Asymptotically Best Causal Effect Identification with Multi-Armed Bandits

Alan Malek
DeepMind London
alanmalek@deepmind.com

Silvia Chiappa
DeepMind London
csilvia@deepmind.com

Abstract

This paper considers the problem of selecting a formula for identifying a causal quantity of interest among a set of available formulas. We assume an online setting in which the investigator may alter the data collection mechanism in a data-dependent way with the aim of identifying the formula with lowest asymptotic variance in as few samples as possible. We formalize this setting by using the best-arm-identification bandit framework where the standard goal of learning the arm with the lowest loss is replaced with the goal of learning the arm that will produce the best estimate. We introduce new tools for constructing finite-sample confidence bounds on estimates of the asymptotic variance that account for the estimation of potentially complex nuisance functions, and adapt the best-arm-identification algorithms of LUCB and Successive Elimination to use these bounds. We validate our method by providing upper bounds on the sample complexity and an empirical study on artificially generated data.

1 Introduction

Many scientific disciplines, including biology, healthcare, and social and behavioral sciences, are concerned with estimating the *causal effect* of some exposure on an outcome of interest from *observational data*. When the structure of the causal relationships between relevant variables is known and satisfies certain conditions, one can use *identification formulas* derived from the do-calculus to express a causal quantity τ as a functional of the observational data distribution [20, 21], after which an estimate of τ can be obtained with an appropriate estimator. For many structures, τ can be identified using several formulas involving different covariates. When choosing which formula to use, the investigator should consider the statistical performance in addition to practical considerations such as the cost of data collection or the difficulty of observing certain covariates. For simplicity of exposition, we only consider one estimator for each identification formula, but the method proposed in this paper can also be used in the more general case of several estimators.

Recently introduced graphical criteria allow the comparison of identification formulas derived from the popular *adjustment criterion* w.r.t. the asymptotic variance of linear and non-parametric estimators [6, 24, 26, 30]. However, information about the causal graph structure alone does not allow comparison of certain adjustment formulas nor of arbitrary identification formulas. In addition, such criteria only focus on statistical performance and disregard practical considerations. A selection method that is applicable to arbitrary identification formulas and accounts for both statistical performance and practical considerations must make use of data.

The goal of this work is to introduce such a method in the setting in which the investigator can actively make observations and can, therefore, first gather some data for the purpose of selecting a formula before collecting a large dataset to be used for estimating τ. For example, the investigator could be selecting which laboratory tests to require for a large observational study or which sensors

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

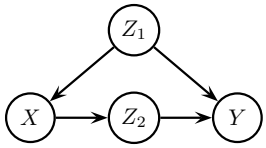
to install for a multi-year environmental study. In both cases, the investigator would like to actively collect some observations to aid in the final decision.

A naïve approach would be for the investigator to include the full set of covariates needed by all identification formulas in each observation of the initial dataset, then make a decision by evaluating the performance of every formula. We instead study a sequential decision approach where the investigator decides which subset of covariates to include in each observation in a data-dependent way, with the hope to concentrate on the covariates needed by formulas that are more likely to be optimal and identify the best formula with fewer samples. We formalize this sequential decision strategy as a best-arm-identification bandit problem [17] where each arm corresponds to an estimator and the classic goal of learning the arm with the best mean is replaced by the goal of learning the estimator with the best long-run statistical performance.

To enable us to meaningfully compare long-run behavior of different estimators, we focus on $\sqrt{n}$-*consistent estimators*, *i.e.* with error $o_p(1/\sqrt{n})$, and use the *asymptotic variance*, which is the leading constant in the error rate [19], as the performance evaluation metric. We adapt well-known bandit algorithms to this goal by introducing finite-sample confidence intervals on the asymptotic variance. In particular, we show how to obtain such intervals for *asymptotically linear estimators* with known *influence functions*, which form a large class of $\sqrt{n}$-consistent estimators that can be constructed for any causal effect [12]. Estimators of causal quantities typically contain high-dimensional *nuisance functions* η, and modern practice is to model such functions with large classes that do not have classical Donsker-like smoothness, such as *e.g.* neural networks, often leading to estimates of η that have slow convergence rate of $\mathcal{O}(n^{-1/4})$ and therefore to loss of asymptotic linearity. Recent work in double machine learning has established Neyman orthogonality as a sufficient condition for asymptotic linearity even when η is estimated at a $\mathcal{O}(n^{-1/4})$ rate [2]. This result allows us to include estimators with complex nuisance functions in our setting. A critical aspect of our setting is that the asymptotic variance needs to be estimated at the same rate as τ; otherwise the samples needed to estimate the asymptotic variance could be larger than the samples needed to just estimate τ directly. Our main theorems show that, even without Neyman orthogonality, $\mathcal{O}(n^{-1/2})$ estimation of the asymptotic variance is possible whenever $\mathcal{O}(n^{-1/2})$ estimation of τ is possible.

The rest of the paper is organized as follows. Section 2 introduces the technical assumptions on the estimators and formalizes the sequential decision problem as a best-arm-identification problem in the multi-armed bandit setting. Section 3 shows how to construct finite-sample confidence sequences (*i.e.* confidence intervals that hold uniformly over sequences of random variables) on the asymptotic variance, which are then used in Section 4 to adapt the *LUCB* and *Successive Elimination* bandit algorithms to our setting. Sample complexity upper bounds are also derived. Finally, Section 5 presents an empirical evaluation of our methods on artificially generated data, showing significant sample complexity reduction with respect to a naïve uniform sampling method.

2 Causal Effect Identification as a Multi-Armed Bandit Problem



Let $W := (X, Y, \mathcal{Z})$ be a tuple of random variables where X corresponds to an exposure, Y to an outcome of interest, and $\mathcal{Z}$ to a set of observable covariates; let p and $p_{\text{do}(x)}$ denote the *observational* and *interventional distributions*, respectively, differing in passively observing X versus intervening on X by fixing its value to x.

We wish to estimate the *causal contrast* $\tau := \sum_x \lambda_x \mu_{\text{do}(x)}$, where $\mu_{\text{do}(x)} := \mathbb{E}_{p_{\text{do}(x)}}[Y|X = x]$ and λ_x is a scalar, in the overidentified setting in which $\mu_{\text{do}(x)}$ can be expressed as a functional of the observational distribution with K different *identification formulas*, each using a different subset of covariates. For example, in the causal DAG above, $\mu_{\text{do}(x)}$ can be identified with the *adjustment criterion* using Z_1, *i.e.* as $\mu_{\text{do}(x)} = \mathbb{E}_p[\mu_x(Z_1)]$ with $\mu_x(Z_1) := \mathbb{E}_p[Y|X = x, Z_1]$, or with the *frontdoor criterion* using Z_2, *i.e.* as $\mu_{\text{do}(x)} = \sum_{z_2} p(Z_2 = z_2|X = x) \sum_{x'} p(X = x')\mu_{x'}(Z_2 = z_2)$. Without loss of generality, we assume that each identification formula is associated with only one estimator $\hat{\tau}_k$ of τ, formally defined as a collection of mappings $\{w_k^i\}_{i=1}^n \to \mathbb{R}$, where $w_k^i \sim p(W_k)$ and $W_k := (X, Y, \mathcal{Z}_k)$.

