Bayesian Causal Bandits with Backdoor Adjustment Prior

Anonymous authors
Paper under double-blind review

Abstract

The causal bandit problem setting is a sequential decision-making framework where actions of interest correspond to interventions on variables in a system assumed to be governed by a causal model. The underlying causality may be exploited when investigating actions in the interest of optimizing the yield of the reward variable. Most existing approaches assume prior knowledge of the underlying causal graph, which is in practice restrictive and often unrealistic. In this paper, we develop a novel Bayesian framework for tackling causal bandit problems that does not rely on possession of the causal graph, but rather simultaneously learns the causal graph while exploiting causal inferences to optimize the reward. Our methods efficiently utilize joint inferences from interventional and observational data in a unified Bayesian model constructed with intervention calculus and causal graph learning. For the implementation of our proposed methodology in the discrete distributional setting, we derive an approximation of the sampling variance of the backdoor adjustment estimator. In the Gaussian setting, we characterize the interventional variance with intervention calculus and propose a simple graphical criterion to share information between arms. We validate our proposed methodology in an extensive empirical study, demonstrating compelling cumulative regret performance against state-of-the-art standard algorithms as well as optimistic implementations of their causal variants that assume strong prior knowledge of the causal structure.

1 Introduction

The multi-armed bandit (MAB) problem is a well-known sequential allocation framework for experimental investigations (Berry & Fristedt, 1985). Classically, the MAB problem formulation features an action set $\mathcal{A}$ consisting of $|\mathcal{A}| = K$ actions, also called arms, typically corresponding to interventions. Each arm $a \in \mathcal{A}$ defines a real-valued distribution for the reward signal, with expected reward μ_a. The objective of an allocation policy is to sequentially pick arms in a manner that maintains a balance between exploration and exploitation in the interest of identifying and obtaining the greatest reward. Maximally and effectively utilizing all available information is imperative, especially when investigating interventions that are either or both resource-demanding and time consuming.

Lattimore et al. (2016) proposed the causal bandit (CB) problem setting wherein a non-trivial probabilistic causal model is assumed to govern the distribution of the reward variable and its covariates (Pearl, 2000). The addition of causal assumptions introduces avenues by which interventional distributions may be inferred from observational distributions and information may be shared between arms. Most works addressing the CB problem exploit strong assumptions as to prior knowledge of the underlying causal model to achieve improvements over standard MAB algorithms. In this work, we develop a Bayesian CB framework that does not require prior knowledge of the underlying causal structure, but instead efficiently utilizes previously available observational data and acquired interventional data to inform exploitation and guide exploration.

1.1 Related Work

In its original formulation by Lattimore et al. (2016), the CB problem presupposes knowledge of the underlying causal graph. Accordingly, most proposed CB algorithms require knowledge of the causal graph structure (Lattimore et al., 2016; Lee & Bareinboim, 2018; Maiti et al., 2021; Yabe et al., 2018), and some additionally assume certain model parameters are given (Lu et al., 2020; Nair et al., 2021). Furthermore, many approaches are dependent on some restrictive form or class of graphs. These assumptions are restrictive and often unrealistic in practice.

More recently, Lu et al. (2021) proposed a central node approach based on the work of Greenewald et al. (2019) that does not assume prior knowledge of the causal graph, but rather asymptotic knowledge of the observational distribution. Their approach is restrictive in terms of structural and distributional assumptions, and while it is generally reasonable to assume that observational data is much more accessible than interventional data (Greenewald et al., 2019), the large-sample observational setting is not often realistic. de Kroon et al. (2022) proposed an estimator using separating sets to share information between arms without assuming prior knowledge of nor requiring discovery of the causal graph. Their methodology makes no attempt to learn the causal graph, and makes use of observational data only to strengthen conditional independence testing to identify separating sets.

Relevant to our work is the estimation of interventional quantities from observational data using intervention calculus (Pearl, 2000). In the CB setting, Lattimore et al. (2016) and Nair et al. (2021) consider graph structures with no confounding such that the interventional distributions are equivalent to conditional distributions, and Maiti et al. (2021) proposed a consistent estimator for the expected reward for discrete variables using both interventional and observational data in the presence of confounding. Our work extends the Bayesian model averaging approach proposed by Pensar et al. (2020) wherein the possible causal effect estimates are averaged across an observational posterior distribution of graphs.

1.2 Our Contributions

We approach the CB problem from a Bayesian perspective, assuming simply that finite samples of observational data are available. Importantly, we do not assume the causal graph is known, nor are we restrictive as to the class of graphs. We design a novel Bayesian CB framework called **B**ayesian **B**ackdoor **B**andit (BBB) that efficiently utilizes the entirety of evidence from an ensemble of observational and interventional data in a unified Bayesian model. Our proposed BBB methodology quantifies the uncertainty in the expected reward estimates as contributed to by the reward signal and the causal model to identify potentially profitable exploration, simultaneously learning the causal graph in addition to and for the purposes of improving estimates to exploit. Through extensive numerical experiments, we validate our methodology by demonstrating compelling empirical performance against both non-causal and causal algorithms. In particular, even with modest amounts of observational data, our BBB approach achieves substantially superior cumulative regret performance compared to standard algorithms, as well as against a generously optimistic version of the causal central node approach proposed by Lu et al. (2021) that assumes large-sample observational data.

Additionally, in detailing the application of our methods to the discrete and Gaussian distributional settings, we propose various developments that are of independent interest. In the discrete setting, we derive an approximation for the sampling variance of the backdoor adjustment probability estimate. In the Gaussian setting, we characterize the interventional variance of a target variable using intervention calculus and correspondingly propose an estimator, and we propose a simple graphical criterion for sharing causal information between arms to perform intervention calculus with jointly observational and interventional data.

The remainder of the paper is arranged as follows. We first review relevant background and notation in Section 2. Then, we develop the formulation of our proposed Bayesian backdoor prior and posterior update in Section 3, discussing the design of informative conditional priors given a graph and Bayesian model averaging across graph structures. In Section 4, we develop our proposed algorithms by applying established MAB algorithms under the BBB framework, and we discuss details regarding the implementation of BBB in the discrete and Gaussian settings in Section 5. Finally, we provide extensive empirical results in Section 6

and conclude with a discussion in Section 7. Appendices A to D contain proofs, additional details and numerical results, and technical derivations.

2 Preliminaries

We consider the setting where the generative model governing a joint probability distribution P of a set of p variables $\mathbf{X} = \{X_1, \ldots, X_p\}$ is a causal Bayesian network (CBN). A CBN consists of its structure $\mathcal{G}$, which takes the form of a directed acyclic graph (DAG), and its parameters $\Theta_{\mathcal{G}}$. Its DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, often referred to as the underlying causal graph, is composed of a set of nodes $\mathbf{V} = \{1, \ldots, p\}$ in one-to-one correspondence with the variables, and a set of directed edges $\mathbf{E}$ oriented such that there are no directed cycles. As is standard in causal literature, we may refer to a node $i \in \mathbf{V}$ and its corresponding variable $X_i \in \mathbf{X}$ interchangeably.

The probability distribution P imposed by the CBN factorizes according to structure $\mathcal{G}$, with local probability distributions defined by $\Theta_{\mathcal{G}}$. In particular, $P(\mathbf{X}) = \prod_{i=1}^{p} P(X_i \mid \mathbf{Pa}_i^{\mathcal{G}}, \theta_i^{\mathcal{G}})$, where $\mathbf{Pa}_i^{\mathcal{G}} = \{X_j : j \to i \in \mathbf{E}\}$ is the parents of X_i in $\mathcal{G}$, and the local parameters $\theta_i^{\mathcal{G}} \in \Theta_{\mathcal{G}}$ specify the conditional probability distribution (CPD) of X_i given its parents. In our work, we assume that $\mathbf{X}$ is a causally sufficient system with no unobserved confounders.

The action set $\mathcal{A}$ consists of $|\mathcal{A}| = K$ arms that correspond to interventions on variables in $\mathbf{X} \setminus Y$, where $Y = X_p$ is the reward variable (Lattimore et al., 2016). In particular, let arm $a \in \mathcal{A}$ correspond to a deterministic atomic intervention denoted $do(X_{\langle a \rangle} = x_a)$ (Pearl, 1995), fixing $X_{\langle a \rangle}$ to some value $x_a \in \mathrm{Dom}(X_{\langle a \rangle})$, where $\langle a \rangle \in \mathbf{V}$ is the node corresponding to the intervened variable. The expected reward of each arm $a \in \mathcal{A}$ is given by $\mu_a \coloneqq \mathrm{E}[Y \mid do(X_{\langle a \rangle} = x_a)]$, and there is some optimal arm $a^* \coloneqq \mathrm{argmax}_{a \in \mathcal{A}} \mu_a$ corresponding to the optimal reward $\mu^* \coloneqq \mu_{a^*}$. Given a horizon of time steps T, let $a_t \in \mathcal{A}$ be the arm pulled by an algorithm at time step $t \in \{1, \ldots, T\}$. Denote by $n_a(t) = \sum_{l=1}^{t} \mathbb{1}\{a_l = a\}$ the number of times arm a has been pulled in t time steps. We take interest in optimizing the cumulative regret, where the objective of the algorithm is to pull arms over T time steps with a balance between exploring different arms and exploiting the reward signal to minimize the expected cumulative regret $\mathrm{E}[R_T] = T\mu^* - \sum_{a \in \mathcal{A}} \mu_a \mathrm{E}[n_a(T)]$.

In our problem formulation, we assume possession of n_0 samples of observational data $\mathcal{D}_0$ prior to investigating arms. We denote by $\mathcal{D}^{(t)}$ the interventional data acquired by pulling arm a_t at time t, and by $\mathcal{D}_a[t] = \bigcup_{l \le t, a_l = a} \mathcal{D}^{(l)}$ the accumulated interventional data from arm a through time t. The combined observational and interventional data accrued through time t is $\mathcal{D}[t] = \mathcal{D}_0 \cup \bigcup_{a \in \mathcal{A}} \mathcal{D}_a[t]$, which we refer to as ensemble data.

We now describe a Bayesian approach to the general MAB problem, with some notation adapted from Kaufmann et al. (2012). The parameters $\Theta_{\mathcal{A}} = (\theta_a)_{a \in \mathcal{A}}$, assumed to mutually independently define the corresponding marginal reward distributions $p_{\theta_a}(y) \coloneqq P[Y = y \mid do(X_{\langle a \rangle} = x_a)]$, jointly follow a modular prior distribution $\Pi^0(\Theta_{\mathcal{A}}) = \prod_{a \in \mathcal{A}} \pi_a^0(\theta_a)$. Typically, $(\pi_a^0)_{a \in \mathcal{A}}$ are chosen to be all equal and uninformative. When arm $a_t \in \mathcal{A}$ is pulled at time step t and a realization $y_t \leftarrow Y \mid do(X_{\langle a_t \rangle} = x_{a_t})$ is observed, the posterior Π^t is computed by updating according to $\pi_{a_t}^t(\theta_{a_t}) \propto p_{\theta_{a_t}}(y_t)\, \pi_{a_t}^{t-1}(\theta_{a_t})$, while $\pi_a^t = \pi_a^{t-1}$ for $a \neq a_t$. For each arm $a \in \mathcal{A}$, the posterior π_a^t induces a posterior distribution for the expected reward μ_a, which is simply a marginal or transformation of π_a^t since, in general, μ_a is a function of θ_a. These posteriors are utilized by Bayesian MAB algorithms, which we discuss and apply under our proposed framework in Section 4.

3 Designing Informative Priors with Intervention Calculus

In this section, we detail the development of our proposed BBB model consisting of an informative prior Π^0 constructed using observational data that seamlessly integrates interventional data to obtain the posterior Π^t. For each arm $a \in \mathcal{A}$, we construct conditional priors $\pi_{a|\mathbf{Z}}^0(\theta_a)$ using the backdoor adjustment given sets $\mathbf{Z} \subseteq \mathbf{X} \setminus X_{\langle a \rangle}$ as follows. If $\mathbf{Z}$ satisfies the backdoor criterion relative to $X_{\langle a \rangle}$ and Y (Pearl, 2000, Definition 3.3.1), then the interventional distribution $Y \mid do(X_{\langle a \rangle} = x_a)$ may be expressed in terms of the

joint observational distribution of $\{X_{\langle a \rangle}, Y\} \cup \mathbf{Z}$ via the backdoor adjustment (Pearl, 2000, Theorem 3.3.2):

$$P[Y = y \mid do(X_{\langle a \rangle} = x_a)] = \sum_{\mathbf{z} \in \mathrm{Dom}(\mathbf{Z})} P(Y = y \mid X_{\langle a \rangle} = x_a, \mathbf{Z} = \mathbf{z}) P(\mathbf{Z} = \mathbf{z}). \tag{1}$$

Eq. (1) provides an avenue through which an estimator for μ_a using observational data may be derived, which we denote $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$ and with which we design an informative prior $\pi^0_{a|\mathbf{Z}}$ such that the induced prior distribution of the expected reward μ_a satisfies

$$\mathrm{E}_{\pi^0_{a|\mathbf{Z}}}[\mu_a] = \hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z}), \quad \mathrm{Var}_{\pi^0_{a|\mathbf{Z}}}[\mu_a] = \hat{\mathrm{SE}}^2\left[\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})\right]. \tag{2}$$

The matching of the prior variance with the sampling variance of the backdoor adjustment estimator endeavors to assign the appropriate prior effective sample size. When arm $a_t = a$ is pulled at time step $t \in \{1, \ldots, T\}$ and a realization of the reward $y_t \leftarrow Y \mid do(X_{\langle a \rangle} = x_a)$ is observed in $\mathcal{D}^{(t)}$, the posterior $\pi^t_{a|\mathbf{Z}}$ is computed by updating according to $\pi^t_{a|\mathbf{Z}}(\theta_a) \propto p_{\theta_a}(y_t)\ \pi^{t-1}_{a|\mathbf{Z}}(\theta_a)$.

Thus far we have taken for granted the possession of adjustment set $\mathbf{Z}$, the validity of which is dependent on the underlying causal structure $\mathcal{G}$ which we assume to be unknown. If $Y \notin \mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}$, then $\mathbf{Z} = \mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}$ satisfies the backdoor criterion relative to $X_{\langle a \rangle}$ and Y, and its uncertainty is quantified by the posterior probability $P(\mathbf{Pa}_{\langle a \rangle} = \mathbf{Z} \mid \mathcal{D}[t])$ given the ensemble data at time t. Accordingly, the posterior of θ_a is determined by averaging over all possible parent sets for $X_{\langle a \rangle}$:

$$\pi^t_a(\theta_a) = \sum_{\mathbf{Z} \subseteq \mathbf{X} \setminus X_{\langle a \rangle}} \pi^t_{a|\mathbf{Z}}(\theta_a) P(\mathbf{Pa}_{\langle a \rangle} = \mathbf{Z} \mid \mathcal{D}[t]), \tag{3}$$

which is the key posterior distribution to be updated at each time step t in the Bayesian CB problem. Note that if $Y \in \mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}$, then $P[Y = y \mid do(X_{\langle a \rangle} = x_a)] = P(Y = y)$ holds straightforwardly for $y \in \mathrm{Dom}(Y)$. Accordingly, if $Y \in \mathbf{Z}$, we compute $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$ with the marginal distribution of Y for the design of $\pi^0_{a|\mathbf{Z}}$.

