Research Article

Harrison H. Li* and Art B. Owen

Double machine learning and design in batch adaptive experiments

https://doi.org/10.1515/jci-2023-0068
received October 10, 2023; accepted July 22, 2024

Abstract: We consider an experiment with at least two stages or batches and $O(N)$ subjects per batch. First, we propose a semiparametric treatment effect estimator that efficiently pools information across the batches, and we show that it asymptotically dominates alternatives that aggregate single batch estimates. Then, we consider the design problem of learning propensity scores for assigning treatment in the later batches of the experiment to maximize the asymptotic precision of this estimator. For two common causal estimands, we estimate this precision using observations from previous batches, and then solve a finite-dimensional concave maximization problem to adaptively learn flexible propensity scores that converge to suitably defined optima in each batch at rate $O_p(N^{-1/4})$. By extending the framework of double machine learning, we show this rate suffices for our pooled estimator to attain the targeted precision after each batch, as long as nuisance function estimates converge at rate $o_p(N^{-1/4})$. These relatively weak rate requirements enable the investigator to avoid the common practice of discretizing the covariate space for design and estimation in batch adaptive experiments while maintaining the advantages of pooling. Our numerical study shows that such discretization often leads to substantial asymptotic and finite sample precision losses outweighing any gains from design.

Keywords: adaptive design, propensity score, pooled estimation, partially linear model, convex optimization

MSC 2020: 62K05

1 Introduction

In sequential experimentation, we can use earlier observations to adjust our treatment allocation policy for subsequent observations and thereby improve the estimation of causal effects in the overall study. For instance, for an experiment with one treatment arm and one control arm, Neyman [1] showed that choosing the number of subjects in each arm to be proportional to the outcome standard deviation of that arm minimizes the variance of the treatment effect estimate based on the difference in means. While these standard deviations are unknown, they can be estimated using the initial data. Then, the Neyman allocation can be approximated to improve the sample efficiency of the remainder of the experiment [2–5].

We study a version of this design problem for an experiment divided into a small, fixed number of stages or *batches*. A salient feature of the batched setting, in contrast to the online setting, is that treatment decisions must be made for an entire batch before any of the outcomes from that batch are observed. In particular, we must choose treatment for all subjects in the first batch without any knowledge of the outcome distribution. However, the design (treatment assignment mechanism) can be updated adaptively for later batches based on the observations from earlier batches to improve the precision of the causal estimate computed at the end of the experiment. We assume knowledge of pre-treatment covariates that can further improve precision. Then,

*** Corresponding author: Harrison H. Li,** Department of Statistics, Stanford University, Stanford, California, United States, e-mail: hli90722@stanford.edu, hli90722@gmail.com
Art B. Owen: Department of Statistics, Stanford University, Stanford, California, United States, e-mail: owen@stanford.edu

we formulate our design problem as choosing a *propensity score* for each batch. The propensity score specifies the probability that a subject receives treatment given their covariates (throughout, we consider the setting of a binary treatment). The propensity score is well known to be a key mathematical object to be estimated in the causal analysis of observational data [6]. In a randomized experiment, it is known and under the control of the investigator. Hence, it can be exploited for design.

Our specific design objective is to choose batch-specific propensity scores to minimize an appropriate scalarization of the asymptotic covariance matrix of a proposed estimator that efficiently pools information across all batches of the experiment. After describing our mathematical notation and setup in Section 2, we present and study an oracle version of our pooled estimator in Section 3. That oracle is given knowledge of some nuisance parameters, typically infinite-dimensional mean or variance functions. It pertains to the so-called "non-adaptive batch experiment," where treatment is assigned with possibly varying but nonrandom propensity scores across batches. In certain cases, our oracle pooled estimator asymptotically dominates the best possible alternative that aggregates single batch estimates, regardless of the per-batch propensities. This justifies designing for the pooled estimator instead of an aggregation-based alternative.

For such a design procedure to be useful, however, we must show that the targeted asymptotic precision of the oracle pooled estimator is attainable in a batch experiment where the propensity scores used are adaptive (data dependent) and the nuisance parameters need to be estimated. We address these challenges in Section 4 by extending the framework of double machine learning (DML) formalized by Chernozhukov et al. [7] to what we call a "convergent split batch adaptive experiment" (CSBAE). In a CSBAE, observations in each batch are split into K folds, and treatment is assigned in each fold according to an adaptive propensity score that only depends on observations from previous batches within the same fold. Within a given batch, the K adaptive propensity scores are further required to converge to a common limit at rate $O_p(N^{-1/4})$ in root mean square (RMS). Our DML extension then shows that by plugging in estimated nuisance functions, we can construct a feasible estimator in a CSBAE that is asymptotically equivalent to the oracle pooled estimator computed on the limiting non-adaptive batch experiment. The nuisance function estimates only need to converge at the rate $o_p(N^{-1/4})$.

Section 5 details a finite-dimensional concave maximization procedure (Algorithm 1) that provably constructs a CSBAE for which the limiting propensities in each batch are sequentially optimal within a function class satisfying standard complexity conditions. Hence, we can effectively design for our pooled estimator in a batch adaptive experiment with theoretical guarantees that our final estimator will, attain the targeted optimal asymptotic precision, even with a fairly flexible propensity score learning method and nonparametric machine learning estimates for nuisance functions.

To the best of our knowledge, existing work either designs for a less efficient alternative to our pooled estimator, or discretizes the covariate space at the design and estimation stages to construct a feasible variant of the pooled estimator on a batch adaptive experiment. Our simulations in Section 6 suggest that the latter approach, in particular, can lead to substantial precision losses that swamp any gains from design, even with a moderate number of continuous covariates. Thus, for the practitioner, we provide an end-to-end design and estimation procedure to efficiently handle continuous covariates in a batch experiment.

1.1 Related work

There has been substantial research interest in adaptive experiment designs in recent years. In many applications, treatment assignments are updated in an attempt to maximize the (expected) response values of either those in the experiment, as in adaptive bandit algorithms [8,9], or those in the superpopulation from which the experimental subjects are assumed to arrive [10,11]. Inference on data collected from these algorithms can be challenging since the treatment assignment rules often do not converge [12–15]. By contrast, in our setting where the goal is purely statistical (maximizing asymptotic precision of the treatment estimate), there is often an optimal propensity score to which one hopes to converge. An interesting direction for further study would be to design for a mixture of both statistical and non-statistical objectives. For example, one might expand the work on tie-breaker designs [16–19] to the setting of adaptive experiments.

The seminal work of van der Laan [20] provides a thorough and general framework for efficient estimation and inference in adaptive experiments where the propensity score converges to a fixed target, or more generally satisfies an appropriate asymptotic stability criterion. This framework was later instantiated by Kato et al. [21] for the special case of average treatment effect (ATE) estimation. That work provides refined technical conditions, regret bounds, and guidance on performing sequentially valid inference for their more specific setting. Due to martingale central limit theorems (CLTs), these estimators admit asymptotic inference via Wald-type confidence intervals, regardless of whether the observations are divided into batches or arise one-at-a-time in an online fashion.

Indeed, the results of Kato et al. [21] and van der Laan and Lendle [22] show how to construct online causal estimators that are asymptotically equivalent to one that used the optimal (precision-maximizing) propensity score starting from the first sample. By contrast, in a batch experiment where the first batch is of non-negligible size, as in many clinical trials or studies of online education [14], the investigator does not have information to learn a "good" propensity score with which to assign treatment in the first batch, and thus will necessarily use a suboptimal propensity score. Consequently, the investigator cannot hope to beat the online estimators by employing batches of non-negligible size (although there is some work suggesting that batches of negligible size can win in higher-order asymptotics [4]). Thus, the present work is most relevant to a setting where the batch structure is externally imposed, as in previous studies [2,14].

There are also some unique considerations in batch experiments that do not arise in the literature focused on the online setting. One is non-stationarity in the covariates across batches. The other is the possibility that there are differing constraints on the fraction of subjects that can receive treatment in each batch. Then, the possible propensity scores that can be used – and therefore the optimal ones – may vary across batches. For example, the Reemployment and Eligibility Assessment (REA) program requires those applying for unemployment insurance to meet with a career counselor. Randomization occurs in weekly batches, with the number of applicants varying widely across weeks and uncontrollable by the experimenter. Staffing constraints limit the number of counseling meetings that can occur each week [23,24]. We find that the estimators of van der Laan et al. [20] and the related literature on covariate-adjusted response-adaptive (CARA) designs [25–27] do not efficiently handle these constraints. As discussed earlier, we will prove a result (Theorem 3.1) that a "pooling" estimator asymptotically dominates one that aggregates separate estimates from each batch. This domination is strict when the (adaptive) propensity scores converge to different limits in each batch, as would hold in the REA experiments due to the staffing constraints. As the existing estimators are asymptotically equivalent to aggregations of estimates computed using a single batch, they become inadmissible. Since our pooled estimator "looks forward" to future batches to plug in an estimate of a batch size-weighted average propensity score, the asymptotic normality of our pooled estimator does not follow from martingale arguments, necessitating a different technical approach.

The precise setting and estimators we study are inspired directly by Hahn et al. [2]. Those authors considered a two-batch experiment to estimate the ATE as precisely as possible. As in our study, there may be constraints on the fraction that can be treated in the two batches, although Hahn et al. [2] did not consider any nonstationarities across batches. Using data from the first batch to estimate variance functions, they estimated the asymptotic variance of a pooled version of the semiparametric efficient ATE estimator of Hirano et al. [28] for a coarsely discretized covariate. Then, they learn a propensity score for the second batch that approximately minimizes this variance. The covariate discretization ensures nuisance functions, and optimal propensities can be estimated at parametric $O_p(N^{-1/2})$ rates without parametric assumptions. Consequently, a feasible version of the pooled estimator, attains the targeted asymptotic variance on the batch adaptive experiment. We generalize this pooling construction beyond the setting of ATE estimation and relax these rate requirements to those described in the previous section. This permits more efficient handling of continuous covariates through nonparametric nuisance function estimates and more flexible adaptive propensity scores. We note that continuous covariates are already efficiently handled by much of the existing work on online adaptive experiments [20–22,29]; our goal is to show how do so in the batched setting of Hahn et al. [2].

Existing work in this vein includes Tabord-Meehan [30], who proposes a method to learn a variance-minimizing stratification of the covariate space of fixed size, avoiding the need to discretize the space prior to

observing the data. Cytrynbaum [31] showed that by performing a form of highly stratified treatment assignment called local randomization, consistent variance function estimates from the first batch make it possible to attain the semiparametric lower bound for ATE estimation in the second batch with optimal propensity score without having to estimate the conditional mean functions. Neither of these approaches, however, maintains the efficiency advantage of pooling, which, to our knowledge, has not been rigorously established before. They also do not immediately extend beyond ATE estimation.

2 Setup and notation

Let $T \geqslant 2$ be the number of batches in the experiment. Each subject $i = 1, \ldots, N_t$ in batch $t = 1, \ldots, T$ has observed covariates $X_{ti} \in \mathcal{X} \subseteq \mathbb{R}^d$ and potential outcomes $Y_{ti}(0), Y_{ti}(1) \in \mathbb{R}$. We place these in the vectors $S_{ti} = (X_{ti}^\top, Y_{ti}(0), Y_{ti}(1))^\top$, which are exogenous in our model. Let $Z_{ti} \in \{0, 1\}$ be the binary treatment indicator for this subject, which is controlled by the investigator. Under the usual stable unit value treatment assumption (SUTVA), the observed outcome is

$$Y_{ti} = Z_{ti} Y_{ti}(1) + (1 - Z_{ti}) Y_{ti}(0). \tag{1}$$

Then, the available data for the subject are $W_{ti} = (X_{ti}^\top, Z_{ti}, Y_{ti})^\top \in \mathcal{W}$. We assume that the vectors $S_{ti} \overset{\text{iid}}{\sim} P^S$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N_t$, for some distribution P^S. Appendix A (see Supplementary material) relaxes this assumption to permit certain forms of non-stationarity across batches, such as covariate shifts. All appendices can be found in the supplementary material.[1] It will be convenient to define the mean and variance functions

$$m_0(z, x) = \mathbb{E}[Y(z) | X = x] \quad \text{and} \quad v_0(z, x) = \operatorname{Var}(Y(z) | X = x), \tag{2}$$

for $z \in \{0, 1\}$ and $x \in \mathcal{X}$. These expectations are taken under P^S.

Let $N = N_1 + \ldots + N_T$ be the total sample size across all batches. Then, as in references [2,32] and others, we consider a proportional asymptotic regime

$$\lim_{N \to \infty} \frac{N_t}{N} = \kappa_t \in (0, 1), \quad t = 1, \ldots, T. \tag{3}$$

Recall from the previous section that throughout, we view the number of batches T and the per-batch sample sizes $N_1, \ldots, N_T$ as externally imposed rather than a design choice. We note that other authors have considered different asymptotic regimes, which lend themselves to different estimators and analysis. For instance, some online settings have lots of small batches with the number of batches T growing at the same rate as N [22].

We summarize the preceding requirements for the data-generating process (DGP) in Assumption 2.1. Assumption 2.1 does not impose any restrictions on the treatment assignment process, which will be discussed at length in subsequent sections.

Assumption 2.1. (DGP) For some fixed number of batches $T \geqslant 2$, the vectors

$$S_{ti} = (X_{ti}, Y_{ti}(0), Y_{ti}(1)), \quad 1 \leqslant t \leqslant T, 1 \leqslant i \leqslant N_t,$$

are independent and identically distributed (i.i.d.) from a distribution P^S. Furthermore, the sample sizes N_t satisfy (3), and the vector $W_{ti} = (X_{ti}, Z_{ti}, Y_{ti})$ is observed where the outcomes Y_{ti} satisfy the SUTVA assumption (1).

We end this section by introducing some mathematical notations. For various $q \geqslant 1$ and probability measures P on a space Ω, it will be useful to consider function norms of the form

$$\|f\|_{q,P} = \left(\int |f(w)|^q \, \mathrm{d}P(w) \right)^{1/q},$$

1 Available at https://github.com/hli90722/double_machine_learning_and_batch_adaptive_experiments.

for $f \in L^q(P)$. We will use propensity scores denoted by $e(\cdot)$ with various subscripts. A propensity score $e(\cdot)$ specifies $e(x) = \Pr(Z = 1|X = x)$, the probability of treatment conditional on covariates. We will typically require propensity scores to lie in $\mathcal{F}_\gamma$, the set of all measurable functions on $\mathcal{X}$ taking on values in the interval $[\gamma, 1 - \gamma]$ for some $\gamma \in [0, 1/2)$. We use $||A||$ to denote the square root of the sum of the squared entries of any vector, matrix, or tensor A. For any integer $p \geqslant 1$, $\mathbb{S}_+^p$ will denote the set of symmetric positive semidefinite $p \times p$ real matrices, and $\mathbb{S}_{++}^p$ will be the set of symmetric positive definite $p \times p$ real matrices. Finally, for any real vector v, we write $v^{\otimes 2} = vv^\top$.

2.1 Estimands and score equations

Consider the setting where $T = 1$ (so we can drop the batch subscript t), and the observations $W_1, \ldots, W_N$ are i.i.d. Suppose additionally that (1) holds along with the unconfoundedness assumption

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp Z_i|X_i, \quad i = 1, \ldots, N.$$

Then, many popular causal estimands $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ are identified by a score equation

$$\mathbb{E}[s(W;\, \theta_0, \nu_0, e_0)] = 0.$$

In this score equation, ν_0 is a vector of possibly infinite-dimensional nuisance parameters lying in a nuisance set $\mathcal{N}$, and $e_0 = e_0(\cdot) : \mathcal{X} \to [0, 1]$ is the propensity score. Following Section 3.1 of [7], we will assume that the score $s(\cdot)$ is *linear* in the sense that

$$s(w;\, \theta, \nu, e) = s_a(w;\, \nu, e)\theta + s_b(w;\, \nu, e), \quad \forall w \in \mathcal{W}, \theta \in \Theta, (\nu, e) \in \mathcal{N} \times \mathcal{F}_\gamma, \tag{4}$$

for some $\gamma \in [0, 1/2)$, $s_a(\cdot, \nu, e) : \mathcal{W} \to \mathbb{R}^{p \times p}$, and $s_b(\cdot, \nu, e) : \mathcal{W} \to \mathbb{R}^p$.

