

PROBABILISTIC GRAPHICAL MODELS APPLICATIONS IN MEDICAL RE-SEARCHES

Emerson Assis de Carvalho

Department of Computing - IFSULDEMINAS Campus Machado
emerson.carvalho@ifsuldeminas.edu.br
<https://orcid.org/0000-0001-5028-2243>

Igor Duarte Rodrigues

Institute of Computer Science - Federal University of Viçosa (UFV)
igordrodrigues@gmail.com
<https://orcid.org/0000-0002-3076-8657>

Fábio Junior Alves

Department of Computing - IFSULDEMINAS Campus Machado
fabio.alves@ifsuldeminas.edu.br
<https://orcid.org/0000-0002-0535-7741>

Guilherme Sousa Bastos

Institute of Systems Engineering and Information Technology, Federal University of Itajubá (UNIFEI)
sousa@unifei.edu.br
<https://orcid.org/0000-0001-8280-2813>

Recebido em: 10/11/2023

Aprovado em: 22/02/2024



DOI: 10.18406/2359-1269v11n12024367

Abstract

The use of intelligent systems in medical research and practice has a long history, becoming an essential support tool nowadays. In this paper, we investigated and presented an overview of the Probabilistic Graphical Models, as well as some applications of them in medical research. We explored these approaches because they are widely used for inference in environments of uncertainty, especially for inter-causal reasoning. Such techniques are established as safe and complete inference mechanisms, which motivated our explanation and analysis. The methods investigated are promising for health research. They can be applied to different types of inference problems, especially causal inference, which is very common in the case of diseases, syndromes, and disorders.

Keywords: Artificial intelligence, probabilistic graphical models, health diagnosis, medical research.

1. Introduction

Science and technology are omnipresent in our everyday lives, becoming the new basis for belief and bringing new ways to improve our quality of life (FEENBERG, 2006). Artificial Intelligence (AI) is the study of intelligent behaviors. Its primary goal is a theory of intelligence that explains the behavior of natural intelligent entities and guides the creation of artificial agents capable of intelligent behaviors, such as prediction problems and decision support (RUSSELL; NORVIG, 2020).

Prediction problems usually involve reasoning with uncertainty. Reasoning under uncertainty has a long history and is a significant issue in AI. Many problems require solutions for a better decision-making process. Such solutions may seek the application of probabilistic methods to construct inference models (GENESERETH; NILSSON, 2012). However, most real-world events are unpredictable, demanding that intelligent applications need to handle partial observability, non-determinism, or any eventuality (RUSSELL; NORVIG, 2020). The uncertainty arises from some information deficiency. Thus, the information may be incomplete, imprecise, incorrect, or contradictory (KLIR, 2006).

Probabilistic methods may involve, for example, Probabilistic Graphical Models (PGMs) such as Bayesian Networks (BNs) and Hidden Markov Models (HMMs), which are among the best methods for reasoning about uncertainty (NEIL; FENTON; NIELSON, 2000). These probabilistic networks allow inter-causal reasoning, a vital aspect that distinguishes them from other inference systems (KJAERULFF; MADSEN, 2013). In inter-causal reasoning, taking evidence about a hypothesis decreases the belief in the competing unobserved hypotheses automatically (KJAERULFF; MADSEN, 2013), establishing a safe and complete inference mechanism (PEARL, 1988).

BNs are graphical representations of causal relationships in a particular domain (HOLMES; JAIN, 2008). A BN is a directed and acyclic graph in which each node corresponds to a random variable. Directed edges connect pairs of nodes, indicating a direct influence of one node over the other. In addition, each node has a conditional probability table that quantifies the effect of its parent nodes over it. By using inference algorithms over BNs, it is possible to estimate beliefs in the context of observed pieces of evidence

(RUSSELL; NORVIG, 2020). Employing rigorous formalism and practical algorithms for probabilistic reasoning, BNs support any reasoning with causal variables, such as diagnosis, prediction, or causal explanation (RUSSELL; NORVIG, 2020; WILLIAMSON, 2002).

HMMs are a double stochastic process, with a non-visible stochastic process (not observable) that can be observed/predicted through another stochastic process that produces the sequence of observations. The hidden processes are a set of states connected by transitions with probabilities. In contrast, the observable (non-hidden) processes are outputs or observable states, each one emitted by each not observable state according to some output of a probability density function. HMMs allow the designing of systems to predict a sequence of related states through a sequence of observations (RABINER, 1989).

Also used for modeling several different problems in medical researches (KROGH; MIAN; HAUSSLER, 1994; MEYER; DURBIN, 2002; TESTA et al., 2015), HMMs have been applied in several different areas of AI, such as Computer Vision (GHAHRAMANI, 2001), Robotics (BERG et al., 2018), Speech and Face Recognition (MUSTAFA; ALLEN; APPIAH, 2019; RAHUL et al., 2019), and Computational Biology (TAMPOSI et al., 2019).

The causal nature of health diseases and the possibility of structuring probabilistic networks as trees motivate the PGMs as a reasonable alternative for medical research. Considering the ability of PGMs to build transparent and efficient inference models in inter-causal domains, this work aims to explain and investigate the use of PGMs (BNs and HMMs) in medical research.

Section 2 provides an overview of AI with emphasis on probabilistic models applied to reasoning under environments of uncertainty. It introduces some necessary concepts of probability theory and presents two of the main probabilistic graphical approaches, highlighting their structure, syntax, semantics, and learning and inference methods. Sections 3 and 4 present works related to AI and PGMs applied to medical research, while Section 5 discusses and concludes this work.

2. Artificial Intelligence

AI concerns designing intelligent computer systems that exhibit characteristics we associate with intelligence (learning, reasoning, understanding language, etc). Most AI comprises the following sub-fields: Machine Learning, Natural Language Processing, Computer Vision, Robotics, and Planning. These subdivisions relate more to the sub-fields practical goals than the technologies each employs (RUSSELL; NORVIG, 2020). For example, Artificial Neural Networks is a Machine Learning technique commonly used in Natural Language Processing, Robotics, Computer Vision, and Planning.

2.1. AI Sub-fields

Machine Learning techniques enable machines to learn by observing data and creating models based on such information. Computers use these models as both a hypothesis about the problem and software capable of solving them. There are three main types of learning: 1) supervised learning, in which the computer observes pairs of input and output data to learn a function that maps from inputs to outputs; 2) unsupervised learning, in

which the computer uses the inputted data to learn patterns but does not receive explicit feedback; and 3) reinforcement learning, in which the computer learns through a set of reinforcements that can be rewards or punishments.

Natural Language Processing techniques enable machines to communicate successfully in natural languages, such as English or Portuguese. However, as natural languages are different from formal languages, a common problem is the language models, which are models to predict the probability distribution of the language expressions.

Intelligent agents can use several sensors to sense the environment (e.g., images, noises, distances, positions, temperatures). Through this perceptual channel, machines receive stimulus and create a representation of the real world. Computer Vision techniques enable machines to perceive objects from the environment through the use of sensors like cameras. Based on external information acquired by sensors (such as images), Computer Vision agents can build a real-world model, known as reconstruction (e.g., creating geometries), or describe distinctions among the objects they “see” (e.g., labeling objects), known as recognition.

Robotics are techniques that enable machines to move and manipulate the physical world. Such machines (robots) usually are equipped with actuators (e.g., arms, grippers, legs, wheels) designed to produce physical forces on the environment and sensors (e.g., gyroscopes, accelerometers, GPS, cameras, radars, lasers) dedicated to perceiving the environment. This robot-environment interaction can change the state of the robot, the state of the environment, and the state of the people around it.

Planning techniques allow finding a sequence of actions to achieve a goal. Given an initial state, a goal (or a set of them), and a set of possible actions, the planning problem synthesizes a sequence of steps to be executed in the initial state to turn the environment into a goal state. Planning has applications in several areas such as Games, Logistics, Robotics, Manufacturing, etc.

In general, real-world applications still require other abilities from so-called intelligent agents, such as automated reasoning and knowledge representation. Automated reasoning refers to performing reasoning sequences electronically and automatically finding suitable reasoning steps to infer new knowledge from a given data, answer questions, and outline new conclusions. Knowledge representation refers to how to represent real-world events (what machines know, hear, or see) in a pattern that machines can use to reason and solve problems (RUSSELL; NORVIG, 2020).

2.2. Probabilistic Graphical Models

Any technique that allows the computer to imitate human behaviors is popularly classified as AI. However, we also have AI techniques based on other biological systems, probability, statistics, and mathematics. Emerging in the 1950s, AI is a relatively new science and engineering field that uses many spheres of human knowledge, such as logic, probability, and mathematics. Its primary goal is a theory of intelligence that explains the behavior of natural intelligent entities and guides the creation of artificial agents capable of smart behaviors. An intelligent agent must be capable of precisely perceive the environment and perform proper actions. An agent is rational if it does the right thing, given its acquired knowledge (RUSSELL; NORVIG, 2020; GENESERETH; NILSSON, 2012).

In the AI context, we can define intelligence as human or rational. Human intelligence is committed to human performance, while rational is a formal and ideal performance measure called rationality. Different techniques are used to pursue these dimensions. Human intelligence approaches usually use empirical science related to psychology, cognitive science, biology, and neuroscience. These approaches involve observations, hypotheses, and the study of how humans behave, how our minds operate, and how human brains process information. Rationalist approaches consist of a combination of formalism from logic, mathematics, statistics, and control theory. These methods aim to create more strict rules for the decision process (RUSSELL; NORVIG, 2020).

2.3. Probabilistic Networks

Most real-world events are unpredictable, demanding that intelligent applications handle the uncertainty from partial observability, nondeterminism, or any eventuality (RUSSELL; NORVIG, 2020). The uncertainty arises from some information deficiency. So the information may be incomplete, imprecise, incorrect, or contradictory (KLIR, 2006).

Deductive Logic is insufficient for reasoning under uncertain environments once it does not attribute a degree of uncertainty to the premises nor conclusions. Then, the Inductive Logic, supported by the Probability Theory, has emerged as a proper alternative for expressing reasoning (WILLIAMSON, 2002), once the nature of the knowledge from which inferences are produced is uncertain and subjective (PEARL, 1986).

The probability theory provides ways to deal with the uncertainty coming from laziness and ignorance. Laziness is due to the extensive work to consider every possible explanation for given evidence. Ignorance is due to a non-complete knowledge about the domain or uncertainty about a particular situation once we can not evaluate all premises. Address uncertainty with numeric degrees of belief solves the qualification problem, which specifies the impossibility of identifying all preconditions needed to succeed in the desired action (RUSSELL; NORVIG, 2020).

