Enhancing Bayesian Networks with Psychometric Models

Iván Pérez CABRERA@UTIA.CAS.CZ
Institute of Information Theory and Automation, Czech Academy of Sciences
Institute of Computer Science, Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University
Jiří Vomlel VOMLEL@UTIA.CAS.CZ
Institute of Information Theory and Automation, Czech Academy of Sciences

Editors: J.H.P. Kwisthout & S. Renooij

Abstract

Bayesian networks (BNs) are a popular framework in education and other fields. In this paper, we consider two-layer BNs, where the first layer consists of hidden binary variables that are assumed to be independent of each other, and the second layer consists of observed binary variables. The variables in the second layer depend on the variables in the first layer. The dependence is characterized by conditional probability tables, which represent Noisy-AND models. We refer to this class of models as BN2A models. We found that these models are also popular in the psychometric community, where they can be found under the name of Cognitive Diagnostic Models (CDMs), which are used to classify test takers into some latent classes according to the similarity of their responses to test questions. This paper shows the relation between some BN2A models and their corresponding CDMs. In particular, we compare the performance of these models on large-scale tests conducted in the Czech Republic in 2022. The BN2A model with general conditional probability tables produced the best absolute fit. However, when we added monotonic constraints to the General model, we obtained better predictive results.

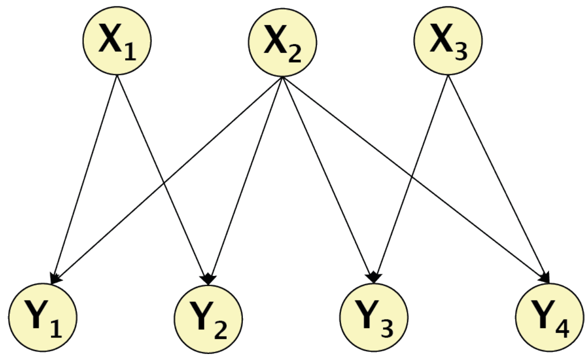
Keywords: Bayesian networks, Parameter Learning, Hidden Variables, BN2A models, Cognitive Diagnostic Modeling, Psychometrics.

1. Introduction

Bayesian networks (Pearl, 1988; Jensen and Nielsen, 2007; Koller and Friedman, 2009) are a popular framework for modelling probabilistic relationships between random variables. We are interested on a special class of Bayesian Networks (BNs) - two-layer BNs, where the first layer consists of hidden variables, which are assumed to be mutually independent, and the second layer consists of observed variables. All variables are assumed to be binary. The variables in the second layer depend only on the variables in the first layer. The dependence is characterised by conditional probability tables (CPTs). In this paper we are interested in CPTs that are represented by Noisy-AND models, the corresponding BN will be called BN2A. In Fig. 1 we give an example of a directed bipartite graph that can define the structure of a BN2A model.

Noisy-AND models are examples from the family of canonical models of CPTs (Henrion, 1987; Díez and Druzdzel, 2006). The study of these models is motivated by practical applications. BN2A models are used in psychometrics for cognitive diagnostic modeling of students. In this case, the hidden variables correspond to the student's skills and the

Figure 1: An example of a directed bipartite graph.



observed variables correspond to the student's responses to test questions. A typical test question requires all related skills to be present, unless a missing skill is compensated by another knowledge or skill. This relationship is well represented by Noisy-AND models.

There is related work that mentions the connection between Bayesian networks and psychometrics, one of the best references is the book: Bayesian Networks in Educational Assessment (Almond et al., 2015), in particular, most of the examples presented in it consider continuous latent variables. The book also discusses the noisy models (with binary latent variables) and describes in detail the DINA (Simple Noisy-AND) model with two observed variables and one hidden variable. In our work: a) we present this model in a general form, and b) we present two other more complex models (the Noisy-AND, and the General model) and show how the psychometric models R-RUM and G-DINA can be derived.

In this work we are interested in BN2A models, but there are other models that can also be used in educational assessment (Almond et al., 2015). One example is the hierarchical latent class (HLC) models, which are Bayesian networks whose structures are rooted trees where the leaf nodes are observed while all other nodes are latent (hidden), the main advantage of these models over those analyzed is that they can model local dependence. Another two other interesting examples from the field of psychometric are logistic regression (LR) models and item response theory (IRT) models, their main difference from our models is that they consider continuous latent variables.

This manuscript is organized as follows. In Section 2 we formally introduce the BN2A models. We present the General model, define the monotonicity constraint, and derive two specific models: The Noisy-AND model and the Simple Noisy-AND model. In Section 3, we describe the dataset used and the method we used to specify the corresponding Q-matrix. The fit of our models is described in Section 4. Finally, we summarize our contribution in Section 5.