When $T > 1$, propensity scores may vary across batches by design or external constraints. For any propensity $e = e(\cdot)$ and integrable function $f : \mathcal{W} \to \mathbb{R}$, we use the subscripted notation $\mathbb{E}_e[f(W)] = \int f(w) \mathrm{d}P_e(w)$, where $P_e = P_e^W$ is the distribution of $W = (X, Z, Y) = (X, Z, ZY(1) + (1 - Z)Y(0))$ induced by $S = (X, Y(0), Y(1)) \sim P^S$ and $Z|X \sim \mathrm{Bern}(e(X))$ under the SUTVA assumption (1). Furthermore, let P^X be the marginal distribution of X under $S \sim P^S$. Then, we will require the following score equations to hold for some $\gamma \in [0, 1/2)$ to identify the our causal estimand θ_0:

$$\mathbb{E}_e[s(W;\, \theta_0, \nu_0, e')] = 0, \quad \forall e, e' \in \mathcal{F}_\gamma. \tag{5}$$

Note that (5) requires the score $s(\cdot)$ to have mean 0 when *any* propensity score $e'(\cdot) \in \mathcal{F}_\gamma$ is plugged in. This plug-in propensity $e'(\cdot)$ may differ from the propensity $e(\cdot) \in \mathcal{F}_\gamma$ used for treatment assignment in the experiment that generated the observations W. Such a robustness property is satisfied by definition so long as the score $s(\cdot)$ is *doubly robust*. It is required to ensure the validity of the pooled estimator that we propose in Section 3. That estimator requires plugging in a mixture propensity score that averages propensities across all batches $t = 1, \ldots, T$. While the identification of θ_0 within each batch is possible by only requiring (5) to hold when $e(\cdot) = e'(\cdot)$, this will not be sufficient to ensure validity of our pooled estimator. We formally restate our requirements on the identification of the estimand θ_0 in Assumption 2.2.

Assumption 2.2. (Estimand identification) The estimand $\theta_0 \in \mathbb{R}^p$ of interest satisfies (5) for some $\gamma \in [0, 1/2)$, some nuisance parameters ν_0 lying in a known convex set $\mathcal{N}$, and some score $s(\cdot)$ satisfying (4).

The first estimand we are motivated by is the ATE, given by $\theta_{0,\mathrm{ATE}} = \mathbb{E}[Y(1) - Y(0)] \in \mathbb{R}$ in our notation. For an investigator interested in modeling how the treatment effect varies with X, they may, instead, wish to estimate the regression parameter $\theta_{0,\mathrm{PL}} \in \mathbb{R}^p$ under a linear treatment effect assumption

$$\mathbb{E}[Y(1) - Y(0)|X] = \psi(X)^\top \theta_{0,\mathrm{PL}}. \tag{6}$$

See Robinson [33] for background on semiparametric estimation of $\theta_{0,\mathrm{PL}}$ under (6), which characterizes the well-known "partially linear model." As shown by Chernozhukov et al. [7], the estimands $\theta_{0,\mathrm{ATE}}$ and $\theta_{0,\mathrm{PL}}$

are identified by score functions s_{AIPW} and s_{EPL}, respectively, which are linear in the sense of (4) and robust in the sense of (5). We write out these results in the following.

Example 2.1. (ATE estimation) Let $\theta_0 = \theta_{0,\mathrm{ATE}}$ be the estimand of interest. Now, consider the augmented inverse propensity weighting (AIPW) score function

$$s_{\mathrm{AIPW}}(W;\, \theta, \nu, e) = m(1, X) - m(0, X) + \frac{Z(Y - m(1, X))}{e(X)} - \frac{(1 - Z)(Y - m(0, X))}{1 - e(X)} - \theta, \tag{7}$$

for nuisance parameter $\nu = (m(0, \cdot), m(1, \cdot))$. For each $\gamma > 0$, it is well known that $\mathbb{E}_e[s_{\mathrm{AIPW}}(W;\, \theta_0, \nu_0, e')] = 0$ for any $e(\cdot), e'(\cdot)$ in $\mathcal{F}_\gamma$ when $\nu_0 = \nu_{0,\mathrm{AIPW}} = (m_0(0, \cdot), m_0(1, \cdot))$ lies in the nuisance set $\mathcal{N} = \mathcal{N}_{\mathrm{AIPW}} = L^1(P^X) \times L^1(P^X)$. Hence, $s_{\mathrm{AIPW}}(\cdot)$ satisfies the score equation (5). This score is also linear because $s_{\mathrm{AIPW}} = s_{\mathrm{AIPW},a}\theta + s_{\mathrm{AIPW},b}$ for

$$\begin{aligned} s_{\mathrm{AIPW},a}(W;\, \nu, e) &= -1 \quad \text{and} \\ s_{\mathrm{AIPW},b}(W;\, \nu, e) &= m(1, X) - m(0, X) + \frac{Z(Y - m(1, X))}{e(X)} - \frac{(1 - Z)(Y - m(0, X))}{1 - e(X)}, \end{aligned}$$

which completes the task of showing that $\theta_{0,\mathrm{ATE}}$ satisfies Assumption 2.2.

Example 2.2. (Partially linear model) Suppose the linear treatment effect assumption (6) holds and $\theta_0 = \theta_{0,\mathrm{PL}}$ is the estimand of interest. Now, consider the weighted least squares score

$$s_{\mathrm{EPL}}(W;\, \theta, \nu, e) = w(X;\, \nu, e)(Z - e(X))(Y - m(0, X) - Z\psi(X)^\top\theta)\psi(X), \tag{8}$$

for nonnegative weights

$$w(X, \nu, e) = (v(0, x)e(x) + v(1, x)(1 - e(x)))^{-1},$$

with nuisance parameter $\nu = (m(0, \cdot), v(0, \cdot), v(1, \cdot))^\top$. Let $\mathcal{F}(\mathcal{X};\, I)$ be the set of all measurable functions $f : \mathcal{X} \to I$. Then, if $\nu_0 = (m_0(0, \cdot), v_0(0, \cdot), v_0(1, \cdot))$ lies in the nuisance set $\mathcal{N} = \mathcal{N}_{\mathrm{EPL}} = L^2(P^X) \times \mathcal{F}(\mathcal{X};\, [c, \infty)) \times \mathcal{F}(\mathcal{X};\, [c, \infty))$ for some $c > 0$, we have $\mathbb{E}_e[s_{\mathrm{EPL}}(W;\, \theta_0, \nu_0, e')] = 0$ for any $e(\cdot), e'(\cdot)$ in $\mathcal{F}_0$. Furthermore, $s_{\mathrm{EPL}}(\cdot)$ is linear, because $s_{\mathrm{EPL}} = s_{\mathrm{EPL},a}\theta + s_{\mathrm{EPL},b}$ for

$$\begin{aligned} s_{\mathrm{EPL},a}(W;\, \nu, e) &= -w(X;\, \nu, e)Z(Z - e(X))\psi(X)\psi(X)^\top \quad \text{and} \\ s_{\mathrm{EPL},b}(W;\, \nu, e) &= w(X;\, \nu, e)(Z - e(X))(Y - m(0, X))\psi(X). \end{aligned}$$

Thus, $\theta_{0,\mathrm{EPL}}$ satisfies Assumption 2.2 with $\gamma = 0$ and the score $s_{\mathrm{EPL}}(\cdot)$. Note that the score equations (5) hold for any nonnegative weight functions $w(\cdot, \nu, e) \in L^1(P^X)$, although the specific choice in $s_{\mathrm{EPL}}(\cdot)$ is semiparametrically efficient [34,35].

Some of our later results pertain to general estimands identified by Assumption 2.2. Others will be specialized to the settings of Examples 2.1 and 2.2.

3 Oracle estimation in non-adaptive batch experiments

Here, we propose and analyze an oracle estimator $\hat{\theta}^*$ of a generic estimand θ_0 satisfying Assumption 2.2 with score $s(\cdot)$. It is an oracle in the sense that it uses the unknown true value of the nuisance parameter ν_0 from the score equations (5). We discuss the feasible estimation of θ_0, including the estimation of ν_0, in Section 4. The estimator $\hat{\theta}^*$ pools observations across all batches $t = 1, \ldots, T$. We then prove a CLT for it in the setting of a *non-adaptive batch experiment* where treatment in each batch $t = 1, \ldots, T$ is assigned according to a fixed (non-random) propensity score $e_t(\cdot)$. We carefully analyze the adaptive setting in Sections 4 and 5.

Definition 3.1. (Non-adaptive batch experiment) A **non-adaptive batch experiment** has DGP satisfying Assumption 2.1 and treatment assignments satisfying

$$Z_{ti} = \mathbf{1}(U_{ti} \leqslant e_t(X_{ti})), \quad t = 1,\dots,T,\ i = 1,\dots,N_t,$$

for some *nonrandom* (i.e., non-adaptive) batch propensity scores $e_1(\cdot),\dots,e_T(\cdot)$ and uniformly distributed random variables $\{U_{ti} | t = 1, \dots,T, i = 1, \dots,N_t\}$ that are i.i.d. and independent of the vectors $\{S_{ti} | t = 1, \dots,T, i = 1, \dots,N_t\}$.

Next, we compare the pooled estimator $\hat{\theta}^*$ to an alternative oracle that performs an optimal linear aggregation of per-batch estimates. It also satisfies a CLT, but we show that our pooling strategy dominates aggregation in terms of efficiency in the setting of Examples 2.1 and 2.2. We remark that some authors (e.g., [30]) refer to this aggregation approach as pooling, but we reserve that term for pooling data, not estimators.

3.1 Pooled oracle estimator

The main idea behind the construction of our oracle estimator $\hat{\theta}^*$ is as follows. After collecting the observations from all batches $t = 1,\dots,T$ in a non-adaptive batch experiment, we ignore the batch structure and pool together the observations across batches. Now, consider a random draw $W = (X, Z, Y)$ from these pooled observations $\{W_{ti} | 1 \leqslant t \leqslant T, 1 \leqslant i \leqslant N_t\}$. In the notation of Section 2, it is straightforward to show that the distribution of W is $P_{e_{0,N}} = \sum_{t=1}^{T}(N_t/N)P_{e_t}$, where $e_{0,N}(\cdot)$ is the mixture propensity score

$$e_{0,N}(x) = \Pr(Z = 1 | X = x) = \sum_{t=1}^{T} \frac{N_t}{N} e_t(x). \tag{9}$$

It will also be helpful to define the limiting mixture propensity score $e_0(\cdot)$ under the proportional asymptotics (3):

$$e_0(x) = \sum_{t=1}^{T} \kappa_t e_t(x). \tag{10}$$

When a particular set of nonrandom batch propensities $e_1(\cdot),\dots,e_T(\cdot)$ is relevant, we omit an additional subscript by letting $\mathbb{E}_{0,N}[f(W)] = \mathbb{E}_{e_{0,N}}[f(W)]$ and $\mathbb{E}_0[f(W)] = \mathbb{E}_{e_0}[f(W)]$. Using this notation, the oracle pooled estimator $\hat{\theta}^*$ is derived by solving the sample analog of the mixture score equations $\mathbb{E}_{0,N}[s(W; \theta_0, \nu_0, e_{0,N})] = 0$ for θ:

$$\hat{\theta}^* = -\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t} s_a(W_{ti}; \nu_0, e_{0,N})\right)^{-1}\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t} s_b(W_{ti}; \nu_0, e_{0,N})\right). \tag{11}$$

This $\hat{\theta}^*$ is only defined when the matrix inverse in (11) exists. That will be the case with probability tending to 1 so long as $\mathbb{E}_0[s_a(W; \nu_0, e_0)]$ is invertible, which we will require in our theoretical results. One such result is a CLT for $\hat{\theta}^*$.

Proposition 3.1. (Oracle CLT) *Let $\theta_0 \in \mathbb{R}^p$ be an estimand satisfying Assumption 2.2 for some $\gamma \in [0, 1/2)$, some score $s(\cdot)$, and some nuisance parameters ν_0. Suppose observations $\{W_{ti} | t = 1, \dots,T, i = 1, \dots,N_t\}$ are collected from a non-adaptive batch experiment with batch propensities $e_1(\cdot),\dots,e_T(\cdot) \in \mathcal{F}_\gamma$, and define $e_{0,N} = e_{0,N}(\cdot)$ and $e_0 = e_0(\cdot)$ as in* (9) *and* (10)*, respectively. Furthermore, assume the following conditions hold*:

(1) *For some sequence $\delta_N \downarrow 0$, we have*

$$(\mathbb{E}_0[||s_a(W; \nu_0, e_{0,N}) - s_a(W; \nu_0, e_0)||^2])^{1/2} \leqslant \delta_N \quad \text{and}$$
$$(\mathbb{E}_0[||s(W; \theta_0, \nu_0, e_{0,N}) - s(W; \theta_0, \nu_0, e_0)||^2])^{1/2} \leqslant \delta_N.$$

(2) $\mathbb{E}_0[s_a(W; \nu_0, e_0)]$ *is invertible and* $\mathbb{E}_0[||s(W; \theta_0, \nu_0, e_0)||^2] < \infty$.

(3) *For some $q > 2$ and $C < \infty$, we have $\mathbb{E}_0[||s(W; \theta_0, \nu_0, e_{0,N})||^q] \leqslant C$ for all sufficiently large N.*

Then, with $\hat{\theta}^*$ *as defined in* (11)*, we have*

$$\sqrt{N}(\hat{\theta}^* - \theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_0),$$

where

$$V_0 = (\mathbb{E}_0[s_a(W;\, \nu_0, e_0)])^{-1}(\mathbb{E}_0[s(W;\, \theta_0, \nu_0, e_0)^{\otimes 2}])(\mathbb{E}_0[s_a(W;\, \nu_0, e_0)])^{-1}.$$

Proof. See Appendix C.1 (Supplementary material). □

For the estimands $\theta_{0,\mathrm{ATE}}$ and $\theta_{0,\mathrm{PL}}$, following (11), we use the scores $s_{\mathrm{AIPW}}(\cdot)$ and $s_{\mathrm{EPL}}(\cdot)$, respectively, to derive the oracle estimates

$$\hat{\theta}^*_{\mathrm{AIPW}} = \frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t} s_{\mathrm{AIPW},b}(W_{ti};\, \nu_{0,\mathrm{AIPW}}, e_{0,N}) \quad (\text{recall } s_{\mathrm{AIPW},a} = -1) \quad \text{and} \tag{12}$$

$$\hat{\theta}^*_{\mathrm{EPL}} = -\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t} s_{\mathrm{EPL},a}(W_{ti};\, \nu_{0,\mathrm{EPL}}, e_{0,N})\right)^{-1}\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t} s_{\mathrm{EPL},b}(W_{ti};\, \nu_{0,\mathrm{EPL}}, e_{0,N})\right), \tag{13}$$

We now specialize the generic oracle CLT of Proposition 3.1 to these two estimators under some regularity conditions.

Assumption 3.1. (Regularity for estimating $\theta_{0,\mathrm{ATE}}$) For some $C < \infty$ and $q > 2$, we have $(\mathbb{E}[|Y(z)|^q])^{1/q} \leqslant C$ and $\mathbb{E}[Y(z)^2|X = x] \leqslant C$ for all $z = 0, 1$ and $x \in \mathcal{X}$.

Assumption 3.2. (Regularity for estimating $\theta_{0,\mathrm{PL}}$) For some $C < \infty$ and $q > 2$, Assumption 3.1 holds. Additionally, $||\psi(x)|| \leqslant C$ for all $x \in \mathcal{X}$, and there exists $c > 0$ such that $\nu_0(z, x) \geqslant c$ for all $z = 0, 1$ and $x \in \mathcal{X}$. Finally, the linear treatment effect assumption (6) holds.

Corollary 3.1. (Oracle CLT for $\hat{\theta}^*_{\mathrm{AIPW}}$) *Suppose Assumption* 3.1 *holds, and let* $\{W_{ti}|t = 1, \ldots, T, i = 1, \ldots, N_t\}$ *be observations from a non-adaptive batch experiment with batch propensities* $e_1(\cdot), \ldots, e_T(\cdot) \in \mathcal{F}_\gamma$ *for some* $\gamma > 0$*. Then,* $\sqrt{N}(\hat{\theta}^*_{\mathrm{AIPW}} - \theta_{0,\mathrm{ATE}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{0,\mathrm{AIPW}})$*, where*

$$V_{0,\mathrm{AIPW}} = \mathbb{E}\left[\frac{\nu_0(1, X)}{e_0(X)} + \frac{\nu_0(0, X)}{1 - e_0(X)} + (m_0(1, X) - m_0(0, X) - \theta_{0,\mathrm{ATE}})^2\right]. \tag{14}$$

Proof. See Appendix C.2 (Supplementary material). □

Corollary 3.2. (Oracle CLT for $\hat{\theta}^*_{\mathrm{EPL}}$) *Suppose Assumption* 3.2 *holds, and let* $\{W_{ti}|t = 1, \ldots, T, i = 1, \ldots, N_t\}$ *be observations from a non-adaptive batch experiment with batch propensities* $e_1(\cdot), \ldots, e_T(\cdot) \in \mathcal{F}_0$*, where* $\mathbb{E}[e_0^2(X)(1 - e_0(X))^2\psi(X)\psi(X)^\top] \in \mathbb{S}^p_{++}$*. Then,* $\sqrt{N}(\hat{\theta}^*_{\mathrm{EPL}} - \theta_{0,\mathrm{PL}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{0,\mathrm{EPL}})$*, where*

$$V_{0,\mathrm{EPL}} = \left(\mathbb{E}\left[\frac{e_0(X)(1 - e_0(X))}{\nu_0(0, X)e_0(X) + \nu_0(1, X)(1 - e_0(X))}\psi(X)\psi(X)^\top\right]\right)^{-1}. \tag{15}$$

Proof. See Appendix C.3 (Supplementary material). □

3.2 Linearly aggregated oracle estimator

When the covariate space $\mathcal{X}$ is finite, the oracle pooled estimator $\hat{\theta}^*_{\mathrm{AIPW}}$ is precisely the estimator proposed by Hahn et al. [2] with the true values of the nuisance functions plugged in for their estimates. The complexity of constructing a *feasible* pooled estimator for an *adaptive* experiment in more general settings has led other authors to, instead, consider single batch estimators that can lose considerable efficiency. For instance,

Cytrynbaum [31] proposed simply discarding the first batch in a two-batch experiment when computing the final estimate, which is clearly inadmissible in our proportional asymptotic regime (3). Section 3.2 of Tabord-Meehan [30] suggests, instead, taking a linear aggregation of estimates computed separately on each batch, as described earlier. We now show that even the best linearly aggregated oracle estimator is asymptotically dominated by our pooled estimators $\hat{\theta}^*_{\mathrm{AIPW}}$ and $\hat{\theta}^*_{\mathrm{EPL}}$. While the study by Tabord-Meehan [30] contains a hypothesis (see their Appendix C.2 [Supplementary material]) that this may be true for ATE estimation, they do not pursue this further as they are unable to construct a feasible pooled estimator attaining the targeted oracle variance when batch propensities are chosen adaptively using their stratification trees. In contrast, our design approach will allow us to construct such an estimator using our extension of DML in Section 4. We note a loosely analogous result in the online bandit setting that tracking a pooled propensity score enables an algorithm reducing asymptotic complexity in best arm identification [36].

