \leq \mathbb{E}(h_t^q) + \mathbb{E}(\theta_0^q).$$

Note that $\mathbb{E}(h_t^q) \propto \mathbb{E}(Y_t^q) < \infty$. Then pick $K^* = K + \mathbb{E}(h_t^q)$ and the condition holds.

Assumption 3 (Consistent variance estimation) We must show that the estimator, $\hat{\sigma}_t^2$, of $\tilde{\sigma}_t^2$ satisfies

$$\frac{\hat{\sigma}_t^2}{\tilde{\sigma}_t^2} \xrightarrow{a.s.} 1.$$

Our estimator is the sample average of the variances estimated thus far. We note that z_t is a square-integrable MDS. Hence, we utilize the Strong Law of Large Numbers for a MDS, and we can establish that the sample average of the squared deviations converges almost surely to the variance of z_t. We establish that $\hat{\sigma}^2(z_t) = \hat{\sigma}^2(h_t)$ by showing

$$\begin{aligned}
\hat{\sigma}^2(z_t) &= \frac{1}{T} \sum_{t=1}^{T} (z_t - \bar{z}_t)^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} \left(h_t - \theta_0 - \frac{1}{T} \sum_{t=1}^{T} (h_t - \theta_0) \right)^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} \left(h_t - \theta_0 + \theta_0 - \bar{h}_T \right)^2 = \hat{\sigma}^2(h_t).
\end{aligned}$$

By the SLLN, $\hat{\sigma}^2(h_t) = \hat{\sigma}^2(z_t) \xrightarrow{a.s.} \mathrm{Var}(z_t) = \mathrm{Var}(h_t)$.

Appendix E. Statement and Proof of Theorem 10

The confidence set produced by Theorem 2 can be computationally expensive, as a grid search is performed over $\theta' \in [-1, 1]$. A significant drawback is a lack of closed-form presentation. We now present a closed-form CS which has slight degradation in performance, but enjoys faster computation. This CS is based on an empirical Bernstein-type process that is shown to be test supermartingale (Waudby-Smith and Ramdas, 2023). Since this process inverts a test supermartingale, the concentration inequality is a looser bound than those produced by test martingales.

Without loss of generality, assume that we observe $Y_t \in [0, 1]$, for all $t \in 1, \ldots, T$, and that the propensity scores, $\pi_t(a_t \mid X_t, \Omega_{t-1})$, are all truncated to fall in $[\frac{1}{k_t}, 1 - \frac{1}{k_t}]$. Following a similar technique as in Waudby-Smith et al. (2024), we define

$$\xi_t = \frac{h_t}{k_t + 1}, \quad \hat{\xi}_{t-1} = \left(\frac{1}{t-1} \sum_{i=1}^{t-1} \xi_i \right) \wedge \frac{1}{k_t + 1}, \quad \text{and} \quad \psi_E(\lambda) = -\log(1 - \lambda) - \lambda. \tag{36}$$

ξ_t can be viewed as a scaled version of h_t. $\hat{\xi}_{t-1}$ is then a sample average of ξ up through observation $t - 1$. By only using previous observations, this value is *predictable*, whereas the quantity $\bar{\xi}_t$, defined below in equation (38), uses the current observation and is therefore not predictable. The scaling in ξ_t and truncation in $\hat{\xi}_{t-1}$ are necessary technical tools to construct a test supermartingale, as shown by Waudby-Smith and Ramdas (2023).

Similarly with the Hedged CS of Theorem 2, there are user-specified parameters, $(\lambda_t)_{t=1}^T$, which have an effect on the finite-sample performance of our forthcoming CS. $(\lambda_t)_{t=1}^T$ can be any $(0, 1)$-valued predictable process. Waudby-Smith et al. (2024) provide an empirically promising setting, inspired by fixed-time empirical Bernstein confidence intervals,

$$\lambda_t = \sqrt{\frac{2 \log(2/\alpha)}{\hat{\sigma}_{t-1}^2 t \log(1 + t)}} \wedge c, \ \text{where } c = 0.5, \tag{37}$$

$$\hat{\sigma}_t^2 = \frac{\sigma_0^2 + \sum_{i=1}^t (\xi_i - \bar{\xi}_i)^2}{t + 1}, \quad \text{and} \quad \bar{\xi}_t = \left(\frac{1}{t} \sum_{i=1}^t \xi_i \right) \wedge \frac{1}{k_t + 1}. \tag{38}$$

$\hat{\sigma}_t^2$ and $\bar{\xi}_t$ can be interpreted as estimates of the mean and variance of ξ. The value σ_0^2 can be viewed as a prior guess for the variance of ξ, and setting $\sigma_0^2 = \frac{1}{4}$ is a reasonable choice. We are now ready to present the CS.

