JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

An Overview of Causal Inference and its Applications in Health-care and Finance using methods such as Bayesian Networks and Granger's Causality

[1]Sameer Mahajan, [2]Vedang Naik, [3]Gayatri Bangar, [4]Rohit Sahoo

[1]Student, MS Data Analytics, Pennsylvania State University, Malvern, Pennsylvania, United States of America,
[3]Student, Department of Information Technology, Terna Engineering College, Nerul, Navi Mumbai, Maharashtra, India,
[2, 4]B.E., Department of Computer Engineering, Terna Engineering College, Nerul, Navi Mumbai, Maharashtra, India.

Abstract : Causal Inference is a discipline that involves discovering the causal connections between various statistically related quantities. Unlike Statistical Analysis, which focuses on answering the question of "what?" Causal inference is concerned with answering the question of "why?". Causal inference has significant applications in the fields of Health-care, Finance, etc. In this paper, we will be exploring some fundamental concepts in Causal Inference and how it is used in the field of medicine and finance. In Section I we will be providing an introduction to the concept of causality, moving to the Section II we will be looking at various aspects of causal inference. Following that we will see basic concepts of causal model. In section IV, V we will be discussing the implementation of causal in the field of health-care and finance**.**

***IndexTerms* - Causal Inference, Health-care, Finance, Bayesian Networks, Granger's Causality.**

I. INTRODUCTION

The question of why has always plagued humanity, fueling its insatiable curiosity and though sometimes it leads to distress it is also the driving force behind discovery and innovation brought forth by the discovery. These questions are often about causes of the current state of being or the outcomes of various types of interventions carried out to change it, regardless of their field. In Health-care, the cause and effect concern various treatments and their ability to better an individual's health, whereas in Finance the questions are about the outcomes of the implementation of an economic policy and such questions concern causality. Moreover, Correlation does not imply causation and misconstruing correlation as causation is caused by cognitive biases like the Availability Heuristic and Motivated Reasoning [1]. The difference between association and causation is that unlike causation, association can be determined by the distribution alone whereas Causal Inference requires that the outcomes of dynamic conditions for example various treatments and interventions, to also be considered [2]. Causal problems depart from statistical problems because the underlying model is not a fixed joint distribution of random variables but a structure that has multiple such distributions [3].

II. CAUSAL INFERENCE

Causal inference is the way to determine the effects of a particular type of intervention performed, we can take for example a treatment(T) that is used to treat stomach aches(C) and the outcomes of decision(Y). The causal graph would like this:

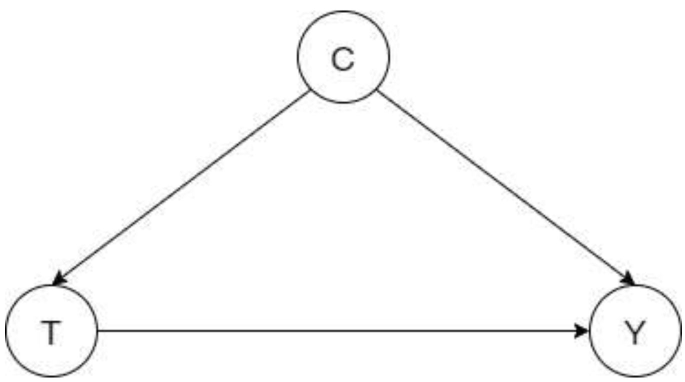

Fig. 1 Causal Graph

Let do(T=1) that the patient receives the treatment T, similarly let do(T=0) signify not receiving the treatment. The outcomes will be $Y_i|_{do(T=1)}$ and $Y_i|_{do(T=0)}$, respectively. For a simpler notation we will use $Y_i(1)$ for the individual getting the treatment and $Y_i(0)$ for the individual not getting the treatment. The causal effect can be then defined as

$$Y_i(1) - Y_i(0) \quad (1)$$

Assuming that $Y_i(1) = 1$ and $Y_i(0) = 0$ that is the outcome of treatment is that the stomach ache goes away when the individual takes the treatment and stays the same if they don't take the treatment.

Thus, $Y_i(1) - Y_i(0) = 1$ ...(2)

There lies a fundamental problem with this statement, that only one of these observations can be calculated, i.e. it is factual, the other cannot be calculated, i.e. it is counterfactual. We cannot know both the effects of taking the treatment and not taking the treatment for an individual. This is referred to as the basic issue of causal inference.

We looked at the individual treatment effect (ITE), and found it to be

$$Y_i(1) - Y_i(0) = 1 \quad (2)$$

Let us look at the effect of mean treatment

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] – E[Y(0)] \quad (3)$$

But,

$$E[Y_i(1) - Y_i(0)] \neq E[Y|T=1] – E[Y|T=0] \quad (4)$$

This is because Y(1) represents a random person in whole population who was assigned the treatment whereas Y|T=1 represents a random person in the subset of the population that took the treatment and these quantities are fundamentally different because of confounding associations that influence the treatment decision. The prior quantity is a causal quantity while the latter is not a causal quantity.

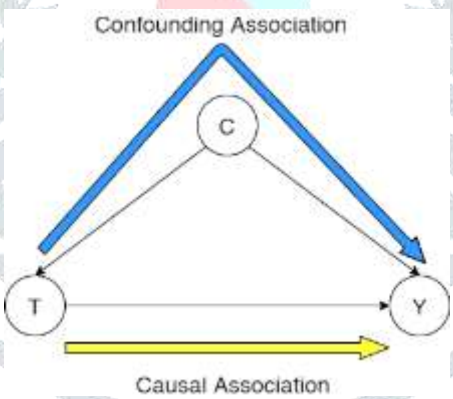

Fig. 2 Confounding Association

The cause influences the treatment, which is not ideal because it adds other associations that are not causal to the outcomes, hence randomized control trials are carried out to have the treatment be independent of the cause. That is T cannot have causal parents and the groups have to be comparable