Then emerged the PGMs, a graph-based representation for compactly encoding a complex distribution over a high dimensional space. Nodes express variables, and edges denote the interactions between them. Known as probabilistic networks, these models allow inter-causal reasoning, a vital aspect that distinguishes them from other automated inference systems (KOLLER; FRIEDMAN, 2009; KJAERULFF; MADSEN, 2013). In the inter-causal reasoning process, taking evidence about a hypothesis decreases the belief in the competing unobserved hypotheses automatically (KJAERULFF; MADSEN, 2013), which constitutes a safe and complete inference mechanism (PEARL, 1988).

There are two graphical families to represent probability distributions. The Bayesian Networks (BNs) (Section 2.6), and the Markov Models (MMs) (Section 2.7). Both models provide the duality of independencies and factorization. However, they differ regarding the set of independencies they can encode and the factorization of the distribution they induce (KOLLER; FRIEDMAN, 2009).

2.4. Basic Probability Theory

Before introducing the graphical models, this subsection aims to introduce some fundamental probability theories suited to the requirements of the probabilistic networks.

A **sample space** is the set of all possible worlds in a specified domain. The possible worlds are mutually exclusive, once two or more possible worlds can not be the case simultaneously. The possible worlds are also exhaustive because one possible world must be the case. For example, the throw of two distinguishable dices has 36 possible worlds $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \dots, (6, 6)\}$. A fully specified sample space associates a probability P (with values between 0 and 1) to each possible world and the total probability of all possible worlds must add up to 1. These associations are called **probability distribution** (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009).

Probability theory names its variables with the first letter in uppercase, and such variables are called aleatory or random variables. An **aleatory variable** is a numerical function defined in a sample space, and it maps from all possible worlds to a set of possible values it can assume. It gives a numerical value X to a phenomenon within the sample space S and is associated with a probability distribution ($P(X)$ in S) such that: Equation 1 (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009):

$$\forall x P(X=x) \geq 0 \text{ and } \sum_x P(X=x) = 1 \quad (1)$$

We can take the sum value obtained from two dices throw as an aleatory variable. Function 2 is the numerical function (x and y are the values obtained from each dice), and Table 1 is the probability distribution for each possible sum.

$$F(x, y) = X = x + y \quad (2)$$

Table 1. Probability distribution of Function 2

X	$P(X)$	X	$P(X)$
2	$\frac{1}{36}$	8	$\frac{1}{7.2}$
3	$\frac{1}{18}$	9	$\frac{1}{9}$
4	$\frac{1}{12}$	10	$\frac{1}{12}$
5	$\frac{1}{9}$	11	$\frac{1}{18}$
6	$\frac{1}{7.2}$	12	$\frac{1}{36}$
7	$\frac{1}{6}$		

A variable can be discrete or continuous. **Discrete** random variables have a finite set of possible values, usually obtained by counting. **Continuous** random variables take any value in a given interval of real numbers, usually obtained by measuring (KJAERULFF; MADSEN, 2013). From this point on, we deal with the discrete variables aspects once they are the type of variables this study suggests.

Dealing with probability distributions of multiple variables requires a special notation. Suppose a domain with the aleatory variables Weather $W =$

$\{sunny, rain, snow\}$ and Traffic $T = \{jam, normal\}$, with probability distributions $P(W) = \{0.6, 0.25, 0.15\}$ and $P(T) = \{0.42, 0.58\}$, respectively. The probabilities of all combinations of the values of W and T produce a matrix $M_{3 \times 2}$ called the **joint probability distribution** of W and T . **Joint probabilities** measure the likelihood of two or more events co-occurring at the same point in time. It can be represented as the probability of the intersection of the co-occurring events. A **full joint probability distribution** determines the distribution for all aleatory variables completely. This full joint distribution is sufficient as a knowledge base to calculate the probability of any possible event in the model (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009). Table 2 shows the joint probability distribution of W and T .

Table 2. Full joint probability distribution of W and T

W	T	
	<i>jam</i>	<i>normal</i>
<i>sunny</i>	0.10	0.50
<i>rain</i>	0.20	0.05
<i>snow</i>	0.12	0.03

A particular subset of possible worlds is called **events**. An example of an event is the list of worlds where two rolled dice add up to 3 $\{(1, 2), (2, 1)\}$. Probabilistic assertions and queries are usually about pre-defined events. The sum of the probabilities associated with each world of an event defines the **event probability**. For example, the probability that two rolled dice add up to 3 is $P(\text{Sum}=3) = P((1, 2)) + P((2, 1)) = 1/36 + 1/36 = 1/18$ (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009). Equation 3 shows the probability for any event E .

$$P(E) = \sum_{w \in E} P(w) \quad (3)$$

We could calculate the probabilities of some events in Table 2. To do so, we must apply Equation 3 to add up the probability values where the desired event is true. For example, the probability of *rain* ($P(W=rain) = 0.20 + 0.05 = 0.25$), and the probability of *snow* and traffic *jam* ($P(W=snow, T=jam) = 0.12$).

Probabilities like $P(W=rain)$ are called **unconditional** or **prior probabilities** once probability theory does not require comprehensive knowledge of the sample space. **Prior probabilities** refer to the belief in the events in the absence of any other information (RUSSELL; NORVIG, 2020).

Most of the time, it is necessary to know the probability of an event given that we have some information already revealed, usually called **evidence**. This probability is called **conditional** or **posterior probability** and is written as $P(X|Y)$ (RUSSELL; NORVIG, 2020). Equation 4 shows how to compute the conditional probability of event X given an event Y .

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} \quad (4)$$

For example, the probability of a traffic jam given that it is raining is:

$$\begin{aligned} P(T=jam|W=rain) &= \frac{P(T=jam, W=rain)}{P(W=rain)} \\ P(T=jam|W=rain) &= 0.2/0.25 = 0.8 \end{aligned} \quad (5)$$

Given a conditional probability, the joint distribution of X and Y can be written following the product rule (RUSSELL; NORVIG, 2020), as in Equation 6.

$$P(X, Y) = P(X|Y) P(Y) \quad (6)$$

We can apply the product rule to compute the joint probability of n variables by a successive product of conditional and joint probabilities of these same variables. Each subsequent product reduces the joint probability to a conditional probability and a shorter joint probability. Equation 7 presents the **chain rule**.

$$\begin{aligned} P(X_1, \dots, X_n) &= P(X_1|X_2, \dots, X_n) P(X_2, \dots, X_n) \\ &= P(X_1|X_2, \dots, X_n) P(X_2|X_3, \dots, X_n) \\ &\quad P(X_3, \dots, X_n) \dots \\ &= P(X_1|X_2, \dots, X_n) P(X_2|X_3, \dots, X_n) P(X_3, \dots, \\ &\quad \dots, X_n) \dots P(X_{n-1}|X_n) P(X_n) \\ &= \left[\prod_{i=1}^{n-1} P(X_i|X_{i+1}, \dots, X_n) \right] P(X_n) \end{aligned} \quad (7)$$

Given observed evidence, to compute the posterior probability for query propositions is known as **probabilistic inference**. We can use the full joint distribution to perform inference. Given the full joint distribution of a model, Equation 8 can be used to answer queries.

$$P(X|E=e) = \alpha P(X, E=e) = \alpha \sum_y P(X, E=e, Y=y) \quad (8)$$

Where:

- $P(X|E=e)$ is what we want to know (query);
- e is the list of observed values;
- y is all possible combinations of the values of the remaining unobserved variables;
- α is a normalization constant.

X , E , and Y are the entire domain set of aleatory variables. $P(X|E=e, Y=y)$ is a subset of the full joint distribution probabilities. The full joint distribution in tabular form does not scale well. However, it is the theoretical foundation to build effective reasoning systems (RUSSELL; NORVIG, 2020).

There is a fundamental property between events known as **independence** (also known as marginal independence or absolute independence). If an event X does not influence an event Y and vice-versa, they are independent events (XY). This independence means that the occurrence of X does not affect the probability of occurrence of Y and

vice-versa. The independence between two events (X and Y) can be written as in Equation 9.

$$\begin{aligned} P(X, Y) &= P(X) P(Y) \text{ or} \\ P(X|Y) &= P(X) \text{ or} \\ P(Y|X) &= P(Y) \end{aligned} \quad (9)$$

The knowledge of the domain supports performing assertions over independent events. Suppose we can split the aleatory variables into independent subsets. In that case, we can factor the full joint distribution into separate joint distributions, which reduces the size of the domain representation and the complexity of the inference model. However, it is difficult to identify independent variables in complex domains once independence will fail if a connection, even indirect, exists between two variables (RUSSELL; NORVIG, 2020).

Although independence is a valuable property, it is difficult to identify fully independent events in real-world domains. It is most common to determine the independence of two events, given a third event. Known as **conditional independence**, this relationship of three variables defines the independence of two variables X and Y , given a third variable Z ($XY|Z$), as in Equation 10.

$$\begin{aligned} P(X, Y|Z) &= P(X|Z) P(Y|Z) \text{ or} \\ P(X|Y, Z) &= P(X|Z) \text{ or} \\ P(Y|X, Z) &= P(Y|Z) \end{aligned} \quad (10)$$

As for absolute independence, conditional independence assertions also allow the decomposition of the full joint distribution. Once conditional independence is more commonly available, it can enable probabilistic systems to scale up. This decomposition of large probabilistic domains into weakly connected subsets makes conditional independence one of the most basic and robust structures of knowledge representation in uncertainty environments (KOLLER; FRIEDMAN, 2009; RUSSELL; NORVIG, 2020).

2.5. Variables Identification

The aleatory variables constitute the entire model basis. Selecting variables is one of the most pervasive selection problems in statistical applications. The problem is the uncertainty about which set of variables should establish the relationship between a variable of interest and a subset of potential explanatory or predictor variables (GEORGE, 2000). Domain experts are usually those who perform the selection of the aleatory variables.

There are some fundamental approaches to address this problem, such as the **clarity test** proposed by Kjaerulff e Madsen (2013). According to the clarity test, a variable A must meet three principles to probe whether it has been clearly defined:

- All possible values in the A domain must be exhaustive and mutually exclusive. If the possible values of A are not mutually exclusive, they should be split into several variables;
- Usually, A should represent a unique set of events with no competing variables. That is, the state of A should not be given deterministically by the state of another variable;
- A must be clearly defined, leaving no ambiguity concerning its semantics.

Kjaerulff e Madsen (2013) also recommends that it is essential to understand the **types of variables** that may arise. The identification and classification of the variables make it easier to connect them.

