
SEMIPARAMETRIC EFFICIENT INFERENCE IN ADAPTIVE EXPERIMENTS

Thomas Cook*

Dept. of Mathematics and Statistics
University of Massachusetts
tjcook@umass.edu

Alan Mishler

J.P. Morgan AI Research
J.P. Morgan Chase & Co.
alan.mishler@jpmchase.com

Aaditya Ramdas

Dept. of Statistics & Data Science and Machine Learning
Carnegie Mellon University
aramdas@cmu.edu

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

Randomized experiments with two treatment arms, also known as A/B tests, are widely used across many domains. Classical statistical tools (*fixed-time* methods) require the analyst to 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 of results, which allows analysts to make repeated decisions about whether to stop or continue 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. When the test statistic is based on a fixed-time method, this can lead to inflated false positive (type-I error) rates. In fact, when used in this fashion, fixed-time methods based on the central limit theorem will in general cause these type-I error rates to go to 1 as $t \rightarrow \infty$, a result that is implied by the law of the iterated logarithm [Robbins, 1952, Johari et al., 2017].

Statistical tools which enable valid inference in this setting are known as *anytime-valid* methods. To illustrate the distinction, consider a confidence interval (CI) for a parameter of interest θ . A $(1 - \alpha)$ CI for θ is an interval $[L_t, U_t]$ based on a sample of size t with the property that

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

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

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

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

The coverage guarantee in (2) is uniform in the sample size, which enables valid inference under data-dependent stopping times. This means that the analyst can continually monitor the experiment and adaptively choose when to stop without inflating the type-I error rate. A CS can be constructed so that time-uniform coverage is guaranteed asymptotically, a notion which is defined rigorously in Waudby-Smith et al. [2023].

Continuous monitoring also enables analysts to adaptively update the policy governing the assignment of subjects to treatment vs. control. Adaptive experiments in general enable more efficient estimation of treatment effects than non-adaptive experiments, such as a traditional experiment which randomly assigns each subject to treatment or control with probability 0.5 [Hahn et al., 2011]. Adaptive designs can therefore enable users to spend less resources running experiments while also minimizing the number of subjects that are exposed to ineffective or possibly harmful treatments. Additionally, mid-stream changes to experimental designs are sometimes imposed by considerations other than statistical efficiency, such as changes to budgets or unexpected impacts of the treatment on a business metric. It is therefore desirable to be able to perform inference under a wide range of adaptive settings.

As an example of this problem setting, consider a pharmaceutical company running a trial to test whether a treatment is effective or not to gain regulatory approval. The pharmaceutical company would like to gain approval quickly by concluding the trial using as few samples as possible. Without prior knowledge of the effect size, choosing an appropriate sample size is difficult. If an overly large sample size is chosen, then the company will have wastefully run the trial longer than necessary, keeping an effective treatment away from patients. However, if the chosen sample size is too small, then the trial may be inconclusive, in which case the entire cost of the trial is wasted. With standard fixed-time inference tools, the company must start the trial from the beginning, or abandon the treatment. If an anytime-valid method is used, then the company can simply continue the trial without worrying about inflating the false positive rate. To protect participants from potentially harmful treatments, it is common to track the results of an experiment as they are observed. It is also possible that a treatment is so overwhelmingly effective that the company would like to conclude the trial immediately and gain regulatory approval. All these settings involve data-dependent stopping times, which require anytime-valid methods rather than fixed-time methods.

In this paper, 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] proposed the Adaptive Augmented IPW (A2IPW) estimator, which, when coupled with a particular adaptive design, yields asymptotically efficient CIs based on the central limit theorem [Hahn et al., 2011]. Furthermore, Kato et al. [2021] analyzed finite-sample regret (in terms of mean squared error) and showed that 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 a similar, but 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 design, the fixed (non-adaptive) design that minimizes the variance of the IPW estimator but is unknown in practice due to dependence on unknown parameters. They provided a treatment assignment policy that achieves sublinear regret, in terms of the estimator’s variance, through the use of a variant of online stochastic projected gradient descent. They also provided an asymptotically-valid Chebyshev-type CI for the ATE. In contrast to our work, their work studied this problem from a design-based potential outcomes framework, which assumes deterministic potential outcomes, and did not consider contexts or covariates. Our work assumes that contexts and potential outcomes are drawn at random. We provide an asymptotically valid Wald-type CI for an A2IPW estimator and provide anytime-valid inference tools.

Our contributions are both theoretical and empirical. Theoretically, we provide semiparametric efficient (approximate) inference at fixed times under weaker (more general) conditions than in the previous literature. Empirically, we adapt state-of-the-art CSs to the setting of adaptive A/B tests and show superior performance in a sequential testing setting using a treatment assignment policy from our fixed-time theory. Explicitly, our contributions are:

- **Theory:** We prove a central limit theorem 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.
- **Empirical Results:** We couple the A2IPW estimator with anytime-valid methods which yield much tighter intervals (more powerful inference) than the methods in Kato et al. [2021]. We derive CSs based on test

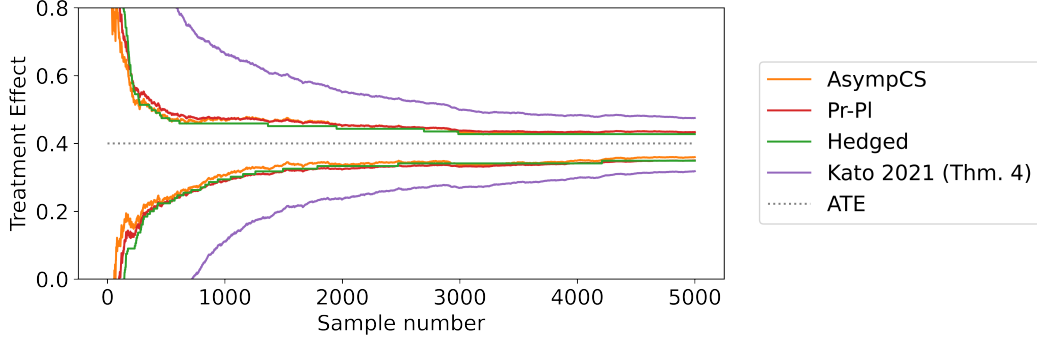


Figure 1: A single run of an experiment with bounded outcomes and the ATE set to 0.4 (simulation setup of Appendix F.4 with $\pi_t \in [0.3, 0.7]$). We propose confidence sequences (AsympCS, Pr-PI, Hedged) that are narrower than previous work Kato et al. [2021].

(super)martingales [Waudby-Smith and Ramdas, 2023], as well as asymptotic CSs [Waudby-Smith et al., 2023].

To offer a demonstrative example, Figure 1 compares our CSs to previous work for a single sequential experiment. Our inference methods provide narrower intervals, leading to higher statistical power.

2 Problem Setting and Fixed-Time Inference

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

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

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 $\mathcal{H}_{t-1}$. We assume that $\{(X_t, Y_t(0), Y_t(1))\}_{t=1}^T$ are independent and identically distributed (iid). However, our treatment assignments are not fixed over time, and depend on previous observations. We define the propensity score $\pi_t(a \mid x, \Omega_{t-1})$ from the experimenter's *policy*, $\pi_t : \mathcal{A} \times \mathcal{X} \times \mathcal{H}_{t-1} \mapsto [0, 1]$. Although the context and potential outcomes are independent over time, the observed outcomes $\{Y_t\}_{t=1}^T$ are dependent due to dependence in the policy.

