

# Efficient Decomposition of Bayesian Networks With Non-graded Variables

Alessandro Magrini

Correspondence: Department of Statistics, Computer Science, Applications – University of Florence, Italy. E-mail: alessandro.magrini@unifi.it

Received: December 9, 2020 Accepted: January 15, 2021 Online Published: February 2, 2021

doi:10.5539/ijsp.v10n2p52 URL: <https://doi.org/10.5539/ijsp.v10n2p52>

## Abstract

Elicitation, estimation and exact inference in Bayesian Networks (BNs) are often difficult because the dimension of each Conditional Probability Table (CPT) grows exponentially with the increase in the number of parent variables. The Noisy-MAX decomposition has been proposed to break down a large CPT into several smaller CPTs exploiting the assumption of causal independence, i.e., absence of causal interaction among parent variables. In this way, the number of conditional probabilities to be elicited or estimated and the computational burden of the joint tree algorithm for exact inference are reduced. Unfortunately, the Noisy-MAX decomposition is suited to graded variables only, i.e., ordinal variables with the lowest state as reference, but real-world applications of BNs may also involve a number of non-graded variables, like the ones with reference state in the middle of the sample space (double-graded variables) and with two or more unordered non-reference states (multi-valued nominal variables). In this paper, we propose the *causal independence decomposition*, which includes the Noisy-MAX and two generalizations suited to double-graded and multi-valued nominal variables. While the general definition of BN implicitly assumes the presence of all the possible causal interactions, our proposal is based on causal independence, and causal interaction is a feature that can be added upon need. The impact of our proposal is investigated on a published BN for the diagnosis of acute cardiopulmonary diseases.

**Keywords:** causal independence, conditional probability table, exact inference, nominal variables, noisy gates, parent divorcing

## 1. Introduction

Bayesian Networks (BNs, Pearl, 1988) provide a formal framework to represent uncertain knowledge and to reason under uncertainty. A BN consists of a directed acyclic graph (DAG), encoding a factorization of the joint probability distribution over a set of random variables with finite sample space, and a Conditional Probability Table (CPT) for each variable, containing its probability distribution for each combination of the values of its parents in the DAG.

A CPT is defined by a number of free parameters which is exponential in the number of parent variables, thus elicitation may require relevant effort for domain experts (Druzdzel and van der Gaag, 2000), and estimation from collected data may be highly inefficient. Also, the DAG of a BN where variables have many parents involves few but very large CPTs, so that the joint tree algorithm for exact inference (Lauritzen and Spiegelhalter, 1988) may not be scalable. A widely adopted solution to this problem is the use of decompositions to break down each large CPT into several smaller CPTs. The most popular among such decompositions is the Noisy-OR, pioneered by Good (1961) and further studied by Pearl (1988). In the Noisy-OR, dichotomous parent variables are assumed to influence the value of a dichotomous response through independent latent causes. Each latent cause is ‘activated’ by a specific parent variable with a certain probability, and a single ‘active’ latent cause is sufficient for the response to change its value from the reference state to the non-reference one. The independence among latent causes implies the absence of interaction among parent variables, an assumption called *causal independence* (Heckerman and Breese, 1996). In Henrion (1987), the Noisy-MAX decomposition was introduced as a generalization of the Noisy-OR to *graded* variables, i.e., ordinal variables with the lowest state as reference. In the Noisy-MAX, latent causes have the same sample space of the response and each one may ‘activate’ one non-reference state of the response, which in turn takes value on the highest among the ‘active’ states. Further elaborations of the Noisy-MAX decomposition were provided by Díez (1993) and Srinivas (1993).

The Noisy-MAX decomposition simplifies elicitation from domain experts and estimation from collected data because, due to the assumption of causal independence, the number of free parameters is linear in the number of parent variables, instead of exponential. Also, parent divorcing can be recursively applied to latent causes until each node in the DAG has no more than two parents. In this way, the number of CPTs is increased, but they have a small size, thus the joint tree algo-

rithm for exact inference is faster and, moreover, it may become scalable also for high dimensional BNs. Unfortunately, real-world applications of BNs may also involve a number of non-graded variables, like ordinal variables with reference state in the middle of the sample space (*double-graded* variables) and variables with two or more unordered non-reference states (*multi-valued nominal* variables).

In this paper, we propose the *causal independence decomposition* for BNs, which includes the Noisy-MAX and two generalizations suited to double-graded and multi-valued nominal variables. The software implementing our proposal is contained in an R (R Core Team, 2020) package available on Github at <https://github.com/alessandromagrini/cibn>. The package can be installed from the R console by typing `install_github("alessandromagrini/cibn")` after loading the `devtools` package.

This paper is structured as follows. Section 2 includes the definition of BN and the notation used in the paper. In Section 3, an overview of the Noisy-MAX decomposition is provided. In Section 4, the causal independence decomposition and its properties are detailed, together with the extension to causal interactions. In Section 5, the impact of our proposal is investigated on a published BN for the diagnosis of acute cardiopulmonary diseases. Section 6 includes concluding remarks.

## 2. Definitions and Notation

In this section, the definition of Bayesian network is provided following Pearl (1988), together with the notation used in the paper.

**Bayesian network.** A Bayesian Network (BN) consists of the following three elements:

1. a set of variables  $V$  with finite sample space;
2. a Directed Acyclic Graph (DAG)  $\mathcal{G}$  defined on  $V$ ;
3. a set of Conditional Probability Tables (CPTs), one for each variable in  $V$ , containing the probability distribution of the variable for each combination of the values of its parents in  $\mathcal{G}$ . □

For each CPT, we denote the response variable as  $Y$  (sample space  $\Omega_Y$ ) and its parent variables as  $X_1, \dots, X_n$  (sample spaces  $\Omega_{X_1}, \dots, \Omega_{X_n}$ ). The states of a variable are labelled by consecutive integer numbers reflecting their order (if one holds), with value 0 assigned to the reference state. If non-reference states are unordered, they are labelled starting from value 1. The cardinality of any set  $S$  is denoted by  $|S|$ . The number of non-reference states of the response variable is denoted as  $n_Y \equiv |\Omega_Y| - 1$ .

Variables can be either random, i.e., defined by a non-degenerate probability distribution, or deterministic, i.e., defined by a deterministic function. Each combination of the values of a variable's parents is called *parent configuration*. A generic realization of a random variable is written in lower case, e.g.,  $v$  denotes a realization of random variable  $V$ . A probability distribution is indicated with the symbol  $p$ , with the random variable to which it refers within brackets, e.g.,  $p(V)$ . The elements of a probability distribution are indicated within angle brackets, e.g.,  $\langle \pi_0, \pi_1, \dots \rangle$ . An unordered set is indicated within curly brackets, e.g.,  $\{X_1, X_2, \dots\}$ . An ordered set is indicated within round brackets, e.g.,  $(0, 1, \dots)$ .

In a DAG, each node is labelled by the name of the variable it refers to, circles represent random variables, double circles represent deterministic variables, and squares indicate variables the response is conditioned on (i.e., parent variables). When possible, the representation through plates (Buntine, 1994) is used: a rectangle contains the nodes to be replicated as many times as shown by the index in the rectangle.

In this paper, we focus on the number of free parameters defining the CPTs in a BN, which determines the efficiency of elicitation and estimation, and on the size (number of cells) of each CPT, which determines the efficiency of exact inference. Proposition 1 states that both these two features increase exponentially with the number of parent variables.

**Proposition 1.** A CPT has size equal to  $(n_Y + 1) \prod_{i=1}^n |\Omega_{X_i}|$  and is defined by a number of free parameters equal to  $n_Y \prod_{i=1}^n |\Omega_{X_i}|$ .

*Proof.* The total number of parent configurations is equal to the product between the cardinalities of each parent's sample space:  $\prod_{i=1}^n |\Omega_{X_i}|$ . Thus, the size of the CPT is equal to the cardinality of the sample space of  $Y$  multiplied by the total number of parent configurations:  $(n_Y + 1) \prod_{i=1}^n |\Omega_{X_i}|$ . Also, since the number of free parameters of each conditional probability distribution of  $Y$  is equal to the number of its non-reference states  $n_Y$ , the total number of free parameters is equal to  $n_Y \prod_{i=1}^n |\Omega_{X_i}|$ . □

### 3. The Noisy-MAX Decomposition

A graded variable is an ordinal variable with the lowest state as reference. According to the notation introduced in Section 2, the sample space of a graded response variable is  $\Omega_Y = (0, 1, \dots, n_Y)$ . The Noisy-MAX decomposition for a graded response variable is defined below following Heckerman and Breese (1996).

**Noisy-MAX decomposition.** *The Noisy-MAX decomposition for a graded response variable  $Y$  with parents  $X_1, \dots, X_n$  consists of the following steps:*

1. latent cause  $\Lambda_0$  is defined with sample space equal to  $\Omega_Y$  and such that:

$$p(\Lambda_0) = \langle \pi_{0,0}, \pi_{0,1}, \dots, \pi_{0,n_Y} \rangle \tag{1}$$

2. for  $i = 1, \dots, n$ , latent cause  $\Lambda_i$  is defined with sample space equal to  $\Omega_Y$ , with  $X_i$  as parent, and such that:

$$\begin{aligned} p(\Lambda_i | X_i = j) &= \langle \pi_{i,j,0}, \pi_{i,j,1}, \dots, \pi_{i,j,n_Y} \rangle & j \in \{\Omega_{X_i} \setminus 0\} & i = 1, \dots, n \\ p(\Lambda_i | X_i = 0) &= \langle 1, 0, \dots, 0 \rangle & i = 1, \dots, n \end{aligned} \tag{2}$$

3. latent causes  $\Lambda_0, \dots, \Lambda_n$  determine the value of  $Y$  through the MAX function. □

The Noisy-MAX decomposition hypothesizes the existence of one latent cause for each parent (i.e.,  $\Lambda_1, \dots, \Lambda_n$ ) plus one for unmodeled causes (i.e.,  $\Lambda_0$ ), and assumes that each latent cause may ‘activate’ one non-reference state of the response variable  $Y$ , which in turn takes value on the highest among the ‘active’ states. Note that the constraint  $p(\Lambda_i | X_i = 0) = \langle 1, 0, \dots, 0 \rangle \forall i = 1, \dots, n$  means that a parent taking value on its reference state cannot cause the response to take value on a non-reference state, a feature called *amechanistic property* in Heckerman and Breese (1996).

