UNIVERSITÀ DEGLI STUDI FIRENZE
DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

A Generalization of the Noisy-MAX Parameterization for Biomedical Applications

Alessandro Magrini, Davide Luciani,
Federico Mattia Stefanini

DISIA WORKING PAPER
2016/06

A Generalization of the Noisy-MAX Parameterization for Biomedical Applications

Alessandro Magrini[a], Davide Luciani[b], Federico M. Stefanini*[a,]

[a] *Department of Statistics, Computer Science, Applications, University of Florence, Italy*
[b] *Mario Negri Institute for Pharmacological Research, Milan, Italy*

Abstract

The development of Bayesian networks for biomedical applications is often difficult because the number of parameters defining each conditional probability table grows exponentially with the increase in the number of parent variables. The Noisy-MAX parameterization have been extensively used to reduce the number of parameters defining a conditional probability table when causes independently influence the response. Unfortunately, the Noisy-MAX parameterization is not suited to a non-ordinal response variable. In this paper, we propose a generalization of the Noisy-MAX parameterization, called SoftDom parameterization, which is suited to a general biomedical response variable influenced by independent causal determinants.

Keywords: Bayesian networks; Conditional probability tables; Nominal response; Partially-ordinal response.

MSC classification codes: 62F15, 62P10

1. Introduction

Bayesian networks have been successfully applied to medical problems due to the opportunity of representing causal understanding of phenomena [14]: once causal relationships among domain variables are established by an acyclic directed graph (DAG), the relevant quantitative information reduces to a set of conditional probability tables, each referring to the relationship between a variable and its parents in the DAG.

Knowledge engineering for non-trivial medical applications of Bayesian networks is challenging because, even though causal notions and conditional independence statements save from specifying the joint probability distribution

*Corresponding author
Email addresses: `magrini@disia.unifi.it` (Alessandro Magrini), `davide.luciani@marionegri.it` (Davide Luciani), `stefanini@disia.unifi.it` (Federico M. Stefanini*)

May 26, 2016

over domain variables, the dimension of conditional probability tables grows exponentially with the increase in the size of parent sets [5].

A widely adopted solution to this problem is the use of parametric models defining conditional probability distributions through a number of parameters which is linear in the number of parent variables. The most popular among such models is the Noisy-OR parameterization, pioneered by [7] and further studied by [17]. In the Noisy-OR, binary parent variables are assumed to influence the value of a binary response through independent latent causes. Each latent cause is 'activated' by a specific parent variable with a certain probability, and a single activated cause is sufficient for the response to change its value from 'absent' to 'present'. By using the Noisy-OR, the dimension of the CPT is linear in the number of parent variables, instead of exponential. In [11], the Noisy-OR parameterization was extended to multi-valued parent variables and unmodeled causes, and the Noisy-MAX parameterization was introduced. In the Noisy-MAX, the response is an ordinal multi-valued variable, each latent cause 'votes' for a state of the response, and a deterministic function selects the state among the voted ones on which the response takes value. Further elaborations of the Noisy-MAX parameterization were provided by [4] and [19]. In [9], the concept of independent causal influence (ICI) was introduced to denote the causal assumption underlying Noisy-OR and Noisy-MAX parameterizations.

Several real-world applications of Bayesian networks to medical problems are based on an extensive use of the Noisy-MAX parameterization (see, for example, [3, 16]). In this paper, we propose the SoftDom parameterization, a generalization of the Noisy-MAX parameterization to address the case in which the value of a response variable is assigned according to a probability distribution depending on the configuration of latent causes. The SoftDom parameterization is suited to a general biomedical response variable influenced by independent causal determinants.

This paper is structured as follows. Section 2 includes a definition of Bayesian network and the notation used hereinafter. In Section 3, we provide an overview of the Noisy-MAX parameterization. In Section 4, the SoftDom parameterization is detailed. Section 5 includes the discussion of our contribution.

2. Bayesian networks

A Bayesian network is a statistical model representing the joint probability distribution over a domain of interest (see, for example, [12]).