Theorem 10 (Predictable plug-in empirical Bernstein CS [Pr-PI]) *Assume we observe data following the data generating process of Section 2.1. Assume that $Y_t \in [0, 1]$ for all $t \in 1, \ldots, T$. Let ξ_t, $\hat{\xi}_{t-1}$, $\psi_E(\lambda)$, λ_t, $\hat{\sigma}_t^2$, and $\bar{\xi}_t$ be defined as in* (36)*,* (37)*, and* (38) *respectively. We have that*

$$C_T^{PrPI-EB} := \frac{\sum_{t=1}^T \lambda_t \xi_t}{\sum_{t=1}^T \lambda_t / (k_t + 1)} \pm \frac{\log(2/\alpha) + \sum_{t=1}^T \left(\xi_t - \hat{\xi}_{t-1} \right)^2 \psi_E(\lambda_t)}{\sum_{t=1}^T \lambda_t / (k_t + 1)},$$

forms a $(1 - \alpha)$ Predictable Plug-In Empirical Bernstein (PrPI-EB) CS for θ_0.

Proof Note that $\xi_t - \hat{\xi}_{t-1} > -1$. Given this fact, Waudby-Smith et al. (2024, Lemma 1) show that the process

$$M_T = \exp\left\{\sum_{t=1}^{T} \lambda_t \left(\xi_t - \frac{\theta_0}{k_t + 1}\right) - \sum_{t=1}^{T} \left(\xi_t - \hat{\xi}_{t-1}\right)^2 \psi_E(\lambda_t)\right\}, \tag{39}$$

is a test supermartingale. Using Ville's inequality, they invert M_t to form a $(1-\alpha)$-lower CS. We define an $(1-\alpha)$-Upper CS by defining $\xi_t = \frac{-h_t}{k_t+1}$, and apply a union bound, which gives the result.

■

Appendix F. Implementation Details

F.1 Experiment Description

We empirically compare our methods to Kato et al. (2021, Thm. 4). We run two simulations with 1000 iterations each: one with Bernoulli outcomes, and one with continuous, bounded outcomes. 5000 total samples are collected for each iteration and intervals are constructed following each sample. We employ sequential sample-splitting on $\hat{f}$ and $\hat{e}$ to avoid double-dipping and overfitting (Waudby-Smith et al., 2023). Sequential sample-splitting permanently allocates each sample to one of two data-folds upon observation. We fit models for $\hat{f}$ and $\hat{e}$ separately on each fold, giving four models in total. Predictions of $\hat{f}$ and $\hat{e}$ are produced from the model fit from the opposite fold. For an individual observation, we estimate the conditional variance by setting $\hat{v}(a,x) = \hat{e}(a,x) - (\hat{f}(a,x))^2$. When determining π_t^{A2IPW}, $\hat{f}$ and $\hat{e}$ are calculated by averaging predictions of the models from both splits, as this calculation occurs prior to observing and assigning the data point to a split. In our simulations, we clip $\hat{v}$ to be no less than 0.01 to avoid division by zero or negative values. During the first 100 samples, $\hat{f}(1, X_t) = 1$, $\hat{f}(0, X_t) = 0$, and $\pi_t = 0.5$. For policy truncation, we set $k_t = \frac{k_{t-1}}{0.999}$ where $k_1 = 2$. Since the method of Kato et al. (2021) does not utilize time-varying bounds (at least in its present form), using the *worst-case* bound for the propensities is a conservative way to guarantee time-uniform validity of their CS. Using our proposed truncation scheme can then make these confidence sequences extremely wide. To remedy this, we observe that setting $k_t = 5$ works well for Kato et al. (2021, Thm. 4).

F.2 $\hat{\theta}^{\text{A2IPW}}$ T-Statistic

Theorem 1 gives an asymptotic distribution for the $\hat{\theta}^{\text{A2IPW}}$ estimator which depends on σ^2. In practice, we typically do not have access to σ^2 and we must estimate this quantity, denoted as $\hat{\sigma}^2$. With $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, we may invoke Slutsky's Theorem, and use $\hat{\sigma}^2$ in place of σ^2. Similarly to Kato et al. (2021), we call this our t-statistic,

$$\frac{\sqrt{T}(\hat{\theta}^{\text{A2IPW}} - \theta_0)}{\hat{\sigma}^2} \xrightarrow{d} N(0,1).$$

