Joint structure learning and causal effect estimation for categorical graphical models

Biometrics, 2024, 80(3), ujae067
https://doi.org/10.1093/biomtc/ujae067
Biometric Methodology

Joint structure learning and causal effect estimation for categorical graphical models

Federico Castelletti [1,*], Guido Consonni[1], Marco L. Della Vedova[2]

[1]Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milan 20123, Italy, [2]Department of Mechanics and Maritime Sciences, Chalmers University of Technology, Hörsalsvägen 7A, Göteborg SE-41296, Sweden

[*]Corresponding author: Federico Castelletti, Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milan 20123, Italy (federico.castelletti@unicatt.it).

ABSTRACT

The scope of this paper is a multivariate setting involving categorical variables. Following an external manipulation of one variable, the goal is to evaluate the causal effect on an outcome of interest. A typical scenario involves a system of variables representing lifestyle, physical and mental features, symptoms, and risk factors, with the outcome being the presence or absence of a disease. These variables are interconnected in complex ways, allowing the effect of an intervention to propagate through multiple paths. A distinctive feature of our approach is the estimation of causal effects while accounting for uncertainty in both the dependence structure, which we represent through a directed acyclic graph (DAG), and the DAG-model parameters. Specifically, we propose a Markov chain Monte Carlo algorithm that targets the joint posterior over DAGs and parameters, based on an efficient reversible-jump proposal scheme. We validate our method through extensive simulation studies and demonstrate that it outperforms current state-of-the-art procedures in terms of estimation accuracy. Finally, we apply our methodology to analyze a dataset on depression and anxiety in undergraduate students.

KEYWORDS: Bayesian inference; categorical data; causal inference; directed acyclic graph; reversible jump Markov chain Monte Carlo.

1 INTRODUCTION

The general framework of this paper is a multivariate setting consisting of categorical variables. Following an external manipulation of one variable, the primary goal is to evaluate the causal effect of this intervention on an outcome of interest. A typical example is represented by a system of variables representing lifestyle, physical and mental features, symptoms, and risk factors, and the outcome is the presence or absence of a disease. All these variables are interconnected in a complex way, which must be learned and taken into account because the effect could propagate along several paths. For an alternative example, consider health data on functioning and disability. Here, the variables represent categories identified according to the International Classification of Functioning, Disability and Health (ICF) (Stucki et al., 2007). ICF categories are organized into 2 parts, each consisting of different components. The first part covers functioning and disability and includes the components "Body Functions and Structures" and "Activities and Participation." The second part covers contextual factors with the components "Environmental Factors" and "Personal Factors." In this framework, Kalisch et al. (2010) analyzed data on patients with spinal cord injury (SCI) resulting from a multicenter, cross-sectional study conducted in 14 countries; see Biering-Sørensen et al. (2006) for further details. After preprocessing, the dataset included around 200 ICF categories mostly from Body Functions/Structures and Activities and Participation. The authors carefully investigated the dependence structure among the variables (categories) and determined the causal effects on a critical outcome variable, "General Health Perception" (ghp), following an intervention on the remaining items present in the dataset. Among the 5 most influential variables on ghp, 4 turned out to belong to the Activities and Participation group (the top one being "Doing housework"), with only 1 ("Sensation of pain") belonging to the Body Functions group. The practical implication of these findings is to inform policies (eg, therapy) available to health care providers.

The analyses carried out on the IFC-SCI data were performed using techniques based on graphical models (Lauritzen, 1996), and more specifically directed acyclic graphs (DAGs) (Koller and Friedman, 2009), where nodes represent variables. This is also the broad methodological framework embraced in our paper. Our distinctive contributions include (i) a Bayesian graphical model for multivariate categorical variables that simultaneously accounts for DAG structure and model parameter uncertainty; (ii) a method for Bayesian Model Averaging (BMA) inference on the causal effects induced by external manipulations of variables; (iii) an efficient computational scheme to perform tasks (i) and (ii).

The rest of this section recaps the basic facts about Bayesian learning of graph structures using observational data and causal inference based on DAG models. To perform causal inference, a DAG model has to be equipped with suitable causal semantics (Pearl, 2000). Alternatively, a (causal) structural equation model

Received: August 13, 2023; Revised: June 13, 2024; Accepted: July 4, 2024

could be employed (Pearl, 1995) but is not discussed in this article. Our causal model is represented by a family of *observational* probability distributions that satisfy the Markov factorization implied by a DAG (Sadeghi, 2017). The term "causal" becomes meaningful through the *do-calculus* (Pearl, 2000), a technique to determine the *interventional distribution* resulting from an external manipulation of variables in the system. A notable feature of the interventional distribution is that it is expressed in terms of the observational distribution, which is estimable from the data; as a consequence, under a few further assumptions (notably that there exist no hidden confounders), causal queries can be answered even when the data are purely observational.

A causal model is predicated on a *given* DAG. In real-world applications, however, the generating DAG is unknown and thus needs to be learned. A difficulty we face is that the true generating DAG is not identifiable in general from purely observational data because its conditional independencies can be encoded in different DAGs that can be grouped into a (Markov) *equivalence class*; identifiability can be reached but this requires specific distributional assumptions; see, for instance, Peters and Bühlmann (2014), Mahdi Mahmoudi and Wit (2018), and Shimizu et al. (2006). Because only a Markov equivalence class can be inferred from data, it follows that there exists a whole *collection* of causal effects (one for each DAG in the class); see Maathuis et al. (2009) for methods to identify these effects in high-dimensional multivariate Gaussian models.

Historically, DAGs were introduced as an engine for probabilistic expert systems with categorical/discrete variables as nodes, and in that setting they acquired the name *Bayesian networks* (Pearl, 1988). Causal discovery for Bayesian networks can be traced back to Heckerman et al. (1995); see also Scutari and Denis (2014) and Roverato (2017) for a more recent account. In this context, Madigan et al. (1996), Castelo and Perlman (2004), and more recently, Castelletti and Peluso (2021) focus on learning equivalence classes. A large part of recent methodological research in causal inference is, however, framed in terms of continuous multivariate distributions (Maathuis et al., 2009; Castelletti and Consonni, 2021). The methodology of Maathuis et al. (2009) was later adapted to categorical distributions by Kalisch et al. (2010). To this end, they first implement the PC algorithm to estimate a Markov equivalence class of DAGs and then compute a battery of possible causal effects for variables of interest using do-calculus rules. Their method relies on a single completed partially DAG (CPDAG) representing the estimated Markov equivalence class of DAGs; accordingly, no uncertainty around such graph estimate is provided, unlike in our approach.