Assumptions on Estimators of τ. We focus on $\sqrt{n}$-*consistent estimators*, *i.e.* with error $o_p(1/\sqrt{n})$, and compare their statistical performance by using the *asymptotic variance*, which is the leading

constant in the error rate [19]. To achieve that, our proposed algorithms in Section 4 require finite-sample confidence interval on the asymptotic variance. While any finite-sample confidence interval can be deployed, we show how to construct such intervals for a very common and often-studied class of $\sqrt{n}$-consistent estimators, namely *asymptotically linear estimators* with known *influence functions* in the present of *nuisance functions*.

Estimators of causal quantities often contain a *nuisance function* η, namely a function that must be estimated in order to estimate τ, but is of no interest otherwise. For example, consider the *augmented inverse probability weighted (AIPW) estimator*, which is a popular estimator for the adjustment criterion that we will use in our experiments, defined as $\hat{\tau}(\mathcal{D}) = \mathbb{E}_{\mathcal{D}}\left[\sum_x \lambda_x \left(\frac{I_x(X)}{\hat{e}_x(\mathcal{Z})}\left(Y - \hat{\mu}_x(\mathcal{Z})\right) + \hat{\mu}_x(\mathcal{Z})\right)\right]$, where $\mathbb{E}_{\mathcal{D}}[\cdot]$ denotes empirical expectation w.r.t. dataset $\mathcal{D} = \{w^i\}_{i=1}^n$, $w^i \sim p(W)$, I_x is the indicator function, and $\hat{\mu}_x, \hat{e}_x$ are estimators of $\mu_x(Z)$ and $e_x(\mathcal{Z}) := p(X = x|\mathcal{Z})$. This estimator has nuisance function $\eta = (\mu_x, e_x)$, which must be estimated before computing $\hat{\tau}(\mathcal{D})$.

Definition 1. *An estimator $\hat{\tau}$ of τ with nuisance function η is asymptotically linear [8, 27] if there exists a function ϕ, called influence function (IF), satisfying $\mathbb{E}_p[\phi(W, \eta, \tau)] = 0$ and $\mathbb{E}_p[\phi^2(W, \eta, \tau)] < \infty$, such that $\sqrt{n}\big(\hat{\tau}(\mathcal{D}) - \tau\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi(w^i, \eta, \tau) + o_p(1)$. By the central limit theorem, $\hat{\tau}$ is $\sqrt{n}$-consistent and asymptotically normal with asymptotic variance $\sigma^2 = \mathbb{E}_p[\phi^2(W, \eta, \tau)]$.*

If the influence function can be decomposed as $\phi(W, \eta, \tau) = \psi(W, \eta) - \tau$, then ψ is called the uncentered influence function (UIF). In this case, $\tau = \mathbb{E}_p[\psi(W, \eta)]$ and $\sigma^2 = \mathrm{var}_p[\psi(W, \eta)]$.

It is important to note that the asymptotic linearity of $\hat{\tau}(\mathcal{D})$ depends on the function class used to fit η and on the rate at which the estimate of η converges: a rate slower than $\mathcal{O}(n^{-1/2})$ implies that $\hat{\tau}$ cannot be $\sqrt{n}$-consistent and, therefore, cannot be asymptotically linear. For example, if η is fit by a high-dimensional or nonparametric function class, one generally expects rate $\mathcal{O}(n^{-1/4})$. Recent work in double machine learning [2] has shown that Neyman orthogonality is a sufficient condition for asymptotic linearity of $\hat{\tau}$ even when η converges at a $\mathcal{O}(n^{-1/4})$ rate; these results permit us to include estimators with high-dimensional or nonparametric nuisance function classes in our framework.

The AIPW estimator has UIF given by $\psi(W, \eta) = \sum_x \lambda_x \left(\frac{I_x(X)}{e_x(\mathcal{Z})}\left(Y - \mu_x(\mathcal{Z})\right) + \mu_x(\mathcal{Z})\right)$ and satisfies Neyman orthogonality under mild conditions on e_x and μ_x. Jung et al. [12, 13] show that asymptotically linear estimators with UIFs exist for all identification formulas, and further they satisfy Neyman orthogonality. Hence, all identification formulas can be included in our framework even with nuisance functions that converge at a $\mathcal{O}(n^{-1/4})$ rate.

Multi-Armed Bandit Formalism. We now have all necessary elements to formalize the sequential decision making approach as a multi-armed bandit problem. For $k \in [K] := \{1, \ldots, K\}$, let $\hat{\tau}_k$ be an asymptotically linear estimator of τ with asymptotic variance σ_k^2 and cost c_k of observing $\mathcal{Z}_k, X$, and Y. Our goal is to identify the estimator with lowest cost-adjusted asymptotic variance, $k^* := \arg\min_k c_k \sigma_k^2$. This scaling arises because guaranteeing $|\hat{\tau}_k - \tau| = \mathcal{O}(\epsilon)$ with high probability requires $n = \mathcal{O}(\sigma_k^2/\epsilon^2)$ samples, which has a cost of $\mathcal{O}(c_k\sigma_k^2/\epsilon^2)$. For$\delta > 0$, and $\epsilon > 0$, our goal is to find an (ϵ, δ)-PAC index $\hat{k}$ which satisfies

$$\mathbb{P}\left(c_{\hat{k}}\sigma_{\hat{k}}^2 \geq \min_k c_k \sigma_k^2 + \epsilon\right) \leq \delta,$$

i.e.$\hat{k}$ has probability at least $1 - \delta$ of being at most ϵ-suboptimal.

We model our setting as a sequential decision problem where the investigator makes a sequence of decisions on rounds $n = 1, 2, \ldots$; at round n, the investigator chooses an index $k_n \in [K]$, obtains observation $w_{k_n}^n \sim p(\mathcal{Z}_k, X, Y)$, updates an the beliefs cost-adjusted asymptotic variances $c_k\sigma_k^2$, and decides whether a (ϵ, δ)-PAC index can be returned. If so, the algorithm terminates.

This sequential decision problem can be seen as a variant of *best-arm-identification* (BAI) in multi-armed bandits (MAB). MAB is a powerful framework for modeling sequential decision problems under uncertainty as a repeated game between an investigator and the environment. The standard BAI goal consists in identifying the arm with the smallest average loss in as few samples as possible. Our setting can be treated as a variant by replacing this goal with the goal of learning the estimator

with the lowest cost-adjusted asymptotic variance. This enables us to leverage algorithms from the BAI literature.