- **Problem or hypothesis variables** are the variables of interest from which we may want to calculate the posterior probability given some evidence (information variables). Usually non-observable, these variables relate to the diagnoses or predictions to be made;
- **Information variables** are usually the observable variables that usually have relevant information to the problem-solving. The author separates these variables into the background and symptom variables:
 - **Background variables** usually are among the network root variables and represent the information available before a problem occurs, holding a causal influence over both the problem variables and the symptom variables;
 - **Symptom variables** are the consequence variables usually available after the occurrence of a problem. These variables are children of the problem variables or background variables.
- **Mediating variables** are usually non-observable. Their posterior probability is not of immediate interest, but they help maintain the essential network independence relationships. They tend to be parents of symptom variables and children of problem variables and background variables.

2.6. Bayesian Networks

Bayesian Networks (also known as Causal Networks, Belief Networks, Causal Probabilistic Networks, Probabilistic Cause-Effect Models, Probabilistic Influence Diagrams, and Graphical Probability Networks) are graphical models of causal relationships in a given domain. Describing dependencies among variables, BNs enable solving logical problems that involve probabilistic concepts, expanding the initial models of knowledge representation and manipulation (HOLMES; JAIN, 2008; NEIL; FENTON; NIELSON, 2000).

Essentially, BNs can represent, concisely, any full joint probability distribution. By employing a rigorous and efficient formalism to uncertain knowledge structuring as well as practical algorithms for probabilistic reasoning, BNs support any reasoning with causal variables, such as diagnosis, prediction, or causal explanation (RUSSELL; NORVIG, 2020; WILLIAMSON, 2002).

BNs are models for knowledge representation consisting of two components: a **qualitative component**, representing the network structure as a Directed Acyclic Graph (DAG), and a **quantitative component**, representing the probabilistic element as a set of conditional probabilities. Both components are fundamental to the definition, construction, and underlying inference process (KJAERULFF; MADSEN, 2013; DARWICHE, 2008).

Figure 1 shows the structure and the Conditional Probability Tables (CPTs) of a BN representing part of the stock exchange domain. The *IR* variable represents the country's interest rate. The interest rate directly impacts the stock market (*SM*) performance. The stock market performance usually indicates how the country's gross domestic product

(GDP) will perform. In addition to internal factors, the state of the stock market (SM) and the performance of the company's economic sector (CS) also impact the stock price (SP) of a particular company.

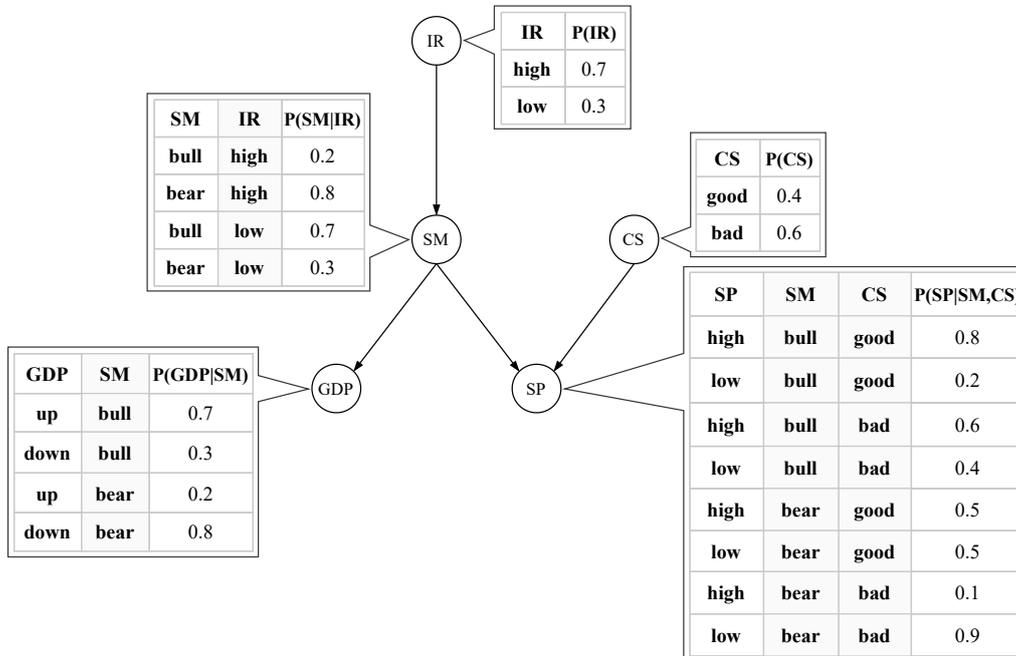


Figure 1. A BN example over five variables. A CPT is associated with each node containing the conditional probabilities of that node given its parents.

2.6.1. Syntax of Bayesian Networks

BNs represent its **qualitative** aspect using graphs that illustrate their probabilistic distributions. A **graph** $G = (V, E)$ consists of a finite set of distinct **vertices** (or **nodes**) $V = \{v_1, v_2, \dots, v_N\}$, and a finite set of **edges** (or **links**) $A = \{a_1, a_2, \dots, a_{N^2}\}$ connecting its vertices (ROSEN, 2017).

The connection pattern between nodes delimits some properties of a graph. The notation $v_1 \rightarrow v_2$ indicates a connection from vertice v_1 to vertice v_2 by a directed edge, which means a **directed graph** (or **digraph**). The notation $v_1 - v_2$ designates a connection from v_1 to v_2 by a not directed edge, which means an **undirected graph**. In a digraph, the edges are unidirectional, indicating that the graph can be traversed only in such directions. On the other hand, in an undirected graph, the edges are bidirectional, indicating that the graph can be traversed in either direction (RUSSELL; NORVIG, 2020; ROSEN, 2017).

A graph is **connected** if there is a path between every pair of its vertices. A directed graph is **acyclic** if any path following the directions of the edges will never produce a closed-loop (cycles). In a **directed multiply connected** graph, there is more than one distinct path between two nodes. There is at most one path between any two nodes in a **directed singly connected** graph (**trees**). In **simple trees**, each node has at most one parent. In **polytrees**, nodes can have more than one parent (RUSSELL; NORVIG, 2020;

ROSEN, 2017). Figure 2 illustrates these types of graphs.

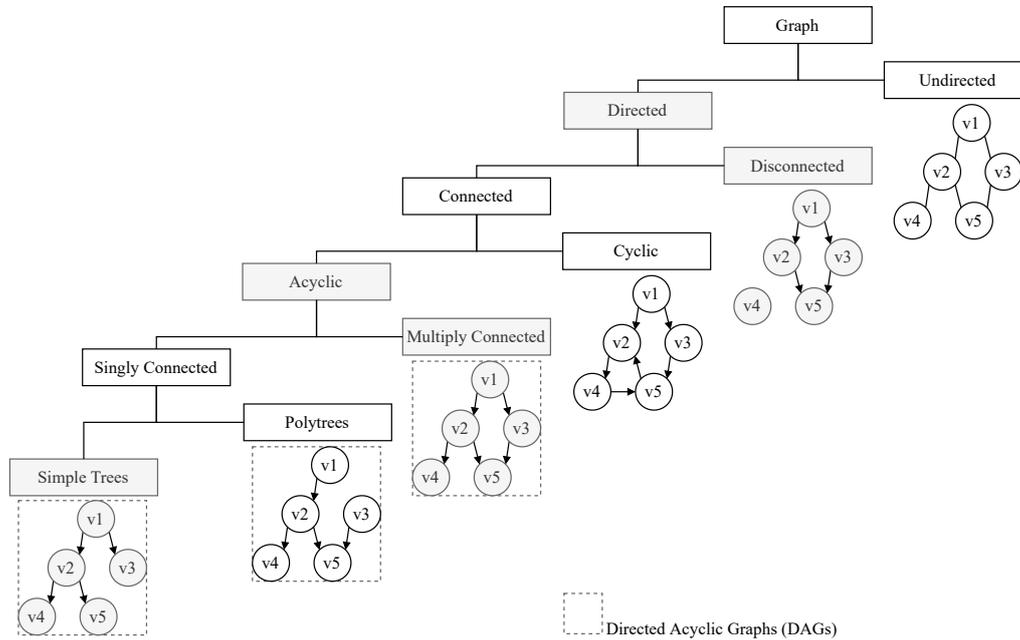


Figure 2. Types of graphs

Directed Acyclic Graphs (DAGs) represent the qualitative aspect of the BNs graphically. Concerning the DAG that represents a BN, vertices represent the aleatory variables, which correspond to the knowledge domain concepts. The directed edges of a DAG represent, in most cases, a dependency relation between the vertices they connect. Thus, the relation $v_1 \rightarrow v_2$ represents a direct dependence of variable v_2 with regard to variable v_1 , meaning typically that v_1 has a direct influence on v_2 .

Some authors point out that the dependency relation is not necessarily a **cause-effect** relationship and could be just some type of association (SCUTARI; DENIS, 2014). Other authors argue the causal relationship, assuming that the dependency relation is a cause-effect relationship (KJAERULFF; MADSEN, 2013). Based on probabilistic properties, other authors argue in favor of both points of view. They argue that the direction of the edges does not need to have a specific meaning. Although they agree that meaningful BNs express cause-effect relationships, once they correspond to more sparse and natural graphs, resulting in a more transparent and significant interpretation (PEARL, 2009; KOLLER; FRIEDMAN, 2009). Bayesian models in which the directed edges represent a causal effect are called **causal models**.

2.6.2. Dependencies and Independencies in Graphs

Dependencies and independencies are crucial for understanding BNs behavior and answering queries once the inference model estimates the probability of unobserved variables through other variables whose state has been observed (NIELSEN; JENSEN, 2009).

There is a direct dependency between X and Y if a directed edge exists between X and Y . Thus, X and Y are correlated regardless of evidence about any other variable.

Given two not directly linked variables, X and Y , a third variable, Z , in the middle of the undirected path determines conditional independence between X and Y . A vertex Z connecting X and Y specifies an indirect dependency between X and Y (NIELSEN; JENSEN, 2009).

The topology of a BN encodes mainly the conditional independence of the model. Figure 3 illustrates the four cases where the vertex Z connects X and Y : a) indirect causal effect; b) indirect evidential effect; c) common cause; and d) common effect.