Fix a non-adaptive batch experiment with batch propensities $e_1(\cdot), \ldots, e_T(\cdot)$. For each batch $t = 1, \ldots, T$, consider an (oracle) estimator $\hat{\theta}^*_{t,\mathrm{AIPW}}$ for $\theta_{0,\mathrm{ATE}}$ computed by solving the empirical analog of the score equations $\mathbb{E}_{e_t}[s_{\mathrm{AIPW}}(W;\, \theta_{0,\mathrm{ATE}}, \nu_{0,\mathrm{AIPW}}, e_t)] = 0$ that averages only those observations in batch t:

$$\hat{\theta}^*_{t,\mathrm{AIPW}} = -\left(\frac{1}{N_t}\sum_{i=1}^{N_t} s_{\mathrm{AIPW},a}(W_{ti};\, \nu_{0,\mathrm{AIPW}}, e_t)\right)^{-1}\left(\frac{1}{N_t}\sum_{i=1}^{N_t} s_{\mathrm{AIPW},b}(W_{ti};\, \nu_{0,\mathrm{AIPW}}, e_t)\right). \tag{16}$$

By applying Proposition 3.1 with a single batch, for each $t = 1, \ldots, T$, we obtain the CLT

$$\sqrt{N_t}(\hat{\theta}^*_{t,\mathrm{AIPW}} - \theta_{0,\mathrm{ATE}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{t,\mathrm{AIPW}}),$$

where

$$\begin{aligned} V_{t,\mathrm{AIPW}} &= A^{-1}_{t,\mathrm{AIPW}} B_{t,\mathrm{AIPW}} A^{-1}_{t,\mathrm{AIPW}}, \quad \text{for} \\ A_{t,\mathrm{AIPW}} &= \mathbb{E}_{e_t}[s_{\mathrm{AIPW},a}(W;\, \nu_{0,\mathrm{AIPW}}, e_t)], \quad \text{and} \\ B_{t,\mathrm{AIPW}} &= \mathbb{E}_{e_t}[s(W;\, \theta_{0,\mathrm{ATE}}, \nu_{0,\mathrm{AIPW}}, e_t)^{\otimes 2}]. \end{aligned}$$

Now, as stated in the study by Slud et al. [37], the asymptotically unbiased linear combination of $\hat{\theta}^*_{1,\mathrm{AIPW}}, \ldots, \hat{\theta}^*_{T,\mathrm{AIPW}}$ with the smallest asymptotic covariance is the inverse covariance weighted estimator

$$\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{AIPW}} = \left(\sum_{t=1}^{T} \kappa_t V^{-1}_{t,\mathrm{AIPW}}\right)^{-1} \sum_{t=1}^{T} \kappa_t V^{-1}_{t,\mathrm{AIPW}} \hat{\theta}^*_{t,\mathrm{AIPW}}.$$

This optimal linearly aggregated estimator $\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{AIPW}}$ satisfies the CLT

$$\sqrt{N}(\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{AIPW}} - \theta_{0,\mathrm{ATE}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V^{(\mathrm{LA})}_{\mathrm{AIPW}}), \quad V^{(\mathrm{LA})}_{\mathrm{AIPW}} = \left(\sum_{t=1}^{T} \kappa_t V^{-1}_{t,\mathrm{AIPW}}\right)^{-1}.$$

Note that the oracle version of the online AIPW estimator of Kato et al. [21], applied to our batched setting so that the propensity scores are only updated between batches, takes the form

$$\hat{\theta}^{*,\mathrm{onl}}_{\mathrm{AIPW}} = \sum_{t=1}^{T} \frac{N_t}{N} \hat{\theta}^*_{t,\mathrm{AIPW}},$$

which is also an unbiased linear combination of the $\hat{\theta}^*_{t,\mathrm{AIPW}}$, so its asymptotic variance is at least $V^{(\mathrm{LA})}_{\mathrm{AIPW}}$.

We can similarly define the linearly aggregated oracle estimator $\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{EPL}}$ for $\theta_{0,\mathrm{PL}}$ based on combining per-batch estimates $\hat{\theta}^*_{t,\mathrm{EPL}}$ from the score $s_{\mathrm{EPL}}(\cdot)$, which satisfies a CLT $\sqrt{N}(\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{EPL}} - \theta_{0,\mathrm{PL}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V^{(\mathrm{LA})}_{\mathrm{EPL}})$. Here, $V^{(\mathrm{LA})}_{\mathrm{EPL}}$ is given by replacing $s_{\mathrm{AIPW}}(\cdot)$ with $s_{\mathrm{EPL}}(\cdot)$, $\nu_{0,\mathrm{AIPW}}$ with $\nu_{0,\mathrm{EPL}}$, and $\theta_{0,\mathrm{ATE}}$ with $\theta_{0,\mathrm{PL}}$ in the definition of $V^{(\mathrm{LA})}_{\mathrm{AIPW}}$. Our main result is then that regardless of the batch propensities, $\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{AIPW}}$ and $\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{EPL}}$ are asymptotically dominated by our pooled estimators $\hat{\theta}_{0,\mathrm{AIPW}}$ and $\hat{\theta}_{0,\mathrm{EPL}}$, respectively. This motivates our work in Section 5 that designs for these estimators. Note that $\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{AIPW}}$ and $\hat{\theta}^{*,(\mathrm{LA})}_{\mathrm{EPL}}$ are asymptotically equivalent to our oracle pooled estimators

$\hat{\theta}^*_{\text{AIPW}}$ and $\hat{\theta}^*_{\text{EPL}}$ when the batch propensities are all the same: $e_1 = ... = e_T = e_{0,N}$. With varying per-batch propensity scores, however, as will typically occur as we adaptively choose the propensities according to our algorithm in Section 5, the linearly aggregated estimators will typically have strictly larger asymptotic variance.

Theorem 3.1. (Pooling dominates linear aggregation) *Under the conditions of Corollary* 3.1, $V_{0,\text{AIPW}} \leqslant V^{(\text{LA})}_{\text{AIPW}}$. *Under the conditions of Corollary* 3.2, $V_{0,\text{EPL}} \leqslant V^{(\text{LA})}_{\text{EPL}}$.

Proof. See Appendix C.4 (Supplementary material). □

The proof of Theorem 3.1 follows from Jensen's inequality. The result can be viewed as analogous to a well-known result in the Monte Carlo literature: when obtaining importance samples from a mixture distribution, using an estimator that normalizes each sample by the mixture density cannot increase the variance compared to an estimator that normalizes by the density of the actual component of the mixture from which the particular sample was drawn. Our Theorem 3.1 shows that replacing the batch propensity scores e_t with the mixture propensity score $e_{0,N}$ cannot increase the (asymptotic) variance of our oracle estimators.

4 Feasible pooled estimation in batch adaptive experiments

The oracle estimator $\hat{\theta}^*$ of (11) depends on nuisance parameters ν_0 that are unknown in practice. Additionally, recall that our CLT for $\hat{\theta}^*$ (Proposition 3.1) holds for a non-adaptive batch experiment. Yet our goal is to choose propensities adaptively in each batch to improve precision. Therefore, we would like to develop a feasible estimator $\hat{\theta}$ that attains the targeted asymptotic variance for experiments where treatment is assigned *adaptively*, even when the nuisance parameters ν_0 must be estimated.

As mentioned earlier, our construction of such a feasible estimator $\hat{\theta}$ is based on extending the DML framework. The main requirements for $\hat{\theta}$ to have the same asymptotic variance as the corresponding oracle are convergence rate guarantees for both nuisance parameter estimates and the adaptive propensities. Our DML extension ensures that these rate requirements can be made sub-parametric, enabling the use of somewhat flexible machine learning methods.

The typical DML setting formalized by Chernozhukov et al. [7] assumes access to a single sample $W_1, \ldots, W_N$ of i.i.d. observations. An example is a non-adaptive batch experiment with $T = 1$ and propensity $e_0(\cdot)$. Then, a standard DML estimator is based on two ingredients: a Neyman orthogonal score and cross-fitting. Neyman orthogonality of the score $s(\cdot)$ at (ν_0, e_0) means a local insensitivity of the score equations to perturbations in (ν_0, e_0) in any direction:

$$\frac{\partial}{\partial \lambda} \mathbb{E}_{e_0}[s(W_i;\, \theta_0, \nu_0 + \lambda(\nu - \nu_0), e_0 + \lambda(e - e_0))] = 0, \quad \forall (\nu, e) \in \mathcal{N} \times \mathcal{F}_\gamma.$$

It is well known that the scores $s_{\text{AIPW}}(\cdot)$ and $s_{\text{EPL}}(\cdot)$ are Neyman orthogonal [7]. Given a Neyman orthogonal score $s(\cdot)$, DML proceeds by constructing an estimator $\hat{\theta}$ by cross-fitting. In cross-fitting, the indices $1, \ldots, N$ are partitioned into K (roughly) equally sized folds $\mathcal{I}_1, \ldots, \mathcal{I}_K$. Then, $\hat{\theta}$ is computed as the solution to the empirical score equations $N^{-1}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} s(W_i;\hat{\theta}, \hat{\nu}^{(-k)}, \hat{e}^{(-k)}) = 0$, where for each $k = 1, \ldots, K$, $\hat{\nu}^{(-k)}$ and $\hat{e}^{(-k)}(\cdot)$ are some estimates of ν_0 and $e_0(\cdot)$, respectively, such that each pair $(\hat{\nu}^{(-k)}, \hat{e}^{(-k)}(\cdot))$ depends only on the observations $\{W_i | i \notin \mathcal{I}_k\}$ outside fold k. The sample splitting ensures that for each $k = 1, \ldots, K$, the estimates $(\hat{\nu}^{(-k)}, \hat{e}^{(-k)})$ are independent of the observations in fold k, which are plugged into these function estimates. Such independence, along with Neyman orthogonality, enables the feasible estimator $\hat{\theta}$ to be equivalent (up to first-order asymptotics) to the oracle $\hat{\theta}^*$ solving $N^{-1}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} s(W_i;\hat{\theta}, \nu_0, e_0) = 0$ (cf. (11)), so long as the estimates $(\hat{\nu}^{(-k)}, \hat{e}^{(-k)})$ converge at certain sub-parametric rates. Many authors have leveraged such sample splitting arguments in some form to avoid imposing further complexity restrictions (e.g., Donsker conditions) on the

nuisance functions, propensity scores, and their estimates, which are needed in traditional arguments based on stochastic equicontinuity for the asymptotic equivalence of feasible estimators to their oracle counterparts [38,39].

Now, suppose the propensity score $\hat{e}_t = \hat{e}_t(\cdot)$ used to assign treatment in batch $t > 1$ is adaptive in the sense of depending on the observations in the previous batches $1,\dots, t-1$. Let $\hat{\nu}^{(t)}$ be an estimate of the nuisance parameters ν_0 computed from the previous batches and the prior observations in the current batch t. Then, for a feasible variant $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$ of the linearly aggregated AIPW estimator of the ATE in Section 3.2 that replaces $\nu_{0,\mathrm{AIPW}}$ and e_t in (16) with $\hat{\nu}^{(t)}$ and $\hat{e}_t$, respectively, the independence property of the previous paragraph holds conditional on the previous observations even without sample splitting. This was used by previous studies [20,22,29] to avoid Donsker conditions in the context of online adaptive experiments (see Sections 4 and 5 of [29] for a more complete discussion).

However, this idea does not apply to our pooled estimator $\hat{\theta}$, due to the need to plug in an estimate of the *mixture* propensity score $e_{0,N}$ defined in (9) into equation (19). The function $e_{0,N}$ differs, in general, from the limit of any of the batch propensity scores $\hat{e}_t$ actually used for treatment assignment in the aforementioned scheme, so we cannot simply plug the actual adaptive propensity scores $\hat{e}_t$ into (19). To restore the independence between estimates of $e_{0,N}$ (which depend on the observations in all T batches) and the observations they are evaluated at, we require sample splitting at the *design* stage, as illustrated in Figure 1. We also need the adaptive propensity scores to converge. Our notion of a CSBAE in Definition 4.1 formalizes this. The main idea is to split the observations in *every* batch $t = 1,\dots, T$ into K folds. Re-using the notation from the standard DML setting with $T = 1$, we let $\mathcal{I}_k$ denote the set of batch and observation indices (t, i) assigned to fold $k = 1,\dots, K$. Then, the adaptive propensity used to assign treatment to a subject in batches $t = 2,\dots, T$ is allowed to only depend on observations in previous batches from the same fold as this subject. To ensure that the adaptivity does not introduce any additional variability (up to the first-order asymptotics) into the final estimator, a CSBAE requires these adaptive propensities to converge to nonrandom limits $e_1(\cdot),\dots, e_T(\cdot)$ at RMS rate $O_p(N^{-1/4})$. While this convergence requirement may appear restrictive, in Section 5, we show how it can be ensured by design by solving an appropriate finite-dimensional concave maximization procedure. Moreover, the limiting propensity scores from this procedure will be provably optimal, in a sense we make more precise in Section 5.

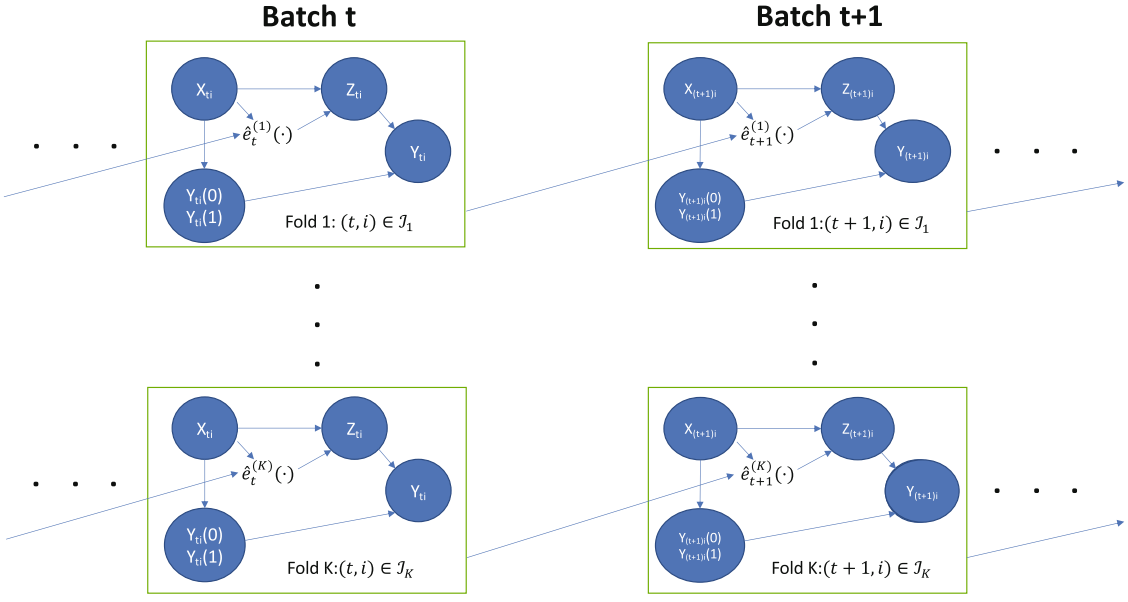

Figure 1: A graphical representation of the dependencies among the observations in two batches of a CSBAE (all notation as in Section 2 and Definition 4.1). Note the lack of vertical arrows, indicating independence between the observations in fold j and fold k for any distinct j, k in $\{1, \dots, K\}$.

