Asymptotically Best Casual Effect Identification with Multi-Armed Bandits

Anonymous Author(s)
Affiliation
Address
email

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 sample complexity bounds and an empirical study on artificially generated data.

1 Introduction

Many scientific disciplines ranging from epidemiology to biology, from healthcare to social and behavioral sciences are concerned with estimating the *causal effect* of some exposure X on an outcome of interest Y. Often, performing interventions on X to look at their effects on Y is unfeasible. This could be the case, for example, when trying to assess impact of education on income: an ideal experiment for answering this question would require randomly selecting the amount of education students acquire, which would not be possible in most scenarios. Instead, the investigator might have access to *observational data* potentially containing spurious correlation between X and Y that confounds the causal effect. The causal effect identification literature [21] provides conditions under which a causal effect can be inferred from observational data, as well as *identification formulas* that express the effect as a functional of the observational data distribution p. For example, given a causal DAG $\mathcal{G}$ underlying the data generation process, the *adjustment criterion* [22, 23, 25] states that, for any set of covariates $\mathcal{Z}$ that block all non-causal paths from X to Y and that are non descendants of nodes on causal paths from X to Y in $\mathcal{G}$, the causal effect of X on Y, $p(Y|\text{do}(X = x))$[1], can be expressed as $\sum_z p(Y = y|X = x, \mathcal{Z} = z)p(\mathcal{Z} = z)$.

In this work, we consider an over-identified setting where there exist many identification formulas that can be used to estimate a causal effect (*e.g.* many covariate adjustment sets), and the investigator would like to select the "best" formula in terms of performance of the associated estimator, as well as practical considerations like the relative cost of observing covariates. We assume an online setting in which the investigator collects information in a data-dependent way with the aim of identifying the best formula in as few samples as possible. Specifically, we assume that the investigator proceeds in

[1] $p(Y|\text{do}(X = x))$ indicates the distribution of Y when we intervene on X by fixing its value to x. This is different from $p(Y|X = x)$ where we only narrow the focus on datapoints for which X takes value x.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

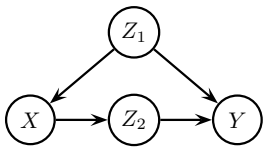
rounds, each consisting in choosing a formula, observing the corresponding covariates, and updating its belief based on the acquired information.

We formalize this setting by using a best-arm-identification bandit framework in which each arm corresponds to an estimator and the standard goal of learning the arm with the lowest loss is replaced with that of learning the estimator that will perform best in the sample-plentiful regime. As measure of estimator performance we use the *asymptotic variance* adjusted with a cost of observing the corresponding covariates. The asymptotic variance is the most prevalent metric for quantifying long-run behavior of estimators [20], and has seen use specifically in the comparison of covariate adjustment sets. The asymptotic variance is also the scale of the leading term of the asymptotic expansion of any asymptotically linear estimator—recent work has shown that a surprisingly large number of estimators, even those with complex nuisance functions, fall into this class [2].

Previous selection approaches focused on either the batch setting (where a single decision is made from a fixed batch of data) or on using the structure of the casual graph underlying the data generation mechanism [6, 7, 16–19, 24, 26, 29]. Recently, Henckel et al. [7], Rotnitzky and Smucler [24], Smucler et al. [26], Witte et al. [29] introduced graphical criteria that enable the comparison of certain sets of adjustment covariates and the identification of asymptotically optimal sets for linear causal models with ordinary least squares estimators and non-linear causal models with non-parametric estimators. Whilst this work represents significant progress with respect to previous methods, selection based on graphical criteria is inherently incomplete, as comparison of certain sets cannot be achieved with information about the graph structure alone. Furthermore, such criteria currently do not extend to formulas more complicated than the adjustment criterion. In order to compare arbitrary formulas, to account for practical constraints such as cost or impossibility to observe certain covariates, or to avoid the requirement of structural knowledge of the causal graph, the asymptotic variance must be estimated directly. This work proposes a best-arm-identification bandit method for achieving this goal in a sample efficient way.

Our contributions are threefold. First, we pose the sequential decision problem of how to choose which observations to collect in order to learn the formula with the best asymptotic variance in the most sample efficient way. Second, we propose framing this problem as a best-arm-identification multi-armed bandits problem where the goal is to find the lowest asymptotic variance instead of the arm with the highest mean. We leverage ideas from semi-parametric estimation theory, specifically the connection between influence functions and asymptotic variance, to frame our bandit problem concretely, as detailed in Section 2. Third, in Section 3 we develop novel finite-sample confidence sequences with nuisance functions for a sample-splitting estimator of the asymptotic variance, establishing that having an uncentered influence function allows for confidence sequences that are insensitive to errors of the nuisance function estimates. We adapt the existing bandit algorithms LUCB and successive elimination to our setting and our variance estimators, and exhibit sample complexity bounds in terms of the nuisance function estimation rate. Finally, in Section 5 we empirically validate our methods on artificially generated data, showing significant sample complexity reduction with respect to a naive uniform sampling method.