Definition 1 *(Bayesian network) A Bayesian network consists of the following:*

1. *a set of variables $\boldsymbol{V}$ with finite sample space;*
2. *a Directed Acyclic Graph (DAG) $\mathcal{G}$ on $\boldsymbol{V}$;*
3. *for each variable in $\boldsymbol{V}$, the probability distribution of the variable for each possible configuration of its parents in $\mathcal{G}$, called conditional probability table (CPT).*

For each CPT, we denote the response variable as Y (sample space Ω_Y) and the collection of its parent variables as $\mathbf{X} = \{X_1, \ldots, X_n\}$ (sample spaces $\Omega_{X_1}, \ldots, \Omega_{X_n}$). In the graphical representation of DAGs, each node will be labelled by the name of the variable it refers to. A generic realization of a variable or a set of variables is written as lower case bold letter, e.g., x or $\boldsymbol{x}$. A probability distribution is written within angle brackets, e.g., $< 0.7, 0.2, 0.1 >$. DAG in Figure 1 represents a response variable with n parents by means of one plate [1], a rectangle collecting the variables to be replicated as many times as shown by the index in it. Squares indicate variables the response is conditioned on. This representation holds for any node in the DAG and whatever the considered response variable, provided that it is not a root node.

3. The Noisy-MAX parameterization

A CPT is defined by a number of parameters equal to $(||\Omega_Y||-1)\prod_{i=1}^{n}||\Omega_{X_i}||$, namely its dimension is exponential in the number of parent variables. The Noisy-MAX parameterization [8] is a representation of a CPT through a number of parameters which is linear in the number of parent variables. In the Noisy-MAX parameterization, the response is an ordinal variable, and both the response and each parent variable admits a reference state, for instance the absence of unfavourable conditions for the system under consideration (*distinguished state* [9], or *neutral state* [2]). States of a variable are labelled by consecutive integer numbers reflecting an eventual order, in particular the reference state is labelled by value 0.

Definition 2 *(Noisy-MAX parameterization) The Noisy-MAX parameterization consists of the following:*

1. *the response Y is an ordinal variable with reference state labelled by value 0, minimal label equal to $y_L \leq 0$, and maximal label equal to $y_R \geq 0$;*
2. *parent variables $X_1, \ldots, X_n$ admit a reference state, labelled by value 0;*
3. *auxiliary variables $\Lambda_1, \ldots, \Lambda_n$ are introduced as intermediary between each parent variable and the response.*
4. *variable Λ_0 is introduced as a parent of Y;*
5. *the probability distribution of Λ_0 is represented by parameter $\boldsymbol{\pi}_0 =< \pi_{0,y_L}, \ldots, \pi_{0,0}, \ldots, \pi_{0,y_R} >$;*
6. *for $i = 1, \ldots, n$, the probability distribution of Λ_i given $X_i = j$ is represented by parameter $\boldsymbol{\pi}_{i,j} =< \pi_{i,j,y_L}, \ldots, \pi_{i,j,0}, \ldots, \pi_{i,j,y_R} >$, $\forall j \in \Omega_{X_i}$, where $\boldsymbol{\pi}_{i,0,0} = 1$ (amechanistic property, [8]);*
7. *the value of Y is assigned by the maximum function of the values taken by variables $\Lambda_0, \ldots, \Lambda_n$:*

$$Y = MAX(\boldsymbol{\lambda}) \; , \qquad \boldsymbol{\lambda} \in \Omega_Y^{n+1}$$

The graphical representation of the Noisy-MAX parameterization is shown in Figure 1. The core assumption of the Noisy-MAX is the existence of latent causes $\mathbf{\Lambda} = \{\Lambda_0, \Lambda_1, \ldots, \Lambda_n\}$ determining the value of the response Y. Each latent cause besides Λ_0 is influenced by a single parent variable; latent cause Λ_0 is jointly influenced by all unmodeled parents at one time. Each latent cause 'votes' for a state of the response, with the constraint that a latent cause cannot vote for a non-reference state if the corresponding parent variable takes value on its reference state. Finally, the response takes value on the higher-ordered state among the voted ones. We refer to this property as *strong dominance*. The special case where the response is a binary variable is widely known as Noisy-OR.

In medical applications, strong dominance may depend on the physiological mechanism underlying the generation of the datum, or on the criteria convened by the study design to collect observations, as shown by the examples below.