The parent set distribution in (3) is obtained according to a posterior distribution of DAG structures:

$$P(\mathbf{Pa}_i = \mathbf{Z} \mid \mathcal{D}[t]) = \sum_{\mathcal{G}' : \mathbf{Pa}^{\mathcal{G}'}_i = \mathbf{Z}} P(\mathcal{G}' \mid \mathcal{D}[t]). \tag{4}$$

The structure posterior is given by $P(\mathcal{G} \mid \mathcal{D}[t]) \propto P(\mathcal{D}[t] \mid \mathcal{G}) P(\mathcal{G})$, where $P(\mathcal{G})$ is the structure prior, and the marginal likelihood $P(\mathcal{D}[t] \mid \mathcal{G}) = \int P(\mathcal{D}[t] \mid \mathcal{G}, \Theta_{\mathcal{G}}) P(\Theta_{\mathcal{G}} \mid \mathcal{G}) d\Theta_{\mathcal{G}}$ is obtained by integrating the likelihood function over the support of a conjugate prior of the parameters as follows. Let $m \in \mathcal{I} \coloneqq \{1, \ldots, M\}$ index the $M = n_0 + t$ samples of data in $\mathcal{D}[t]$, and let $\mathcal{O}_i \subseteq \mathcal{I}$ represent the data points for which X_i is not fixed by intervention. We make standard assumptions of global and local parameter independence and parameter modularity (see Heckerman et al. (1995) and Friedman & Koller (2003) for details). These allow us to express the marginal likelihood as $P(\mathcal{D}[t] \mid \mathcal{G}) = \prod_{i=1}^{p} P(x_i[\mathcal{O}_i] \mid \mathbf{pa}^{\mathcal{G}}_i[\mathcal{O}_i])$, where $x_i[\cdot]$ and $\mathbf{pa}^{\mathcal{G}}_i[\cdot]$ represent indexed samples of X_i and $\mathbf{Pa}^{\mathcal{G}}_i$ in $\mathcal{D}[t]$, respectively. Assuming a conjugate prior and complete data, each conditional likelihood $P(x_i[\mathcal{O}_i] \mid \mathbf{pa}^{\mathcal{G}}_i[\mathcal{O}_i])$ can be calculated in closed form by integrating over the parameters:

$$P(x_i[\mathcal{O}_i] \mid \mathbf{pa}^{\mathcal{G}}_i[\mathcal{O}_i]) = \int \left[\prod_{m \in \mathcal{O}_i} P\left(x_i[m] \mid \mathbf{pa}^{\mathcal{G}}_i[m], \theta^{\mathcal{G}}_i\right) \right] P(\theta^{\mathcal{G}}_i) d\theta^{\mathcal{G}}_i \tag{5}$$

where $\theta^{\mathcal{G}}_i = \theta_{X_i | \mathbf{Pa}^{\mathcal{G}}_i}$ is the parameters specifying the conditional distribution of X_i given its parents (Eaton & Murphy, 2007).

Assuming the distribution P is faithful to $\mathcal{G}$ (that is, all and only the conditional independence relationships in P are entailed by $\mathcal{G}$), the posterior probability $P(\mathcal{G} \mid \mathcal{D}[t])$ will concentrate around the Markov equivalence class with increasing samples of observational data n_0. The equivalence class consists of the identification of all direct edge connections (that is, the skeleton of $\mathcal{G}$) and some edge orientations called compelled edges, but even with infinite observational data, in general, not all edge orientations are identifiable without interventional data. The effect on $P(\mathcal{G} \mid \mathcal{D}[t])$ of pulling arm $a \in \mathcal{A}$ and observing interventional data according

to the intervention $do(X_{\langle a \rangle} = x_a)$ is primarily though not limited to that of clarifying the orientation of the edges incident to $X_{\langle a \rangle}$ in $\mathcal{G}$.

Considering that the DAG space grows super-exponentially with the number of variables (Robinson, 1977), computation of the parent set probabilities $P(\mathbf{Pa}_i \mid \mathcal{D}[t])$ is admittedly challenging, even when the maximum number of parents is restricted. It is generally computationally advantageous to additionally assume a structure prior satisfying modularity, that is $P(\mathcal{G}) = \Pi_{i=1}^{p} P(\mathbf{Pa}_i^{\mathcal{G}})$, so that the posterior distribution is proportional to decomposable weights consisting of the product of local scores depending only on a node and its parents (Friedman & Koller, 2003). This property of score decomposability is crucial for the efficient implementation of Markov Chain Monte Carlo (MCMC) methods in which the probability distribution of features in $\mathcal{G}$ may be estimated by sampling DAGs from a Markov chain with stationary distribution $P(\mathcal{G} \mid \mathcal{D}[t])$ (Madigan et al., 1995; Friedman & Koller, 2003; Kuipers & Moffa, 2017). Particularly useful for our purposes is an algorithm developed by Pensar et al. (2020) to compute the exact parent set probabilities for a graph in time $O(3^p p)$ that also takes advantage of score decomposability.

4 Bayesian Backdoor Bandit Algorithms

In this section, we apply our proposed BBB framework to several state-of-the-art MAB algorithms, namely upper confidence bound (UCB), Thompson sampling (TS), and Bayesian UCB (Bayes-UCB). Each method is concerned with designing and computing some criterion $U_a(t)$ to maintain a balance between exploration and exploitation when selecting arms according to $a_t \in \operatorname{argmax}_{a \in \mathcal{A}} U_a(t)$. In what follows, we briefly introduce these methods and discuss their application under the BBB framework.

The general UCB family of algorithms operates under the principle of optimism in the face of uncertainty (Lai & Robbins, 1985). Arms that have not been investigated as many times as others have more uncertain reward estimates and thus optimistically have potential for greater reward, motivating the design of a padding function $F_a(t)$ for computing the selection criterion $U_a(t) = \hat{\mu}_a(t) + F_a(t)$. Intuitively, the combination of the expected reward estimate $\hat{\mu}_a(t)$ and the uncertainty $F_a(t)$ maintains a balance between high confidence exploitation and potentially profitable exploration. Perhaps the most well-known and typically the default instantiation of UCB algorithms is UCB1 (Agrawal, 1995; Auer et al., 2002) which computes the following criterion:

$$U_a(t) = \hat{\mu}_a(t-1) + c\sqrt{\log(t-1)/n_a(t-1)}. \tag{6}$$

The confidence tuning parameter $c > 0$, discussed in Sutton & Barto (2018), controls the desired degree of exploration, where $c = \sqrt{2}$ in Auer et al. (2002). Hereafter, when discussing UCB, we refer to the policy expressed by the criterion in (6).

In what we refer to as BBB-UCB, we estimate the expected reward with the posterior mean $\mathrm{E}_{\pi_a^{t-1}}[\mu_a]$ with respect to (3), and we replace $1/n_a(t-1)$ with the posterior variance $\mathrm{Var}_{\pi_a^{t-1}}[\mu_a] \sim 1/(n_0 + n_a(t-1))$. In particular, for each arm $a \in \mathcal{A}$, we compute

$$U_a(t) = \mathrm{E}_{\pi_a^{t-1}}[\mu_a] + c\sqrt{\mathrm{Var}_{\pi_a^{t-1}}[\mu_a] \log(t)}, \tag{7}$$

where, for outer expectation with respect to the parent set distribution $\mathbb{P}_{t-1}(\mathbf{Z}) \coloneqq P(\mathbf{Pa}_i = \mathbf{Z} \mid \mathcal{D}[t-1])$ in (4),

$$\mathrm{E}_{\pi_a^{t-1}}[\mu_a] = \mathrm{E}_{\mathbb{P}_{t-1}}\left[\mathrm{E}_{\pi_{a|\mathbf{Z}}^{t-1}}[\mu_a]\right], \quad \mathrm{Var}_{\pi_a^{t-1}}[\mu_a] = \mathrm{E}_{\mathbb{P}_{t-1}}\left[\mathrm{Var}_{\pi_{a|\mathbf{Z}}^{t-1}}[\mu_a]\right] + \mathrm{Var}_{\mathbb{P}_{t-1}}\left[\mathrm{E}_{\pi_{a|\mathbf{Z}}^{t-1}}[\mu_a]\right].$$

The Bayesian procedures of Bayes-UCB and TS are especially amenable to straightforward application under the BBB framework. These methods follow the Bayesian MAB formulation introduced in Section 2, typically taking as input uninformative priors in Π^0 that are equivalent for each arm. At each time step t, TS samples the expectations from the posterior $U_a(t) \leftarrow \pi_a^{t-1}(\mu_a)$, effectively selecting arm $a \in \mathcal{A}$ with probability equal to the posterior probability that μ_a is the highest expectation (Thompson, 1933). Bayes-UCB instead

computes for each arm at time t an upper quantile of μ_a based on its posterior distribution induced by π_a^{t-1}:

$$U_a(t) = Q\left(1 - \frac{1}{t(\log T)^c}, \pi_a^{t-1}\right), \tag{8}$$

where $Q(r, \rho)$ is the quantile function defining $P_\rho(X \leq Q(r, \rho)) = r$ for probability distribution ρ and random variable $X \sim \rho$, and c is a constant for computing the quantile used in the theoretical analysis of Bayes-UCB, with $c = 0$ empirically preferred (Kaufmann et al., 2012). For the BBB variants of Bayes-UCB and TS, we need simply to supply our designed backdoor adjustment prior Π^0 and make appropriate Bayesian updates to obtain the posterior Π^t.

We present our proposed BBB methodology applied to Bayes-UCB, TS, and UCB in Algorithm 1.

Algorithm 1 `BBB-Alg`$(T, \mathcal{A}, \mathcal{D}_0, c)$

Require: Horizon T, action set $\mathcal{A}$, observational data $\mathcal{D}_0$, confidence level c

Compute the observational parent set posteriors (4)
for all $a \in \mathcal{A}$ and $\mathbf{Z} \subseteq \mathbf{X} \setminus X_{\langle a \rangle}$ **do**
    Compute $\pi^0_{a|\mathbf{Z}}$ according to (2)
end for
for all $t = 1, \ldots, T$ **do**
    for all $a \in \mathcal{A}$ **do**
        Compute criterion $U_a(t)$ according to `Alg`:
        - `Bayes-UCB`: $U_a(t) = Q(1 - 1/(t(\log T)^c), \pi_a^{t-1})$ as in (8)
        - `TS`: Sample $U_a(t) \leftarrow \pi_a^{t-1}(\mu_a)$
        - `UCB`: $U_a(t) = \mathrm{E}_{\pi_a^{t-1}}[\mu_a] + c\sqrt{\mathrm{Var}_{\pi_a^{t-1}}[\mu_a] \log(t)}$ as in (7)
    end for
    Pull arm $a_t \in \operatorname{argmax}_{a \in \mathcal{A}} U_a(t)$ and observe $\mathcal{D}^{(t)}$
    for all $\mathbf{Z} \subseteq \mathbf{X} \setminus X_{\langle a \rangle}$ where $a = a_t$ **do**
        Update $\pi^t_{a|\mathbf{Z}}$ according to $\pi^t_{a|\mathbf{Z}}(\theta_a) \propto p_{\theta_a}(y_t)\ \pi^{t-1}_{a|\mathbf{Z}}(\theta_a)$
    end for
    Compute or update the parent set posteriors (4)
end for

5 Implementation Details

5.1 Nonparametric Discrete Setting

We now detail the application of our proposed construction of $\pi^0_{a|\mathbf{Z}}$ to the setting where the CPDs are multinomials, with each variable $X_i \in \mathbf{X}$ probabilistically attaining its states depending on the attained state configuration of its parents $\mathbf{Pa}_i^{\mathcal{G}}$. The reward variable $Y = X_p$ is a binary variable with $\mathrm{Dom}(Y) = \{0, 1\}$. If $Y \notin \mathbf{Z}$, μ_a may be estimated with observational data through straightforward empirical estimation of (1):

$$\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z}) = \hat{P}[Y = 1 \mid do(X_{\langle a \rangle} = x_a)] = \frac{1}{n_0} \sum_{\mathbf{z}} \frac{n_0[1, x_a, \mathbf{z}] n_0[\mathbf{z}]}{n_0[x_a, \mathbf{z}]}, \tag{9}$$

where $n_0[1, x_a, \mathbf{z}]$ represents the number of the n_0 samples of $\mathcal{D}_0$ in which $Y = 1$, $X = x_a$, and $\mathbf{Z} = \mathbf{z}$, with corresponding definitions for $n[x_a, \mathbf{z}]$ and $n[\mathbf{z}]$.

Analysis of the sampling distribution of (9) is admittedly challenging. To design an appropriately weighted informative prior as proposed in (2), we require some characterization of the sampling variability of $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$. Hence, we derive an approximation of the variance of (9), $\hat{\mathrm{SE}}^2[\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})]$. We accomplish this by first re-expressing the joint counts $n_0[\cdot]$ as sums of elements of a multinomial random vector. The term within the sum may then be expressed as a product and ratio of intersecting random quantities, which we approximate

through a first-order Taylor series expansion. The details of the derivation are delegated to Appendix D. It is appropriate to acknowledge that Maiti et al. (2021) proposed a provably unbiased strategy for empirical estimation of (1) through splitting the sample into independent partitions. However, this approach suffers from severe loss of precision through what some may consider underutilization of the observed data. In our experiments detailed in Appendix C.1, we find the empirical performance of (9) in our applications to be acceptable. We additionally provide extensive empirical validation of our derived approximation, demonstrating coverage probabilities comparable to empirical estimates of the sampling variability for modest sample sizes.