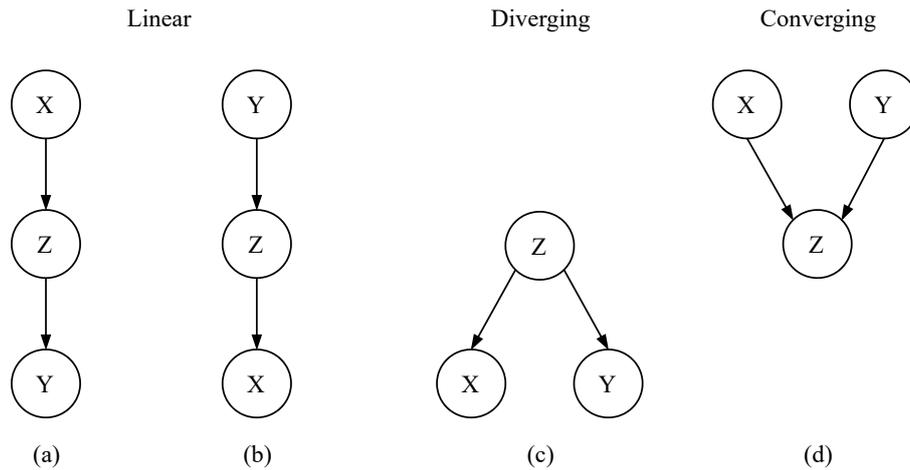


Figure 3. D-connection types between vertices/variables.

Evidence can be forwarded through the variables of **linear** (serial) connections unless the state of a variable in the middle is known. In the linear connection shown in Figure 3a, if the state of Z is known the cause X can not influence the effect Y . In the linear connection shown in Figure 3b, if the state of Z is known the effect X can not evidence the cause Y (or the effect X can evidence the cause Y only if Z is unknown) (NIELSEN; JENSEN, 2009; CHARNIAK, 1991).

Evidence can pass between all children of a parent variable (vertex) Z in **diverging** connections unless the state of Z is known. In the diverging connection shown in Figure 3c, evidence can pass between X and Y unless the state of Z is known (X is correlated with Y if and only if Z is not known) (NIELSEN; JENSEN, 2009; CHARNIAK, 1991).

It is impossible to infer anything about the parents of a variable Z in **converging** connections unless something is known about Z or its descendants. In the converging connection presented in Figure 3d, if something is known about Z or its descendants, evidence in X can tell us something about Y and vice-versa (NIELSEN; JENSEN, 2009).

In linear and diverging connections, X and Y are independent only if the state of Z is known. Thus X and Y are **d-separated** given Z (**d** connotes “directional”). In converging connections, X and Y are independent and d-separated only if the state of Z or any of its descendants are not known (NIELSEN; JENSEN, 2009; CHARNIAK, 1991).

There are still other general conditional independence properties. As shown in Figure 4, the **Markov condition** states that a variable X is conditionally independent of its non-descendants (ND_1 and ND_2), given its parents (P_1 and P_2).

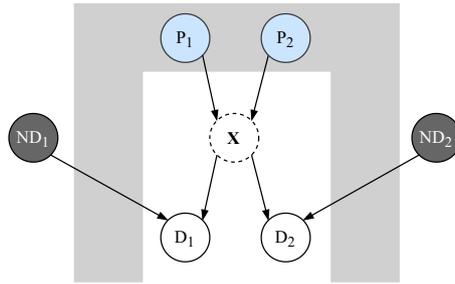


Figure 4. Conditional independence of non-descendants.

As shown in Figure 5, a variable X is conditionally independent of all other variables in the BN given its **Markov blanket**. The Markov blanket of a variable X is the set composed of its parents, children, and children’s parents. Markov blankets follow the d-separation property since the Markov blanket of a variable d-separates it from all other variables.

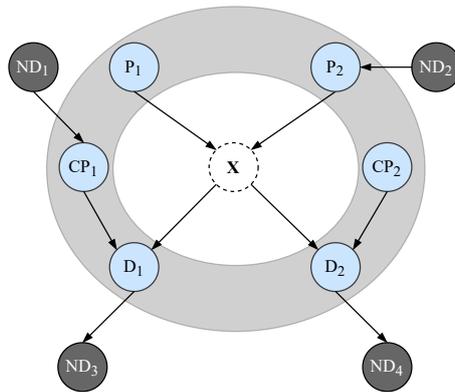


Figure 5. Conditional independence given the Markov blanket of a variable.

Grays areas in Figure 4 and 5 represent evidence, these areas “block” probability propagation (RUSSELL; NORVIG, 2020).

2.6.3. Semantics of Bayesian Networks

Sucintly, BNs are DAGs in which each vertice corresponds to an aleatory variable. Directed edges connecting pairs of vertices indicate a direct influence of one vertice (parent) over another (child). The **qualitative aspect** of the BNs specifies the correspondence between their syntax with the joint probability distribution over the BN variables. Once the topology of a BN was specified, a conditional distribution must be computed as the local probability for each variable given its parents. As a probabilistic model, each vertice has a CPT that quantifies the effects of its parents on it. The topology and the local probability define the full joint distribution for all variables of a BN (RUSSELL; NORVIG, 2020).

As mentioned before, the full joint probability distribution of a domain will increase as the number of its variables grows. However, given the topology of a BN, only the conditional probabilities for the vertices involved in direct dependencies are required,

which means the probability for every node given all possible combinations of its parents. A complete example of a BN (topology and CPTs) can be seen in Figure 1.

The edges in a BN specify the independence assumptions that must hold between the random variables. These assumptions determine what probability information is required to specify the probability distribution among the network's random variables. Each node X_i has an associated probability $P(X_i|parents(X_i))$ that quantifies the effect of its parents on it (NIELSEN; JENSEN, 2009; RUSSELL; NORVIG, 2020). Therefore, the chain rule can be reduced, as shown in Equation 11.

$$P(x_i|x_{i+1}, \dots, x_n) = P(x_i|parents(X_i)) \quad (11)$$

Suppose a BN which contains n variables $\{X_1, \dots, X_n\}$. The product of the relevant elements of the local conditional distributions represents each entry in the joint probability distribution table, as in Equation 12. Thus, BNs allow defining the joint distribution based only on their conditional probabilities, reducing the number of probability values needed substantially (NIELSEN; JENSEN, 2009; RUSSELL; NORVIG, 2020).

$$P(X_1=x_1, \dots, X_n=x_n) = \prod_{i=1}^n P(x_i|parents(X_i)) \quad (12)$$

Equation 13 presents the **Bayes theorem**, which is the probabilistic basis of the BNs. The Bayes theorem allows computing unknown probabilities from known and stable ones. This simple equation underlies all modern AI approaches for probabilistic inference by helping to simplify the intermediate calculations (RUSSELL; NORVIG, 2020).

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)} \quad (13)$$

Where:

- $P(h)$ is the prior probability of a hypothesis h ;
- $P(D)$ is the prior probability of the observed data D ;
- $P(D|h)$ is the conditional probability of D given h ; and
- $P(h|D)$ is the posterior probability of h given D . It is the belief in the model after seeing the data.

Given the topology and the conditional probabilities of a BN, it is possible to infer the probability of any variable in the network applying basically the Bayes theorem together with some new evidence. Thus, it is possible to take action or search for further evidence to increase the network's confidence (RUSSELL; NORVIG, 2020).

2.6.4. Learning in Bayesian networks

Learning is the task of estimate and select models. Usually, the topology and the probabilities required to define a BN are given by specialists, preview studies, or obtained with experiments and calculus. It is also possible to reach the topology and the statistical information using methods that extract them from the data available (CHARNIAK, 1991; HECKERMAN, 2008; HECKERMAN; GEIGER; CHICKERING, 1995).

There are different learning approaches such as **manual** methods supported by the experience of domain experts, known as supervised learning; **(semi-)automatic** methods that learn from data, known as unsupervised learning; or a combination of both approaches, which combines observed data with experts' experience (KJAERULFF; MADSEN, 2013).

The manual construction of a BN is usually a challenging task. It requires distinct expertise such as model engineering abilities and a comprehensive understanding of the problem domain. The model elicitation process requires: 1) a solid problem definition; 2) a careful identification of the relevant variables; 3) a precise definition of dependences/independences relationships among the chosen variables; and 4) a proper elicitation of many conditional and prior probabilities (KJAERULFF; MADSEN, 2013).

As the parameters of a BN are determined by its structure, creating a BN always proceeds in three consecutive stages. The first step refers to the selection of the variables of interest. The second step refers to identifying the causal, functional or informational relations among the variables to construct the network structure (DAG). The last step refers to estimating the set of conditional and prior probabilities for all network nodes.

2.6.5. Developing the BN Structure

The network structure development defines the dependency relationship between the selected variables. There are two main creational approaches: a **basic approach** based on the natural causal ordering among the previously mentioned types of variables and the **Neil method** proposed by Neil, Fenton e Nielson (2000).

The **basic approach** maintains a causal perspective in the model construction, once this causality is crucial to construct influence diagrams. Such a causal approach may lead to a more suitable representation of the dependence and independence relations and a more reliable estimate of the conditional probabilities (KJAERULFF; MADSEN, 2013).

Thus, the next step in a BN construction process involves identifying and verifying causal links among the selected variables. According to the types of variables, Figure 6 gives an overall view of the causal dependence relations of a BN.

The **Neil method** creates the network structure based on five commonly occurring substructures. These substructures are known as idioms, and their semantics and syntax represent different methods of uncertain reasoning, covering the vast majority of substructures that can occur in a BN. As described by Neil, Fenton e Nielson (2000), the five idioms are:

- **Definitional or synthesis** integrates many variables into a single variable aiming to organize the BN;
- **Cause/consequence** models cause-effect mechanisms;
- **Measurement** models the uncertainty associated with the accuracy of a measurement instrument;
- **Induction** models inductive reasoning based on populations of similar or exchangeable members;

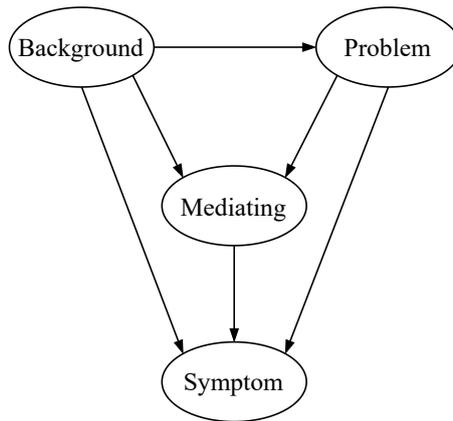


Figure 6. Typical overall causal structure of a BN. Adapted from Kjaerulff e Madsen (2013).

- **Reconciliation** models the reconciliation of results from competing measurement or prediction systems.

According to the types of variables defined in Section 2.5, the variables classification depends on their position in the DAG structure. Which of the idioms to chose depends on how we perceive the relationships among the variables. However, the cause/-consequence idiom is the most frequently used substructure. Thus, considering that the relations among the subset of variables are best described using one or more cause/consequence relations is a good starting point (KJAERULFF; MADSEN, 2013). Neil, Fenton e Nielson (2000) present a guide to choosing the proper idiom.