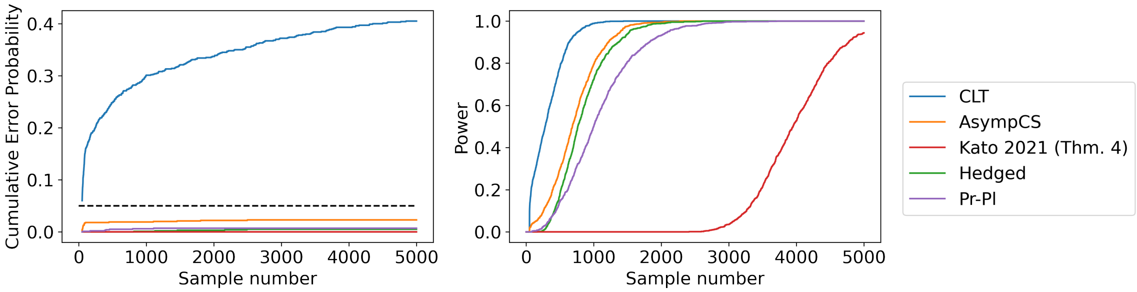

Figure 2: Utilizing an kNN Regressor for the protocol used in Figure 1. The policy used for Pr-Pl is modified to be truncated within $[0.2, 0.8]$.

In Assumption 3 of Appendix D, we show that our variance estimator converges almost surely, implying convergence in probability. Our asymptotic CI is,

$$C_T := \bar{h}_t \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}^2}{\sqrt{T}},$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(h_t - \bar{h}_T\right)^2$.

F.3 Bernoulli Outcome Simulation

We simulate $(X_t, A_t, Y_t)_{t=1}^{T=5000}$, where,

$$\mathbf{X}_t \sim N([\mathbf{0}_3], \mathbf{I}_3),$$

$$\boldsymbol{\beta}^T = \left[-2, -3, 5\right],$$

$$\pi_t = \left(\frac{\sqrt{\hat{v}(1, \mathbf{X}_t)}}{\sqrt{\hat{v}(1, \mathbf{X}_t)} + \sqrt{\hat{v}(0, \mathbf{X}_t)}}\right),$$

$$k_t = \frac{k_{t-1}}{.999},\ k_1 = 2 \text{ if method not Kato, else } k_t = 5,$$

$$A_t \sim \text{Bernoulli}\left(\left(\pi_t \vee \frac{1}{k_t}\right) \wedge (1 - \frac{1}{k_t})\right),$$

$$p_t = 0.9 \times \text{logit}\left(0.5 + \mathbf{X}_t\boldsymbol{\beta}\right) + 0.1A_t,$$

$$Y_t \sim \text{Bernoulli}\left(p = p_t\right).$$

With the data generating process above, $\theta_0 = 0.1$. We ran two separate simulations, where one used k-Nearest Neighbors Regressor (kNN) and the other used Random Forest Regressor (RF) for $\hat{f}$ and $\hat{e}$. We ran 1000 iterations using the DGP above, results for the simulation when RF is used are shown in Figure 1. We provide results for the simulation using kNN in Figure 2.

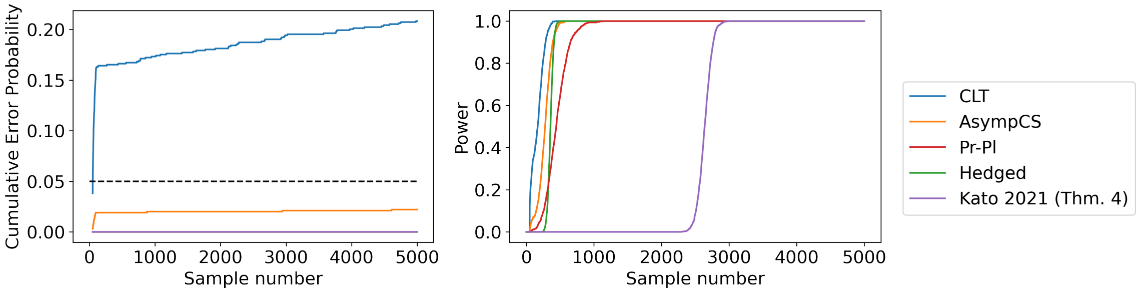

Figure 3: Results for simulation described in Appendix F.4 using a Random Forest Regressor.