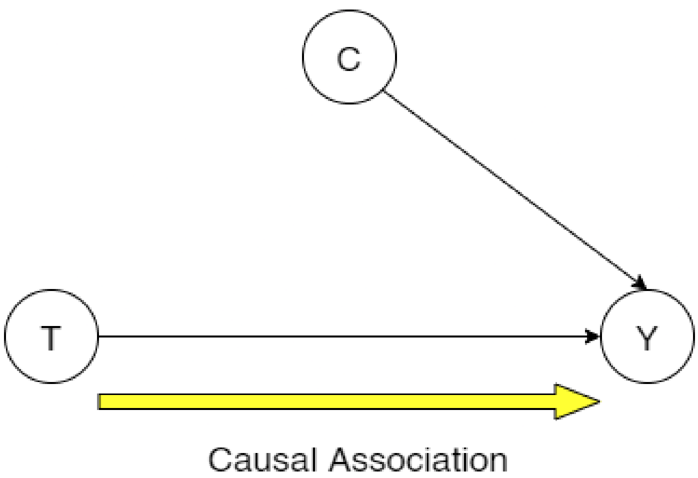

Fig. 3 Causal Association

ATE when there are no causal parents for T, i.e. no confounding associations:

$$E[Y_i(1) - Y_i(0)] = E[Y|T=1] - E[Y|T=0] \tag{5}$$

2.1 Ignorability/Exchangeability

To ensure that ATE is equal to associational difference, we need ignorability/exchangeability, i.e. we need Y to be independent of T

$$(Y(1), Y(0)) \perp\!\!\!\perp T \tag{6}$$

If Y is independent of T,

$$E[Y(1)] - E[Y(0)] = E[Y(0)|T=1] - E[Y(0)|T=0] \tag{7}$$

becomes,

$$E[Y(1)] - E[Y(0)] = E[Y|T=1] - E[Y|T=0] \tag{8}$$

The quantities of E[Y(1)] and E[Y(0)] are causal, whereas the quantities E[Y|T=1] and E[Y|T=0] are statistical. Thus, they are identifiable because a causal quantity that can be computed purely from a statistical quantity is an identifiable causal quantity.[4]

The way this can be done is through RCTs where the causal graph would look like Fig 3.

Exchangeability is not readily available in observational data and conditional exchangeability/unconfoundedness needs to be used.[5] Consider the causal graph in Fig. 1
Conditional Exchangeability is

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid C \tag{9}$$

Therefore, Conditional ATE is

$$E[Y(1) - Y(0) \mid C] = E[Y(1) \mid C] - E[Y(0) \mid C] \tag{10}$$

$$E[Y(1) - Y(0) \mid C] = E[Y(1) \mid T=1, C] - E[Y(0) \mid T = 0, C] \tag{11}$$

$$E[Y(1) - Y(0) \mid C] = E[Y \mid T = 1, C] - E[Y \mid T = 0, C] \tag{12}$$

For calculating actual ATE we need an adjustment formula:

$$E[Y(1) - Y(0)] = E_C E[Y(1) - Y(0) \mid C] \tag{13}$$

$$E[Y(1) - Y(0)] = E_C[E[Y \mid T = 1, C] - E[Y \mid T = 0, C]] \tag{14}$$

Conditional Exchangeability is an untestable assumption because we can never truly know all the if all the confounding factors have been accounted for.

2.2 Positivity

Another condition that needs to be met is Positivity, that is there needs to be at least some entities for each layer of confounding factors [5] that is for all covariates x present in the population of interest

$$0 < P(T = 1 \mid X = x) < 1 \tag{15}$$

2.3 No Interference

The third assumption is the no interference assumption, that is the treatment assignment of other people is considered to have no effect on the possible results of one individual [6].

2.4 Consistency

The fourth and final assumption is consistency, this implies that the exposure is described with sufficient clarity that alternative versions of the exposure have no different influence on the result [7], i.e. there should not be no two versions of a treatment

$$T = t \Rightarrow Y = Y(t) \tag{16}$$

Combining these assumptions to find ATE we get:

Assuming no interference,

$$E[Y(1) – Y(0)] = E[Y(1)] - E[Y(0)] \quad \text{(Linearity of Expectation)}$$

$$E[Y(1) – Y(0)] = E_x[E[Y(1) \mid X] – E[Y(0) \mid X]] \quad \text{(Law of iterated expectations)}$$

$$E[Y(1) – Y(0)] = E_x[E[Y(1) \mid T = 1, X] – E[Y(0) \mid T =0 \mid X]] \quad \text{(Unconfoundedness)}$$

$$E[Y(1) – Y(0)] = E_x[E[Y|T = 1, X] – E[Y \mid T = 0, X] \quad \text{(Consistency)}$$

III. Causal Models

Causal models are the way to convert a Causal Estimand to a Statistical Estimand by defining the Causal assumptions.

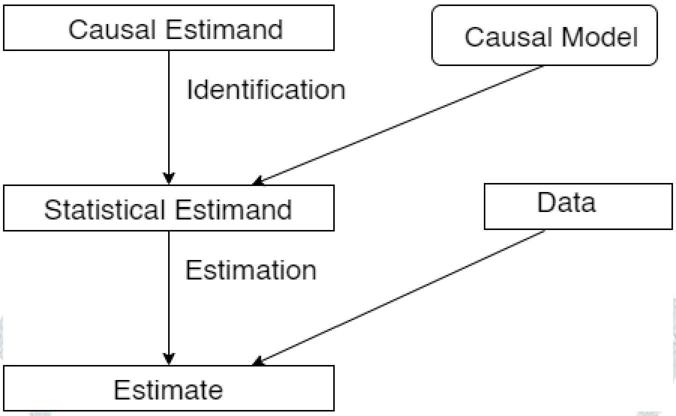

Fig. 4 Role of Causal Model in Causal Inference

The causal mechanism is the effect a parent in a Causal Graph has on its child. The modularity assumption states that an intervention on node X_i should only cause changes in the mechanism $P(X_i \mid pa_i)$ and all other mechanisms should remain unchanged.

3.1 Backdoor Adjustment

There are two types of data that we encounter while doing causal inference, observational data and interventional data. To maintain the assumption of modularity we need to block backdoor paths from the treatment to the result.