2.6.6. Inference in Bayesian Networks

BNs answer questions concerning the nature of their data through the use of partial **queries**. These queries are performed through techniques known as inference, probabilistic reasoning or belief updating. Given a BN B with n variables $\{X_1, \dots, X_n\}$, a partial question $Q = \{B, A, E\}$ consists of computing the conditional probability $P(A|E=e)$ where:

- A is a target set of non-observed variables;
- The evidence $E=e$ is a set of k observed variables $E = (E_1=e_1, \dots, E_k=e_k)$;
- Variables in X not included in A nor E constitute the set of hidden variables H .

Evidence may combine multiple and not always perfect sources of information. Thus, the observation can be uncertain and imprecise, which generates what is known as uncertain evidence. Therefore, there are different types of evidence, such as hard evidence and probabilistic evidence (virtual evidence and soft evidence) (MRAD et al., 2015).

The classic notion of evidence is hard or regular evidence that precisely specifies the state of a random variable. It is an observation that a variable A definitely has a particular value (e.g., $A=1$) (PEARL, 1988).

Virtual evidence, also known as likelihood evidence, corresponds to the cases where the observation is uncertain. It is usually interpreted as evidence with uncertainty

and is commonly represented as a likelihood ratio. A likelihood $P(A)$ represents virtual evidence of a variable A as in Equation 14 (PEARL, 1988).

$$P(A) = (P(a_1|a_1), \dots, P(a_n|a_n)) \quad (14)$$

Where $P(a_i|a_i)$ is the probability of observe A in the state a_i if it really is in the state a_i .

Soft evidence is usually interpreted as evidence of uncertainty and is represented as a probability distribution of one or more variables. There is uncertainty concerning the precise value of a variable A , but certainty regarding its probability distribution $P(A)$. Since $P(A)$ distribution is a certain observation, updating network belief should preserve it (VALTORTA; KIM; VOMLEL, 2002).

This preservation of the local distribution of the evidence variable is the main difference between soft evidence and virtual evidence, once virtual evidence does not require this preservation. Belief in virtual evidence is not fixed and can be modified by further evidence on other variables (MRAD et al., 2015).

There are three main categories of partial queries: Conditional Probability Query (CPQ), Maximum a Posteriori (MAP), and Most Probable Explanation (MPE) or Marginal MAPs (SCUTARI; DENIS, 2014; KOLLER; FRIEDMAN, 2009).

Koller e Friedman (2009) classify **CPQs** as:

- **Causal, deductive or predictive reasoning:** that estimates the probability of a variable given the observation of non-descending variables (from causes to effects). In the BN example in Figure 1, $P(SP|SM, CS)$ represents this type of query;
- **Evidence, abductive or explanation reasoning:** that estimates the probability of a variable given the observation of descending variables (from effects to causes). In the BN example in Figure 1, $P(IR|GDP, SP)$ represents this type of query;
- **Inter-causal reasoning:** that addresses the interaction of causing variables with regard to the same effect variable. It refers to the decrease in the belief of competing hypotheses once observed the occurrence of one or several hypotheses. In the BN example in Figure 1, $P(CS|SM, SP)$ represents this type of query.

MPAs and MPEs consist of identifying the most likely configuration for all variables in A that maximize the posterior probability of E . In MPE, A coincides with all remaining variables in the subset $\{X - E\}$. In MAP, A is a strict subset of “hypothesis” variables in $\{X - E\}$. Thus, MPEs and MAPs calculate the most probable assignment for A (a^*) in a model X given evidence $E=e$, as in Equation 15 (DARWICHE, 2008).

$$(A|E=e) = a^* = \underset{A}{\operatorname{argmax}} P(A|E) \quad (15)$$

All these inference problems are complex. The decision version of MPEs, CPQs, and MAPs are known to be NP-complete, PP-complete, and NP^{PP}-complete¹, respectively. There are exact and approximate algorithms for answering these queries. All exact

¹NP-, PP-, and NP^{PP}-complete are classifications for the complexity of common problems in computer science. These classifications usually describe the amount of computer time (elementary operations performed) and space an algorithm takes to run (OZTOK; CHOI; DARWICHE, 2016; PAPADIMITRIOU, 1994).

inference algorithms have an exponential complexity regarding the BN treewidth. Approximate inference algorithms usually are not sensitive to the BN treewidth and can be pretty efficient regardless of the BN topology. However, the approximate methods usually present issues regarding the quality of answers they compute, which is commonly related to the amount of time scheduled by the algorithm (DARWICHE, 2008). The most suitable inference algorithm will depend on the accuracy required and the computational cost.

Variable elimination is the simplest algorithm for exact inference in PGMs, and is very efficient on models whose DAG representation is a tree. Belief propagation is another algorithm to satisfy CPQs with exact inference when the DAG is a tree (RUSSELL; NORVIG, 2020; PEYRARD et al., 2019).

2.7. Markov Models

It may be necessary to model dynamic systems that allow reasoning about the state of the world as it evolves. These systems states are also represented as a set of aleatory variables, whose values at time t are a snapshot of the relevant system attributes. It is possible to model BNs representing a temporal probability model, known as Dynamic Bayesian Networks (DBNs). DBNs model stochastic processes over time intervals (RUSSELL; NORVIG, 2020; KOLLER; FRIEDMAN, 2009).

DBMs are not the first temporal method of reasoning under uncertainty. Hidden Markov Models have great popularity due to their compact representation, fast learning, and fast inference techniques (RUSSELL; NORVIG, 2020). According to Koller e Friedman (2009), the Hidden Markov Models are the simplest nontrivial type of these **state-observation temporal models**.

In probability theory, a Markov Model is a **Stochastic Process** (SP) which consists of a family of variables that evolve regarding some parameter, usually time. An SP is represented by $\{X_t \mid t \in T\}$, where:

- T is the parametric space, formed by a set of ordered values (e.g., time);
- t is a given value in T ; and
- Each X_t is an aleatory variable. The set of its possible values is called the states space, and its specific values at any given time are the process states.

In general, SPs are used to study the evolution of phenomena or systems. Given an initial condition, all system evolution is unknown, having several possible trajectories for its evolution. The SPs analysis determines the probability distributions for each set of aleatory variables, using them to predict future behaviors (states) given past behaviors (states). In contrast with deterministic models, those specified by a set of equations that describe exactly how a system evolves, the evolution of stochastic models is random, and if the process runs several times (realizations of the process), it will not give the same results (JELINEK, 1997; RABINER, 1989).

Let $\{X_0, X_1, \dots, X_t, \dots, X_T\}$ be a sequence of stochastic variables, where $(0 \leq t \leq T)$ represents a discrete time order, defined for the same discrete and finite state space. If nothing else is considered, the joint probability of these stochastic variables is given by the chain rule (JELINEK, 1997), as shown in Equation 16.

$$\begin{aligned}
 P(X_0, X_1, \dots, X_T) &= \prod_{t=0}^T P(X_t | X_0, X_1, \dots, X_{t-1}) \\
 &= P(X_0)P(X_1|X_0)P(X_2|X_0, X_1) \dots \\
 &\quad P(X_T|X_0, X_1, X_2, \dots, X_{T-1})
 \end{aligned} \tag{16}$$

An SP is taken as **Markovian** if it satisfies the property shown in Equation 17.

$$P(X_t | X_0, X_1, X_2, \dots, X_{t-1}) = P(X_t | X_{t-1}) \tag{17}$$

When dealing with a **markovian process**, the Equation 16 can be simplified, as shown in Equation 18.

$$\begin{aligned}
 P(X_0, X_1, \dots, X_T) &= \prod_{t=0}^T P(X_t | X_{t-1}) \\
 &= P(X_0)P(X_1|X_0)P(X_2|X_1)P(X_3|X_2) \dots \\
 &\quad P(X_T|X_{T-1})
 \end{aligned} \tag{18}$$

2.8. Markov Chains

Markovian processes in discrete state spaces are known as Markov Chains (MCs). An MC is a memoryless SP whose future state depends only on its current state, disregarding past states. Satisfying what is known as **Markov property**, a MC X_t is a SP where given a value of X_t , the values of X_s ($t < s$) are not influenced by the values of X_u ($u < t$). Or, more succinctly, successive steps are statistically independent (REICHL, 2016).

Grinstead e Snell (1998) made an interesting description of MCs by defining it as a set of **states** $S = \{s_1, s_2, \dots, s_r\}$ in a **process**. The process starts in one of these states and **moves** successively from one state to another. Each move is called a **step**. If the chain is in a current state s_i , then it moves to a state s_j at the next step with a probability denoted by p_{ij} , and this probability does not depend upon which states the chain was before the current state s_i . The probabilities p_{ij} are called **transition probabilities**. The process can remain in the state it is in, and this occurs with probability p_{ii} . An **initial probability distribution**, defined on S , specifies the starting state and is calculated as a vector π that indicates the initial probability of each state.

This probability distribution of the states transitions is typically represented in a **transition matrix**. If a MC has N possible states, its transition matrix will be an $N \times N$ matrix, where each entry N_{ij} is the transition probability from state i to state j . The transition matrix must be stochastic, which is a matrix where entries in each row must add up to exactly one ($\sum_{j=1}^n P_{ij} = 1$) since each row represents its probability distribution. The transition matrix probabilities can vary over time or be stationary (when its probabilities are time-independent). The π vector and the hidden states (BuM, StM, BeM) in Figure 7 illustrate an MC.

Through the transition matrix it is possible to obtain the absolute probability of the system states after a given number of transitions. The probability of a system composed by: 1) N states ($1, 2 \dots N$); 2) a transition matrix $A_{N \times N}$; and 3) an initial state vector π_0 , stay in one of its N states after k transitions is seen in equation 19.

$$\pi_k = \pi_0 (A_{N \times N})^k \tag{19}$$

Where:

- Each A_{ixj}^k position is the probability of staying in state j , since it started in state i , after k transitions; and
- π_k has the probabilities of staying in each state after k transitions when considering the initial state vector π_0 .

2.8.1. Hidden Markov Models

Most Markovian processes consist of states that can be directly observed. However, HMMs are used to model Markovian processes that generate **indirectly observable states** through the transitions between the states of the MC that govern the process, but which can not be directly observed. HMMs are double-layered SPs with a nonvisible SP that can be observed through another SP that produces the sequence of observations (RABINER, 1989).

The **hidden process** is a set of states connected by transitions with probabilities (an MC). In contrast, the **observable process** is a set of outputs or visible states emitted by each not observable state according to some output of a probability density function. The challenge is to determine the hidden states from the visible states (RABINER, 1989).

The fundamental difference between HMMs and the rest of the Markovian processes is how the system is observed. HMMs have an indirect observation of the states, carried out by inference since the observable ones are probabilistic functions regarding the states of the chain or regarding the transition between these states. In contrast, the rest of the markovian processes has direct observation, where the observable ones are the states themselves.