The graphical representation of the Noisy-MAX decomposition is shown in Figure 1. Proposition 2 states that the Noisy-MAX decomposition is defined by a number of free parameters which is linear in the number of parent variables, instead of exponential.

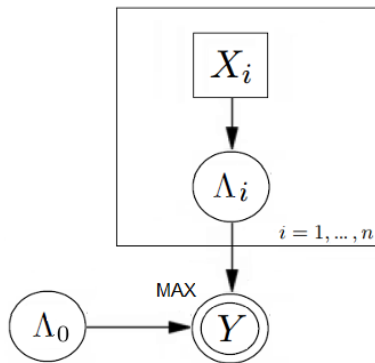


Figure 1. Graphical representation of the Noisy-MAX decomposition

**Proposition 2.** *The Noisy-MAX decomposition is defined by a number of free parameters equal to  $n_Y (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1))$ .*

*Proof.* The only random variables in the Noisy-MAX decomposition are latent causes  $\Lambda_0, \dots, \Lambda_n$ , each with  $n_Y$  non-reference states. Since  $\Lambda_0$  has no parents, its probability distribution is defined by  $n_Y$  free parameters. For  $i = 1, \dots, n$ , latent cause  $\Lambda_i$  has only  $X_i$  as parent, which has  $|\Omega_{X_i}| - 1$  non-reference states, thus the probability distribution of  $\Lambda_i$  is defined by  $n_Y (|\Omega_{X_i}| - 1)$  free parameters. Thus, the total number of free parameters is equal to  $n_Y (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1))$ . □

The linearity of the number of free parameters in the number of parent variables is a good property when CPTs are elicited from domain experts, as they may focus on the influence on the response of one parent at a time, and/or estimated from collected data, as estimates have higher efficiency. Note that, in the Noisy-MAX decomposition, probabilities refer to the states of latent causes rather than of observable variables. For instance, parameter  $\pi_{i,j,l}$  can be elicited from a domain expert by asking a question like: ‘What is the probability that the event represented by variable  $X_i$  taking value  $j$  causes the event represented by variable  $Y$  taking value  $l$ ?’.

A further property of the Noisy-MAX decomposition is that parent divorcing can be recursively applied to latent causes  $\Lambda_0, \dots, \Lambda_n$  in order to obtain an arbitrary (but no less than two) maximum number of parents for each node in the DAG. Auxiliary nodes introduced by parent divorcing are determined by the MAX function. We refer to the Noisy-MAX decomposition where parent divorcing is applied to obtain a maximum of two parents per node as *maximal* Noisy-MAX decomposition (Figure 2). The maximal Noisy-MAX decomposition introduces  $2n$  new nodes in the DAG as stated by Proposition 3, thus replacing the original CPT with several new CPTs of smaller size as stated by Proposition 4. As a consequence, the joint tree algorithm for exact inference is faster and, moreover, it may become scalable also for high dimensional BNs.

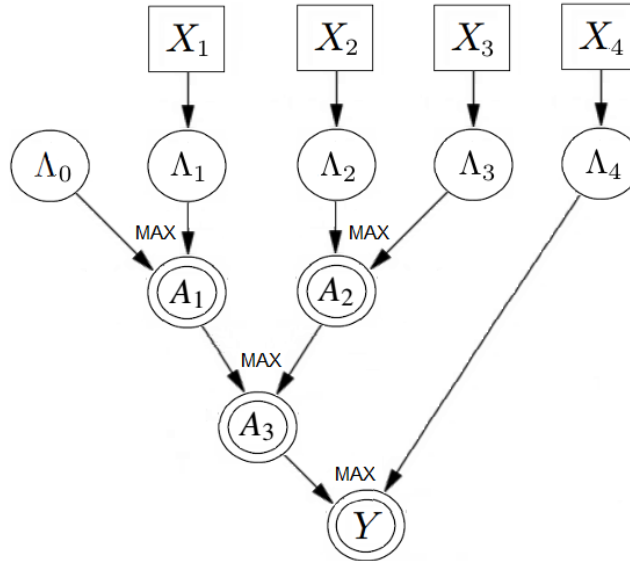


Figure 2. Maximal Noisy-MAX decomposition for a response variable with four parents. Auxiliary nodes are denoted by letter A

**Proposition 3.** *The number of new nodes introduced in the DAG by the maximal Noisy-MAX decomposition is equal to  $2n$ .*

*Proof.* By definition, the Noisy-MAX decomposition introduces  $n + 1$  new nodes (i.e., latent causes  $\Lambda_0, \dots, \Lambda_n$ ) in the DAG, and parent divorcing achieving the maximal decomposition requires  $n - 1$  additional nodes. Thus, the total number of new nodes introduced in the DAG by the maximal Noisy-MAX decomposition is equal to  $(n + 1) + (n - 1) = 2n$ .  $\square$

**Proposition 4.** *The maximal Noisy-MAX decomposition replaces the original CPT of size  $(n_Y + 1) \prod_{i=1}^n |\Omega_{X_i}|$  with one CPT of size  $n_Y + 1$ ,  $n$  CPTs of size  $(n_Y + 1)|\Omega_{X_i}|$  ( $i = 1, \dots, n$ ) and  $n$  CPTs of size  $(n_Y + 1)^3$ .*

*Proof.* The maximal Noisy-MAX decomposition introduces  $2n$  new nodes in the DAG (Proposition 3), thus implying  $2n + 1$  CPTs: one for each of the  $n + 1$  latent causes, one for each of the  $n - 1$  auxiliary nodes, and one for  $Y$ . Latent cause  $\Lambda_0$  has  $n_Y + 1$  states and no parents, thus its CPT has size  $n_Y + 1$ . For  $i = 1, \dots, n$ , latent cause  $\Lambda_i$  has  $n_Y + 1$  states and  $X_i$  as parent, thus its CPT has size  $(n_Y + 1)|\Omega_{X_i}|$ . Auxiliary nodes and  $Y$  have  $n_Y + 1$  states and, as parents, two nodes among latent causes or auxiliary nodes, thus their respective CPT has size  $(n_Y + 1)^3$ .  $\square$

Since each latent cause is influenced at most by a single parent variable, the Noisy-MAX decomposition implicitly assumes the absence of causal interaction among parents, i.e., that they independently influence the response variable. Such property is called *causal independence* in Heckerman and Breese (1996).

#### 4. The Causal Independence Decomposition

The Noisy-MAX decomposition is suited to graded response variables only, but real-world applications of BNs may also involve a number of non-graded variables, like double-graded and multi-valued nominal ones. In this section, we propose the Causal Independence Decomposition (CID), which includes the Noisy-MAX and two generalizations suited to double-graded and multi-valued nominal variables. These two decompositions are detailed in Subsections 4.1 and 4.2. Subsection 4.3 provides the main properties of the CID, while the extension to causal interactions is addressed in Subsection 4.4.

4.1 Causal Independence Decomposition for a Double-graded Response Variable

A double-graded variable is an ordinal variable with reference state in the middle of the sample space. According to the notation introduced in Section 2, the sample space of a double-graded response variable is:

$$\Omega_Y = \left(-\frac{n_Y}{2}, \dots, -1, 0, 1, \dots, \frac{n_Y}{2}\right) \tag{3}$$

The definition of the CID for a double-graded response variable is provided below.

**CID for a double-graded response variable.** *The CID for a double-graded response variable  $Y$  with parents  $X_1, \dots, X_n$  consists of the following steps:*

1. two latent causes  $\Lambda_0^{(L)}$  and  $\Lambda_0^{(R)}$  are defined with sample space:

$$\Omega_{\Lambda_0^{(L)}} = \left(-\frac{n_Y}{2}, \dots, -1, 0\right) \quad \Omega_{\Lambda_0^{(R)}} = \left(0, 1, \dots, \frac{n_Y}{2}\right) \tag{4}$$

and such that:

$$\begin{aligned} p(\Lambda_0^{(L)}) &\propto \langle \pi_{0,-\frac{n_Y}{2}}, \dots, \pi_{0,-1}, \pi_{0,0} \rangle \\ p(\Lambda_0^{(R)}) &\propto \langle \pi_{0,0}, \pi_{0,1}, \dots, \pi_{0,\frac{n_Y}{2}} \rangle \end{aligned} \tag{5}$$

2. for  $i = 1, \dots, n$ , two latent causes  $\Lambda_i^{(L)}$  and  $\Lambda_i^{(R)}$  are defined with sample space:

$$\Omega_{\Lambda_i^{(L)}} = \left(-\frac{n_Y}{2}, \dots, -1, 0\right) \quad \Omega_{\Lambda_i^{(R)}} = \left(0, 1, \dots, \frac{n_Y}{2}\right) \quad i = 1, \dots, n \tag{6}$$

with  $X_i$  as parent, and such that:

$$\begin{aligned} p(\Lambda_i^{(L)} | X_i = j) &\propto \langle \pi_{i,j,-\frac{n_Y}{2}}, \dots, \pi_{i,j,-1}, \pi_{i,j,0} \rangle \quad j \in \{\Omega_{X_i} \setminus 0\} \quad i = 1, \dots, n \\ p(\Lambda_i^{(R)} | X_i = j) &\propto \langle \pi_{i,j,0}, \pi_{i,j,1}, \dots, \pi_{i,j,\frac{n_Y}{2}} \rangle \quad j \in \{\Omega_{X_i} \setminus 0\} \quad i = 1, \dots, n \\ p(\Lambda_i^{(L)} | X_i = 0) &= \langle 0, \dots, 0, 1 \rangle \quad i = 1, \dots, n \\ p(\Lambda_i^{(R)} | X_i = 0) &= \langle 1, 0, \dots, 0 \rangle \quad i = 1, \dots, n \end{aligned} \tag{7}$$

3. variable  $\xi^{(L)}$  is defined with sample space  $\left(-\frac{n_Y}{2}, \dots, -1, 0\right)$  and such that latent causes  $\Lambda_0^{(L)}, \dots, \Lambda_n^{(L)}$  determine its value through the MIN function;
4. variable  $\xi^{(R)}$  is defined with sample space  $\left(0, 1, \dots, \frac{n_Y}{2}\right)$  and such that latent causes  $\Lambda_0^{(R)}, \dots, \Lambda_n^{(R)}$  determine its value through the MAX function;
5. variables  $\xi^{(L)}$  and  $\xi^{(R)}$  determine the value of  $Y$  through the SUM function. □

The CID for a double-graded response variable hypothesizes the existence of two sets of latent causes: one of type  $L$  and another one of type  $R$ . Latent causes of type  $L$  may ‘activate’ one non-reference state in the left side of  $\Omega_Y$ , and the lowest ‘active’ state is stored into variable  $\xi^{(L)}$ , while latent causes of type  $R$  may ‘activate’ one non-reference state in the right side of  $\Omega_Y$ , and the highest ‘active’ state is stored into variable  $\xi^{(R)}$ . Finally, the value of  $Y$  is determined as a balance between the ‘active’ states in the left and right sides of  $\Omega_Y$  by summing  $\xi^{(L)}$  and  $\xi^{(R)}$ .