S meets the back-door criterion [8] if
i) it stops every path from X to Y with an arrow into X, and
(ii) no node in S is a descendant of X. Then

$$Pr (Y \mid do(X = x)) = \sum_s Pr (Y |X = x, S = s) Pr (S = s) \quad (17)$$

3.2 Front-door Adjustment

S meets the front-door requirement if
(i) it blocks all directed pathways from X to Y,
(ii) there are no unobstructed back-door routes from X to S, (iii) X prevents all back-door paths from S to Y. Then,

$$Pr (Y |do(X = x)) = \sum_s Pr (S = s|X = x) \sum_{x'} Pr (Y |X = x', S = s) Pr (X = x') \quad (18)$$

3.3 Structural Causal Model

A structural causal model is a set of Structural Equations. Structural Equations are written as follows: If A causes B, then

$$B := f_B(A, U_B)$$

Here U_B are the unknown confounding factors that might influence B.

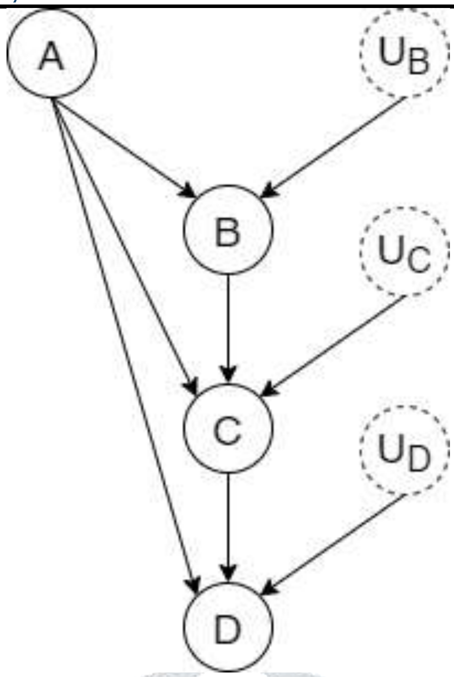

Fig. 5 Causal Graph with unknown confounding factors

For a causal model that is represented by the above graph, the structural causal model will be:

$$M: \quad \begin{aligned} B &:= f_B(A, U_B) \\ C &:= f_C(A, B, U_C) \\ D &:= f_D(A, B, C, U_D) \end{aligned}$$

.

IV. Causal Inference in healthcare

According to the patient in question, the main aim of causal inference applied with machine learning in healthcare is to discover better matching drugs and personalized diagnosis. A particular diagnosis or a drug may work for some patient and may not work for other, while both experiencing the same symptoms [9]. In 2017, the US alone reported approximately more than 1.7 million cancer cases. With an expectancy of more than 600,000, the Agency of Health-Care Research and Quality added that it cost about 80.2 billion USD as an Economic Impact due to cancer [10]. Another such disease is Alzheimer's Disease which comes under the Neurological Diseases [11].

Causal machine learning and simulation to this increasing data of patients and help discover what new treatment methods work for who and at what time. Causal machine learning helps in simulation and asking many if-else questions to then provide a conclusive diagnosis for a particular individual. Thus, precision medicine considers many factors, including individual environment, lifestyle, and genes [12]. Precision medicine does not only forecast risks and consequences but also weigh interventions. Thus, in addition to risk assessment and mortality predictions using machine learning, causal inference can also help in intervention, decision support, and precision medicare [13]. Interventions are key to causal inference as it measures how a situation would be in case of a particular occurrence. And thus, we not only take into consideration the association of various parameters but also ask the question that why did this association originate in the first place.

Electronic Health Records (EHR) with some reassessment to maintain internal validity can be used for causal inference frameworks. Although single stage, contextual knowledge is absent in EHR, bias such as sampling bias and protopathic bias, selection bias, or indication bias can make the inference and causality less accurate, leading to fatal consequences [13]. Randomized Controlled Trials (RCT) can reduce many biases observed in EHR [14] making them valid for causal inference but this is not possible always [15] and thus RCTs cannot be considered as the ideal standard for inference and there isn't a perfect standard yet [16,17,18,19]. Causal Inference, explanation, modeling go hand in hand as inference takes a set of observations as an input producing causal relationships. These relationships can be used to create causal models. In addition to this, causal models can also use prior knowledge. An explanation is achieved when one observation is combined with previous causal relationships to explain a certain occurrence [16]. The explanation is important with causality as we need to understand and answer why a particular patient was diagnosed with a disease and what history and factors lead to this conclusion. In the subsequent paragraphs, we shall cover few approaches to large-scale inference.

4.1 Bayesian Networks

Probabilistic models can be used to determine and quantify relationships between variables. This requires a large amount of data and tracking probabilities in fully conditional models may get cumbersome. The problem is usually that the domain information needed to completely define the dependency on random variables is missing. If available, it may be impractical to calculate the complete conditional probability. Thus, one intermediate method is to develop a model that retains a known conditional dependence between random and conditional variables, therefore bridging the gap between a completely conditional model and a fully conditionally independent model. Bayesian networks are a probabilistic graphical model that captures the known conditional dependency in a graphical model with directed edges. The conditional independence in the model is described by all missing connections. Bayesian networks are therefore less restrictive than the Naive Bayes classifier's overall assumption of conditional independence, but more tractable than discarding all assumptions of conditional independence [30]. Two popular graphical models are the Hidden Markov Model and Bayesian Networks. The Hidden Markov Model is a graphical model with undirected graphical edges that indicates that the graph is cycling (HMM). Bayesian networks are more restricted in that the diagram's edges only point in one direction. As a result, cycles are not conceivable, and the graph is referred to as a directed acyclic graph (DAG). Undirected graphs are better at describing soft restrictions between random variables, but guided graphs are better at conveying causal connections between random variables [31]. A probabilistic graphical model, such as the Bayesian network, enables the definition of a probabilistic model for a complicated issue by describing all of the conditional assumptions for independent variables, allowing for the existence of unknown variables.