Since the reward variable under arm a is a Bernoulli random variable with probability parameter $\mu_a = P[Y = 1 \mid do(X_{\langle a \rangle} = x_a)]$, we assume a conjugate prior $\pi^0_{a|\mathbf{Z}} = \text{Beta}(\alpha_0, \beta_0)$ for $\theta_a = \mu_a$ designed according to (2), resulting in prior hyperparameters

$$\alpha_0 = \hat{\mu}_{a,\text{bda}}(\mathbf{Z}) \left(\frac{\hat{\mu}_{a,\text{bda}}(\mathbf{Z})[1 - \hat{\mu}_{a,\text{bda}}(\mathbf{Z})]}{\hat{\text{SE}}^2 [\hat{\mu}_{a,\text{bda}}(\mathbf{Z})]} - 1 \right), \qquad \beta_0 = \alpha_0 \left(\frac{1 - \hat{\mu}_{a,\text{bda}}(\mathbf{Z})}{\hat{\mu}_{a,\text{bda}}(\mathbf{Z})} \right).$$

5.2 Gaussian Unit Deviation Setting

In this section, we consider the setting where the causal model may be expressed as a set of Gaussian structural equations:

$$X_j = \sum_{i=1}^{p} \beta_{ij} X_i + \varepsilon_j, \quad \varepsilon_j \sim \text{N}(0, \sigma_j^2), \quad j = 1, \ldots, p. \tag{10}$$

There is no intercept term, which is analogous to having prior knowledge of the observational means, and we consider interventions $x_a \in \{-1, 1\}$, which may be interpreted as investigating unit deviations from the observational means. In this setting, the causal effect of $X_{\langle a \rangle}$ on Y is given by

$$\psi_{\langle a \rangle} \coloneqq \text{E}[Y \mid do(X_{\langle a \rangle} = x' + 1)] - \text{E}[Y \mid do(X_{\langle a \rangle} = x')]$$

for any $x' \in \mathbb{R}$, derived via a special case of (1). Note that in our problem formulation, $Y \mid do(X_{\langle a \rangle} = 1)$ and $-Y \mid do(X_{\langle a \rangle} = -1)$ are identically distributed, so all data generated from interventions on $X_{\langle a \rangle}$ may be combined to estimate $\psi_{\langle a \rangle}$. Since $\mu_a = x_a \psi_{\langle a \rangle}$, we focus our efforts on estimating and modeling $\psi_{\langle a \rangle}$. Accordingly, in constructing our priors using intervention calculus, we design priors $\pi^0_{\langle a \rangle|\mathbf{Z}}$ for $\theta_{\langle a \rangle}$ corresponding to estimating $\psi_{\langle a \rangle}$, and allow $\pi^0_{a|\mathbf{Z}}$ to be the induced priors for θ_a corresponding to $\mu_a = \psi_{\langle a \rangle} x_a$, detailed as follows.

If $Y \notin \mathbf{Z}$, then a consistent estimator of $\psi_{\langle a \rangle}$, denoted $\hat{\psi}_{\langle a \rangle,\text{bda}}$, may be obtained with observational data by the least squares regression

$$Y = \psi_{\langle a \rangle} X_{\langle a \rangle} + \boldsymbol{\gamma}^\top \mathbf{Z} + e, \quad e \sim \text{N}(0, \eta^2), \tag{11}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathbf{Z}|}$ is the coefficients of the parents $\mathbf{Z}$ (Maathuis et al., 2009; Pensar et al., 2020), and some dependence on a is omitted for simplicity. Correspondingly, we express the desired interventional distribution as $Y \mid do(X_{\langle a \rangle} = x_a) \sim \text{N}(\psi_{\langle a \rangle} x_a, \omega^2)$. Claiming no prior knowledge of the interventional variance, we assume a Normal-inverse-gamma (N-Γ^{-1}) conjugate prior $\pi^0_{\langle a \rangle|\mathbf{Z}}$ for $\theta_{\langle a \rangle} = (\psi_{\langle a \rangle}, \omega^2)$:

$$\psi_{\langle a \rangle} \mid \omega^2 \sim \text{N}\left(m_0, \omega^2 \nu_0^{-1}\right), \quad \omega^2 \sim \Gamma^{-1}(u_0, v_0). \tag{12}$$

Since in general, the residual variance η^2 in (11) is not equivalent to ω^2, we propose the following to estimate ω^2 from observational data.

Proposition 1. *Suppose that* $\mathbf{X}$ *follows the causal structural equation model (SEM) in* (10)*. Let* $Y, X \in \mathbf{X}$*, and denote by* ψ *the causal effect of* X *on* Y*. Then for any* $x \in \text{Dom}(X)$*,*

$$\text{Var}[Y \mid do(X = x)] = \text{Var}[Y - \psi X].$$

Note that the variance on the right side is with respect to the observational distribution of $\mathbf{X}$. Intuitively, subtracting by ψX negates the noise variances σ^2 in (10) propogated through and from X to Y. We include a detailed proof for Proposition 1 in Appendix A.

Thus, to estimate ω^2 from $\mathcal{D}_0$, we propose the estimator $\hat{\omega}^2 = \sum_i (\tilde{y}_i - \bar{\tilde{y}})^2/(n_0 - |\mathbf{Z}| - 2)$ where $\tilde{y}_i$ are realizations of $\tilde{Y} \coloneqq Y - \hat{\psi}_{\langle a \rangle, \text{bda}}(\mathbf{Z}) X_{\langle a \rangle}$ in $\mathcal{D}_0$, and $n_0 - |\mathbf{Z}| - 2$ is the degrees of freedom resulting from estimating $\bar{\tilde{y}}$ in addition to $|\mathbf{Z}| + 1$ coefficients in (11). Accordingly, we design the prior $\omega^2 \sim \Gamma^{-1}(u_0, v_0)$ to have prior mean $\mathrm{E}[\omega^2] = v_0/(u_0 - 1) = \hat{\omega}^2$, resulting in hyperparameters $u_0 = (n_0 - |\mathbf{Z}|)/2$ and $v_0 = \sum_i (\tilde{y}_i - \bar{\tilde{y}})^2/2$. After marginalizing out ω^2, $\psi_{\langle a \rangle} \sim t_{2a_0}(m_0, v_0(u_0\nu_0)^{-1})$, so we set $\mathrm{E}_{\pi^0_{\langle a \rangle | \mathbf{Z}}}[\psi_{\langle a \rangle}] = m_0 = \hat{\psi}_{\langle a \rangle, \text{bda}}(\mathbf{Z})$ and solve to obtain $\nu_0 = v_0/(u_0 \hat{\mathrm{SE}}^2[\hat{\psi}_{\langle a \rangle, \text{bda}}(\mathbf{Z})])$.

To maximally utilize the ensemble data, we further generalize the estimation of $\psi_{\langle a \rangle}$ via regression in (11) to include eligible samples of intervention data. This is achieved through the following proposition, which we prove in Appendix A. This result does not rely on any parametric assumptions for the underlying causal model, assuming simply that $\mathbf{X}$ follows a general linear SEM with DAG $\mathcal{G}$ (Pearl, 2000).

Proposition 2. *Suppose that $\mathbf{X}$ follows a linear SEM with a DAG $\mathcal{G}$, and $X, Y \in \mathbf{X}$. Suppose that $W \in \mathbf{X} \setminus \{X, Y\}$ does not block any directed path from X to Y in $\mathcal{G}$. Then for any $w \in \mathbb{R}$,*

$$\frac{\partial}{\partial x} \mathrm{E}[Y \mid do(X = x)] = \frac{\partial}{\partial x} \mathrm{E}[Y \mid do(X = x), do(W = w)].$$

Proposition 2 asserts a simple graphical criterion which, if satisfied, defines an avenue by which information can be shared between arms. In our work, we check the graphical criterion for estimating the causal effect of $X_{\langle a \rangle}$ on Y with interventional data generated from intervening on X_j as follows. Using another algorithm proposed by Pensar et al. (2020) for computing exact ancestor posterior probabilities, we consider the criterion satisfied at time step t if the event that X_j blocks a directed path from $X_{\langle a \rangle}$ to Y has low posterior probability:

$$P(X_{\langle a \rangle} \rightsquigarrow X_j \rightsquigarrow Y \mid \mathcal{D}[t]) \leq \min\{P(X_{\langle a \rangle} \rightsquigarrow X_j \mid \mathcal{D}[t]), P(X_j \rightsquigarrow Y \mid \mathcal{D}[t])\} \leq \tau \tag{13}$$

where $X_j \rightsquigarrow Y$ denotes that X_j is an ancestor of Y and the threshold is set to $\tau = 0.1$ in our application. If (13) holds at time step t, we combine the observational data and the data from interventions on X_j when conducting the regression (11). While independent samples of observational and interventional data are not guaranteed to have identically distributed errors in the regression (11), we provide extensive empirical validation of our proposed regression with ensemble data in Appendix C.2, confirming indistinguishable performance for the purposes of estimating $\psi_{\langle a \rangle}$ and its sampling variability compared to that of purely observational data.

6 Numerical Experiments

We conducted extensive numerical experiments to empirically validate our proposed methodology. For our experiments, we generated CBN models with $p = 10$ variables. The structures were randomly generated according to a process adapted from de Kroon et al. (2022), and the reward variable was designated to have $|\mathbf{Pa}_p^{\mathcal{G}}| = 3$ parents. The conditional probability distributions of each CBN were likewise generated randomly. Atomic interventions as described in Section 5 were allowed on all variables excluding the reward variable, with the discrete variables assumed to be binary, resulting in $|\mathcal{A}| = 2(p - 1) = 18$ actions. Additional experimental details are provided in Appendix B.

6.1 Cumulative Regret Comparisons

We evaluated our BBB methodology against algorithms designed to optimize cumulative regret, including popular standard MAB algorithms Bayes-UCB, TS, and UCB that do not utilize causal assumptions (see Section 4). Additionally, we compared against what can be interpreted as a highly optimistic version of the central node approach by Lu et al. (2021), introduced in Section 1, by presupposing knowledge of the

direct causes of the reward variable. In particular, for Bayes-UCB*, TS*, and UCB*, we executed the respective algorithms over the reduced action set $\mathcal{A}' = \{a \in \mathcal{A} : \langle a \rangle \in \mathbf{Pa}_p^{\mathcal{G}}\}$. Accordingly, for the cases where $\langle a^* \rangle \notin \mathbf{Pa}_p^{\mathcal{G}}$, we redefined the optimal intervention to $a^* = \text{argmax}_{a \in \mathcal{A}'} \mu_a$ when evaluating the regret of TS* and (Bayes-)UCB*.

Using the process described above, we generated 100 CBN models for each distributional setting. For each CBN, we executed the competing methods 10 times but our BBB methods only 5 times due to their greater computational expense, with $T = 5000$ time steps. The results presented are averaged across all simulations for each time step, with the cumulative regret normalized by the optimal reward μ^* to ensure that each CBN model contributes comparably. In preference to the competing methods, we tuned for their best-performing parameters where relevant and applied them to our BBB implementations.

The empirical cumulative regret results in Figure 1 demonstrate that in both the discrete and Gaussian settings and for all algorithms, our BBB methodology is able to reliably outperform the non-causal variants with finite samples of observational data. The improvement increases monotonically with increasing sample sizes of observational data (n_0). While corresponding variants of Bayes-UCB and TS perform comparably, UCB achieves substantially lower regret because the parameter c in (6) was tuned to maintain a balance between exploration and exploitation that is most empirically preferred. In particular, UCB is able to avoid excessive exploration by scaling its padding term with a relatively small constant, whereas Bayes-UCB maintains a relatively high minimum exploration rate according to its formulation in (8), as does TS.

In comparison to the optimistic central node versions of the algorithms, BBB generally achieves lower cumulative regret with $n_0 \geq 800$ in the discrete setting and $n_0 \geq 40$ in the Gaussian setting. Recall that,

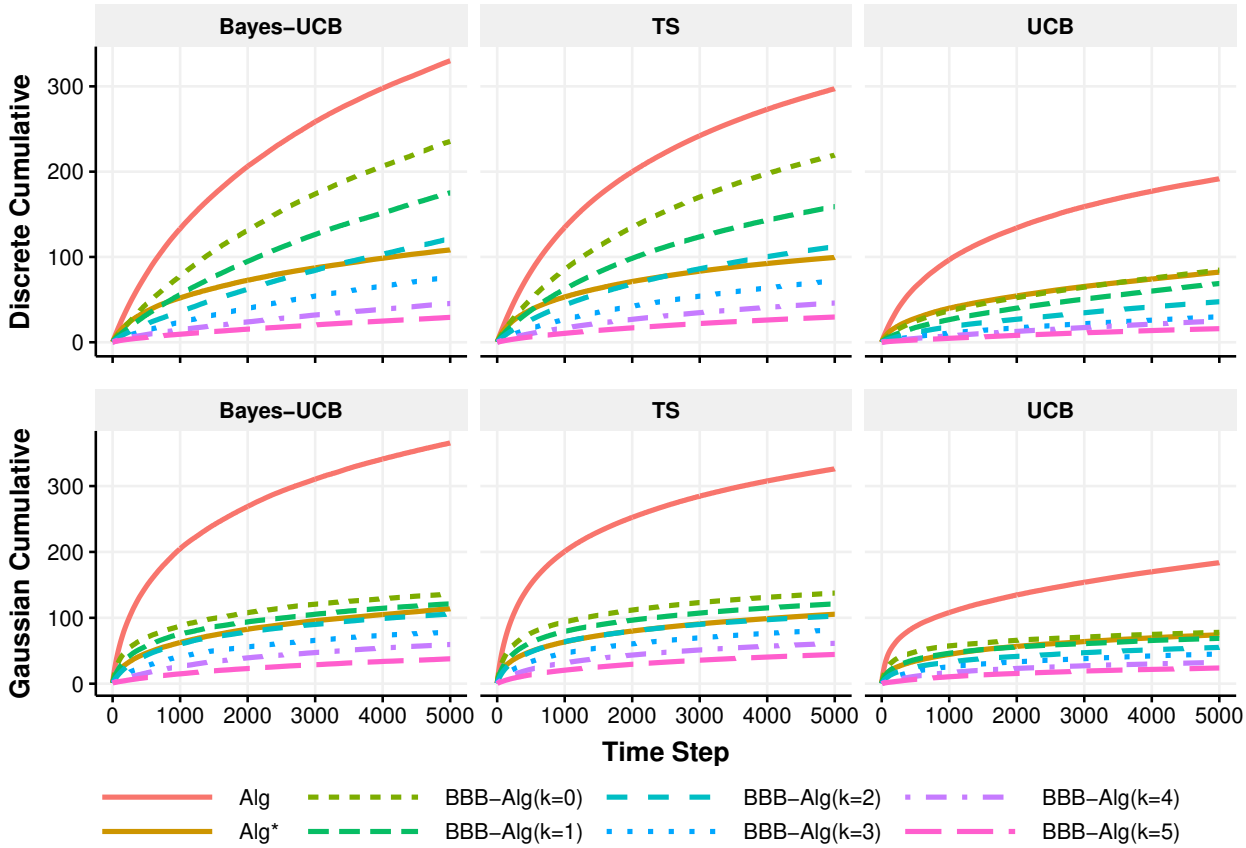

Figure 1: Cumulative regret results against $T = 5000$ time steps comparing Alg, Alg*, and BBB-Alg for Alg $\in \{\text{Bayes-UCB}, \text{TS}, \text{UCB}\}$. BBB methods were executed with $n_0 = 100 \cdot 2^k$ in the discrete setting and $n_0 = 10 \cdot 2^k$ in the Gaussian setting

in practical applications, the central node approach relies on the availability of large-sample observational data as well as a sequence of interventions to recover the reward generating variables $\mathbf{Pa}_p^{\mathcal{G}}$. Based on our simulation settings, this reduces the action set from $|\mathcal{A}| = 18$ arms to only $|\mathcal{A}'| = 2|\mathbf{Pa}_p^{\mathcal{G}}| = 6$ arms, and we additionally restrict $a^* \in \mathcal{A}'$ to evaluate the regret. Furthermore, the regret results reported for these methods do not include the interventions required to identify $\mathbf{Pa}_p^{\mathcal{G}}$, thus representing a kind of best case scenario for the central node approach. In contrast, our methodology derives substantial benefit from modest amounts of observational data samples n_0.