Remark 1. Although we assume $\{X_t, Y_t(0), Y_t(1)\}_{t=1}^T$ to be iid, our results can be extended to a non-iid setting in which we estimate the running mean of the individual treatment effects, $\frac{1}{T} \sum_{t=1}^T \theta_t$, where $\theta_t = (Y_t(1) - Y_t(0))$ denotes the individual treatment effect for subject t . As a special case of that setting, when $\{X_t, Y_t(0), Y_t(1)\}_{t=1}^T$ are iid, we recover the ATE, since $\frac{1}{T} \sum_{t=1}^T \theta_t \rightarrow \theta_0$.

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 a predetermined sample size and consider proceeding with further data collection, for example because they consider the results inconclusive. This type of data-dependent stopping 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.

2.2 Estimating the Average Treatment Effect

Additional notation: Our notation follows Kato et al. [2021] with minor modification. Let a be a treatment in $\mathcal{A}$. We denote $\mathbb{E}[Y_t(a) \mid x]$, $\mathbb{E}[Y_t^2(a) \mid x]$, $\text{Var}(Y_t(a) \mid x)$, 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.² 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) estimator 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]. Recently, 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 π_t , which can be updated at each time point t based on the accumulated data Ω_{t-1} . The A2IPW estimator, given that T subjects have been observed, is defined as $\hat{\theta}_T^{\text{A2IPW}} = \frac{1}{T} \sum_{t=1}^T h_t$, where

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

Hadad et al. [2021] presented a more general form of an adaptive AIPW estimator. In their definition, the individual iterates are weighted by adaptive evaluation weights to guarantee asymptotic normality of the weighted average. $\hat{\theta}_T^{\text{A2IPW}}$ can therefore be viewed as a special case in which the evaluation weights are set to be equal.

Hahn et al. [2011] showed that the policy that minimizes the semiparametric lower bound of the asymptotic variance for regular estimators of the ATE is π^{AIPW} , defined as

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

Hahn et al. [2011] derived this lower bound in the context of a two-stage experimental design. Armstrong [2022] showed that for estimating the ATE of a binary treatment, no further first-order asymptotic efficiency gain is possible in a purely adaptive experiment. This policy depends on unknown quantities of the underlying data generating process. The policy proposed in Kato et al. [2021] estimates the unknown quantities from the observed data and is defined 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 0.5. 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 define our policy as

$$\pi_t^{\text{A2IPW}}(1 \mid X_t, \Omega_{t-1}) = \left(\pi_t^{\text{A2IPW, Kato}}(1 \mid X_t, \Omega_{t-1}) \vee \frac{1}{k_t} \right) \wedge \left(1 - \frac{1}{k_t} \right), \quad (3)$$

where $k_t \in [2, \infty)$ is a user-chosen truncation parameter. Since our theorem below holds in a more general setting, we can apply this truncation to arbitrary policies $\tilde{\pi}_t$, denoting

$$\pi_t = \left(\tilde{\pi}_t \vee \frac{1}{k_t} \right) \wedge \left(1 - \frac{1}{k_t} \right). \quad (4)$$

Note that setting $k_t = 2$ results in $\pi_t(A_t = 0 \mid X_t, \Omega_{t-1}) = \pi_t(A_t = 1 \mid X_t, \Omega_{t-1}) = 0.5$, and $k_t \rightarrow \infty$ results in the non-truncated policy $\tilde{\pi}_t$. 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. Our empirical results show that truncation can improve finite-sample performance for well-chosen k_t , which aligns with the results in the preceding work.

²In general, $\hat{f}$ can be any arbitrary estimator. In Theorem 1, we simply require $\hat{f}$ to be consistent for f .

2.3 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 central limit theorem to show $\hat{\theta}_T^{\text{A2IPW}}$ is asymptotically Gaussian. They further showed that $\hat{\theta}_T^{\text{A2IPW}}$ 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 $\tilde{\pi}_t : \mathcal{A} \times \mathcal{X} \mapsto (0, 1)$ be an arbitrary sequence of policies, and let π_t be the corresponding truncated policies as defined in (4). Assume $\frac{1}{\pi(a|x)} < C_1 < \infty$ and $v(a, x) < C_2 < \infty$ for all $x \in \mathcal{X}$ and $a \in \{0, 1\}$ for some constants C_1 and C_2 . Assume $k_t \|\hat{f}_t - f\|_2 = o_{\mathbb{P}}(1)$ and $k_t \|\pi_t - \pi\|_2 = o_{\mathbb{P}}(1)$ for some policy π . Further, assume that $\text{Var}(\hat{f}_t | \Omega_{t-1}) < C_3 < \infty$ and $\text{Var}(\pi_t | \Omega_{t-1}) < C_4 < \infty$ for all $x \in \mathcal{X}$, $a \in \{0, 1\}$, and $t \in \{1, 2, \dots\}$, for some constants C_3 and C_4 . 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 $\pi(a | x)$, given by

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

In particular, if we have $\pi = \pi^{\text{AIPW}}$, then $\hat{\theta}_T^{\text{A2IPW}}$ is semiparametric efficient.

Proof is provided in Appendix A. Note that Kato et al. [2021] assumed that Y_t and $\hat{f}_t$ are uniformly bounded and that $\hat{f}_t$ and π_t converge pointwise. By contrast, we only assume that $v(a, x)$ are uniformly bounded, that $\hat{f}_t$ and π_t converge in ℓ_2 norm, and that $\hat{f}_t$ and π_t have uniformly bounded conditional variances. Kato et al. [2021] also assumed that π_t is uniformly bounded away from 0, whereas we only make this assumption on π , the stationary policy which π_t converges to. By utilizing truncation, we can satisfy the assumption that the policies π_t (at fixed times) are uniformly bounded. However, truncation is still relevant to relax convergence assumptions. Our proof uses a MDS central limit theorem given by Dvoretzky [1972], which is used in a similar fashion by Zhang et al. [2021]. This form of a MDS central limit theorem states conditions based on the second moment of z_t conditional on Ω_{t-1} , and lends itself directly to weakening pointwise convergence assumptions on $\hat{f}$ and π_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 $k_t \max(\|\hat{f}_t - f\|_2, \|\pi_t - \pi\|_2) = o_{\mathbb{P}}(1)$). If, instead, 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 our proposed policy π_t^{A2IPW} .

Remark 2 (Semiparametric Efficiency). *Assume that we set k_t such that $\lim_{t \rightarrow \infty} k_t > \sup \frac{1}{\pi^{\text{AIPW}}}$. 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 2 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 2 depends on the rate that $\hat{v}_t$ converges to v . In practice this rate is unobservable, and it is worth acknowledging this limitation. Addressing this limitation is an interesting direction for future research.

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 the potential outcomes follow a distribution without bounded support.

3 Anytime-Valid Inference in Adaptive Experiments

We now construct confidence sequences (CSs) for the ATE that utilize the A2IPW estimator. Kato et al. [2021] developed such CSs via concentration inequalities based on the law of the iterated logarithm (LIL) as derived in Balsubramani [2015] and Balsubramani and Ramdas [2016], which are today known to be loose in constants [Howard et al., 2021]. The concentration inequality derived for $\hat{\theta}_T^{\text{A2IPW}}$ [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 $\{h_t\}_{t=1}^{T-1}$.