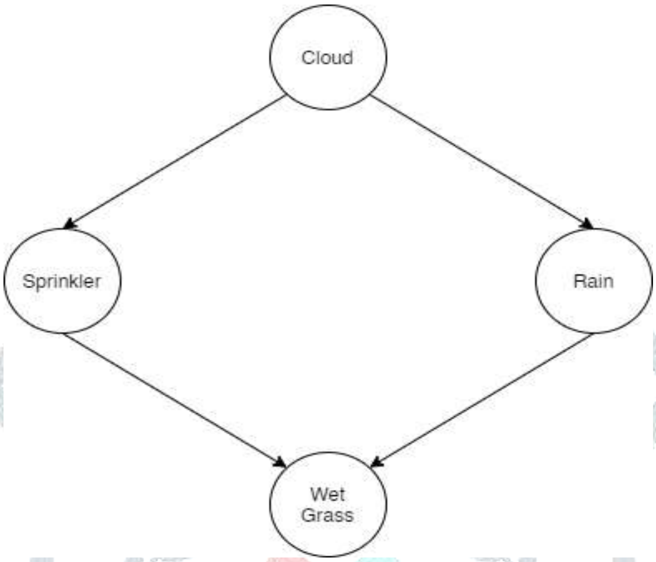

Fig. 6 Bayesian Network

In the above example the probability of Wet Grass is conditionally dependent on Rain. Moreover, the probability of Wet Grass is may also be conditionally dependent on Sprinkler as both have the potential to cause Wet Grass. Probability of Sprinkler when conditionally dependent on Cloud, Rain will be equal to the probability of Sprinkler when it is conditionally dependent on Cloud as Sprinkler is conditionally independent of Rain and hence Bayesian Networks follow the local Markov property. Bayesian Networks are being used in various application such as [32]:

4.1.1 Automated Insight

An application of Business Intelligence is to go through a large number of reports in a way to find the means to a better decision-making process for the firm. Many times, manual analysis may lead to irrelevant and non-essential findings. Bayesian Networks can help identify if a piece of information in the report is essential for analysis or not, thus giving automated insights. This approach can be semi-automated or fully automated.

4.1.2 Anomalous Patterns

The data patterns represented by the dataset are sometimes not as expected or usual. In problems such as disease outbreaks, it is difficult to build a pattern using datasets. Predictions are generally made by comparing data patterns of current data with a baseline distribution. But, the prevalence of various patterns in health care records, such as trends triggered by the day of the week and seasonal fluctuations in temperature and weather, makes determining the baseline a challenging task. Creating the baseline distribution without considering these patterns will result in unacceptably high false-positive counts and long detection periods. Wong et al. [36] presented WSARE 3.0, which replaced baseline distribution with Bayesian Networks taking into account the joint distribution of the data and conditioning on characteristics that are responsible for the trends.

4.1.3 Anomaly Detection

Supervised problems of anomaly detection require a large dataset containing labeled data. The dataset must have inputs and outcomes as well. Adding to this balanced data helps in more accurate classification results. It is difficult to detect future anomalies using past anomalies. A labeled dataset is not present in unsupervised anomaly detection. Because deviations emerge from hostile or rare occurrences of unknown distributions, anomaly identification is seen as an unsupervised learning issue. However, the prediction performance of unsupervised anomaly detection alone frequently falls short of the needed detection rates in a variety of activities, and labeled data is required to assist model construction [33]. In this context, Görnitz et al. [33] proposed a semi-supervised anomaly detection approach. Since they can accommodate high-dimensional data that humans find difficult to understand, Bayesian networks are well adapted for anomaly detection. Although certain anomalies are easily observable when individual variables are plotted, many anomalies are much more complex and are dependent on the association of many variables and the Bayesian network helps in portraying such associations.

4.1.4 Time series

In many applications, analysis can be done by using the data itself. But, In the case of forecasting a time series, each previous value becomes a possible attribute for use in the network model representing the mechanism that generated the time series. This increase in statistical complexity, along with the potential to improve the forecast model using data sources other than the time series being forecasted, emphasizes the need for an autonomous approach for defining the forecast model. Novobilski et al. [34] suggested a technique for updating a Naive Bayesian classifier by looking for connections in the data that will result in a model for the underlying process that creates the time series data. Dagum et al. [35] worked in the same context by using Bayesian Networks to figure out more realistic dependencies for forecasting using time series.

Some more applications include Decision automation, Cost-based decision making, Decision support, and Decision making under uncertainty [32].

The idea of inference using Bayesian Networks is to figure out the set of graphs that represent the data to its best potential. Thus, it is learning the structure of a Directed Acyclic Graph (DAG) from data. Supposing we have tabular data on Lung cancer, Bronchitis, Fatigue, and Chest X-rays of a set of patients we might learn a Directed Acyclic Graph like:

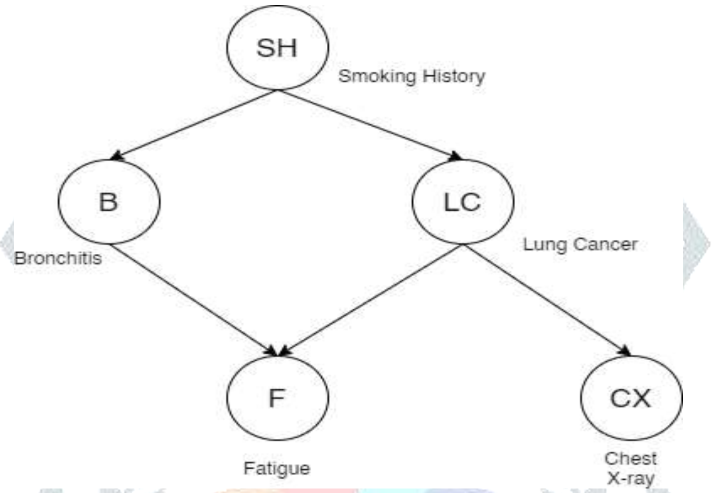

Fig 7. DAG example

There are two ways to learn the structure:

1. Score based approach:

- DAGs are awarded scores depending on the data.
- Finally, we select the DAG with the highest score.

2. Constraint-based approach:

- We attempt to learn the DAG using conditional independencies based on the data.
- The DAG's edges reflect causal dependency. Edges that are missing indicate independence.