Example 1 *Consider a variable representing the occurrence of cough in a patient with states 'absence of cough', 'dry cough' and 'productive cough'. Dry cough is always manifested over no cough, because it represents an alteration of the healthy condition of a patient. Productive cough is always manifested over dry cough, since, in the presence of the former, the latter is prevented by the mucus in the respiratory airways.*

Example 2 *Findings of left ventricle dyssynergia are often reduced to three alternative categories, based on the isolate occurrence of dyskinesia, on the occurrence of akinesia disregarding dyskinesia, and on the occurrence of aneurysm disregarding dyskinesia and akinesia. As such, the following strict order relation holds: 'absence of dyssynergia', 'dyskinesia', 'akinesia', 'aneurism'.*

Parameters of the Noisy-MAX parameterization can be elicited from domain experts by asking questions on the probability of causal mechanisms. For instance, parameter $\pi_{i,j,l}$ can be elicited by asking a question of the type: '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?'.

The number of parameters required by the Noisy-MAX parameterization is equal to $(||\Omega_Y|| - 1) \cdot (1 + \sum_{i=1}^{n}(||\Omega_{X_i}|| - 1))$, namely it is linear in the number of parent variables. The CPT implied by the Noisy-MAX parameterization can be computed by applying the following formula [2]:

$$\Pr(Y \leq y \mid \boldsymbol{x}) = \left(\sum_{k=y_L}^{y} \pi_{0,k} \right) \cdot \prod_{i=1}^{n} \left(\sum_{k=y_L}^{y} \pi_{i,j:X_i=j,k} \right) \quad (1)$$

4. The SoftDom parameterization

In general, a response variable may be characterized by partial order, or even no order, on its states so that strong dominance is not a natural assumption. Also, at the level of detail captured by the model, uncertainty may still affect

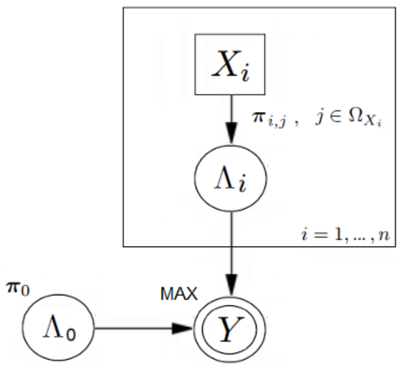

Figure 1: (a) Graphical representation of the Noisy-MAX parameterization (parameters are shown besides pertaining edges).

the response variable even if the values taken by latent causes are known. In this case, it seems natural to introduce a probability distribution on the states receiving at least one vote from latent causes, no matter whether the response is an ordinal variable or not. The SoftDom parameterization is built on this intuition.

Definition 3 *(SoftDom parameterization) The SoftDom parameterization consists of the following:*

1. *the response Y and its parents $X_1, \ldots, X_n$ admit a reference state labelled by value 0. The minimal and the maximal label of Y are respectively equal to $y_L \leq 0$ and $y_R \geq 0$;*
2. *auxiliary variables $\Lambda_1, \ldots, \Lambda_n$ are introduced as intermediary between each parent variable and the response.*
3. *variable Λ_0 is introduced as a parent of Y;*
4. *the probability distribution of Λ_0 is represented by parameter $\boldsymbol{\pi}_0 =< \pi_{0,y_L}, \ldots, \pi_{0,0}, \ldots, \pi_{0,y_R} >$;*
5. *for $i = 1, \ldots, n$, the probability distribution of Λ_i given $X_i = j$ is represented by parameter $\boldsymbol{\pi}_{i,j} =< \pi_{i,j,y_L}, \ldots, \pi_{i,j,0}, \ldots, \pi_{i,j,y_R} >$, $\forall j \in \Omega_{X_i}$, where $\boldsymbol{\pi}_{i,0,0} = 1$ (amechanistic property, [8]);*
6. *each variable in set $\boldsymbol{\Upsilon} = \{\Upsilon_{y_L}, \ldots, \Upsilon_{y_R}\}$ is a deterministic function of the values taken by variables $\Lambda_0, \ldots, \Lambda_n$:*