The graphical representation of the CID for a double-graded response variable is shown in Figure 3. The maximal CID is obtained by applying parent divorcing to latent causes (nodes denoted by letter  $\Lambda$ ) until  $\xi^{(L)}$  and  $\xi^{(R)}$  have a maximum of two parents. Auxiliary nodes introduced by parent divorcing as intermediary between latent causes of type  $L$  and  $\xi^{(L)}$  are determined by the MIN function, while those between latent causes of type  $R$  and  $\xi^{(R)}$  are determined by the MAX function. Figure 4 displays the maximal CID for a double-graded response variable with three parents.

4.2 Causal Independence Decomposition for a Multi-valued Nominal Response Variable

A multi-valued nominal variable is a variable with two or more unordered non-reference states. According to the notation introduced in Section 1, the sample space of a multi-valued nominal response variable is  $\Omega_Y = (0, 1, \dots, n_Y)$ , where labels of non-reference states  $1, \dots, n_Y$  do not reflect any real order. Note that a dichotomous variable with unordered states can be considered as graded, provided that one of the two states can be chosen as reference. The definition of the CID for a multi-valued nominal response variable is provided below.

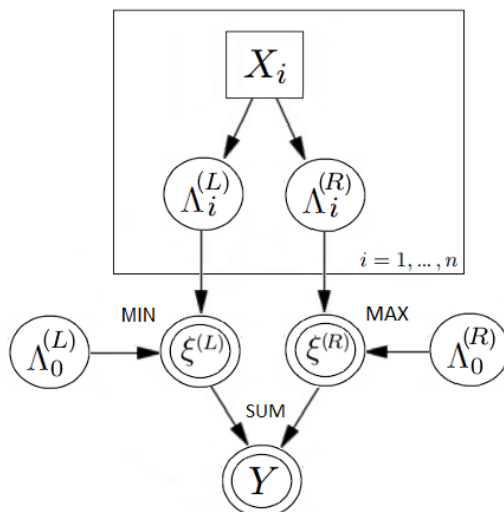


Figure 3. Graphical representation of the CID for a double-graded response variable

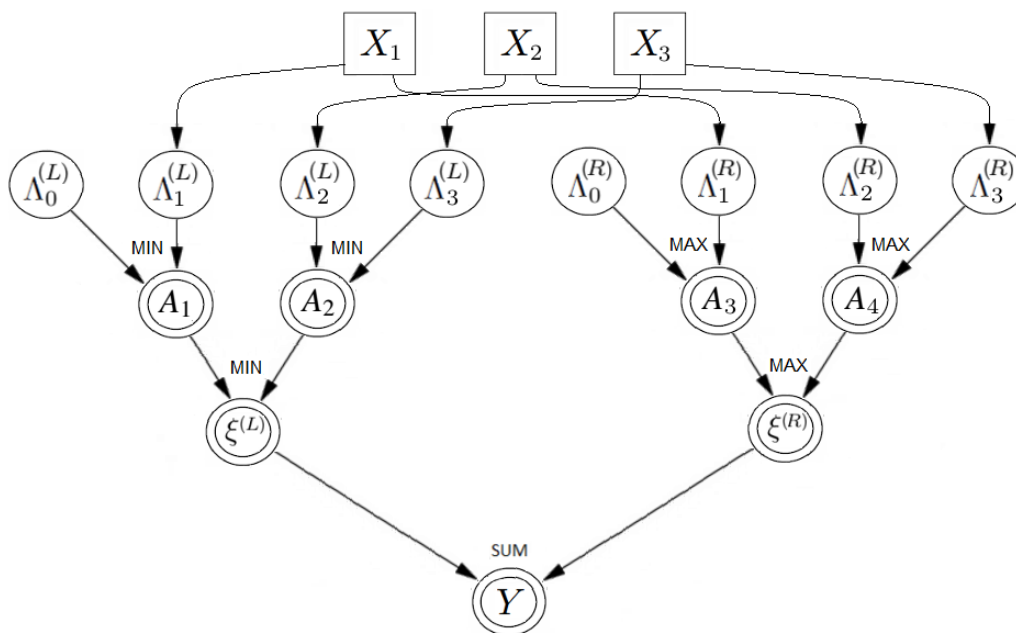


Figure 4. Maximal CID for a double-graded response variable with three parents. Auxiliary nodes are denoted by letter A

**CID for a multi-valued nominal response variable.** The CID for a multi-valued nominal response variable  $Y$  with parents  $X_1, \dots, X_n$  consists of the following steps:

- for  $l = 1, \dots, n_Y$ , latent cause  $\Lambda_0^{(l)}$  is defined with sample space  $(0, 1)$  and such that:

$$p(\Lambda_0^{(l)}) = \langle 1 - \pi_{0,l}, \pi_{0,l} \rangle \quad l = 1, \dots, n_Y \tag{8}$$

- for  $i = 1, \dots, n$  and for  $l = 1, \dots, n_Y$ , latent cause  $\Lambda_i^{(l)}$  is defined with sample space  $(0, 1)$ , with  $X_i$  as parent, and such that:

$$p(\Lambda_i^{(l)} | X_i = j) = \langle 1 - \pi_{i,j,l}, \pi_{i,j,l} \rangle \quad j \in \{\Omega_{X_i} \setminus 0\} \quad i = 1, \dots, n \quad l = 1, \dots, n_Y \tag{9}$$

$$p(\Lambda_i^{(l)} | X_i = 0) = \langle 1, 0 \rangle \quad i = 1, \dots, n \quad l = 1, \dots, n_Y$$

- for  $l = 1, \dots, n_Y$ , variable  $\xi^{(l)}$  is defined with sample space  $(0, 1)$  and such that latent causes  $\Lambda_0^{(l)}, \dots, \Lambda_n^{(l)}$  determine its value through the MAX function;

4. variables  $\xi^{(1)}, \dots, \xi^{(n_Y)}$  determine the value of  $Y$  through the following function:

$$g : \begin{cases} Y = l & \text{if } \exists l > 0 : \xi^{(l)} = 1 \wedge \xi^{(m)} = 0 \forall m \neq l \\ Y = 0 & \text{otherwise} \end{cases} \tag{10}$$

□

The CID for a multi-valued nominal response variable assumes the existence of one set of dichotomous latent causes for each non-reference state of the response. Each set of latent causes is merged through the MAX function into variable  $\xi^{(l)}$  ( $l = 1, \dots, n_Y$ ) to determine the ‘active’ non-reference states, and the response takes value on a non-reference state if and only if there is a single ‘active’ non-reference state.

The graphical representation of the CID for a multi-valued nominal response variable is shown in Figure 5. The maximal CID is obtained by applying parent divorcing to latent causes (nodes denoted by letter  $\Lambda$ ) until  $\xi^{(1)}, \dots, \xi^{(n_Y)}$  have a maximum of two parents. Figures 6 and 7 display the maximal CID for a multi-valued nominal response variable with three parents in the case, respectively, of two and three non-reference states. Note that the maximal CID applied to a multi-valued nominal response variable reduces the number of parents to two for each node, excepting for node  $Y$  which maintains  $n_Y$  parents, although each of them has only two states (see Figure 7).

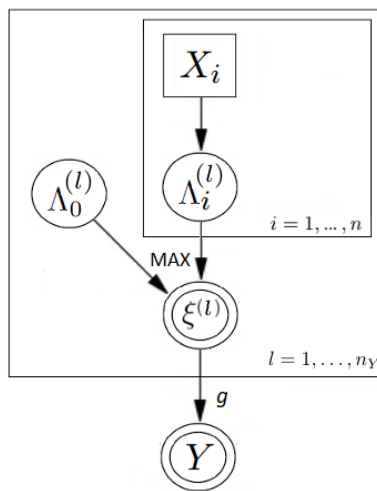


Figure 5. Graphical representation of the CID for a multi-valued nominal response variable

#### 4.3 Properties of the Causal Independence Decomposition

The CID is defined by the same number of free parameters as the Noisy-MAX decomposition whichever the type of response variable, as stated by Proposition 5.