Three key assumptions are needed to infer these graphs from data: faithfulness, CMC, and causal sufficiency [16].

Bayesian networks have major applications in Epidemiological data as they can be used to detect and predict the causes, risk factors, frequency, patterns related to diseases. Twardy et al. [20] use Epidemiological data for risk assessment of coronary heart disease by building Bayesian Networks and applying a Bayesian Network learner. Su et al. [21] begin their discussion of the use of Bayesian Networks for discovering relationships between Genetics, environment, and disease by first describing the Bayesian Network approach and its applicability to disease detection taking into account factors such as Genes and the patient's environment. They then describe some methods for learning structure from observational data. Mani et al. [22] in their Bayesian Local Causal Discovery (BLCD) algorithm attempted to figure out a causal link between infant deaths. BLCD could find six probable causes such as Heart Malformations, Hydrocephalus, Weight gain, Microcephalus, Diaphragmatic Hernia, and Apgar Score out of which according to them three seemed to be plausible. Adding to this Bayesian Networks may be disadvantageous as it doesn't guarantee a cause and effect relationship as it encodes only directional information and not bi-directional information. Also, as DAG is data-dependent, if data has three depending parameters making a cyclic relationship then Bayesian Networks will not be able to model this.

4.2 Dynamic Bayesian Networks

Any event is characterized by a certain time frame and it is not generally represented at a particular point of time or instance. The concept of time is extremely essential in the field of AI, Data Science for reasoning as it gives meaning to the data patterns over time. Similarly, almost every disease has time as a major factor in play. A disease cannot be represented in a single-stage but has to be represented over multiple stages of time. Although Bayesian Networks succeed in the causal inference of variables, they fail to consider the temporal components. Dynamic Bayesian Networks represent a structure that changes or evolves dynamically over time. This model would allow users to track and update the system over time, as well as forecast future system behavior. Changing the nature of the static BN to model dynamically can then be viewed as adapting it to a dynamic model [37]. A state-based model depicts the state of each variable at discrete time intervals; hence, a dynamic Bayesian network consists of a sequence of time slices, each of which displays the value of each variable at time t. Edges between multiple variables from different time slices show direction following the flow of time. As in Bayesian Network, DBN have a CPT for each node. DBN is predicated on two assumptions. The first is that the network adheres to the first-order Markov Model, which implies that the state variable t is reliant on the state variable at (t-1). Second, the model's structure and parameters stay constant, i.e., the values may vary but the parameters remain constant.

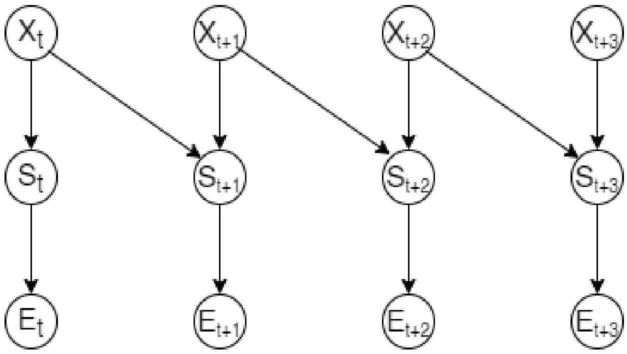

Fig. 8 Dynamic Bayesian Network Example

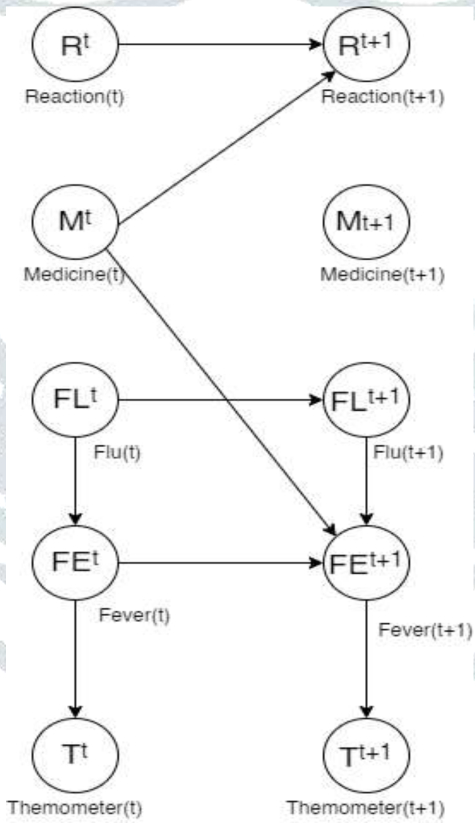

Fig. 9 Dynamic Bayesian Network for Fever

Charitos et al. [23] improved a decision-theoretic network used to help clinicians in detecting and treating pneumonia patients. They used a Dynamic Bayesian Network model to overcome the complexities of ventilator-associated pneumonia diagnosis. A significant disadvantage of the decision-theoretic network was solved by utilizing DBN, since pneumonia could be represented across a series of time and therefore the progression of pneumonia is recorded over a specific period. Dynamic Bayesian Networks also have applications in neuroscience. Eldawlatly et al. [24] illustrated how DBNs may be used to estimate the effective connection between spiking cortical neurons based on their observed spike patterns. DBNs may deduce the underlying non-linear and time-varying causal relationships between these neurons and differentiate between mono and polysynaptic connections between them under specific constraints regulating their putative connectivity. Zou et. al. [25] used DBNs to predict gene regulatory networks. In addition to this, they overcame two of DBN's drawbacks i.e. relative low accuracy and large computational time by limiting potential regulators and using time difference initial change of a regulator and its potential target gene. Despite their widespread use, DBNs have two limitations: the first is that there are no techniques for verifying complicated relationships, and the second is that each link between each time slice is inferred independently [16].

