Financial Informatics

Swakkhar Shatabda

February 15, 2020

Contents

1 Introduction **3**
1.1 Problems in Finance and Business 3
1.2 Information Science 4
1.3 Random Variables 4
1.3.1 Discrete Random Variables 5
1.3.2 Continuous Random Variables 5
1.3.3 Distribution 5
1.4 Probability Distribution 5
1.4.1 Axioms of Probability 6
1.5 An Example 6
1.6 Joint Probability 6
1.7 Conditional Probability 7
1.8 Independence 8
1.9 Conditional Independence 8

2 Bayesian Net **10**
2.1 Enumeration from Joint Distribution 10
2.1.1 Marginal Probability 10
2.1.2 Conditional Probability 11
2.2 Chain Rule 12
2.3 Bayes Theorem 12
2.4 A simple network 13
2.5 Representation of a Bayesian Net 14
2.5.1 A Fraud Detection Problem 15
2.6 Markov Condition 16
2.7 Bayesian Inference Using Enumeration 17

3 Naive Bayes Classifier **18**

4 Markov Models **19**

5 Exploratory Data Analysis **20**
5.1 Histograms 20
5.1.1 Number of Bins and width 20

5.1.2 Variable Bin Widths 20
5.2 Kernel Density Estimator 21

6 Time Series Analysis **22**

7 Linear Regression **23**

8 Multiple Linear Regression **24**

9 Logistic Regression **25**

10 Artificial Neural Network **26**

11 Recurrent Neural Network **27**
11.1 Example Papers to Study 27

12 Conclusion **28**

Chapter 1

Introduction

Financial Informatics is the application of Computer and Information Science to the storage, retrieval and processing of financial data and information [1]. Informatics concerns [2]:

1. organizing data into information,
2. learning knowledge from information
3. learning new information from existing information and knowledge, and
4. making decisions based on the knowledge and information learned.

1.1 Problems in Finance and Business

The problems in finance and business domain are of the hardest nature due to several uncertainties that are present. Financial Informatics has to deal with a lots of uncertainties coming from:

- the financial markets due to the **statistical volatility** of the prices of securities and returns,
- political risks such as changes in regulation or **market structure** that alter strategic priorities,
- **technological risks** such as new product innovation, and model risks such as new mathematical techniques or software implementations.

Problems in finance and business are amongst the hardest problems to be solved on computer systems:

- The analysis of this data has to be carried out in **real time** and the data bubbles through data nurseries (**markets and technical data**) and has to be excavated from data tombs (**fundamental data**)

- The analysis has to include methods developed for **(nonlinear) dynamical systems**, for self-organising systems, for approximate reasoning, for adapting to newer types of data related to innovative financial instruments.

- In addition to the fundamental and technical/market data, we now have the nebulously defined **behavioural data**. This data relates to behaviour of individual stakeholders (traders, regulators, commentators, and increasingly other computers), governmental policies, changes in fashion, unexpected events, market related announcements and so on. All this is regarded as news in the economics/finance literature.

- The quantification of this affect data is a key ontological and computational problem involve information extraction, data mining, fuzzy logic, knowledge representation and many aspects of intelligent systems.

1.2 Information Science

Information systems provides the computational power to achieve efficiency and accuracy in multiple types of financial analyses and complex calculations. The modern computer systems (pc, laptop, smartphone, cloud, etc) are hugely influencing the financial systems.

- Cloud Computing
- Cyber Security
- Cryptocurrency
- Internet of Things
- Data Science

1.3 Random Variables

Random Variable is a mapping/function of the outcome of an experiment to a real value. Examples:

1. Price of a stock
2. Maximum Temperature of last week
3. Number of employees in the company
4. Gender of a sales person/customer
5. Types of companies
6. Amount of money in your bank account

Each of these experiments have got outcomes. These outcomes are of two types: countable and of infinite. The values that a random variable can take are said to be in a set called Domain of the random variable. Random variables are of two types: discrete and continuous.

1.3.1 Discrete Random Variables

Discrete variables are those that have a countable number of values with in any range/duration. Example: Type of asset, Seniority level of an employee, Number of days, etc.

1.3.2 Continuous Random Variables

Continuous variables on the other hand can take infinite number of values in a range. Example: Number of years, amount of money, etc.

Random variables are generally denoted by a capital Letter and its values by small letters.

1.3.3 Distribution

An ensemble of numbers in called a distribution if they sum to 1 and they are within the range $[0, 1]$.

- $0, 0, 0, 0, 1, 0, 0, 0$
- $0.2, 0.2, 0.3, 0.3$
- $1, 0, -1, 1$
- $10, -9, 0$

If we normalize (divide them by their sum) any ensemble of positive integers we get a distribution that is proportional to the original ensemble.