Most BAI algorithms fall into the following categories: (1) *action elimination algorithms*, such as *Successive Elimination* (SE) [3], which keep a set of arms that could be the optimal and alternate between sampling all the arms in this set and using confidence intervals to prune the set; (2) *optimistic algorithms*, including *Upper Confidence Bound* (UCB) and *LUCB* [14], which first construct the most optimistic problem instance that is consistent with the confidence intervals (*e.g.* by assuming the smallest mean values) and then act greedily as if this instance was true; (3) *Track-and-Stop-style algorithms* [16], which compute an asymptotic lower bound on the sample complexity and try to keep the empirical sampling proportion close to the sampling proportion achieving the lower bound.

Track-and-Stop-style algorithms require the sampling distribution of all the arms to be from a parametric family and therefore are not appropriate for our setting. Instead, action elimination and optimistic algorithms can be implemented whenever one has finite-sample confidence sets. We focus on LUCB and SE, as both algorithms can be implemented using arbitrary confidence sets (*i.e.* that are not a known function of n).

The next section is dedicated to developing the necessary tools for estimating two-sided confidence bounds on σ_k^2 that hold for all formulas and all rounds simultaneously. As the section concerns a single estimator, the k subscripts are dropped.

3 Finite-Sample Confidence Sequences for the Asymptotic Variance

This section introduces tools for constructing finite-sample confidence sequences on the asymptotic variance σ^2 of the estimator of τ which depend on assumptions on the distribution of W, smoothness properties of the influence function, and the rate at which $\hat{\eta}$ converges to η.

3.1 Confidence Sequences

A confidence sequences, defined for a stochastic process, essentially provides a confidence interval that holds uniformly over time.

Definition 2. *For a stochastic process $\{\xi_n\}_{n\geq 1} \in \mathbb{R}$ and coverage level $\alpha > 0$, a confidence sequence of level α is defined by a sequence of real numbers $\{u_n\}_{n\geq 1}$, referred to as a boundary sequence at level α, satisfying $\mathbb{P}(\forall n \geq 1 : \xi_n \leq u_n) > 1 - \alpha$.*

Instead of using a confidence sequence, one could control ξ_n by defining a confidence interval for every n and taking a union bound. However, to maintain a total probability of error of α, the confidence interval for ξ_n would need to hold with probability $O(\alpha/n^2)$, which would result in a confidence interval width that is a logarithmic factor larger [10].

3.2 Sample-Splitting Estimator of σ^2

For an estimator $\hat{\tau}$ of τ with IF ϕ, nuisance function η with estimator $\hat{\eta}$, and asymptotic variance σ^2, we propose and analyze the following sample-splitting estimator $\hat{\sigma}^2$ of σ^2 on dataset $\mathcal{D}$. We first divide $\mathcal{D}$ into three folds $\mathcal{D}^\eta$, $\mathcal{D}^\tau$ and $\mathcal{D}^\sigma$. We then use $\mathcal{D}^\eta$ to obtain an estimate $\hat{\eta}(\mathcal{D}^\eta)$ of η, compute $\hat{\tau}(\mathcal{D}^\tau, \hat{\eta}(\mathcal{D}^\eta))$ as an arbitrary solution in $\{\tau' : \mathbb{E}_{\mathcal{D}^\tau}[\phi(W, \hat{\eta}(\mathcal{D}^\eta), \tau')] = 0\}$, and evaluate

$$\hat{\sigma}^2(\mathcal{D}) := \mathbb{E}_{\mathcal{D}^\sigma}[\phi^2(W, \hat{\eta}(\mathcal{D}^\eta), \hat{\tau}(\mathcal{D}^\tau, \hat{\eta}(\mathcal{D}^\eta))]. \tag{1}$$

In the case where $\hat{\tau}$ has UIF ψ, the estimation procedure can be simplified by dividing $\mathcal{D}$ into two folds $\mathcal{D}^\eta$ and $\mathcal{D}^\sigma$ and computing

$$\hat{\sigma}^2(\mathcal{D}) := \mathrm{var}_{\mathcal{D}^\sigma}[\psi(W, \hat{\eta}(\mathcal{D}^\eta))] = \mathbb{E}_{\mathcal{D}^\sigma}[\phi^2(W, \hat{\eta}(\mathcal{D}^\eta), \hat{\tau}(\mathcal{D}^\sigma, \hat{\eta}(\mathcal{D}^\eta))], \tag{2}$$

where $\mathrm{var}_{\mathcal{D}^\sigma}$ indicates empirical variance w.r.t. $\mathcal{D}^\sigma$. In both cases, data splitting alleviates the bias in $\hat{\sigma}^2$ by forcing it to be independent of the bias in $\hat{\eta}$ (and of the bias of $\hat{\tau}$ in the IF case). The UIF case is more sample efficient because we do not need to directly compute $\hat{\tau}$.

3.3 Confidence Sequences for $|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2|$

We want to study the behavior of an estimator $\hat{\tau}$ evaluated on larger and larger datasets, as one would find in a bandit problem. We consider datasets $\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \ldots$, where $\mathcal{D}_n$ is obtained by augmenting

$\mathcal{D}_{n-1}$ with new samples. For example, in the UIF case in which $\mathcal{D}_n = \mathcal{D}_n^\eta \cup \mathcal{D}_n^\sigma$, $\mathcal{D}_{n-1}^\eta \subseteq \mathcal{D}_n^\eta$, and $\mathcal{D}_{n-1}^\sigma \subseteq \mathcal{D}_n^\sigma$. We wish to obtain a confidence sequence on the estimation error of the asymptotic variance, $|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2|$, in terms the estimation error of the nuisance function $\left\|\hat{\eta}^2(\mathcal{D}_n^\eta) - \eta\right\|$ and other problem parameters.

The main result of this section is that the estimation error of the sample-splitting estimators defined in Eqs. (1) and (2) only scales with $\mathcal{O}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|^2)$; this allows for $|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2| = \mathcal{O}(n^{-1/2})$ even when $\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\| = \mathcal{O}(n^{-1/4})$. We present the UIF case first.