**Proposition 5.** *The CID is defined by a number of free parameters equal to  $n_Y (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1))$ , whichever the type of response variable.*

*Proof.* The only random variables in the CID are latent causes (nodes denoted by letter  $\Lambda$ ), whichever the type of response variable. The CID for a graded response variable equates to the Noisy-MAX decomposition, which, according to Proposition 2, is defined by a number of free parameters equal to  $n_Y (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1))$ . In the case of a double-graded response variable, latent causes have  $\frac{n_Y}{2}$  non-reference states and include:  $\Lambda_0^{(L)}$  and  $\Lambda_0^{(R)}$  with no parents;  $\Lambda_i^{(L)}$  and  $\Lambda_i^{(R)}$ , each with  $X_i$  as parent ( $i = 1, \dots, n$ ). Thus, the number of free parameters is equal to  $2 \cdot \frac{n_Y}{2} \cdot (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1)) = n_Y (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1))$ . In the case of a multi-valued nominal response variable, latent causes have one non-reference state and include:  $\Lambda_0^{(1)}, \dots, \Lambda_0^{(n_Y)}$  with no parent variables;  $\Lambda_i^{(1)}, \dots, \Lambda_i^{(n_Y)}$ , each with  $X_i$  as parent ( $i = 1, \dots, n$ ). Thus, the number of free parameters is equal to  $n_Y (1 + \sum_{i=1}^n (|\Omega_{X_i}| - 1))$ . □

Analogously to the maximal Noisy-MAX decomposition, the maximal CID introduces a number of new nodes in the DAG as stated by Proposition 6, thus replacing the original CPT with several new CPTs of smaller size as stated by Proposition

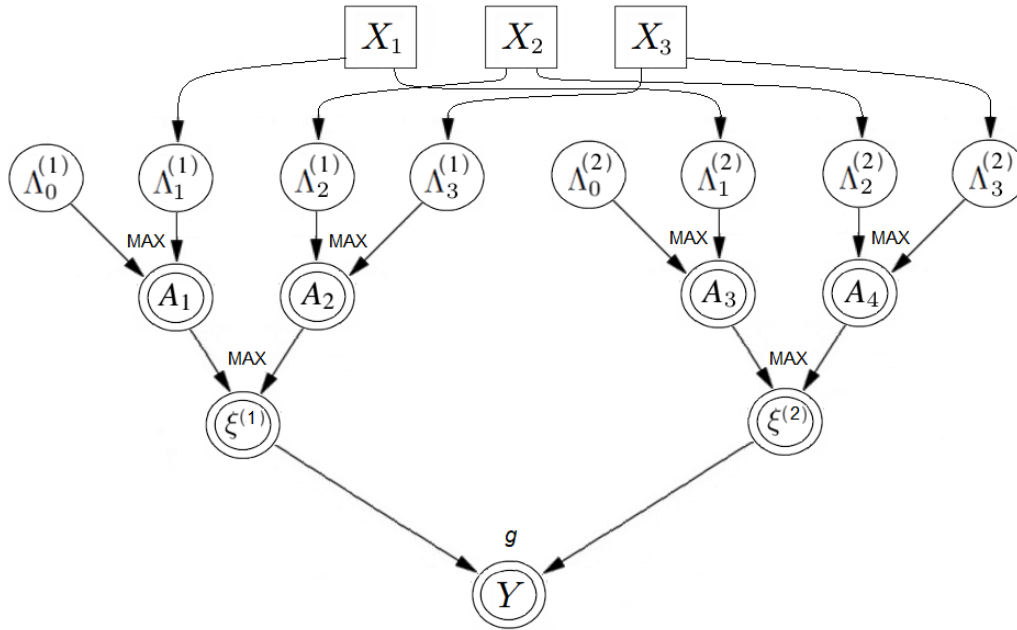


Figure 6. Maximal CID for a multi-valued nominal response variable with two non-reference states and three parents. Auxiliary nodes are denoted by letter A

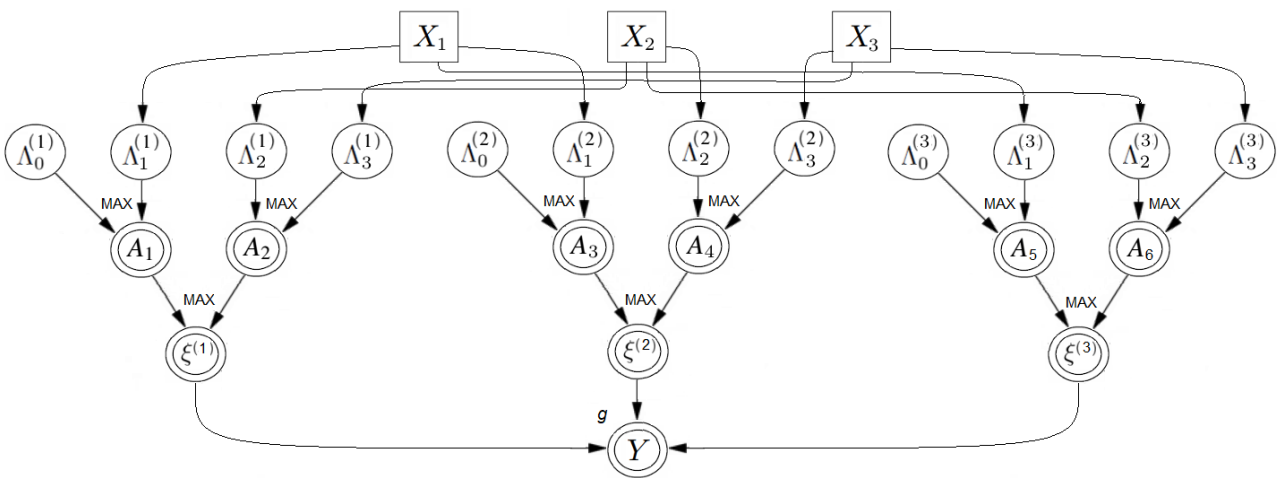


Figure 7. Maximal CID for a multi-valued nominal response variable with three non-reference states and three parents. Auxiliary nodes are denoted by letter A

7. As a consequence, the joint tree algorithm for exact inference is generally faster and, moreover, it may become scalable also for high dimensional BNs. The properties of the maximal CID are summarized in Table 1.

**Proposition 6.** *The number of new nodes introduced in the DAG by the maximal CID is equal to  $2n$  for a graded response, to  $2(2n + 1)$  for a double-graded response, and to  $n_Y(2n + 1)$  for a multi-valued nominal response.*

*Proof.* The CID for a graded response variable equates to the Noisy-MAX decomposition, which, according to Proposition 3, introduces  $2n$  new nodes. By definition, the CID for a double-graded response variable introduces two sets of new nodes  $\Lambda_i^{(L)}$  and  $\Lambda_i^{(R)}$  ( $i = 0, \dots, n$ ), each requiring  $n - 1$  nodes to achieve the maximal CID, plus two further additional nodes  $\xi^{(L)}$  and  $\xi^{(R)}$ , for a total of  $2[(n + 1) + (n - 1)] + 2 = 2(2n + 1)$  new nodes. By definition, the CID for a multi-valued nominal response variable introduces  $n_Y$  sets of new nodes  $\Lambda_i^{(l)}$  ( $i = 0, \dots, n; l = 1, \dots, n_Y$ ), each requiring  $n - 1$  nodes to achieve the maximal CID, plus  $n_Y$  further additional nodes  $\xi^{(l)}$  ( $l = 1, \dots, n_Y$ ), for a total of  $n_Y[(n + 1) + (n - 1)] + n_Y = n_Y(2n + 1)$  new nodes.  $\square$



Table 1. Properties of the maximal CID

Type of response	# free parameters	# new nodes	# CPTs	CPT size
Graded	$n_Y \left(1 + \sum_{i=1}^n ( \Omega_{X_i}  - 1)\right)$	$2n$	1	$n_Y + 1$
			$n$	$(n_Y + 1) \Omega_{X_i}  \quad i = 1, \dots, n$
			$n$	$(n_Y + 1)^3$
Double-graded	$n_Y \left(1 + \sum_{i=1}^n ( \Omega_{X_i}  - 1)\right)$	$2(2n + 1)$	2	$\frac{n_Y}{2} + 1$
			$2n$	$\left(\frac{n_Y}{2} + 1\right) \Omega_{X_i}  \quad i = 1, \dots, n$
			$2n$	$\left(\frac{n_Y}{2} + 1\right)^3$
			1	$(n_Y + 1)\left(\frac{n_Y}{2} + 1\right)^2$
Multi-valued nominal	$n_Y \left(1 + \sum_{i=1}^n ( \Omega_{X_i}  - 1)\right)$	$n_Y(2n + 1)$	$n_Y$	2
			$n_Y \cdot n$	$2 \Omega_{X_i}  \quad i = 1, \dots, n$
			$n_Y \cdot n$	$2^3$
			1	$(n_Y + 1)2^{n_Y}$

**Proposition 7.** *The maximal CID replaces the original CPT of size  $(n_Y + 1) \prod_{i=1}^n |\Omega_{X_i}|$  with several new CPTs as follows:*

- *graded response: one CPT of size  $n_Y + 1$ ,  $n$  CPTs of size  $(n_Y + 1)|\Omega_{X_i}|$  ( $i = 1, \dots, n$ ) and  $n$  CPTs of size  $(n_Y + 1)^3$ ;*
- *double-graded response: 2 CPTs of size  $\frac{n_Y}{2} + 1$ ,  $2n$  CPTs of size  $\left(\frac{n_Y}{2} + 1\right)|\Omega_{X_i}|$  ( $i = 1, \dots, n$ ),  $2n$  CPTs of size  $\left(\frac{n_Y}{2} + 1\right)^3$  and one CPT of size  $(n_Y + 1)\left(\frac{n_Y}{2} + 1\right)^2$ ;*
- *multi-valued nominal response:  $n_Y$  CPTs of size 2,  $n_Y \cdot n$  CPTs of size  $2|\Omega_{X_i}|$  ( $i = 1, \dots, n$ ),  $n_Y \cdot n$  CPTs of size  $2^3$  and one CPT of size  $(n_Y + 1)2^{n_Y}$ .*

*Proof (graded response).* The maximal CID for a graded variable equates to the maximal Noisy-MAX decomposition, which, according to Proposition 4, implies one CPT of size  $n_Y + 1$ ,  $n$  CPTs of size  $(n_Y + 1)|\Omega_{X_i}|$  ( $i = 1, \dots, n$ ) and  $n$  CPTs of size  $(n_Y + 1)^3$ . □

*Proof (double-graded response).* The maximal CID for a double-graded variable introduces  $2(2n + 1)$  new nodes (Proposition 6), thus implying  $4n + 3$  CPTs: one for each of the  $n + 1$  latent causes  $\Lambda_i^{(L)}$  ( $i = 0, \dots, n$ ), one for each of the  $n + 1$  latent causes  $\Lambda_i^{(R)}$  ( $i = 0, \dots, n$ ), one for each of the  $2(n - 1)$  auxiliary nodes to achieve the maximal decomposition, one for  $\xi^{(L)}$ , one for  $\xi^{(R)}$  and one for  $Y$ . Since  $\Lambda_0^{(L)}$  and  $\Lambda_0^{(R)}$  have  $\frac{n_Y}{2} + 1$  states and no parents, their respective CPT has size  $\frac{n_Y}{2} + 1$ . For  $i = 1, \dots, n$ , latent causes  $\Lambda_i^{(L)}$  and  $\Lambda_i^{(R)}$  have  $\frac{n_Y}{2} + 1$  states and  $X_i$  as parent, thus their respective CPT has size  $\left(\frac{n_Y}{2} + 1\right)|\Omega_{X_i}|$ . Auxiliary nodes to achieve the maximal CID, as well as  $\xi^{(L)}$  and  $\xi^{(R)}$  have  $\frac{n_Y}{2} + 1$  states and, as parents, two nodes among latent causes and auxiliary nodes, thus their respective CPT has size  $\left(\frac{n_Y}{2} + 1\right)^3$ . Finally,  $Y$  has  $\xi^{(L)}$  and  $\xi^{(R)}$  as parents, thus its CPT has size  $(n_Y + 1)\left(\frac{n_Y}{2} + 1\right)^2$ . □