In contrast, we present CSs for the ATE based on more recent, state-of-the-art methods for inference of means of random variables in sequential settings [Waudby-Smith and Ramdas, 2023, Waudby-Smith et al., 2024]. All our sequences are *fully empirical*, meaning they do not depend on unknown parameters. We will see that these methods empirically yield much tighter intervals than those previously derived. This section thus yields the first practically tight and theoretically sound CSs for using the semiparametric efficient A2IPW estimator in adaptive experiments.

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 accumulated wealth of a gambler playing 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 the respective capital process has not exceeded an improbable level of wealth for a fair game parameterized by θ' . By constructing the capital process at θ_0 to be a test martingale, the probability that the capital process *ever* exceeds the value $\frac{1}{\alpha}$ is bounded by α [Ville, 1939]. This gives time-uniform type-1 error control, allowing the analyst to reject any θ' for which the respective capital process exceeds $\frac{1}{\alpha}$.

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 $\pi_t(1 \mid X_t, \Omega_{t-1}) \in [\frac{1}{k_t}, 1 - \frac{1}{k_t}]$ for all $t \in 1, \dots, T$. If we define*

$$\begin{aligned} \mathcal{K}_T^+(\theta') &:= \prod_{t=1}^T (1 + \lambda_t(\theta')(h_t - \theta')), & \mathcal{K}_T^-(\theta') &:= \prod_{t=1}^T (1 - \lambda_t(\theta')(h_t - \theta')), \\ \mathcal{M}_T(\theta') &:= \frac{\mathcal{K}_T^+(\theta') + \mathcal{K}_T^-(\theta')}{2}, \end{aligned}$$

then

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

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 and other details can be found in Appendix C. Before providing intuition for the proof, we note that although we assume $Y_t \in [0, 1]$, our result holds for any bounded Y_t by rescaling.

$\mathcal{K}_T^+(\theta')$ 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. 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 . For brevity, we defer the details of setting λ_t to Appendix C.3.

3.2 Empirical Bernstein Confidence Sequences

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. In this section we 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 a 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 $t \in 1, \dots, T$, and that the propensity scores, π_t , 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

$$\xi_t = \frac{h_t}{k_t + 1}, \quad \hat{\xi}_{t-1} = \left(\frac{1}{t-1} \sum_{i=1}^{t-1} \xi_i \right) \wedge \frac{1}{k_t + 1}, \quad \text{and} \quad \psi_E(\lambda) = -\log(1 - \lambda) - \lambda. \quad (5)$$

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 (7), 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 to 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 CIs,

$$\lambda_t = \sqrt{\frac{2 \log(2/\alpha)}{\hat{\sigma}_{t-1}^2 t \log(1+t)}} \wedge c, \quad \text{where } c = 0.5, \quad (6)$$

$$\hat{\sigma}_t^2 = \frac{\sigma_0^2 + \sum_{i=1}^t (\xi_i - \bar{\xi}_t)^2}{t+1}, \quad \text{and} \quad \bar{\xi}_t = \left(\frac{1}{t} \sum_{i=1}^t \xi_i \right) \wedge \frac{1}{k_t + 1}. \quad (7)$$

$\hat{\sigma}_t^2$ 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 3 (Predictable Plug-In Empirical Bernstein CS [Pr-PI]). *Assume we observe data following the data generating process of Section 2.1. Assume that $Y_t \in [0, 1]$ and $\pi_t(1 | X_t, \Omega_{t-1}) \in [\frac{1}{k_t}, 1 - \frac{1}{k_t}]$ for all $t \in 1, \dots, T$. Let $\xi_t, \hat{\xi}_{t-1}, \psi_E(\lambda), \lambda_t, \hat{\sigma}_t^2$, and $\bar{\xi}_t$ be defined as in (5), (6), and (7) respectively. We have that*

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

forms a $(1 - \alpha)$ CS for θ_0 .

Proof is provided in Appendix D, and follows the truncation technique used in Waudby-Smith et al. [2024]. When k_t does not grow quickly, meaning propensities are truncated more aggressively, intervals tend to be narrower at small t . When k_t grows quickly, then finite sample performance is sacrificed for performance at large t . The effect of truncation is consistent with the Hedged-CS.

3.3 Asymptotic Confidence Sequences

Because of their stronger time-uniform guarantees, the CSs in the preceding section produce intervals that have larger widths than their fixed-time CI counterparts. In the fixed-time setting, exact error coverage is not guaranteed, and the analyst must rely on asymptotic coverage guarantees. Waudby-Smith et al. [2023] introduce the sequential analogue of asymptotic CIs, asymptotic CSs (AsympCS), by defining a sequence of intervals which converges to some (unknown) CS. We now define our AsympCS for θ_0 .

Theorem 4 (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 $\text{Var}(h_t)$, and $\rho > 0$ be a user-specified parameter. For all $t \in 1, \dots, T$, we have that*

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

forms a $(1 - \alpha)$ -AsympCS for θ_0 . Further, the width of C_T^{AsympCS} at time T can be (approximately) minimized by setting

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

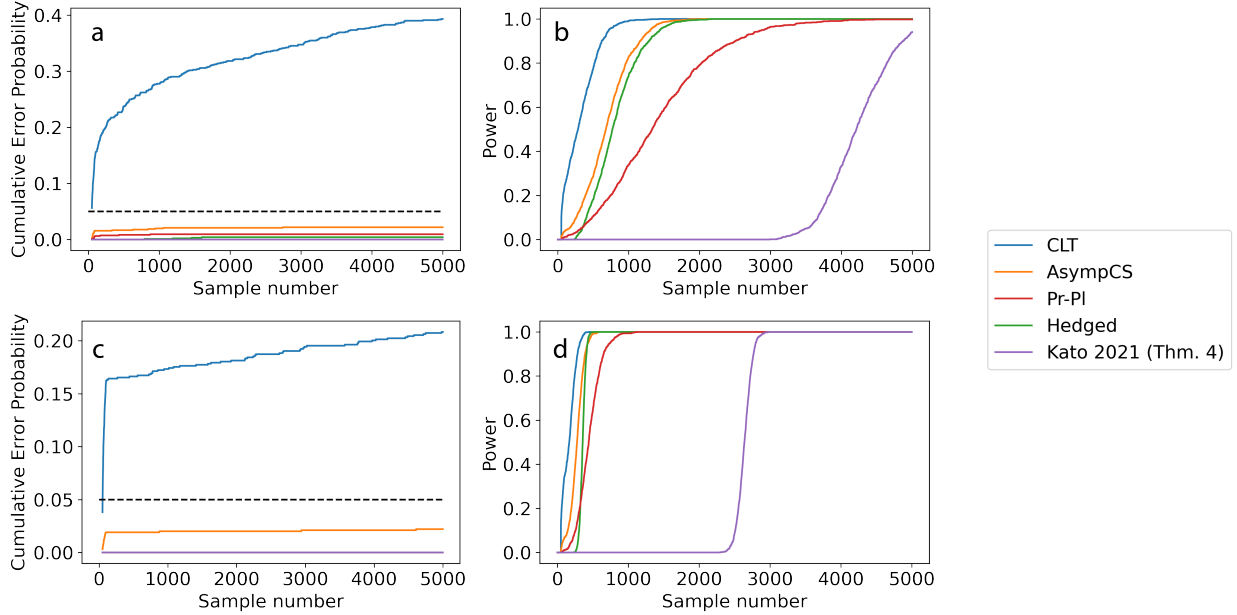


Figure 2: Cumulative error probability (a, c) and power (b, d) as functions of sample size, of experiments from Appendix F.3 and Appendix F.4. The first row corresponds to the experiment with Bernoulli outcome, and the bottom row corresponds to the experiment with bounded, continuous outcomes. Intervals based on the CLT (Theorem 1), AsympCS (Theorem 4), Pr-PI (Theorem 3), Hedged (Theorem 2), and [Kato et al., 2021, Theorem 4] begin at $t = 50$.