2. BN2A models

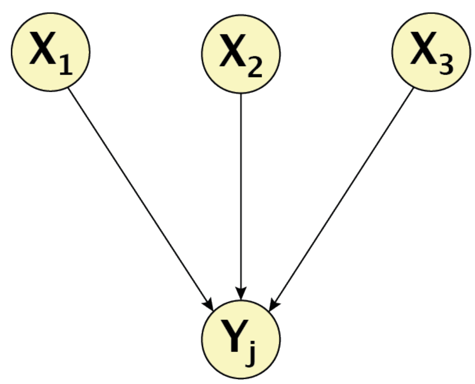
In this manuscript, we use $\mathbf{X}$ to denote the vector $(X_1, \ldots, X_K)$ of K hidden variables, and similarly $\mathbf{Y}$ to denote the vector $(Y_1, \ldots, Y_L)$ of L observed dependent variables. BN2A models are characterized by being represented by a bipartite graph, where the first layer of variables is hidden while the second layer corresponds to observed variables. The hidden variables will be referred also as skills due to application to educational domain. The observed dependent variables corresponds to exam questions. All variables are assumed to be binary, taking states from $\{0, 1\}$.

The main component of the presented BN2A models are conditional probability tables (CPTs) specified in the form of a Noisy-AND model. The three models of interest are presented in the following: The General model, the Noisy-AND model, and the Simple Noisy-AND model.

2.1. General model

To introduce the General model, we will start with an example. Consider an observed variable Y_j whose value depends on three hidden variables (X_1, X_2, and X_3) represented on Fig. 2. In the context of educational assessment, the observed variable represents a test question, while the hidden variables represent skills required to answer this question correctly.

Figure 2: BN2A model with three hidden variables and one observed variable.



The CPT of the General model (Table 1) involves the main effects and all possible interactions between three skills measured by the question:

- $q_{j,0}$ represents the probability of answering question Y_j correctly when all of the required skills are present, and it is traditionally called *leak* probability.
- q_{j,k_1} represents the main effect of one skill k_1 on question Y_j.
- q_{j,k_1k_2} represents the interaction effect of two skills k_1 and k_2 on question Y_j.
- $q_{j,k_1k_2k_3}$ represents the interaction effect of all three skills on question Y_j.

In this example, there are three main effects ($q_{j,1}$, $q_{j,2}$, and $q_{j,3}$), three 2-way interaction effects ($q_{j,12}$, $q_{j,13}$, and $q_{j,23}$), and one 3-way interaction effect ($q_{j,123}$). The range of q effects is from 0 to 1 and they can be interpreted as penalization factors since they reduce the probability of answering correctly.

Then, according to Table 1, the probability to answer correctly the question Y_j given that only skills X_1 and X_2 are mastered is equal to $q_{j,0} \cdot q_{j,3} \cdot q_{j,13} \cdot q_{j,23} \cdot q_{j,123}$ (the product of *leak* probability of question Y_j and all effects that consider the non-mastered skill X_3).

x_1	x_2	x_3	$P(Y_j = 1 \mid x_1, x_2, x_3)$
1	1	1	$q_{j,0}$
0	1	1	$q_{j,0} \cdot q_{j,1} \cdot q_{j,12} \cdot q_{j,13} \cdot q_{j,123}$
1	0	1	$q_{j,0} \cdot q_{j,2} \cdot q_{j,12} \cdot q_{j,23} \cdot q_{j,123}$
1	1	0	$q_{j,0} \cdot q_{j,3} \cdot q_{j,13} \cdot q_{j,23} \cdot q_{j,123}$
0	0	1	$q_{j,0} \cdot q_{j,1} \cdot q_{j,2} \cdot q_{j,12} \cdot q_{j,13} \cdot q_{j,23} \cdot q_{j,123}$
0	1	0	$q_{j,0} \cdot q_{j,1} \cdot q_{j,3} \cdot q_{j,12} \cdot q_{j,13} \cdot q_{j,23} \cdot q_{j,123}$
1	0	0	$q_{j,0} \cdot q_{j,2} \cdot q_{j,3} \cdot q_{j,12} \cdot q_{j,13} \cdot q_{j,23} \cdot q_{j,123}$
0	0	0	$q_{j,0} \cdot q_{j,1} \cdot q_{j,2} \cdot q_{j,3} \cdot q_{j,12} \cdot q_{j,13} \cdot q_{j,23} \cdot q_{j,123}$

Table 1: CPT for General model with three hidden variables.