4.3 Temporal Approach

Bayesian Networks and Dynamic Bayesian Networks are unable to decipher the relationships from the given data which can consider the factors of time and sequence of factors contributing to a particular disease. When the objective is to infer disease causality for disease detection and patient diagnosis, the time and sequence of variables contributing to disease become critical. Instead of "a causes b," the temporal method might infer correlations such as "Smoking and asbestos exposure can cause lung cancer for 1 to 3 years."[16]. In this context, Kleinberg et. al. [26] developed a new algorithm based on a framework that combines concepts of philosophical causality with algorithms based on model checks and statistical methods for the test of various hypotheses. In general, a set of logical formulae is created and verified using data, and then a measure of causal significance is computed that compares probable causes with alternative explanations to evaluate an average difference between a cause and its impact likelihood. The test is applied to a set of data from a time series and provides a set of key associations rather than the structure of a graph. To do so, a collection of logical formulas is first constructed utilizing context information or generated by constructing all feasible logical formulas among the variables in the dataset up to a maximum number indicating probable causal links.

V. Causal Inference in finance

Machine learning in finance works best for some use-cases in which a particular prediction is involved and there are fewer regulations, laws, and restrictions involved.

5.1 Machine learning applications in Finance

5.1.1 Credit Scoring:

Credit risk assessment is used to measure a person's credit risk and creditworthiness. Loan applicants are frequently assessed using a credit scoring system, which is frequently based on logistic regression (LR). It is taught through the use of historical records, such as credit history [38]. The main aim is to classify a loan-seeking entity into Good or Bad. A Good predicted entity is likely to repay the financial loan. A certain threshold can be set for credit score for making the loan decision [39]. Credit scoring is generally used when a customer requires a small amount of loan. A set of questions are asked as inputs to the credit scoring model. Based on the vast historical data on loan defaults and payments a credit score is evaluated which is used by banks before lending out loans. Kennedy et. al. [39] proposes an in-depth study on Machine Learning algorithms and their usage for credit scoring. Since credit history is confidential and data is not available in abundance Niu et al. [40] suggested a novel technique to credit rating based on social networks. The purpose of this research is to assess the accuracy of social network data in predicting loan default. They gathered social network data from borrowers' mobile phones and used logistic regression to examine the link between social network data and loan default.

5.1.2 Algorithmic Trading:

Machine learning help in stock forecasting by using statistical models. The algorithms use real-time stock market data to predict the future behavior of the particular stock's portfolio. According to Morgan Stanley research from 2012, computer algorithms performed 84 percent of all stock trades in the United States Stock Exchange, while human investors performed only 16 percent [41][42]. A large amount of information needs to be processed to analyze the behavior of one particular stock. When hundreds of stocks are involved with millions of customers machine learning performs better in analyzing them simultaneously in real-time. Ning Lu et al. [42] investigate different Machine Learning methods such as Logistic Regression, Nave Bayes, and Support Vector Machines to forecast the performance of S&P 500 companies. Deep Learning is also implemented for algorithmic trading use cases. Chen et. al. [43] implemented Long Short-term Memory Deep Neural Network to forecast the stock price of Intel Corporation. Théate et. al. [44] proposes a novel Deep Reinforcement Learning trading strategy for maximizing the resulting Sharpe ratio efficiency indicator across a wide range of capital markets.

5.1.3 Portfolio Management:

Portfolio Management is an online wealth management tool that uses statistical points to maximize client assets and automated algorithms. For example, customers achieve their financial objectives to save a certain amount of time. The consultant then assigns current asset variants and prospects to investments. The management of portfolios includes the development and supervision of selected investments in accordance with long-term financial objectives and risk tolerance for investors. Asset management activities include portfolio creation, risk management, resource management, infrastructure and implementation, and sales and marketing. Machine learning techniques such as processing of unstructured data, supervised learning, validation techniques, unsupervised learning, and reinforcement learning can be used for the above-mentioned tasks. Derek Snow [47] explores these techniques in-depth for portfolio management. Many investment firms have made use of Artificial Intelligence and Machine Learning to enable asset managers to evaluate risk and manage portfolios of their clients. BlackRock Aladdin [45] and Axyon SynFinance [46] are two such financial investment firms.

5.1.4 Fraud Detection:

Using machine learning banks can track a lot of transactions from each account simultaneously. A suspicion may arise if a customer's bank transactions don't fit the pattern based on its historical data concerning payments, loans, etc. Any malicious behavior may be detected quickly and frauds can be avoided efficiently using machine learning. Machine Learning-based fraud detection helps find hidden and complex correlations of data, Automated fraud detection, and real-time processing and has several benefits over the rule-based approach [48]. For the identification of frauds, Babu et al. [49] presented a strategy for dealing with severely skewed financial data that utilizes a convolutional network. This research study uses the collaboration of convolutional neural network layers to decrease misclassification and false positives in an attempt to develop a model for identifying credit card fraud with a high degree of accuracy. They earned a 99.62 percent accuracy rate.

While machine learning performs very well for the above-mentioned use-cases, it isn't always the go-to approach in finance. Many use-cases in finance consist of a large number of rules, regulations, and just prediction will not solve the purpose. Things need to make sense in such cases and machine learning prediction techniques generally perform badly. A problem in which multiple time series are to be considered to reach a particular solution like what would be the GDP of nation A based on the GDP time series of B. Many such time series move simultaneously like the stock market. In order to get accurate forecasting, the interrelation of the two time series is to be found out. Another example could be if a city A with poor revenue wants to increase its revenue, it would try to imitate that time-series pattern of a rich revenue city B. Thus, a situation in which two time series are linked such that one time series can be predicted using the lagged values of a different time series causality comes into the picture and here we say that one-time series granger causes the other. Causality in finance can be applied using Granger's causality which not only helps in accurate forecasting but can also reason as to why certain things happen with respect to others leading to more precise, less biased, and more reliable estimators of causal effects.

5.2 Granger's Causality

Granger's Causality is one of the main methods in the literature to analyze directional connectivity, that is, whether a factor influences another and in what way. The idea of Granger's causality is that given the two time series A and B can we predict the current values of A using past values of B, better than the using past values of A itself. And if the answer is yes then we say that there is a Granger causal interaction of B on A. Moreover, the prediction might not work from A to B as direction is vital in Granger's causality.