Theorem 1. *Consider an L-Lipschitz UIF ψ, an upper bound $\tilde{\tau}$ on $|\tau|$, and let $\alpha > 0$. Let $\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \ldots$ be a sequence of datasets with $\mathcal{D}_n = \mathcal{D}_n^\eta \cup \mathcal{D}_n^\sigma$, and assume that there exists boundary sequences of level α $u_n^{(\psi,1)}$, $u_n^{(\psi,2)}$, and u_n^η for $\left|\mathbb{E}_{\mathcal{D}_n^\sigma}[\psi(W,\eta)] - \mathbb{E}[\psi(W,\eta)]\right|$, $\left|\mathbb{E}_{\mathcal{D}_n^\sigma}[\psi(W,\eta)]^2 - \mathbb{E}[\psi(W,\eta)]^2\right|$, and $\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|$, respectively. The estimator $\hat{\sigma}^2$ defined in Eq. (2) satisfies*

$$\mathbb{P}\left(\forall n \geq 1 : \left|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2\right| \leq 2L^2(u_n^\eta)^2 + u_n^{(\psi,2)} + \left(u_n^{(\psi,1)}\right)^2 + 2\tilde{\tau}u_n^{(\psi,1)}\right) \geq 1 - 3\alpha.$$

That is, a confidence sequence for $\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|$ and tail control of $\psi(W,\eta)$ are sufficient for control on $\left|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2\right|$. Note that we only require control of $\psi(W,\eta)$ at the true η. Also note that only $(u_n^\eta)^2$ appears, which justifies our claim of $\mathcal{O}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|^2)$ scaling. Computing this bound is an essential subroutine, outlined on the right.

CSUpdate (UIF version)

Input u_n^η, u_n^1, u_n^2, $\hat{\eta}$, L, $\tilde{\tau}$, $\mathcal{D}^\eta$, $\mathcal{D}^\sigma$, ψ
Compute $\hat{\eta}(\mathcal{D}^\eta)$
Compute $\hat{\sigma}^2 \leftarrow \mathrm{var}_{\mathcal{D}^\sigma}[\psi(W,\hat{\eta}(\mathcal{D}^\eta)]$
Using $\alpha = \delta/K$, $n_1 = |\mathcal{D}_k^\eta|$, $n_2 = |\mathcal{D}_k^\sigma|$,
$\beta \leftarrow 2L^2(u_{n_1}^\eta)^2 + u_{n_2}^2 + \left(u_{n_2}^1\right)^2 + 2\tilde{\tau}u_{n_2}^1$
Return $\hat{\sigma}^2$, β

Proof Outline. For this outline, let $\hat{\eta}_n := \hat{\eta}(\mathcal{D}_n^\eta)$, $\hat{\sigma}_n^2 := \hat{\sigma}^2(\mathcal{D}_n)$, and $\mathbb{E}_n := \mathbb{E}_{\mathcal{D}_n^\sigma}$. Using the identity $\mathrm{var}_n[\psi(W,\hat{\eta}_n)] = \mathbb{E}_n[\psi(W,\hat{\eta}_n)^2] - \mathbb{E}_n[\psi(W,\hat{\eta}_n)]^2$, we can expand $\hat{\sigma}_n^2 - \sigma^2$, as

$$\begin{aligned}\hat{\sigma}_n^2 - \sigma^2 = {} & \mathbb{E}_n\left[(\psi(W,\hat{\eta}_n) - \psi(W,\eta))^2\right] + 2\mathbb{E}_n\left[\psi(W,\eta)(\psi(W,\hat{\eta}_n) - \psi(W,\eta))\right] \\ & - \mathbb{E}_n[\psi(W,\hat{\eta}_n) - \psi(W,\eta)]\mathbb{E}_n[\psi(W,\hat{\eta}_n) + \psi(W,\eta)] \\ & + \left(\mathbb{E}_n[\psi^2(W,\eta)] - \mathbb{E}[\psi^2(W,\eta)]\right) + \left(\mathbb{E}_n[\psi(W,\eta)]^2 - \mathbb{E}[\psi(W,\eta)]^2\right).\end{aligned}$$

Since ψ is L-Lipschitz, the first term can be bounded by $L^2\|\hat{\eta}_n - \eta\|^2$. The second and third terms can be simplified with Cauchy-Schwarz, the first order terms cancel out, and the result is in another $L^2\|\hat{\eta}_n - \eta\|^2$ term. The forth term is controlled by $u_n^{(\psi,2)}$, and the final term can bounded as

$$\begin{aligned}\left|\mathbb{E}_n[\psi(W,\eta)]^2 - \mathbb{E}[\psi(W,\eta)]^2\right| &\leq \left|\left(\mathbb{E}[\psi(W,\eta)] - u_n^{(\psi,1)}\right)^2 - \mathbb{E}[\psi(W,\eta)]^2\right| \\ &\leq \left(u_n^{(\psi,1)}\right)^2 + 2\left|u_n^1\mathbb{E}[\psi(W,\eta)]\right| \leq (u_n^{(\psi,1)})^2 + 2\tilde{\tau}u_n^{(\psi,1)}.\end{aligned}$$

□

A similar result can be obtained without an UIF as long as one can obtain an additional confidence sequence on $\hat{\tau}(D_n^\tau, \hat{\eta}(D_n^\eta))$. See Appendix F for a precise statement and the proof. In this case, the CSUpdate subroutine needs to be modified to use a third data fold $\mathcal{D}_n^\tau$ and the appropriate $\hat{\sigma}^2$ and confidence sequence from Theorem 2.

Theorem 2. *Consider an L-Lipschitz IF ϕ, an upper bound $\tilde{\tau}$ on $|\tau|$, and let $\alpha > 0$. Let $\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \ldots$ be a sequence of datasets with $\mathcal{D}_n = \mathcal{D}_n^\eta \cup \mathcal{D}_n^\tau \cup \mathcal{D}_n^\sigma$, and assume that there exists boundary sequences of level α as in Theorem 1 with ϕ replacing ψ throughout, and that u_n^τ is a boundary sequences of level α for $|\hat{\tau}(\mathcal{D}_n^\tau, \hat{\eta}(\mathcal{D}_n^\eta)) - \eta|$. The estimator $\hat{\sigma}^2$ defined in Eq. (1) satisfies*

$$\mathbb{P}\left(\forall n \geq 1 : \left|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2\right| \leq 2L^2(u_n^\eta + u_n^\tau)^2 + u_n^{(\phi,2)} + (u_n^{(\phi,1)})^2 + 2\tilde{\tau}u_n^{(\phi,1)}\right) \geq 1 - 4\alpha.$$

Combined with a result that $|\hat{\tau}(\mathcal{D}_n^\tau, \hat{\eta}(\mathcal{D}_n^\eta)) - \tau| = \mathcal{O}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|^2)$ (see e.g. [2, 4]), we have that $|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2| = \mathcal{O}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|^2)$.

We emphasize that the $\mathcal{O}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|^2)$ scaling is surprising, as a naïve argument would produce a dependence on $\mathcal{O}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|)$ instead, as

$$\begin{aligned}\hat{\sigma}^2 - \sigma^2 &= \mathbb{E}_n[\phi^2(W, \hat{\eta}, \hat{\tau})] - \mathbb{E}[\phi^2(W, \eta, \tau)] \\ &= \mathbb{E}_n\left[(\phi(W, \hat{\eta}, \hat{\tau}) - \phi(W, \eta, \tau))^2 + 2\phi(W, \eta, \tau)(\phi(W, \hat{\eta}, \hat{\tau}) - \phi(W, \eta, \tau))\right] \\ &\quad + \mathbb{E}_n[\phi^2(W, \eta, \tau)] - \mathbb{E}[\phi^2(W, \eta, \tau)],\end{aligned}$$

and the second term is $\mathcal{O}(\|\hat{\eta} - \eta\|)$ under a simple bound (*e.g.* Cauchy-Schwarz).