Proof is provided in Appendix E. The value of ρ that minimizes interval width at a certain sample size follows from Waudby-Smith et al. [2023, Appendix B.2]. 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. The only user-specified parameter is ρ , a positive scalar which specifies at what intrinsic time the AsympCS should be tightest, with lower values corresponding to tightness at earlier times. We set this parameter to 0.5 in our simulations; for more details, see Waudby-Smith et al. [2023]. 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 4 can be extended to time-varying treatment effects, which we leave for future work.

4 Empirical Results

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 CSs extremely wide. To remedy this, we observe that setting $k_t = 5$ works well for Kato et al. [2021, Thm. 4]. Figure 2 shows results for these simulations when a Random Forest is used for $\hat{f}$ and $\hat{e}$, and are discussed in detail in the subsequent subsections. Appendix F provides additional results using a k-Nearest Neighbors model.

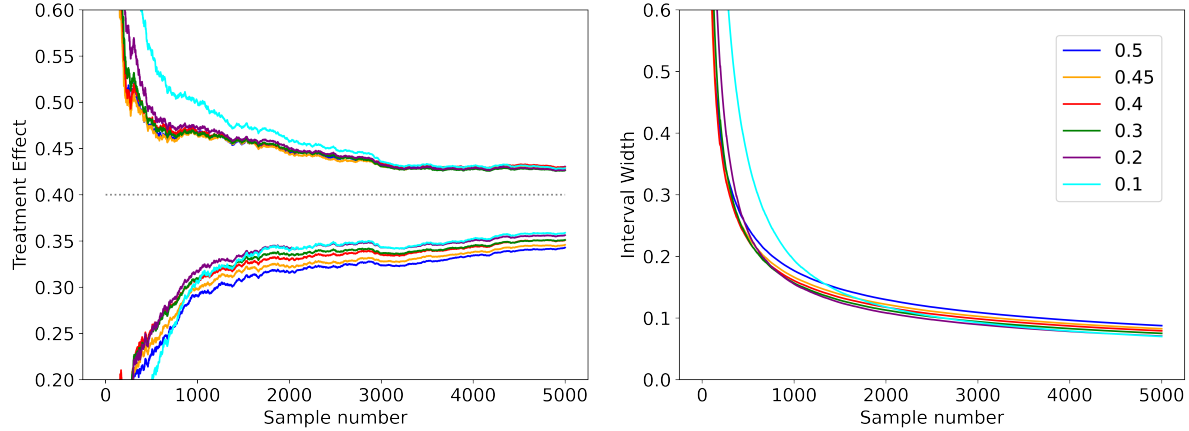


Figure 3: When π_t is bounded in a narrower range, intervals produced by a Pr-PI CS are narrower at smaller t .

4.1 Bernoulli Outcomes

In Figure 2, plots (a) and (b) show aggregated results across 1000 iterations of a simulation with Bernoulli outcomes where the $ATE = 0.1$. Full details of the data generating process can be found in Appendix F.3. Our methods provide significantly narrower intervals due to leveraging tighter concentration inequalities, as well as using time-varying truncation. Performance between our proposed CSs is inline with expectations from the CS literature. Specifically, the AsympCS provides the narrowest interval, at the expense of higher miscoverage probabilities. Since we begin constructing intervals at $t = 50$, we avoid experiencing severe interval miscoverage early in the experiment, which prevents inflating the cumulative miscoverage rate. By doing so, we have empirically controlled the time-uniform error probability at level $\alpha = 0.05$. The Hedged-CS provides tighter intervals than the closed form Pr-PI. This is because the Hedged-CS inverts a test martingale instead of a test supermartingale. Waudby-Smith and Ramdas [2023] note that this removes a source of conservatism in generating CSs; however, this comes at a higher computational cost.

4.2 Bounded Outcomes

Plots (c) and (d) in Figure 2 show aggregated results across 1000 iterations of a simulation with a bounded, continuous outcomes where the $ATE = 0.1$. The data generating process can be found in Appendix F.4. Results in this section follow an identical pattern to the results in the Bernoulli experiment.

The data generating process yields data which is noticeably heteroskedastic. This implies that π^{AIPW} will produce values that are close to 0 and 1. However, when t is small, $\hat{f}$ and $\hat{e}$ can be noisy estimates of the true, unobservable f and e . In this case, the truncation schedule plays an interesting balance of preventing noisy estimates of f and e from inducing high variance in $\hat{h}_t$, while still allowing π_t to converge to the optimal policy at an appreciable rate.

4.3 Effect of Truncation Schemes

The policy studied in this work is deemed optimal because it minimizes the asymptotic variance of an unbiased estimator, the A2IPW estimator. The width of a CI based on the CLT has a direct dependence 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 3.

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 $\hat{\theta}_T^{\text{A2IPW}}$, 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 Section 4.2, where we modify k_t to be constant. Specifically we set $k_t = 1/\pi_{t,min}$ and we vary $\pi_{t,min} \in \{0.5, 0.45, 0.40, 0.30, 0.20, 0.10\}$. For an explicit data generating process, see Appendix F.5. 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 3. 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.

5 Conclusion

We have provided both confidence intervals (CIs) and confidence sequences (CSs) for the ATE in adaptive experiments using the A2IPW estimator. The CI based on the CLT achieves the semiparametric lower bound of the asymptotic variance under weaker assumptions than in previous work. The CSs with time-uniform error control surpass the performance of previous work considerably. Our methods apply to arbitrary adaptive designs, but we also propose a particular policy truncation scheme that preserves the asymptotic efficiency of the A2IPW estimator while improving finite sample performance. We emphasize that the inference tools (the CIs and CSs) and our proposed policy are individual contributions which do not require the use of one another.

This work provides a clear framework for using the A2IPW estimator in adaptive experiments. There are many interesting directions for further research. As mentioned in Section 4.3, the truncation scheme used can have a considerable impact on finite sample performance. In the context of bounded random variables, Shekhar and Ramdas [2023] derive lower bounds on the width of a CS and show that betting-based confidence sets are nearly optimal. In our adaptive experiment, the analyst sets the bounds of the observed random variable h_t , and could explore minimizing the lower bound of the width of our CSs using their results. Separately, our adaptive design focuses purely on efficient statistical inference. In bandit experiments, an analyst may wish to minimize the *regret*, so that patient welfare is maximized by treatment assignments. An interesting direction for future work is incorporating a notion of regret in the treatment assignments, such as the scheme proposed by Simchi-Levi and Wang [2023]. Lastly, we have only considering performing inference on the ATE. An interesting line of future work would be extending these results to other causal estimands.