To simplify the subsequent mathematical notation we introduce the following indicator function:

$$\chi(x_S) = \begin{cases} 1 & \text{if } x_i = 1 \text{ for all } i \in S \\ 0 & \text{otherwise.} \end{cases}$$

Let Y_j be an observed variable representing an exam question and $pa(Y_j)$ be the subset of indexes of related hidden variables (skills) from $\mathbf{X} = (X_1, \ldots, X_K)$. They are referred to as the parents of Y_j. We define the probability of answering correctly the question Y_j given $pa(Y_j)$ as:

$$P(Y_j = 1 \mid \mathbf{x}_{pa(Y_j)}) = q_{j,0} \cdot \prod_{S \subseteq pa(Y_j)} (q_{j,S})^{1-\chi(x_S)} \quad , \tag{1}$$

where

- $q_{j,0}$ represents the probability of answering question Y_j correctly when all of the required skills are present, (*leak* probability), and
- $q_{j,S}$ represents the $|S|$-way interaction effect of skill subset S on question Y_j.

Let K_j be the total of parents of Y_j, then, there is $\binom{K_j}{0} = 1$ *leak* effect, there are $\binom{K_j}{1} = K_j$ main effects, $\binom{K_j}{2}$ two-way interaction effects, and so on. Then, the model has $\sum_{i=0}^{K_j} \binom{K_j}{i} = 2^{K_j}$ parameters for the question Y_j. If we compute the logarithm of $P(Y_j = 1 \mid \mathbf{x}_{pa(Y_j)})$ from (1) considering the indicator function $\chi(x_S)$ instead of its complement $1 - \chi(x_S)$, and group the summands by the cardinality of the interaction effects, we get the expression

$$\delta_{j,0} + \sum_{\substack{S \subseteq pa(Y_j) \\ |S|=1}} \delta_{j,S} \cdot \chi(x_S) + \sum_{\substack{S \subseteq pa(Y_j) \\ |S|=2}} \delta_{j,S} \cdot \chi(x_S) + \cdots + \delta_{j,1,2\ldots K_j} \cdot \chi(x_{pa(Y_j)})$$

This expression is known in psychometrics as the General Deterministic Input Noisy-AND Model (G-DINA). It is a popular framework to classify subjects according to their skill mastering (de la Torre, 2011; Ma and de la Torre, 2020; Gu and Xu, 2021).

The General model is not restrictive, students with fewer required skills for a question Y can have a higher probability of answering the question correctly. To adress this issue, it is possible to add monotonicity constraints to the model. In the context of BNs, the concept of monotonicity constraint has been discussed in the literature for a long time (Wellman, 1990; Druzdzel and Henrion, 1993). More recent papers in this topic are (Restificar and Dietterich, 2013), (Masegosa et al., 2016) and (Plajner and Vomlel, 2020).

Bayesian networks model the probabilistic influences between its variables. Considering binary variables X and Y, a positive qualitative influence of a variable X on a variable Y along an arc $X \to Y$ in the network means that the occurrence of X increases the probability of Y occurs, assuming that the values of the other parents of Y remain the same. It means that

$$P(Y = 1|X = 1, \mathbf{z}) \quad \geq \quad P(Y = 1|X = 0, \mathbf{z})$$

for any combination of values $\mathbf{z}$ for the set of parents of Y other than X (Masegosa et al., 2016). In the context of educational testing, a positive influence is commonly assumed since mastering the skill X increases the probability of answering correctly the question Y. In this manuscript we will refer to this positive influence as monotonic constraint. In Section 4 we compare the fit of the General model with and without monotonic constraints.

2.2. Noisy-AND model

From the General model, we can derive the Noisy-AND model if we omit the interaction effects between the hidden variables, i.e., if we consider the influence of each hidden variable on an observed variable Y_j does not depend on the values of the other hidden variables. This assumption offers the advantage of having a simpler model (a reduced number of parameters) that is easier to interpret and, under certain conditions, can be identifiable (Pérez and Vomlel, 2024).

Analogously to the General model, we can introduce the Noisy-AND model with the structure presented on Fig. 2 (an observed variable Y_j whose value depends on three hidden variables (X_1, X_2, and X_3). The CPT of the Noisy-AND model (Table 2) involves only the *leak* probability and the main effects of the three skills measured by the question:

- $q_{j,0}$ represents the probability of answering question Y_j correctly when all of the required skills are present (*leak* probability), while
- q_{j,k_1} represents the main effect of one skill k_1 on answering question Y_j.

In contrast with the General model, for the question Y_j, the Noisy-AND model has only $K_j + 1$ parameters, $\binom{K_j}{1} = K_j$ main effects and $\binom{K_j}{0} = 1$ *leak* effect.