Most Neural Networks are probabilistic methods. They work in a discriminative approach to take inputs from a high-dimensional space and map it to a lower-dimensional space. On the other hand, HMMs are statistical methods that work in a generative approach that models conditional dependencies of hidden states. Each state has a probability distribution regarding the observations. An HMM hidden state is the entity's identity that caused each observation, and this hidden cause is translated statistically into the observed data. Through the forward-backward algorithms, it is possible to find the conditional distribution over the hidden states (CAPPÉ; MOULINES; RYDÉN, 2006; RABINER, 1989).

Described for the first time in the late 1960s and early 1970s (BAUM; PETRIE, 1966; BAUM; EAGON, 1967), HMM applications began to be used in word recognition in the middle 1970s (BAKER, 1975). HMMs appear in the literature under various names, such as Hidden Markov Processes, Markov Sources, Hidden Markov Chains, and Probabilistic Functions of Markov Chains. HMM's first applications focused on speech and handwriting recognition and DNA sequencing, reaching, later, great importance in bioinformatics.

2.8.2. Hidden Markov Models Structure

An HMM structure is characterized by:

- T : the observation sequence length;
- N : the number of distinct states in the model;
- S : a set of states. Individual states are labeled $\{1, 2, \dots, N\}$ and the state at time t as Q_t ;
- M : the number of distinct observable symbols in the model.
- V : a set of symbols. Individual symbols are denoted as $\{v_1, v_2, \dots, v_M\}$;
- $A = \{a_{ij}\}$: the transition probability distribution from state a , where: $a_{ij} = P[q_{t+1} = j | q_t = i], 1 \leq i, j \leq N$ (a_{ij} can be read as $P(\text{state } q_j \text{ at } t + 1 | \text{state } q_i \text{ at } t)$);
- B : a $N \times M$ probability distribution matrix which relates the states of the set S (rows) to the observable symbols of the set V (columns). $B = \{b_j(k)\}$ defines the observation probability distribution of symbols in the state $j, \{j_1, j_2, \dots, j_N\}$, where: $b_j(k) = P[O_t = v_k | q_t = j], 1 \leq k \leq M$. As A, B is stochastic and its probabilities $b_j(k)$ are time independent ($b_j(k)$ can be read as $P(\text{observation } k \text{ at } t | \text{state } q_j \text{ at } t)$);
- $\pi = \{\pi_i\}$: the initial state distribution, where: $\pi_i = P[q_1 = i], 1 \leq i \leq N$.

Thus, the HMM specification requires the definition of two model parameters (N and M), a symbol observation specification, and the definition of three sets of probability distribution A, B , and π . The complete set of model parameters is defined as $\lambda = (A, B, \pi)$. This set of parameters defines the measure of probability for $O, P(O|\lambda)$, where O is a set of observed states.

A different graphical notation depicts the HMMs structure. Directed (generally cyclic) graphs represent the HMMs transition/emission model, in which vertices denote the different states and edges indicate the transitions/emissions between states (KOLLER; FRIEDMAN, 2009).

Figure 7 presents the structure of an HMM that represents part of the stock exchange domain. The three hidden variables that form the hidden MC represent the stock market states, Bull Market (BuM), Bear Market (BeM), and Stagnant Market (StM). The edges between these hidden states represent the possible transitions. The values next to each edge indicate the transition probabilities between the hidden states. The two observable symbols represent two critical economic indicators: a high interest rate (HIR) and a growing gross domestic product (GGDP). The dashed edges arriving at the observable states represent the possible emissions. The values next to each dashed edge indicate the emission probabilities from hidden to observable states. This HMM would make it possible to predict the stock market direction by observing the economic indicators.

The parameters of the HMM displayed in Figure 7 are listed below:

- $N = 3$;
- $S = \{\text{BuM}, \text{StM}, \text{BeM}\}$;
- $M = 2$;
- $V = \{\text{HIR}, \text{GGDP}\}$;
- $A =$

	BuM	StM	BeM	
$\left[\begin{array}{ccc} 0.2 & 0.2 & 0.6 \\ 0.5 & 0.2 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{array} \right]$				BuM
				StM
				BeM

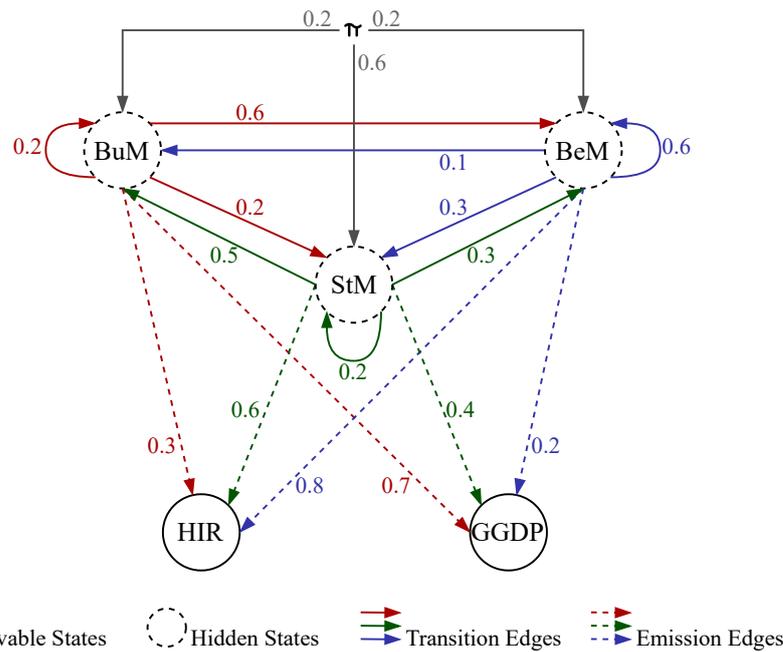


Figure 7. An HMM example over five states (three hidden/non-observable states and two observable states).

• $B =$

HIR	GGDP	
[0.3	0.7] BuM
[0.6	0.4] StM
[0.8	0.2] BeM

• $\pi =$

BuM	StM	BeM
[0.2	0.6	0.2

There are three main **problems** we can solve using HMMs:

1. **Evaluation problem:** given an observation sequence O , and a model λ , how to calculate the probability of O be produced by the model ($P(O|\lambda)$);
2. **Best sequence of states:** given an observation sequence O , and a model λ , how to calculate an optimal state sequence Q for a given sequence of observations;
3. **Training:** how to adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$.

These three problems are traditionally solved, respectively, by **Forward-backward**, **Viterbi** and **Baum-Welch** or **K-Means** algorithms (CAPPÉ; MOULINES; RYDÉN, 2006; RABINER, 1989).

3. AI Applications in Medical Researches

Several AI techniques have been applied in medical research. Machine Learning architectures are being applied in different biomedical areas, such as public and medical

health management, bio and medical imaging, developmental evaluations and interventions, and brain and body machine interface (PAIVA et al., 2022b, 2022a; QUEIROZ et al., 2020; PHELLAN et al., 2019; ZEMOURI; ZERHOUNI; RACOCEANU, 2019; LITJENS et al., 2017; LEE et al., 2017; SANTANA et al., 2016). Current and potential uses of AI in healthcare also include dermatology, ophthalmology, radiology, histopathology, and nuclear medicine (LEE et al., 2019).

Some researches involving the use of intelligent systems applied to autism propose the formulation of diagnostic methods based on magnetic resonance imaging (RODRIGUES et al., 2022; SANTANA et al., 2022; HEINSFELD et al., 2018; BHAUMIK et al., 2018; KHOSLA et al., 2018; LIAO; LU, 2018; ZHAO et al., 2018; DVORNEK; VENTOLA; DUNCAN, 2018; DEKHIL et al., 2018b, 2018a; HAZLETT et al., 2017; EMERSON et al., 2017; DVORNEK et al., 2017; YAHATA et al., 2016), however, most of them do not use any protocol for report results, such as the GSRS protocol (RODRIGUES et al., 2023), implicating in sat-backs for result comparisons.

Moreover, some research focuses on early prediction approaches from behavioral and developmental measures (BUSSU et al., 2018), the use of robots and other AI techniques applied to the therapy processes of ASD children (ALVES et al., 2020), wearable assistive technologies (BENSSASSI et al., 2018), approaches to predicting autism risk genes (BRUEGGEMAN; KOOMAR; MICHAELSON, 2020; LIN et al., 2018; LE; VAN, 2017), methods to reveal differences in regional brain structure between autistic and typical development people (GÓRRIZ et al., 2019), and modeling the diagnostic heterogeneity of autism (LOMBARDO; LAI; BARON-COHEN, 2019).

4. PGMs Applications in Medical Researches

HMMs have been used for modeling several different problems in medical researches, including finds genes in E.coli DNA (KROGH; MIAN; HAUSSLER, 1994), a model comparative ab initio prediction of gene structures (MEYER; DURBIN, 2002), and gene prediction in fungal genomes using RNA-seq transcripts (TESTA et al., 2015).

There are approaches also using HMMs to diagnose cancer (MANOGARAN et al., 2018), for genotype imputation (BROWNING; BROWNING, 2009; LI et al., 2010; MARCHINI et al., 2007; HOWIE; DONNELLY; MARCHINI, 2009; MARCHINI; HOWIE, 2010), and to investigate heart abnormalities (FAHAD et al., 2018; DWIVEDI; IMTIAZ; RODRIGUEZ-VILLEGAS, 2018; SARAÇOĞLU, 2012; CHAUHAN et al., 2008; WANG et al., 2007; UĞUZ; ARSLAN; TÜRKÖĞLU, 2007).

Regarding mental disorders and diagnosis, HMMs have been applied to evaluate the pronunciation quality and acquisition of language skills (SCHIPOR; PENTIUC; SCHIPOR, 2012; SAZ et al., 2009), to forecast a possible future diagnosis from infants with a high risk of autism (ALIE et al., 2011), to diagnose emotion-related mental diseases (GUO et al., 2017), and to recognize the stereotyped gestures which are typical of autistic people (CAMADA; CERQUEIRA; LIMA, 2017).

Carvalho et al. (2020) used HMMs, together with the data on autism heritability, to develop a model to investigate the likelihood of autistic parents generating autistic children. The model was built and validated using statistical data from the association of gender with the recurrence of autism among siblings and statistical data from the association of genetic factors with autism.

Applied in several areas, BNs are among the AI methods that have been most successful in practical applications for medicine. The most common approaches are for medical diagnoses, such as diagnosing diseases of the lymph node (HECKERMAN; HORVITZ; NATHWANI, 1992), heart disease diagnosis (SPIEGELHALTER; FRANKLIN; BULL, 2013; SAHEKI, 2005), and computerized tongue diagnosis (ZHANG; ZHANG; ZHANG, 2017).