Autoregressive models can help us understand Granger's causality. We forecast the variable of interest in a multiple regression model using a linear combination of predictors. In an autoregression model, we forecast the variable of interest by utilizing a linear combination of the variable's previous values. Autoregression is defined as a regression of a variable against itself [51]. (19,20) is a univariate autoregressive model, while (19,21) represents a bivariate autoregressive model (21). These aid in calculating Granger's causality (22)

$$\mathbf{X}_t = a_1\mathbf{X}_{t-1} + a_2\mathbf{X}_{t-2} + \cdots + a_n\mathbf{X}_{t-n} + e_{tx}$$

$$\mathbf{X}_t = \sum_{n=1}^{k} a_n\mathbf{X}_{t-n} + e_t \tag{19}$$

$$\mathbf{Y}_t = b_1\mathbf{Y}_{t-1} + b_2\mathbf{Y}_{t-2} + \cdots + b_n\mathbf{Y}_{t-n} + e_{ty}$$

$$\mathbf{X}_t = \sum_{n=1}^{k} a_n\mathbf{X}_{t-n} + e_t \tag{20}$$

$$\mathbf{X}_t = a_1\mathbf{X}_{t-1} + \cdots + b_1\mathbf{Y}_{t-1} + \cdots + \epsilon_{txy}$$

$$\mathbf{X}_t = \sum_{n=1}^{k} a_n\mathbf{X}_{t-n} + \sum_{n=1}^{k} b_n\mathbf{y}_{t-n} + \epsilon_t \tag{21}$$

$$Granger's\ Causality = \log\left(\frac{\mathrm{Var}[e]}{\mathrm{Var}[\epsilon]}\right) \tag{22}$$

A time series X(t) granger generates a time series Y(t) if the lagged values of X(t) aid in forecasting future values of Y(t) [27,28].

$$Y_t = F(Y_{(t-p)}, X_{tp}) \tag{23}$$

We don't just compute Granger's causality for a complete range of time series. Instead, we compute GC for various small ranges called windows due to the ever-changing nature and dynamics of the time series. We keep on sliding the window to get a graph for granger's causality over the complete time series. One can have a small window or a long window and each has its own pros and cons. A small window model estimation is bad but it has better temporal precision whereas in a long window model estimation is good as we have more data to fit in the model but temporal precision is bad. A typical time window is about 100ms to a few seconds. Another parameter is the model order which is the time which we are incorporating in past values of a time series to predict the current value of another time series. A higher model order is more sensitive to history but has bad model estimation and requires longer computation time whereas a lower model order has better model estimation with faster computation time but it is less sensitive to historical values.

Granger causality happens when two conditions are met. Firstly, cause happens prior to the effect and secondly, the cause has information about the future values of its effect. This signifies that Y(t) should be a function of X(t-1) but not X(t). The lag factor is important in granger's causality. In contrast to Dynamic Bayesian Networks, the method does not seek to find the collection of relationships that best describe a specific dataset, but rather assesses the importance of each relationship which avoids overfitting of data. Granger causality is a "bottom-up" technique in which the data-generating processes in each time series are assumed to be independent variables; the data sets are then evaluated to see if they are connected. A "top-down" approach, on the other hand, assumes that the processes are not independent; the data sets are then examined to see whether they were created independently from each other [50]. Although being vital in causality between 2 time series, Granger's causality fails when a third time series Z(t) that is if X(t) affects Y(t) through Z(t) then we may not be able to forecast and find granger's causality. Moreover, Granger's causality may not be the true causality for a given problem or use case. Chen et. al. [29] considered non-linear extensions to the concept of Granger and proposed an approach which is applied to several chaotic time series as well as to non-linear signals for extended Granger causalities allowing us to evaluate the causal link between 3 or more time series.

VI. ACKNOWLEDGEMENT

We would like to thank Brady Neal from causalcourse.com for providing an excellent introductory course in Causal Inference for the general public.

VII. References

[1] Blanco, F., & Matute, H. The illusion of causality: A cognitive bias underlying pseudoscience. In A. B. Kaufman & J. C. Kaufman (Eds.), Pseudoscience: The conspiracy against science (2018) (p. 45–75).

[2] J. Pearl, "Causal inference in statistics: An overview", Statistics Surveys, vol. 3, no., 2009.

[3] Jonas Peters, Dominik Janzing, and Bernhard Schlkopf. Elements of Causal Inference: Foundations and Learning Algorithms. The MIT Press. 2017

[4] Chen, Hua et al. "Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data." Biometrics vol. 65,3 (2009): 675-82. doi:10.1111/j.1541-0420.2008.01120.x

[5] Hernán, M., Taubman, S. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. Int J Obes 32, S8–S14 (2008).

[6] Hudgens, Michael G, and M Elizabeth Halloran. "Toward Causal Inference With Interference." Journal of the American Statistical Association vol. 103,482 (2008): 832-842.

[7] Rehkopf, David H et al. "The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose is Not a Rose." Current epidemiology reports vol. 3,1 (2016): 63-71.

[8] Shantanu Gupta & Zachary C. Lipton & David Childers, 2020. "Estimating Treatment Effects with Observed Confounders and Mediators," Papers 2003.11991, arXiv.org, revised Jun 2020.

[9] World Health Organization. n.d. *Key facts about cancer*. [online] Available at: <https://www.who.int/cancer/about/facts/en/>.

[10] "Economic Impact of Cancer", *Cancer.org*. [Online]. Available: https://www.cancer.org/cancer/cancer-basics/economic-impact-of-cancer.html.

[11] "Facts and Figures", *Alzheimer's Disease and Dementia*. [Online]. Available: https://www.alz.org/alzheimers-dementia/facts-figures.

[12] "What Are Precision Medicine and Personalized Medicine?", *healthitanalytics.com*, 2018. [Online]. Available https://healthitanalytics.com/features/what-are-precision-medicine-and-personalized-medicine.

[13] Prosperi, M., Guo, Y., Sperrin, M. *et al.* Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat Mach Intell* **2,** (2020) 369–375.

[14] Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M. Randomized controlled trials: part 17 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2011

[15] Schulz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. 1995.