2 Causal Effect Identification as a Bandit Problem



Problem Setting. We consider the problem of an investigator wishing to compute a causal contrast $\tau := \sum_x \lambda_x \mu_{\text{do}(x)}$, where $\mu_{\text{do}(x)} := \mathbb{E}[Y|\text{do}(X = x)]$, in a setting in which K different formulas are available to express the contrast as a functional of the observational distribution p. For example, in the causal DAG in the left figure, $\mu_{\text{do}(x)}$ can be expressed as a functional of p using the adjustment criterion, *i.e.* $\mu_{\text{do}(x)} = \mathbb{E}_p[\mu_x(Z_1)]$ with $\mu_x(Z_1) := \mathbb{E}_p[Y|X = x, Z_1]$, the *frontdoor criterion*, *i.e.* $\mu_{\text{do}(x)} = \sum_{z_2} p(Z_2 = z_2|X = x) \sum_{x'} p(X = x')\mu_{x'}(Z_2 = z_2)$, or a formula that uses both Z_1 and Z_2. We assume that formula k is associated to an estimator $\hat{\tau}_k$ of τ, *i.e.* a mapping $\mathcal{D}_k = \{w_k^i\}_{i=1}^n \to \mathbb{R}$, where $w_k^i \sim p(W_k)$, $W_k := (\mathcal{Z}_k, X, Y)$, and $\mathcal{Z}_k$ is the set of covariates required by the formula. The investigator would like to understand which formula leads to the estimator with best performance in the asymptotic regime—as measured by the asymptotic variance—by collecting information in a data-dependent way, using as few samples as possible and potentially taking into account the relative cost of observing the associated covariates.

Assumptions on Estimators of τ. Many estimators for causal effect identification formulas have been developed over the years that can work for specific settings and address model miss-specifications. A common feature shared across causal effect estimators is the requirement to estimate a *nuisance function* η, *i.e.* a function that is only used as a means to estimate τ. For example, the *augmented inverse probability weighted estimator* (AIPWE), which is a popular estimator for the adjustment criterion, is defined as $\hat{\tau}(\mathcal{D}) = \mathbb{E}_n\left[\sum_x \lambda_x \left(\frac{I_x(X)}{\hat{e}_x(\mathcal{Z})}\left(Y - \hat{\mu}_x(\mathcal{Z})\right) + \hat{\mu}_x(\mathcal{Z})\right)\right]$, where $\mathbb{E}_n[\cdot]$ denotes empirical expectation, I_x the indicator function, and $\hat{\mu}_x, \hat{e}_x$ estimators of the nuisance function $\eta = (\mu_x, e_x)$ where $e_x(\mathcal{Z}) := p(X = x|\mathcal{Z})$.

Asymptotic Linearity. We only consider estimators with the desirable property of *asymptotic linearity* [9, 27], meaning that there exists a function ϕ, called *influence function*, 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 $\sigma^2 = \mathbb{E}_p[\phi^2(W, \eta, \tau)]$. Focusing on asymptotically linear estimators ensures that we only consider estimators that converge at a $\sqrt{n}$ rate.

Uncentered Influence Function. We also only consider estimators whose influence function can be decomposed as $\phi(W, \eta, \tau) = \psi(W, \eta) - \tau$, where ψ is called *uncentered influence function (UIF)* (this is the case for the AIPWE). In this case, $\tau = \mathbb{E}_p[\psi(W, \eta)]$ and the asymptotic variance can be written as $\sigma^2 = \mathrm{var}_p[\psi(W, \eta)]$. Many well-known estimators satisfy this property. In particular, Jung et al. [12, 13] derive such estimators from any causal effect identification formula.

Sample-Splitting Estimator of the Asymptotic Variance. The goal is to compare estimators of τ by their cost-adjusted asymptotic variance. We propose the following sample-splitting approach to obtain an estimate $\hat{\sigma}^2(\mathcal{D})$ of the asymptotic variance σ^2 using dataset $\mathcal{D}$. We randomly split $\mathcal{D}$ into two folds $\mathcal{D}^\eta$ and $\mathcal{D}^\sigma$, obtain an estimate $\hat{\eta}(\mathcal{D}^\eta)$ of the nuisance function η, and set

$$\hat{\sigma}^2(\mathcal{D}) := \mathrm{var}_{\mathcal{D}^\sigma}[\psi(W, \hat{\eta}(\mathcal{D}^\eta))], \tag{1}$$

where $\mathrm{var}_{\mathcal{D}^\sigma}$ indicates the empirical variance on fold $\mathcal{D}^\sigma$. Data splitting alleviates the bias in $\hat{\sigma}^2$ by forcing it to be independent of the bias in $\hat{\eta}$. In addition, as we will see in Section 3, when combined with UIFs, this sample-splitting procedure produces confidence bounds that scale with $\|\hat{\eta}(\mathcal{D}^\eta) - \eta\|^2$, meaning that we can use a non-parametric estimator $\hat{\eta}$ with slow error rate $\mathcal{O}(n^{-1/4})$ but still be able to estimate σ^2 at the fast parametric rate $\mathcal{O}(n^{-1/2})$.