Regarding mental disorders, Palmer, Lawson e Hohwy (2017) gathered Bayesian approaches to autism within a framework that extends from simple to complex Bayesian inference models. Given that the ASD core features relate to how individuals interact with the world around them, they propose that ASD is characterized by a greater weighting of sensory information in updating probabilistic representations of the environment. Thus, ASD may relate to finer mechanisms involved in the adjustment of sensory perception, and the hypotheses regarding atypical sensory weighting in ASD have direct implications for behavior regulation. They base their work on a theory called predictive processing, in which top-down and bottom-up messages passing across the cerebral cortex implement hierarchical probabilistic inference on the sensory stimulation causes. The hypothesis regarding ASD is that the incoming sensory signals are weighted more highly when integrated with the brain's existing model of the environment, such that neural processes like perception are dictated to a greater extent by the present sensory data rather than prior or contextual information.

Carvalho (2022) proposed BN models capable of estimating the risk of autism among the family members given some evidence. Structured like a family tree, the BNs are capable of, given some evidence, for example, the autism diagnosis of one family member, estimate the risk of ASD among other family members.

Other BN human applications include automated language (CHARNIAK; GOLDMAN, 1990) and text understanding (GOLDMAN, 1991), describing the interaction between genes (FRIEDMAN et al., 2000), control of Computer Vision systems (LEVITT; AGOSTA; BINFORD, 1990), pathology finder (HECKERMAN, 1990), genetic models (SILBERSTEIN et al., 2013), and clinical support (POURRET; NAÏM; MARCOT, 2008).

5. Discussions and Conclusions

This paper presented an overview of the AI sub-fields, with emphasis on the PGMs. We further explored these approaches because they are widely used for inference in environments of uncertainty. Moreover, an overview concerning probability theory also was necessary due to its importance to the probabilistic models. We dedicated special attention to understanding the models' fundamentals, how they work, and what they can do.

We started showing the vast dimension of the AI field by succinctly defining it and describing its main sub-fields. Then, we presented the probabilistic networks, which allow inter-causal reasoning to build inference models. We also introduced both the probability theory basics and the graphs fundamentals since these techniques underlie the development of the graphical models and the inference process.

Both Bayesian and Markovian approaches seem the most suitable methods to model the complex nature of systems with a cause-and-effect association, which is very

common in health, especially related to disease prediction and diagnosis. Thus, BNs and HMMs were sufficiently explored once these two types of methodologies can infer unknown states given a piece of evidence.

Those uses of PGMs could also be extrapolated to many applications related to bio-medicine, once many of them are mainly composed by statistical characteristics, such as genetic heritability, set of body measurements to reach diagnoses, and protein bind sites for drug effects on disease control.

For cases where it is possible to selecting variable by the identification process described in this work, and a sample representation can be used to extract statistical data is available, the PGMs are highly recommended, given the simplicity of implementation and previous efficiency on literature for similar problems.

However, as probabilistic models, the quality of Bayesian and Markov models fundamentally depends on an accurately defined set of probabilities to support the models' construction. Such data may not always be available or correct. Therefore, estimates from probabilistic systems must always be taken as directive information once nondeterministic models provide them.

References

- ALIE, D. et al. Analysis of eye gaze pattern of infants at risk of autism spectrum disorder using markov models. In: IEEE. *2011 IEEE Workshop on Applications of Computer Vision (WACV)*. [S.l.], 2011. p. 282–287.
- ALVES, F. J. et al. Applied behavior analysis for the treatment of autism: A systematic review of assistive technologies. *IEEE Access*, IEEE, v. 8, p. 118664–118672, 2020.
- BAKER, J. K. *Stochastic modeling as a means of automatic speech recognition*. [S.l.], 1975.
- BAUM, L. E.; EAGON, J. A. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, v. 73, n. 3, p. 360–363, 1967.
- BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, JSTOR, v. 37, n. 6, p. 1554–1563, 1966.
- BENSSASSI, E. M. et al. Wearable assistive technologies for autism: opportunities and challenges. *IEEE Pervasive Computing*, IEEE, v. 17, n. 2, p. 11–21, 2018.
- BERG, J. et al. Action recognition in assembly for human-robot-cooperation using hidden markov models. *Procedia CIRP*, Elsevier, v. 76, p. 205–210, 2018.
- BHAUMIK, R. et al. Predicting autism spectrum disorder using domain-adaptive cross-site evaluation. *Neuroinformatics*, Springer, v. 16, n. 2, p. 197–205, 2018.
- BROWNING, B. L.; BROWNING, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics*, Elsevier, v. 84, n. 2, p. 210–223, 2009.



- BRUEGGEMAN, L.; KOOMAR, T.; MICHAELSON, J. J. Forecasting risk gene discovery in autism with machine learning and genome-scale data. *Scientific reports*, Nature Publishing Group, v. 10, n. 1, p. 1–11, 2020.
- BUSSU, G. et al. Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: A longitudinal cross-domain classifier analysis. *Journal of autism and developmental disorders*, Springer, p. 1–16, 2018.
- CAMADA, M. Y.; CERQUEIRA, J. J.; LIMA, A. M. N. Stereotyped gesture recognition: An analysis between hmm and svm. In: IEEE. *2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA)*. [S.l.], 2017. p. 328–333.
- CAPPÉ, O.; MOULINES, E.; RYDÉN, T. *Inference in hidden Markov models*. New York, NY: Springer Science & Business Media, 2006.
- CARVALHO, E. A. *Estimating the family bias to autism: a bayesian approach*. Tese (PhD Dissertation) — Federal University of Itajubá, Itajubá, MG. Brazil, 2022.
- CARVALHO, E. A. et al. Hidden markov models to estimate the probability of having autistic children. *IEEE Access*, IEEE, v. 8, p. 99540–99551, 2020.
- CHARNIAK, E. Bayesian networks without tears. *AI magazine*, v. 12, n. 4, p. 50–50, 1991.
- CHARNIAK, E.; GOLDMAN, R. Plan recognition in stories and in life. In: *Machine Intelligence and Pattern Recognition*. [S.l.]: Elsevier, 1990. v. 10, p. 343–351.
- CHAUHAN, S. et al. A computer-aided mfcc-based hmm system for automatic auscultation. *Computers in biology and medicine*, Elsevier, v. 38, n. 2, p. 221–233, 2008.
- DARWICHE, A. Bayesian networks. In: *Handbook of knowledge representation*. 1. ed. United Kingdom: Elsevier, 2008. p. 467–499.
- DEKHIL, O. et al. Identifying personalized autism related impairments using resting functional mri and ados reports. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland, 2018. p. 240–248.
- DEKHIL, O. et al. Using resting state functional mri to build a personalized autism diagnosis system. In: IEEE. *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. [S.l.], 2018. p. 1381–1385.
- DVORNEK, N. C.; VENTOLA, P.; DUNCAN, J. S. Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks. In: IEEE. *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. [S.l.], 2018. p. 725–728.
- DVORNEK, N. C. et al. Identifying autism from resting-state fmri using long short-term memory networks. In: SPRINGER. *International Workshop on Machine Learning in Medical Imaging*. Cham, Switzerland, 2017. p. 362–370.
- DWIVEDI, A. K.; IMTIAZ, S. A.; RODRIGUEZ-VILLEGAS, E. Algorithms for automatic analysis and classification of heart sounds: a systematic review. *IEEE Access*, IEEE, v. 7, p. 8316–8345, 2018.

- EMERSON, R. W. et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science translational medicine*, American Association for the Advancement of Science, v. 9, n. 393, p. eaag2882, 2017.
- FAHAD, H. et al. Microscopic abnormality classification of cardiac murmurs using anfis and hmm. *Microscopy research and technique*, Wiley Online Library, v. 81, n. 5, p. 449–457, 2018.
- FEENBERG, A. What is philosophy of technology? In: *Defining technological literacy*. [S.l.]: Springer, 2006. p. 5–16.
- FRIEDMAN, N. et al. Using bayesian networks to analyze expression data. *Journal of computational biology*, Mary Ann Liebert, Inc., v. 7, n. 3-4, p. 601–620, 2000.
- GENESERETH, M. R.; NILSSON, N. J. *Logical foundations of artificial intelligence*. [S.l.]: Morgan Kaufmann, 2012.
- GEORGE, E. I. The variable selection problem. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 95, n. 452, p. 1304–1308, 2000.
- GHAHRAMANI, Z. An introduction to hidden markov models and bayesian networks. In: *Hidden Markov models: applications in computer vision*. Singapore: World Scientific Publishing, 2001. p. 9–41.
- GOLDMAN, R. *A Probabilistic Approach to Language Understanding*, "Department of Computer Science. [S.l.], 1991.
- GÓRRIZ, J. M. et al. A machine learning approach to reveal the neurophenotypes of autisms. *International journal of neural systems*, World Scientific, v. 29, n. 07, p. 1850058, 2019.
- GRINSTEAD, C. M.; SNELL, J. L. *Introduction to probability*. 2. ed. Rhode Island, USA: American Mathematical Society, 1998.
- GUO, K. et al. Eeg-based emotion classification using innovative features and combined svm and hmm classifier. In: IEEE. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.], 2017. p. 489–492.
- HAZLETT, H. C. et al. Early brain development in infants at high risk for autism spectrum disorder. *Nature*, Nature Publishing Group, v. 542, n. 7641, p. 348, 2017.
- HECKERMAN, D. Probabilistic similarity networks. *Networks*, Wiley Online Library, v. 20, n. 5, p. 607–636, 1990.
- HECKERMAN, D. A tutorial on learning with bayesian networks. *Innovations in Bayesian networks*, Springer, p. 33–82, 2008.
- HECKERMAN, D.; GEIGER, D.; CHICKERING, D. M. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, Springer, v. 20, n. 3, p. 197–243, 1995.
- HECKERMAN, D.; HORVITZ, E.; NATHWANI, B. *Toward Normative Expert Systems: Part I, the Pathfinder Project*. Knowledge Systems Laboratory, Medical Computer Science. [S.l.]: Stanford University, 1992.
- HEINSFELD, A. S. et al. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, Elsevier, v. 17, p. 16–23, 2018.