3.4 Sub-Gaussian Random Variables

Confidence sequences for empirical expectations around their means are well established in the literature, see e.g. [7]. As an illustration, this section derives a confidence sequence for the common assumption that the variables are sub-Gaussian.

Definition 3. *A random variable W is λ sub-Gaussian if there exists a constant λ such that $\mathbb{E}\left[e^{t(W - \mathbb{E}[W])}\right] \leq e^{\lambda \frac{t^2}{2}}$ and all $t \in \mathbb{R}$. A random variable W is ν sub-exponential with scale c if there exist constants ν, c such that $\mathbb{E}\left[e^{t(W - \mathbb{E}[W])}\right] \leq e^{\nu \frac{t^2}{2}}$ $\forall t \in [0, 1/c)$.*

Lemma 1. *If W is λ sub-Gaussian, then W^2 is $8\lambda^2$ sub-exponential with scale $c = 2\lambda$.*

Intuitively, sub-Gaussian random variables have tails no heavier than a Gaussian with variance λ, and sub-exponential random variables have tails no heavier than a χ^2 distribution. Sub-Gaussianity is a common assumption in the bandit literature satisfied by many applications; for example, a B-bounded random variable is B^2 sub-Gaussian.

Confidence sequences for sub-Gaussian and and sub-Exponential random variables can be found in the literature, see e.g. [7]. Then, we may use the fact that $\left|\mathbb{E}_{\mathcal{D}_n^\sigma}[\psi(W, \eta)] - \mathbb{E}[\psi(W, \eta)]\right|$ is sub-Gaussian and and $\left|\mathbb{E}_{\mathcal{D}_n^\sigma}[\psi(W, \eta)]^2 - \mathbb{E}[\psi(W, \eta)]^2\right|$ is sub-Exponential to derive the following bound for our variance estimation error. See Appendix C for more details.

Corollary 1. *Let $\alpha \in (0, 1)$ and assume the same setting as Theorem 1, and additionally that $\psi(W, \eta)$ is λ sub-Gaussian. Then, with a certain choice of parameters, the confidence sequence of Lemma 5 guarantees that, for any $m > 0$ and $n \geq (18.6\lambda \log(\lambda n/m) + \log(2/\alpha)) \vee m/\lambda'$,*

$$\mathbb{P}\left(\exists n \geq 1 : \left|\hat{\sigma}^2(\mathcal{D}_n) - \sigma^2\right| \geq 2L^2(u_n^\eta)^2 + \frac{5\left(\sqrt{2\lambda} + \tilde{\tau}\right)}{8}\sqrt{\frac{1}{n}\left(2\lambda \log(\lambda' n/m) + \log\frac{2}{\alpha}\right)}\right) \leq \alpha.$$

3.5 Confidence Sequences for $\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|$

Our main theorem presents a confidence sequence on the asymptotic variance in terms of the nuisance function estimation error, $\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|$. There are many estimators $\hat{\eta}$ that have confidence sequences established in the literature, such as least squares estimators [28]. However, for estimators with a confidence interval but no confidence sequence, we can always take a union bound over n.

To make this intuition precise, suppose that we have a function $R(n, \alpha)$ such that, for any $n > 1$ and $\alpha \in (0, 1)$, $\mathbb{P}(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\| \geq R(n, \alpha)) \leq \alpha$. Then $u_n^\eta = R(n, 6\alpha/(\pi n)^2)$ is a boundary sequence of level α, verified by $\mathbb{P}\left(\exists n > 1 : \|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\| \geq u_n^\eta\right) \leq \sum_{n=1}^\infty \mathbb{P}\left(\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\| \geq R(n, 6\alpha/(\pi n)^2)\right) \leq \sum_{n=1}^\infty \frac{6\alpha}{\pi n^2} \leq \alpha$.

Hence, without loss of generality, our algorithms and results can be stated in terms of a boundary function u_n^η. Typically, if $R(n, \alpha) = \mathcal{O}(n^\nu \log(1/\alpha))$, the above union bound construction only adds log terms to the final bound, which is generally sufficient for most applications.

4 CS-LUCB and CS-SE Algorithms

This section introduces adaptations of LUCB and SE to use the confidence sequences derived in Section 3, referred to as CS-LUCB and CS-SE. For ease of exposition, we present the UIF case (the IF case required keeping a third data fold and modify $\mathrm{CSUpdate}$ as described in Section 3.3).

CS-LUCB and CS-SE are summarized in Algorithms 1 and 2. These algorithms take as input, for $k \in [K]$, the UIF ψ_k, the estimator $\hat{\eta}_k$ of the nuisance function η_k, the cost c_k, and the boundary sequences $u_k := (u^{\eta}_{k,n},\ u^{1}_{k,n}, u^{2}_{k,n})$ on $\|\hat{\eta}_k - \eta_k\|$, $\psi_k(W_k, \eta_k)$, and $\psi_k(W_k, \eta_k)^2$, respectively.

Every round, CS-LUCB samples observations for the estimator with the lowest cost-adjusted asymptotic variance and for the estimator with the lowest lower confidence bound in the remaining estimators. Intuitively, samples from these estimators are the most informative. CS-LUCB continues until the confidence sequences of the best estimator is separated from the rest, up to the error tolerance.

CS-SE keeps a set S of plausibly best estimators. At each round, it obtains samples for every estimator in S, updates the confidence sequences, then removes all estimators with lower bounds higher than the upper bound of the best (as there estimators are no longer plausible the best) from S. The algorithm terminates if either all non-optimal estimators are removed or all remaining estimators are within ϵ of each other.

In the remainder of the section, we prove that both algorithms return (ϵ, δ)-PAC indices if the confidence sequence width approaches zero and provide upper bounds on the sample complexities when an upper bound on the width is available.

We define the gap of estimator k to be $\Delta_k := c_k\sigma_k^2 - \min_{k'} c_{k'}\sigma_{k'}^2$. The random variable $\beta_k(n)$ is the confidence width returned by CSUpdate for estimator k after n updates, which might be random. We assume that $\beta_k(n)$ is independent of $\beta_j(m)$ for all $j \neq k$.