Obtaining scaling with $\|\hat{\eta}(\mathcal{D}^\eta) - \eta\|^2$ is also the subject of double/debiased machine learning [2, 4], which shows that a property called *Neyman orthogonality* is sufficient for a sample-splitting estimator to converge at rate $\mathcal{O}(n^{-1/2})$, even when $\hat{\eta}$ has rate $\mathcal{O}(n^{-1/4})$. An implication of our results is that Neyman orthogonality is not sufficient for $\mathcal{O}(n^{-1/2})$ rates of variance estimation; one must have an UIF as well. However, it is possible to extend our framework to estimators without an UIF if one is willing to tolerate a slower rate.

Experimental Protocol. We now have all necessary elements to state the general structure of the experimental protocol. For $k = 1, \ldots, K$, let $\hat{\tau}_k$ be an asymptotically linear estimator of τ with nuisance function η_k, UIF $\psi_k(W_k, \eta_k)$, asymptotic variance $\sigma_k^2 = \mathrm{var}_p[\psi_k(W_k, \eta_k)]$, and cost c_k of observing covariates $\mathcal{Z}_k$.

The goal is to identify the $\hat{\tau}_{k^*}$ with lowest cost-adjusted asymptotic variance, *i.e.* such that $k^* = \arg\min_k c_k \sigma_k^2$. This scaling arises because guaranteeing $|\hat{\sigma}_k^2 - \sigma_k^2| = \epsilon$ 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 $\delta > 0$, $\epsilon > 0$, index k^* is (ϵ, δ)-PAC if

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

i.e. if k^* has probability at least $1 - \delta$ of being at most ϵ-suboptimal. The experimental protocols proceeds in rounds. At round t, the investigator chooses an index $k_t \in [K] := \{1, \ldots, K\}$, obtains observation $w_{k_t}^t = (z_{k_t}^t, x^t, y^t) \sim p(\mathcal{Z}_k, X, Y)$, updates the cost-adjusted estimator $\hat{\sigma}_k^2$, and decides whether a (ϵ, δ)-PAC index can be returned.

Experimental Protocol

Given: Estimators $\hat{\tau}_1, \ldots, \hat{\tau}_K$, costs $c_1, \ldots c_K$, $\epsilon > 0$, $\delta > 0$
for $t = 1, 2, \ldots$: **do**
  Choose $k_t \in [K]$
  Obtain observation $w_{k_t}^t$
  Choose whether to stop sampling
end
Return: (ϵ, δ)-PAC index k^*

To implement the experimental protocol, we propose adapting well-known best-arm-identification bandit algorithms to estimating two-sided confidence bounds on σ_k^2 that hold for all formulas and all rounds simultaneously. The next section is dedicated to developing the necessary tools for our adaptations.

3 Finite-Sample Confidence Sequences for the Asymptotic Variance

This section introduces new tools for constructing finite-sample confidence bounds on $\sigma^2 = \mathrm{var}_p[\psi(W, \eta)]$. The bounds will be determined by assumptions on the distribution of W, by smoothness properties of η, and by how quickly $\hat{\eta}$ converges to η. We focus on *confidence sequences*, which are confidence intervals that hold uniformly for a stochastic process: we derive confidence sequences for which the true and empirical means of random variables meeting certain tail assumptions can be uniformly bounded for all sample sizes simultaneously.

3.1 Confidence Sequences

Definition 1. *A random variable W is λ sub-Gaussian if there exists a constant λ such that $\mathbb{E}\left[e^{t(W-\mathbb{E}[W])}\right] \leq e^{\lambda \frac{t^2}{2}}$ $\forall t \in \mathbb{R}$. 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 $c = 2\lambda$.*

We can think of sub-Gaussian random variables as having tails no heavier than a Gaussian with variance λ, and of sub-exponential random variables as having tails no heavier than a χ^2 distribution. We emphasize that sub-Gaussianity is a very common assumption in the bandit literature and that this assumption is satisfied in many applications; for example, a B-bounded random variable is B^2 sub-Gaussian, but sub-Gaussian random variables need not be bounded. Well established concentration inequalities exist for both tail behaviors. However, we know from previous work that applying concentration bounds to $\mathbb{E}_n[W] - \mathbb{E}[W]$ pointwise (separately for many n with a union bound) leads to suboptimal sample complexity guarantees in bandit problems [10]. This issue can be addressed by employing confidence sequences, which are typically constructed from $\mathbb{E}_n[W]$ and problem parameters and contain $\mathbb{E}[W]$ at all times with high probability.