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

References

- Timothy B. Armstrong. Asymptotic efficiency bounds for a class of experimental designs. *arXiv preprint 2205.02726*, 2022.
- 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.
- Vitor Hadad, David A. Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences*, 118(15), 2021.
- Jinyong Hahn, Keisuke Hirano, and Dean Karlan. Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, 29(1):96–108, 2011.
- Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2), 2021.
- Ramesh Johari, Pete Koomen, Leonid Pekelis, and David Walsh. Peeking at A/B Tests: Why it matters, and what to do about it. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017.
- Masahiro Kato, Takuya Ishihara, Junya Honda, and Yusuke Narita. Efficient adaptive experimental design for average treatment effect estimation. *arXiv preprint 2002.05308*, 2021.
- M. Loève. *Probability Theory*. Graduate texts in mathematics. Springer, 1977.
- Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 2023.
- Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527 – 535, 1952.
- 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.
- Shubhanshu Shekhar and Aaditya Ramdas. On the near-optimality of betting confidence sets for bounded means. *arXiv preprint 2310.01547*, 2023.
- David Simchi-Levi and Chonghuan Wang. Multi-armed bandit experimental design: Online decision-making and adaptive inference. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206, 2023.
- 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*, 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.

A Proof of Theorem 1

A.1 High-Level Roadmap

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.2 Assumptions

- IID Contexts and Potential Outcomes : $\{X_t, Y_t(0), Y_t(1)\}_{t=1}^T$ are independent and identically distributed.
- 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 all $x \in \mathcal{X}$, and $a \in \{0, 1\}$, where C_1 is a constant such that $C_1 < \infty$.
- Finite Conditional Variance : $v(a, x) < C_2$ for all $x \in \mathcal{X}$, and $a \in \{0, 1\}$, where C_2 is a constant such that $C_2 < \infty$.
- Finite Conditional Variance of Predictions : $\text{Var}(\hat{f}_{t-1}(a, X_t) \mid \Omega_{t-1}) < C_3$ for some $C_3 < \infty$ for all $x \in \mathcal{X}$, $a \in \{0, 1\}$, and $t \in \{1, 2, \dots\}$.
- Finite Conditional Variance of Policy : $\text{Var}(\pi_t \mid \Omega_{t-1}) < C_4$ for some $C_4 < \infty$ for all $x \in \mathcal{X}$, $a \in \{0, 1\}$, and $t \in \{1, 2, \dots\}$.

A.3 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.

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

The second required condition is $\mathbb{E}|z_t| < \infty$. By assumption $\mathbb{E}(z_t^2 \mid \Omega_{t-1}) < M < \infty$, where M is some constant. This follows by uniformly bounded variance assumptions, since $\mathbb{E}(z_t \mid \Omega_{t-1}) = 0$. Since $\mathbb{E}(z_t^2) = \mathbb{E}(\mathbb{E}(z_t^2 \mid \Omega_{t-1}))$, it follows that $\mathbb{E}(z_t^2) < \infty$, since the expectation of a uniformly bounded variable is bounded. This implies existence of the first moment.

A.4 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. $\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\mathcal{P}, \eta} [z_t^2 \mid \Omega_{t-1}] \xrightarrow{p} \sigma^2$ uniformly over $\mathcal{P} \in \mathbb{P}$, where σ^2 is a constant $0 < \sigma^2 < \infty$, and that,

2. for any $\epsilon > 0$, $\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\mathcal{P}, \eta} [z_t(\mathcal{P})^2 \mathbb{1} [|z_t(\mathcal{P})| > \epsilon] \mid \Omega_{t-1}] \xrightarrow{p} 0$ 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.4.1 Condition 1 (Conditional Variance)

We wish to show that

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

This is equivalent to showing

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

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

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

$$+ 2\mathbb{E}^{t-1} \left[(f(1, X_t) - f(0, X_t) - \hat{f}_{t-1}(1, X_t) + \hat{f}_{t-1}(0, X_t)) (\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0) \right] \quad (9)$$

$$+ \mathbb{E}^{t-1} \left[(\hat{f}_{t-1}(1, X_t) - \hat{f}_{t-1}(0, X_t) - \theta_0)^2 - (f(1, X_t) - f(0, X_t) - \theta_0)^2 \right]. \quad (10)$$

We now consider terms (8), (9) and (10) 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.2.

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

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

Lemma 2 (Convergence of (9)). *Under the assumptions of Theorem 1, we have*

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

Lemma 3 (Convergence of (10)). *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 1, 2, and 3, 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. So far we have shown that $\mathbb{E}^{t-1} [z_t^2] \xrightarrow{p} \sigma^2$, but we wish to show that $\frac{1}{T} \sum_{t=1}^T (\mathbb{E}^{t-1} [z_t^2] - \sigma^2) \xrightarrow{p} 0$.

Let $a_t = \mathbb{E}^{t-1} [z_t^2]$ and $a = \sigma^2$. It follows that $\mathbb{E}[a_t] = \mathbb{E}[|a_t|] < M < \infty$ for all t and some constant M , where the equality holds since $a_t > 0$. This uniform boundedness implies that a_t is uniformly integrable. Since it is also true that $a_t \xrightarrow{p} a$, then by the L^R convergence theorem, we have that $a_t \rightarrow a$ in L^1 , which implies $\mathbb{E} [|\mathbb{E}^{t-1}[z_t^2] - \sigma^2|] \rightarrow 0$ [Loève, 1977].

Kato et al. [2021] show through Markov's inequality that $\frac{1}{T} \sum_{t=1}^T (\mathbb{E}^{t-1} [z_t^2] - \sigma^2) \xrightarrow{p} 0$ if $\mathbb{E} [|\mathbb{E}^{t-1}[z_t^2] - \sigma^2|] \rightarrow 0$. So condition 1 is satisfied.

A.4.2 Condition 2 (Conditional Lindeberg)

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

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

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,

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

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

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

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

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

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

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 1

The term considered in Lemma 1, specifically term (8), 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.

$$\begin{aligned} & \mathbb{E}^{t-1} \left[\frac{\left(Y_t(a) - \hat{f}_{t-1}(a, X_t) \right)^2}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{(Y_t(a) - f(a, X_t))^2}{\pi(a \mid X_t)} \right] \\ &= \mathbb{E}^{t-1} \left[\frac{\left(Y_t(a) - \hat{f}_{t-1}(a, X_t) + f(a, X_t) - f(a, X_t) \right)^2}{\pi_t(a \mid X_t, \Omega_{t-1})} - \frac{(Y_t(a) - f(a, X_t))^2}{\pi(a \mid X_t)} \right] \end{aligned} \quad (11)$$

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

$$= \mathbb{E}^{t-1} \left[\frac{(Y_t(a) - f(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} + \frac{2(Y_t(a) - f(a, X_t))(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} \right. \\ \left. + \frac{(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{(Y_t(a) - f(a, X_t))^2}{\pi(a | X_t)} \right] \quad (13)$$

$$= \mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \\ + 2\mathbb{E}^{t-1} \left[\frac{(Y_t(a) - f(a, X_t))(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} + \frac{(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} \right]. \quad (14)$$

Above, (11) simultaneously adds and subtracts $f(a, X_t)$, while (12) and (13) square the binomial term. In equation (14), we factor $(Y_t(a) - f(a, X_t))^2$ from the first and final terms, and utilize linearity of expectation. Continuing,

$$\mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \\ + 2\mathbb{E}^{t-1} \left[\frac{(Y_t(a) - f(a, X_t))(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} + \frac{(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} \right] \\ \leq \mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \quad (15)$$

$$+ 2\mathbb{E}^{t-1} \left[\frac{(Y_t(a) - f(a, X_t))(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} \right] + \mathbb{E}^{t-1} \left[k_t (f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2 \right] \\ = \mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \quad (16) \\ + 2\mathbb{E}^{t-1} \left[\frac{(Y_t(a) - f(a, X_t))(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} \right] + k_t (\|\hat{f}(a, X_t) - f(a, X_t)\|_2)^2.$$