Algorithm 1 CS-LUCB

Input $\epsilon > 0$, $\delta > 0$, $\Delta_n > 1$,$\tilde{\tau} > 0$
$\{\psi_k, \hat{\eta}_k, u_k, c_k : k \in [K]\}$
for *k=1,…, K* **do**
  Obtain Δ_n new samples $\mathcal{D}$
  Add half of $\mathcal{D}$ to $\mathcal{D}_k^{\eta}$ and half to $\mathcal{D}_k^{\sigma}$
  $\hat{\sigma}_k^2, \beta_k \leftarrow \mathrm{CSUpdate}(u_k, \hat{\eta}_k, \psi_k, L, \tilde{\tau}, \mathcal{D}_k^{\eta}, \mathcal{D}_k^{\sigma})$
end
for $t = 1, 2, \ldots$ **do**
  $l_t \leftarrow \arg\min_{k\in[K]} c_k\hat{\sigma}_k^2$
  $u_t \leftarrow \arg\min_{k\neq l_t} c_k(\hat{\sigma}_k^2 - \beta_k)$
  if $c_{l_t}\left(\hat{\sigma}_{l_t}^2 + \beta_{l_t}\right) \leq c_{u_t}\left(\hat{\sigma}_{u_t}^2 - \beta_{u_t}\right) - \epsilon$ **then**
    Return $\hat{k} = l_t$
  end
  for $k \in u_t, l_t$ **do**
    Obtain Δ_n new samples $\mathcal{D}$
    Add half of $\mathcal{D}$ to $\mathcal{D}_k^{\eta}$ and half to $\mathcal{D}_k^{\sigma}$
    $\hat{\sigma}_k^2, \beta_k \leftarrow \mathrm{CSUpdate}(u_k, \hat{\eta}_k, \psi_k, L, \tilde{\tau}, \mathcal{D}_k^{\eta}, \mathcal{D}_k^{\sigma})$
  end
end

Algorithm 2 CS-SE

Input $\delta > 0$, $\epsilon > 0$, $\Delta_n > 1$, $\tilde{\tau} > 0$
$\{\psi_k, \hat{\eta}_k, u_k, c_k, : k \in [K]\}$
$S \leftarrow [K]$, and $\mathcal{D}_k^{\eta} \leftarrow \emptyset, \mathcal{D}_k^{\sigma} \leftarrow \emptyset\ \forall k \in [K]$
while $|S| > 1$ **do**
  for $k \in S$ **do**
    Obtain Δ_n new samples $\mathcal{D}_k^{\Delta}$
    Add half of $\mathcal{D}_k^{\Delta}$ to $\mathcal{D}_k^{\eta}$ and half to $\mathcal{D}_k^{\sigma}$
    $\hat{\sigma}_k^2, \beta_k \leftarrow \mathrm{CSUpdate}(u_k, \hat{\eta}_k, \psi_k, L, \tilde{\tau}, \mathcal{D}_k^{\eta}, \mathcal{D}_k^{\sigma})$
  end
  $l_t \leftarrow \arg\min_{k\in S} c_k\hat{\sigma}_k^2$
  $R \leftarrow \{k \in S : c_{k^*}(\hat{\sigma}_{l_t}^2 + \beta_{l_t}) \leq c_k(\hat{\sigma}_k^2 - \beta_k)\}$
  $S \leftarrow S \setminus R$
  if $\beta_k \leq \frac{\epsilon}{2}$ *for all* $k \in S$ **then**
    $S \leftarrow \arg\min_{k\in S} c_k\hat{\sigma}_k^2$
  end
end
Return $\hat{k} = S$

Theorem 3. *Assume that the conditions of Theorem 1 hold and that* $u^{\eta}_{k,n}, u^{(\psi,1)}_{k,n}, u^{(\psi,2)}_{k,n} \to 0$ *for all* $k \in [K]$. *Then both CS-LUCB and CS-SE run with* $u_k = (u^{\eta}_{k,n}, u^{(\psi,1)}_{k,n}, u^{(\psi,2)}_{k,n})$ *and* $\alpha = \epsilon/3$ *return an* (ϵ, δ)*-PAC estimator.*

If we have a deterministic upper bound $B_k(n, \delta)$ *such that, for all* $\delta > 0$, $\mathbb{P}(\beta_k(n) \leq B_k(n, \delta)) \geq 1 - \delta$, *then both algorithms terminate in at most* $\sum_{k\in[K]}$ *min* $\left\{n : B_k(n, \delta/K) \leq \frac{\Delta_k}{4} \vee \frac{\epsilon}{2}\right\}$ *samples.*

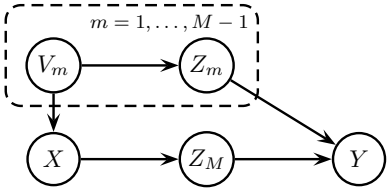
If, additionally, there exists constants ν_η, $\nu_{(\psi,1)}$, *and* $\nu_{\psi,2)}$ *such that* $u^{\theta}_{k,n} \leq \mathcal{O}(n^{-\nu_\theta}\log(nK/\delta))$ *for all* $\theta \in \{\eta, (\psi, 1), (\psi, 2)\}$ *and all* $k \in [K]$, *then the sample complexity is*

$$\mathcal{O}\left(\sum_{k=1}^{K}(\Delta_k \vee \epsilon)^{-1/\nu}\left(\log\frac{K}{\delta(\Delta_k \vee \epsilon)^{1/\nu}}\right)^{1/\nu}\right),$$

with probability at least $1-\delta$, where $\nu = min\{2\nu_\eta, \nu_{(\psi,1)}, \nu_{(\psi,2)}\}$. In particular, if $\psi(W,\eta)$ is sub-Gaussian, we recover the sample complexity results (up to log factors) of [3, 14] under the mild condition of $\nu_\eta \geq 1/4$.

Discussion. To aid comparison with the BAI literature, we have provided sample complexity bounds for the special case in which $u^\theta_{k,n} = \mathcal{O}(\sqrt{n^{-\nu\theta}\log(n/\delta)})$. The most common choice in the bandit literature assumes that the data is sub-Gaussian, meaning that we may take $\nu_\theta = \frac{1}{2}$. In this case, our sample complexity results are within a log-factor of the optimal, and they are within a log-factor of the lower bound when $\epsilon = 0$. Algorithms achieving this optimal rate include lil'UCB [11] and Exponential Gap Elimination [15]. Unfortunately, these algorithms require the widths of the confidence regions for the arm mean estimates to decrease at the same rate, which is not an assumption we can make in our setting (as the rates in our setting depend on properties of the estimator and its influence function).

5 Experiments



We study the empirical sample complexities of the CS-LUCB and CS-SE algorithms w.r.t. uniform sampling for the task of selecting an identification formula for the causal DAG on the right, where the plate notation indicates that the path $X \leftarrow V_m \rightarrow Z_m \rightarrow Y$ is repeated $M-1$ times. Specifically, we consider the 2^{M-1} identification formulas obtained from the adjustment criterion using either V_m or Z_m, for all $m \in [M-1]$, and the identification formula obtained from the frontdoor criterion using Z_M, giving a total of $2^{M-1}+1$ different formulas to choose from.