F.4 Bounded Continuous Outcomes Simulation

We now consider simulations with a continuous response. Data was simulated as,

$$X_i \sim \text{Uniform}(0,1), \;\; \text{for } i \in \{1,2,3\},$$

$$\pi_t = \left(\frac{\sqrt{\hat{v}(1,\mathbf{X}_t)}}{\sqrt{\hat{v}(1,\mathbf{X}_t)} + \sqrt{\hat{v}(0,\mathbf{X}_t)}} \right),$$

$$k_t = \frac{k_{t-1}}{.999}, \; k_1 = 2 \text{ if method not Kato, else } k_t = 5,$$

$$A_t \sim \text{Bernoulli}\left(\left(\pi_t \vee \frac{1}{k_t} \right) \wedge (1 - \frac{1}{k_t}) \right),$$

$$\boldsymbol{\beta}^T = [-0.04, -0.01, 0.05],$$

$$\epsilon_0 \sim \text{Uniform}(-0.05, 0.05,)$$

$$Y_0 = 0.4 + \mathbf{X}\boldsymbol{\beta} + \epsilon_0,$$

$$\epsilon_1 \sim \text{Uniform}(-4.5\mathbf{X}\boldsymbol{\beta}, 4.5\mathbf{X}\boldsymbol{\beta}),$$

$$Y_1 = 0.4 + \mathbf{X}\boldsymbol{\beta} + \theta_0 + \epsilon_1.$$

In our simulations we set $\theta_0 = 0.1$. Again, we use kNN and RF Regressors to estimate $\hat{f}$ and $\hat{e}$. We ran 1000 iterations using the DGP above, results for the simulation when RF is used are shown in Figure F.4. We provide results for the simulation using kNN in Figure F.4.

F.5 Selecting ρ for an AsympCS

When constructing an AsympCS, the analyst must select a value for ρ. If the analyst wishes to minimize width of the interval produced at a specific sample size, T, then the analyst can accomplish this by setting

$$\rho = \sqrt{\frac{-2\log\alpha + \log(-2\log\alpha + 1)}{T}}.$$

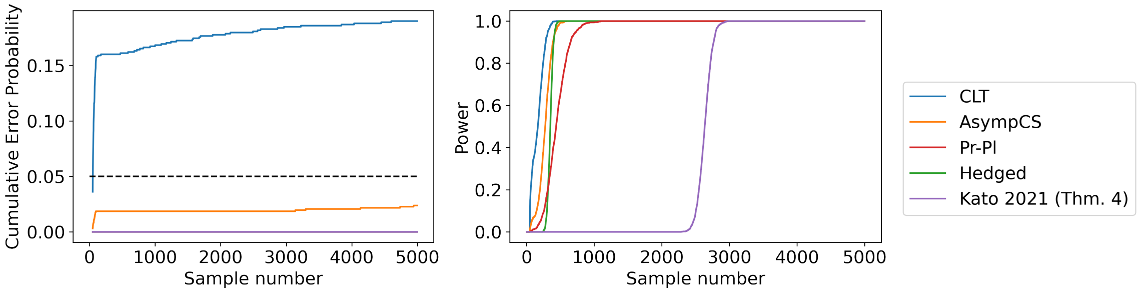

Figure 4: Results for simulation described in Appendix F.4 using a k-Nearest Neighbor regressor.

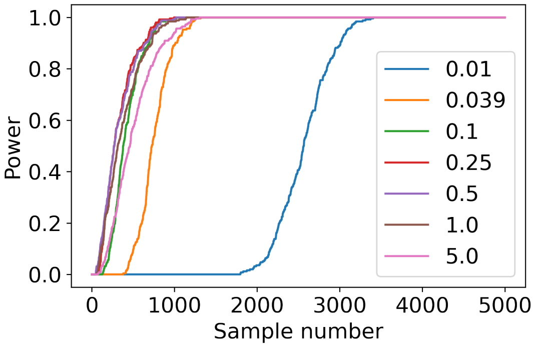

Figure 5: Power curves for AsympCSs constructed using different values of ρ. Curves are based on 256 iterations of the simulation setup described in Appendix F.3.

In practice, the analyst may not have prior knowledge of the effect size magnitude or may not know how long the experiment could last. In this case, it may not be clear for which value of T that ρ should be tuned to. In our simulations we begin constructing CSs at a sample size of $T = 50$. For the sake of simplicity in presentation, we chose to set $\rho = 0.5$ across all experiments. Setting $\rho = 0.5$ yields an AsympCS with tight intervals approximately at the start of inference. To understand the effect of setting $\rho = 0.5$ on the performance of the AsympCS, we performed 256 iterations of the Bernoulli outcome simulation described in Appendix F.3 while varying ρ. We found that setting $\rho = 0.5$ for this scenario is a reasonable choice and the resulting AsympCS produces intervals with widths that allow for high power early in the experiment. Figure 5 shows power curves of the AsympCSs constructed using different levels of ρ.