Definition 4.1. (CSBAE) A CSBAE is an experiment with DGP satisfying Assumption 2.1 where each observation index (t, i) is assigned to one of K folds $\mathcal{I}_1, \ldots, \mathcal{I}_K$. The fold assignments are such that $n_{t,k} = |\{(t, i) \in \mathcal{I}_k | i = 1, \ldots, N_t\}|$, the number of observations in batch t assigned to fold k, satisfies $|n_{t,k} - N_t/K| \leqslant 1$ for all $t = 1, \ldots, T$, $k = 1, \ldots, K$. Now, let $P_{N,t}^{X,(k)}$ be the empirical distribution of $\{X_{ti} | (t, i) \in \mathcal{I}_k, 1 \leqslant i \leqslant N_t\}$, the covariates in batch t and fold k, and define $\mathcal{S}_t^{X,(k)}$ to be the σ-algebra generated by the covariates $\{X_{ti} | (t, i) \in \mathcal{I}_k, 1 \leqslant i \leqslant N_t\}$ in batch t and fold k along with the observations $\{W_{ui} | (u, i) \in \mathcal{I}_k, u = 1, \ldots, t-1, i = 1, \ldots, N_u\}$ in fold k and any of the previous batches $1, \ldots, t-1$. We further require the following for each batch $t = 1, \ldots, T$ and fold $k = 1, \ldots, K$:

(1) Treatment is assigned according to an adaptive propensity $\hat{e}_t^{(k)}(\cdot)$ that is measurable with respect to $\mathcal{S}_t^{X,(k)}$, i.e., the treatment indicators can be represented as

$$Z_{ti} = \mathbf{1}(U_{ti} \leqslant \hat{e}_t^{(k)}(X_{it})), \quad (t, i) \in \mathcal{I}_k, \tag{17}$$

where $\{U_{ti} : 1 \leqslant t \leqslant T, 1 \leqslant i \leqslant N_t\}$ is a collection of i.i.d. uniformly distributed random variables independent of the vectors $\{S_{ti} | t = 1, \ldots, T, i = 1, \ldots, N_t\}$.

(2) For some nonrandom propensity $e_t(\cdot)$, the adaptive propensity $\hat{e}_t^{(k)}(\cdot)$ satisfies

$$\|\hat{e}_t^{(k)} - e_t\|_{2, P_{N,t}^{X,(k)}} = O_p(N^{-1/4}), \quad t = 1, \ldots, T, k = 1, \ldots, K. \tag{18}$$

Remark 4.1. The left-hand side of equation (18) uses an L^2 norm on the empirical distribution $P_{N,t}^{X,(k)}$ of the covariates of the subjects that will be assigned treatment according to the learned propensity $\hat{e}_t^{(k)}(\cdot)$. These covariates will also be used to learn $\hat{e}_t^{(k)}(\cdot)$ itself in our propensity learning procedure of Section 5. Thus, we can interpret (18) as a rate requirement on the "in-sample" convergence of $\hat{e}_t^{(k)}(\cdot)$. In Section 5, we provide precise conditions under which (18) is guaranteed to hold by design.

Given a CSBAE, our feasible estimator is

$$\hat{\theta} = -\left(\frac{1}{N}\sum_{k=1}^{K}\sum_{(t,i)\in\mathcal{I}_k} s_a(W_{ti}; \hat{\nu}^{(-k)}, \hat{e}^{(-k)})\right)^{-1}\left(\frac{1}{N}\sum_{k=1}^{K}\sum_{(t,i)\in\mathcal{I}_k} s_b(W_{ti}; \hat{\nu}^{(-k)}, \hat{e}^{(-k)})\right). \tag{19}$$

Analogous to the standard (single batch) DML setting, for each $k = 1, \ldots, K$, $\hat{\nu}^{(-k)}$ and $\hat{e}^{(-k)}(\cdot)$ are some estimates of the nuisance parameters ν_0 and the mixture propensity $e_{0,N}(\cdot)$ defined in (9), respectively, that depend only on the observations $\{W_{ti} | (t, i) \notin \mathcal{I}_k\}$ outside fold k. These observations are fully independent of the observations in fold k (across *all* batches $t = 1, \ldots, T$) by the construction of a CSBAE. As in the single batch case, given $o_p(N^{-1/4})$ convergence of the estimators $\hat{\nu}^{(-k)}$ to ν_0, this independence along with (18) enables a DML-style argument that $\hat{\theta}$ is asymptotically equivalent to the oracle $\hat{\theta}^*$ under (18) computed on a counterfactual non-adaptive batch experiment with propensities $e_1(\cdot), \ldots, e_T(\cdot)$. This argument proceeds by coupling the treatment indicators Z_{ti} in the CSBAE with counterfactual treatment indicators $\tilde{Z}_{ti} = \mathbf{1}(U_{ti} \leqslant e_t(X_{ti}))$.

Assumption 4.1. (Score properties and convergence rates for estimating nuisance parameters and the mixture propensity in a CSBAE) Observations $\{W_{ti} : 1 \leqslant t \leqslant T, 1 \leqslant i \leqslant N_t\}$ are collected from a CSBAE with limiting batch propensities $e_1(\cdot), \ldots, e_T(\cdot)$. Additionally, the estimand θ_0 of interest is identified as in Assumption 2.2 by some score $s(\cdot)$, nuisance parameters $\nu_0 \in \mathcal{N}$, and $\gamma \in [0, 1/2)$, such that the propensity collection $\mathcal{F}_\gamma$ contains $e_1(\cdot), \ldots, e_T(\cdot)$. Defining $W_{ti}(z) = (Y_{ti}(z), X_{ti}, z)$ for $z = 0, 1$, the score $s(\cdot)$ has the following properties:

(a) The matrix $\mathbb{E}_0[s_a(W; \nu_0, e_0)]$ is invertible and $\mathbb{E}_0[\|s(W; \theta_0, \nu_0, e_0)\|^2] < \infty$.

(b) The mapping $\lambda \mapsto \mathbb{E}_{0,N}[s(W; \theta_0, \nu_0 + \lambda(\nu - \nu_0), e_{0,N} + \lambda(e - e_{0,N})]$ is twice continuously differentiable on $[0, 1]$ for each $(\nu, e) \in \mathcal{T} = \mathcal{N} \times \mathcal{F}_\gamma$.

(c) All propensities $e(\cdot) \in \mathcal{F}_\gamma$ satisfy

$$\mathbb{E}[s(W_{ti}(1); \theta_0, \nu_0, e) - s(W_{ti}(0); \theta_0, \nu_0, e) | X_{ti}] = 0, \tag{20}$$

for all $t = 1, \ldots, T$ and $i = 1, \ldots, N_t$.

Also, there exist estimators $\hat{\nu}^{(-k)}$ and $\hat{e}^{(-k)}(\cdot)$ of ν_0 and the mixture propensity $e_{0,N}(\cdot)$ defined in (9), respectively, that depend only on the observations outside fold k of the CSBAE. Next, there are nonrandom subsets $\mathcal{T}_N \subseteq \mathcal{T}$ containing $(\nu_0, e_{0,N}(\cdot))$, such that for all $k = 1,\ldots,K$, $\Pr((\hat{\nu}^{(-k)}, \hat{e}^{(-k)}(\cdot)) \in \mathcal{T}_N) \to 1$ as $N \to \infty$. The sets $\mathcal{T}_N$ shrink quickly enough for the following to hold for all $(\nu, e) \in \mathcal{T}_N$, all $\lambda \in (0,1)$ and all $z \in \{0,1\}$ when N is sufficiently large:

$$\left\| \frac{\partial}{\partial \lambda} \mathbb{E}_{0,N}[s(W;\, \theta_0, \nu_0 + \lambda(\nu - \nu_0), e_{0,N} + \lambda(e - e_{0,N}))]|_{\lambda=0} \right\| \leqslant N^{-1/2}\delta_N, \tag{21}$$

$$\left\| \frac{\partial^2}{\partial \lambda^2} \mathbb{E}_{0,N}[s(W;\, \theta_0, \nu_0 + \lambda(\nu - \nu_0), e_{0,N} + \lambda(e - e_{0,N}))] \right\| \leqslant N^{-1/2}\delta_N, \tag{22}$$

$$(\mathbb{E}_0[||s_a(W;\, \nu, e) - s_a(W;\, \nu_0, e_0)||^2])^{1/2} \leqslant \delta_N, \tag{23}$$

$$(\mathbb{E}_0[||s(W;\, \theta_0, \nu, e) - s(W;\, \theta_0, \nu_0, e_0)||^2])^{1/2} \leqslant \delta_N, \tag{24}$$

$$(\mathbb{E}_0[||s_a(W(z);\, \nu, e)||^q])^{1/q} \leqslant C, \tag{25}$$

$$(\mathbb{E}_0[||s(W(z);\, \theta_0, \nu, e)||^q])^{1/q} \leqslant C. \tag{26}$$

Finally, letting $\mathcal{S}^{(-k)}$ be the σ-algebra generated by the observations $\{W_{ti} : (t,i) \notin \mathcal{I}_k\}$ outside fold k across all batches 1 through T, we require

$$S_t^{(-k)}(z) = o_p(N^{-1/4}), \quad z \in \{0,1\}, k = 1,\ldots,K, \tag{27}$$

where

$$S_t^{(-k)}(z) = \sqrt{\frac{1}{n_{t,k}} \sum_{i:(t,i)\in\mathcal{I}_k} ||\mathbb{E}[s(W_{ti}(z);\, \hat{\nu}^{(-k)}, \hat{e}^{(-k)}) - s(W_{ti}(z);\, \nu_0, e_{0,N})|\mathcal{S}^{(-k)}, X_{ti}]||^2}\,.$$

Theorem 4.1. (Feasible CLT for a CSBAE) *Suppose Assumption* 4.1 *holds. Then, for* $\hat{\theta}$ *defined in* (19), *there exists a non-adaptive batch experiment with propensities* $e_1(\cdot),\ldots, e_T(\cdot)$ *for which*

$$\hat{\theta} = \hat{\theta}^* + o_p(N^{-1/2}),$$

where $\hat{\theta}^*$ *is the oracle* (11) *computed on this non-adaptive batch experiment. Then,* $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_0)$, *where* V_0 *is the limiting covariance derived in Proposition* 3.1.

Proof. See Appendix C.5 (Supplementary material). □

In Assumption 4.1, the conditions (a) and (b) along with inequalities (21) through (26) are direct extensions of Assumptions 3.1 and 3.2 in the study by Chernozhukov et al. [7] for ordinary DML ($T = 1$). Equations (20) and (27) are additional requirements that enable the dependence across batches in a CSBAE to be sufficiently weak so that $\hat{\theta}$ computed on the CSBAE is asymptotically equivalent to a version of it computed on the limiting non-adaptive batch experiment.

We can show that Assumption 4.1 is satisfied for estimating $\theta_{0,\mathrm{ATE}}$ and $\theta_{0,\mathrm{PL}}$ with $s_{\mathrm{AIPW}}(\cdot)$ and $s_{\mathrm{EPL}}(\cdot)$, respectively, under simple rate conditions on nuisance parameter and propensity estimation rates that mirror those in Section 5 of Chernozhukov et al. [7] for single-batch DML estimators. Then, we apply Theorem 4.1 to construct feasible pooled estimators $\hat{\theta}_{\mathrm{AIPW}}$ and $\hat{\theta}_{\mathrm{EPL}}$ as special cases of equation (19), which attain the oracle asymptotic variances $V_{0,\mathrm{AIPW}}$ and $V_{0,\mathrm{EPL}}$ defined in (14) and (15), respectively.

Corollary 4.1. (Feasible pooled estimation of $\theta_{0,\mathrm{ATE}}$ in a CSBAE) *Suppose observations are collected from a CSBAE for which the regularity conditions of Assumption* 3.1 *hold for some* $q > 2$ *and* $C < \infty$ *and the limiting batch propensities* $e_1(\cdot),\ldots, e_T(\cdot)$ *are in* $\mathcal{F}_\gamma$ *for some* $\gamma > 0$. *Additionally, for each* $k = 1,\ldots,K$, *suppose we have*

estimates $\hat{m}^{(-k)}(\cdot)$ *and* $\hat{e}^{(-k)}(\cdot)$ *of the mean function* $m_0(\cdot)$ *and mixture propensity* $e_{0,N}(\cdot)$*, respectively, both depending only on the observations outside fold k, such that the following are true*:

(1) $||\hat{m}^{(-k)}(z,\cdot) - m_0(z,\cdot)||_{2,P^X} = o_p(N^{-1/4}), \quad z = 0, 1,$

(2) $||\hat{e}^{(-k)} - e_{0,N}||_{2,P^X} = O_p(N^{-1/4}),$

(3) $||\hat{m}^{(-k)}(z,\cdot) - m_0(z,\cdot)||_{q,P^X} \leqslant C, \quad z = 0, 1,$

(4) $\hat{e}^{(-k)}(\cdot) \in \mathcal{F}_\gamma$ *with probability tending to* 1.

Then, $N^{1/2}(\hat{\theta}_{\mathrm{AIPW}} - \theta_{0,\mathrm{ATE}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{0,\mathrm{AIPW}})$.

Proof. See Appendix C.6 (Supplementary material). □

Corollary 4.2. (Feasible estimation of $\theta_{0,\mathrm{PL}}$ in a CSBAE) *Suppose observations are collected from a CSBAE for which the regularity conditions of Assumption* 3.2 *hold for some* $q > 2$ *and* $0 < c < C < \infty$*, and for which the limiting batch propensities* $e_1(\cdot), \ldots, e_T(\cdot)$ *are in* $\mathcal{F}_0$ *with* $\mathbb{E}[e_0^2(X)(1 - e_0(X))^2\psi(X)\psi(X)^\top] \in \mathbb{S}^p_{++}$*. For each* $k = 1, \ldots, K$*, assume we have estimates* $\hat{m}^{(-k)}(0,\cdot)$*,* $\hat{v}^{(-k)}(\cdot,\cdot)$*, and* $\hat{e}^{(-k)}(\cdot)$ *of the mean function* $m_0(0,\cdot)$*, the variance function* $v_0(\cdot,\cdot)$*, and the mixture propensity* $e_{0,N}(\cdot)$*, respectively, all depending only on the observations outside fold k, such that the following are true*:

(1) $||\hat{m}^{(-k)}(0,\cdot) - m_0(0,\cdot)||_{2,P^X} = o_p(N^{-1/4}),$

(2) $||\hat{v}^{(-k)}(z,\cdot) - v_0(z,\cdot)||_{2,P^X} = o_p(1), \quad z = 0, 1,$

(3) $||\hat{e}^{(-k)} - e_{0,N}||_{2,P^X} = O_p(N^{-1/4}),$

(4) $||\hat{m}^{(-k)}(0,\cdot) - m_0(0,\cdot)||_{q,P^X} \leqslant C,$

(5) $\inf_{x\in\mathcal{X}} \hat{v}^{(-k)}(z,x) \geqslant c, \quad z = 0, 1.$

Then, $N^{1/2}(\hat{\theta}_{\mathrm{EPL}} - \theta_{0,\mathrm{PL}}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{0,\mathrm{EPL}})$.

Proof. See Appendix C.7 (Supplementary material). □

In addition to requiring consistency of both propensity score estimates and mean function estimates, typical DML estimators for θ_{ATE} and θ_{PL} require the product of their RMS convergence rates to be faster than $N^{-1/2}$ in order for the estimator to remain asymptotically equivalent to the oracle. Since the propensity scores are known in experimental settings, typical DML estimators in randomized experiments will plug in these known propensity scores so that the product rate requirement is automatically satisfied, and only $o_p(1)$ convergence is needed for the mean function estimates [7,29,40]. This is weaker than the $o_p(N^{-1/4})$ convergence required in Corollaries 4.1 and 4.2. However, we remind the reader that the estimator $\hat{\theta}$ in our batch setting requires estimating the limiting mixture propensity score $e_{0,N}$, which is not known even after conditioning on any subset of observations. By default (and in our implementation), one would estimate $e_{0,N}$ by a weighted average of the adaptive propensities learned out of fold. Since our goal is to choose propensity scores to maximize precision, we aim to enable an investigator to choose propensities that are as flexible as possible. In particular, in the next section, we show how an investigator can choose adaptive propensity scores that optimize over certain nonparametric function classes. The trade-off for optimizing over these fairly large function classes is that we can only guarantee sub-parametric ($O_p(N^{-1/4})$) rates of convergence in these adaptive propensity scores. This then motivates the $o_p(N^{-1/4})$ requirement in mean function convergence imposed in Corollaries 4.1 and 4.2 to achieve the needed product rate convergence.