The remaining part of this paper is structured as follows. Section 2 presents relevant notation, the model formulation, and the allied priors; Section 3 specifies the causal effect as the main parameter of inference; and Section 4 details our computational strategy leading up to a BMA estimate of the causal effect. The performance of our method, including comparisons with alternative approaches, is presented in Section 5, while Section 6 presents an application to depression and anxiety data. The final section offers a brief discussion together with possible future developments.

2 BAYESIAN INFERENCE OF CATEGORICAL DAG MODELS

2.1 Categorical data and notation

Let $X = (X_j,\ j \in V)^\top$, $V = \{1, \ldots, q\}$, be a $(q, 1)$ vector of categorical random variables with X_j taking values in the corresponding set of levels $\mathcal{X}_j$, whose generic element is x_j. We let $\mathcal{X} := \times_{j\in V}\mathcal{X}_j$ be the product space of the sets of levels. The collection of joint probabilities $\boldsymbol{\pi} = \{\pi_x,\ x \in \mathcal{X}\}$ can be arranged in a q-way contingency table of probabilities, where each cell refers to a specific level $x \in \mathcal{X}$. For any given $S \subseteq V$, we let $X_S = (X_j,\ j \in S)$ be the sub-vector of X with components indexed by S, and $x_S \in \mathcal{X}_S := \times_{j\in S}\mathcal{X}_j$ one of its levels. We then let $\pi^S_{x_S} = \Pr(X_S = x_S \mid \boldsymbol{\pi})$ be the corresponding marginal joint probability for variables in S. We instead write $\theta^{j\mid S}_{x_j\mid x_S} = \Pr(X_j = x_j \mid X_S = x_S, \boldsymbol{\pi})$ to denote the conditional probability for variable X_j evaluated at x_j, given configuration x_S of variables in S, $j \notin S$.

Consider now n observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}$ from X, where $\boldsymbol{x}^{(i)} = (x^{(i)}_1, \ldots, x^{(i)}_q)^\top \in \mathcal{X}$ for $i = 1, \ldots, n$. For any $x \in \mathcal{X}$, we can compute the count $n_x = \sum_{i=1}^n \mathbb{1}(\boldsymbol{x}^{(i)} = x)$, that is, the number of observations that are equal to x, and organize the resulting collection of values in a q-way contingency table of counts $\boldsymbol{N} = \{n_x, x \in \mathcal{X}\}$. In addition, for any $x_S \in \mathcal{X}_S$, we let $n^S_{x_S} = \sum_{i=1}^n \mathbb{1}(\boldsymbol{x}^{(i)}_S = x_S)$ and $\boldsymbol{N}_S = \{n^S_{x_S}, x_S \in \mathcal{X}_S\}$ be the allied $|S|$-way marginal contingency table of counts.

2.2 Model formulation

Let $\mathcal{D} = (V, E)$ be a DAG, with set of nodes V, one for each of the q variables, and $E \subseteq V \times V$ its set of directed edges. If $u \neq v$ and $(u, v) \in E$, then $(v, u) \notin E$, and we say that $\mathcal{D}$ contains the directed edge $u \to v$, where u is a *parent* of v; equivalently, v is a *child* of u. The set of all parents of u in $\mathcal{D}$ is written $\mathrm{pa}_{\mathcal{D}}(u)$, while $\mathrm{fa}_{\mathcal{D}}(u) = u \cup \mathrm{pa}_{\mathcal{D}}(u)$ identifies the *family* of u. In the remainder of this section and in Section 3, we reason *conditionally* on a single given DAG, which, for simplicity, is omitted from our notation. Under $\mathcal{D}$, and for any level $x \in \mathcal{X}$, the joint probability function of the random vector X factorizes as

$$p(x) = \Pr\left(X_1 = x_1, \ldots, X_q = x_q\right) = \prod_{j=1}^{q} p\big(X_j = x_j \mid X_{\mathrm{pa}(j)} = x_{\mathrm{pa}(j)}\big). \tag{1}$$

Under a family of probability distributions and given i.i.d. realizations $\{\boldsymbol{x}^{(i)}, i = 1, \ldots, n\}$, the likelihood function becomes

$$p(\boldsymbol{X} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\prod_{x\in\mathcal{X}}\left\{\Pr\left(X^{(i)}_1 = x^{(i)}_1, \ldots, X^{(i)}_q = x^{(i)}_q \mid \boldsymbol{\theta}\right)\right\}^{\mathbb{1}(\boldsymbol{x}^{(i)}=x)}\right\} = \prod_{j=1}^{q}\left\{\prod_{k\in\mathcal{X}_{\mathrm{pa}(j)}}\left\{\prod_{m\in\mathcal{X}_j}\left\{\theta^{j\mid \mathrm{pa}(j)}_{m\mid k}\right\}^{n^{\mathrm{fa}(j)}_{(m,k)}}\right\}\right\}, \tag{2}$$

where $\boldsymbol{X}$ is the (n, q) observed data matrix whose ith row is $(\boldsymbol{x}^{(i)})^\top$. Importantly, the model in (2) is identifiable because it belongs to an exponential family; see Consonni and Massam (2012, Lemma 2.1). For related results, see also Massam and Wesołołowski (2016). Notice that Equation 2 depends on the

raw observations $\boldsymbol{X}$ through the counts $\boldsymbol{N}$, which are the sufficient statistics.

2.3 Parameter prior distributions

We now proceed by assigning a prior distribution to $\boldsymbol{\theta}$. Specifically, consider for each $j \in V$ and each $x_{\mathrm{pa}(j)} \in \mathcal{X}_{\mathrm{pa}(j)}$ the allied set of parameters $\left(\theta^{j\,|\,\mathrm{pa}(j)}_{x_j\,|\,x_{\mathrm{pa}(j)}},\, x_j \in \mathcal{X}_j\right) := \boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_{x_{\mathrm{pa}(j)}}$, where each element is a $|\mathcal{X}_j|$-dimensional vector of conditional probabilities for variable X_j given configuration $x_{\mathrm{pa}(j)}$ of its parents. Clearly, for each $x_{\mathrm{pa}(j)}$, the equality $\sum_{x_j \in \mathcal{X}_j} \theta^{j\,|\,\mathrm{pa}(j)}_{x_j\,|\,x_{\mathrm{pa}(j)}} = 1$ holds. Moreover, let $\left(\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_{x_{\mathrm{pa}(j)}},\, x_{\mathrm{pa}(j)} \in \mathcal{X}_{\mathrm{pa}(j)}\right) := \boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}$ be the collection of conditional probabilities for node j. We introduce the following independence assumptions (Geiger and Heckerman, 1997):

- (G) $\perp\!\!\!\perp_{j \in V} \boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}$ (*global* parameter independence);
- (L) $\perp\!\!\!\perp_{x_{\mathrm{pa}(j)} \in \mathcal{X}_{\mathrm{pa}(j)}} \boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_{x_{\mathrm{pa}(j)}}$ for each variable j (*local* parameter independence).