Inequality (15) follows since our policy is truncated within $[\frac{1}{k_t}, 1 - \frac{1}{k_t}]$. Equation (16) 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 (16),

$$\mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \\ + 2\mathbb{E}^{t-1} \left[\frac{(Y_t(a) - f(a, X_t))(f(a, X_t) - \hat{f}_{t-1}(a, X_t))^2}{\pi_t(a | X_t, \Omega_{t-1})} \right] + k_t (\|\hat{f}(a, X_t) - f(a, X_t)\|_2)^2 \\ = \mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \quad (17) \\ + 2 \left(\mathbb{E}^{t-1} \left[\frac{Y_t(a)}{\pi_t(a | X_t, \Omega_{t-1})} (f(a, X_t) - \hat{f}(a, X_t))^2 \right] - \mathbb{E}^{t-1} \left[\frac{f(a, X_t)}{\pi_t(a | X_t, \Omega_{t-1})} (f(a, X_t) - \hat{f}(a, X_t))^2 \right] \right) \\ + k_t (\|\hat{f}(a) - f(a)\|_2)^2$$

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

$$\begin{aligned}
 &- \mathbb{E}^{t-1} \left[\mathbb{E} \left[\frac{f(a, X_t)}{\pi_t(a | X_t, \Omega_{t-1})} \left(\hat{f}(a, X_t) - f(a, X_t) \right)^2 \mid X_t, \Omega_{t-1} \right] \right] \\
 &+ k_t \left(\|\hat{f}(a, X_t) - f(a, X_t)\|_2 \right)^2,
 \end{aligned} \tag{19}$$

where equation (17) expands the term of interest, and terms (18) and (19) apply the law of iterated expectation. Conditioning on X_t , the only non-constant term in terms (18) and (19) is $Y_t(a)$, whose conditional expectation on X_t is $f(a, X_t)$. Therefore, the terms (18) and (19) reduce to 0. Simplifying, we have

$$\begin{aligned}
 &\mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \\
 &+ 2 \left(\mathbb{E}^{t-1} \left[\mathbb{E} \left[\frac{Y_t(a)}{\pi_t(a | X_t, \Omega_{t-1})} \left(f(a, X_t) - \hat{f}(a, X_t) \right)^2 \mid X_t, \Omega_{t-1} \right] \right] \right. \\
 &\quad \left. - \mathbb{E}^{t-1} \left[\mathbb{E} \left[\frac{f(a, X_t)}{\pi_t(a | X_t, \Omega_{t-1})} \left(\hat{f}(a, X_t) - f(a, X_t) \right)^2 \mid X_t, \Omega_{t-1} \right] \right] \right) \\
 &+ k_t \left(\|\hat{f}(a, X_t) - f(a, X_t)\|_2 \right)^2 \\
 &= \mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \\
 &+ k_t \left(\|\hat{f}(a, X_t) - f(a, X_t)\|_2 \right)^2.
 \end{aligned} \tag{20}$$

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 (20) can be further simplified to

$$\begin{aligned}
 &\mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] \\
 &+ k_t \left(\|\hat{f}(a, X_t) - f(a, X_t)\|_2 \right)^2 \\
 &= \mathbb{E}^{t-1} \left[(Y_t(a) - f(a, X_t))^2 \left(\frac{1}{\pi_t(a | X_t, \Omega_{t-1})} - \frac{1}{\pi(a | X_t)} \right) \right] + o_{\mathbb{P}}(1) \\
 &= \mathbb{E} \left[\mathbb{E} \left[\frac{\pi(a | X_t) - \pi_t(a | X_t, \Omega_{t-1})}{\pi(a | X_t) \pi_t(a | X_t, \Omega_{t-1})} (Y_t(a) - f(a, X_t))^2 \mid X_t, \Omega_{t-1} \right] \mid \Omega_{t-1} \right] + o_{\mathbb{P}}(1)
 \end{aligned} \tag{21}$$

$$= \mathbb{E} \left[\frac{\pi(a | X_t) - \pi_t(a | X_t, \Omega_{t-1})}{\pi(a | X_t) \pi_t(a | X_t, \Omega_{t-1})} \mathbb{E} \left[(Y_t(a) - f(a, X_t))^2 \mid X_t, \Omega_{t-1} \right] \mid \Omega_{t-1} \right] + o_{\mathbb{P}}(1) \tag{22}$$

$$\leq C_1 C_2 k_t \mathbb{E}^{t-1} [\pi(a | X_t) - \pi_t(a | X_t, \Omega_{t-1})] + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1). \tag{23}$$

Equation (21) follows from the law of total expectation. In equation (22) $\pi_t(a | X_t, \Omega_{t-1})$ 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 (23) then utilizes our policy truncation and our assumption that $\frac{1}{\pi}$ is uniformly bounded by C_1 . 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 $v(a, x)$. By assumption, $v(a, x)$ is bounded uniformly by $C_2 < \infty$. The bound 23 reduces to $o_{\mathbb{P}}(1)$, since convergence in ℓ_2 implies convergence in ℓ_1 , and the Lemma is proved.

B.2 Proof of Lemma 2

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,

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

$$\begin{aligned} & \leq \sqrt{\mathbb{E}^{t-1} \left[\left(f(1, X_t) - \hat{f}_{t-1}(1, X_t) \right)^2 \right] \mathbb{E}^{t-1} \left[(f(1, X_t) - f(0, X_t) - \theta_0)^2 \right]} \\ & \quad + \sqrt{\mathbb{E}^{t-1} \left[\left(f(0, X_t) - \hat{f}_{t-1}(0, X_t) \right)^2 \right] \mathbb{E}^{t-1} \left[(f(1, X_t) - f(0, X_t) - \theta_0)^2 \right]}, \end{aligned} \quad (25)$$

where equation (24) separates terms from different treatments and utilizes the linearity of expectation. Bound (25) then follows from applying the Cauchy-Schwarz inequality. We conclude by showing

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

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 3

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,

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

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

Equation (26) arises from the fact that $a^2 - b^2 = (a + b)(a - b)$ for real numbers a, b . Equation (27) 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 (27), giving

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

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

Equation (28) completes this step, and rearranges terms so that we may use the assumption of convergence of regression. Next, distributing the second term in equation (28), along with the use of the linearity of expectation gives

$$\begin{aligned}
& \mathbb{E}^{t-1} \left[\left(\left(\hat{f}(1, X_t) - f(1, X_t) \right) + \left(f(0, X_t) - \hat{f}(0, X_t) \right) + 2(f(1, X_t) - f(0, X_t) - \theta_0) \right) \right. \\
& \quad \left. \times \left(\left(\hat{f}(1, X_t) - f(1, X_t) \right) + \left(f(0, X_t) - \hat{f}(0, X_t) \right) \right) \right] \\
& = \mathbb{E}^{t-1} \left[\left(\left(\hat{f}(1, X_t) - f(1, X_t) \right) + \left(f(0, X_t) - \hat{f}(0, X_t) \right) \right)^2 \right] \\
& \quad + 2\mathbb{E}^{t-1} \left[\left(f(1, X_t) - f(0, X_t) - \theta_0 \right) \left(\left(\hat{f}(1, X_t) - f(1, X_t) \right) + \left(f(0, X_t) - \hat{f}(0, X_t) \right) \right) \right]. \tag{29}
\end{aligned}$$