Indeed, we find that our BBB methods are able to perform competitively against the competing methods even when the latter are given n_0 time steps to explore arms before incurring regret. To compensate for the fact that BBB utilizes n_0 samples of observational data prior to investigating arms, we present the results where the competing algorithms TS(*) and (Bayes-)UCB(*) are given a head start of $n_0 \in \{100 \cdot 2^k : k = 0, 1, \ldots, 5\}$ time steps to explore arms before incurring regret. The results for the discrete setting are shown in Figure 2. The Gaussian results are omitted because $n_0 \leq 320$ is relatively small, so the head start does not offer substantial benefit to the competing methods. In all cases, BBB still significantly outperforms the standard algorithms TS and (Bayes-)UCB given the head start. Given sufficient samples of observational data, BBB still performs comparably to if not better than the optimistic central node variants in terms of cumulative regret, for which the head start is only an additional unwarranted advantage given that they already require significantly more observational data as well as additional interventions.

6.2 Structure Identification

In addition to the cumulative regret performance, it is of interest to consider the structure identification behavior of the BBB approach in our experiments. We measure the concentration of the posterior probability across DAGs $\mathcal{G}$ with respect to the underlying causal graph $\mathcal{G}^*$ using the edge support sum of absolute errors

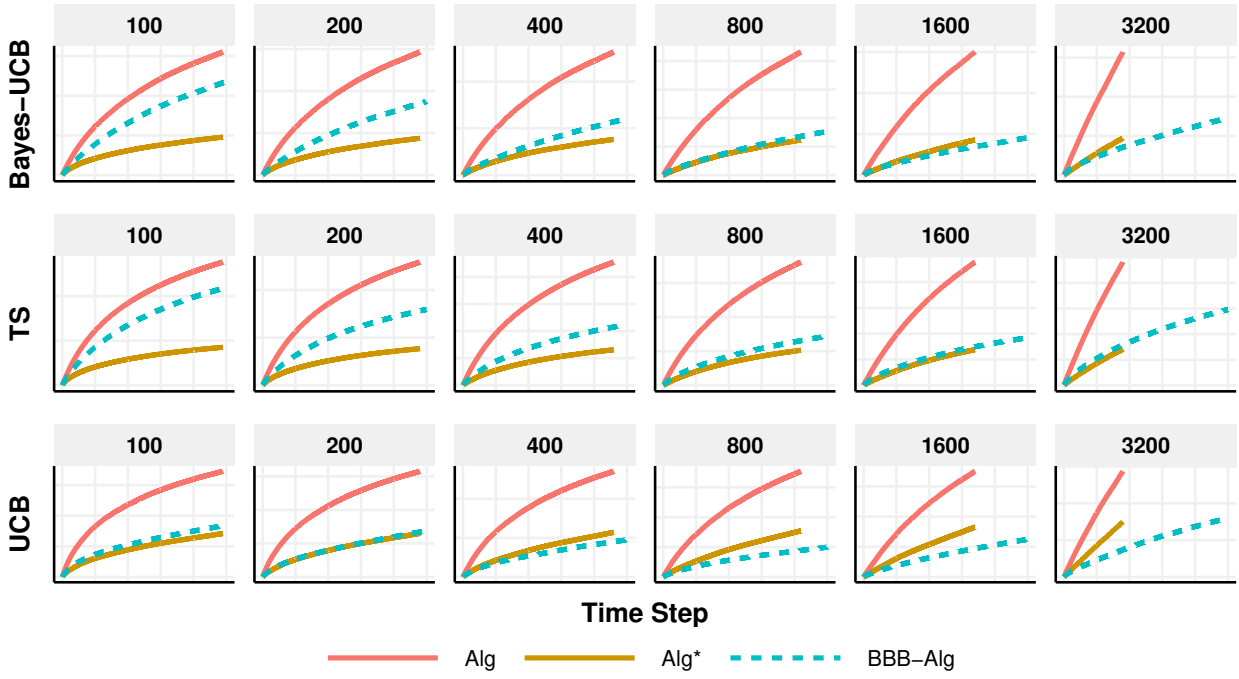

Figure 2: Discrete cumulative regret for Alg $\in \{\text{Bayes-UCB}, \text{TS}, \text{UCB}\}$ with a head start of $n_0 \in \{100 \cdot 2^k : k = 0, 1, \ldots, 5\}$ time steps for competing methods

(ESSAE), which is given at time t by

$$\sum_{i=1}^{p} \sum_{j \neq i} \left| P(j \in \mathbf{Pa}_i^{\mathcal{G}} \mid \mathcal{D}[t]) - \mathbb{1}\left\{ j \in \mathbf{Pa}_i^{\mathcal{G}^*} \right\} \right|.$$

This quantity may be understood as a probabilistic version of the structural hamming distance, a common metric in Bayesian network structure learning literature. Lower ESSAE corresponds to greater concentration of the posterior probability around the causal graph $\mathcal{G}^*$. The results are provided in Figure 3.

In the discrete results for BBB-Bayes-UCB and BBB-TS, the initial ESSAE is unsurprisingly lower for the larger sample sizes, but the trend quickly reverses as the time steps progress. This effect is also observed occurring in the Gaussian results, but at an accelerated pace. This behavior is perhaps best understood in complement to the cumulative regret results in Figure 1. If P is faithful to $\mathcal{G}$, then if n_0 is large, the structure prior $P(\mathcal{G} \mid \mathcal{D}_0)$ is expected to concentrate around the Markov equivalence class, which entails identification of the skeleton and in general, partial identification of the orientations. Additionally, the conditional priors $\pi^0_{a|\mathbf{Z}}$ are precise models, allowing BBB to quickly identify and select arm(s) $a \in \mathcal{A}$ with small regret $\mu^* - \mu_a$, which has the effect of clarifying the orientation of edges incident to such $X_{\langle a \rangle}$. The policies take no interest in determining the orientation of the remaining edges if the uncertainty does not indicate potential to identify more profitable actions. In contrast, when n_0 is small, the greater uncertainty in both the structure prior and the conditional priors encourage the exploration of many different arms, thus incurring greater cumulative regret. In addition to clarifying the orientation of the incident edges, selecting arm(s) $a \in \mathcal{A}$ contributes to identifying the direct edge connections excluding those from $\mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}$ to $X_{\langle a \rangle}$, as can be seen in (5). Thus, the skeleton is recovered and more edge orientations are identified than in the case where n_0 is large, achieving lower ESSAE at the cost of greater cumulative regret. Notably, while this reversal appears to be absent in

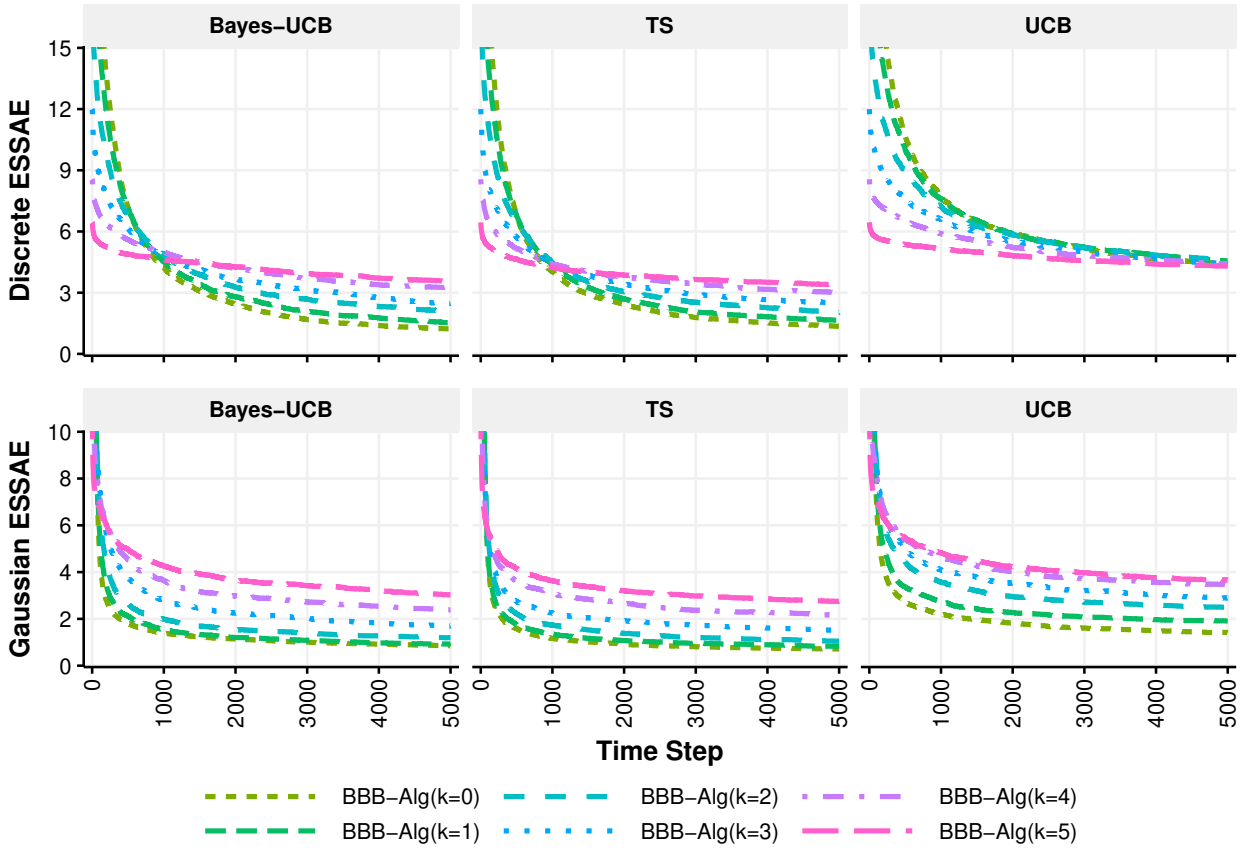

Figure 3: Discrete ($n_0 = 100 \cdot 2^k$) and Gaussian ($n_0 = 10 \cdot 2^k$) results of the ESSAE of the full graph structure for BBB-Alg, Alg $\in$ {Bayes-UCB, TS, UCB}

the discrete results for BBB-UCB, in actuality it has simply not yet been realized even after $T = 5000$ time steps due to the small exploration constant c in (7).

7 Discussion

In this paper, we proposed the BBB framework for enhancing experimental investigations with observational data. BBB consists of an aggregation of various strategies for estimating and modeling the parameters of interest with jointly interventional and observational data in order to efficiently utilize all available data to inform exploitation and exploration. Applied in our methodology but also of independent interest, we derived a well-performing approximation for the variance of the discrete backdoor adjustment estimator, and in the Gaussian setting, we characterized the interventional variance using the observational distribution and proposed a simple graphical criterion for sharing information between arms. We empirically validated our proposed algorithms through extensive numerical experiments against standard MAB algorithms as well as a generously optimistic version of a recently proposed CB approach.

A substantial limitation of our method is the computational expense of computing the parent set probabilities and the conditional parameter posteriors corresponding to each parent set. However, such an investment is justified when interventions are particularly expensive or time-consuming. In these settings, it is crucial to utilize all available evidence, motivating future work in scaling our methods to feasibly operate for larger causal systems. Note that parent set probabilities with trivial support may be thresholded to zero and the corresponding conditional posteriors need not be updated. This significantly reduces the computational load of causal parameter modeling, especially for systems with sparse structures, so the structure posterior tends to be the limiting factor. However, we emphasize that exact computation is not always necessary. Instead, approximations may be attained efficiently through MCMC, and posterior sampling methods such as BBB-TS need only to sample a single DAG from the posterior distribution of causal graphs. Kuipers et al. (2022) applied their hybrid MCMC approach, in which DAGs are sampled from a restricted space estimated by a structure learning algorithm, on systems with up to 200 variables.

In addition to scaling, there are a number of interesting directions for future investigations. Though we validated our proposed BBB methodology numerically by demonstrating compelling empirical performance against well-studied algorithms, we leave formal theoretical analysis of the BBB framework to future work. Finally, it would be interesting to consider how to share information between arms in the discrete setting as in Proposition 2 with an equally simple graphical criterion.

References

Rajeev Agrawal. Sample mean based index policies by $O(\log n)$ regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078, 1995. https://doi.org/10.2307/1427934.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine learning*, 47(2):235–256, 2002. https://doi.org/10.1023/A:1013689704352.

Donald A Berry and Bert Fristedt. Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability). *London: Chapman and Hall*, 5(71-87):7–7, 1985. https://link.springer.com/book/10.1007/978-94-015-3711-7.

Arnoud de Kroon, Joris Mooij, and Danielle Belgrave. Causal bandits without prior knowledge using separating sets. In *First Conference on Causal Learning and Reasoning*, 2022. https://openreview.net/forum?id=50eDSLScz7r.

Daniel Eaton and Kevin Murphy. Exact Bayesian structure learning from uncertain interventions. In Marina Meila and Xiaotong Shen (eds.), *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, volume 2 of *Proceedings of Machine Learning Research*, pp. 107–114, San Juan, Puerto Rico, 21–24 Mar 2007. PMLR. https://proceedings.mlr.press/v2/eaton07a.html.

Nir Friedman and Daphne Koller. Being Bayesian About Network Structure. A Bayesian Approach to Structure Discovery in Bayesian Networks. *Machine learning*, 50(1):95–125, 2003. https://doi.org/10.

1023/A:1020249912095.

Kristjan Greenewald, Dmitriy Katz, Karthikeyan Shanmugam, Sara Magliacane, Murat Kocaoglu, Enric Boix Adsera, and Guy Bresler. Sample Efficient Active Learning of Causal Trees. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. https://proceedings.neurips.cc/paper/2019/file/5ee5605917626676f6a285fa4c10f7b0-Paper.pdf.

David Heckerman, Dan Geiger, and David M Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995. https://doi.org/10.1007/BF00994016.

Emilie Kaufmann, Olivier Cappe, and Aurelien Garivier. On Bayesian Upper Confidence Bounds for Bandit Problems. In Neil D. Lawrence and Mark Girolami (eds.), *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pp. 592–600, La Palma, Canary Islands, 21–23 Apr 2012. PMLR. https://proceedings.mlr.press/v22/kaufmann12.html.

Jack Kuipers and Giusi Moffa. Partition MCMC for Inference on Acyclic Digraphs. *Journal of the American Statistical Association*, 112(517):282–299, 2017. https://doi.org/10.1080/01621459.2015.1133426.

Jack Kuipers, Polina Suter, and Giusi Moffa. Efficient Sampling and Structure Learning of Bayesian Networks. *Journal of Computational and Graphical Statistics*, 0(0):1–12, 2022. https://doi.org/10.1080/10618600.2021.2020127.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985. ISSN 0196-8858. https://doi.org/10.1016/0196-8858(85)90002-8.

Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal Bandits: Learning Good Interventions via Causal Inference. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. https://proceedings.neurips.cc/paper/2016/file/b4288d9c0ec0a1841b3b3728321e7088-Paper.pdf.

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper/2018/file/c0a271bc0ecb776a094786474322cb82-Paper.pdf.