Furthermore, we assume for each $\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_k$, with $j \in V$ and $k \in \mathcal{X}_{\mathrm{pa}(j)}$,

$$\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_k \sim \mathrm{Dir}\left(\boldsymbol{a}^{j\,|\,\mathrm{pa}(j)}_k\right), \tag{3}$$

a Dirichlet distribution with hyperparameter $\boldsymbol{a}^{j\,|\,\mathrm{pa}(j)}_k = \left(a^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k} > 0,\, m \in \mathcal{X}_j\right)$, whose probability density function is given by

$$\begin{aligned} p\left(\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_k\right) &= \frac{\Gamma\left(\sum_{m \in \mathcal{X}_j} a^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k}\right)}{\prod_{m \in \mathcal{X}_j} \Gamma\left(a^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k}\right)} \prod_{m \in \mathcal{X}_j} \left\{\theta^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k}\right\}^{a^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k} - 1} \\ &= h\left(\boldsymbol{a}^{j\,|\,\mathrm{pa}(j)}_k\right) \prod_{m \in \mathcal{X}_j} \left\{\theta^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k}\right\}^{a^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k} - 1}, \end{aligned} \tag{4}$$

where $h(\cdot)$ is the prior normalizing constant. Equation 4, together with (G) and (L), determines a prior on the overall DAG-parameter

$$\boldsymbol{\theta} = \left\{\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_k,\ j \in V,\ k \in \mathcal{X}_{\mathrm{pa}(j)}\right\}, \tag{5}$$

which factorizes as

$$\begin{aligned} p(\boldsymbol{\theta}) &= \prod_{j=1}^{q} \left\{ \prod_{k \in \mathcal{X}_{\mathrm{pa}(j)}} p\left(\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_k\right) \right\} \\ &= \prod_{j=1}^{q} \left\{ \prod_{k \in \mathcal{X}_{\mathrm{pa}(j)}} p\mathrm{Dir}\left(\boldsymbol{\theta}^{j\,|\,\mathrm{pa}(j)}_k \,\middle|\, \boldsymbol{a}^{j\,|\,\mathrm{pa}(j)}_k\right) \right\}. \end{aligned} \tag{6}$$

The choice of the hyperparameters in (6) requires care especially when several DAGs are entertained and the purpose is DAG model selection. In particular, assuming faithfulness, observational data cannot distinguish between Markov equivalent DAGs; accordingly, the prior on the parameter $\boldsymbol{\theta}$ should guarantee that any two equivalent DAGs are assigned the same *marginal* likelihood; this is the rationale behind the procedure for prior elicitation introduced by Heckerman et al. (1995) leading to their Bayesian Dirichlet equivalent uniform score (BDeu); see also Geiger and Heckerman (2002). Specifically, these authors show that the default choice

$$a^{j\,|\,\mathrm{pa}(j)}_{m\,|\,k} = \frac{a}{|\mathcal{X}_{\mathrm{fa}(j)}|}, \quad j \in V, \quad m \in \mathcal{X}_j, \quad k \in \mathcal{X}_{\mathrm{pa}(j)}, \tag{7}$$

with $a > 0$, guarantees DAG score equivalence. Besides ensuring this *compatibility* requirement, the proposed model provides closed-form expressions for posterior distributions of parameters and marginal likelihoods. We will leverage this feature in Section 4 to develop a Markov chain Monte Carlo (MCMC) sampler targeting the posterior over the space of DAGs and parameters.

3 CAUSAL EFFECTS

The DAG factorization (1) is also called the *observational* (or *pre-intervention*) distribution. Consider now two variables, X_v and $X_h := Y\,(h \neq v)$, where the latter is a response of interest. We are interested in the (total) *causal effect* on Y of an intervention on X_v. In particular, we consider a *hard* intervention on X_v, consisting in the action of forcing its value to a given level $\tilde{x}$, denoted $\mathrm{do}(X_v = \tilde{x})$. Under a hard intervention, the *post-intervention* distribution (Pearl, 2000) is given by the truncated factorization

$$\begin{aligned} &p\left(x \,|\, \mathrm{do}(X_v = \tilde{x})\right) \\ &\quad = \begin{cases} \prod_{j \neq v} p\left(X_j = x_j \,|\, X_{\mathrm{pa}(j)} = x_{\mathrm{pa}(j)}\right) & \text{if } x_v = \tilde{x} \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{8}$$

where each term $p(X_j = x_j \,|\, \cdot)$ is the corresponding (pre-intervention) conditional distribution of Equation 1 and again we omit subscript $\mathcal{D}$ to ease the notation. Assuming for simplicity that both X_v and Y are binary taking values in $\{0, 1\}$, the causal effect on Y resulting from an intervention on X_v can be defined as

$$c_v = \mathbb{E}\left(Y \,|\, \mathrm{do}(X_v = 1)\right) - \mathbb{E}\left(Y \,|\, \mathrm{do}(X_v = 0)\right). \tag{9}$$

Moreover, it can be shown (Pearl, 2000, Theorem 3.2.3) that

$$\begin{aligned} c_v = &\sum_{k \in \mathcal{X}_{\mathrm{pa}(v)}} \mathbb{E}\left(Y \,|\, X_v = 1, X_{\mathrm{pa}(v)} = k\right) \Pr\left(X_{\mathrm{pa}(v)} = k\right) \\ &- \sum_{k \in \mathcal{X}_{\mathrm{pa}(v)}} \mathbb{E}\left(Y \,|\, X_v = 0, X_{\mathrm{pa}(v)} = k\right) \Pr\left(X_{\mathrm{pa}(v)} = k\right), \end{aligned} \tag{10}$$

where the expectations can be alternatively written in terms of conditional probabilities of Y being a success because of its binary nature. Equation 10 uses the set of parents as an adjustment set; however, alternative sets are also available (Pearl, 2000; Henckel et al., 2022). Under model (2) the causal effect becomes

$$\gamma_v(\boldsymbol{\theta}) = \sum_{k \in \mathcal{X}_{\mathrm{pa}(v)}} \left\{ \left(\theta^{Y\,|\,\mathrm{fa}(v)}_{1\,|\,(1,k)} - \theta^{Y\,|\,\mathrm{fa}(v)}_{1\,|\,(0,k)}\right) \theta^{\mathrm{pa}(v)}_k \right\}. \tag{11}$$