In general, for the Noisy-AND model, we define the probability of answering correctly the question Y_j given $pa(Y_j)$ as

$$P(Y_j = 1|\mathbf{x}_{pa(Y_j)}) = q_{j,0} \cdot \prod_{i \in pa(Y_j)} (q_{j,i})^{1-\chi(x_i)} \tag{2}$$

x_1	x_2	x_3	$P(Y_j = 1 \mid x_1, x_2, x_3)$
1	1	1	$q_{j,0}$
0	1	1	$q_{j,0} \cdot q_{j,1}$
1	0	1	$q_{j,0} \cdot q_{j,2}$
1	1	0	$q_{j,0} \cdot q_{j,3}$
0	0	1	$q_{j,0} \cdot q_{j,1} \cdot q_{j,2}$
0	1	0	$q_{j,0} \cdot q_{j,1} \cdot q_{j,3}$
1	0	0	$q_{j,0} \cdot q_{j,2} \cdot q_{j,3}$
0	0	0	$q_{j,0} \cdot q_{j,1} \cdot q_{j,2} \cdot q_{j,3}$

Table 2: CPT for Noisy-AND model with three hidden variables.

From (2), we can see that if the skill X_i is not mastered ($x_i = 0$) then the value $\chi(x_i)$ is equal to 1, and as a result, the penalization factor $q_{j,i}$ is present when the probability of answering question Y_j correctly is computed.

If we compute the logarithm of $P(Y_j = 1 \mid \mathbf{x}_{pa(Y_j)})$ on (2) considering the indicator function $\chi(x_i)$ instead of its complement $1 - \chi(x_i)$, we get the expression

$$\delta_{j,0} + \sum_{i \in pa(Y_j)} \delta_{j,i} \cdot \chi(x_i)$$

This expression is known as Reduced Reparameterized Unified Model (R-RUM) in psychometrics, and has been popular in the last two decades (Hartz, 1996; Culpepper and Chen, 2019).

2.3. Simple Noisy-AND model

There is an even simpler model that can be derived from the general model, we call it Simple Noisy-AND, and considers only two parameters:

- g_j represents the probability of answering question Y_j correctly when the student does not master all the skills required for question Y_j, it is usually called *guessing* parameter.
- s_j represents the probability of answering question Y_j incorrectly when the student masters all the skills required for question Y_j, it is usually called *slipping* parameter.

For the Simple Noisy-AND model, we define the probability of answering correctly the question Y_j given $pa(Y_j)$ as:

$$P(Y_j = 1 \mid \mathbf{x}_{pa(Y_j)}) = (g_j)^{\chi(\mathbf{x}_{pa(Y_j)})} \cdot (1 - s_j)^{1 - \chi(\mathbf{x}_{pa(Y_j)})} \tag{3}$$

From (3), we can see that if the student does not masters all the skills required by question Y_j, then the value $\chi(\mathbf{x}_{pa(Y_j)})$ is equal to 1, and as a result, the probability of answering question Y_j correctly is g_j. In other words, it is possible to answer correctly by guessing.

By computing the logarithm of $P(Y_j = 1 \mid \mathbf{x}_{pa(Y_j)})$ from (3), we get the expression

$$\chi(\mathbf{x}_{pa(Y_j)}) \cdot \log(g_j) + (1 - \chi(\mathbf{x}_{pa(Y_j)})) \cdot \log(1 - s_j)$$

and by factoring $\chi(\mathbf{x}_{pa(Y_j)})$, we obtain

$$\delta_{j,0} + \delta_{j,1,2\ldots K_j} \cdot \chi(\mathbf{x}_{pa(Y_j)}) \ ,$$

where $\delta_{j,0} = \log(1-s_j)$ and $\delta_{j0}+\delta_{j,1,2\ldots K_j} = \log(g_j)$. This expression is known in psychometrics as the Deterministic Input Noisy-AND Model (DINA). Regardless of its simplicity, it is a popular framework and many researchers are still analyzing these models nowadays (Gu, 2023; Gu and Xu, 2021; Ma and de la Torre, 2020).

3. The CERMAT dataset

3.1. Questions and skills

The Ministry of Education, Youth and Sports of the Czech Republic has established an experimental verification of knowledge and skills in secondary school mathematics. The catalog of requirements for the mathematics exam includes nine main topics, and their respective representation is determined by CERMAT (Center for the Determination of Educational Results). The nine topics evaluated in the exam are: *Numerical sets*, *Algebraic expressions*, *Equations and inequalities*, *Functions*, *Sequences and series*, *Planimetry*, *Stereometry*, *Analytic geometry*, and *Combinatorics, probability and statistics*.