1.4 Probability Distribution

For any random variable if you associate a distribution with all the values it can take, we can call it a probability distribution. For example: Say there is a random variable called X which denotes the number of courses taken by SS in a year. It is either 0,1,2 or 3. This is definitely a case of discrete random variable. Now lets associate a distribution $0, 0, 1, 0$ to this values. This distribution, we can call it a probability distribution of X.
$P(X = 0) = 0$
$P(X = 1) = 0$
$P(X = 2) = 1$
$P(X = 3) = 0$

1.4.1 Axioms of Probability

1. $\sum_x P(X = x) = 1$
2. $\forall_x 0 \leq P(X = x) \leq 1$

1.5 An Example

Consider the following data:

Case	Gender	Height (inches)	Wage ($)
1	female	64	30,000
2	female	64	30,000
3	female	64	40,000
4	female	64	40,000
5	female	68	30,000
6	female	68	40,000
7	male	64	40,000
8	male	64	50,000
9	male	68	40,000
10	male	68	50,000
11	male	70	40,000
12	male	70	50,000

Here, we can define 3 random variables. Lets try to associate probability distributions with them.

Lets find $P(G), P(H)$ and $P(W)$. The three variables have respectively 2,3 and 3 possible values. Thus distribution table of these variables will have that many number of rows summing to 1. We can use the relative frequency property to estimate the distributions for each of these variables.

G	$P(G)$
m	$\frac{6}{12} = 0.5$
f	$\frac{6}{12} = 0.5$

H	$P(H)$
64	$\frac{6}{12} = 0.5$
68	$\frac{4}{12} = 0.334$
70	$\frac{2}{12} = 0.166$

W	$P(W)$
$30k$	$\frac{3}{12} = 0.25$
$40k$	$\frac{6}{12} = 0.5$
$50k$	$\frac{3}{12} = 0.25$

1.6 Joint Probability

Joint probability means that both event can happen together, denoted as $P(X, Y)$. Here we can use the relative frequency approach once again to find join probability distribution for any combination of the variables. For the previous example, lets consider two variables Gender and Wage. We are trying to find $P(G, W)$. These variables have 2 and 3 possible values for each. That is why they will yield $2 \times 3 = 6$ combinations. Thus the joint probability distribution will have six values summing to 1.

G	W	$P(G, W)$
m	30k	$\frac{0}{12} = 0.0$
m	40k	$\frac{3}{12} = 0.25$
m	50k	$\frac{3}{12} = 0.25$
f	30k	$\frac{3}{12} = 0.25$
f	40k	$\frac{3}{12} = 0.25$
f	50k	$\frac{0}{12} = 0.0$

Similar table / distribution is possible from three variables. In the case of $P(G, H, W)$, the number of rows will be 18. For n binary or two values variables number of rows are 2^n.

G	H	W	$P(G, H, W)$
m	64	30k	$\frac{0}{12}$
m	64	40k	$\frac{1}{12}$
m	64	50k	$\frac{1}{12}$
m	68	30k	$\frac{0}{12}$
m	68	40k	$\frac{1}{12}$
m	68	50k	$\frac{1}{12}$
m	70	30k	$\frac{0}{12}$
m	70	40k	$\frac{1}{12}$
m	70	50k	$\frac{1}{12}$
f	64	30k	$\frac{2}{12}$
f	64	40k	$\frac{2}{12}$
f	64	50k	$\frac{0}{12}$
f	68	30k	$\frac{1}{12}$
f	68	40k	$\frac{1}{12}$
f	68	50k	$\frac{0}{12}$
f	70	30k	$\frac{0}{12}$
f	70	40k	$\frac{0}{12}$
f	70	50k	$\frac{0}{12}$

Look at the 18 combinations here. Due to only 12 samples, several of the values are 0. Thus we can assume for such tables with large number of variables we often need a lots of samples if we are to calculate or estimate joint probability distribution using relative frequency approach. Another trouble is the exponential number of rows. It makes the storage of such tables into the memory a challenging issue. There are solution to both of the problems. But first lets get an idea on the strength of this distribution.

1.7 Conditional Probability

Conditional probability of a random variable is the probability when the outcome of the other variable is known. For example, suppose I is a variable denoting the increase of price of stock for a certain company share and B denotes its directors buying shares from the market. Lets assume both are binary variables. Now, conditional probability $P(I|B = yes)$ or $P(I|B = no)$ denotes two conditional probability distributions, one considering the directions bought company shares and the other the opposite. Knowing the value of one variable outcome might change / not change the probability distribution of the other variable. Lets try to calculate the probability distribution of W given the other variable G. There are definitely two cases: male and female.