[16] Samantha Kleinberg, George Hripcsak, A review of causal inference for biomedical informatics, Journal of Biomedical Informatics, Volume 44, Issue 6, 2011, Pages 1102-1112

[17] Cartwright N. Are RCTs the gold standard? Biosocieties 2007;2(01):11-20

[18] Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. J Eval Clin Pract 2010;16(2):260–6.

[19] Mackenzie F, Grossman J. The randomized controlled trial: gold standard, or merely standard? Perspect Biol Med 2005;

[20] Twardy C, Nicholson A, Korb K, McNeil J. Epidemiological data mining of cardiovascular Bayesian networks. Electron J Health Inform 2006;

[21] Su, C., Andrew, A., Karagas, M.R. et al. Using Bayesian networks to discover relations between genes, environment, and disease. BioData Mining 6, 6 (2013)

[22] Mani S, Cooper GF. Causal discovery using a Bayesian local causal discovery algorithm. Studies in Health Technology and Informatics. 2004;):731-735.

[23] T. Charitos, L. van der Gaag, S. Visscher, K. Schurink and P. Lucas, "A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients", *Expert Systems with Applications*, vol. 36, no. 2,2009. pp. 1249-1258,

[24] S. Eldawlatly, Y. Zhou, R. Jin and K. Oweiss, "On the Use of Dynamic Bayesian Networks in Reconstructing Functional Neuronal Networks from Spike Train Ensembles", Neural Computation, vol. 22, no. 1,2010, pp. 158-189.

[25] Zou M, Conzen SD. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics. 2005.

[26] Samantha Kleinberg and Bud Mishra. 2009. The temporal logic of causal structures. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. AUAI Press, Arlington, Virginia, USA, 303–312.

[27] Granger, C. W. J. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” Econometrica, vol. 37, no. 3, 1969, pp. 424–438.

[28] C.W.J. Granger, Testing for causality: A personal viewpoint, Journal of Economic Dynamics and Control, Volume 2, 1980, Pages 329-352,

[29] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, Mingzhou Ding, Analyzing multiple nonlinear time series with extended Granger causality, Physics Letters A, Volume 324, Issue 1, 2004, Pages 26-35,

[30] Thomas M. Mitchell. 1997. Machine Learning (1st. ed.). McGraw-Hill, Inc., USA.

[31] Bishop, Christopher M. Pattern Recognition and Machine Learning. New York :Springer, 2006.

[32] "Introduction to Bayesian networks", Bayesserver.com. [Online]. Available: https://www.bayesserver.com/docs/introduction/bayesian-networks.

[33] N. Goernitz, M. Kloft, K. Rieck and U. Brefeld, "Toward Supervised Anomaly Detection", Journal of Artificial Intelligence Research, vol. 46,2013 pp. 235-262.

[34] A. Novobilski and F. Kamangar, "Inferencing Bayesian Networks from Time Series Data Using Natural Selection", Conference: Proceedings of the Thirteenth International Florida Artificial Intelligence Research Society Conference, 2000,

[35] P. Dagum, A. Galper, E. Horvitz and A. Seiver, "Uncertain reasoning and forecasting", International Journal of Forecasting, vol. 11, no. 1, 1995, pp. 73-87.

[36] Weng-Keen Wong, Andrew Moore, Gregory Cooper, and Michael Wagner. 2003. Bayesian network anomaly pattern detection for disease outbreaks. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning (ICML'03). AAAI Press, 808–815.

[37] Mihajlović, V. and M. Petkovic. “Dynamic Bayesian Networks: A State of the Art.” CTIT technical report series (2001): n. pag.

[38] "Using machine learning in credit risk modelling - deepsense.ai", deepsense.ai, 2021. [Online]. Available: https://deepsense.ai/using-machine-learning-in-credit-risk-modelling-to-reduce-risk-costs/.

[39] Kennedy, K. (2013). *Credit scoring using machine learning*. Doctoral thesis. Technological University Dublin.

[40] Niu B, Ren J, Li X. Credit Scoring Using Machine Learning by Combing Social Network Information: Evidence from Peer-to-Peer Lending. Information. 2019;

[41] T. Demos, "‘Real’ investors eclipsed by fast trading", Ft.com, 2012. [Online]. Available: https://www.ft.com/content/da5d033c-8e1c-11e1-bf8f-00144feab49a.

[42] N. Lu, "A Machine Learning Approach to Automated Trading", Boston College Computer Science Senior Thesis, 2016. Available: https://www.bc.edu/content/dam/files/schools/cas_sites/cs/pdf/academics/honors/16Lu.pdf.

[43] Chen, G., Chen, Y., & Fushimi, T. “Application of Deep Learning to Algorithmic Trading”. 2017.

[44] T. Théate and D. Ernst, "An application of deep reinforcement learning to algorithmic trading", Expert Systems with Applications, vol. 173, 2021, p. 114632.

[45] About Us | BlackRock", BlackRock. [Online]. Available: https://www.blackrock.com/corporate.

[46] "Axyon AI | Deep Learning solutions for asset management and trading", Axyon.ai. [Online]. Available: https://axyon.ai/.

[47] Snow, Derek, Machine Learning in Asset Management (July 16, 2019). JFDS: https://jfds.pm-research.com/content/2/1/10 , or http://dx.doi.org/10.2139/ssrn.3420952I

[48] "Fraud Detection: How Machine Learning Systems Help Reveal Scams in Fintech, Healthcare, and eCommerce", AltexSoft. [Online]. Available: https://www.altexsoft.com/whitepapers/fraud-detection-how-machine-learning-systems-help-reveal-scams-in-fintech-healthcare-and-ecommerce/.

[49] A. M. Babu and A. Pratap, "Credit Card Fraud Detection Using Deep Learning," 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2020, pp. 32-36,

[50] "Granger Causality: Definition, Running the Test - Statistics How To", Statistics How To. [Online]. Available: https://www.statisticshowto.com/granger-causality/.

[51] "Autoregressive models | Forecasting: Principles and Practice (2nd ed)", Otexts.com. [Online]. Available: https://otexts.com/fpp2/AR.html.