$$\Upsilon_k(\boldsymbol{\lambda}) = \mathbb{1}_{\Lambda_0 = k \vee \ldots \vee \Lambda_n = k} \;, \qquad \boldsymbol{\lambda} \in \Omega_Y^{n+1}, \quad k = y_L, \ldots, y_R$$

   where $\mathbb{1}$ is the indicator function.
7. *the probability distribution of Y given $\boldsymbol{\Upsilon} = \boldsymbol{\upsilon}$ is represented by parameter $\omega_{\boldsymbol{\upsilon}} =< \omega_{\boldsymbol{\upsilon}}^{(y_L)}, \ldots, \omega_{\boldsymbol{\upsilon}}^{(0)}, \ldots, \omega_{\boldsymbol{\upsilon}}^{(y_R)} >$, $\forall \boldsymbol{\upsilon} \in \{\{0,1\}^{y_R - y_L + 1} \setminus \{0, \ldots, 0\}\}$, where:*

$$\Upsilon_k = 0 \implies \omega_{\boldsymbol{\upsilon}}^{(k)} = 0, \quad \forall k = y_L, \ldots, y_R$$

The graphical representation of the SoftDom parameterization is shown in Figure 2. The soft dominance property generalizes strong dominance in the assignment of a value to the response. First, votes from latent causes are recorded by variables in $\boldsymbol{\Upsilon}$ by marking each state with value 0 if it receives no vote, or 1 if it receives at least one vote. Second, the response variable takes value according to a probability distribution on the states marked with value 1 by variables in $\boldsymbol{\Upsilon}$. We refer to these probability distributions as 'assigning distributions'.

Example 3 *Chest pain is typically classified by assuming that a patient reports only one among several type of chest pain, like chest pain increasing with breathing movements, fixed chest pain, oppressive chest pain and stabbing chest pain. If latent causes vote for only fixed chest pain and chest pain increasing with breathing movements, the latter is manifested over the former. Instead, no criteria are in use to anticipate which type of chest pain is manifested if latent causes vote for only oppressive chest pain and stabbing chest pain, thus a reasonable option is to assume an uniform assigning distribution.*

Example 4 *Consider a response variables made by a list of alternative diseases. In this case, the value of the response may depend on a probability distribution on the voted states. Such representation is often involved when the simultaneous occurrence of diseases is very rare, like in the framework of similarity networks [10]. A reasonable option is to assume uniform assigning distributions, in order to reflect the lack of knowledge on the mechanism making diseases competing each other.*

The Noisy-MAX parameterization is obtained from the SoftDom parameterization if all assigning distributions degenerate into MAX functions. The proposition below provides a formula to compute the CPT implied by the SoftDom parameterization.

Proposition 1 *The conditional probability distribution of a response variable Y given parent configuration $\boldsymbol{x}$ under the SoftDom parameterization is:*

$$\Pr(Y = y \mid \boldsymbol{x}) = \sum_{\{\boldsymbol{\upsilon}:\upsilon_y=1\}} \omega_{\boldsymbol{\upsilon}}^{(y)} \sum_{\{\boldsymbol{\lambda}:\ \{\Upsilon_{y_L}(\boldsymbol{\lambda}),\ldots,\Upsilon_0(\boldsymbol{\lambda}),\ldots,\Upsilon_{y_R}(\boldsymbol{\lambda})\}=\boldsymbol{\upsilon}\}} \pi_{0,\lambda_0} \prod_{i=1}^{n} \pi_{i,j:X_i=j,\lambda_i} \tag{2}$$

Proof. The result follows by marginalizing variables in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Upsilon}$ out from the joint probability distribution (see Figure 2):

$$\Pr(Y = y \wedge \boldsymbol{\Lambda} = \boldsymbol{\lambda} \wedge \boldsymbol{\Upsilon} = \boldsymbol{\upsilon} \mid \boldsymbol{X} = \boldsymbol{x}) =$$

$$\Pr(Y = y \mid \boldsymbol{\Upsilon} = \boldsymbol{\upsilon}) \cdot \prod_{k=y_L}^{y_R} \Pr(\Upsilon_k = \upsilon_k \mid \boldsymbol{\Lambda} = \boldsymbol{\lambda}) \cdot \Pr(\Lambda_0 = \lambda_0) \cdot \prod_{i=1}^{n} \Pr(\Lambda_i = \lambda_i \mid \boldsymbol{X} = \boldsymbol{x})$$