The exam is composed of 30 questions, 19 of them are open questions (where the student must write his/her own procedure to give the correct answer) and 11 questions are closed questions (in which the student will choose the correct answer from a set of plausible options). In the exam, not all the questions have the same value, for this reason, we binarize the grade obtained in each question before performing our analysis, so a value of 1 means the question was answered correctly and 0 means the answer was incorrect. The datasets used in this study are publicly available in the statistical section of the CERMAT website: vysledky.cermat.cz/statistika. There are two evaluation periods: Spring and Fall.

The dataset we used contains information from the Czech high school final exam from the spring of 2022. This dataset is representative as the sample size is large ($N = 12709$) and students come from all regions of the Czech Republic.

3.2. Q-Matrix specification

For any BN2A model (equivalently, CDM), it is necessary to specify the underlying graph of the model, i.e., the relationship between questions and skills. This is done using an incidence matrix, called a Q-matrix. The Q-matrix encodes the underlying graph of the model. It is a $l \times k$ matrix where the rows represent questions and the columns represent the measured skills. If a skill X_j is required by a question Y_i, then the entry (i, j) of the Q-matrix is equal to 1, otherwise the assigned value is 0.

In this study the Q-matrix was designed in two steps: First, we analyzed the question wording of the Czech high school final exam (Spring 2022), and based on that we defined a theoretical framework that includes six mathematical skills, they are listed in Table 3.

Second, the questions were coded by the first author of this paper and an external education specialist. Agreement was measured using Cohen's Kappa coefficient. On a first attempt, 22 of the 30 questions showed substantial or complete agreement (Cohen's Kappa greater than 0.55). Differences were discussed and resolved in a subsequent meeting. We

Skill	Description
X_1	Perform operations with arithmetic and algebraic expressions.
X_2	Solve different types of equations (e.g., quadratic, system of linear equations)
X_3	Recognize and evaluate different types of functions (e.g., logarithmic, trigonometric)
X_4	Solve geometric problems in two and three dimensions.
X_5	Solve problems of combinatorics and probability.
X_6	Interpret word problems in algebraic language.

Table 3: Skills proposed for the Czech high school final exam (Spring 2022)

obtained the Q-matrix shown in Table 4. In particular, we can observe that the first skill is required for all questions.

	Skill							Skill					
Question	X_1	X_2	X_3	X_4	X_5	X_6	**Question**	X_1	X_2	X_3	X_4	X_5	X_6
Y_1	1	0	0	0	0	0	Y_{16}	1	0	0	0	0	0
Y_2	1	0	0	0	0	0	Y_{17}	1	0	0	0	0	0
Y_3	1	0	0	0	0	1	Y_{18}	1	1	0	0	0	1
Y_4	1	0	0	0	0	1	Y_{19}	1	1	0	0	0	1
Y_5	1	0	0	1	0	1	Y_{20}	1	0	0	0	0	1
Y_6	1	0	0	1	0	1	Y_{21}	1	0	0	0	0	1
Y_7	1	0	0	1	0	1	Y_{22}	1	1	0	0	0	0
Y_8	1	0	0	0	0	0	Y_{23}	1	0	1	1	0	0
Y_9	1	1	0	0	0	0	Y_{24}	1	0	0	1	0	1
Y_{10}	1	1	0	0	0	0	Y_{25}	1	0	0	1	0	1
Y_{11}	1	0	1	1	0	0	Y_{26}	1	0	0	1	0	0
Y_{12}	1	1	1	0	0	0	Y_{27}	1	0	1	0	0	0
Y_{13}	1	1	1	0	0	0	Y_{28}	1	0	0	0	1	1
Y_{14}	1	0	0	0	1	1	Y_{29}	1	0	0	0	0	0
Y_{15}	1	0	0	0	1	1	Y_{30}	1	0	0	1	0	0

Table 4: Q-matrix for the Czech high school final exam (Spring 2022)

4. Experiments

The R-package GDINA (Ma and de la Torre, 2020) was used to estimate the four models (General, Monotone General, Noisy-AND, and Simple Noisy-AND). In this package, Marginal Maximum Likelihood method with Expectation-Maximization (MMLE/EM) algorithm is used for item parameter estimation. The model fit was evaluated for each of the proposed models, and then the skill classification results were compared across the four models.

4.1. Model Fit

The relative fit statistics of the BN2A models were compared. In Table 5 it can be seen that the General model is the best model according to the presented criteria (AIC = -382 430; BIC = -384 121; CAIC = -384 348) and similar results were obtained when considering the General model with monotonicity constraints. The second best model is the Noisy-AND model, while the least suitable is the Simple Noisy-AND model.