The first term in equation (29) converges in probability by assumption. For the second term, we distribute $f(1, X_t) - f(0, X_t) - \theta_0$ yielding

$$\mathbb{E}^{t-1} \left[\left(\left(\hat{f}(1, X_t) - f(1, X_t) \right) + \left(f(0, X_t) - \hat{f}(0, X_t) \right) \right)^2 \right] \tag{30}$$

$$\begin{aligned}
& + 2\mathbb{E}^{t-1} \left[\left(f(1, X_t) - f(0, X_t) - \theta_0 \right) \left(\left(\hat{f}(1, X_t) - f(1, X_t) \right) + \left(f(0, X_t) - \hat{f}(0, X_t) \right) \right) \right] \\
& = \mathbb{E}^{t-1} \left[\left(2(f(1, X_t) - f(0, X_t) - \theta_0) \right) \left(f(1, X_t) - \hat{f}(1, X_t) \right) \right] \tag{31}
\end{aligned}$$

$$+ \mathbb{E}^{t-1} \left[\left(2(f(1, X_t) - f(0, X_t) - \theta_0) \right) \left(f(0, X_t) - \hat{f}(0, X_t) \right) \right] + o_{\mathbb{P}}(1). \tag{32}$$

Applying the Cauchy-Schwarz inequality to each term in equation (32) gives

$$\mathbb{E}^{t-1} \left[\left(2(f(1, X_t) - f(0, X_t) - \theta_0) \right) \left(f(1, X_t) - \hat{f}(1, X_t) \right) \right] \tag{33}$$

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

$$\leq \sqrt{\mathbb{E}^{t-1} \left[\left(2(f(1, X_t) - f(0, X_t) - \theta_0) \right)^2 \right]} \sqrt{\mathbb{E}^{t-1} \left[\left(f(1, X_t) - \hat{f}(1, X_t) \right)^2 \right]} \tag{34}$$

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

$$= \sqrt{\mathbb{E}^{t-1} \left[\left(2(f(1, X_t) - f(0, X_t) - \theta_0) \right)^2 \right]} \|\hat{f} - f\|_2 \tag{35}$$

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

Equation (35) follows from the bound (34) by definition. Since $\|\hat{f} - f\|_2 = o_{\mathbb{P}}(1)$, Equation (35) reduces to $o_{\mathbb{P}}(1)$, and the lemma is proved.

C Proof of Theorem 2

C.1 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.2 Constructing a $(1 - \alpha)$ Betting-CS

Lemma 4. *Assume we observe data following the data generating process of Section 2.1. Assume that $Y_t \in [0, 1]$ $\forall t \in 1, \dots, T$. Suppose that $\pi_t(1 \mid X_t, \Omega_{t-1}) \in [\frac{1}{k_t}, 1 - \frac{1}{k_t}]$ for all $t \in 1, \dots, T$, then*

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

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 (36) is a test martingale,

$$M_T(\theta_0) := \prod_{t=1}^T (1 + \lambda_t(\theta_0)(h_t - \theta_0)). \quad (36)$$

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 $\mathbb{E}^{T-1}(M_T(\theta_0)) = M_{T-1}(\theta_0)$.

$M_T(\theta_0)$ is non-negative if $(1 + \lambda_t(\theta_0)(h_t - \theta_0)) > 0 \forall t \in 1, \dots, T$. 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

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

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,

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

$M_T(\theta_0)$ is therefore a test martingale. By Ville's inequality for non-negative supermartingales,

$$\mathbb{P}\left(\exists T \in \mathbb{N}^+, M_T(\theta_0) \geq \frac{1}{\alpha}\right) \leq \alpha.$$

It follows that the set

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

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

C.3 Hedged-CS

Following suggested values from Waudby-Smith and Ramdas [2023], we set

$$\begin{aligned}\lambda_t &= \sqrt{\frac{2 \log(2/\alpha)}{\hat{\sigma}_{t-1}^2 t \log(1+t)}} \wedge c, \text{ where } c = 0.5, \\ \hat{\theta}_t &= \frac{\frac{1}{2} + \sum_{i=1}^{t-1} h_i}{t}, \\ \hat{\sigma}_t^2 &= \frac{\frac{1}{4} + \sum_{i=1}^t (h_i - \hat{\theta})^2}{t}.\end{aligned}\tag{37}$$

We define

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

where $m = 0.5$ (in general, $m \in [0, 1]$). Letting $\lambda_t(\theta') = \lambda_t$ as defined in Equation (37), 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 4,

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

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

$$\begin{aligned}L_T^{\text{Hedged}} &= \sup_{t \in \{1, \dots, T\}} \inf_T \left\{ \theta' \in [-1, 1] : \mathcal{M}_T(\theta') < \frac{1}{\alpha} \right\}, \\ U_T^{\text{Hedged}} &= \inf_{t \in \{1, \dots, T\}} \sup_T \left\{ \theta' \in [-1, 1] : \mathcal{M}_T(\theta') < \frac{1}{\alpha} \right\}.\end{aligned}$$

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

D Proof of Theorem 3

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

$$M_T = \exp \left\{ \sum_{t=1}^T \lambda_t \left(\xi_t - \frac{\theta_0}{k_t + 1} \right) - \sum_{t=1}^T \left(\xi_t - \hat{\xi}_{t-1} \right)^2 \psi_E(\lambda_t) \right\},\tag{38}$$

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. $\square$

E Proof of Theorem 4

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 their sum diverges.

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

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

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

As is noted in Waudby-Smith et al. [2023], this equation is satisfied if $1/K \leq \mathbb{E}|h_t - \theta_0|^q < K$ 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 $\bar{\sigma}_t^2$ satisfies

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

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

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

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

F Implementation Details

F.1 Performance Metrics

We explicitly define ‘‘Cumulative Error Probability’’ and ‘‘Power’’, the performance metrics shown in Figures 2, 4, and 5. These are both functions of the sample size.

At time T , the Cumulative Error Probability (Error) is defined as

$$\text{Error}(T) := \mathbb{P}(\exists t \in \{50, \dots, T\} \text{ s.t. } \theta_0 \notin [L_t, U_t]).$$

We can empirically estimate this probability over 1000 repetitions of our simulation. For each iteration, we construct confidence sets at each time. We denote the confidence set at time T for iteration i as $[L_{i,T}, U_{i,T}]$. We define our estimate as

$$\widehat{\text{Error}}(T) := \frac{1}{1000} \sum_{i=1}^{1000} \mathbb{1}[\exists t \in \{50, \dots, T\} \text{ s.t. } \theta_0 \notin [L_{i,t}, U_{i,t}]].$$

In Figure 2, we denote the null hypothesis as $\theta_{H_0} = 0$. At time T , we denote the power as

$$\text{Power}(T) := \mathbb{P}(\exists t \in \{50, \dots, T\} \text{ s.t. } \theta_{H_0} \notin [L_t, U_t]).$$

Our empirical estimate of this function is

$$\widehat{\text{Power}}(T) := \frac{1}{1000} \sum_{i=1}^{1000} \mathbb{1}[\exists t \in \{50, \dots, T\} \text{ s.t. } \theta_{H_0} \notin [L_{i,t}, U_{i,t}]].$$