□

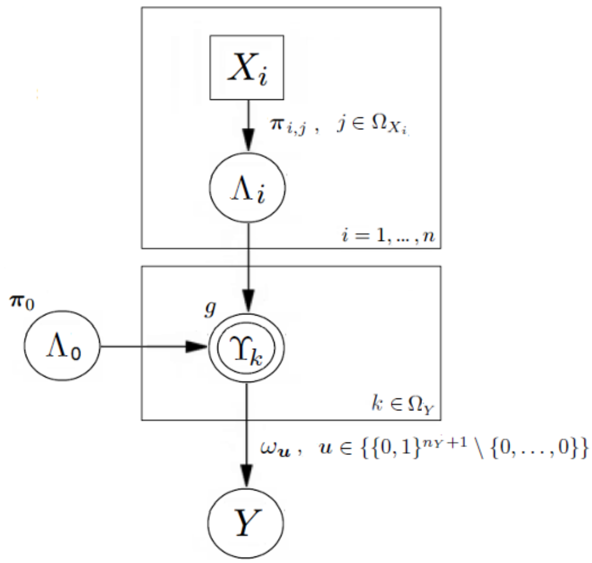

Figure 2: The SoftDom parameterization (parameters are shown besides pertaining edges).

The SoftDom parameterization is defined by the same parameters defining the Noisy-MAX parameterization, plus a number of additional parameters describing the assigning distributions. In practice, we expect that most of assigning distributions are MAX functions, uniform distributions, or probability distributions degenerated into a single state, as shown by examples above, thus requiring no significant effort to be elicited from medical experts. However, in the worst case, the number of parameters defining assigning distributions is equal to $\sum_{k=2}^{y_R-y_L+1}\binom{y_R-y_L+1}{k} - 1 = 2^{y_R-y_L+1} - y_R + y_L - 2$, namely it is exponential in the cardinality of Ω_Y (Table 1). We addressed this issue by designing a special case of the SoftDom parameterization, describing assigning distributions with a number of parameters which is linear in the cardinality of Ω_Y.

Definition 4 *(Noisy-SCORE) The Noisy-SCORE parameterization is a SoftDom parameterization such that:*

$$\omega_{\boldsymbol{\upsilon}} = \begin{cases} < \frac{\upsilon_{y_L}\phi_{y_L}}{\sum_{k=y_L}^{y_R}\upsilon_k}, \dots, \frac{\upsilon_{y_R}\phi_{y_R}}{\sum_{k=y_L}^{y_R}\upsilon_k} > & \textit{if } \sum_{k=y_L}^{y_R}\upsilon_k > 0 \\ < \frac{\mathbb{1}_{\upsilon_{y_L}=0}}{\sum_{k=y_L}^{y_R}\mathbb{1}_{\upsilon_k=0}}, \dots, \frac{\mathbb{1}_{\upsilon_{y_R}=0}}{\sum_{k=y_L}^{y_R}\mathbb{1}_{\upsilon_k=0}} > & \textit{otherwise} \end{cases}$$

$\forall \boldsymbol{\upsilon} \in \{0,1\}^{y_R-y_L+1} \setminus \{0,\dots,0\}$, *where* $\phi_k \geq 0$, $\forall k = y_L, \dots, y_R$.

The Noisy-SCORE is a SoftDom parameterization where a positive score is attached to each state of the response, and assigning distributions are such that each state receiving at least one vote from latent causes has a probability

Table 1: Number of assigning distributions to be specified for the SoftDom parameterization as a function of the number of states of the response.

Cardinality of Ω_Y	# assigning distributions to be specified
2	1
3	4
4	11
5	26
6	57
7	120
8	247
9	502
10	1013

of selection proportional to its score. A state with null score can be selected only if latent causes vote for no states with positive scores. The Noisy-SCORE cannot degenerate into the Noisy-MAX, but can approximate it at the desired level of precision, as shown by the proposition below.