F.6 Effect of Truncation Schemes

The policy studied in this work is deemed optimal because it minimizes the asymptotic variance of an unbiased estimator. The width of a CI based on the CLT has a direct depen-

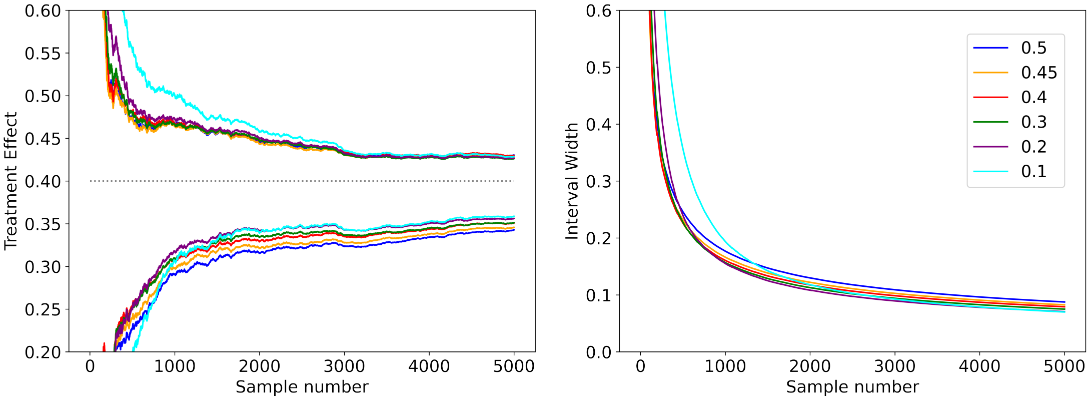

Figure 6: When π_t is bounded in a narrower range, intervals produced by a Pr-PI CS are narrower at smaller t.

dence on the asymptotic variance of the estimator. Naturally, minimizing the asymptotic variance leads to a sense of optimal inference, by minimizing the mean squared error (MSE). In our proof of Theorem 1, we make use of k_t to bound propensity scores away from 0 and 1. In turn, we require that $k_t\|\hat{f}_t - f\|_2 = o_{\mathbb{P}}(1)$ and $k_t\|\pi_t - \pi\|_2 = o_{\mathbb{P}}(1)$. The rate at which k_t increases is limited by the rates that $\|\hat{f}_t - f\|_2 \xrightarrow{p} 0$ and $\|\pi_t - \pi\|_2 \xrightarrow{p} 0$. We also use truncation as a technical tool when considering bounds on $(\lambda_t)_{t=1}^T$ in Theorem 2 and Theorem 10.

Our primary concern in this work lies in anytime-valid inference, and as such, greater attention towards the width of the intervals produced by our CSs at fixed times is warranted. Since the propensity scores set by our policy appear in the denominator of $\hat{\theta}_T^{\mathrm{A2IPW}}$, propensity scores near 0 or 1 can make h_t arbitrarily large. The CSs with fixed-time error control considered in this paper make use of the boundedness of h_t. Particularly, the proofs make use of an underlying test (super)martingale, which by construction, is non-negative. For example, non-negativity is guaranteed by scaling $\lambda_t(\theta')$ such that $\lambda_t(\theta')(h_t - \theta') > -1$ for the Hedged CS. Temporarily subscribing to the betting analogy of Waudby-Smith and Ramdas (2023), an inherent trade-off arises where the analyst must balance the allowable size of their bet, $\lambda_t(\theta')$, with the bounds of the evidence presented by nature, $(h_t - \theta')$. The opportunity to observe large evidence comes at the cost of placing small bets.

This effect is noted explicitly by Waudby-Smith et al. (2024, Remark 2). In our setting, their intuition implies that faster growth in k_t will yield a smaller asymptotic variance at the cost of having wider intervals at small t. In this section, we empirically show that a departure from our optimal policy through truncation will yield narrower intervals at finite times.

We consider a simulation that follows a similar set up to that used in Appendix F.4, where we modify k_t to be constant. Specifically we set $k_t = 1/\pi_{t,min}$ and we vary $\pi_{t,min} \in \{0.5, 0.45, 0.40,\ 0.30, 0.20, 0.10\}$. We note that π^{AIPW}, can be close to 0 or 1, and as a result, we truncate the *optimal* policy. Results of a single iteration are shown in Figure 6.

More aggressive truncation (larger values of $\pi_{t,min}$) leads to narrower intervals for small t, however, once t is sufficiently large, less aggressive truncation (smaller values of $\pi_{t,min}$) provides narrower intervals. These results suggest that optimizing an adaptive policy for statistical inference at finite times is an interesting direction for future work.