Definition 2. *A confidence sequence for W with coverage at level α is a sequence of intervals $\mathcal{C}_1, \mathcal{C}_2, \ldots \subseteq \mathbb{R}$, measurable w.r.t. $w^1, w^2, \ldots$, satisfying $\mathbb{P}(\forall n \geq 1 : \mathbb{E}[W] \in \mathcal{C}_n) > 1 - \alpha$.*

If suffices to take confidence intervals of the form $\mathcal{C}_n := [\mathbb{E}_n[W] - u_n, \mathbb{E}_n[W] + u_n]$, for a sequence of potentially random variables u_n, measurable w.r.t. $w^1, w^2, \ldots$, called a *boundary function*.

Definition 3. *We say that a boundary function u_n has correct coverage at level α for W if $\mathcal{C}_n := [\mathbb{E}_n[W] - u_n, \mathbb{E}_n[W] + u_n]$ is a confidence sequence for W with coverage at level α.*

Boundary functions with correct coverage at level α satisfy $\mathbb{P}\left(\exists n \geq 1 : |\mathbb{E}_n[W] - \mathbb{E}[W]| \geq u_n\right) \leq \alpha$, therefore giving an upper bound for the deviations between the empirical and true expected value of W for all values of n simultaneously. In contrast, using a separate confidence interval for every value of n would require a union bound over sample sizes, inflating u_n by a log-factor.

There is an established literature of deriving confidence sequences for random variables under various assumptions on tail behavior (see *e.g.* Howard et al. [8] and references therein). For our purposes, it suffices to present the specific confidence sequences that we will use, but we emphasize that many other options are available. The following lemma is a simplified version of more general results in Howard et al. [8] (for completeness, a proof is provided in the appendix).

Lemma 2. *Assume that W is a λ sub-exponential random variable with scale c. Let $\gamma > 0$ and $m > 0$ be scalar parameters, and let $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be an increasing function with summable reciprocals. With $\ell(v) := \log(h(\log_\gamma(v/m)) + \log(2/\alpha)$, define the boundary function*

$$u_n(a, b) = \frac{\gamma^{1/4} + \gamma^{-1/4}}{\sqrt{2}n} \sqrt{(an \vee m)\ell(an \vee m)} + b\frac{\sqrt{\gamma} + 1}{n}\ell(an \vee m).$$

Then $u_n(\lambda, c)$ has correct coverage at level α for W. If W is a λ sub-Gaussian random variable, then $u_v(\lambda) = u_n(\lambda, 0)$ has correct coverage at level α for W.

3.2 Finite-sample Confidence Sequences

In this section we state our main technical result: a confidence sequence for the sample-splitting estimator defined in Eq. (1). Detailed proofs can be found in the appendix.

Theorem 1. *Let $\alpha > 0$. Assume that ψ is L-Lipschitz, and let $\tilde{\tau}$ be an upper bound on $|\tau|$. Let $\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \dots$ be a sequence of datasets with $\mathcal{D}_n = \mathcal{D}_n^\eta \cup \mathcal{D}_n^\sigma$ and assume that u_n^1, u_n^2, and u_n^η are boundary functions with correct coverage at level $\alpha/3$ for $\mathbb{E}_{\mathcal{D}_n^\sigma}[\psi(W,\eta)]$, $\mathbb{E}_{\mathcal{D}_n^\sigma}[\psi(W,\eta)]^2$, and $\|\hat{\eta}(\mathcal{D}_n^\eta) - \eta\|$, respectively. Then the sample-splitting estimator $\hat{\sigma}^2(\mathcal{D}_n)$ defined in Eq. (1) satisfies*

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

In words, the theorem states that boundary functions with correct coverage for the nuisance function and tail control of $\psi(W,\eta)$ are sufficient for establishing boundary functions with correct coverage for $\hat{\sigma}^2$. We point out a few noteworthy features of this theorem. First, we only require control of $\psi(W,\eta)$ at the true η. Second, the dependence on u_n^η is of order two; hence, a slow rate of nuisance estimation, *e.g.* $u_n^\eta = \mathcal{O}(n^{-1/4})$, does not destroy the ability to estimate the asymptotic variance at rate $\mathcal{O}(n^{-1/2})$. This favorable rate is a consequence of having an UIF—otherwise, we can only obtain an $\mathcal{O}(u_n^\eta)$ result.

The following corollary is a more specific version of the theorem for the tails assumptions in the previous section. The proof can be found in the appendix.