Proposition 2 *The Noisy-SCORE with* $\phi_k = 10^{(\tau+1)k}a$*,* $\forall k = y_L, \ldots, y_R$*,* $a > 0$*, approximates the Noisy-MAX at the first* τ *decimals.*

Proof. The constraints $\phi_k = 10^{(\tau+1)k}a$, $a > 0$, forces any state following another to have a score higher by $\tau + 1$ orders of magnitude. Thus $\omega_{\boldsymbol{\upsilon}}^{(m)} = 1$, with $m = \text{MAX}\{k : \upsilon_k = 1\}$, at the first τ decimals, $\forall \boldsymbol{\upsilon} \in \{0,1\}^{y_R - y_L + 1} \setminus \{0, \ldots, 0\}$.

□

Examples below illustrate the application of the Noisy-SCORE parameterization to real-world medical response variables.

Example 5 *The Noisy-SCORE parameterization with null score on the reference state and uniform scores on non-reference states is equivalent to the SoftDom parameterization proposed in Example 4.*

Example 6 *Consider a response variable qualifying the metabolic status of a patient as 'normal', 'hypoglicemic', 'ketoacidotic' and 'hyperosmolar'. Soft dominance holds for such variable, because state 'normal' is lower-ordered than the other states, but these cannot be ranked each other. The SoftDom parameterization would require up to* 33 *parameters to describe* 11 *assigning distributions, instead, the Noisy-SCORE parameterization would require only* 4 *parameters. For instance, with* $\boldsymbol{\phi} = \{0.7,\ 7,\ 37.7,\ 54.6\}$ *the following assigning distributions are obtained:*

$$\omega_{\{1,1,0,0\}} =< 0.1, 0.9, 0, 0 >, \quad \omega_{\{1,0,1,0\}} =< 0.02, 0, 0.98, 0 >$$

$$\omega_{\{1,0,0,1\}} =< 0.01, 0, 0, 0.99 >, \quad \omega_{\{0,1,1,0\}} =< 0, 0.16, 0.84, 0 >$$
$$\omega_{\{0,1,0,1\}} =< 0, 0.11, 0, 0.89 >, \quad \omega_{\{0,0,1,1\}} =< 0, 0, 0.41, 0.59 >$$
$$\omega_{\{1,1,1,0\}} =< 0.02, 0.15, 0.83, 0 >, \quad \omega_{\{1,1,0,1\}} =< 0.01, 0.11, 0, 0.88 >$$

$$\omega_{\{1,0,1,1\}} =< 0.01, 0, 0.40, 0.59 >, \quad \omega_{\{0,1,1,1\}} =< 0, 0.07, 0.38, 0.55 >$$
$$\omega_{\{1,1,1,1\}} =< 0.01, 0.07, 0.38, 0.55 >$$

5. Discussion

The Noisy-MAX parameterization [8] eases the elicitation of conditional probability tables (CPTs) in Bayesian networks, because, due to the assumption of independence of causal influences (ICI), the number of parameters required to specify a CPT is linear in the number of parent variables, instead of exponential. In the Noisy-MAX parameterization, a strict order relation holds on the sample space of the response, so that it is possible to define the value taken by the response as a deterministic function of the values taken by latent causes. We referred to this property as strong dominance.

In medical applications of Bayesian networks, it is instead useful to deal with soft dominance, that is the value of the response is selected according to a probability distribution over the states receiving at least one vote from latent causes. In general, soft dominance may characterize a medical response variable no matter whether a strict order relation holds on its states. At this purpose, we proposed the SoftDom parameterization as a generalization of the Noisy-MAX parameterization, which is suited to a general biomedical response variable influenced by independent causal determinants.

The SoftDom parametrization was motivated by the development of a Bayesian network for the diagnosis of cardiopulmonary diseases [15], nevertheless other fields of application may be envisioned. For example, since a cell's activity is organized as a network of interacting modules [18], independent activators and repressors genes may be hypothesized to regulate the concentration of key enzymes in the phosphorilation and metylation processes.