While the online and linearly aggregated estimators do not require the more stringent $o_p(N^{-1/4})$ convergence rate in the mean function estimates (as discussed earlier, those estimators can avoid sample splitting altogether and simply plug-in the actual adaptive propensities used in each batch), our numerical studies in Section 6 suggest that this difference does not show up in finite samples. Indeed, there we tend to see the opposite effect; there appear to be finite sample precision advantages to our pooled estimator beyond those predicted by the greater asymptotic precision of $\hat{\theta}$ over $\hat{\theta}^{\mathrm{LA}}$. This is in line with the empirical "paradox"

observed by Kato et al. [29] in the context of off-policy evaluation, where using estimated propensity scores in an AIPW-style estimator leads to improved finite sample performance over using the known propensity scores, despite the stronger mean function estimation rates needed with estimated propensity scores.

5 Batch adaptive learning of the optimal propensity score

We now discuss how to learn adaptive propensity scores $\hat{e}_t^{(k)}(\cdot)$ that satisfy (18) with limiting propensity scores $e_t(\cdot)$ that maximize the asymptotic precision of the final estimators $\hat{\theta}_{\text{AIPW}}$ and $\hat{\theta}_{\text{EPL}}$ constructed in the previous section, as measured by their asymptotic covariance matrices $V_{0,\text{AIPW}}$ and $V_{0,\text{EPL}}$. This generates a CSBAE on which, by Theorem 4.1, feasible estimators $\hat{\theta}_{0,\text{AIPW}}$ and $\hat{\theta}_{0,\text{EPL}}$ achieve the targeted asymptotic variances $V_{0,\text{AIPW}}$ and $V_{0,\text{EPL}}$.

While $V_{0,\text{AIPW}}$ is scalar and so there is no ambiguity in what it means to maximize the asymptotic precision of $\hat{\theta}_{\text{AIPW}}$, when $\theta_{0,\text{PL}}$ is multivariate ($p > 1$), $V_{0,\text{EPL}}$ is a matrix. To handle the multivariate setting, we follow classical literature on experimental design in regression models by scalarizing an *information matrix* $\mathcal{I} \in \mathbb{S}_+^p$ (typically an inverse covariance matrix) using an *information function* $\Psi : \mathbb{S}_+^p \to \mathbb{R} \cup \{-\infty\}$. See textbooks such as refs. [41,42] for more background. We generically write $\mathcal{I} = \mathcal{I}(e, \eta)$ to emphasize that in our setting, $\mathcal{I}$ will be indexed by a propensity score $e = e(\cdot)$ in some function class $\mathcal{F}_*$ along with some unknown nuisance functions $\eta = \eta(\cdot)$ to be estimated. Then, the design objective is to learn an optimal propensity $e^*(\cdot)$, in the sense of maximizing $\Psi(\mathcal{I})$:

$$e^*(\cdot) \in \arg\max_{e \in \mathcal{F}_*} \Psi(\mathcal{I}(e;\, \eta)). \tag{28}$$

The nuisance functions η in the information matrix are distinct from the nuisance functions ν in the score function. Examples of η for information matrices based on $V_{0,\text{AIPW}}$ and $V_{0,\text{EPL}}$ are given in Section 5.2.

5.1 Generic convergence rates with concave maximization

We start by considering a feasible procedure to learn the propensity $e^*(\cdot)$ of (28) when the information matrix $\mathcal{I}$ and the function class $\mathcal{F}_*$ being optimized over generically satisfy Assumption 5.1. An explanation of how to apply this generic procedure to construct an appropriate CSBAE that designs for the estimators $\hat{\theta}_{\text{AIPW}}$ and $\hat{\theta}_{\text{EPL}}$ is deferred to Section 5.2. Here, we instead focus on a precise exposition of our propensity learning procedure and the technical assumptions needed to ensure convergence, without explicitly invoking any notation or setup from previous sections. We begin with some generic structure on the information matrix $\mathcal{I}$ and the function class $\mathcal{F}_*$. Roughly, we require the information matrix to be strongly concave in $e(\cdot)$ and the function class $\mathcal{F}_*$ to be not too complex. We also allow for interval constraints on the per-batch expected proportion of subjects treated, as previewed in Section 1.1.

Assumption 5.1. (Generic optimization setup) The information matrix $\mathcal{I} = \mathcal{I}(e, \eta)$ in (28) takes the form

$$\mathcal{I}(e;\, \eta) = \mathbb{E}_{X \sim P}[f(e(X), \eta(X))], \tag{29}$$

where $\eta : \mathcal{X} \to \mathcal{W}$ is a vector of possibly unknown functions taking values in some compact set $\mathcal{W} \subseteq \mathbb{R}^r$, and P is a **covariate distribution** from which i.i.d. observations $X_1, \ldots, X_n \in \mathcal{X}$ are drawn. Additionally, the function $f : [0,1] \times \mathcal{W} \to \mathbb{S}_+^p$ appearing in (29) satisfies the following properties:

- There exists an extension of $f(\cdot,\cdot)$ with continuous second partial derivatives to an open neighborhood of $\mathbb{R}^{r+1}$ containing $(\delta, 1-\delta) \times \mathcal{W}$ for some $\delta < 0$.
- For some $c > 0$ we have

$$-f''(e, w) \in \mathbb{S}_+^p, \quad \text{and} \quad \operatorname{tr}(f''(e, w)) \leqslant -c, \quad \forall (e, w) \in [0,1] \times \mathcal{W}, \tag{30}$$

  where $f''(e, w)$ denotes the second partial derivative of $f(\cdot,\cdot)$ with respect to the first argument, evaluated at (e, w).

Next, the collection of propensity scores $\mathcal{F}_*$ to be optimized over takes the form $\mathcal{F}_* = \mathcal{F}_*(m_L, m_H; P) = \{e \in \mathcal{E} | m_L \leqslant \mathbb{E}_P[e(X)] \leqslant m_H\}$, for some **base propensity class** $\mathcal{E} \subseteq \mathcal{F}_0$ and some known **budget constraints** (m_L, m_H) with $0 \leqslant m_L \leqslant m_H \leqslant 1$. The base propensity class $\mathcal{E}$ is convex and closed in $L^2(P)$ and additionally satisfies the following properties for all $n \geqslant 1$:

(1) There exists $C < \infty$ for which

$$\int_0^1 \sqrt{\log \mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n))}\, \mathrm{d}\varepsilon \leqslant C \quad \forall n \quad \text{w.p.1}, \tag{31}$$

where $\log \mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n))$ is the metric entropy of the base propensity class $\mathcal{E}$ in $L^2(P_n)$, as defined in Appendix B.1 (see Supplementary material), and P_n is the empirical distribution on the observations $X_1, \ldots, X_n$.

(2) Given any $x_1, \ldots, x_n \in \mathcal{X}$, there exists a convex set $E_n \subseteq [0,1]^n$, possibly depending on $x_1, \ldots, x_n$, such that
– For every $e \in \mathcal{E}$, we must have $(e(x_1), e(x_2), \ldots, e(x_n)) \in E_n$, and
– For every $(e_1, e_2, \ldots, e_n) \in E_n$, there exists $e \in \mathcal{E}$ with $e(x_i) = e_i$ for all $i = 1, \ldots, n$.

(3) There exists $e_*(\cdot) \in \mathcal{F}_*$ such that $\mathbb{E}_P[f(e_*(X), \eta(X))] \geqslant cI$ for some $c > 0$.

(4) There exist $e_L(\cdot), e_H(\cdot) \in \mathcal{E}$ such that $\mathbb{E}_P[e_L(X)] > m_L$ and $\mathbb{E}_P[e_H(X)] < m_H$.

Remark 5.1. If (30) only holds for (e, w) in $[\gamma, 1-\gamma] \times \mathcal{W}$ for some $\gamma \in (0, 1/2)$ (instead of on all of $[0,1] \times \mathcal{W}$) and the base propensity class $\mathcal{E}$ is chosen to lie within $\mathcal{F}_\gamma$, then one can rewrite (29) as

$$\mathcal{I}(e;\, \eta) = \mathbb{E}_{X \sim P}[\tilde{f}(e(X), \eta(X))], \quad \tilde{f}(e, w) = f(L_\gamma(e), w),$$

where $L_\gamma(e) = \gamma + (1 - 2\gamma)e$ is an invertible linear mapping. Then, the remainder of Assumption 5.1 holds as stated with f replaced by $\tilde{f}$ and the base propensity class $\mathcal{E}$ replaced by $\tilde{\mathcal{E}} = \{L_\gamma^{-1}(e(\cdot)) | e(\cdot) \in \mathcal{E}\}$, and so all results below depending on Assumption 5.1 hold by considering optimization over $\tilde{\mathcal{E}}$ instead of $\mathcal{E}$.

Similar to the idea of empirical risk minimization in supervised learning (ERM, e.g., [43,44]), when Assumption 5.1 is satisfied, our learning procedure replaces the unknown population expectation appearing in the objective (28) by a sample average over observations $X_1, \ldots, X_n \sim iidP$, where the generic sample size n diverges. In our designed CSBAE, these observations will be the covariates of those subjects in a given batch $t = 1, \ldots, T$ of a CSBAE within a given fold $k = 1, \ldots, K$, so n will be identified with the quantity $n_{t,k}$ of Definition 4.1. The main computational procedure is then a finite-dimensional optimization over the values of the propensity score at the n points $X_1, \ldots, X_n$:

$$(\hat{e}_1, \ldots, \hat{e}_n) \in \underset{(e_1, \ldots, e_n) \in F_n}{\arg\max}\, \Psi\left\{ n^{-1} \sum_{i=1}^n f(e_i, \hat{\eta}(X_i)) \right\}, \tag{32}$$

where $\hat{\eta}(\cdot)$ is an estimate of the nuisance function $\eta(\cdot)$ in (29), and the optimization set $F_n = F_n(m_L, m_H) \subseteq \mathbb{R}^n$ is defined by

$$F_n(m_L, m_H) = \left\{ (e_1, \ldots, e_n) \in E_n | m_L \leqslant \frac{1}{n} \sum_{i=1}^n e_i \leqslant m_H \right\} \subseteq \mathbb{R}^n, \tag{33}$$

where (m_L, m_H) are some budget constraints as in Assumption 5.1 and E_n is as in the numbered condition (2) of that assumption.

We convert the vector $(\hat{e}_1, \ldots, \hat{e}_n)$ in (32) to a propensity score $\hat{e}(\cdot)$ by taking $\hat{e}(\cdot)$ to be any member of the base propensity class $\mathcal{E}$ of Assumption 5.1 with $\hat{e}(X_i) = \hat{e}_i$ for each $i = 1, \ldots, n$. The existence of such a propensity $\hat{e}(\cdot)$ is guaranteed by the numbered condition (2) in Assumption 5.1. Then, as in the ERM literature, we use empirical process arguments to guarantee the learned propensity $\hat{e}(\cdot)$ converges to $e^*(\cdot)$ at rate $O_p(n^{-1/4})$ under appropriate restrictions on the complexity of the base propensity class $\mathcal{E}$. Our main such restriction is the finite entropy integral requirement in (31). Examples of function classes satisfying this condition can be found in the literature on empirical processes. A sufficient condition is that $\log \mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n)) \leqslant K\varepsilon^{-2+\delta}$ for some

$\delta > 0$, some $K < \infty$, and all $\varepsilon > 0$. Examples 5.1 through 5.4 show that this requirement is loose enough to admit some fairly rich function classes.

Example 5.1. (Monotone) Suppose $d = 1$, and let $\mathcal{E}$ be the set of nondecreasing functions in $\mathcal{F}_0$. By Lemma 9.11 of [45], we know that $\log \mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n)) \leqslant K/\varepsilon$ for each $\varepsilon > 0$ and some positive universal constant $K < \infty$.

Example 5.2. (Lipschitz) Again, suppose $d = 1$, and let $\mathcal{E}$ be the set of L-Lipschitz functions in $\mathcal{F}_0$ for some fixed $L > 0$. If $\mathcal{X}$ is a bounded closed interval, then the discussion preceding Example 5.11 of Rothe [46] shows that $\log \mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n)) \leqslant K/\varepsilon$ for each $\varepsilon > 0$ and some positive universal constant $K < \infty$ (which may depend on L).

Example 5.3. (VC-subgraph class) Let $\mathcal{E}$ be any subset of $\mathcal{F}_0$ that is closed and convex in $L^2(P)$, and whose subgraphs are a Vapnik–Chervonenkis (VC) class, meaning they have a finite VC dimension V. A special case is a fully parametric class like $\{x \mapsto \theta^\top \xi(x) | \theta^\top \mathbf{1}_p \leqslant 1, \theta \geqslant 0\}$, where $\xi(x) \in [0,1]^p$ is a known set of basis functions and $\mathbf{1}_p \coloneqq (1, \ldots, 1) \in \mathbb{R}^p$. Then, by Theorem 2.6.7 of [47], $\mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n)) \leqslant K(1/\varepsilon)^{2V-2}$ for some universal $K > 0$ depending on the VC dimension V of the subgraphs. Note that K may depend on p.

Example 5.4. (Symmetric convex hull of VC-subgraph class) Let $\mathcal{E}_0$ be a VC-subgraph class of functions. The symmetric convex hull of $\mathcal{E}_0$ is defined as

$$\operatorname{sconv}(\mathcal{E}_0) = \left\{ \sum_{i=1}^m \omega_i e_i \;\middle|\; e_i \in \mathcal{E}_0, \sum_{i=1}^m |\omega_i| \leqslant 1 \right\}.$$

Now, suppose $\mathcal{E}$ is contained within $\overline{\operatorname{sconv}(\mathcal{E}_0)}$, the pointwise closure of $\operatorname{sconv}(\mathcal{E}_0)$. Let $V < \infty$ be the VC dimension of the collection of subgraphs of functions in $\mathcal{E}_0$. Then, by Theorem 2.6.9 of [47], we have $\log \mathcal{N}(\varepsilon, \mathcal{E}, L^2(P_n)) \leqslant K(1/\varepsilon)^{2(1-1/V)}$ for all $\varepsilon > 0$. For example, with $\operatorname{expit}(x) = \exp(x)/(1 + \exp(x))$, we can take

$$\mathcal{E} = \left\{ \sum_{i=1}^m \omega_i \operatorname{expit}(\theta_i^\top \varphi(x)) \;\middle|\; \omega_i \geqslant 0, \sum_{i=1}^m \omega_i \leqslant 1 \right\}, \tag{34}$$

where $\varphi(\cdot)$ is any vector of p real-valued basis functions, m can be made arbitrarily large, and $\theta_1, \ldots, \theta_m \in \mathbb{R}^p$ are arbitrary. This choice of $\mathcal{E}$ is evidently a closed and convex subset of $\operatorname{sconv}(\mathcal{E}_0)$ with $\mathcal{E}_0 = \{\operatorname{expit}(\theta^\top \varphi(x)) | \theta \in \mathbb{R}^p\}$. Note the collection $\mathcal{E}_0$ is, indeed, a VC-subgraph class by Lemmas 2.6.15 and 2.6.17 of [47], as each function in $\mathcal{E}_0$ is the composition of the monotone function $\operatorname{expit}(\cdot)$ with the p-dimensional vector space of functions $\{x \mapsto \theta^\top \varphi(x) | \theta \in \mathbb{R}^p\}$.

Next, we construct sets E_n that satisfy condition (2) of Assumption 5.1 for the base propensity classes in Examples 5.1 to 5.4. For the set of monotone functions in one dimension (Example 5.1), we can take

$$E_n = \{(e_1, \ldots, e_n) | 0 \leqslant e_{\pi(1)} \leqslant e_{\pi(2)} \leqslant \ldots \leqslant e_{\pi(n)} \leqslant 1\},$$

where $\pi(\cdot)$ is the inverse of the function that maps each $i \in \{1, \ldots, n\}$ to the rank of x_i among $x_1, \ldots, x_n$ (with any ties broken in some deterministic way). For the set of L-Lipschitz functions (Example 5.2), we can take

$$E_n = \{(e_1, \ldots, e_n) | |e_{\pi(i)} - e_{\pi(i-1)}| \leqslant L(x_{\pi(i)} - x_{\pi(i-1)}),\ i = 2, \ldots, n\}.$$

For the parametric class in Example 5.3, we can take

$$E_n = \{(\theta^\top \xi(x_1), \ldots, \theta^\top \xi(x_n)) | \theta^\top \mathbf{1}_p \leqslant 1, \theta \geqslant 0\}.$$

Finally, for the class (34) in Example 5.4, we take

$$E_n = \left\{ \sum_{i=1}^m \omega_i (\operatorname{expit}(\theta_i^\top \varphi(x_1)), \ldots, \operatorname{expit}(\theta_i^\top \varphi(x_n))) \;\middle|\; \omega_i \geqslant 0, \sum_{i=1}^m \omega_i \leqslant 1 \right\}.$$

The convergence rates of $\hat{e}(\cdot)$ to $e^*(\cdot)$ will be proven using strong concavity of the design objective (28) on the space $\mathcal{F}^*$. This is ensured by Assumption 5.1 along with the following conditions on the information function $\Psi(\cdot)$:

Assumption 5.2. (Information function regularity) The information function $\Psi : \mathbb{S}_+^p \to \mathbb{R} \cup \{-\infty\}$ is concave, continuous, and nondecreasing with respect to the semidefinite ordering $\succcurlyeq$ on $\mathbb{R}^{p\times p}$ and satisfies the following conditions:

(a) For every $k > 0$, $\inf_{B \succcurlyeq kI} \Psi(B) > \sup_{A \in \mathbb{S}_+^p \setminus \mathbb{S}_{++}^p} \Psi(A) =: \Psi_0$.