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, for $u_n(a,b)$ and $u_n(a)$ as defined in Lemma 2 and $n' = |\mathcal{D}_n^\sigma|$,*

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

For a concrete bound, take boundary functions of Lemma 2 with $\gamma = 2$, $\lambda' = \lambda \vee 8\lambda^2$, and $h(k) = 2^{2k+1}$, as suggested by Howard et al. [8]. Then, for any $n \geq (18.6\lambda\log(\lambda n/m) + \log(2/\alpha)) \vee m/\lambda'$,

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

The above results are central to our bandit construction, since they allow us to form confidence sequences around the asymptotic variance for every identification formula that we consider.

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

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

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

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

□

Confidence Sequences for σ^2. Combing the above theorems with Eq. (1) yields a confidence sequence described in the algorithm on the right. This algorithm requires the inputs necessary for Theorem 1 as well as $\hat{\eta}$ and ψ and will be used in our bandit algorithms.

CSUpdate

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

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

In the same vein as the double/debiased machine learning literature, our main theorem presents a bound on our estimator in terms of the nuisance function estimation error, $\|\hat{\eta}(\mathcal{D}_n) - \eta\|$. Confidence sequences exist for the estimation errors of well-known estimators, such as least squares estimators [28]. However, we can adapt a pointwise guarantee for our setting. For readability, we explicitly state the two assumptions that we may use.

Definition 4. *Let $\hat{\eta}(\mathcal{D}_n)$ be an estimate of the nuisance function η calculated from dataset $\mathcal{D}_n$. We say that we have uniform control over $\|\hat{\eta}(\mathcal{D}_n) - \eta\|$ if there exists a boundary function $u_n^\eta(\alpha)$ with correct coverage at level α,* i.e.

$$\mathbb{P}\left(\exists n \geq 1 : \|\hat{\eta}(\mathcal{D}_n) - \eta\| \geq u_n^\eta(\alpha)\right) \leq \alpha.$$

We say that we have pointwise control if there exists a function $R(n, \alpha)$ such that, for any n,

$$\mathbb{P}(\|\hat{\eta}(\mathcal{D}_n) - \eta\| \geq R(n, \alpha)) \leq \alpha.$$

Pointwise control is weaker than uniform control. However, for $\alpha_n = 6\alpha/(\pi n)^2$, $R(n, 6\alpha/(\pi n)^2)$ is a boundary function since $\mathbb{P}\left(\exists n \geq 1 : \|\hat{\eta}(\mathcal{D}_n) - \eta\| \geq R(n, 6\alpha/(\pi n)^2)\right)$ is bounded above by

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\|\hat{\eta}(\mathcal{D}_n) - \eta\| \geq R(n, 6\alpha/(\pi n)^2)\right) \leq \sum_{n=1}^{\infty} \frac{6\alpha}{\pi n^2} \leq \alpha.$$

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 Confidence Sequence LUCB and SE Algorithms

This section introduces bandit algorithms using the confidence sequences developed above. We focus on adapting the two well-known algorithms *lower-upper confidence bounds* (LUCB) and *successive elimination* (SE) [3] to using confidence sequences (CS)—most other best-arm-identification algorithms, developed using Hoeffding bounds, can be similarly adapted [14][2].

CS-LUCB and CS-SE–summarized in Algorithms 1 and 2—take as input, $\forall k = 1, \ldots, K$, the influence function ψ_k, the estimator $\hat{\eta}_k$ of the nuisance function η_k, the cost per observation c_k, and the boundary functions $u_k := (u_{k,n}^\eta, u_{k,n}^1, u_{k,n}^2)$ on $\|\hat{\eta}_k - \eta_k\|$, $\psi_k(W_k, \eta_k)$, and $\psi_k(W_k, \eta_k)^2$.

At each round, CS-LUCB samples the estimator with the lowest $c_k \hat{\sigma}_k^2$ and the estimator with the lowest bound among the remaining estimators. Intuitively, these arms are two most likely to be miss-classified. The algorithm proceeds until there is a clear separation between the two.

CS-SE keeps a set S of plausibly best arms. At each round, it collects new data, updates the confidence sequences of all the arms in S, and then removes from S the set R of arms which are not plausibly the best, namely those that have lower bounds higher than the upper bound of the best. Eventually, all non-optimal arms are removed and the algorithm terminates.

Algorithm 1 CS-LUCB

Input $\epsilon > 0, \delta > 0, \Delta_n > 1,$
$\{\psi_k, \hat{\eta}_k, u_k, c_k : k \in [K]\}$
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**
  $B_t \leftarrow \{k : \hat{\sigma}_k^2 = \min\{\hat{\sigma}_k^2 : k \in [K]\}\}$
  $l_t \leftarrow \arg\max_{k \in B_t} \hat{\sigma}_k^2 + \beta_k$
  $u_t \leftarrow \arg\min_{k \notin B_t} \hat{\sigma}_k^2 - \beta_k$
  if $\hat{\sigma}_{l_t}^2 - \beta_{l_t} \leq \hat{\sigma}_{u_t}^2 + \beta_{u_t} - \epsilon$ **then**
    Return $\hat{k} = k_t$
  end
  for $k \in u_t, l_t$ **do**
    Obtain Δ_n new samples $\mathcal{D}$
    Add half of $\mathcal{D}$ to $\mathcal{D}_{k_t}^\eta$ 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

[2] We are not the first to consider bounds beyond Hoeffding for these algorithms: see *e.g.* Kaufmann and Kalyanakrishnan [15].