Notice that the univariate θ-parameters involved in (11) are *not* the components of the overall DAG-parameter $\boldsymbol{\theta}$ in (5) because

the conditional distribution of $Y \mid X_{\text{fa}(v)}$ does not appear in general in the factorization (2). Yet $\gamma_v(\cdot)$ is a function of $\boldsymbol{\theta}$, so that inference on the causal effect can be retrieved from the posterior distribution of $\boldsymbol{\theta}$, which is the subject of the next section; see also Web Appendix A for examples. When X_v is polytomous, one can define a battery of causal effects. Typically, one would choose a *reference* level for X_v, $\widetilde{m}$ say, and then apply (9) for pairs $(X_v = m, X_v = \widetilde{m})$ with $m \neq \widetilde{m}$. On the other hand, when the levels of the response Y are more than 2, the conditional expectation in (9) should be replaced by the probability that Y attains a suitable benchmark level. Alternatively, a collection of causal effects, one for each level of Y, can be computed and then analyzed to gauge sensitivity.

4 POSTERIOR INFERENCE

Let $\mathcal{S}_q$ be the set of all DAGs with q nodes. In this section, we also regard DAG $\mathcal{D}$ as uncertain and introduce a reversible jump MCMC scheme for joint posterior inference on the DAG structure and the allied parameter. Let $p(\mathcal{D})$ be a prior on $\mathcal{D} \in \mathcal{S}_q$, which will be specified in Section 4.1. Our target is the joint posterior distribution

$$p(\boldsymbol{\theta}, \mathcal{D} \mid \boldsymbol{X}) \propto p(\boldsymbol{X} \mid \boldsymbol{\theta}, \mathcal{D})\, p(\boldsymbol{\theta} \mid \mathcal{D})\, p(\mathcal{D}), \tag{12}$$

where we now emphasize the dependence on DAG $\mathcal{D}$ both in the likelihood and prior.

4.1 Prior on DAG $\mathcal{D}$

We assign a prior on DAGs belonging to $\mathcal{S}_q$ as follows. For a given DAG $\mathcal{D} = (V, E) \in \mathcal{S}_q$, let $\boldsymbol{S}^{\mathcal{D}}$ be the $0-1$ *adjacency matrix* of its skeleton, which is the underlying undirected graph obtained after removing the orientation of all its edges. For each (u, v)-element of $\boldsymbol{S}^{\mathcal{D}}$, we have $\boldsymbol{S}^{\mathcal{D}}_{u,v} = 1$ if and only if $(u, v) \in E$ or $(v, u) \in E$, zero otherwise. Conditionally on a prior probability of inclusion $\eta \in (0, 1)$ we assume, for each $u > v$, $\boldsymbol{S}^{\mathcal{D}}_{u,v} \mid \eta \overset{\text{iid}}{\sim} \text{Ber}(\eta)$, which implies $p(\boldsymbol{S}^{\mathcal{D}} \mid \eta) = \eta^{|\boldsymbol{S}^{\mathcal{D}}|}(1-\eta)^{\frac{q(q-1)}{2} - |\boldsymbol{S}^{\mathcal{D}}|}$, where $|\boldsymbol{S}^{\mathcal{D}}|$ is the number of edges in $\mathcal{D}$ (equivalently in its skeleton), and $q(q-1)/2$ is the maximum number of edges in a DAG on q nodes. We then assume $\eta \sim \text{Beta}(c, d)$, so that, by integrating out η, the resulting prior on $\boldsymbol{S}^{\mathcal{D}}$ is

$$p(\boldsymbol{S}^{\mathcal{D}}) = \frac{\Gamma\left(|\boldsymbol{S}^{\mathcal{D}}| + c\right)\Gamma\left(\frac{q(q-1)}{2} - |\boldsymbol{S}^{\mathcal{D}}| + d\right)}{\Gamma\left(\frac{q(q-1)}{2} + c + d\right)} \cdot \frac{\Gamma(c+d)}{\Gamma(c)\,\Gamma(d)}. \tag{13}$$

Finally, we set $p(\mathcal{D}) \propto p(\boldsymbol{S}^{\mathcal{D}})$ for each $\mathcal{D} \in \mathcal{S}_q$.

4.2 MCMC scheme and posterior summaries

To approximate the posterior (12), we develop an MCMC scheme. This is presented in Web Appendix B and is based on a Partial Analytic Structure (PAS) algorithm, which iteratively updates DAG $\mathcal{D}$ and the DAG-parameter $\boldsymbol{\theta}$ by sampling from their full conditional distributions.

Its output is a collection of DAGs and corresponding DAG-parameters $\left\{\left(\boldsymbol{\theta}^{(1)}, \mathcal{D}^{(1)}\right), \ldots, \left(\boldsymbol{\theta}^{(S)}, \mathcal{D}^{(S)}\right)\right\}$, approximately sampled from (12), where S is the number of final MCMC iterations. An approximate marginal posterior distribution over the DAG space $\mathcal{S}_q$ can be computed as

$$\widehat{p}(\mathcal{D} \mid \boldsymbol{X}) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\left(\mathcal{D}^{(s)} = \mathcal{D}\right) \tag{14}$$

for any $\mathcal{D} \in \mathcal{S}_q$, where $\mathbb{1}(\cdot)$ is the indicator function, and whose expression corresponds to the MCMC frequency of visits of $\mathcal{D}$. In addition, for any directed edge (u, v), we can estimate a marginal posterior probability of edge inclusion (PPI) as

$$\widehat{p}(u \to v \mid \boldsymbol{X}) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\left(u \to v \in \mathcal{D}^{(s)}\right). \tag{15}$$

Starting from the previous quantities, single DAG estimates summarizing the MCMC output can be recovered: a maximum a posteriori estimate, corresponding to the DAG with the highest posterior probability (14) or a median probability model (MPM) estimate, obtained by including only those edges whose PPI (15) is greater than 0.5.