Yangyi Lu, Amirhossein Meisami, Ambuj Tewari, and William Yan. Regret Analysis of Bandit Problems with Causal Background Knowledge. In Jonas Peters and David Sontag (eds.), *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pp. 141–150. PMLR, 03–06 Aug 2020. https://proceedings.mlr.press/v124/lu20a.html.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal Bandits with Unknown Graph Structure. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24817–24828. Curran Associates, Inc., 2021. https://proceedings.neurips.cc/paper/2021/file/d010396ca8abf6ead8cacc2c2f2f26c7-Paper.pdf.

Marloes H. Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A):3133 – 3164, 2009. https://doi.org/10.1214/09-AOS685.

David Madigan, Jeremy York, and Denis Allard. Bayesian Graphical Models for Discrete Data. *International Statistical Review / Revue Internationale de Statistique*, 63(2):215–232, 1995. ISSN 03067734, 17515823. http://www.jstor.org/stable/1403615.

Aurghya Maiti, Vineet Nair, and Gaurav Sinha. Causal Bandits on General Graphs. *arXiv preprint arXiv:2107.02772*, 2021. https://arxiv.org/abs/2107.02772.

Vineet Nair, Vishakha Patil, and Gaurav Sinha. Budgeted and Non-Budgeted Causal Bandits. In Arindam Banerjee and Kenji Fukumizu (eds.), *Proceedings of The 24th International Conference on Artificial Intel-*

ligence and Statistics, volume 130 of *Proceedings of Machine Learning Research*, pp. 2017–2025. PMLR, 13–15 Apr 2021. https://proceedings.mlr.press/v130/nair21a.html.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 12 1995. ISSN 0006-3444. https://doi.org/10.1093/biomet/82.4.669.

Judea Pearl. *Causality*. Cambridge University Press, 2000.

Johan Pensar, Topi Talvitie, Antti Hyttinen, and Mikko Koivisto. A Bayesian Approach for Estimating Causal Effects from Observational Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):5395–5402, Apr. 2020. https://ojs.aaai.org/index.php/AAAI/article/view/5988.

Robert W Robinson. Counting unlabeled acyclic digraphs. In Charles H. C. Little (ed.), *Combinatorial Mathematics V*, pp. 28–43. Springer, Berlin, Heidelberg, 1977. ISBN 978-3-540-37020-8. https://doi.org/10.1007/BFb0069178.

Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. MIT Press, second edition, 2018.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 12 1933. ISSN 0006-3444. https://doi.org/10.1093/biomet/25.3-4.285.

Akihiro Yabe, Daisuke Hatano, Hanna Sumita, Shinji Ito, Naonori Kakimura, Takuro Fukunaga, and Ken-ichi Kawarabayashi. Causal Bandits with Propagating Inference. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5512–5520. PMLR, 10–15 Jul 2018. https://proceedings.mlr.press/v80/yabe18a.html.

A Proofs

In this section, we prove Proposition 1 and Proposition 2.

Proof of Proposition 1. Let $\mathbf{Z}$ be the parent set of X. Then by a special case of (1),

$$\begin{aligned} p(y \mid do(x)) &= \int p(y \mid x, \mathbf{z}) p(\mathbf{z}) d\mathbf{z} \\ &= \int \phi(y \mid \psi x + \boldsymbol{\gamma}^\top \mathbf{z}, \sigma^2) \phi(\mathbf{z} \mid 0, \Sigma_{\mathbf{Z}}) d\mathbf{z} \\ &= \phi(y \mid \psi x, \boldsymbol{\gamma}^\top \Sigma_{\mathbf{Z}} \boldsymbol{\gamma} + \sigma^2), \end{aligned}$$

where $\phi(\cdot \mid \mu, \Sigma)$ is the probability density function of $\mathrm{N}(\mu, \Sigma)$ and $\Sigma_{\mathbf{Z}}$ is the covariance matrix of $\mathbf{Z}$. Thus,

$$Y \mid do(X = x) \sim \mathrm{N}(\psi x, \boldsymbol{\gamma}^\top \Sigma_{\mathbf{Z}} \boldsymbol{\gamma} + \sigma^2).$$

Now representing $[Y \mid X, \mathbf{Z}]$ by a linear regression:

$$Y = \psi X + \boldsymbol{\gamma}^\top \mathbf{Z} + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2) \perp \mathbf{Z} \sim N(0, \Sigma_{\mathbf{Z}})$. Then we have

$$\begin{aligned} \mathrm{Var}(Y - \psi X) &= \mathrm{Var}(\boldsymbol{\gamma}^\top \mathbf{Z} + \varepsilon) \\ &= \boldsymbol{\gamma}^\top \Sigma_{\mathbf{Z}} \boldsymbol{\gamma} + \sigma^2 = \mathrm{Var}(Y \mid do(X = x)). \end{aligned}$$

□

Proof of Proposition 2. The result follows straightforwardly from a simple graphical argument. Let $\Xi^{\mathcal{G}}_{XY}$ denote the distinct directed paths from X to Y in the causal graph $\mathcal{G}$ given the model (10), where $\xi \in \Xi^{\mathcal{G}}_{XY}$

consists of all the directed edges $i \to j \in \mathbf{E}$ on the given path from X to Y. Then the causal effect of X on Y can be expressed as the sum of propagated direct effects along all directed paths from X to Y:

$$\psi_{XY} \coloneqq \frac{\partial}{\partial x}\mathrm{E}[Y \mid do(X = x)] = \sum_{\xi \in \Xi^{\mathcal{G}}_{XY}} \prod_{i \to j \in \xi} \beta_{ij}.$$

We denote the variables under the intervention $do(W = w)$, $w \in \mathbb{R}$ as $\tilde{\mathbf{X}}$, with resulting causal model

$$\tilde{X}_j = \sum_{i=1}^{p} \tilde{\beta}_{ij}\tilde{X}_i + \tilde{\varepsilon}_j, \quad j = 1 \ldots, p,$$

where

$$\tilde{\beta}_{ij} = \begin{cases} 0 & \text{if } X_j = W \\ \beta_{ij} & \text{otherwise,} \end{cases} \qquad \tilde{\varepsilon}_j = \begin{cases} w & \text{if } X_j = W \\ \varepsilon_j & \text{otherwise.} \end{cases}$$

The corresponding causal graph for $\tilde{\mathbf{X}}$ is the mutilated graph $\tilde{\mathcal{G}}$ resulting from deleting all edges into W. The causal effect of $\tilde{X}$ on $\tilde{Y}$ is then

$$\psi_{\tilde{X}\tilde{Y}} \coloneqq \frac{\partial}{\partial x}\mathrm{E}[\tilde{Y} \mid do(\tilde{X} = x)] = \frac{\partial}{\partial x}\mathrm{E}[Y \mid do(X = x), do(W = w)] = \sum_{\xi \in \Xi^{\tilde{\mathcal{G}}}_{XY}} \prod_{i \to j \in \xi} \tilde{\beta}_{ij}.$$

Since W does not block any directed path from X to Y, the mutilated graph $\tilde{\mathcal{G}}$ retains all the directed paths from X to Y in $\mathcal{G}$, so $\Xi^{\tilde{\mathcal{G}}}_{XY} = \Xi^{\mathcal{G}}_{XY}$. By the same reasoning, $\tilde{\beta}_{ij} = \beta_{ij}$ for all $i \to j \in \xi$ where $\xi \in \Xi^{\tilde{\mathcal{G}}}_{XY}$. Therefore, for any $w \in \mathbb{R}$,

$$\frac{\partial}{\partial x}\mathrm{E}[Y \mid do(X = x)] = \frac{\partial}{\partial x}\mathrm{E}[Y \mid do(X = x), do(W = w)].$$

□

B Experimental Details

In this section, we include details regarding the experiments discussed in Section 6. The complete code for reproducing our results has been made available at the following link:

`https://anonymous.4open.science/r/bcb0`

For our experiments, we generated CBN models for $p = 10$ variables with reward variable $Y = X_p$. In order to investigate interesting structures with diverse non-trivial confounding relationships, we randomly generated graph structures using the following process adapted from de Kroon et al. (2022). Given a fixed topological sort of the variables $X_1 \prec \cdots \prec X_p$ where the reward variable is $Y = X_p$, we sequentially considered nodes in reverse topological order: $i = p-1, \ldots, 1$. We uniformly sampled the maximum out-degree of X_i, denoted d_i, from 1 to $p - i$. Then, for d_i times, we randomly selected X_j from $\{X_j \in \mathbf{X} : X_i \prec X_j\}$, adding $X_i \to X_j$ to the graph only if the edge was not already present and $|\mathbf{Pa}^{\mathcal{G}}_j| < 3$. We imposed an additional requirement that $|\mathbf{Pa}^{\mathcal{G}}_p| = 3$, randomly adding parents if necessary. If the generated structure consisted of multiple disconnected components, we rejected the structure and reattempted the process.

The conditional probability distributions of each CBN were likewise generated randomly. For discrete networks, the variables were all assumed to be binary, and the conditional probability tables were randomly generated uniformly and normalized, and were accepted only if for every edge $X_j \to X_i$, there is a sufficiently large causal effect, with $|P[X_i = x_i \mid do(X_j = x_j)] - P(X_i = x_i)| \geq 0.05$ for some $x_i \in \mathrm{Dom}(X_i)$ and $x_j \in \mathrm{Dom}(X_j)$. Additionally, we required the marginal probability of any single discrete level to be at least 0.01, and that the reward signal of the optimal intervention a^* be sufficiently large with respect to

the observational mean: $\mu^* - \mathrm{E}[Y] \geq 0.05$. For Gaussian networks, according to the model expressed in (10), we sampled coefficients uniformly from $[-1, -0.5] \cup [0.5, 1]$ for $X_i \in \mathbf{Pa}_j^{\mathcal{G}}$ and standard deviations from $[\sqrt{0.5}, 1]$, and we normalized the system to have unit variance. Note that in the Gaussian setting, there are effectively $|\mathcal{A}| = 9$ actions given that interventional data on the same variable may be combined as discussed in Section 5.2, which we implement for the competing methods as well. We found that $\langle a^* \rangle \in \mathbf{Pa}_p^{\mathcal{G}}$ held for 98% of the discrete models that we randomly generated, though only for 65% of the random Gaussian models. As discussed in Section 6, we artificially enforced $\langle a^* \rangle \in \mathbf{Pa}_p^{\mathcal{G}}$ when evaluating the regret of TS* and (Bayes-)UCB*.

For Bayes-UCB(*), the best quantile constant in (8) was $c = 0$, in agreement with the empirical recommendation by Kaufmann et al. (2012). The best exploration parameter for UCB in (6) was $c = 1/(2\sqrt{2})$ for UCB(*) in the discrete setting. In the Gaussian setting, UCB and UCB* preferred $c = 1/2$ and $c = 1/\sqrt{2}$, respectively, the latter of which we applied for BBB. We used standard uninformative priors for TS(*), with $\alpha_0 = \beta_0 = 1$ for the Beta prior and $m_0 = 0$, $\nu_0 = 1$, and $u_0 = v_0 = 1$ for the N-Γ^{-1} prior. For BBB, we computed exact parent set probabilities (4) using the program[1] implementing the efficient algorithm developed by and applied in Pensar et al. (2020), restricting the maximum size of parent sets to three and using the Bayesian Dirichlet equivalent uniform and Bayesian Gaussian equivalent scores. For the Gaussian setting, we checked the graphical criterion in Proposition 2 according to (13) with $\tau = 0.1$.

While we focused in Section 3 on designing the marginal posteriors according to (3), a notable difference between our proposed Bayesian CB framework and the Bayesian MAB approach described in Section 2 is that in our design, the posterior distribution is not modular, with the marginals $(\pi_a^t)_{a \in \mathcal{A}}$ mutually dependent on the distribution of graph structures. However, because of software limitations and for simplicity, we sampled the criterion $U_a(t)$ for each arm independently in the implementation of BBB-TS in our experiments (line 7 in Algorithm 1). Although preliminary results have shown the difference in empirical performance to be negligible, a more precise implementation would first sample a DAG $\mathcal{G}$ from the posterior distribution $P(\mathcal{G} \mid \mathcal{D}[t])$ and subsequently for each arm $a \in \mathcal{A}$, sample $U_a(t)$ from $\pi^t_{a|\mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}}$.

C Additional Experiments

Here, we present the results from additional experiments designed to evaluate firstly our proposed approximation of the sampling variance of the discrete backdoor adjustment estimator (9), and secondly the application of Proposition 2 by way of Gaussian backdoor adjustment with jointly interventional and observational data.

C.1 Discrete Backdoor Adjustment and Variance

In this section, we describe and present experiments evaluating the behavior of $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$ where $\mathbf{Z} = \mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}$ as in (9), as well as our proposed approximation of its variance, derived in detail in Appendix D. Four variance estimation methods were investigated. In the naive approach, $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$ is treated as a conditional proportion as is the case when $|\mathbf{Z}| = 0$, and the variance is estimated as $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})[1 - \hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})]/n[x_a]$ where $n[x_a]$ is the number of samples of data where $X_{\langle a \rangle} = x_a$. The sampling approach estimates the variance from samples from the population distribution, and the bootstrap approach conducts resampling from each sample distribution, each with 10^3 repetitions.

The generation of discrete CBNs for the simulation scenarios was designed as follows. The graph structure was generated simply by initializing a structure where there is a direct edge from the intervened node $X_{\langle a \rangle}$ to the reward variable Y and $X_{\langle a \rangle}$ has $|\mathbf{Z}| = m$ parents. For each parent $X_j \in \mathbf{Z}$, an edge $X_j \to Y$ was randomly added with 0.5 probability to create backdoor paths. Finally, conditional probability tables were generated uniformly as described in Section 6.

For observational sample sizes $n_0 \in \{100 \cdot 2^k : k = 0, 1, \ldots, 5\}$ and parent set sizes $|\mathbf{Z}| \in \{0, 1, 2, 3\}$, 10^3 scenarios were created by randomly generating CBNs as described above and the methods were assessed under each scenario through the following process. First, 10^6 datasets were generated, each with n_0 samples of observational data, and for each dataset, $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$ was computed for some arbitrary $x_a \in \mathrm{Dom}(X_{\langle a \rangle})$.

[1] Pensar et al. (2020) provided their code under the MIT License at `https://github.com/jopensar/BIDA`.