Graphical criteria suggest that, when not considering costs, using Z_m instead of V_m yields an adjustment formula with lower asymptotic variance (see Appendix A.2). Therefore, we assumed costs of including Z_m or V_m equal to 3 and 1, respectively (and no cost for X and Y). These values were chosen to reflect a setting in which post-exposure variables and variables further from X are harder to measure.

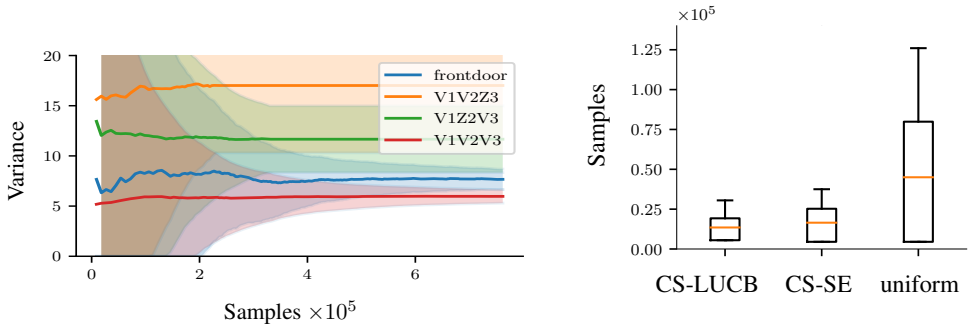
We assumed binary exposure X, and causal contrast $\tau := \sum_x \lambda_x \mu_{\text{do}(x)}$ with $\lambda_0 = -1, \lambda_1 = 1$, corresponding to the average treatment effect. The adjustment formulas were estimated with the AIPW estimator, and the frontdoor formula with a Neyman orthogonal estimator derived from Fulcher et al. [5, Theorem 1] (see Appendix A.3 for details).

5.1 Linear Model

In our first experiment, we considered observational data generated by the following linear model. For $m \in [M-1]$, $V_m \sim \mathcal{N}(0, I_2)$, $Z_m = A_m V_m + \epsilon_{z,m}$ for matrix $A_m \in \mathbb{R}^{3\times 2}$ and $\epsilon_{z,m} \sim \mathcal{N}(0, .1I_3)$, $X \sim \text{Bernoulli}\left(1\Big/\left(1+e^{-\sum_{m=1}^{M-1} B_m^\top V_m}\right)\right)$ for vectors $B_m \in \mathbb{R}^3$, and Z_M sampled from a categorical distribution conditioned on X. Specifically, for support points $\mathcal{S} := \{s_i \in \mathbb{R}^2 : i \in [10]\}$ and vectors q_1 and q_0 in the 9-simplex, we set $p(Z_M = s_j | X = i) = q_i(j)$, where $q_i(j)$ in the jth element of q_i. Finally, $Y = \sum_{m=1}^M C_m^\top Z_m + \epsilon_y$ for vector $C_m \in \mathbb{R}^2$ and $\epsilon_y \sim \mathcal{N}(0, .1)$.

Specific observational distributions were obtained by sampling each element of A_m, B_m, and C_m, from $\mathcal{N}(0, U_m^2/4)$, $\mathcal{N}(0, U_m^2)$, and $\mathcal{N}(0, (2-U_m)^2)$, respectively, with $U_m \sim \text{Uniform}[.1, .9]$. The purpose of this sampling scheme was to produce p where the correlation between V_m and X and the correlation between Z_m and Y are not both large. This choice widens the gaps to the best estimator and results in a more interesting bandit problem. B_M had elements independently sampled from $\mathcal{N}(0, 1/4)$.

To generate the distribution of Z_M, we sampled each support point of $\mathcal{S}$ from a standard 2-dimensional normal, sampled $G_1, G_0 \sim \mathcal{N}(0, I_{10})$, then computed $q_1 = \text{softmax}(.4G_1)$ and $q_0 = \text{softmax}(.4G_0)$. We then computed $\tau = \sum_{i=1}^{10}(q_1(i) - q_0(i))B_M^\top s_i$ and resampled $\mathcal{S}$, q_0, and q_1 until τ was larger than 1, which avoided the hardest instances and allowed us to run more simulations. As harder instances increase the sample complexity of uniform sampling the most, this resampling step does not inflate the advantages of CS-LUCB or CS-SE.



(a) (b)

Figure 1: (a): Example of confidence sequences corresponding to the four best estimators for CS-SE, up to the $8.1 * 10^5$ samples required to find the optimal estimator. (b): Box plot comparing sample complexities of CS-LUCB, CS-SE, and uniform sampling.

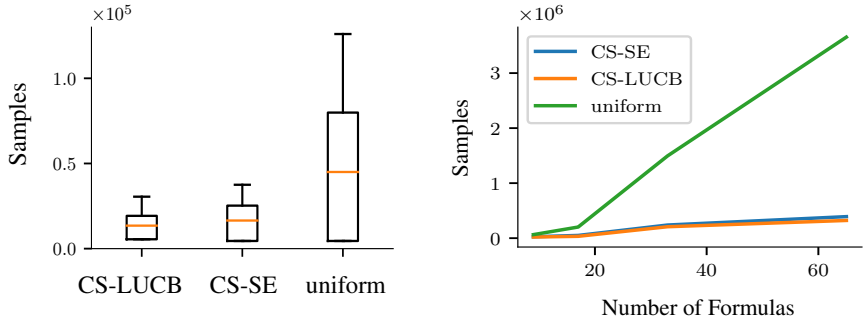
The nuisance function $(\mu_x(\mathcal{Z}) := \mathbb{E}_p[Y|X = x, \mathcal{Z}], e_x(\mathcal{Z}) = p(X|\mathcal{Z}))$ of the AIPW estimator was fit with with ridge regression and logistic regression, respectively. We used known confidence sequences for linear regression (see Abbasi-Yadkori et al. [1, Theorem 2] or Lemma 7 in Appendix D), and confidence intervals for logistic regression were approximated by a standard central limit theorem (CLT) confidence interval, $\left[\hat{\sigma}^2 + z_{\alpha_n/2}\frac{std(\hat{\sigma}^2)}{\sqrt{n}}, \hat{\sigma}^2 + z_{1-\alpha_n/2}\frac{std(\hat{\sigma}^2)}{\sqrt{n}}\right]$, with $\delta = 0.1$ and $\alpha_n = 6\delta/\pi n^2$, following construction described in Section 3.5. We used the estimator in Appendix A.3 for the front-door criterion; the nuisance function $(\mathbb{E}_p[Y|Z = z, X = x], p(X = x), p(X = x|Z = z))$ is fit using ridge regression for the first term and the MLE for the others (which corresponds to the empirical counts since the random variables are categorical).