For a given node $v \in \{2, \ldots, q\}$, consider now the causal effect of $\text{do}(X_v = \tilde{x})$ on Y, represented by the parameter $\gamma_v(\boldsymbol{\theta})$ in (11). For each draw $\boldsymbol{\theta}^{(s)}$ from the posterior (12), we can first recover $\gamma_v\left(\boldsymbol{\theta}^{(s)}\right)$ using Equation 11. An estimate of $\gamma_v(\boldsymbol{\theta})$ is then

$$\widehat{\gamma}_v^{\text{BMA}} = \frac{1}{S}\sum_{s=1}^{S} \gamma_v\left(\boldsymbol{\theta}^{(s)}\right), \tag{16}$$

which implicitly performs BMA through the MCMC frequencies of the visited DAGs.

5 SIMULATION STUDY

We illustrate the performance of our methodology through simulation. Specifically, we consider different scenarios in which we vary the number of variables $q \in \{10, 20\}$ and the sample size $n \in \{200, 500, 1000, 2000\}$. For each choice of q, we randomly generate $G = 50$ DAGs with probability of edge inclusion $2/q$ reflecting sparsity. Under each DAG, a dataset consisting of n observations from q categorical variables is generated as described in Web Appendix C. Eventually, a collection of $G = 50$ DAGs and allied datasets is available under each scenario defined by $\{q, n\}$. In the same section of the Web Appendix, we provide details on the computation of the true causal effect γ_v^* for each node $v \in \{2, \ldots, q\}$, and with node $Y = X_1$ as the response.

5.1 Results

We apply our MCMC scheme to approximate the joint posterior distribution in (12). To this end, we let the number of MCMC iterations S vary in the set $\{5000, 10\,000\}$ for, respectively, $q \in \{10, 20\}$, disregarding from the output a burn-in period of size $B \in \{1000, 2000\}$ for the two values of q, respectively. Moreover, we set the common hyperparameter of the Dirichlet prior in (7) as $a = 1$ and $c = d = 1$ in the $\text{Beta}(c, d)$ prior for the probability of edge inclusion η leading to the prior on DAG-space $p(\mathcal{D})$; see Section 4.1.

We start by evaluating the global performance of our method in learning the underlying graphical structure. Specifically, we first estimate the posterior probabilities of edge inclusion as in (15) for each pair of distinct nodes (u, v) and

TABLE 1 Simulations. Average (w.r.t. 50 simulations) Structural Hamming Distance (SHD), sensitivity (SEN), and specificity (SPE) indexes, computed under each scenario defined by number of variables $q \in \{10, 20\}$ and sample size $n \in \{200, 500, 1000, 2000\}$.

		$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
$q = 10$	SHD	6.35	5.22	4.50	4.15
	SEN	56.15	69.62	78.28	82.36
	SPE	96.23	95.92	95.47	95.70
$q = 20$	SHD	15.47	12.55	12.10	11.40
	SEN	50.27	66.15	73.02	74.37
	SPE	98.10	97.95	97.50	97.61

TABLE 2 Simulations. Average (w.r.t. 50 simulations and intervened nodes) absolute error (AE) between true and estimated causal effect (values multiplied by 100), computed under each scenario defined by number of variables $q \in \{10, 20\}$ and sample size $n \in \{200, 500, 1000, 2000\}$.

	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
$q = 10$	4.46	3.70	3.50	3.28
$q = 20$	2.17	1.80	1.74	1.65

produce an MPM estimate of the DAG, $\widehat{\mathcal{D}}$. The latter is compared with the true DAG $\mathcal{D}$ in terms of sensitivity (SEN) and specificity (SPE) indexes, respectively, defined as $\text{SEN} = \text{TP}/(\text{TP} + \text{FN})$, $\text{SPE} = \text{TN}/(\text{TN} + \text{FP})$, where TP, TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively, which can be recovered from the 0-1 adjacency matrix of the estimated graphs. As an overall summary, we also consider the Structural Hamming Distance (SHD), defined as the number of insertions and deletions of flips needed to transform the estimated graph into the true graph. Results, averaged w.r.t. the $G = 50$ simulations under each scenario defined by q and n, are summarized in Table 1. Both the SHD and SEN metrics suggest that the accuracy of our method in recovering the true DAG improves as the number of available data grows; moreover, the SPE index attains high levels even for the smallest value of n, and is essentially stable as the sample size grows; accordingly, the method shows an overall appreciable performance.

We now consider causal effect estimation. To this end, we produce the collection of BMA estimates $\widehat{\gamma}_v^{\text{BMA}}$, $v \in \{2, \ldots, q\}$ according to Equation 16. We compare each BMA estimate with the corresponding true causal effect γ_v^*, and compute the absolute error (AE)

$$\text{AE}_v = |\gamma_v^* - \widehat{\gamma}_v^{\text{BMA}}|. \tag{17}$$

Results are summarized in Table 2, where we report for each value of q and n the average value of the AE $\times$ 100 (computed across the 50 simulated DAGs and nodes $v = 2, \ldots, q$). By increasing the sample size the difference between estimated and true causal effect progressively reduces.

5.2 Comparisons with PC algorithm, HC, and IDA approach

In this section, we compare the performance of our Bayesian methodology with the IDA (identification when DAG is absent) approach of Maathuis et al. (2009), originally introduced for Gaussian data and adapted to a categorical setting in Kalisch et al. (2010). IDA estimates first a CPDAG using the PC algorithm (Spirtes et al., 2000; Kalisch and Bühlmann, 2007). The latter is based on a sequence of conditional independence tests that we implement for significance level $\alpha \in \{1\%, 5\%, 10\%\}$. The resulting CPDAG represents a Markov equivalence class of DAGs; although these are equivalent in terms of conditional independencies, they can lead in principle to distinct causal effects for the same intervention. Accordingly, Maathuis et al. (2009) propose two different strategies for causal effect estimation. The first enumerates all DAGs in the equivalence class and for each one estimates the causal effect. As this approach is computationally expensive, even for moderate values of q, a second algorithm (hereinafter considered), which only outputs the *distinct* causal effects within a given equivalence class, is implemented. Finally, an average causal effect, computed across all distinct causal effects compatible with the estimated CPDAG, is returned. Each of the distinct causal effect coefficients is computed as in Equation 11 upon replacing marginal and conditional probabilities with the corresponding sample proportions. We refer to the resulting estimate as γ_v^{IDA}. Finally, notice that the PC algorithm provides a CPDAG estimate, rather than a DAG. For comparison purposes, we then recover from our MPM DAG estimate the representative CPDAG.