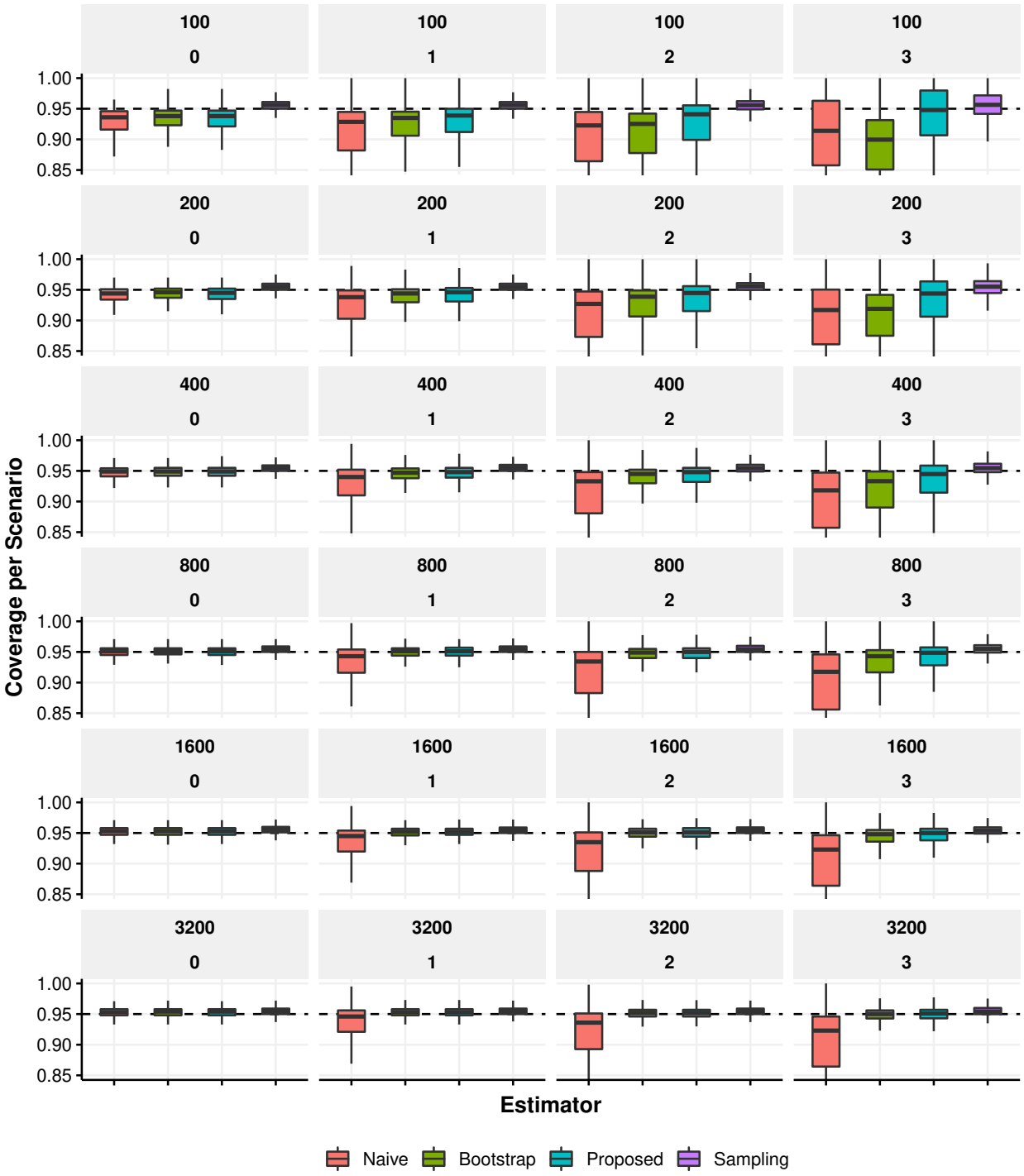

Figure 4: Coverage probability per scenario using various estimators of $\mathrm{Var}[\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})]$ across $n_0 \in \{100 \cdot 2^k : k = 0, 1, \ldots, 5\}$ samples of observational data and $|\mathbf{Z}| \in \{0, 1, 2, 3\}$ adjustment set sizes

Then, for each of the four methods, the variance was estimated corresponding to the first 10^3 estimates of $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$, and from those the 2 standard deviation interval coverage probability of the true μ_a was computed.

The estimator $\hat{\mu}_{a,\mathrm{bda}}(\mathbf{Z})$ itself was found to be generally unbiased, with the average of the 10^6 estimates deviating from the true μ_a by less than 2% in over 99% of the 24,000 scenarios. The coverage probability

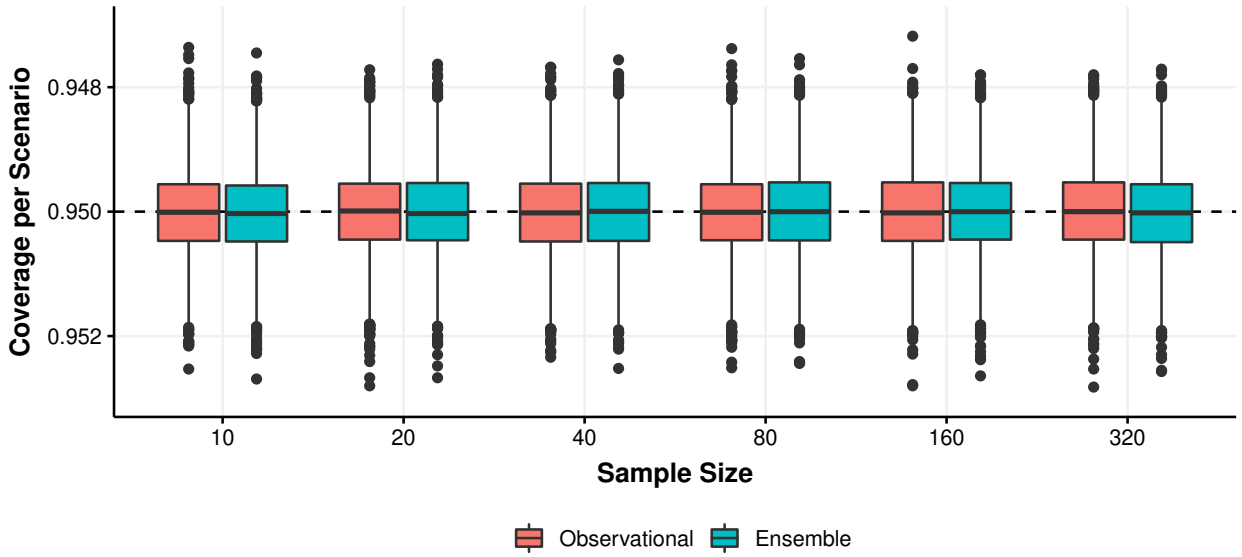

Figure 5: Coverage probability per simulation scenario across sample sizes for observational and ensemble data generating methods

results are shown in Figure 4, where each boxplot visualizes the coverage probability of a method across 10^3 scenarios randomly generated under the given simulation setting. The outliers and invalid values, which typically corresponded to extreme scenarios, were removed. The naive approach is only correct when $|\mathbf{Z}| = 0$ and performs poorly when otherwise. The general results may be summarized as Naive $<$ Bootstrap $\approx$ Proposed $<$ Sampling, though our proposed estimator appears to outperform the bootstrap approach for larger $|\mathbf{Z}|$ and perform comparably with the population sampling approach for larger n_0.

C.2 Gaussian Backdoor Adjustment with Ensemble Data

In this section, we empirically validate our methodology of conducting the regression (11) with jointly interventional and observational data to estimate $\psi_{\langle a \rangle}$, as discussed in Section 5.2. In particular, we compare the coverage probability of $\hat{\psi}_{\langle a \rangle,\text{bda}}(\mathbf{Z})$ where $\mathbf{Z} = \mathbf{Pa}^{\mathcal{G}}_{\langle a \rangle}$ estimated using purely observational data and ensemble data. The ensemble data was generated by allowing each data sample to be generated by one of the possible interventions $\{do(X_j = x_j) : X_j \in \mathbf{Z}, x_j \in \{-1, 1\}\}$ or by passive observation, with equal probability given to each of the $2|\mathbf{Z}| + 1$ options.

For sample sizes $n \in \{10 \cdot 2^k : k = 0, 1, \ldots, 5\}$ and parent set sizes $|\mathbf{Z}| \in \{1, 2, 3, 4\}$, 10^3 scenarios were created by randomly generating CBNs. The network structures were generated as described in Appendix C.1, and the parameters as in Section 6. Each data generation method was evaluated for each scenario by generating 10^5 datasets with n samples and estimating $\hat{\psi}_{\langle a \rangle,\text{bda}}(\mathbf{Z})$ and $\hat{\text{SE}}^2[\hat{\psi}_{\langle a \rangle,\text{bda}}(\mathbf{Z})]$ for each dataset by conducting the regression (11). From those estimates, 95% confidence interval coverage probabilities were computed for each scenario.

The average of the 10^5 estimates of $\hat{\psi}_{\langle a \rangle,\text{bda}}(\mathbf{Z})$ deviated from the true $\psi_{\langle a \rangle}$ by at most 0.9% across all 24,000 simulation scenarios for both data generation methods. The coverage probability results are shown in Figure 5. Since the results did not vary across parent set sizes, each boxplot visualizes the coverage probability of a method across the $4,000$ simulation scenarios at each sample size. It is easy to see equivalent performance of the estimator computed with ensemble data compared to observational data, with consistent coverage across all sample sizes.

D Derivation of the Discrete Backdoor Adjustment Variance Approximation

In this section, we derive the approximation of the sampling variance of (9):

$$\hat{\mu}_{a,\text{bda}}(\mathbf{Z}) = \frac{1}{n_0} \sum_{\mathbf{z}} \frac{n_0[1, x_a, \mathbf{z}] n_0[\mathbf{z}]}{n_0[x_a, \mathbf{z}]}.$$

D.1 Introduction

For simplicity, we redefine some notation. The backdoor adjustment to estimate the interventional distribution of $Y \mid do(X = x)$ with parent set $\mathbf{Z} = \mathbf{Pa}_X^{\mathcal{G}}$ with r parent configurations is given by:

$$P[Y = y \mid do(X = x)] = \sum_{\mathbf{z}} P(Y = y \mid X = x, \mathbf{Z} = \mathbf{z}) P(\mathbf{Z} = \mathbf{z}).$$

Empirically, given n samples of observational data, this quantity is estimated using counts:

$$\hat{P}[Y = y \mid do(X = x)] = \sum_{\mathbf{z}} \frac{n[y, x, \mathbf{z}]}{n[x, \mathbf{z}]} \frac{n[\mathbf{z}]}{n} = \frac{1}{n} \sum_{\mathbf{z}} \frac{n[y, x, \mathbf{z}] n[\mathbf{z}]}{n[x, \mathbf{z}]} \tag{14}$$

where $n[y, x, \mathbf{z}]$ represents the number of samples in which $Y = y$, $X = x$, and $\mathbf{Z} = \mathbf{z}$, with corresponding definitions for $n[x, \mathbf{z}]$ and $n[\mathbf{z}]$. The joint probability distribution of X, Y, and $\mathbf{Z}$ may be lumped into a multinomial random vector $\mathbf{N} = (N_1, N_1{}', N_1{}'', \ldots, N_r, N_r{}', N_r{}'') \in \mathbb{R}^{3r}$ where for $i = 1, \ldots, r$,

$$N_i = n[y, x, \mathbf{z}_i], \quad N_i{}' = n[\neg y, x, \mathbf{z}_i], \quad N_i{}'' = n[\neg x, \mathbf{z}_i].$$

Note that $N_i + N_i{}' + N_i{}'' = n[\mathbf{z}_i]$, so $\sum_{i=1}^{r} (N_i + N_i{}' + N_i{}'') = n$, so $\mathbf{N}$ may be thought of as a repartitioning of the joint probability distribution of X, Y, and $\mathbf{Z}$ into $3r$ disjoint levels:

$$\begin{aligned}
&\mathbf{N} = (N_1, N_1{}', N_1{}'', \ldots, N_r, N_r{}', N_r{}'') \sim \text{Multinom}(n, \mathbf{p}), \\
&\mathbf{p} = (p_1, p_1{}', p_1{}'', \ldots, p_r, p_r{}', p_r{}''), \quad \text{where} \\
&p_i = \mathrm{E}\left[\frac{n[y, x, \mathbf{z}_i]}{n}\right], \quad p_i{}' = \mathrm{E}\left[\frac{n[\neg y, x, \mathbf{z}_i]}{n}\right], \quad p_i{}'' = \mathrm{E}\left[\frac{n[\neg x, \mathbf{z}_i]}{n}\right] \quad \text{for } i = 1, \ldots, r.
\end{aligned} \tag{15}$$

The advantage of such a representation is so that for each $\mathbf{z}_i$, the term within the summation may be expressed as a function of three disjoint elements of a multinomial random vector:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{r} \frac{n[y, x, \mathbf{z}_i] n[\mathbf{z}_i]}{n[x, \mathbf{z}_i]} &= \frac{1}{n} \sum_{i=1}^{r} \frac{n[y, x, \mathbf{z}_i] \left(n[y, x, \mathbf{z}_i] + n[\neg y, x, \mathbf{z}_i] + n[\neg x, \mathbf{z}_i]\right)}{n[y, x, \mathbf{z}_i] + n[\neg y, x, \mathbf{z}_i]} \\
&= \frac{1}{n} \sum_{i=1}^{r} \frac{N_i (N_i + N_i{}' + N_i{}'')}{N_i + N_i{}'}.
\end{aligned} \tag{16}$$

Note that each term is not straightforward to compute. An obvious challenge is that the denominator of each term in the summation in (16) can be zero, so there is no analytical solution for its mean, variance, and covariance.

D.2 Taylor Series Expansion for Ratio Distribution

To circumvent this challenge, we approximate the ratio in (16) with the Taylor series approximation. We begin by defining

$$\begin{aligned}
M_i &= \frac{N_i (N_i + N_i{}' + N_i{}'')}{n^2}, \\
W_i &= \frac{N_i + N_i{}'}{n}, \\
Q_i &= f(M_i, W_i) = \frac{M_i}{W_i}.
\end{aligned}$$

This allows us to express the variance of (16) in terms of Q_i:

$$\begin{aligned}\mathrm{Var}\left[\hat{P}[Y=y \mid do(X=x)]\right] &= \mathrm{Var}\left[\sum_{i=1}^{r} Q_i\right] \\ &= \sum_{i}^{r} \mathrm{Var}\left[Q_i\right] + 2\sum_{i=1}^{r}\sum_{j>i} \mathrm{Cov}\left[Q_i, Q_j\right].\end{aligned} \tag{17}$$

By Taylor series expansion around $\mu_i = (\mu_{M_i}, \mu_{W_i}) = (\mathrm{E}[M_i], \mathrm{E}[W_i])$:

$$\begin{aligned}Q_i &= f(M_i, W_i) \\ &\approx f(\mu_i) + (M_i - \mu_{M_i})\frac{\partial f}{\partial M_i}(\mu_i) + (W_i - \mu_{W_i})\frac{\partial f}{\partial W_i}(\mu_i) \\ &\quad + \frac{1}{2}(M_i - \mu_{M_i})^2\frac{\partial^2 f}{\partial M_i^2}(\mu_i) + \frac{1}{2}(W_i - \mu_{W_i})^2\frac{\partial^2 f}{\partial W_i^2}(\mu_i) \\ &\quad + (M_i - \mu_{M_i})(W_i - \mu_{W_i})\frac{\partial^2 f}{\partial M_i \partial W_i}(\mu_i),\end{aligned} \tag{18}$$

where

$$\begin{aligned}\frac{\partial f}{\partial M_i}(M_i, W_i) &= \frac{1}{W_i}, & \frac{\partial^2 f}{\partial M_i^2}(M_i, W_i) &= 0, \\ \frac{\partial f}{\partial W_i}(M_i, W_i) &= -\frac{M_i}{W_i^2}, & \frac{\partial^2 f}{\partial W_i^2}(M_i, W_i) &= \frac{2M_i}{W_i^3}, \\ \frac{\partial^2 f}{\partial M_i \partial W_i}(M_i, W_i) &= \frac{\partial^2 f}{\partial W_i \partial M_i}(M_i, W_i) = \frac{1}{W_i^2}\end{aligned} \tag{19}$$