*Proof (multi-valued nominal response).* The maximal CID for a multi-valued nominal variable introduces  $n_Y(2n + 1)$  new nodes (Proposition 6), thus implying  $n_Y(2n + 1) + 1$  CPTs: one for each of the  $n_Y(n + 1)$  latent causes  $\Lambda_i^{(l)}$  ( $i = 0, \dots, n; l = 1, \dots, n_Y$ ), one for each of the  $n_Y(n - 1)$  auxiliary nodes to achieve the maximal CID, one for each of the  $n_Y$  variables  $\xi^{(l)}$  ( $l = 1, \dots, n_Y$ ) and one for  $Y$ . For  $l = 1, \dots, n_Y$ , latent cause  $\Lambda_0^{(l)}$  has 2 states and no parents, thus its CPT has size 2. For  $i = 1, \dots, n$  and for  $l = 1, \dots, n_Y$ , latent cause  $\Lambda_i^{(l)}$  has 2 states and  $X_i$  as parent, thus its CPT has size  $2|\Omega_{X_i}|$ . Auxiliary nodes to achieve the maximal CID, as well as variables  $\xi^{(l)}$  ( $l = 1, \dots, n_Y$ ) have 2 states and, as parents, two nodes among latent causes and auxiliary nodes, thus their respective CPT has size  $2^3$ . Finally,  $Y$  has  $\xi^{(1)}, \dots, \xi^{(n_Y)}$  as parents, thus its CPT has size  $(n_Y + 1)2^{n_Y}$ . □

#### 4.4 Extension to Causal Interactions

Suppose that, for a response variable  $Y$ , causal interaction holds among the variables in a subset  $X_S$  of the parents  $X_1, \dots, X_n$ . The following two-step technique allows to apply the CID in this case.

1. A new multi-valued nominal variable  $Z$  is created with variables in  $X_S$  as parents, sample space equal to the cartesian product of the sample spaces of variables in  $X_S$ , and such that  $Z$  takes value on a particular combination of states  $x_S$  if and only if each variable in  $X_S$  takes value on the respective state in  $x_S$ . The neutral state of  $Z$  is the one combining the neutral states of variables in  $X_S$ .
2. The edges connecting variables in  $X_S$  to  $Y$  are deleted and an edge from  $Z$  to  $Y$  is added.

After this technique is implemented, causal independence holds among the new parents of  $Y$ , thus the CID can be applied. As an example, suppose that  $X_S = \{X_1, X_2\}$  and  $\Omega_{X_1} = \Omega_{X_2} = (0, 1)$ . In this case, we create the multi-valued nominal variable  $Z$  as follows:

$$Z = \begin{cases} (0, 0) & \text{if } X_1 = 0 \wedge X_2 = 0 \\ (0, 1) & \text{if } X_1 = 0 \wedge X_2 = 1 \\ (1, 0) & \text{if } X_1 = 1 \wedge X_2 = 0 \\ (1, 1) & \text{if } X_1 = 1 \wedge X_2 = 1 \end{cases} \quad (11)$$

where the reference state is  $(0, 0)$ , because it combines the reference states of  $X_1$  and  $X_2$ . Figure 8 illustrates the technique for a graded response variable, but it is identically implemented for a double-graded or a multi-valued nominal one.

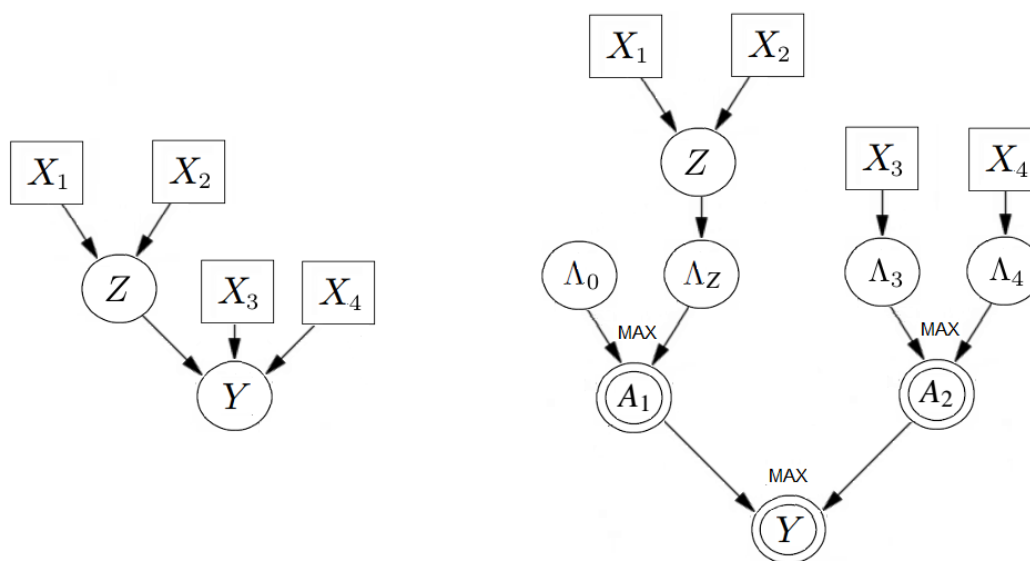


Figure 8. Illustration of the technique allowing to apply the CID in presence of causal interactions. Here,  $Y$  is a graded variable with parents  $X_1, X_2, X_3$  and  $X_4$ , with a causal interaction holding between  $X_1$  and  $X_2$ . Node  $Z$  is introduced as intermediary between the interacting parents and  $Y$  (left panel). In this way, the new set of parents  $\{Z, X_3, X_4\}$  satisfies causal independence and the maximal CID can be applied (right panel)

### 5. Practical Application

We investigate the impact of the maximal CID on the BN for the diagnosis of acute cardiopulmonary diseases developed by Magrini et al. (2018). The BN contains 278 variables, which are distinguished by the authors into dichotomous, polytomous and continuous. The reference state and an eventual order on the states is established for each dichotomous and polytomous variable, while a reference range is defined for continuous variables, which can be at the left side of the sample space (restricted continuous variable), or in its middle (non-restricted continuous variable).

We considered dichotomous, ordinal polytomous and restricted continuous variables as graded (249 variables), non-restricted continuous variables as double-graded (12 variables), and non-ordinal polytomous variables as multi-valued nominal (17 variables). Most variables have a single non-reference state (207 variables), all the double-graded variables have four non-reference states, and, among multi-valued nominal variables, most have two or three non-reference states

(14 variables out of 17). Also, there are 10 variables with at least two sets of interacting parents, for a total of 36 sets, all with cardinality equal to two. The BN is characterized by a high structural complexity: the mean size of the parent sets (inner degree) is 2.1 with a maximum of 10, the mean size of the child set (outer degree) is 2.1 with a maximum of 31, and the mean size of Markov blankets (i.e., number of parents, children and parents of the children) is 8.5 with a maximum of 50. The distribution of the main structural characteristics of the BN is summarized in Table 2.

Table 2. Distribution of the main structural characteristics of the BN developed by Magrini et al. (2018)

	Type of response	Min.	1st quartile	Median	Mean	3rd quartile	Max.
# non-reference states	All	1	1	1	1.4	2	5
	Graded	1	1	1	1.2	1	5
	Double-graded	4	4	4	4.0	4	4
	Multi-valued nominal	2	2	2	3.0	3	5
# parents	All	0	1	2	2.1	3	10
	Graded	0	1	1	2.0	3	10
	Double-graded	1	1	2	2.5	3	7
	Multi-valued nominal	1	2	4	4.1	6	9
# children	All	0	0	1	2.1	3	31
	Graded	0	0	1	2.0	3	31
	Double-graded	0	0	1.5	2.1	3	7
	Multi-valued nominal	0	0	3	3.1	3	16
Markov blanket size	All	1	1	5.5	8.5	13	50
	Graded	1	1	5	8.2	12	48
	Double-graded	1	2	5	7.3	14	17
	Multi-valued nominal	1	5	13	13.8	18	50

An illustration of the maximal CID applied to three variables in the BN is provided in the Appendix, and the decomposed BN is available as an R object at <https://github.com/alessandromagrini/cibn>. Table 3 summarizes the impact of the maximal CID on the BN with regard to the number of new nodes, the number of free parameters and the size of CPTs. We see that, at the cost of increasing by 5.7 times the number of nodes (from 278 to 1574) and by 6.6 times the number of CPTs (from 278 to 1828), the number of free parameters is reduced by 23.5 times (from 47560 to 2023), while the size of the CPTs is decreased by 24.1 times in mean (from 234 to 9.7) and by 108 times in maximum (from 23328 to 216).

### 6. Concluding Remarks

We have proposed an extension of the Noisy-MAX decomposition to non-graded variables, called Causal Independence Decomposition (CID). Our proposal maintains the two desirable properties of the Noisy-MAX: linearity of the number of free parameters with respect to the number of parent variables and significant reduction of the size of CPTs. The first property is important for elicitation from domain experts and estimation from collected data, as the addition of a new parent variable entails an increase in the number of parameters which is proportional to the number of that parent’s non-reference states. The second property is important for exact inference, because speed and scalability of the joint tree algorithm depend inversely on the size of CPTs.