Model	Simple Noisy-AND	Noisy-AND	Monotone General	General
Parameters	123	161	227	227
Loglik	-197 859	-192 190	-191 036	-190 988
AIC	-395 964	-384 703	-382 526	-382 430
BIC	-396 881	-385 903	-384 217	-384 121
CAIC	-397 004	-386 064	-384 444	-384 348

Table 5: Relative Fit Statistics

Additionally, in Table 6, we present two known statistics for measuring the absolute fit of each model, the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMSR) (Maydeu-Olivares and Joe, 2014). SRMSR is a measurement that assesses the approximate fit of large models when the data are ordinal. For a pair of items Y_i and Y_j, the residual correlation is the sample correlation minus the expected correlation. For both RMSEA and SRMSR, a smaller value indicates a better absolute model data fit. Simulation studies suggest that RMSEA < 0.03 indicates excellent fit, $0.03 <$ RMSEA < 0.045 a good fit, and RMSEA > 0.045 poor fit, analogously, SRMSR < 0.05 indicates good model fit (Shi et al., 2021).

In our experiment we can see that, with the exception of Simple Noisy-AND, in general all models show an excellent fit with respect to RMSEA and a good fit with respect to SRMSR.

	Simple Noisy-AND	Noisy-AND	Monotone General	General
RMSEA	0.0458	0.0306	0.0275	0.0275
SRMSR	0.0779	0.0497	0.0425	0.0424

Table 6: Comparing models using two Absolute Fit Statistics

4.2. Skill classification

Table 7 presents the prior probability that students master each of the six skills estimated by each of the BN2A models. Considering that skills with low prior probability correspond to skills that are expected to be difficult and skills with high prior probability correspond to skills that are expected to be easy, the results can be interpreted to determine the skill difficulty. In general, it can be observed that, with the exception of the simple model, the prior probability of each skill is similar in the rest of the models. In particular, it can be seen that skill X_5 (Solve problems of combinatorics and probability) has the lowest prior probability, this is consistent with the fact that the related concepts in the CERMAT test curriculum are taught at the end of secondary education because they are more complex. On the other hand, skill X_1 (Perform operations with arithmetic and algebraic expressions) has the highest prior probability, and is consistent with the fact that the related concepts are the basis of the CERMAT test curriculum.

	Simple Noisy-AND	Noisy-AND	Monotone General	General
$p(X_1 = 1)$	0.7053	0.7040	0.6334	0.6350
$p(X_2 = 1)$	0.6776	0.4380	0.4403	0.4249
$p(X_3 = 1)$	0.6470	0.3985	0.3839	0.3816
$p(X_4 = 1)$	0.7038	0.5113	0.5161	0.5015
$p(X_5 = 1)$	0.3001	0.1886	0.1685	0.1566
$p(X_6 = 1)$	0.7120	0.4403	0.5021	0.5121

Table 7: Prior probabilities of each skill estimated for each model

In psychometrics, BN2A models belong to a family called latent class models that classify students into some latent classes according to their responses to test questions. With 6 underlying skills, the students are classified into 64 (i.e., 2^6) latent classes, also called: skill profiles. The proportion of the first five most likely skill profiles for the four models are presented in Table 8, in which value 1 indicates mastery of the skill and value 0 indicates non-mastery of the skill. For instance, a skill profile of **100000** (the most representative profile in three of the four models) indicates that the students master skill X_1 and do not master the rest of the skills.

In particular, it can be seen that the most representative profiles in the Noisy-AND, Monotone General, and General models are the same (**100000**, **000000**, **111111**, **100100**, and **111101**) and classify more than 50% of the sample. Another interesting observation is that these profiles include both the profile of students who had not mastered any of the skills and the profile of students who had mastered all six skills.