For the purpose of structure learning underlying the IDA approach, we also consider a Hill Climbing (HC) score-based method (Russell and Norvig, 2009). HC is an optimized greedy search algorithm that explores the space of DAGs by single-arc additions, removals, and reversals. We implement HC with both the Bayesian Information Criterion (HC BIC) and the Bayesian Dirichlet equivalent uniform score (HC BDeu) of Heckerman et al. (1995); see also Russell and Norvig (2009). Both HC BIC and HC BDeu output a DAG estimate, for which we construct the representative CPDAG. Then, the IDA approach for causal effect estimation is applied as described above.

Figure 1 summarizes the distribution of SHD computed across the 50 simulations under each method and for different values of q and n. In general, it appears that the results of our method improve as the sample size grows, while for PC and HC, this holds only for moderate sample sizes (from 200 to 500), because when n increases from 1000 to 2000 the performance slightly deteriorates. Our Bayesian method adapted to output an MPM-based CPDAG is therefore highly competitive with all three versions of PC and outperforms both HC BIC and HC BDeu; moreover, it shows an overall better performance across sample sizes when considering the median value of the distribution, while variability is comparable to mildly larger.

Finally, we consider causal effect estimation and report in Figure 2, for each method and different combinations of q and n, the boxplot of the AE, again computed across the 50 simulated DAGs and nodes subject to intervention. While all methods improve as n grows for both values of q, our Bayesian methodology based on a BMA estimate of the causal effect outperforms the IDA method under all scenarios. The relative inaccuracy of IDA is strictly related to the poor performance of both the PC and HC algorithms in recovering the true CPDAG. This in turn affects the correct identification of the set of distinct causal effects leading to the IDA estimate. By contrast, our BMA output is typically

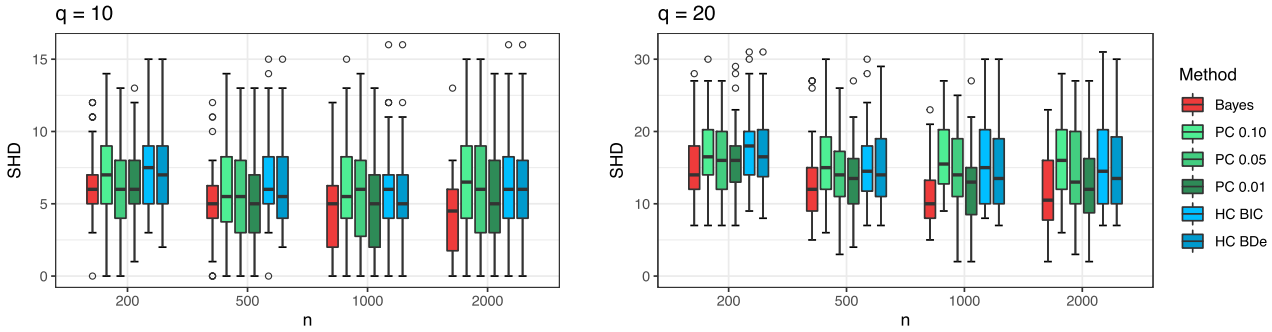

FIGURE 1 Simulations. Structural Hamming Distance (SHD) between true and estimated CPDAGs for number of nodes $q \in \{10, 20\}$ and increasing samples sizes $n \in \{200, 500, 1000, 2000\}$. Methods under comparison are as follows: our Bayesian proposal (Bayes) leading to the MPM CPDAG estimate, the PC algorithm implemented for significance levels $\alpha \in \{0.10, 0.05, 0.01\}$ (respectively, PC 0.10, PC 0.05, PC 0.01), and the Hill Climbing algorithm with BIC and BDeu scores (respectively, HC BIC, HC BDeu).

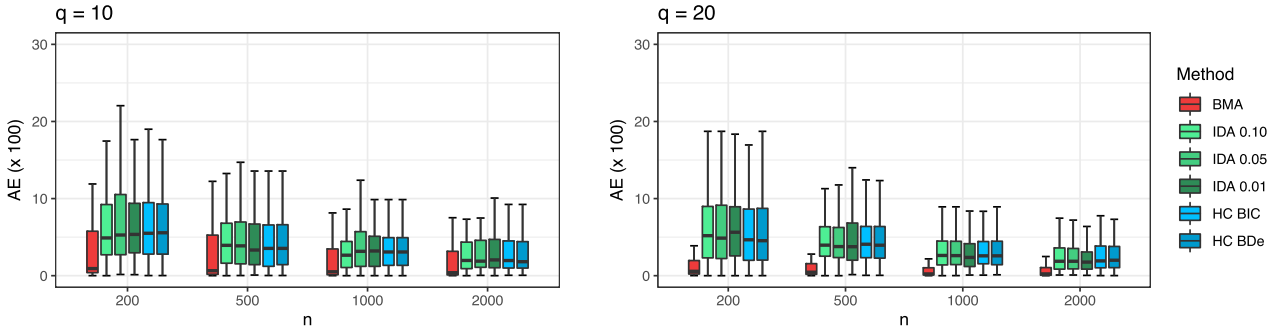

FIGURE 2 Simulations. Absolute error (AE) between true and estimated causal effects (values multiplied by 100) for number of nodes $q \in \{10, 20\}$ and increasing samples sizes $n \in \{200, 500, 1000, 2000\}$. Methods under comparison are as follows: our Bayesian proposal with the BMA causal effect estimate (BMA) and the IDA method based on the PC algorithm implemented for significance levels $\alpha \in \{0.10, 0.05, 0.01\}$ (respectively, PC 0.10, PC 0.05, and PC 0.01) and on the Hill Climbing algorithm with BIC and BDeu scores (respectively, HC BIC and HC BDeu).

based on a larger collection of DAGs, which, although possibly outside the equivalence class of the true CPDAG for some fraction of the iterations, may well lead to a causal effect that is closer to the true value because of the similarity in the corresponding causal pathway.

6 APPLICATION TO ANXIETY AND DEPRESSION DATA

We consider a dataset relative to a study on depression and anxiety in undergraduate students. Depression represents a serious illness especially among young people, which can be identified through several symptoms such as feelings of melancholy and emptiness, disturbed sleep, or loss of interest in social activities. In addition, it is strictly related to anxiety disorders and stress. Several therapies for the treatment of depression and anxiety have been proposed, and many of these have shown beneficial effects on patients in terms of a complete or partial restore of social behavior and mental conditions.