In the remainder of the section we perform the usual best-arm identification analysis, a) by proving that both algorithms return (ϵ, δ)-PAC indices, and b) by providing an upper bound on the sample complexities as a function of the problem instance parameters and u_n^η (the proofs can be found in the appendix). We use the notation $\Delta_k := \sigma_k^2 - \min_k \sigma_k^2$.

Algorithm 2 CS-SE

Input $\delta > 0$, $\Delta_n > 1$
$\{\psi_k, \hat{\eta}_k, u_{k,n}, c_k, : k \in [K]\}$
$S \leftarrow [K]$, $\mathcal{D}_k^\eta, \mathcal{D}_k^\sigma \leftarrow \emptyset \; \forall k \in [K]$
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
  $k^* \leftarrow \arg\min_k c_k \hat{\sigma}_k^2$
  $R \leftarrow \{k \in S : c_{k^*}(\hat{\sigma}_{k^*}^2 + \beta_{k^*}) \leq c_k(\hat{\sigma}_k^2 - \beta_k)\}$
  $S \leftarrow S \setminus R$
end
Return $\hat{k} = S$

Theorem 2. *Assume that $u_{k,n}^\eta \to 0$ and that the conditions of Theorem 1 hold. Then CS-LUCB returns a (ϵ, δ)-PAC index.*

If, additionally, there exists constants ν_η, ν_1,and ν_2 such that $u_{k,n}^\theta \leq \mathcal{O}(n^{\nu_\theta} \log(nK/\delta))$ for all $\theta \in \{\eta, 1, 2\}$, the sample complexity is

$$\mathcal{O}\left(H^{\nu,\epsilon/2} \log \frac{H^{\nu,\epsilon/2}}{\delta}\right)$$

for $\nu = \max\{2\nu_\theta, \nu_1, \nu_2\}$ and complexity term $H^{\nu,\epsilon/2} = \sum_{k \neq k^*} (\Delta_k \vee \epsilon/2)^{1/\nu}$.

Theorem 3. *Assume that the estimate for η is consistent (i.e. $u_{k,n}^\eta \to 0$) and that the conditions of Theorem 1 hold. Then CS-SE returns a $(0, \delta)$-PAC index.*

If, additionally, there exists constants ν_η, ν_1, and ν_2 such that $u_{k,n}^\theta \leq \mathcal{O}(n^{\nu_\theta} \log(nK/\delta))$ for all $\theta \in \{\eta, 1, 2\}$, the sample complexity is

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

where $\nu = \max\{2\nu_\theta, \nu_1, \nu_2\}$.

We can see that the sample complexity changes depending on the relative relation of the three nuisance function estimates. Evaluating $\nu = \max\{2\nu_\theta, \nu_1, \nu_2\}$ under the conditions of Corollary 1 yields the following corollary.

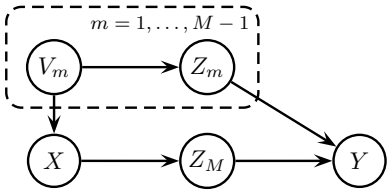
Corollary 2. *Under the conditions of Corollary 1 and additionally assuming that $u_{k,n}^\eta = \mathcal{O}(n^{\nu_\eta} \log(nK/\delta))$, CS-SE returns a $(0, \delta)$-PAC index with sample complexity*

$$\mathcal{O}\left(\sum_{k=1}^{K} \Delta_k^{-2} \left(\log \frac{K}{\delta \Delta_k}\right)^2\right) \quad \textit{or} \quad \mathcal{O}\left(\sum_{k=1}^{K} \Delta_k^{1/\nu_\eta} \left(\log \frac{K}{\delta \Delta_k}\right)^{-1/\nu_\eta}\right),$$

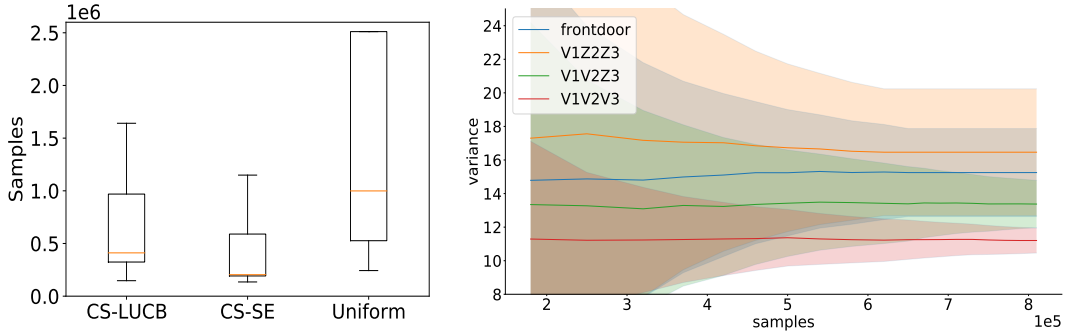
with the first bound only if $\nu_\eta \leq -1/4$.