The CID is maximally efficient if nominal variables have no more than two non-reference states and all the sets of interacting parents have cardinality equal to two, situation where the DAG can be decomposed until each node has no more than two parents. In the Bayesian network for the diagnosis of acute cardiopulmonary diseases developed by Magrini et al. (2018), the number of nominal variables with more than two non-reference states is small compared to the dimension of the model, and all the sets of interacting parents have cardinality equal to two. The CID has a very significant impact in this case, with the number of free parameters and the mean size of CPTs reduced by 23.5 and 24.1 times, respectively.

Oppositely to the general definition of Bayesian network implicitly assuming the presence of all the possible causal interactions, the CID is based on causal independence, and causal interaction is a feature that can be added upon need. Evidence in favour of a good fit of the Noisy-MAX decomposition obtained by Zagorecki and Druzdzal (2013) on several published Bayesian networks gives validity to the assumptions on which the CID relies. Moreover, the characteristics of the high dimensional Bayesian network developed by Magrini et al. (2018) fit very well to these assumptions: nominal variables with more than two reference states and variables with interacting parents are the minor part, and most sets of interacting parents have cardinality equal to two.

Table 3. Impact of the maximal CID on the BN developed by Magrini et al. (2018)

		Before maximal CID						
	Type of response	Total	Min.	1st quartile	Median	Mean	3rd quartile	Max.
# free parameters	All	47560	1	3	6	171.1	24	19440
	Graded	11899	1	2	4	47.8	12	3840
	Double-graded	4464	20	20	78	372.0	135	3456
	Multi-valued nominal	31197	4	18	48	1835.1	648	19440
CPT size	All	# CPTs: 278	2	6	10	234.0	36	23328
	Graded	# CPTs: 249	4	8	8	80.4	24	5760
	Double-graded	# CPTs: 12	25	25	97.5	465.0	169	4320
	Multi-valued nominal	# CPTs: 17	6	27	72	2322.2	864	23328
		After maximal CID						
	Type of response	Total	Min.	1st quartile	Median	Mean	3rd quartile	Max.
# new nodes	All	1574	1	2	4	5.7	6	85
	Graded	1000	1	2	2	4.0	6	20
	Double-graded	144	6	6	10	12.0	14	30
	Multi-valued nominal	430	6	10	18	25.3	34	85
# free parameters	All	2023	1	3	4	7.3	8	80
	Graded	1316	1	2	3	5.3	6	36
	Double-graded	356	20	20	30	29.7	36	52
	Multi-valued nominal	351	4	10	15	20.7	24	80
CPT size	All	# CPTs: 1828	2	4	8	9.7	8	216
	Graded	# CPTs: 1225	2	4	6	9.5	8	216
	Double-graded	# CPTs: 156	3	9	15	18.4	27	45
	Multi-valued nominal	# CPTs: 447	2	4	6	7.2	8	192

**Acknowledgments**

The authors acknowledge the partial financial support provided by the “Dipartimenti Eccellenti 2018-2022” ministerial funds, Italy.

**References**

Buntine, W. L. (1994). Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 2, 159–225. <https://doi.org/10.1613/jair.62>

Díez, F. J. (1993). Parameter adjustment in Bayes networks: The generalized Noisy-OR gate. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 99–105, Washington, US-DC. Morgan Kaufmann. <https://doi.org/10.1016/B978-1-4832-1451-1.50016-0>

Druzdzel, M. J., & van der Gaag, L. C. (2000). Building probabilistic networks: Where do the numbers come from? *IEEE Transactions on Knowledge and Data Engineering*, 12(4), 481–486. <https://doi.org/10.1109/TKDE.2000.868901>.

Good, I. (1961). A causal calculus (I). *British Journal of Philosophy of Science*, 11, 305–318. <https://doi.org/10.1093/bjps/XI.44.305>

Heckerman, D., & Breese, J. S. (1996). Causal independence for probability assessment and inference using Bayesian networks. *Systems, Man and Cybernetics*, 26(6), 826–831. <https://doi.org/10.1109/3468.541341>

Henrion, M. (1987). Some practical issues in constructing belief networks. In *Proceedings of the 3rd Conference on Uncertainty in Artificial Intelligence*, pages 161–173, Seattle, US-WA. Morgan Kaufmann.

Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B (Methodological)*, 50(2), 157–224. <https://doi.org/10.1023/A:1008935617754>

Magrini, A., Luciani, D., & Stefanini, F. M. (2018). A probabilistic network for the diagnosis of acute cardiopulmonary diseases. *Biometrical Journal*, 60(1), 174–195. <https://doi.org/10.1002/bimj.201600206>

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, San Francisco, US-CA. <https://doi.org/10.1016/B978-0-08-051489-5.50008-4>

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, AT.

Srinivas, S. (1993). A generalization of the Noisy-Or model. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 208–215, Washington, US-DC. Morgan Kaufmann.

Zagorecki, A., & Druzdzal, M. J. (2013). Knowledge engineering for Bayesian networks: How common are Noisy-MAX distributions in practice? *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(1), 186–195. <https://doi.org/10.1109/TSMCA.2012.2189880>

**Appendix**

In this section, we provide an illustration of the maximal CID applied to three variables in the BN developed by Magrini et al. (2018): one graded variable (Subsection A1), one double-graded variable (Subsection A2) and one multi-valued nominal variable (Subsection A3).

*A1. Illustration on a graded variable*

*Peripheral Edema* is a graded variable with sample space: (0 = ‘absent’, 1 = ‘unilateral’, 2 = ‘bilateral’), labelled as *Edem\_g* and here denoted as *Y*. It has four parent variables:

- *Lower Limbs Fractures*, a graded variable with sample space: (0 = ‘absent’, 1 = ‘present’), labelled as *Fr* and here denoted as *X<sub>1</sub>*;
- *Right Heart Failure*, a graded variable with sample space: (0 = ‘absent’, 1 = ‘moderate’, 2 = ‘severe’), labelled as *RHF* and here denoted as *X<sub>2</sub>*;
- *Lower Limbs Deep Vein Thrombosis*, a graded variable with sample space: (0 = ‘absent’, 1 = ‘present’), labelled as *TVPi* and here denoted as *X<sub>3</sub>*;
- *Chronic Cardiac Muscle Disease*, a graded variable with sample space: (0 = ‘absent’, 1 = ‘initial’, 2 = ‘advanced’), labelled as *Cardiomio* and here denoted as *X<sub>4</sub>*.

The total number of parent configurations is  $\prod_{i=1}^4 |\Omega_{X_i}| = 2 \cdot 3 \cdot 2 \cdot 3 = 36$ , thus the CPT of *Y* without assuming causal independence consists of 36 conditional probability distributions, each of size  $|\Omega_Y| = n_Y + 1 = 3$  and defined by  $|\Omega_Y| - 1 = n_Y = 2$  free parameters, leading to a CPT size of  $3 \cdot 36 = 108$  and a total number of free parameters equal to  $2 \cdot 36 = 72$ . After applying the maximal CID, which in this case coincides with the Noisy-MAX decomposition, the DAG is augmented as shown in Figure 2, with the following free parameters:

<i>X<sub>i</sub> = j</i>	<i>Y</i>	
	1	2
Unmodeled causes	$\pi_{0,1}$	$\pi_{0,2}$
<i>X<sub>1</sub> = 1</i>	$\pi_{1,1,1}$	$\pi_{1,1,2}$
<i>X<sub>1</sub> = 2</i>	$\pi_{1,2,1}$	$\pi_{1,2,2}$
<i>X<sub>2</sub> = 1</i>	$\pi_{2,1,1}$	$\pi_{2,1,2}$
<i>X<sub>3</sub> = 1</i>	$\pi_{3,1,1}$	$\pi_{3,1,2}$
<i>X<sub>3</sub> = 2</i>	$\pi_{3,2,1}$	$\pi_{3,2,2}$
<i>X<sub>4</sub> = 1</i>	$\pi_{4,1,1}$	$\pi_{4,1,2}$

and the following CPTs:

	$\Lambda_0$		
	0	1	2
	$(1 - \pi_{0,1} - \pi_{0,2})$	$\pi_{0,1}$	$\pi_{0,2}$

<i>X<sub>1</sub></i>	$\Lambda_1$		
	0	1	2
0	1	0	0
1	$(1 - \pi_{1,1,1} - \pi_{1,1,2})$	$\pi_{1,1,1}$	$\pi_{1,1,2}$
2	$(1 - \pi_{1,2,1} - \pi_{1,2,2})$	$\pi_{1,2,1}$	$\pi_{1,2,2}$

<i>X<sub>2</sub></i>	$\Lambda_2$		
	0	1	2
0	1	0	0
1	$(1 - \pi_{2,1,1} - \pi_{2,1,2})$	$\pi_{2,1,1}$	$\pi_{2,1,2}$

$X_3$	$\Lambda_3$			$X_4$	$\Lambda_4$		
	0	1	2		0	1	2
	0	1	0		0	0	0
1	$(1 - \pi_{3,1,1} - \pi_{3,1,2})$	$\pi_{3,1,1}$	$\pi_{3,1,2}$	1	$(1 - \pi_{4,1,1} - \pi_{4,1,2})$	$\pi_{4,1,1}$	$\pi_{4,1,2}$
2	$(1 - \pi_{3,2,1} - \pi_{3,2,2})$	$\pi_{3,2,1}$	$\pi_{3,2,2}$				

$\Lambda_0$	$\Lambda_1$	$A_1$			$\Lambda_2$	$\Lambda_3$	$A_2$			$A_1$	$A_2$	$A_3$			$A_3$	$\Lambda_4$	$Y$		
		0	1	2			0	1	2			0	1	2			0	1	2
0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0
0	2	0	0	1	0	2	0	0	1	0	2	0	0	1	0	2	0	0	1
1	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0
1	1	0	1	0	1	1	0	1	0	1	1	0	1	0	1	1	0	1	0
1	2	0	0	1	1	2	0	0	1	1	2	0	0	1	1	2	0	0	1
2	0	0	0	1	2	0	0	0	1	2	0	0	0	1	2	0	0	0	1
2	1	0	0	1	2	1	0	0	1	2	1	0	0	1	2	1	0	0	1
2	2	0	0	1	2	2	0	0	1	2	2	0	0	1	2	2	0	0	1

A2. Illustration on a double-graded variable

Arterial Blood Gas  $paCO_2$  is a double-graded variable with sample space: ( $-2 =$  ‘very low’,  $-1 =$  ‘low’,  $0 =$  ‘normal’,  $1 =$  ‘high’,  $2 =$  ‘very high’), labelled as CO2 and here denoted as  $Y$ . It has three parent variables:

- Minute ventilation, a double-graded variable with sample space: ( $-2 =$  ‘very low’,  $-1 =$  ‘low’,  $0 =$  ‘normal’,  $1 =$  ‘high’,  $2 =$  ‘very high’), labelled as minVent and here denoted as  $X_1$ ;
- Lung perfusion, a graded variable with sample space: ( $0 =$  ‘normal’,  $1 =$  ‘moderate’,  $2 =$  ‘severe’), labelled as pulmPerf and here denoted as  $X_2$ .
- Pulmonary shunt, a graded variable with sample space: ( $0 =$  ‘normal’,  $1 =$  ‘moderate’,  $2 =$  ‘severe’), labelled as Shu and here denoted as  $X_3$ .