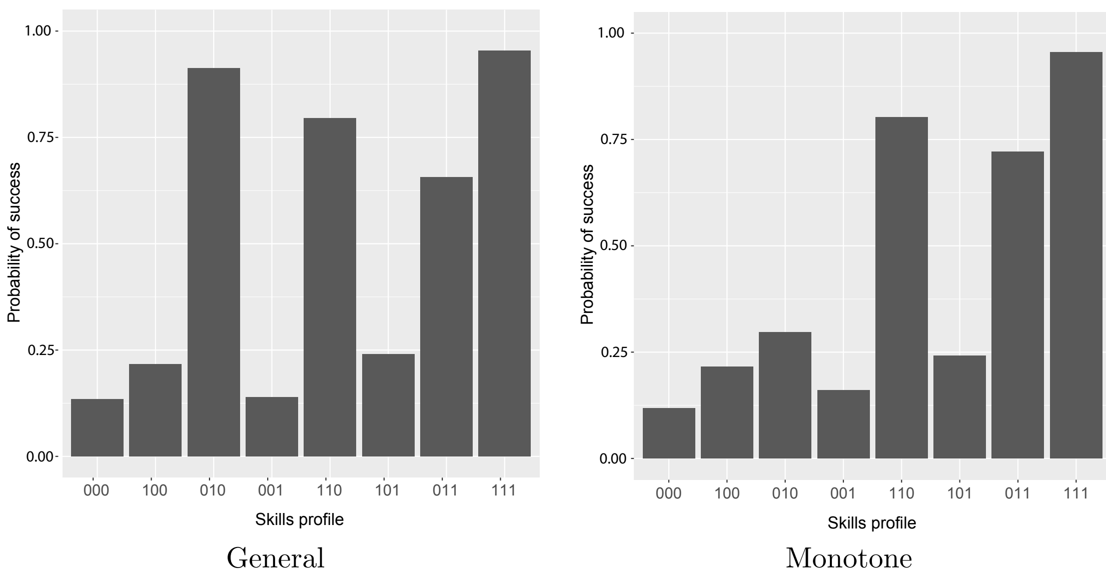
	Simple Noisy-AND		Noisy-AND		Monotone General		General	
	class	prob	class	prob	class	prob	class	prob
1^{st}	**111101**	0.2715	**100000**	0.1723	**100000**	0.1550	**100000**	0.1602
2^{nd}	**111111**	0.1582	**000000**	0.1461	**000000**	0.1140	**000000**	0.1182
3^{rd}	**100101**	0.0469	**100100**	0.1056	**111111**	0.0937	**111111**	0.0953
4^{th}	**111100**	0.0389	**111111**	0.1026	**100100**	0.0926	**100100**	0.0872
5^{th}	**111001**	0.0376	**111101**	0.0963	**111101**	0.0792	**111101**	0.0830

Table 8: The five most representative skill profiles for each model

It is interesting to note that after we learned the general model, 13 of the 30 questions did not satisfy the monotonicity condition, but the fits of both models (with and without the monotonicity constraint) are very similar.

As an example, Figure 3 shows the probabilities of answering question 14 correctly given the three required skills. For this question, the general model does not satisfy the monotonicity condition, the probability of answering correctly with only the second skill ($P(010) = 0.913$) is higher than the probability of answering correctly with the first and second skills ($P(110) = 0.795$).

Figure 3: Comparison of the probabilities of correct answers to Question 14 for the general and monotone models (the height of the columns corresponds to the values in the tables presented below the bar plots).



Model	P(000)	P(100)	P(010)	P(001)	P(110)	P(101)	P(011)	P(111)
General	0.134	0.216	0.913	0.139	0.795	0.240	0.656	0.954
Monotone	0.117	0.215	0.297	0.160	0.802	0.241	0.721	0.955

Note that both general models (monotone and not necessarily monotone) have the same number of parameters (227, in our experiment), and the log-likelihood of the general model without the monotonicity constraint cannot be worse than the log-likelihood of a model with the same parameters but with the constraint. Therefore, the BIC of the general model without the monotonicity constraint cannot be worse.

4.3. Prediction accuracy

In this paper, we also use an alternative evaluation method. We performed a 10-fold cross-validation for the general model (with and without the monotonicity constraint). In each iteration, we randomly selected 20 questions for each subject, took their values as evidence, and then inferred the result of the remaining 10 questions. We compute the average percentage of correct predictions of the actual values from each testing dataset. The prediction accuracy is presented in Table 9 for each iteration. It can be seen that in 7 out of the 10 iterations of cross-validation, the results were better for the model that considers the monotonicity constraints.

Iteration	1	2	3	4	5	6	7	8	9	10
General	0.658	0.665	**0.682**	0.663	**0.663**	0.662	**0.666**	0.668	0.669	0.664
Monotone	**0.669**	**0.676**	0.678	**0.668**	0.662	**0.665**	0.665	**0.671**	**0.670**	**0.675**

Table 9: Accuracy of prediction for the general model and the model with monotonicity constraints for each iteration of the 10-fold cross validation.

5. Concluding Remarks

In this paper, we have presented four BN2A models: General model, Monotone General model, Noisy-AND model, and Simple Noisy-AND model. They have been studied in psychometrics under different names. We show how these models are useful for classifying students according to their skills, but it is important to mention that they have also been used in other different fields such as Medicine or Biology.