- HOLMES, D. E.; JAIN, L. C. Introduction to bayesian networks. In: *Innovations in Bayesian Networks*. [S.l.]: Springer, 2008. p. 1–5.
- HOWIE, B. N.; DONNELLY, P.; MARCHINI, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*, Public Library of Science, v. 5, n. 6, p. e1000529, 2009.
- JELINEK, F. *Statistical methods for speech recognition*. [S.l.]: MIT press, 1997.
- KHOSLA, M. et al. 3d convolutional neural networks for classification of functional connectomes. *arXiv preprint arXiv:1806.04209*, 2018.
- KJAERULFF, U. B.; MADSEN, A. L. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. 2. ed. New York NY: Springer, 2013.
- KLIR, G. J. *Uncertainty and information: foundations of generalized information theory*. Kybernetes, Emerald Group Publishing Limited, 2006.
- KOLLER, D.; FRIEDMAN, N. *Probabilistic graphical models: principles and techniques*. Cambridge, MA: The MIT press, 2009.
- KROGH, A.; MIAN, I. S.; HAUSSLER, D. A hidden markov model that finds genes in e.coli dna. *Nucleic Acids Research*, Oxford University Press, v. 22, n. 22, p. 4768–4778, 1994.
- LE, D.-H.; VAN, N. T. Meta-analysis of whole-transcriptome data for prediction of novel genes associated with autism spectrum disorder. In: *Proceedings of the 8th International Conference on Computational Systems-Biology and Bioinformatics*. [S.l.: s.n.], 2017. p. 56–61.
- LEE, J.-G. et al. Deep learning in medical imaging: general overview. *Korean journal of radiology*, v. 18, n. 4, p. 570–584, 2017.
- LEE, L. I. et al. The current state of artificial intelligence in medical imaging and nuclear medicine. *BJR Open*, The British Institute of Radiology., v. 1, p. 20190037, 2019.
- LEVITT, T. S.; AGOSTA, J. M.; BINFORD, T. O. Model-based influence diagrams for machine vision. In: *Machine Intelligence and Pattern Recognition*. [S.l.]: Elsevier, 1990. v. 10, p. 371–388.
- LI, Y. et al. Mach: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology*, Wiley Online Library, v. 34, n. 8, p. 816–834, 2010.
- LIAO, D.; LU, H. Classify autism and control based on deep learning and community structure on resting-state fmri. In: *IEEE. Advanced Computational Intelligence (ICACI), 2018 Tenth International Conference on*. [S.l.], 2018. p. 289–294.
- LIN, Y. et al. A machine learning approach to predicting autism risk genes: Validation of known genes and discovery of new candidates. *bioRxiv*, Cold Spring Harbor Laboratory, p. 463547, 2018.
- LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017.

- LOMBARDO, M. V.; LAI, M.-C.; BARON-COHEN, S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular psychiatry*, Nature Publishing Group, v. 24, n. 10, p. 1435–1450, 2019.
- MANOGARAN, G. et al. Machine learning based big data processing framework for cancer diagnosis using hidden markov model and gm clustering. *Wireless personal communications*, Springer, v. 102, n. 3, p. 2099–2116, 2018.
- MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, Nature Publishing Group, v. 11, n. 7, p. 499, 2010.
- MARCHINI, J. et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics*, Nature Publishing Group, v. 39, n. 7, p. 906, 2007.
- MEYER, I. M.; DURBIN, R. Comparative ab initio prediction of gene structures using pair hmms. *Bioinformatics*, Oxford University Press, v. 18, n. 10, p. 1309–1318, 2002.
- MRAD, A. B. et al. An explication of uncertain evidence in bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, SPRINGER VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, v. 43, n. 4, p. 802–824, 2015.
- MUSTAFA, M. K.; ALLEN, T.; APPIAH, K. A comparative review of dynamic neural networks and hidden markov model methods for mobile on-device speech recognition. *Neural Computing and Applications*, Springer, v. 31, n. 2, p. 891–899, 2019.
- NEIL, M.; FENTON, N.; NIELSON, L. Building large-scale bayesian networks. *The Knowledge Engineering Review*, Cambridge University Press, v. 15, n. 3, p. 257–284, 2000.
- NIELSEN, T. D.; JENSEN, F. V. *Bayesian networks and decision graphs*. 2. ed. New York, NY: Springer Science & Business Media, 2009.
- OZTOK, U.; CHOI, A.; DARWICHE, A. Solving pppp-complete problems using knowledge compilation. In: *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning*. [S.l.: s.n.], 2016. p. 94–103.
- PAIVA, V. A. et al. Gass-metal: identifying metal-binding sites on protein structures using genetic algorithms. *Briefings in Bioinformatics*, Oxford University Press, v. 23, n. 5, p. bbac178, 2022.
- PAIVA, V. de A. et al. Protein structural bioinformatics: An overview. *Computers in Biology and Medicine*, Elsevier, v. 147, p. 105695, 2022.
- PALMER, C. J.; LAWSON, R. P.; HOHWY, J. Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological bulletin*, American Psychological Association, v. 143, n. 5, p. 521, 2017.
- PAPADIMITRIOU, C. H. *Computational Complexity*. [S.l.]: Addison-Wesley, 1994.
- PEARL, J. Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, Elsevier, v. 29, n. 3, p. 241–288, 1986.
- PEARL, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1. ed. San Francisco CA: Morgan Kaufmann, 1988.

- PEARL, J. *Causality: Models, Reasoning, and Inference*. 2. ed. New York, NY: Cambridge University Press, 2009.
- PEYRARD, N. et al. Exact or approximate inference in graphical models: why the choice is dictated by the treewidth, and how variable elimination can be exploited. *Australian & New Zealand Journal of Statistics*, Wiley Online Library, v. 61, n. 2, p. 89–133, 2019.
- PELLAN, R. et al. Automatic detection of age-and sex-related differences in human brain morphology. In: *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM) 27th ANNUAL MEETING & EXHIBITION*. [S.l.: s.n.], 2019.
- POURRET, O.; NAÏM, P.; MARCOT, B. *Bayesian networks: a practical guide to applications*. [S.l.]: John Wiley & Sons, 2008.
- QUEIROZ, F. C. et al. ppigremlin: a graph mining based detection of conserved structural arrangements in protein-protein interfaces. *BMC bioinformatics*, Springer, v. 21, p. 1–25, 2020.
- RABINER, L. R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, Ieee, v. 77, n. 2, p. 257–286, 1989.
- RAHUL, M. et al. An efficient technique for facial expression recognition using multi-stage hidden markov model. In: *Soft Computing: Theories and Applications*. Singapore: Springer, 2019. p. 33–43.
- REICHL, L. E. *A modern course in statistical physics*. [S.l.]: John Wiley & Sons, 2016.
- RODRIGUES, I. D. et al. Machine learning and rs-fmri to identify potential brain regions associated with autism severity. *Algorithms*, v. 15, n. 6, 2022. ISSN 1999-4893. <http://dx.doi.org/10.3390/a1506019510.3390/a15060195>.
- RODRIGUES, I. D. et al. Grsr-a guideline for reporting studies results for machine learning applied to electroencephalogram data. *Revista Brasileira de Computação Aplicada*, v. 15, n. 2, p. 22–35, 2023.
- ROSEN, K. H. *Handbook of discrete and combinatorial mathematics*. 2. ed. Boca Raton, FL: CRC press, 2017.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. [S.l.]: Pearson Education, 2020. v. 4.
- SAHEKI, A. H. *Construção de uma rede Bayesiana aplicada ao diagnóstico de doenças cardíacas*. Tese (Doutorado) — Universidade de São Paulo, 2005.
- SANTANA, C. A. et al. Gremlin: A graph mining strategy to infer protein-ligand interaction patterns. In: IEEE. *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*. [S.l.], 2016. p. 28–35.
- SANTANA, C. P. et al. rs-fmri and machine learning for asd diagnosis: a systematic review and meta-analysis. *Scientific Reports*, Nature, v. 12, n. 6030, 2022. <http://dx.doi.org/10.1038/s41598-022-09821-610.1038/s41598-022-09821-6>.
- SARAÇOĞLU, R. Hidden markov model-based classification of heart valve disease with pca for dimension reduction. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 25, n. 7, p. 1523–1528, 2012.

- SAZ, O. et al. Tools and technologies for computer-aided speech and language therapy. *Speech Communication*, Elsevier, v. 51, n. 10, p. 948–967, 2009.
- SCHIPOR, O.; PENTIUC, S.; SCHIPOR, M. Automatic assessment of pronunciation quality of children within assisted speech therapy. *Elektronika ir Elektrotechnika*, v. 122, n. 6, p. 15–18, 2012.
- SCUTARI, M.; DENIS, J.-B. *Bayesian networks: with examples in R*. 1. ed. Boca Raton, FL: Chapman and Hall/CRC, 2014.
- SILBERSTEIN, M. et al. A system for exact and approximate genetic linkage analysis of snp data in large pedigrees. *Bioinformatics*, Oxford University Press, v. 29, n. 2, p. 197–205, 2013.
- SPIEGELHALTER, D. J.; FRANKLIN, R. C.; BULL, K. Assessment, criticism and improvement of imprecise subjective probabilities for a medical expert system. *arXiv preprint arXiv:1304.1529*, 2013.
- TAMPOSIS, I. A. et al. Semi-supervised learning of hidden markov models for biological sequence analysis. *Bioinformatics*, Oxford University Press, v. 35, n. 13, p. 2208–2215, 2019.
- TESTA, A. C. et al. Codingquarry: highly accurate hidden markov model gene prediction in fungal genomes using rna-seq transcripts. *BMC genomics*, BioMed Central, v. 16, n. 1, p. 170, 2015.
- UĞUZ, H.; ARSLAN, A.; TÜRKÖĞLU, İ. A biomedical system based on hidden markov model for diagnosis of the heart valve diseases. *Pattern recognition letters*, Elsevier, v. 28, n. 4, p. 395–404, 2007.
- VALTORTA, M.; KIM, Y.-G.; VOMLEL, J. Soft evidential update for probabilistic multi-agent systems. *International Journal of Approximate Reasoning*, Elsevier, v. 29, n. 1, p. 71–106, 2002.
- WANG, P. et al. Phonocardiographic signal analysis method using a modified hidden markov model. *Annals of Biomedical Engineering*, Springer, v. 35, n. 3, p. 367–374, 2007.
- WILLIAMSON, J. *Probability logic. Handbook of the Logic of Argument and Inference*. [S.l.]: North-Holland, 2002.
- YAHATA, N. et al. A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nature communications*, Nature Publishing Group, v. 7, p. 11254, 2016.
- ZEMOURI, R.; ZERHOUNI, N.; RACOCEANU, D. Deep learning in the biomedical applications: Recent and future status. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 9, n. 8, p. 1526, 2019.
- ZHANG, D.; ZHANG, H.; ZHANG, B. Computerized tongue diagnosis based on bayesian networks. In: *Tongue Image Analysis*. [S.l.]: Springer, 2017. p. 265–280.
- ZHAO, Y. et al. 3d deep convolutional neural network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland, 2018. p. 172–180.