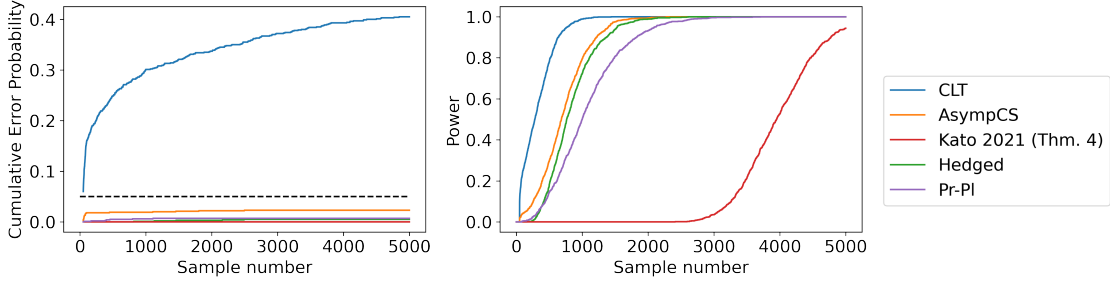


Figure 4: Utilizing a kNN regressor for the protocol used in Figure 2. The policy used for Pr-PI is modified to be truncated within $[0.2, 0.8]$.

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

In Section 2.3, 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,

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

In Assumption 3 of Appendix E, 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 $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (h_t - \bar{h}_T)^2$.

F.3 Bernoulli Outcome Simulation

In Section 4.1 and Figure 2 plots (a) and (b), we simulate $(X_t, A_t, Y_t)_{t=1}^{T=5000}$, where

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

$$\beta^T = [-2, -3, 5],$$

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

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

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

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

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

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 employ sample-splitting and cross-fitting in an effort to avoid over fitting. For the first 50 samples, we let $\pi_t(1 | X_t, \Omega t - 1) = 0.5$ while sufficient samples are collected to give reliable regression estimates. For the regression estimates used in h_t , we use sample means conditioned on A_t until $t = 50$. We ran 1000 iterations using the DGP above, results for the simulation when RF is used are shown in Figure 2. We provide results for the simulation using kNN in Figure 4.

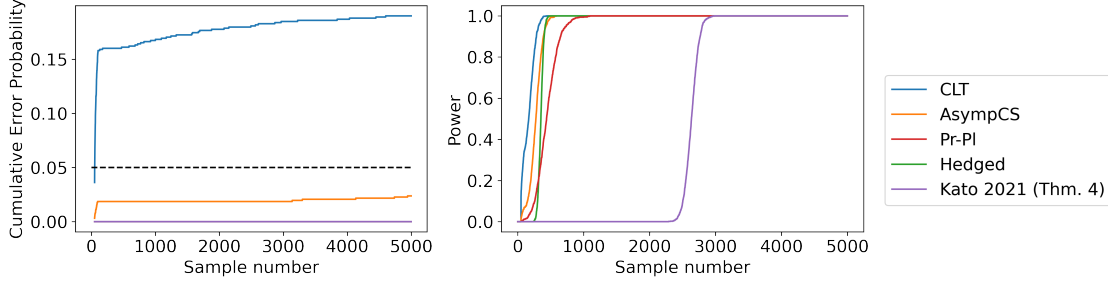
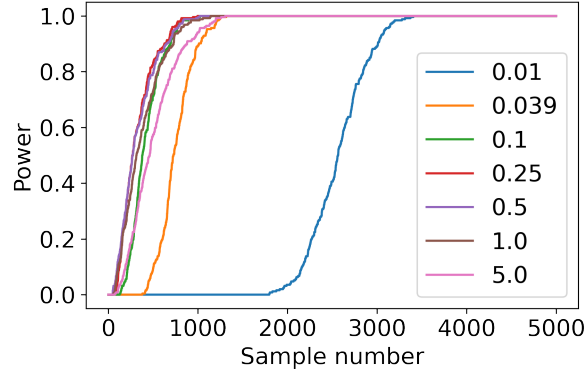


Figure 5: Results for simulation described in Appendix F.4 using a k-Nearest Neighbor regressor.


 Figure 6: 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.

F.4 Bounded Continuous Outcomes Simulation

We now consider simulations with a continuous response, as described in Section 4.2 and shown in Figure 2, plots (c) and (d). Data was simulated as

$$\begin{aligned}
 X_i &\sim \text{Uniform}(0, 1), \text{ for } i \in \{1, 2, 3\}, \\
 \pi_t &= \left(\frac{\sqrt{\hat{v}_{t-1}(1, \mathbf{X}_t)}}{\sqrt{\hat{v}_{t-1}(1, \mathbf{X}_t)} + \sqrt{\hat{v}_{t-1}(0, \mathbf{X}_t)}} \right), \\
 k_t &= \frac{k_{t-1}}{.999}, k_1 = 2 \text{ if method not Kato, else } k_t = 5, \\
 A_t &\sim \text{Bernoulli} \left(\left(\pi_t \vee \frac{1}{k_t} \right) \wedge \left(1 - \frac{1}{k_t} \right) \right), \\
 \beta^T &= [-0.04, -0.01, 0.05], \\
 \epsilon_0 &\sim \text{Uniform}(-0.05, 0.05), \\
 Y_0 &= 0.4 + \mathbf{X}\beta + \epsilon_0, \\
 \epsilon_1 &\sim \text{Uniform}(-4.5\mathbf{X}\beta, 4.5\mathbf{X}\beta), \\
 Y_1 &= 0.4 + \mathbf{X}\beta + \theta_0 + \epsilon_1.
 \end{aligned}$$

In our simulations we set $\theta_0 = 0.1$. Again, we use kNN and random forest (RF) regressors to estimate $\hat{f}$ and $\hat{e}$. Similarly, we employ sample-splitting and cross-fitting. For the first 50 samples, we again let $\pi_t(1 | X_t, \Omega t - 1) = 0.5$. For the regression estimates used in h_t , we use sample means conditioned on A_t until $t = 50$. We ran 1000 iterations using the DGP above, results for the simulation when RF is used are shown in Figure 2. We provide results for the simulation using kNN in Figure 5.

F.5 Effect of Truncation on Inference

Although the modifications made to the DGP in Appendix F.3 are only minor made to the simulation described in Section 4.3 and shown in Figure 3, we provide an explicit DGP for completeness. We assume that we have oracle access to v , allowing us to calculate π^{AIPW} . We set k_t to be a constant. The DGP is

$$\begin{aligned} \mathbf{X}_t &\sim N([\mathbf{0}_3], \mathbf{I}_3), \\ \boldsymbol{\beta}^T &= [-2, -3, 5], \\ \pi_t &= \left(\frac{\sqrt{v(1, \mathbf{X}_t)}}{\sqrt{v(1, \mathbf{X}_t)} + \sqrt{v(0, \mathbf{X}_t)}} \right), \\ k_t &= \frac{1}{\pi_{t, \min}}, \\ A_t &\sim \text{Bernoulli} \left(\left(\pi_t \vee \frac{1}{k_t} \right) \wedge \left(1 - \frac{1}{k_t} \right) \right), \\ p_t &= 0.1 \times \text{logit}(0.5 + \mathbf{X}_t \boldsymbol{\beta}) + 0.4 A_t, \\ Y_t &\sim \text{Bernoulli}(p = p_t). \end{aligned}$$

F.6 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

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

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 T 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 6 shows power curves of the AsympCSs constructed using different levels of ρ .