Results. We considered $M = 3$, obtaining 8 adjustment formulas and a frontdoor formula. We generated 10 observational distributions and, for each distribution, ran the algorithms 5 times from scratch on independent data. A typical run of CS-SE is shown in Fig. 1(a): confidence sequences of the four best formulas are plotted, and one can identify visually when sub-optimal estimators are removed from $\mathcal{S}$. The algorithm terminates as soon as the two lowest confidence sequences no longer intersect. In Fig. 1(b), we provide a box plot for the distribution of sample complexities for the three algorithms. The orange line, box, and whiskers indicates the median, the IQR (the 25th percentile to the 75th percentile), and the outlier-filtered range (points with values $1.5\times$IQR greater than the 75th percentile are removed), respectively. We see that some instances can be difficult for all algorithms (if, for example, many of the gaps are small), but CS-LUCB and CS-SE display substantial reduction in sample complexity w.r.t. the uniform sampling algorithm, on average needing 34% and 51% the number of samples to terminate.

5.2 Nonlinear Model

A key result from Theorems 1 and 2 is the second-order dependence on the rate of η estimation, which allows the use of high-dimensional, nonparametric estimators for η. To showcase this, we considered a more complex non-linear generation process. Given non-linear functions f_m, g_m, h_m for $m \in [M-1]$ and g_M, we generated observations as in Section 5.1, except with $X \sim \text{Bernoulli}\left(1\Big/\left(1 + e^{-\sum_{m=1}^{M-1} h_m(V_m)}\right)\right)$, $Z_m = f_m(V_m) + \epsilon_{z,m}$, and $Y = \sum_{m=1}^{M} g_m(Z_m) + \epsilon_y$. We sampled the nonlinear functions to be element-wise from a Gaussian process prior with RBF kernel on 10^d equidistant points in $(-2, 2)^d$, where d is the dimension of the domain. We fit all the non-linear functions with gradient-boosted regression trees with the same kernel. The confidence sequences for the nuisance functions were approximated using the same CLT-based construction used for logistic regression.

Results. Similar to the linear case, we generated 10 random observational distributions with 9 formulas and, for each distribution, ran each algorithm 5 times from scratch on independent data. We



(a) (b)

Figure 2: (a) Box plot of sample complexities with 9 formulas. (b) Sample complexities as a function of the number of formulas

found a similar reduction in sample complexity as in linear setting (see Fig. 2(a) for another box plot). We also explored how the sample complexity changes with the number of formulas. We sampled an observational distribution with 3, 4, 5, and 6 $X \leftarrow V_m \rightarrow Z_m \rightarrow Y$ paths, corresponding to 7, 15, 31, and 63 formulas; for each distribution, we ran the algorithms from scratch 4 times on independent data and plotted the average sample complexities in Fig. 2(b). Our theory from Section 4 predicts that the sample complexity is determined not by the number of estimators but by the reciprocals of the squared gaps, which increases sub-linearly with the number of estimators as many of them will have large gaps. In contrast, the sample complexity of the uniform algorithm should increase linearly. We see this prediction confirmed by our experiments.

6 Discussion

Much of the literature on causal inference from observational data has focused either on *identifying* causal effects using structural knowledge of the causal graph underlying the data generation mechanism or on the design or selection of sample efficient *estimators*. The problem of selecting an identification formula using the efficiency of a corresponding estimator is only starting to receive attention.

Despite the recent progress of graphical criteria, comparing arbitrary formulas with statistical and practical considerations, such as cost or impossibility to observe certain covariates, remains an open problem. This work attempts to provide a first functional answer to this problem: instead of trying to derive a solution from graph properties, we have provided a practical method that uses observational data to select a formula. When the graphical criteria do not apply, such as when the graph is partially unknown, or latent variables exist, or when costs need to be considered, our methods can help the investigator to reach a conclusion in a sample efficient way.

Our methods have the limitations of only guaranteeing to find the asymptotically optimal formula and relying on the availability of influence functions to quantify the asymptotic variance. To the best of our knowledge, more data-driven approaches to estimating the variance of estimators, such as resampling methods, do not have any finite-sample guarantees and are not appropriate for a bandit algorithm. Still, selecting between more general classes of estimators is an interesting direction.

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

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

[3] Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(6), 2006.

[4] Dylan J. Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *arXiv preprint arXiv:1901.09036*, 2019.

[5] Isabel R. Fulcher, Ilya Shpitser, Stella Marealle, and Eric J. Tchetgen. Robust inference on population indirect causal effects: the generalized front door criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):199–214, 2020.

[6] Leonard Henckel, Emilija Perković, and Marloes H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv preprint arXiv:1907.02435*, 2019.

[7] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *arXiv preprint arXiv:1810.08240*, 2018.

[8] Hidehiko Ichimura and Whitney K. Newey. The influence function of semiparametric estimators. *arXiv preprint arXiv:1508.01378*, 2015.

[9] Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014.

[10] Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *Artificial Intelligence and Statistics*, pages 240–248, 2016.

[11] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

[12] Yonghan Jung, Jin Tian, and Elias Bareinboim. Estimating identifiable causal effects through double machine learning. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 2021.

[13] Yonghan Jung, Jin Tian, and Elias Bareinboim. Estimating identifiable causal effects on Markov equivalence class through double machine learning. Technical Report R-71, Causal Artificial Intelligence Lab, Columbia University, 2021.

[14] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning*, pages 227–234, 2012.

[15] Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*, pages III–1238–III–1246, 2013.

[16] Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Conference on Learning Theory*, pages 228–251, 2013.

[17] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

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

[19] Whitney K. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62 (6):1349–1382, 1994.

[20] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[21] Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell. *Causal Inference in Statistics: A Primer*. John Wiley & Sons, 2016.

[22] Emilija Perkovic, Markus Kalisch, and H. Marloes Maathuis. Interpreting and using CPDAGs with background knowledge. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*, 2017.

[23] Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H. Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18(220):1–62, 2018.

[24] Andrea Rotnitzky and Ezequiel Smucler. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *Journal of Machine Learning Research*, 21 (188):1–86, 2020.

[25] Ilya Shpitser, Tyler VanderWeele, and James M. Robins. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, 2010.

[26] Ezequiel Smucler, F Sapienza, and Andrea Rotnitzky. Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika*, 2021.

[27] Aad W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.

[28] H Victor, Michael J Klass, and Tze Leung Lai. Theory and applications of multivariate self-normalized processes. *Stochastic Processes and their Applications*, 119(12):4210–4227, 2009.

[29] Martin J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press, 2019.

[30] Janine Witte, Leonard Henckel, Marloes H. Maathuis, and Vanessa Didelez. On efficient adjustment in causal graphs. *Journal of Machine Learning Research*, 21(246):1–45, 2020.