Throughout, we focus on confidence sequences with $\mathcal{O}(\sqrt{n^{-1} \log(n/\delta)})$ upper bounds instead of iterated-log upper bounds. It is known that the optimal best-arm-identification rates are of the form $\sum_k \Delta_k^{-2} \log\log(1/\delta\Delta_k^2)$ [11], and we can obtain rates of this form if we used confidence sequences with an iterated log. However, we focus on single-log bounds because they are more common for bounding the $\|\hat{\eta} - \eta\|$ term and, for practical sample sizes, the single-log confidence sequences are generally tighter [8].

5 Experiments



This section presents an evaluation of the proposed CS-LUCB and CS-SE algorithms on the causal DAG on the left figure, containing one frontdoor path $X \to Z_M \to Y$, and $M - 1$ backdoor paths $X \leftarrow V_m \to Z_m \to Y$, as indicated with a plate notation.



(a) (b)

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

More specifically, we considered the following structural equation model: $V_m \sim \mathcal{N}(0, I_2)$, where $\mathcal{N}$ indicates a Gaussian distribution and I_d the d-dimensional identity matrix; $Z_m = A_m V_m + \epsilon_z$ for matrix A_m and $\epsilon_z \sim \mathcal{N}(0, .1 * I_3)$; $X \sim Bernoulli\left(1/\left(1 + e^{-\sum_{m=1}^{M-1} \delta_m V_m}\right)\right)$ for vectors $\delta_1, \ldots, \delta_{M-1}$; Z_M sampled from a categorical distribution with 10 points of support $\{s_i : i \in [10]\}$ and probability vector p_1 if $X = 1$ and p_0 if $X = 0$; and $Y = \sum_{m=1}^{M} B_m Z_m + \epsilon_y$ for matrix B_m and $\epsilon_y \sim \mathcal{N}(0, .1)$.

Specific model instances were obtained by sampling the parameters as follows. Backdoor paths were sampled by first taking $U_m \sim \text{Uniform}[.1, .9]$, and by then setting each coordinate of $\delta_m \sim \mathcal{N}(0, U_m^2)$, each coordinate of $B_m \sim N(0, (2 - U_m)^2)$, and each coordinate of $A_m \sim \mathcal{N}(0, U^2/4)$. The purpose of this sampling scheme was to anti-correlate the strength of paths $V_m \to X$ and $Z_m \to Y$; since all adjustment sets must include either V_m or Z_m, having a random model where one path is, on average, stronger than the other tends to create a larger separation of the asymptotic variances and a more interesting problem instance for the bandit algorithm. B_M had coordinates independently sampled from $\mathcal{N}(0, 1/4)$. The support points of Z_M were standard 2-dimensional normals, and p_1 and p_0 were chosen randomly from the simplex but filtered so that the causal contrast effect was not too small. More details are given in the appendix.

As causal contrast $\tau := \sum_x \lambda_x \mu_{\text{do}(x)}$ we considered the average treatment effect (ATE) (*i.e.* $\lambda = (1, -1)$), which for this particular case is given by $\tau = \sum_{i=1}^{10} (p_1(i) - p_0(i)) B_M s_i$, and the adjustment and frontdoor criteria as identification formulas for a total of 9 formulas. Following our framework, each estimator was calculated by finding the value of τ that causes the uncentered efficient influence function to be zero.

For the adjustment criterion $\mu_{\text{do}(x)} = \mathbb{E}_p[\mu_x(\mathcal{Z})]$ with $\mu_x(\mathcal{Z}) := \mathbb{E}_p[Y|X = x, \mathcal{Z}]$, we used the AIPWE estimator (see Section 2), where the components of the nuisance function $e_x(\mathcal{Z}) = p(X = 1|\mathcal{Z} = z)$ and $\mu_x(\mathcal{Z}) = \mathbb{E}_p[Y|X = x, \mathcal{Z}]$ were estimated with logistic regression (with regularization to keep the predictions away from the boundary) and with linear regression respectively.