The dataset was collected from $n = 787$ undergraduate students at the University of Lahore. Variables in the analyzed dataset include depression diagnosis (`depr`, the absence/presence of depressive status), anxiety diagnosis (`anx`, the absence/presence of anxiety disorder), and 2 related variables indicating the administration or not of a therapy against depression or anxiety (`depr treat` and `anx treat`, respectively), besides other features such as `gender`, body max index (`bmi`, a categorical variable with 2 levels, normal/abnormal), suicidal instinct (`suicidal`), and 2 variables linked to daytime sleepiness: `sleep` and its measure based on the Epworth scale (`epworth`). Most variables are recorded as binary; scores were instead dichotomized.

We implement our method for structure learning and causal effect estimation by running $S = 40\,000$ iterations of our MCMC scheme after a burn-in period of 5000 runs. We summarize the output by reporting, for each directed edge $u \to v$ and each pair of variables in the dataset, the corresponding posterior probability of inclusion (Equation 15). Results are displayed in the heat map reported in the left-side panel of Figure 3. In addition, we provide a summary of the posterior distribution over the DAG space by constructing the MPM DAG estimate. The CPDAG representing the Markov equivalence class of the estimated graph, which is reported in the right-side panel of Figure 3, is highly sparse as it contains only 10 edges, together with 3 unrelated components (in addition to the separate variable BMI): One involving the anxiety-depression diagnosis/measurement

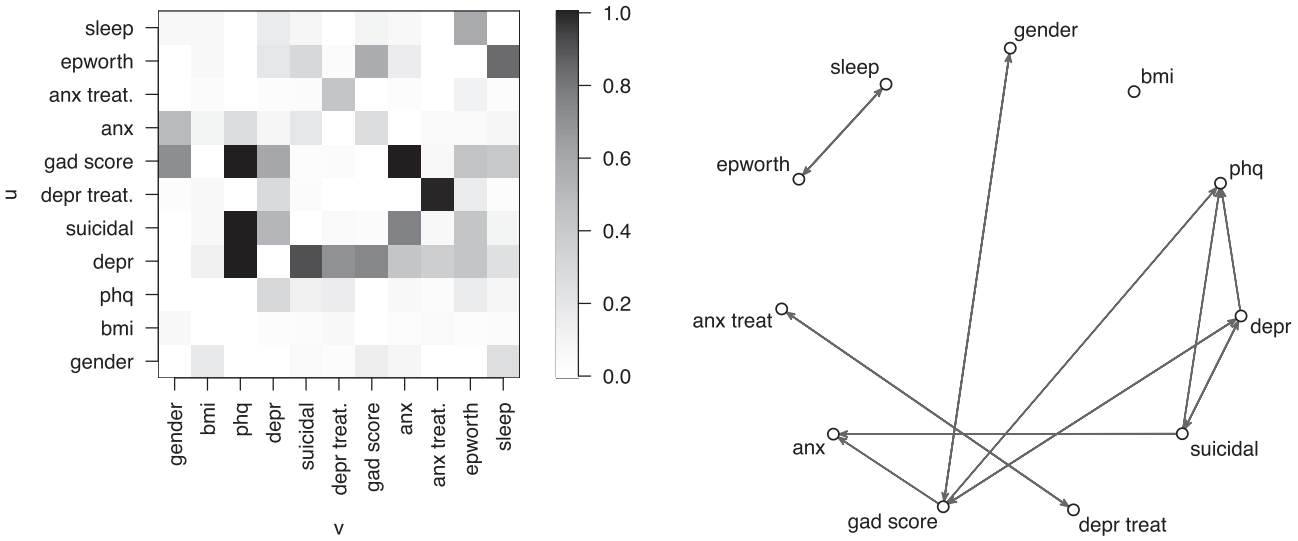

FIGURE 3 Anxiety and depression data. Heat map with estimated posterior probabilities of edge inclusion for each edge $u \to v$ (left panel); CPDAG representing the equivalence class of the median probability DAG estimate (right panel).

TABLE 3 Anxiety and depression data. Posterior summaries (mean, standard deviation, and posterior probabilities of negative, null, and positive values) for the 2 causal effect coefficients considered in the study.

	$\mathbb{E}(\gamma_v \mid \boldsymbol{X})$	$\mathrm{sd}(\gamma_v \mid \boldsymbol{X})$	$P(\gamma_v < 0 \mid \boldsymbol{X})$	$P(\gamma_v = 0 \mid \boldsymbol{X})$	$P(\gamma_v > 0 \mid \boldsymbol{X})$
Depression	−0.117	0.243	0.578	0.320	0.102
Anxiety	−0.126	0.222	0.818	0.141	0.041

variables, one involving the 2 treatment variables, and finally the sleepiness block.

Variables that appear to be directly linked to depression status are the `phq` (Patient Health Questionnaire Score) and `gad score` (Generalized Anxiety Disorder Index), here included as binary variables with levels high and low, besides `suicidal`. On the other hand, both `gender` and `bmi` do not seem to influence directly the depression or anxiety status.

We now focus on causal effect estimation. Specifically, it is of interest to evaluate the efficacy of the two therapies for depression and anxiety. Accordingly, we consider `depr` as the response of interest Y in our causal-effect analysis and evaluate the causal effect on `depr` of an intervention on `depr treat` (X_v); similarly, we consider the same analysis for intervention target `anx treat` and response variable `anx`. Notice that the estimated CPDAG exhibits no causal paths from either `depr treat` and `anx treat` to `depr` and `anx`, respectively. However, this is subject to an overall degree of uncertainty because of the presence of many directed edges whose individual probability of edge inclusion is below the 0.5 threshold, and yet lead to causal paths contributing to non-zero causal effects in our BMA estimate; see the heat map in Figure 3. We emphasize that any alternative method based on a *single* graph estimate such as the CPDAG in Figure 3 would fail to provide causal effects that are different from zero.

We recover from our MCMC output the posterior distribution of the two causal effect parameters computed according to Equation 11. Summaries in terms of posterior mean, standard deviation, and probabilities of a causal effect being null, negative, or positive are reported in Table 3.

The posterior means of the two parameters, namely our BMA estimates, are both around −0.12, suggesting that both therapies have a beneficial effect on the status of depression and anxiety. It also appears that the probability of negative causal effect (suggesting that the therapy lessens depression or anxiety) is higher than 80% for variable `anx treat`, while it is substantially smaller for `depr treat` (only around 58%). Accordingly, the probabilities of null-or-positive effects (which would imply ineffective treatments) are just around 18% for the anxiety treatment, while not negligible (42%) for depression treatment. In conclusion, these summaries provide useful statistical information to evaluate the effectiveness of the two therapies.