Given (18), we obtain an approximate expected value:

$$\mathrm{E}[Q_i] \approx f(\mu_i) + \frac{1}{2}\frac{\partial^2 f}{\partial M_i^2}(\mu_i)\mathrm{Var}[M_i] + \frac{1}{2}\frac{\partial^2 f}{\partial W_i^2}(\mu_i)\mathrm{Var}[W_i] + \frac{\partial^2 f}{\partial M_i \partial W_i}(\mu_i)\mathrm{Cov}[M_i, W_i]. \tag{20}$$

For variance and covariance, we use a simpler approximation:

$$Q_i = f(M_i, W_i) \approx f(\mu_i) + (M_i - \mu_{M_i})\frac{\partial f}{\partial M_i}(\mu_i) + (W_i - \mu_{W_i})\frac{\partial f}{\partial W_i}(\mu_i), \tag{21}$$

resulting in

$$\begin{aligned}\mathrm{Var}[Q_i] \approx \frac{\partial f}{\partial M_i}(\mu_i)^2\mathrm{Var}[M_i] + \frac{\partial f}{\partial W_i}(\mu_i)^2\mathrm{Var}[W_i] \\ + 2\frac{\partial f}{\partial M_i}(\mu_i)\frac{\partial f}{\partial W_i}(\mu_i)\mathrm{Cov}[M_i, W_i],\end{aligned} \tag{22}$$

and

$$\begin{aligned}\mathrm{E}[Q_i Q_j] &\approx f(\mu_i)f(\mu_j) \\ &\quad + \frac{\partial f}{\partial M_i}(\mu_i)\frac{\partial f}{\partial M_j}(\mu_j)\mathrm{Cov}[M_i, M_j] + \frac{\partial f}{\partial M_i}(\mu_i)\frac{\partial f}{\partial W_j}(\mu_j)\mathrm{Cov}[M_i, W_j] \\ &\quad + \frac{\partial f}{\partial W_i}(\mu_i)\frac{\partial f}{\partial M_j}(\mu_j)\mathrm{Cov}[W_i, M_j] + \frac{\partial f}{\partial W_i}(\mu_i)\frac{\partial f}{\partial W_j}(\mu_j)\mathrm{Cov}[W_i, W_j],\end{aligned}$$

so

$$\begin{aligned}\mathrm{Cov}[Q_i, Q_j] &= \mathrm{E}[Q_i Q_j] - \mathrm{E}[Q_i]\mathrm{E}[Q_j] \\ &= \frac{\partial f}{\partial M_i}(\mu_i)\frac{\partial f}{\partial M_j}(\mu_j)\mathrm{Cov}[M_i, M_j] + \frac{\partial f}{\partial M_i}(\mu_i)\frac{\partial f}{\partial W_j}(\mu_j)\mathrm{Cov}[M_i, W_j] \\ &\quad + \frac{\partial f}{\partial W_i}(\mu_i)\frac{\partial f}{\partial M_j}(\mu_j)\mathrm{Cov}[W_i, M_j] + \frac{\partial f}{\partial W_i}(\mu_i)\frac{\partial f}{\partial W_j}(\mu_j)\mathrm{Cov}[W_i, W_j].\end{aligned} \tag{23}$$

In what follows, we first derive important quantities from the multinomial distribution in Appendix D.3 and apply them to compute the quantities in (17).

D.3 Multinomial Derivations

For this subsection, in an abuse of notation, let $\mathbf{N} = (N_1, \ldots, N_r) \sim \text{Multinom}(n, \mathbf{p})$ and $u, v, w, x \in \{1, \ldots, r\}$ are distinct values. It is well-known that $\text{E}[N_u] = np_u$, $\text{Var}[N_u] = np_u(1-p_u)$, and $\text{Cov}(N_u, N_v) = -np_up_v$. Furthermore,

$$\begin{aligned} \text{E}[N_uN_v] &= \text{Cov}[N_u, N_v] + \text{E}[N_u]\text{E}[N_v] \\ &= n(n-1)p_up_v, \end{aligned} \tag{24}$$

and the first four moments from derivating the moment generating function are:

$$\begin{aligned} \text{E}[N_u] &= np_u, \\ \text{E}[N_u^2] &= n(n-1)p_u^2 + \text{E}[N_u] \\ &= np_u[1 + (n-1)p_u], \\ \text{E}[N_u^3] &= n(n-1)[(n-2)p_u^3 + 2p_u^2] + \text{E}[N_u^2] \\ &= np_u\left[1 + (n-1)p_u(3 + (n-2)p_u)\right], \\ \text{E}[N_u^4] &= n(n-1)(n-2)\left[(n-3)p_u^4 + 3p_u^3\right] + 2n(n-1)\left[(n-2)p_u^3 + 2p_u^2\right] + \text{E}[N_u^3] \\ &= np_u\left[1 + (n-1)p_u(7 + (n-2)p_u[6 + (n-3)p_u])\right]. \end{aligned} \tag{25}$$

Define indicator random variable U_i such that $U_i = 1$ if the outcome for trial i is $u \in \{1, \ldots, r\}$ and $U_i = 0$ otherwise. Similarly define V_i for $v \neq u$, W_i for $w \neq v \neq u$, and X_i for $x \neq w \neq v \neq u$. Then N_u, N_v, N_w, and N_x may be expressed as

$$N_u = \sum_{i=1}^{n} U_i, \quad N_v = \sum_{i=1}^{n} V_i, \quad N_w = \sum_{i=1}^{n} W_i, \quad N_x = \sum_{i=1}^{n} X_i.$$

We are interested in $\text{E}[N_u^2N_v^2]$, $\text{E}[N_u^3N_v]$, $\text{E}[N_u^2N_vN_w]$, $\text{E}[N_uN_vN_wN_x]$, $\text{E}[N_u^2N_v]$, and $\text{E}[N_uN_vN_w]$.

$$
\begin{aligned}
\mathrm{E}[N_u^2 N_v^2] &= \mathrm{E}\left[\left(\sum_{i=1}^{n} U_i\right)^2 \left(\sum_{i=1}^{n} V_i\right)^2\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} U_i U_j V_k V_l\right] && \text{by distributing} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathrm{E}\left[U_i U_j V_k V_l\right] && \text{by linearity of expectation} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j}} \mathrm{E}\left[U_i U_j V_k V_l\right] && \text{since } U_i V_i = 0 \text{ for all } i = 1, \ldots, n \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j}} \mathrm{E}\left[U_i U_j\right]\mathrm{E}\left[V_k V_l\right] && \text{by independence between trials} \\
&= \sum_{i=j}\sum_{\substack{k=l\\ k\neq i}} \mathrm{E}\left[U_i U_j\right]\mathrm{E}\left[V_k V_l\right] + \sum_{i}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq k\\ l\neq i\\ l\neq j}} \mathrm{E}\left[U_i U_j\right]\mathrm{E}\left[V_k V_l\right] \\
&\quad + \sum_{i=j}\sum_{k\neq i}\sum_{\substack{l\neq k\\ l\neq i}} \mathrm{E}\left[U_i U_j\right]\mathrm{E}\left[V_k V_l\right] + \sum_{k=l}\sum_{i\neq k}\sum_{\substack{j\neq i\\ j\neq k}} \mathrm{E}\left[U_i U_j\right]\mathrm{E}\left[V_k V_l\right] && \text{reexpressed} \\
&= \sum_{i}\sum_{\substack{k=l\\ k\neq i}} \mathrm{E}[U_i^2]\mathrm{E}[V_k^2] + \sum_{i}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq k\\ l\neq i\\ l\neq j}} \mathrm{E}[U_i]\mathrm{E}[U_j]\mathrm{E}[V_k]\mathrm{E}[V_l] && \text{reexpressed; independence; and} \\
&\quad + \sum_{i}\sum_{k\neq i}\sum_{\substack{l\neq k\\ l\neq i}} \mathrm{E}[U_i^2]\mathrm{E}[V_k V_l] + \sum_{k}\sum_{i\neq k}\sum_{\substack{j\neq i\\ j\neq k}} \mathrm{E}[U_i]\mathrm{E}[U_j]\mathrm{E}[V_k^2] && \text{since } \mathrm{E}[U_i U_j] = \mathrm{E}[U_i]\mathrm{E}[U_j],\ i \neq j \\
&= n(n-1)p_u p_v + n(n-1)(n-2)(n-3)p_u^2 p_v^2 && \text{since } \mathrm{E}[U_i^2] = \mathrm{E}[U_i] = p_u \\
&\quad + n(n-1)(n-2)p_u p_v^2 + n(n-1)(n-2)p_u^2 p_v \\
&= n(n-1)p_u p_v\left[1 + (n-2)(p_u + p_v + (n-3)p_u p_v)\right] && \text{simplified.}
\end{aligned}
$$

Hence,

$$
\mathrm{E}[N_u^2 N_v^2] = n(n-1)p_u p_v\left[1 + (n-2)(p_u + p_v + (n-3)p_u p_v)\right]. \tag{26}
$$

Following the same derivation strategy,

$$
\mathrm{E}[N_u^3 N_v] = n(n-1)p_u p_v\left[1 + (n-2)p_u(3 + (n-3)p_u)\right], \tag{27}
$$

$$
\mathrm{E}[N_u^2 N_v N_w] = n(n-1)(n-2)p_u p_v p_w\left[1 + (n-3)p_u\right], \tag{28}
$$

$$
\mathrm{E}[N_u N_v N_w N_x] = n(n-1)(n-2)(n-3)p_u p_v p_w p_x, \tag{29}
$$

$$
\mathrm{E}[N_u^2 N_v] = n(n-1)p_u p_v\left[1 + (n-2)p_u\right], \tag{30}
$$

$$
\mathrm{E}[N_u N_v N_w] = n(n-1)(n-2)p_u p_v p_w. \tag{31}
$$

D.4 Numerator and Denominator of Ratio

We now turn to the task of deriving expressions for $\mathrm{Var}[M_i]$, $\mathrm{Var}[W_i]$, and $\mathrm{Cov}[M_i, W_i]$ in order to compute (22), and additionally for $\mathrm{Cov}[M_i, M_j]$, $\mathrm{Cov}[M_i, W_j]$, $\mathrm{Cov}[W_i, M_j]$, and $\mathrm{Cov}[W_i, W_j]$ for (23). For this subsection, return to the notation for $\mathbf{N}$ expressed in (15).

The distribution of $W_i = n^{-1}(N_i + N_i')$ is most simple. By the lumping property of multinomial random vectors,

$$\begin{aligned} \mathrm{E}[W_i] &= p_i + p_i', \\ \mathrm{Var}[W_i] &= \frac{(p_i + p_i')(1 - p_i - p_i')}{n}, \\ \mathrm{Cov}[W_i, W_j] &= -\frac{(p_i + p_i')(p_j + p_j')}{n}. \end{aligned} \tag{32}$$

The distribution of $M_i = n^{-2}N_i(N_i + N_i' + N_i'')$ is more challenging. From (25) and (24), the expectation is given by:

$$\begin{aligned} \mathrm{E}[M_i] &= n^{-2}\mathrm{E}[N_i(N_i + N_i' + N_i'')] \\ &= n^{-2}\left(\mathrm{E}[N_i^2] + \mathrm{E}[N_i N_i'] + \mathrm{E}[N_i N_i'']\right) \\ &= n^{-2}\left(np_i[1 + (n-1)p_i] + n(n-1)p_i p_i' + n(n-1)p_i p_i''\right) \\ &= n^{-1}p_i[1 + (n-1)(p_i + p_i' + p_i'')]. \end{aligned} \tag{33}$$

Next, the variance is given by:

$$\begin{aligned} \mathrm{Var}[M_i] &= n^{-4}\mathrm{Var}[N_i(N_i + N_i' + N_i'')] \\ &= n^{-4}\mathrm{Var}[N_i^2 + N_i N_i' + N_i N_i''] \\ &= n^{-4}\big(\mathrm{Var}[N_i^2] + \mathrm{Var}[N_i N_i'] + \mathrm{Var}[N_i N_i''] \\ &\qquad + 2\mathrm{Cov}[N_i^2, N_i N_i'] + 2\mathrm{Cov}[N_i^2, N_i N_i''] + 2\mathrm{Cov}[N_i N_i', N_i N_i'']\big). \end{aligned}$$

The terms in the expression above are given below. From the moments of the multinomial distribution (25):

$$\begin{aligned} \mathrm{Var}[N_i^2] &= \mathrm{E}[N_i^4] - \mathrm{E}[N_i^2]^2 \\ &= np_i\left[1 + (n-1)p_i(7 + (n-2)p_i[6 + (n-3)p_i])\right] - (np_i[1 + (n-1)p_i])^2 \\ &= np_i\left[1 + (n-1)p_i(7 + (n-2)p_i[6 + (n-3)p_i]) - np_i(1 + (n-1)p_i)^2\right]. \end{aligned}$$

From (26) and (24):

$$\begin{aligned} \mathrm{Var}[N_i N_i'] &= \mathrm{E}[N_i^2 {N_i'}^2] - \mathrm{E}[N_i N_i']^2 \\ &= n(n-1)p_i p_i'[1 + (n-2)(p_i + p_i' + (n-3)p_i p_i')] - [n(n-1)p_i p_i']^2 \\ &= n(n-1)p_i p_i'[1 + (n-2)(p_i + p_i' + (n-3)p_i p_i') - n(n-1)p_i p_i'], \\ \mathrm{Var}[N_i N_i''] &= n(n-1)p_i p_i''[1 + (n-2)(p_i + p_i'' + (n-3)p_i p_i'') - n(n-1)p_i p_i'']. \end{aligned}$$

From (27), (25), and (24):

$$\begin{aligned} \mathrm{Cov}[N_i^2, N_i N_i'] &= \mathrm{E}[N_i^3 N_i'] - \mathrm{E}[N_i^2]\mathrm{E}[N_i N_i'] \\ &= n(n-1)p_i p_i'\left[1 + (n-2)p_i(3 + (n-3)p_i)\right] \\ &\quad - np_i[1 + (n-1)p_i]n(n-1)p_i p_i' \\ &= n(n-1)p_i p_i'\left[1 + (n-2)(3p_i + (n-3)p_i^2) - np_i(1 + (n-1)p_i)\right] \\ \mathrm{Cov}[N_i^2, N_i N_i''] &= n(n-1)p_i p_i''\left[1 + (n-2)(3p_i + (n-3)p_i^2) - np_i(1 + (n-1)p_i)\right]. \end{aligned}$$

From (28) and (24):

$$\begin{aligned} \mathrm{Cov}[N_i N_i', N_i N_i''] &= \mathrm{E}[N_i^2 N_i' N_i''] - \mathrm{E}[N_i N_i']\mathrm{E}[N_i N_i''] \\ &= n(n-1)(n-2)p_i p_i' p_i''[1 + (n-3)p_i] - n(n-1)p_i p_i' n(n-1)p_i p_i'' \\ &= n(n-1)p_i p_i' p_i''\left[(n-2)[1 + (n-3)p_i] - n(n-1)p_i\right]. \end{aligned}$$