The total number of parent configurations is  $\prod_{i=1}^3 |\Omega_{X_i}| = 5 \cdot 3 \cdot 3 = 45$ , thus the CPT of  $Y$  without assuming causal independence consists of 45 conditional probability distributions, each of size  $|\Omega_Y| = n_Y + 1 = 5$  and defined by  $|\Omega_Y| - 1 = n_Y = 4$  free parameters, leading to a CPT size of  $5 \cdot 45 = 225$  and a total number of free parameters equal to  $4 \cdot 45 = 180$ . After applying the maximal CID, the DAG is augmented as shown in Figure 4, with the following free parameters:

$X_i = j$	$Y$			
	-2	-1	1	2
Unmodeled causes	$\pi_{0,-2}$	$\pi_{0,-1}$	$\pi_{0,1}$	$\pi_{0,2}$
$X_1 = -2$	$\pi_{1,-2,-2}$	$\pi_{1,-2,-1}$	$\pi_{1,-2,1}$	$\pi_{1,-2,2}$
$X_1 = -1$	$\pi_{1,-1,-2}$	$\pi_{1,-1,-1}$	$\pi_{1,-1,1}$	$\pi_{1,-1,2}$
$X_1 = 1$	$\pi_{1,1,-2}$	$\pi_{1,1,-1}$	$\pi_{1,1,1}$	$\pi_{1,1,2}$
$X_1 = 2$	$\pi_{1,2,-2}$	$\pi_{1,2,-1}$	$\pi_{1,2,1}$	$\pi_{1,2,2}$
$X_2 = 1$	$\pi_{2,1,-2}$	$\pi_{2,1,-1}$	$\pi_{2,1,1}$	$\pi_{2,1,2}$
$X_2 = 2$	$\pi_{2,2,-2}$	$\pi_{2,2,-1}$	$\pi_{2,2,1}$	$\pi_{2,2,2}$
$X_3 = 1$	$\pi_{3,1,-2}$	$\pi_{3,1,-1}$	$\pi_{3,1,1}$	$\pi_{3,1,2}$
$X_3 = 2$	$\pi_{3,2,-2}$	$\pi_{3,2,-1}$	$\pi_{3,2,1}$	$\pi_{3,2,2}$

Note:  $\pi_{0,0}$  and  $\pi_{i,j,0}$  ( $i = 1, 2, 3; j \in \Omega_{X_i}$ ) are obtained by subtraction from 1.

and the following CPTs (the ones of variables denoted by letter  $\Lambda$  require normalization to sum 1 by row):

$\Lambda_0^{(L)}$			$\Lambda_0^{(R)}$		
-2	-1	0	0	1	2
$\pi_{0,-2}$	$\pi_{0,-1}$	$\pi_{0,0}$	$\pi_{0,0}$	$\pi_{0,1}$	$\pi_{0,2}$

$\Lambda_1^{(L)}$				$\Lambda_1^{(R)}$			
$X_1$	-2	-1	0	$X_1$	0	1	2
-2	$\pi_{1,-2,-2}$	$\pi_{1,-2,-1}$	$\pi_{1,-2,0}$	-2	$\pi_{1,-2,0}$	$\pi_{1,-2,1}$	$\pi_{1,-2,2}$
-1	$\pi_{1,-1,-2}$	$\pi_{1,-1,-1}$	$\pi_{1,-1,0}$	-1	$\pi_{1,-1,0}$	$\pi_{1,-1,1}$	$\pi_{1,-1,2}$
0	0	0	1	0	1	0	0
1	$\pi_{1,1,-2}$	$\pi_{1,1,-1}$	$\pi_{1,1,0}$	1	$\pi_{1,1,0}$	$\pi_{1,1,1}$	$\pi_{1,1,2}$
2	$\pi_{1,2,-2}$	$\pi_{1,2,-1}$	$\pi_{1,2,0}$	2	$\pi_{1,2,0}$	$\pi_{1,2,1}$	$\pi_{1,2,2}$

$\Lambda_2^{(L)}$				$\Lambda_2^{(R)}$			
$X_2$	-2	-1	0	$X_2$	0	1	2
0	0	0	1	0	1	0	0
1	$\pi_{2,1,-2}$	$\pi_{2,1,-1}$	$\pi_{2,1,0}$	1	$\pi_{2,1,0}$	$\pi_{2,1,1}$	$\pi_{2,1,2}$
2	$\pi_{2,2,-2}$	$\pi_{2,2,-1}$	$\pi_{2,2,0}$	2	$\pi_{2,2,0}$	$\pi_{2,2,1}$	$\pi_{2,2,2}$

$\Lambda_3^{(L)}$				$\Lambda_3^{(R)}$			
$X_3$	-2	-1	0	$X_3$	0	1	2
0	0	0	1	0	1	0	0
1	$\pi_{3,1,-2}$	$\pi_{3,1,-1}$	$\pi_{3,1,0}$	1	$\pi_{3,1,0}$	$\pi_{3,1,1}$	$\pi_{3,1,2}$
2	$\pi_{3,2,-2}$	$\pi_{3,2,-1}$	$\pi_{3,2,0}$	2	$\pi_{3,2,0}$	$\pi_{3,2,1}$	$\pi_{3,2,2}$

$\Lambda_0^{(L)}$					$\Lambda_2^{(L)}$					$\Lambda_0^{(R)}$					$\Lambda_2^{(R)}$				
$\Lambda_1^{(L)}$	-2	-1	0	$A_1$	$\Lambda_3^{(L)}$	-2	-1	0	$A_2$	$\Lambda_1^{(R)}$	0	1	2	$A_3$	$\Lambda_3^{(R)}$	0	1	2	$A_4$
-2	-2	1	0	0	-2	-2	1	0	0	0	0	1	0	0	0	0	1	0	0
-2	-1	1	0	0	-2	-1	1	0	0	0	1	0	1	0	0	1	0	1	0
-2	0	1	0	0	-2	0	1	0	0	0	2	0	0	1	0	2	0	0	1
-1	-2	1	0	0	-1	-2	1	0	0	1	0	0	1	0	1	0	1	0	0
-1	-1	0	1	0	-1	-1	0	1	0	1	1	0	1	0	1	1	0	1	0
-1	0	0	1	0	-1	0	0	1	0	1	2	0	0	1	1	2	0	0	1
0	-2	1	0	0	0	-2	1	0	0	2	0	0	0	1	2	0	0	0	1
0	-1	0	1	0	0	-1	0	1	0	2	1	0	0	1	2	1	0	0	1
0	0	0	0	1	0	0	0	0	1	2	2	0	0	1	2	2	0	0	1

$\xi^{(L)}$					$\xi^{(R)}$					$Y$						
$A_1$	$A_2$	-2	-1	0	$A_3$	$A_4$	0	1	2	$\xi^{(L)}$	$\xi^{(R)}$	-2	-1	0	1	2
-2	-2	1	0	0	0	0	1	0	0	-2	0	1	0	0	0	0
-2	-1	1	0	0	0	1	0	1	0	-2	1	0	1	0	0	0
-2	0	1	0	0	0	2	0	0	1	-2	2	0	0	1	0	0
-1	-2	1	0	0	1	0	0	1	0	-1	0	0	1	0	0	0
-1	-1	0	1	0	1	1	0	1	0	-1	1	0	0	1	0	0
-1	0	0	1	0	1	2	0	0	1	-1	2	0	0	0	1	0
0	-2	1	0	0	2	0	0	0	1	0	0	0	0	1	0	0
0	-1	0	1	0	2	1	0	0	1	0	1	0	0	0	1	0
0	0	0	0	1	2	2	0	0	1	0	2	0	0	0	0	1

A3. Illustration on a multi-valued nominal variable

Prophylaxis or anticoagulation is a multi-valued nominal variable with sample space: (0 = ‘no’, 1 = ‘heparine’, 2 = ‘anticoagulants’), labelled as Prof and here denoted as Y. It has three parent variables:

- Chronic atrial arrhythmia, a graded variable with sample space: (0 = ‘absent’, 1 = ‘present’), labelled as Arit\_sopra\_cron and here denoted as X<sub>1</sub>;

- *Previous episode of deep venous thrombosis or pulmonary embolism*, a graded variable with sample space: (0 = ‘no’, 1 = ‘yes’), labelled as *preTVP* and here denoted as  $X_2$ ;
- *Surgery*, a multi-valued nominal variable with sample space: (0 = ‘no’, 1 = ‘general’, 2 = ‘orthopedic’), labelled as *Chir* and here denoted as  $X_3$ .