(b) $\Psi(\cdot)$ is twice continuously differentiable on $\mathbb{S}_{++}^p$, such that for all $0 < k < K$, there exists $C > 0$ such that $\|\nabla\Psi(A) - \nabla\Psi(B)\| \leqslant C\|A - B\|$ whenever $KI \succcurlyeq A \succcurlyeq kI$ and $KI \succcurlyeq B \succcurlyeq kI$.

(c) For every $0 < k < K$, $kI \preccurlyeq A \preccurlyeq KI$ implies $\tilde{k}I \preccurlyeq \nabla\Psi(A) \preccurlyeq \tilde{K}I$ for some $0 < \tilde{k} < \tilde{K}$.

(d) For every $K < \infty$ and $\tilde{\Psi}_0 > \Psi_0$, there exists $k > 0$ such that for all $0 \preccurlyeq A \preccurlyeq KI$ with $\Psi(A) \geqslant \tilde{\Psi}_0$, we have $A \succcurlyeq kI$.

We can show that Assumption 5.2 is satisfied by two common information functions: the "A-optimality" function $\Psi_a(\cdot) = -\mathrm{tr}((\cdot)^{-1})$ with $\Psi_a(M) := -\infty$ whenever M is singular, and the "D-optimality" function $\Psi_d(\cdot) = \log(\det(\cdot))$. The A-optimality criterion corresponds to minimizing the average (asymptotic) variance of the components of the estimand, while D-optimality corresponds to minimizing the volume of the ellipsoid spanned by the columns of the (asymptotic) covariance matrix.

Lemma 5.1. *The information functions $\Psi_d(\cdot)$ and $\Psi_a(\cdot)$ satisfy Assumption* 5.2.

Proof. See Appendix C.8 (Supplementary material). □

There are some common information functions that do not satisfy Assumption 5.2. For example, the "E-optimality" function $\Psi_e(\cdot) = \lambda_{\min}(\cdot)$, where $\lambda_{\min}(M)$ refers to the smallest eigenvalue of $M \in \mathbb{R}^{p\times p}$, is not differentiable. Similarly, the function $\Psi_c(\cdot) = -c^\top(\cdot)^{-1}c$ (for some fixed $c \in \mathbb{R}^p$), corresponding to "c-optimality," does not satisfy condition (c). We leave open the question of whether the $O_p(n^{-1/4})$ convergence rate of $\hat{e}(\cdot)$ to $e^*(\cdot)$ in Lemma 5.2 can be extended to these (and other) information functions using different techniques, and now, prove this rate under Assumptions 5.1 and 5.2.

Lemma 5.2. (Convergence of generic concave maximization routine) *Suppose Assumption* 5.1 *holds for some information matrix $\mathcal{I}$ of the form* (29)*, covariate distribution P, and budget constraints (m_L, m_H). Furthermore, assume that for some sequence $\alpha_n \downarrow 0$, we have an estimate $\hat{\eta}(\cdot)$ of $\eta(\cdot)$ satisfying*

$$\hat{\eta}(x) \in \mathcal{W}, \quad \forall x \in \mathcal{X}, \quad \text{and} \quad \|\hat{\eta} - \eta\|_{2,P_n} = O_p(\alpha_n), \tag{35}$$

where $\eta(\cdot)$ is defined by $\mathcal{I}$ by (29)*. Then, for any information function $\Psi(\cdot)$ satisfying Assumption* 5.2*, the following statements are true*:

(1) *There exists an optimal propensity function $e^*(\cdot)$ satisfying* (28)*, which is unique P-almost everywhere.*

(2) *There exist optimal finite sample treatment probabilities $(\hat{e}_1, \ldots, \hat{e}_n) \in [0,1]^n$ satisfying* (32)*, where $F_n = F_n(m_L, m_H)$ is defined as in* (33)*.*

(3) *For any such optimal probabilities $(\hat{e}_1, \ldots, \hat{e}_n)$, there exists a propensity score $\hat{e}(\cdot) \in \mathcal{E}$ for which $\hat{e}(X_i) = \hat{e}_i$ for each $i = 1, \ldots, n$. Any such function $\hat{e}(\cdot)$ satisfies both $\|\hat{e} - e^*\|_{2,P} = O_p(n^{-1/4} + \alpha_n)$ and $\|\hat{e} - e^*\|_{2,P_n} = O_p(n^{-1/4} + \alpha_n)$.*

Proof. See Appendix C.9 (Supplementary material). □

5.2 Convergence of batch adaptive designs

We now leverage Lemma 5.2 to develop a procedure (Algorithm 1) that can learn adaptive propensities $\hat{e}_t^{(k)}(\cdot)$ with the convergence guarantees (18) so that when used for treatment assignment, they lead to a CSBAE

with limiting propensities that optimize objectives of the form (28) with information matrices based on $V_{0,\mathrm{AIPW}}^{-1}$ and $V_{0,\mathrm{EPL}}^{-1}$. This shows we can effectively design for the estimators $\hat{\theta}_{\mathrm{AIPW}}$ and $\hat{\theta}_{\mathrm{EPL}}$.

Algorithm 1 CSBAE for estimating $\theta_{0,\mathrm{ATE}}$ resp. $\theta_{0,\mathrm{PL}}$

Require Base propensity class $\mathcal{E}$ satisfying conditions in Assumption 5.1, initial propensity $e_1(\cdot) \in \mathcal{F}_{\varepsilon_1}$ for some $\varepsilon_1 > 0$, information function $\Psi(\cdot)$ satisfying Assumption 5.2, number of folds $K \geqslant 2$

for batch $t = 1, \ldots, T$ **do**
  Observe subject covariates $X_{t1}, \ldots, X_{tN_t}$
  Split subject indices $i = 1, \ldots, N_t$ into K folds $\mathcal{I}_1, \ldots, \mathcal{I}_K$, of size as equal as possible
  for fold $k = 1, \ldots, K$ **do**
    Label the covariates in batch t, fold k by $X_{t1}^{(k)}, \ldots, X_{tn_{t,k}}^{(k)}$
    if $t = 1$ **then**
      Set $(\hat{e}_{t1}^{(k)}, \ldots, \hat{e}_{tn_{t,k}}^{(k)}) = (e_1(X_{t1}^{(k)}), \ldots, e_1(X_{tn_{t,k}}^{(k)}))$
    else
      Compute $(\hat{e}_{t1}^{(k)}, \ldots, \hat{e}_{tn_{t,k}}^{(k)})$ using right-hand side of (41) resp. (44) given batch t budget constraints $m_{L,t} \leqslant m_{H,t}$
    end if
    Draw $(U_{t1}^{(k)}, \ldots, U_{tn_{t,k}}^{(k)}) \overset{\mathrm{iid}}{\sim} \mathrm{Unif}(0,1)$
    Assign treatments according to $Z_{ti}^{(k)} = \mathbf{1}(U_{ti}^{(k)} \leqslant \hat{e}_{ti}^{(k)})$, $i = 1, \ldots, n_{t,k}$
    Choose and store any $\hat{e}_t^{(k)}(\cdot) \in \mathcal{E}$ with $\hat{e}_t^{(k)}(X_{ti}^{(k)}) = \hat{e}_{ti}^{(k)}$, $i = 1, \ldots, n_{t,k}$
  end for
  Observe outcomes $\{W_{ti} | 1 \leqslant i \leqslant N_t\}$ in batch t
  Compute and store $N^{1/4}$-consistent estimates $(\hat{v}_{1:t}^{(1)}(\cdot,\cdot), \ldots, \hat{v}_{1:t}^{(K)}(\cdot,\cdot))$ of $v_0(\cdot,\cdot)$ where for each fold $k = 1, \ldots, K$, $\hat{v}_{1:t}^{(k)}(\cdot,\cdot)$ depends only on the observations $\{W_{ui} : (u,i) \in \mathcal{I}_k, 1 \leqslant u \leqslant t\}$ in fold k and batches $1, \ldots, t$
end for
Compute final estimator $\hat{\theta}$ via (19), using $(s_{a,\mathrm{AIPW}}, s_{b,\mathrm{AIPW}})$, resp. $(s_{a,\mathrm{PL}}, s_{b,\mathrm{PL}})$.

For simplicity, we assume treatment in the first batch is assigned according to a non-random propensity $e_1(\cdot) \in \mathcal{F}_{\varepsilon_1}$ for some $\varepsilon_1 > 0$. We let $\hat{e}_1^{(k)}(\cdot) = e_1(\cdot)$ for $k = 1, \ldots, K$. Then, for later batches $t = 2, \ldots, T$, the target propensities are taken to be one of the following, for an information function $\Psi(\cdot)$ satisfying Assumption 5.2:

$$e_{t,\mathrm{AIPW}}^*(\cdot) \in \underset{e_t(\cdot) \in \mathcal{F}_{*,t}}{\arg\max}\, \Psi(V_{0:t,\mathrm{AIPW}}^{-1}) \quad \text{or} \quad e_{t,\mathrm{EPL}}^*(\cdot) \in \underset{e_t(\cdot) \in \mathcal{F}_{*,t}}{\arg\max}\, \Psi(V_{0:t,\mathrm{EPL}}^{-1}), \tag{36}$$

where $V_{0:t,\mathrm{AIPW}}$ and $V_{0:t,\mathrm{EPL}}$ are the asymptotic variances of the oracle pooled estimators $\hat{\theta}_{\mathrm{AIPW}}$ and $\hat{\theta}_{\mathrm{EPL}}$ of (12) and (13), respectively, *when computed using observations in a non-adaptive batch experiment with only t batches and propensities* $e_1(\cdot), \ldots, e_t(\cdot)$. By (14), we can compute

$$V_{0:t,\mathrm{AIPW}} = V_{0:t,\mathrm{AIPW}}(e_t;\, \eta_{0,\mathrm{AIPW}}) = \mathbb{E}_{P^X}\left[\frac{v_0(1,X)}{e_{0:t}(X)} + \frac{v_0(0,X)}{1 - e_{0:t}(X)} + (\tau_0(X) - \theta_0)^2\right], \tag{37}$$

where $\eta_{0,\mathrm{AIPW}}(x)$ includes the components $(v_0(0,x), v_0(1,x), \tau_0(x), \theta_0)$. Similarly, by (15), we have

$$V_{0:t,\mathrm{EPL}}(e_t;\, \eta_{0,\mathrm{EPL}}) = \left(\mathbb{E}_{P^X}\left[\frac{e_{0:t}(X)(1 - e_{0:t}(X))}{v_0(0,X)e_{0:t}(X) + v_0(1,X)(1 - e_{0:t}(X))}\psi(X)\psi(X)^\top\right]\right)^{-1}, \tag{38}$$

where $\eta_{0,\mathrm{EPL}}(x)$ includes the components $(v_0(0,x), v_0(1,x))$. In both of the preceding equations, the dependence on the batch t propensity score $e_t(\cdot)$ is through the mixture $e_{0:t}(\cdot)$ given by

$$e_{0:t}(x) \coloneqq \left(\sum_{u=1}^{t} \kappa_u\right)^{-1}\left(\sum_{u=1}^{t-1} \kappa_u e_u(x) + \kappa_t e_t(x)\right), \quad x \in \mathcal{X}. \tag{39}$$

Finally, the optimization set $\mathcal{F}_{*,t}$ in (36) is

$$\mathcal{F}_{*,t} = \mathcal{F}_*(m_{L,t}, m_{H,t}; P^X) = \{e(\cdot) \in \mathcal{E} | m_{L,t} \leqslant \mathbb{E}_{P^X}[e(X)] \leqslant m_{H,t}\}, \tag{40}$$

which satisfies all the conditions in Assumption 5.1 with covariate distribution P^X, budget constraints $(m_{L,t}, m_{H,t})$, and base propensity class $\mathcal{E}$.

We do not target the final covariances $V_{0,\mathrm{AIPW}} = V_{0:T,\mathrm{AIPW}}$ and $V_{0,\mathrm{EPL}} = V_{0:T,\mathrm{EPL}}$ in our discussion here since the sample sizes and budget constraints in future batches may not be known. If they are known in advance, then we can learn propensities for all future batches simultaneously at the time batch 2 covariates are observed, and indeed, target $V_{0:T,\mathrm{AIPW}}$ or $V_{0:T,\mathrm{EPL}}$ at that stage.

Now, suppose we split our observations into K folds as in a CSBAE, and for notational simplicity, we re-index the covariates in each batch $t = 1,\dots,T$, fold $k = 1,\dots,K$ as $X_{t1}^{(k)},\dots,X_{tn_{t,k}}^{(k)}$. Then, following (32), we can estimate $e_{t,\mathrm{AIPW}}^*(\cdot)$ for each batch $t \geqslant 2$ within each fold $k = 1,\dots,K$ by computing

$$(\hat{e}_{t1,\mathrm{AIPW}}^{(k)},\dots,\hat{e}_{tn_{t,k},\mathrm{AIPW}}^{(k)}) \in \underset{(e_1,\dots,e_{n_{t,k}})\in F_{n_{t,k}}}{\arg\max}\ \Psi((\hat{V}_{0:t,\mathrm{AIPW}}^{(k)})^{-1}), \quad k = 1,\dots,K, \tag{41}$$

where $F_{n_{t,k}} = F_{n_{t,k}}(m_{L,t}, m_{H,t})$ is defined as in (33), and the estimate $\hat{V}_{0:t,\mathrm{AIPW}}^{(k)}$ of $V_{0:t,\mathrm{AIPW}}$ is given by

$$\begin{aligned}\hat{V}_{0:t,\mathrm{AIPW}}^{(k)} &= \hat{V}_{0:t,\mathrm{AIPW}}^{(k)}(e_1,\dots,e_{n_{t,k}};\ \hat{e}_1^{(k)}(\cdot),\dots,\hat{e}_{t-1}^{(k)}(\cdot),\hat{v}_{1:(t-1)}^{(k)}(\cdot,\cdot)) \\ &= -\frac{1}{n_{t,k}}\sum_{i=1}^{n_{t,k}} \frac{\hat{v}_{1:(t-1)}^{(k)}(1, X_{ti}^{(k)})}{\hat{e}_{(0:t)i}^{(k)}} + \frac{\hat{v}_{1:(t-1)}^{(k)}(0, X_{ti}^{(k)})}{1-\hat{e}_{(0:t)i}^{(t)}},\end{aligned} \tag{42}$$

where each estimate $\hat{v}_{1:(t-1)}^{(k)}(z,\cdot)$ of the variance function $v_0(z,\cdot)$ is computed using only the observations from batches $u = 1,\dots,t-1$ within fold k. Note that any plug-in estimate of $\tau_0(\cdot)$ and θ_0 does not affect the optimization (41) and so can be omitted. The dependence of $\hat{V}_{0:t,\mathrm{AIPW}}^{(k)}$ on the optimization variables $e_1,\dots,e_{n_{t,k}}$ is through the mixture quantities

$$\hat{e}_{(0:t)i}^{(k)} \coloneqq \frac{1}{N_{1:t}}\left(\sum_{u=1}^{t-1} N_u \hat{e}_u^{(k)}(X_{ti}^{(k)}) + N_t e_i\right), \quad i = 1,\dots,n_{t,k}, \tag{43}$$

where for $u = 1,\dots,t-1$, $\hat{e}_u^{(k)}(\cdot)$ is the (possibly adaptive) propensity used to assign treatment in batch u, fold k. By comparing (43) and (39), we see that for each batch $u = 1,\dots,t$, N_u/N is being used as a plug-in estimate of κ_u and the adaptive propensity score $\hat{e}_u^{(k)}(\cdot)$ is used as a plug-in estimate of its limit $e_u(\cdot)$. Finally, as in the conclusion of Lemma 5.2, the learned adaptive propensity $\hat{e}_{t,\mathrm{AIPW}}^{(k)}(\cdot)$ to be used for treatment assignment in batch t, fold k is taken to be any choice in the predetermined base collection $\mathcal{E}$ with $\hat{e}_{t,\mathrm{AIPW}}^{(k)}(X_{ti}^{(k)}) = \hat{e}_{ti,\mathrm{AIPW}}^{(k)}$ for all $i = 1,\dots,n_{t,k}$.