In our work, based on real data from a large-scale assessment, we fit the four models mentioned and found that for this experiment, the model with the best fit was the General model. It is important to note that this is not always the case; there have been studies where models with fewer parameters fit better. When we estimated the parameters of the General model, we noticed that almost all of the questions requiring three skills did not satisfy the monotonicity constraints. Therefore, we decided to also estimate the general model under these conditions. The fit of the general model with and without monotonicity constraints to the training data was similar. Additionally, we performed a cross-validation to compare the predictive power of both models. In 7 out of 10 iterations, we obtained better predictions considering monotonicity constraints.

We plan to extend our analysis to other datasets of CERMAT test results for periods other than 2022. Our proposal is to compare BN2A models with models allowing dependent skills, and with respect to BN2A models, not only learn the model parameters but also their structure (i.e., the Q-matrix) from the data, related research has already been submitted for publication (Pérez and Vomlel, under review).

Acknowledgments

This work was supported by the grant 22-11101S of the Czech Science Foundation. We are grateful to Jan Netık for his invaluable assistance in analyzing the CERMAT evaluation, and to the anonymous reviewers for their valuable feedback and insightful comments, which greatly contributed to improving the quality of this paper.

References

R. G. Almond, R. J. Mislevy, L. Steinberg, D. Yan, and D. Williamson. *Bayesian Networks in Educational Assessment*. Springer Publishing Company, Incorporated, 2015. ISBN 149392124X.

S. A. Culpepper and Y. Chen. Development and application of an exploratory reduced reparameterized unified model. *Journal of Educational and Behavioral Statistics*, 44(1): 3–24, 2019. ISSN 10769986, 19351054. URL http://www.jstor.org/stable/45278331.

J. de la Torre. The generalized DINA model framework. *Psychometrika*, 76:179–199, 2011.

F. J. Díez and M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01, UNED, Madrid, Spain, 2006.

J. Druzdzel and M. Henrion. Efficient Reasoning in Qualitative Probabilistic Networks. *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 548–553, 1993.

Y. Gu. Generic Identifiability of the DINA Model and Blessing of Latent Dependence. *Psychometrika*, 88(1):117–131, March 2023. doi: 10.1007/s11336-022-09886-. URL https://ideas.repec.org/a/spr/psycho/v88y2023i1d10.1007_s11336-022-09886-2.html.

Y. Gu and G. Xu. Sufficient and necessary conditions for the identifiability of the Q-matrix. *Statistica Sinica*, 31:449–472, 2021. doi: 10.5705/ss.202018.0410.

S. M. Hartz. *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. PhD thesis, University of Illinois at Urbana-Champaign, 1996.

M. Henrion. Some practical issues in constructing belief networks. In *Proceedings of the Third Conference on Uncertainty in Artificial Intelligence (UAI-87)*, pages 161–173. Elsevier Science Publishers B.V. (North Holland), 1987.

F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer New York, NY, 2 edition, 2007. doi: 10.1007/978-0-387-68282-2.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

W. Ma and J. de la Torre. GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, 93(14):1–26, 2020. doi: 10.18637/jss.v093.i14.

A. R. Masegosa, A. J. Feelders, and L. C. van der Gaag. Learning from Incomplete Data in Bayesian Networks with Qualitative Influences. *International Journal of Approximate Reasoning*, 69:18–34, 2016.

A. Maydeu-Olivares and H. Joe. Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4):305–328, 2014. doi: 10.1080/00273171.2014.911075. URL https://doi.org/10.1080/00273171.2014.911075.

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

I. Pérez and J. Vomlel. On identifiability of BN2A networks. In Z. Bouraoui and S. Vesic, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2023*, volume 14294 of *Lecture Notes in Computer Science*, pages 136–148, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-45608-4. doi: https://doi.org/10.1007/978-3-031-45608-4_11.

I. Pérez and J. Vomlel. BN2A models: Identifiability and structural learning. *International Journal of Approximate Reasoning*, under review.

M. Plajner and J. Vomlel. Learning bipartite Bayesian networks under monotonicity restrictions. *International Journal of General Systems*, 49(1):88–111, 2020. doi: 10.1080/03081079.2019.1692004.

A. C. Restificar and T. G. Dietterich. Exploiting Monotonicity via Logistic Regression in Bayesian Network Learning. *Technical Report. Corvallis*, 2013.

Q. Shi, W. Ma, A. Robitzsch, M. A. Sorrel, and K. Man. Cognitively diagnostic analysis using the g-dina model in r. *Psych*, 3(4):812–835, 2021. ISSN 2624-8611. doi: 10.3390/psych3040052. URL https://www.mdpi.com/2624-8611/3/4/52.

M. P. Wellman. Fundamental Concepts of Qualitative Probabilistic Networks. *Artificial Intelligence*, 44 (3):257–303, 1990.