Hence, $\mathrm{Var}[M_i]$ is derived:

$$\begin{aligned}\mathrm{Var}[M_i] = n^{-4}\big(&np_i\left[1+(n-1)p_i(7+(n-2)p_i[6+(n-3)p_i])-np_i(1+(n-1)p_i)^2\right]\\&+n(n-1)p_ip_i'[1+(n-2)(p_i+p_i'+(n-3)p_ip_i')-n(n-1)p_ip_i']\\&+n(n-1)p_ip_i''[1+(n-2)(p_i+p_i''+(n-3)p_ip_i'')-n(n-1)p_ip_i'']\\&+2n(n-1)p_ip_i'\left[1+(n-2)(3p_i+(n-3)p_i^2)-np_i(1+(n-1)p_i)\right]\\&+2n(n-1)p_ip_i''\left[1+(n-2)(3p_i+(n-3)p_i^2)-np_i(1+(n-1)p_i)\right]\\&+n(n-1)p_ip_i'p_i''\left[(n-2)[1+(n-3)p_i]-n(n-1)p_i\right]\big).\end{aligned} \tag{34}$$

Next, consider $\mathrm{Cov}[M_i, M_j]$.

$$\begin{aligned}\mathrm{Cov}[M_i,M_j] &= n^{-4}\mathrm{Cov}\left[N_i(N_i+N_i'+N_i''),N_j(N_j+N_j'+N_j'')\right]\\&= n^{-4}\mathrm{Cov}\left[N_i^2+N_iN_i'+N_iN_i'',N_j^2+N_jN_j'+N_jN_j''\right]\\&= n^{-4}\big(\mathrm{Cov}[N_i^2,N_j^2]\\&\quad+\mathrm{Cov}[N_i^2,N_jN_j']+\mathrm{Cov}[N_i^2,N_jN_j'']+\mathrm{Cov}[N_iN_i',N_j^2]+\mathrm{Cov}[N_iN_i'',N_j^2]\\&\quad+\mathrm{Cov}[N_iN_i',N_jN_j']+\mathrm{Cov}[N_iN_i',N_jN_j'']\\&\quad+\mathrm{Cov}[N_iN_i'',N_jN_j']+\mathrm{Cov}[N_iN_i'',N_jN_j'']\big).\end{aligned}$$

The terms in the expression above are given below. From (26) and (25):

$$\begin{aligned}\mathrm{Cov}[N_i^2,N_j^2] &= \mathrm{E}[N_i^2N_j^2]-\mathrm{E}[N_i^2]\mathrm{E}[N_j^2]\\&= n(n-1)p_ip_j\left[1+(n-2)(p_i+p_j+(n-3)p_ip_j)\right]\\&\quad-np_i[1+(n-1)p_i]np_j[1+(n-1)p_j]\\&= np_ip_j[(n-1)(1+(n-2)(p_i+p_j+(n-3)p_ip_j))\\&\qquad-n(1+(n-1)p_i)(1+(n-1)p_j)]\end{aligned}$$

From (28), (25), and (24):

$$\begin{aligned}\mathrm{Cov}[N_i^2,N_jN_j'] &= \mathrm{E}[N_i^2N_jN_j']-\mathrm{E}[N_i^2]\mathrm{E}[N_jN_j']\\&= n(n-1)(n-2)p_ip_jp_j'[1+(n-3)p_i]-np_i(1+(n-1)p_i)n(n-1)p_jp_j'\\&= n(n-1)p_ip_jp_j'\left[(n-2)[1+(n-3)p_i]-n(1+(n-1)p_i)\right],\\\mathrm{Cov}[N_i^2,N_jN_j''] &= n(n-1)p_ip_jp_j''\left[(n-2)[1+(n-3)p_i]-n(1+(n-1)p_i)\right],\\\mathrm{Cov}[N_iN_i',N_j^2] &= n(n-1)p_jp_ip_i'\left[(n-2)[1+(n-3)p_j]-n(1+(n-1)p_j)\right],\\\mathrm{Cov}[N_iN_i'',N_j^2] &= n(n-1)p_jp_ip_i''\left[(n-2)[1+(n-3)p_j]-n(1+(n-1)p_j)\right].\end{aligned}$$

From (29) and (24):

$$\begin{aligned}\mathrm{Cov}[N_iN_i',N_jN_j'] &= \mathrm{E}[N_iN_i'N_jN_j']-\mathrm{E}[N_iN_i']\mathrm{E}[N_jN_j']\\&= n(n-1)(n-2)(n-3)p_ip_i'p_jp_j'-n(n-1)p_ip_i'n(n-1)p_jp_j'\\&= n(n-1)p_ip_i'p_jp_j'\left[(n-2)(n-3)-n(n-1)\right],\\\mathrm{Cov}[N_iN_i',N_jN_j''] &= n(n-1)p_ip_i'p_jp_j''\left[(n-2)(n-3)-n(n-1)\right],\\\mathrm{Cov}[N_iN_i'',N_jN_j'] &= n(n-1)p_ip_i''p_jp_j'\left[(n-2)(n-3)-n(n-1)\right],\\\mathrm{Cov}[N_iN_i'',N_jN_j''] &= n(n-1)p_ip_i''p_jp_j''\left[(n-2)(n-3)-n(n-1)\right].\end{aligned}$$

Hence, $\mathrm{Cov}[M_i, M_j]$ is derived:

$$\begin{aligned}
&\mathrm{Cov}[M_i, M_j] \\
&\quad = n^{-4}\big(np_ip_j\big[(n-1)(1+(n-2)(p_i+p_j+(n-3)p_ip_j)) \\
&\qquad\qquad - n(1+(n-1)p_i)(1+(n-1)p_j)\big] \\
&\qquad + n(n-1)p_ip_jp_j{}'\left[(n-2)[1+(n-3)p_i] - n(1+(n-1)p_i)\right] \\
&\qquad + n(n-1)p_ip_jp_j{}''\left[(n-2)[1+(n-3)p_i] - n(1+(n-1)p_i)\right] \\
&\qquad + n(n-1)p_jp_ip_i{}'\left[(n-2)[1+(n-3)p_j] - n(1+(n-1)p_j)\right] \\
&\qquad + n(n-1)p_jp_ip_i{}''\left[(n-2)[1+(n-3)p_j] - n(1+(n-1)p_j)\right] \\
&\qquad + n(n-1)p_ip_i{}'p_jp_j{}'\left[(n-2)(n-3) - n(n-1)\right] \\
&\qquad + n(n-1)p_ip_i{}'p_jp_j{}''\left[(n-2)(n-3) - n(n-1)\right] \\
&\qquad + n(n-1)p_ip_i{}''p_jp_j{}'\left[(n-2)(n-3) - n(n-1)\right] \\
&\qquad + n(n-1)p_ip_i{}''p_jp_j{}''\left[(n-2)(n-3) - n(n-1)\right]\big) \\
&\quad = n^{-3}p_ip_j\big[(n-1)\big(1+(n-2)(p_i+p_j+(n-3)p_ip_j) \\
&\qquad\qquad + (p_j{}'+p_j{}'')[(n-2)[1+(n-3)p_i] - n(1+(n-1)p_i)] \\
&\qquad\qquad + p_jp_i(p_i{}'+p_i{}'')[(n-2)[1+(n-3)p_j] - n(1+(n-1)p_j)] \\
&\qquad\qquad + (p_i{}'+p_i{}'')(p_j{}'+p_j{}'')[(n-2)(n-3) - n(n-1)]\big) \\
&\qquad - n(1+(n-1)p_i)(1+(n-1)p_j)\big]
\end{aligned} \tag{35}$$

Finally, we turn our attention to $\mathrm{Cov}[M_i, W_i]$, $\mathrm{Cov}[M_i, W_j]$, and $\mathrm{Cov}[W_i, M_j]$. Beginning with $\mathrm{Cov}[M_i, W_i]$:

$$\begin{aligned}
\mathrm{Cov}[M_i, W_i] &= n^{-3}\mathrm{Cov}\left[N_i(N_i + N_i{}' + N_i{}''), N_i + N_i{}'\right] \\
&= n^{-3}\mathrm{Cov}\left[N_i^2 + N_iN_i{}' + N_iN_i{}'', N_i + N_i{}'\right] \\
&= n^{-3}\big(\mathrm{Cov}[N_i^2, N_i] + \mathrm{Cov}[N_i^2, N_i{}'] \\
&\qquad + \mathrm{Cov}[N_iN_i{}', N_i] + \mathrm{Cov}[N_iN_i{}', N_i{}'] + \mathrm{Cov}[N_iN_i{}'', N_i] + \mathrm{Cov}[N_iN_i{}'', N_i{}']\big).
\end{aligned}$$

The terms in the expression above are given below. From (25):

$$\begin{aligned}
\mathrm{Cov}[N_i^2, N_i] &= \mathrm{E}[N_i^3] - \mathrm{E}[N_i^2]\mathrm{E}[N_i] \\
&= np_i\left[1+(n-1)p_i(3+(n-2)p_i)\right] - np_i[1+(n-1)p_i]np_i \\
&= np_i\left[1+(n-1)p_i(3+(n-2)p_i) - np_i(1+(n-1)p_i)\right] \\
&= np_i\left[1+p_i((n-1)[3-2p_i]-n)\right].
\end{aligned}$$

From (30) and (25):

$$\begin{aligned}
\mathrm{Cov}[N_i^2, N_i{}'] &= \mathrm{E}[N_i^2N_i{}'] - \mathrm{E}[N_i^2]\mathrm{E}[N_i{}'] \\
&= n(n-1)p_ip_i{}'[1+(n-2)p_i] - np_i(1+(n-1)p_i)np_j \\
&= np_ip_i{}'\left[(n-1)(1+(n-2)p_i) - n(1+(n-1)p_i)\right].
\end{aligned} \tag{36}$$

From (30) and (25):

$$\begin{aligned}
\mathrm{Cov}[N_iN_i{}', N_i] &= \mathrm{E}[N_i^2N_i{}'] - \mathrm{E}[N_iN_i{}']\mathrm{E}[N_i] \\
&= n(n-1)p_ip_i{}'[1+(n-2)p_i] - n(n-1)p_ip_i{}'np_i \\
&= n(n-1)p_ip_i{}'[1-2p_i], \\
\mathrm{Cov}[N_iN_i{}', N_i{}'] &= n(n-1)p_i{}'p_i[1-2p_i{}'], \\
\mathrm{Cov}[N_iN_i{}'', N_i] &= n(n-1)p_ip_i{}''[1-2p_i].
\end{aligned}$$

From (31), (24), and (25):

$$\begin{aligned}\mathrm{Cov}[N_i N_i'', N_i'] &= \mathrm{E}[N_i N_i'' N_i'] - \mathrm{E}[N_i N_i'']\mathrm{E}[N_i'] \\ &= n(n-1)(n-2)p_i p_i'' p_i' - n(n-1)p_i p_i'' n p_i' \\ &= -2n(n-1)p_i p_i'' p_i'.\end{aligned} \tag{37}$$

Hence, $\mathrm{Cov}[M_i, W_i]$ is derived:

$$\begin{aligned}\mathrm{Cov}[M_i, W_i] = n^{-3}\big(&np_i[1 + p_i((n-1)[3-2p_i] - n)] \\ &+ np_i p_i'[(n-1)(1+(n-2)p_i) - n(1+(n-1)p_i)] \\ &+ n(n-1)p_i p_i'[1-2p_i] \\ &+ n(n-1)p_i' p_i[1-2p_i'] \\ &+ n(n-1)p_i p_i''[1-2p_i] \\ &- 2n(n-1)p_i p_i'' p_i'\big). \\ = n^{-2}p_i\big(&1 + p_i((n-1)[3-2p_i] - n) \\ &+ p_i'[(n-1)(1+(n-2)p_i) - n(1+(n-1)p_i)]\big) \\ &+ n^{-2}(n-1)(p_i p_i'[2-2p_i-2p_i'] + p_i p_i''[1-2p_i-2p_i']).\end{aligned} \tag{38}$$

Then, moving on to $\mathrm{Cov}[M_i, W_j]$ and $\mathrm{Cov}[W_i, M_j]$:

$$\begin{aligned}\mathrm{Cov}[M_i, W_j] &= n^{-3}\mathrm{Cov}\left[N_i(N_i + N_i' + N_i''), N_j + N_j'\right] \\ &= n^{-3}\mathrm{Cov}\left[N_i^2 + N_i N_i' + N_i N_i'', N_j + N_j'\right] \\ &= n^{-3}\big(\mathrm{Cov}[N_i^2, N_j] + \mathrm{Cov}[N_i^2, N_j'] \\ &\qquad + \mathrm{Cov}[N_i N_i', N_j] + \mathrm{Cov}[N_i N_i', N_j'] + \mathrm{Cov}[N_i N_i'', N_j] + \mathrm{Cov}[N_i N_i'', N_j']\big).\end{aligned}$$

The terms in the expression above are given below. From (36):

$$\begin{aligned}\mathrm{Cov}[N_i^2, N_j] &= np_i p_j\left[(n-1)(1+(n-2)p_i) - n(1+(n-1)p_i)\right], \\ \mathrm{Cov}[N_i^2, N_j'] &= np_i p_j'\left[(n-1)(1+(n-2)p_i) - n(1+(n-1)p_i)\right].\end{aligned}$$

From (37):

$$\begin{aligned}\mathrm{Cov}[N_i N_i', N_j] &= -2n(n-1)p_i p_i' p_j, \\ \mathrm{Cov}[N_i N_i', N_j'] &= -2n(n-1)p_i p_i' p_j', \\ \mathrm{Cov}[N_i N_i'', N_j] &= -2n(n-1)p_i p_i'' p_j, \\ \mathrm{Cov}[N_i N_i'', N_j'] &= -2n(n-1)p_i p_i'' p_j'.\end{aligned}$$

Hence, $\mathrm{Cov}[M_i, W_j]$ and $\mathrm{Cov}[W_i, M_j]$ are derived:

$$\begin{aligned}\mathrm{Cov}[M_i, W_j] &= n^{-3}\big[np_i(p_j + p_j')[(n-1)(1+(n-2)p_i) - n(1+(n-1)p_i)] \\ &\qquad - 2n(n-1)p_i(p_i' p_j + p_i' p_j' + p_i'' p_j + p_i'' p_j')\big] \\ &= n^{-2}p_i(p_j + p_j')\big[(n-1)(1+(n-2)p_i - 2(p_i' + p_i'')(p_j + p_j')) \\ &\qquad - n(1+(n-1)p_i)\big], \\ \mathrm{Cov}[W_i, M_j] &= n^{-2}p_j(p_i + p_i')\big[(n-1)(1+(n-2)p_j - 2(p_j' + p_j'')(p_i + p_i')) \\ &\qquad - n(1+(n-1)p_j)\big].\end{aligned} \tag{39}$$

Thus, all quantities necessary to compute (17) are derived.