The SoftDom parameterization is defined by the same parameters defining the Noisy-MAX parameterization, plus a number of additional parameters, whose number is exponential in the number of states of the response. Although we expect that most of such parameters are extreme or uniform probability values requiring no significant effort to be elicited from medical experts, we developed a specialization of the SoftDom parameterization, called Noisy-SCORE, requiring a number of additional parameters which is linear in the number of states of the response. The Noisy-SCORE is a very flexible conditional model because it approximates strong dominance at the desired level of precision, and, as shown by several numerical examples, it accurately represents soft dominance if the probability of selection for each state is approximately proportional all over the possible configuration of votes.

Acknowledgments

This work was partially supported by the University of Florence, funding framework "Progetto strategico di ricerca di base per l'anno 2015", grant "Disegno e analisi di studi sperimentali e osservazionali per le decisioni in ambito epidemiologico, socio-economico, ambientale e tecnologico".

References

[1] W. L. Buntine. Operations for Learning with Graphical Models. *Journal of Artificial Intelligence Research*, **2**: 159-225, 1994.

[2] F. J. Díez, M. J. Druzdzel. Canonical Probabilistic Models for Knowledge Engineering. *Technical Report CISIAD-06-01*, UNED, Madrid (ES), 2006.

[3] F. J. Díez, J. Mira, E. Iturralde, S. Zubillaga. DIAVAL, a Bayesian Expert System for Echocardiography. *Artificial Intelligence in Medicine*, **10**: 59-73, 1997.

[4] F. J. Díez. Parameter Adjustment in Bayes Networks: The Generalized Noisy-OR Gate. *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 99-105. Washington (US-DC), 1993.

[5] M. J. Druzdzel, L. C. van der Gaag. Building Probabilistic Networks: Where Do the Numbers Come From? *IEEE Transactions on Knowledge and Data Engineering*, **12**(4): 481-486, 2000.

[6] S. F. Galan and F. J. Dìez. Modeling Dynamic Causal Interactions with Bayesian Networks: Temporal Noisy Gates. *Proceedings of the 2nd International Workshop on Causal Networks*, pages 1-5, Berlin (DE), 2000.

[7] I. Good. A causal calculus. *British Journal of Philosophy of Science*, **11**: 305-318, 1961.

[8] D. Heckerman, J. S. Breese. Causal Independence for Probability Assessment and Inference Using Bayesian Networks. *Systems, Man and Cybernetics*, **26**(6): 826-831, 1996.

[9] D. Heckerman, J. S. Breese. A New Look at Causal Independence. *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, pages 286-292, Seattle (US-WA), 1994.

[10] D. Heckerman. Similarity Networks for the Construction of Multiple-Faults Belief Networks. *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, pages 51-64, Cambridge (US-MA), 1990.

[11] M. Henrion. Some Practical Issues in Constructing Belief Networks. *Uncertainty in Artificial Intelligence*, **3**:161-173, 1989.

[12] K. B. Korb, A. E. Nicholson. Bayesian Artificial Intelligence. CRC press, 2nd edition, 2010.

[13] J. F. Lemmer and D. E. Gossink. Recursive Noisy OR - A Rule for Estimating Complex Probabilistic Interactions. *IEEE Systems, Man, and Cybernetics Society*, Part B, **34**(6): 2252-2261, 2004.

[14] P. J. F. Lucas, L. C. van der Gaag, A. Abu-Hanna. Bayesian Networks in Biomedicine and Health-Care Artificial Intelligence in Medicine, 30(3): 201-214, 2004.

[15] D. Luciani, S. Cavuto, L. Antiga, M. Miniati, S. Monti, M. Pistolesi, G. Bertolini. Bayes Pulmonary Embolism Assisted Diagnosis: A New Expert System For Clinical Use. *Emergence Medicine Journal*, **24**(3): 157-164, 2007

[16] A. Onisko, M. J. Druzdzel, H. Wasyluk. Learning Bayesian Networks Parameters from Small Data Sets: Applications of Noisy-OR Gates. *International Journal of Approximate Reasoning*, **27**: 165-182, 2001.

[17] J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.

[18] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, N. Friedman. Module Networks: Identifying Regulatory Modules and Their Condition-Specific Regulators from Gene Expression Data. *Nature Genetics*, **34**(2): 166-176 (2003).

[19] S. Srinivas. A Generalization of Noisy-Or Model. *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 208-215, 1993.