For the frontdoor criterion $\mu_{\text{do}(x)} = \sum_z p(\mathcal{Z} = z|X = x) \sum_{x'} p(X = x')\mu_{x'}(\mathcal{Z} = z)$, we used the unique efficient influence function derived in Fulcher et al. [5] for the population intervention indirect effect (corresponding to the ATE in our causal DAG). This is given by $\psi(W, \eta) = (Y - \mathbb{E}_p[Y|\mathcal{Z}]) \left(\frac{p(\mathcal{Z}|X=1)}{p(\mathcal{Z}|X)} - \frac{p(\mathcal{Z}|X=0)}{p(\mathcal{Z}|X)}\right) + \left(\frac{X}{p(X=1)} - \frac{1-X}{p(X=0)}\right) \mathbb{E}_p[Y|\mathcal{Z}] + \sum_{z'} \mathbb{E}_p[Y|z'] \left(p(\mathcal{Z} = z'|X = 1) - p(\mathcal{Z} = z'|X = 0)\right) - \sum_{z'} \mathbb{E}_p[Y|\mathcal{Z} = z'] \left(\frac{Xp(\mathcal{Z}=z'|X=1)}{p(X=1)} - \frac{(1-X)p(\mathcal{Z}=z'|X=0)}{p(X=0)}\right)$. The components of the nuisance function $\mathbb{E}_p[Y|\mathcal{Z}], p(\mathcal{Z}|X)$, and $p(X)$ were estimated using linear regression and simple counting

respectively. Confidence sequences for linear regression are known; see Abbasi-Yadkori et al. [1, Theorem 2] or Lemma 7 in the appendix.

We set V_m to have cost 1 and Z_m to have cost 3, for $m = 1, \ldots, M$—these values were chosen to reflect a setting in which post-treatment variables and variables further from X are harder to measure. The cost of each covariate set was set to the sum of the costs associated with every variable. We randomly generated 10 different probability distributions by sampling 10 sets of parameters values for $A_1, \ldots, A_M$, $B_1, \ldots, B_M$, and $\delta_1, \ldots, \delta_{M-1}$. For each distribution, we compared the sample complexity of CS-LUCB, CS-SE, and a naive uniform sampling baseline (which samples all formulas equally until the lowest-variance formula has a confidence set that does not intersect the others) on 4 independent datasets generated from the distribution. All 40 runs are summarized in the box plot of Fig. 1 (a). The sampled probability distributions displayed a wide gamut of behaviors; for example, no formula uniformly dominated (see the appendix for the details). Compared to the uniform sampling baseline, CS-LUCB and CS-SE display substantial reduction in sample complexity, with CS-SE having a slight advantage over CS-LUCB. In comparison with uniform sampling, CS-SE and CS-LUCB needed, on average, 34% and 51% the number of samples to reach a conclusion.

Fig. 1(b) shows an example of confidence sequences corresponding to the four best arms for CS-SE, up to the $8.1 * 10^5$ samples required to find the optimal arm. One can clearly see that the confidence sequences shrink and that the algorithm terminates as soon as the red sequence no longer intersects the green sequence. One can also observe when the sequences stop shrinking, indicating that enough samples were acquired to eliminate that arm completely.

6 Discussion

Much of the literature on causal inference from observational data has focused either on the question of *identification* of causal effects using structural knowledge of the causal graph underlying the data generation mechanism or on the question of 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 costs need to be accounted for, 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 of 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 therefore not appropriate for a bandit algorithm.

Instead of a best-arm-identification problem that focuses on finding the arm with the best long term behavior (which we quantify by the asymptotic variance), we could consider the problem of trying to obtain the best estimate of the causal effect given a limited budget. This problem is more analogous to the cumulative regret minimizing bandit problem and its exploration will be the subject of future research.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] In the Discussion section.
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] In Section 2.
   (b) Did you include complete proofs of all theoretical results? [Yes] We give sufficiently detailed proofs in the main text and complete proofs in the appendix.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] The code is proprietary, but will be open sourced after publication. The code is a simple modification of standard bandits algorithm, for which open source libraries are available.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] In the main text.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] As the experiments do not require large-scale resources.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

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] F. Richard Guo and Emilija Perković. Efficient least squares for estimating total effects under linearity and causal sufficiency. *arXiv preprint arXiv:2008.03481*, 2021.

[7] 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.

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

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

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

[16] Jack Kuipers and Giusi Moffa. The variance of causal effect estimators for binary V-structures. *arXiv preprint arXiv:2004.09181*, 2020.

[17] Manabu Kuroki. Selection of post-treatment variables for estimating total effect from empirical research. *Journal of the Japan Statistical Society*, 30(2):115–128, 2000.

[18] Manabu Kuroki and Masami Miyakawa. Covariate selection for estimating the causal effect of control plans by using causal diagrams. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):209–222, 2003.

[19] Manabu Kuroki and Hisayoshi Nanmo. Variance formulas for estimated mean response and predicted response with external intervention based on the back-door criterion in linear structural equation models. *AStA Advances in Statistical Analysis*, 104(4):667–685, 2020.

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

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

[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] 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.