The total number of parent configurations is  $\prod_{i=1}^3 |\Omega_{X_i}| = 2 \cdot 2 \cdot 3 = 12$ , thus the CPT of  $Y$  without assuming causal independence consists of 12 conditional probability distributions, each of size  $|\Omega_Y| = n_Y + 1 = 3$  and defined by  $|\Omega_Y| - 1 = n_Y = 2$  free parameters, leading to a CPT size of  $3 \cdot 12 = 36$  and a total number of free parameters equal to  $2 \cdot 12 = 24$ . After applying the maximal CID, the DAG is augmented as shown in Figure 6, with the following free parameters:

$X_i = j$	$Y$	
	1	2
Unmodeled causes	$\pi_{0,1}$	$\pi_{0,2}$
$X_1 = 1$	$\pi_{1,1,1}$	$\pi_{1,1,2}$
$X_2 = 1$	$\pi_{2,1,1}$	$\pi_{2,1,2}$
$X_3 = 1$	$\pi_{3,1,1}$	$\pi_{3,1,2}$
$X_3 = 2$	$\pi_{3,2,1}$	$\pi_{3,2,2}$

and the following CPTs:

<table border="1"> <thead> <tr> <th colspan="2"><math>\Lambda_0^{(1)}</math></th> </tr> <tr> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td><math>(1 - \pi_{0,1})</math></td> <td><math>\pi_{0,1}</math></td> </tr> </tbody> </table>	$\Lambda_0^{(1)}$		0	1	$(1 - \pi_{0,1})$	$\pi_{0,1}$	<table border="1"> <thead> <tr> <th><math>X_1</math></th> <th colspan="2"><math>\Lambda_1^{(1)}</math></th> </tr> <tr> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td><math>(1 - \pi_{1,1,1})</math></td> <td><math>\pi_{1,1,1}</math></td> </tr> </tbody> </table>	$X_1$	$\Lambda_1^{(1)}$			0	1	0	1	0	1	$(1 - \pi_{1,1,1})$	$\pi_{1,1,1}$	<table border="1"> <thead> <tr> <th><math>X_2</math></th> <th colspan="2"><math>\Lambda_2^{(1)}</math></th> </tr> <tr> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td><math>(1 - \pi_{2,1,1})</math></td> <td><math>\pi_{2,1,1}</math></td> </tr> </tbody> </table>	$X_2$	$\Lambda_2^{(1)}$			0	1	0	1	0	1	$(1 - \pi_{2,1,1})$	$\pi_{2,1,1}$	<table border="1"> <thead> <tr> <th><math>X_3</math></th> <th colspan="2"><math>\Lambda_3^{(1)}</math></th> </tr> <tr> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td><math>(1 - \pi_{3,1,1})</math></td> <td><math>\pi_{3,1,1}</math></td> </tr> <tr> <td>2</td> <td><math>(1 - \pi_{3,2,1})</math></td> <td><math>\pi_{3,2,1}</math></td> </tr> </tbody> </table>	$X_3$	$\Lambda_3^{(1)}$			0	1	0	1	0	1	$(1 - \pi_{3,1,1})$	$\pi_{3,1,1}$	2	$(1 - \pi_{3,2,1})$	$\pi_{3,2,1}$																																																			
$\Lambda_0^{(1)}$																																																																																																			
0	1																																																																																																		
$(1 - \pi_{0,1})$	$\pi_{0,1}$																																																																																																		
$X_1$	$\Lambda_1^{(1)}$																																																																																																		
	0	1																																																																																																	
0	1	0																																																																																																	
1	$(1 - \pi_{1,1,1})$	$\pi_{1,1,1}$																																																																																																	
$X_2$	$\Lambda_2^{(1)}$																																																																																																		
	0	1																																																																																																	
0	1	0																																																																																																	
1	$(1 - \pi_{2,1,1})$	$\pi_{2,1,1}$																																																																																																	
$X_3$	$\Lambda_3^{(1)}$																																																																																																		
	0	1																																																																																																	
0	1	0																																																																																																	
1	$(1 - \pi_{3,1,1})$	$\pi_{3,1,1}$																																																																																																	
2	$(1 - \pi_{3,2,1})$	$\pi_{3,2,1}$																																																																																																	
<table border="1"> <thead> <tr> <th colspan="2"><math>\Lambda_0^{(2)}</math></th> </tr> <tr> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td><math>(1 - \pi_{0,2})</math></td> <td><math>\pi_{0,2}</math></td> </tr> </tbody> </table>	$\Lambda_0^{(2)}$		0	1	$(1 - \pi_{0,2})$	$\pi_{0,2}$	<table border="1"> <thead> <tr> <th><math>X_1</math></th> <th colspan="2"><math>\Lambda_1^{(2)}</math></th> </tr> <tr> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td><math>(1 - \pi_{1,1,2})</math></td> <td><math>\pi_{1,1,2}</math></td> </tr> </tbody> </table>	$X_1$	$\Lambda_1^{(2)}$			0	1	0	1	0	1	$(1 - \pi_{1,1,2})$	$\pi_{1,1,2}$	<table border="1"> <thead> <tr> <th><math>X_2</math></th> <th colspan="2"><math>\Lambda_2^{(2)}</math></th> </tr> <tr> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td><math>(1 - \pi_{2,1,2})</math></td> <td><math>\pi_{2,1,2}</math></td> </tr> </tbody> </table>	$X_2$	$\Lambda_2^{(2)}$			0	1	0	1	0	1	$(1 - \pi_{2,1,2})$	$\pi_{2,1,2}$	<table border="1"> <thead> <tr> <th><math>X_3</math></th> <th colspan="2"><math>\Lambda_3^{(2)}</math></th> </tr> <tr> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td><math>(1 - \pi_{3,1,2})</math></td> <td><math>\pi_{3,1,2}</math></td> </tr> <tr> <td>2</td> <td><math>(1 - \pi_{3,2,2})</math></td> <td><math>\pi_{3,2,2}</math></td> </tr> </tbody> </table>	$X_3$	$\Lambda_3^{(2)}$			0	1	0	1	0	1	$(1 - \pi_{3,1,2})$	$\pi_{3,1,2}$	2	$(1 - \pi_{3,2,2})$	$\pi_{3,2,2}$																																																			
$\Lambda_0^{(2)}$																																																																																																			
0	1																																																																																																		
$(1 - \pi_{0,2})$	$\pi_{0,2}$																																																																																																		
$X_1$	$\Lambda_1^{(2)}$																																																																																																		
	0	1																																																																																																	
0	1	0																																																																																																	
1	$(1 - \pi_{1,1,2})$	$\pi_{1,1,2}$																																																																																																	
$X_2$	$\Lambda_2^{(2)}$																																																																																																		
	0	1																																																																																																	
0	1	0																																																																																																	
1	$(1 - \pi_{2,1,2})$	$\pi_{2,1,2}$																																																																																																	
$X_3$	$\Lambda_3^{(2)}$																																																																																																		
	0	1																																																																																																	
0	1	0																																																																																																	
1	$(1 - \pi_{3,1,2})$	$\pi_{3,1,2}$																																																																																																	
2	$(1 - \pi_{3,2,2})$	$\pi_{3,2,2}$																																																																																																	
<table border="1"> <thead> <tr> <th><math>\Lambda_0^{(1)}</math></th> <th><math>\Lambda_1^{(1)}</math></th> <th colspan="2"><math>A_1</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	$\Lambda_0^{(1)}$	$\Lambda_1^{(1)}$	$A_1$				0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	<table border="1"> <thead> <tr> <th><math>\Lambda_2^{(1)}</math></th> <th><math>\Lambda_3^{(1)}</math></th> <th colspan="2"><math>A_2</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	$\Lambda_2^{(1)}$	$\Lambda_3^{(1)}$	$A_2$				0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	<table border="1"> <thead> <tr> <th><math>\Lambda_0^{(2)}</math></th> <th><math>\Lambda_1^{(2)}</math></th> <th colspan="2"><math>A_3</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	$\Lambda_0^{(2)}$	$\Lambda_1^{(2)}$	$A_3$				0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	<table border="1"> <thead> <tr> <th><math>\Lambda_2^{(2)}</math></th> <th><math>\Lambda_3^{(2)}</math></th> <th colspan="2"><math>A_4</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	$\Lambda_2^{(2)}$	$\Lambda_3^{(2)}$	$A_4$				0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1
$\Lambda_0^{(1)}$	$\Lambda_1^{(1)}$	$A_1$																																																																																																	
		0	1																																																																																																
0	0	1	0																																																																																																
0	1	0	1																																																																																																
1	0	0	1																																																																																																
1	1	0	1																																																																																																
$\Lambda_2^{(1)}$	$\Lambda_3^{(1)}$	$A_2$																																																																																																	
		0	1																																																																																																
0	0	1	0																																																																																																
0	1	0	1																																																																																																
1	0	0	1																																																																																																
1	1	0	1																																																																																																
$\Lambda_0^{(2)}$	$\Lambda_1^{(2)}$	$A_3$																																																																																																	
		0	1																																																																																																
0	0	1	0																																																																																																
0	1	0	1																																																																																																
1	0	0	1																																																																																																
1	1	0	1																																																																																																
$\Lambda_2^{(2)}$	$\Lambda_3^{(2)}$	$A_4$																																																																																																	
		0	1																																																																																																
0	0	1	0																																																																																																
0	1	0	1																																																																																																
1	0	0	1																																																																																																
1	1	0	1																																																																																																
<table border="1"> <thead> <tr> <th><math>A_1</math></th> <th><math>A_2</math></th> <th colspan="2"><math>\xi^{(1)}</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	$A_1$	$A_2$	$\xi^{(1)}$				0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	<table border="1"> <thead> <tr> <th><math>A_3</math></th> <th><math>A_4</math></th> <th colspan="2"><math>\xi^{(2)}</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	$A_3$	$A_4$	$\xi^{(2)}$				0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	<table border="1"> <thead> <tr> <th><math>\xi^{(1)}</math></th> <th><math>\xi^{(2)}</math></th> <th colspan="3"><math>Y</math></th> </tr> <tr> <td></td> <td></td> <th>0</th> <th>1</th> <th>2</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>1</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> <td>0</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td>1</td> <td>1</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	$\xi^{(1)}$	$\xi^{(2)}$	$Y$					0	1	2	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	1	1	1	0	0																			
$A_1$	$A_2$	$\xi^{(1)}$																																																																																																	
		0	1																																																																																																
0	0	1	0																																																																																																
0	1	0	1																																																																																																
1	0	0	1																																																																																																
1	1	0	1																																																																																																
$A_3$	$A_4$	$\xi^{(2)}$																																																																																																	
		0	1																																																																																																
0	0	1	0																																																																																																
0	1	0	1																																																																																																
1	0	0	1																																																																																																
1	1	0	1																																																																																																
$\xi^{(1)}$	$\xi^{(2)}$	$Y$																																																																																																	
		0	1	2																																																																																															
0	0	1	0	0																																																																																															
0	1	0	0	1																																																																																															
1	0	0	1	0																																																																																															
1	1	1	0	0																																																																																															

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).