Learning $e_{t,\mathrm{EPL}}^*(\cdot)$ is exactly analogous; first, we compute

$$(\hat{e}_{t1,\mathrm{EPL}}^{(k)},\dots,\hat{e}_{tn_{t,k},\mathrm{EPL}}^{(k)}) \in \underset{(e_1,\dots,e_{n_{t,k}})\in F_{n_{t,k}}}{\arg\max}\ \Psi((\hat{V}_{0:t,\mathrm{EPL}}^{(k)})^{-1}), \tag{44}$$

where

$$\begin{aligned}\hat{V}_{0:t,\mathrm{EPL}}^{(k)} &= \hat{V}_{0:t,\mathrm{EPL}}^{(k)}(e_1,\dots,e_{n_{t,k}};\ \hat{e}_1^{(k)}(\cdot),\dots,\hat{e}_{t-1}^{(k)}(\cdot),\hat{v}_{1:(t-1)}^{(k)}(\cdot,\cdot)) \\ &= \frac{1}{n_{t,k}}\sum_{i=1}^{n_{t,k}} \frac{\hat{e}_{(0:t)i}^{(k)}(1-\hat{e}_{(0:t)i}^{(k)})}{\hat{v}_{1:(t-1)}^{(k)}(0, X_{ti}^{(k)})\hat{e}_{(0:t)i}^{(k)} + \hat{v}_{1:(t-1)}^{(k)}(1, X_{ti}^{(k)})(1-\hat{e}_{(0:t)i}^{(k)})}\psi(X_{ti}^{(k)})\psi(X_{ti}^{(k)})^\top.\end{aligned} \tag{45}$$

Then, we assign treatment with any propensity $\hat{e}^{(k)}_{t,\mathrm{EPL}}(\cdot)$ in the base propensity class $\mathcal{E}$ satisfying $\hat{e}_{t,\mathrm{EPL}}(X^{(k)}_{ti}) = \hat{e}^{(k)}_{ti,\mathrm{EPL}}$ for all $i = 1,\dots, n_{t,k}$.

The main additional regularity condition required to ensure the adaptive propensities $\hat{e}^{(k)}_{t,\mathrm{AIPW}}(\cdot)$ and $\hat{e}^{(k)}_{t,\mathrm{EPL}}$ above converge at the desired $O_p(N^{-1/4})$ RMS rate to $e^*_{t,\mathrm{AIPW}}$ and $e^*_{t,\mathrm{EPL}}$ of (36) is the same rate of convergence in the estimates $\hat{\eta}^{(k)}_{\mathrm{AIPW}}(\cdot)$ and $\hat{\eta}^{(k)}_{\mathrm{EPL}}(\cdot)$ of the nuisance parameters $\eta_{0,\mathrm{AIPW}}(\cdot)$ and $\eta_{0,\mathrm{EPL}}(\cdot)$. We also strengthen the sample size asymptotics (3) by requiring

$$\frac{N_t}{N} = \kappa_t + O(N^{-1/4}), \quad t = 1,\dots, T. \tag{46}$$

Theorem 5.1. (Convergence of Algorithm 1) *Suppose $T \geqslant 2$, fix a batch $t \in \{2, \dots, T\}$ and suppose Assumption* 2.1 *holds along with* (46). *Further assume treatment in batches $1,\dots, t-1$ is assigned according to a CSBAE where the batch 1 propensities are $\hat{e}^{(1)}_1(\cdot) = \dots = \hat{e}^{(K)}_1(\cdot) = e_1(\cdot) \in \mathcal{F}_{\varepsilon_1}$ for some $\varepsilon_1 > 0$, and that for each fold $k = 1,\dots, K$:*

- *There exists an estimator $\hat{v}^{(k)}(\cdot)$ of the variance function $v_0(\cdot)$ depending only on the observations $\{W^{(k)}_{ui} | 1 \leqslant u \leqslant t-1\}$ in batches $1,\dots, t-1$ assigned to fold k, such that $||\hat{v}^{(k)}(z,\cdot) - v_0(z,\cdot)||_{2,P^X} = O_p(N^{-1/4})$ for $z = 0, 1$.*
- *There are universal constants $0 < c < C < \infty$ for which*

$$c \leqslant \inf_{(z,x)\in\{0,1\}\times\mathcal{X}} \min(\hat{v}^{(k)}(z,x), v_0(z,x)) \leqslant \sup_{(z,x)\in\{0,1\}\times\mathcal{X}} \max(\hat{v}^{(k)}(z,x), v_0(z,x)) \leqslant C.$$

- *The information function $\Psi(\cdot)$ satisfies Assumption* 5.2.

Let $\mathcal{E} \subseteq \mathcal{F}_0$ be any base propensity class satisfying the conditions of Assumption 5.1, *and define $P^{(k),X}_{N,t}$ to be the empirical distribution on the covariates $X^{(k)}_{t1},\dots, X^{(k)}_{tn_{t,k}}$ in batch t, fold k. Then, for any budget constraints $0 \leqslant m_{L,t} \leqslant m_{H,t} \leqslant 1$ and each fold $k = 1,\dots, K$, the following holds*:

(1) (*Design for $\hat{\theta}_{\mathrm{AIPW}}$*) *There exists a target propensity $e^*_{t,\mathrm{AIPW}}(\cdot) \in \mathcal{E}$ satisfying* (36) *that is unique P^X-almost surely. Additionally, there exists a solution $(\hat{e}^{(k)}_{t1,\mathrm{AIPW}}, \dots, \hat{e}^{(k)}_{tn_{t,k},\mathrm{AIPW}})$ to* (41); *any such solution has the property that any propensity $\hat{e}^{(k)}_t(\cdot) \in \mathcal{E}$ with $\hat{e}^{(k)}_t(X^{(k)}_{ti}) = \hat{e}^{(k)}_{ti,\mathrm{AIPW}}$ for $i = 1,\dots, n_{t,k}$ satisfies*

$$||\hat{e}^{(k)}_t - e^*_{t,\mathrm{AIPW}}||_{2,P^X} + ||\hat{e}^{(k)}_t - e^*_{t,\mathrm{AIPW}}||_{2,P^X_{N,t}} = O_p(N^{-1/4}).$$

If additionally, the linear treatment effect assumption (6) *holds for some basis function $\psi(X)$ containing an intercept, then*:

(2) (*Design for $\hat{\theta}_{\mathrm{EPL}}$*) *There exists a target propensity $e^*_{t,\mathrm{EPL}}(\cdot) \in \mathcal{E}$ satisfying* (36) *that is unique P^X-almost surely. Furthermore, there exists a solution $(\hat{e}^{(k)}_{t1,\mathrm{EPL}}, \dots, \hat{e}^{(k)}_{tn_{t,k},\mathrm{EPL}})$ to* (44); *any such solution has the property that any propensity $\hat{e}^{(k)}_t(\cdot) \in \mathcal{E}$ with $\hat{e}^{(k)}_t(X^{(k)}_{ti}) = \hat{e}^{(k)}_{ti,\mathrm{EPL}}$ for $i = 1,\dots, n_{t,k}$ satisfies*

$$||\hat{e}^{(k)}_t - e^*_{t,\mathrm{EPL}}||_{2,P^X} + ||\hat{e}^{(k)}_t - e^*_{t,\mathrm{EPL}}||_{2,P^X_{N,t}} = O_p(N^{-1/4}).$$

Proof. See Appendix C.10 (Supplementary material). □

6 Numerical simulations

We implement Algorithm 1 to construct some synthetic CSBAEs that illustrate the finite sample performance of our proposed methods. For simplicity, we consider $T = 2$ batches throughout. Our evaluation metric is the average mean squared error of the estimators $\hat{\theta}_{\mathrm{AIPW}}$ and $\hat{\theta}_{\mathrm{EPL}}$ computed at the end of each CSBAE. As a baseline, we also compute feasible variants of the linearly aggregated estimators $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$ and $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{EPL}}$ computed on a "simple

randomized controlled trial (RCT)": i.e., a non-adaptive batch experiment with a constant propensity score in each batch. We additionally consider the approach of Hahn et al. [2] for design and estimation of $\theta_{0,\mathrm{ATE}}$. This is equivalent to using Algorithm 1 for design (without sample splitting, i.e., $K = 1$) and using the pooled $\hat{\theta}_{\mathrm{AIPW}}$ as the final estimator, but with the covariates X replaced everywhere by a coarse discretization $S = S(X)$. As a hybrid, we also consider using the discretized covariates S for design but computing the final estimates $\hat{\theta}_{\mathrm{AIPW}}$ and $\hat{\theta}_{\mathrm{EPL}}$ using the full original covariate X. To separately attribute efficiency gains to design and pooling, we also consider the linearly aggregated estimators $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$ and $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{EPL}}$ when a modification of Algorithm 1 that targets $V_{2,\mathrm{AIPW}}$ and $V_{2,\mathrm{EPL}}$ (the asymptotic variances of the estimators $\hat{\theta}_{2,\mathrm{AIPW}}$ and $\hat{\theta}_{2,\mathrm{EPL}}$ given in Section 3, depending only on observations in batch 2) is used for design.

We consider four DGPs, distinguished by whether the covariate dimension d is 1 or 10 and whether the conditional variance functions $v_0(\cdot,\cdot)$ are homoskedastic (with $v_0(0,x) = v_0(1,x) = 1$ for all $x \in \mathcal{X}$) or heteroskedastic (with $v_0(0,x) = v_0(1,x)/2 = \exp((\mathbf{1}_d^\top x)/(2\sqrt{d}))$. The scaling by the covariate dimension d in the heteroskedastic variance functions ensures the variance $v_0(z,X)$ is independent of the covariate dimension d. In all of the DGPs, the covariates are i.i.d. spherical Gaussian, i.e., $P^X = \mathcal{N}(0, I_d)$. The outcome mean functions are taken to be $m_0(0,x) = m_0(1,x) = \mathbf{1}_d^\top x$ for $\mathbf{1}_d = (1,\ldots,1)' \in \mathbb{R}^d$. For estimating $\theta_{0,\mathrm{PL}}$, we use the basis functions $\psi(x) = (1, x^\top)^\top \in \mathbb{R}^p$, where $p = d + 1$. Note $\theta_{0,\mathrm{ATE}} = \theta_{0,\mathrm{PL}} = 0$. The potential outcomes $Y(0)$, $Y(1)$ are generated as follows:

$$(\varepsilon(0), \varepsilon(1))|X \sim \mathcal{N}((0,0)^\top, \mathrm{diag}(v_0(0,X), v_0(1,X)))$$
$$Y(z) = m_0(z,X) + \varepsilon(z), \quad z = 0,1.$$

For each DGP, we run Algorithm 1 with $K = 2$ folds, batch sample sizes $N_1 = N_2 = 1{,}000$, information function $\Psi = \Psi_a(\cdot)$ (corresponding to A-optimality), and treatment fraction constraints $m_{L,t} = m_{H,t} = 0.2$ for $t = 1, 2$. In Appendix D (see Supplementary material), we present additional simulation results, where $m_{L,1} = m_{H,1}$ remain at 0.2 but $m_{L,2} = m_{H,2} = 0.4$, so that the treatment budget for the second batch has increased. The initial propensity score $e_1(\cdot)$ is taken to be constant (i.e., $e_1(x) = 0.2$ for all $x \in \mathcal{X}$), and the base propensity class $\mathcal{E}$ is the set of all 1-Lipschitz functions taking values in $[0,1]$ when $d = 1$ (cf. Example 5.2). When $d = 10$, we take $\mathcal{E}$ as in (34), with $\{\theta_1, \ldots, \theta_m\}$ the collection of vectors $(a_1, \ldots, a_{11})' \in \mathbb{R}^{11}$ with each coordinate $a_i \in \{-2, -1, 0, 1, 2\}$ and no more than two of the a_i's nonzero.

The discretization used to implement the approach of Hahn et al. [2] partitions $\mathbb{R}^d$ into four bins based on the quartiles of $\mathbf{1}_d^\top X$. Note that this partition is along the single dimension along which the variance functions $v_0(\cdot,\cdot)$ vary in the heteroskedastic DGPs, so we would expect this to perform better than in practice, where the structure of the variance functions is not known (it could possibly be learned, as in Tabord-Meehan [30]). The choice of four bins is based on the experiments of Hahn et al. [2], which find minimal performance difference between two and six bins for the DGPs they consider. We denote their "binned" AIPW estimator by $\hat{\theta}^{(\mathrm{bin})}_{\mathrm{AIPW}}$. Recall, as indicated earlier, that this is equivalent to $\hat{\theta}_{\mathrm{AIPW}}$ when the only available covariate is the binned $S(X)$ and no cross-fitting is used. For $\hat{\theta}^{(\mathrm{bin})}_{\mathrm{AIPW}}$, the conditional means $\tilde{m}_0(z,s) = \mathbb{E}(Y(z)|S = s)$, $z = 0, 1$ are estimated nonparametrically by sample outcome means among all units with $Z = z$ and $S = s$ in the appropriate batch (es) and fold(s). Similarly, the conditional variances $\tilde{v}_0(z,s) = \mathrm{Var}(Y(z)|S = s)$ are estimated by sample outcome variances.

For all simulations, the concave maximizations are performed using the CVXR software [48] and the MOSEK solver [49]. All estimates $\hat{m}(z,\cdot)$ of the mean function $m_0(z,\cdot)$, $z = 0, 1$ are computed by fitting a generalized additive model (GAM) to the outcomes Y and covariates X from the observations with treatment indicators equal to z in the appropriate fold(s) and batch(es). The GAMs use a thin-plate regression spline basis [50] and the degrees of freedom are chosen using the generalized cross-validation procedure implemented in the `mgcv` package in R [51]. Variance function estimates $\hat{v}(z,\cdot)$ are computed by first computing $\hat{m}(z,\cdot)$ as above on the appropriate observations, then fitting a GAM on these same observations to predict the squared residuals $(Y - \hat{m}(z,X))^2$ from X.

6.1 ATE

Table 1 shows the performance of the various design and estimation procedures for $\theta_{0,\mathrm{ATE}}$ that we consider. Each entry is a "relative efficiency": i.e., a ratio of the MSE of the baseline approach (which computes the linearly aggregated $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$ on a simple RCT) to the MSE of the relevant approach. The simulated relative efficiencies in the table estimate these MSEs by averaging the squared error of each estimator over 1,000 simulations. The 90% confidence intervals for the true finite sample relative efficiency are computed using 10,000 bootstrap replications of these 1,000 simulations. Finally, the asymptotic relative efficiencies in Table 1 are computed by estimating the asymptotic variance of each estimator using the appropriate formula, i.e., $V_{0,\mathrm{AIPW}}$ for $\hat{\theta}_{\mathrm{AIPW}}$ and $V^{(\mathrm{LA})}_{\mathrm{AIPW}}$ for $\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$. The computation is based on a non-adaptive batch experiment with second batch propensity $e_2(\cdot)$ equal to the average of the learned propensities from the relevant approach across the 1,000 simulations, as a closed form solution for the limiting $e_2(\cdot)$ is not easily obtained in general. All expectations over the covariate distribution are computed using Monte Carlo integration.

For both of the homoskedastic DGPs, it is straightforward to show using Jensen's inequality that $e_2^*(x) = 0.2$ for all x. It can be further shown that the unpooled and pooled estimators are asymptotically equivalent. Thus, for the homoskedastic DGPs, there is no asymptotic efficiency gain to be had. With unequal budget constraints, there is some asymptotic benefit to pooling with homoskedastic variance functions (Appendix D.1, see Supplementary material); however, the optimal design will still be the simple RCT. Nonetheless, we do observe some finite sample benefits to pooling in Table 1. For example, the simulated relative efficiency of $\hat{\theta}_{\mathrm{AIPW}}$ on the simple RCT is significantly larger than 1 for both $d = 1$ and $d = 10$. We attribute this to improved nuisance estimates for the pooled estimator, as discussed further in Appendix D.2 (see Supplementary material). As one might expect, this finite sample improvement from pooling is apparently offset by variance in both the flexible and

Table 1: Simulated and asymptotic relative efficiencies (as defined in Section 6) of the various design and estimation approaches for $\theta_{0,\mathrm{ATE}}$ that we study numerically under each of the four DGP's described in the text

DGP	Estimator	Design	Sim. rel. eff. (90% CI)	Asymp. rel. eff.
$d = 1$, Homoskedastic	$\hat{\theta}_{\mathrm{AIPW}}$	Flexible	0.989 (0.965, 1.013)	1.000
	$\hat{\theta}_{\mathrm{AIPW}}$	Binned	1.011 (0.977, 1.046)	0.999
	$\hat{\theta}_{\mathrm{AIPW}}$	Simple RCT	1.016 (1.003, 1.029)	1.000
	$\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$	Flexible	0.984 (0.971, 0.998)	0.999
	$\hat{\theta}^{(\mathrm{bin})}_{\mathrm{AIPW}}$	Binned	0.919 (0.877, 0.961)	0.885
$d = 1$, Heteroskedastic	$\hat{\theta}_{\mathrm{AIPW}}$	Flexible	1.016 (0.968, 1.064)	1.048
	$\hat{\theta}_{\mathrm{AIPW}}$	Binned	1.046 (0.997, 1.096)	1.041
	$\hat{\theta}_{\mathrm{AIPW}}$	Simple RCT	0.997 (0.979, 1.016)	1.000
	$\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$	Flexible	1.008 (0.971, 1.045)	1.024
	$\hat{\theta}^{(\mathrm{bin})}_{\mathrm{AIPW}}$	Binned	1.001 (0.946, 1.057)	0.963
$d = 10$, Homoskedastic	$\hat{\theta}_{\mathrm{AIPW}}$	Flexible	1.039 (0.990, 1.089)	1.000
	$\hat{\theta}_{\mathrm{AIPW}}$	Binned	1.011 (0.959, 1.065)	1.000
	$\hat{\theta}_{\mathrm{AIPW}}$	Simple RCT	1.081 (1.052, 1.110)	1.000
	$\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$	Flexible	1.047 (1.018, 1.077)	1.000
	$\hat{\theta}^{(\mathrm{bin})}_{\mathrm{AIPW}}$	Binned	0.477 (0.439, 0.519)	0.417
$d = 10$, Heteroskedastic	$\hat{\theta}_{\mathrm{AIPW}}$	Flexible	1.081 (1.023, 1.143)	1.036
	$\hat{\theta}_{\mathrm{AIPW}}$	Binned	1.050 (0.986, 1.117)	1.000
	$\hat{\theta}_{\mathrm{AIPW}}$	Simple RCT	1.102 (1.070, 1.135)	1.000
	$\hat{\theta}^{(\mathrm{LA})}_{\mathrm{AIPW}}$	Flexible	0.991 (0.952, 1.030)	1.024
	$\hat{\theta}^{(\mathrm{bin})}_{\mathrm{AIPW}}$	Binned	0.651 (0.597, 0.708)	0.595

binned design procedures. Still, in the homoskedastic DGPs, our adaptive approaches do not show significant finite sample performance decline relative to the baseline, which here is an oracle. With unequal treatment constraints, we further obtain asymptotic efficiency gains from pooling (Appendix D.1, Supplementary material).