7 DISCUSSION

In this paper, we have considered multivariate categorical observations and proposed a novel graphical model framework for causal inference. Specifically, our Bayesian methodology combines structure learning and parameter inference for categorical DAG models. From a computational perspective, we implemented an MCMC scheme based on a PAS algorithm to approximate the joint posterior distribution over DAG structures and DAG-parameters. Starting from this MCMC output, the full posterior distribution of the causal effects between any pairs of variables of interest can be recovered, and eventually summarized through BMA, which naturally incorporates uncertainty around the (unknown) underlying DAG model. We evaluated our method through simulation studies, and demonstrated that it outperforms alternative state-of-the-art strategies for

causal effect estimation. Additionally, our method employs exact formulae based on conditional probabilities when computing causal effects, and does not require further assumptions unlike in Kalisch et al. (2010, Supplement).

Our model formulation is based on the assumption of i.i.d. sample observations from a *single* categorical graphical model (which, however, is unknown, or rather uncertain from a Bayesian perspective). This assumption can be relaxed in two different directions to allow for heterogeneity among individuals belonging to different subgroups of the same population. When groups are known beforehand, one can consider a model comprising *multiple* distinct graphical structures coupled with a Markov random field prior that encourages common edges between groups, and a spike-and-slab prior on network relatedness parameters (Castelletti et al., 2020). Causal effect estimation at group-specific level would benefit from borrowing information across subjects belonging to distinct, yet related groups.

On the other hand, when subgroups are not available a priori, one can set up a *mixture* model, either with a finite or an infinite number of components, allowing for joint posterior inference on DAGs, parameters as well as clustering. A Bayesian nonparametric Dirichlet Process mixture of Gaussian DAG models is considered in Castelletti and Consonni (2023) for causal inference under heterogeneity. Their general framework can be adapted to categorical DAGs and would lead to causal effect estimates at cluster as well as subject-specific level.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices referenced in Sections 2–5 are available with this paper at the Biometrics website on Oxford Academic. These include (A) details on the computation of causal effects from DAG-parameters, (B) a presentation of our MCMC scheme, (C) a description of data generation for the simulation studies of Section 5, (D) additional simulations with polytomous variables, and (E) analyses of sensitivity to prior hyperparameters. R code implementing our methodology is also available at the Biometrics website on Oxford Academic, and at https://github.com/FedeCastelletti/bayes_structure_causal_categorical_graphs.

FUNDING

Work carried out within MUR-PRIN grant 2022 SMNNKY—CUP J53D23003870008, funded by the European Union—Next Generation EU. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. Partial support from UCSC (D1 and 2019-D.3.2 research grants) is also acknowledged by F.C. and G.C.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The dataset analyzed in Section 6 is publicly available at https://www.kaggle.com/datasets/ under the name *Depression and anxiety data*.

REFERENCES

Biering-Sørensen, F., Scheuringer, M., Baumberger, M., Charlifue, S., Post, M., Montero, F. et al. (2006). Developing core sets for persons with spinal cord injuries based on the International Classification of Functioning, Disability and Health as away to specify functioning. *Spinal Cord*, 44, 541–546.

Castelletti, F. and Consonni, G. (2021). Bayesian inference of causal effects from observational data in Gaussian graphical models. *Biometrics*, 77, 136–149.

Castelletti, F. and Consonni, G. (2023). Bayesian graphical modeling for heterogeneous causal effects. *Statistics in Medicine*, 42, 15–32.

Castelletti, F., La Rocca, L., Peluso, S., Stingo, F. C. and Consonni, G. (2020). Bayesian learning of multiple directed networks from observational data. *Statistics in Medicine*, 39, 4745–4766.

Castelletti, F. and Peluso, S. (2021). Equivalence class selection of categorical graphical models. *Computational Statistics and Data Analysis*, 164, 107304.

Castelo, R. and Perlman, M. D. (2004). Learning essential graph Markov models from data. In: *Advances in Bayesian Networks, Volume 146 of Studies in Fuzziness and Soft Computing* (ed. Gámez, J. A., Moral, S. and Salmerón, A.), 255–269, Berlin: Springer.

Consonni, G. and Massam, H. (2012). Parametrizations and reference priors for multinomial decomposable graphical models. *Journal of Multivariate Analysis*, 105, 380–396.

Geiger, D. and Heckerman, D. (1997). A characterization of the Dirichlet distribution through global and local parameter independence. *The Annals of Statistics*, 25, 1344–1369.

Geiger, D. and Heckerman, D. (2002). Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. *The Annals of Statistics*, 30, 1412–1440.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20, 197–243.

Henckel, L., Perković, E. and Maathuis, M. H. (2022). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society, Series B*, 84, 579–599.

Kalisch, M. and Bühlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8, 613–636.

Kalisch, M., Fellinghauer, B. A., Grill, E., Maathuis, M. H., Mansmann, U., Bühlmann, P. et al. (2010). Understanding human functioning using graphical models. *BMC Medical Research Methodology*, 10, 1–10.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques.* Cambridge, Massachusetts, USA: The MIT Press.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford, UK: Oxford University Press.

Maathuis, M. H., Kalisch, M. and Bühlmann, P. (2009). Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37, 3133–3164.

Madigan, D., Andersson, S. A., Perlman, M. D. and Volinsky, C. T. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Communication in Statistics—Theory Methods*, 25, 2493–2519.

Mahdi Mahmoudi, S. and Wit, E. C. (2018). Estimating causal effects from nonparanormal observational data. *International Journal of Biostatistics*, 14, 20180030.

Massam, H. and Wesołołowski, J. (2016). A new prior for discrete DAG models with a restricted set of directions. *Annals of Statistics*, 44, 1010–1037.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82, 669–688.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Peters, J. and Bühlmann, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika*, 101, 219–228.

Roverato, A. (2017). *Graphical Models for Categorical Data. SemStat Elements*. Cambridge, UK: Cambridge University Press.

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice-Hall.

Sadeghi, K. (2017). Faithfulness of probability distributions and graphs. *Journal Machine Learning Research*, 18, 1–29.

Scutari, M. and Denis, J.-B. (2014). *Bayesian Networks: With Examples in R*. New York, USA: Chapman and Hall/CRC.

Shimizu, S., Hoyer, P. O., Hyvärinen, A. and Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7, 2003–2030.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction and Search*. 2nd edn. 1–16, Cambridge, MA: The MIT Press.

Stucki, G., Cieza, A. and Melvin, J. (2007). The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy. *Journal of Rehabilitation Medicine*, 39, 279–285.

Received: August 13, 2023; Revised: June 13, 2024; Accepted: July 4, 2024