We note that using the discretized covariate $S(X)$ in place of X for both design and estimation, as in Hahn et al. [2], leads to a substantial loss of efficiency. Indeed, when $d = 10$, the (asymptotic and simulated) variance of the estimator $\hat{\theta}^{(\text{bin})}_{\text{AIPW}}$ is more than double that of our baseline under the homoskedastic DGP, both asymptotically and in our finite sample simulations. This efficiency loss occurs because the discretized $S(X)$ explains much less of the variation in the potential outcomes $Y(z)$ than the original X. We expect greater precision losses from this discretization at the estimation stage when the variance functions $v_0(z, \cdot)$ vary substantially within the strata defined by $S(X)$.

For the heteroskedastic DGPs, we see modest asymptotic efficiency gains from both pooling and design. Design using the flexible base propensity class leads to about a 2.4% asymptotic efficiency gain for both $d = 1$ and $d = 10$, while pooling provides an additional 1–2% gain. These small asymptotic gains appear to be largely canceled out at our sample sizes by the finite sample variability in learning propensity scores, limiting the net finite sample gains from design. Of course, with greater heteroskedasticity and/or differences between $v_0(0, \cdot)$ and $v_0(1, \cdot)$, we would expect greater efficiency gains from design; in our simulations, we have chosen to keep these differences within common ranges in social science studies as per Blackwell et al. [3].

6.2 Partially linear model

Unlike for estimating $\theta_{0,\text{ATE}}$, for estimating $\theta_{0,\text{PL}}$, we see clear efficiency gains over the baseline from *design* when $d = 1$ (Table 2). For instance, the linearly aggregated estimator $\hat{\theta}^{(\text{LA})}_{\text{EPL}}$ exhibits a 5.6% asymptotic efficiency gain as a result of the flexible design; replacing this with the pooled estimator $\hat{\theta}_{\text{EPL}}$ then yields a total asymptotic gain of 10.0% over the baseline, even in the homoskedastic DGP. The analogous gains for the heteroskedastic DGP are slightly larger. We once again observe a substantial finite sample benefit to pooling, with the simulated relative efficiency of the approaches using the pooled $\hat{\theta}_{\text{EPL}}$ tending to be larger than the

Table 2: Simulated and asymptotic relative efficiencies (as defined in Section 6) of the various design and estimation approaches for $\theta_{0,\text{EPL}}$ that we study numerically under each of the four DGP's described in the text

DGP	Estimator	Design	Sim. rel. eff. (90% CI)	Asymp. rel. eff.
$d = 1$, Homoskedastic	$\hat{\theta}_{\text{EPL}}$	Flexible	1.150 (1.103, 1.198)	1.100
	$\hat{\theta}_{\text{EPL}}$	Binned	1.057 (1.009, 1.107)	1.026
	$\hat{\theta}_{\text{EPL}}$	Simple RCT	1.038 (1.018, 1.058)	1.000
	$\hat{\theta}^{(\text{LA})}_{\text{EPL}}$	Flexible	1.049 (1.016, 1.083)	1.056
$d = 1$, Heteroskedastic	$\hat{\theta}_{\text{EPL}}$	Flexible	1.310 (1.214, 1.412)	1.112
	$\hat{\theta}_{\text{EPL}}$	Binned	1.049 (0.972, 1.133)	0.979
	$\hat{\theta}_{\text{EPL}}$	Simple RCT	1.084 (1.014, 1.159)	1.000
	$\hat{\theta}^{(\text{LA})}_{\text{EPL}}$	Flexible	0.983 (0.880, 1.076)	1.073
$d = 10$, Homoskedastic	$\hat{\theta}_{\text{EPL}}$	Flexible	1.229 (1.203, 1.256)	1.023
	$\hat{\theta}_{\text{EPL}}$	Binned	1.177 (1.147, 1.206)	1.000
	$\hat{\theta}_{\text{EPL}}$	Simple RCT	1.228 (1.204, 1.251)	1.000
	$\hat{\theta}^{(\text{LA})}_{\text{EPL}}$	Flexible	1.023 (1.011, 1.035)	1.019
$d = 10$, Heteroskedastic	$\hat{\theta}_{\text{EPL}}$	Flexible	0.997 (0.963, 1.032)	1.071
	$\hat{\theta}_{\text{EPL}}$	Binned	0.932 (0.899, 0.965)	1.002
	$\hat{\theta}_{\text{EPL}}$	Simple RCT	0.982 (0.952, 1.014)	1.000
	$\hat{\theta}^{(\text{LA})}_{\text{EPL}}$	Flexible	0.969 (0.942, 0.996)	1.060

asymptotic relative efficiency. We attribute this to both improved use of nuisance estimates by the pooled estimator (as in the ATE case) as well as a more fundamental finite sample efficiency boost due to the fact that the asymptotic variance $V_{0,\text{EPL}}$ is not exact for the oracle pooled $\hat{\theta}^*_{\text{EPL}}$ in finite samples, whereas $V_{0,\text{AIPW}}$ is exact for $\hat{\theta}^*_{\text{AIPW}}$ (see Appendix D.2). When $d = 10$, design introduces some more salient finite sample variance from the errors in the concave maximization procedure. This is offset in both the homoskedastic and heteroskedastic DGPs by the asymptotic gains from the flexible design, so that when $\hat{\theta}_{\text{EPL}}$ is used, the flexible design ultimately performs similarly to the simple RCT in finite samples.

Even if we use $\hat{\theta}_{\text{EPL}}$ and hence the original covariate(s) X in the final estimation step, we see that the "binned" design, which uses only $S(X)$ for choosing the second batch propensity score struggles to learn a substantially better propensity score than the simple RCT in all DGPs. Indeed, in the heteroskedastic DGP with $d = 1$, we see a 2.1% asymptotic efficiency *loss* relative to the baseline from using the binned design. By comparison, there is an 11.2% asymptotic efficiency gain from using the flexible design. The efficiency loss can occur with the binned design because the objective in (28) changes when working in terms of the discretized covariate $S(X)$ instead of X. In other words, even though the simple RCT propensity $e_2(x) = 0.2$ is within the class of propensities that can be chosen by the binned design, it is worse according to the binned objective based on $S(X)$, but not according to the objective based on the original X.

7 Discussion

We view our primary technical contribution in this article to be a careful extension of the DML framework that enables both estimation and design in batch experiments based on pooled treatment effect estimators. This allows the investigator to take advantage of the efficiency gains from pooling and design without needing to make strong parametric assumptions or to discretize their covariates. As our numerical study in Section 6 shows, the latter can more than wipe out any efficiency gains from design.

Related to our work is the extensive literature on combining observational data with a (single batch) randomized experiment. In that setting a primary concern is mitigating bias from unobserved confounders in the observational data (e.g., [52,53]). By contrast, in our setting, unconfoundedness holds by design in each batch of the experiment. It would be useful to examine if the ideas from the present work can be extended to the observational setting where confounding bias is a concern.

Acknowledgement: The authors thank Stefan Wager, Lihua Lei, and Kevin Guo for comments that improved the content of this article.

Funding information: H.L. was partially supported by the Stanford Interdisciplinary Graduate Fellowship (SIGF). This work was also supported by the NSF under Grant DMS-2152780.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

References

[1] Neyman J. On the two different aspects of the representative model: The method of stratified sampling and the method of purposive selection. J R Stat Soc. 1934;97(4):558.

[2] Hahn J, Hirano K, Karlan D. Adaptive experimental design using the propensity score. J Business Econ Stat. 2011;29(1):96–108.

[3] Blackwell M, Pashley NE, Valentino D. Batch adaptive designs to improve efficiency in social science experiments. Harvard University; 2022. https://www.mattblackwell.org.

[4] Zhao J. Adaptive Neyman allocation. SSRN; 2023. https://arxiv.org/abs/2309.08808.
[5] Dai J, Gradu P, Harshaw C. Clip-OGD: An experimental design for adaptive Neyman allocation in sequential experiments. 2023. arXiv:2305.17187.
[6] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
[7] Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. Double/debiased machine learning for treatment and structural parameters. Econ J. 2018 01;21(1):C1–C68. doi: https://doi.org/10.1111/ectj.12097.
[8] Russo DJ, Van Roy B, Kazerouni A, Osband I, Wen Z, et al. A tutorial on Thompson sampling. Foundations and Trends in Machine Learning. 2018;11(1):1–96.
[9] Hao B, Lattimore T, Szepesvari C. Adaptive exploration in linear contextual bandit. In: International Conference on Artificial Intelligence and Statistics. PMLR; 2020. p. 3536–45.
[10] Xu L, Honda J, Sugiyama M. A fully adaptive algorithm for pure exploration in linear bandits. In: International Conference on Artificial Intelligence and Statistics. PMLR; 2018. p. 843–51.
[11] Kasy M, Sautmann A. Adaptive treatment assignment in experiments for policy choice. Econometrica. 2021;89(1):113–32.
[12] Luedtke AR, van der Laan MJ. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. Ann Stat. 2016 Apr;44(2):713–42.
[13] Hadad V, Hirshberg DA, Zhan R, Wager S, Athey S. Confidence intervals for policy evaluation in adaptive experiments. Proc Nat Acad Sci. 2021;118(15):e2014602118.
[14] Zhang K, Janson L, Murphy S. Inference for batched bandits. Adv Neural Inform Process Syst. 2020;33:9818–29.
[15] Zhang K, Janson L, Murphy S. Statistical inference with M-estimators on adaptively collected data. Adv Neural Inform Process Syst. 2021;34:7460–71.
[16] Owen AB, Varian H. Optimizing the tie-breaker regression discontinuity design. Electronic J Stat. 2020;14:4004–27.
[17] Morrison TP, Owen AB. Optimality in multivariate tie-breaker designs. 2022. arXiv:2202.10030.
[18] Li HH, Owen AB. A general characterization of optimal tie-breaker designs. Ann Stat. 2023;51(3):1030–57.
[19] Kluger DM, Owen AB. Kernel regression analysis of tie-breaker designs. Electron J Stat. 2023;17(1):243–90.
[20] van der Laan NJ. The construction and analysis of adaptive group sequential designs. UC Berkeley Division of Biostatistics Working Paper Series. 2008 Mar. https://biostats.bepress.com/ucbbiostat/paper232.
[21] Kato M, Ishihara T, Honda J, Narita Y. Efficient adaptive experimental design for average treatment effect estimation. 2020. arXiv:200205308.
[22] van der Laan NJ, Lendle S. Online targeted learning. UC Berkeley Division of Biostatistics Working Paper Series. 2014 Sep. https://biostats.bepress.com/ucbbiostat/paper330.
[23] Michaelides M, Poe-Yamagata E, Benus J, Tirumalasetti D. Impact of the reemployment and eligibility assessment (REA) initiative in Nevada. LLC in Oakland, CA: IMPAQ International; 2012.
[24] Michaelides M, Mian P. Low-cost randomized control trial study of the Nevada reemployment and eligibility assessment (REA) program. LLC in Oakland, CA: IMPAQ International; 2020.
[25] Zhang LX, Hu F, Cheung SH, Chan WS. Asymptotic properties of covariate-adjusted response-adaptive designs. Ann Stat. 2007 Jul;35(3):1166–82.
[26] Chambaz A, van der Laan MJ, Zheng W. Targeted covariate-adjusted response-adaptive lasso-based randomized controlled trials. Modern adaptive randomized clinical trials: statistical, operational, and regulatory aspects. London, UK: Chapman and Hall; 2014. p. 345–68.
[27] Zhu H, Zhu H. Covariate-adjusted response-adaptive designs based on semiparametric approaches. Biometrics. 2023;79(4):2895–906.
[28] Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica. 2003;71(4):1161–89.
[29] Kato M, McAlinn K, Yasui S. The adaptive doubly Robust estimator and a paradox concerning logging policy. In: Advances in neural information processing systems. Vol. 34. Red Hook, NY: Curran Associates, Inc.; 2021. p. 1351–64.
[30] Tabord-Meehan M. Stratification trees for adaptive randomisation in randomised controlled trials. Rev Econ Studies. 2022 Dec;90(5):2646–73.
[31] Cytrynbaum M. Designing representative and balanced experiments by local randomization. 2021. arXiv:2111.08157.
[32] Che E, Namkoong H. Adaptive experimentation at scale: Bayesian algorithms for flexible batches. 2023. arXiv:2303.11582.
[33] Robinson PM. Root-n-consistent semiparametric regression. Econom J Econ Soc. 1988;56:931–54.
[34] Chamberlain G. Efficiency bounds for semiparametric regression. Econom J Econ Soc. 1992;60(3):567–96.
[35] Ma Y, Chiou JM, Wang N. Efficient semiparametric estimator for heteroscedastic partially linear models. Biometrika. 2006;93(1):75–84.
[36] Garivier A, Kaufmann E. Optimal best arm identification with fixed confidence. In: Conference on Learning Theory. PMLR; 2016. p. 998–1027.
[37] Slud E, Vonta I, Kagan A. Combining estimators of a common parameter across samples. Stat Theory Related Fields. 2018;2(2):158–71.
[38] Klaassen CA. Consistent estimation of the influence function of locally asymptotically linear estimators. Ann Stat. 1987;15(4):1548–62.

[39] Zheng W, van der Laan MJ. Cross-validated targeted minimum-loss-based estimation. In: Targeted Learning. New York, NY: Springer New York; 2011. p. 459–74. doi: https://link.springer.com/10.1007/978-1-4419-9782-1_27.
[40] Rothe C. The value of knowing the propensity score for estimating average treatment effects. IZA Discussion Papers. 2016. https://www.ssrn.com/abstract=2797560.
[41] Atkinson A, Donev A, Tobias R. Optimum experimental designs, with SAS. vol. 34. Oxford: Oxford University Press; 2007.
[42] Pukelsheim F. Optimal design of experiments. Philadelphia, PA: SIAM; 2006.
[43] Vapnik V. Principles of risk minimization for learning theory. Advances in Neural Information Processing Systems. 1991;4:831–8.
[44] Montanari A, Saeed BN. Universality of empirical risk minimization. In: Conference on Learning Theory. PMLR; 2022. p. 4310–2.
[45] Kosorok MR. Introduction to empirical processes and semiparametric inference. New York, NY: Springer; 2008.
[46] Wainwright MJ. High-dimensional statistics: A non-asymptotic viewpoint. vol. 48. Cambridge: Cambridge University Press; 2019.
[47] van der Vaart AW, Wellner JA. Weak convergence and empirical processes. Springer Series in Statistics. New York: Springer; 1996.
[48] Fu A, Narasimhan B, Boyd S. CVXR: An R package for disciplined convex optimization. 2017. arXiv:1711.07582.
[49] MOSEK ApS. MOSEK Rmosek package 10.1.9; 2022. https://docs.mosek.com/latest/rmosek/index.html.
[50] Wood SN. Thin-plate regression splines. J R Stat Soc (B). 2003;65(1):95–114.
[51] Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Amer Stat Assoc. 2004;99(467):673–86.
[52] Rosenman ET, Owen AB. Designing experiments informed by observational studies. J Causal Inference. 2021;9(1):147–71.
[53] Gagnon-Bartsch J, Sales A, Wu E, Botelho A, Erickson J, Miratrix L, et al. Precise unbiased estimation in randomized experiments using auxiliary observational data. J Causal Inference. 2023 Aug;11(1):20220011.