W	$P(W\|G = male)$
30k	$\frac{0}{6} = 0$
40k	$\frac{3}{6} = 0.5$
50k	$\frac{3}{6} = 0.5$

W	$P(W\|G = female)$
30k	$\frac{3}{6} = 0.5$
40k	$\frac{3}{6} = 0.5$
50k	$\frac{0}{6} = 0$

How did we find it? We followed the same relative frequency approach to find this values with the condition that now we look at only those samples that satisfies the given condition.

1.8 Independence

Sometimes two random variables are independent of each other. Suppose, there are two variables: X and Y. We say X is independent of Y if any knowledge of X does not change the probability distribution of Y. Lets assume initially the probability distribution was $P(Y)$ and after we gained the knowledge of X, the conditional probability is now $P(Y|X)$. Now if this knowledge have no effect, i.e., $P(Y|X) = p(X)$ we say $Y \perp X$ i.e. Y is independent of X and vice-versa. Lets examine one example. Suppose there is a roll of dice.

$X = \{\text{the outcome is at least 3}\}$

$Y = \{\text{the outcome is an odd number}\}$

here, $P(Y = yes) = 0.5, P(Y = no) = 0.5$. When X is known, i.e. the number is at least 3,

$P(Y = yes|X = yes) = 0.5, P(Y = no|X = yes) = 0.5$

and also if the number is not at least 3,

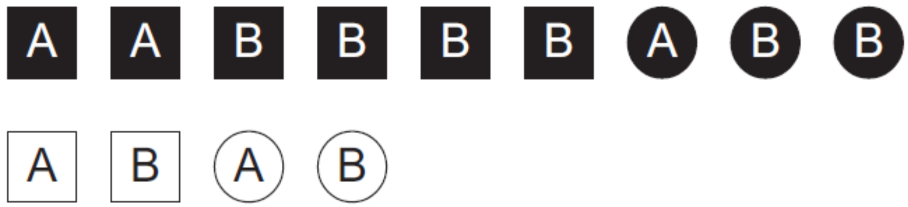
$P(Y = yes|X = no) = 0.5, P(Y = no|X = no) = 0.5$

Note that the conditional probability of Y remains same. Thus we can conclude, $X \perp Y$.

In the previous example, compare G, W and we see that $P(W) \neq P(W|G)$, that is why $G \not\perp W$.

1.9 Conditional Independence

Let us explain the idea of conditional independence using three random variables. $X \perp Y|Z$ if $P(X|Z) = P(X|Y, Z)$. Here note that, if Z is known, then the knowledge of Y does not change the conditional distribution of $X|Z$ which is already conditioned by Z. Consider the following example:



Lets define three variables, A (alphabet = A), B (color = black), C (shape = circle).

Here $P(A) = \frac{5}{13}$

$P(A|C = square) = \frac{3}{8}$

we can conclude $A \not\perp C$

But, $P(A|B = yes) = \frac{3}{9} = \frac{1}{3}$

Now if we add the knowledge of the shape,

$P(A|B = yes, C = square) = \frac{2}{6} = \frac{1}{3}$

$P(A|B = yes, C = circle) = \frac{1}{3}$

the conditional probability distribution does not change. Here we conclude, $A \perp C|B$. That it though they are not independent, they are conditionally independent given B is known. This will lead us to one of the most important structures in graphical models.

Chapter 2

Bayesian Net

In this chapter, we will start first from how to use the joint distribution and then move on to Bayes theorem and Bayesian Nets.

2.1 Enumeration from Joint Distribution

From the joint distribution of any combination of variables it is possible to enumerate and aggregate the probabilities to find the marginal or conditional probabilities. Later, we will extend this idea for Bayesian Enumeration technique.

2.1.1 Marginal Probability

$$P(X) = \sum_y P(X, Y)$$

This formula indicates that marginal probability of X is a sum of all joint probabilities considering all combinations of Y. Lets verify that with one example.

Consider the following join probability distribution of $P(G, H)$

G	H	$P(G,H)$
m	64	$\frac{2}{12} = \frac{1}{6}$
m	68	$\frac{2}{12} = \frac{1}{6}$
m	70	$\frac{2}{12} = \frac{1}{6}$
f	64	$\frac{4}{12} = \frac{1}{3}$
f	68	$\frac{2}{12} = \frac{1}{6}$
f	70	0

Now suppose we wish to calculate marginal of H. Here is the formula:
$P(H = 64) = P(H = 64, G = m) + P(H = 64, G = f)$
$P(H = 68) = P(H = 68, G = m) + P(H = 68, G = f)$
$\cdots$ and so on

G	H	$P(G,H)$
m	64	$\frac{2}{12} = \frac{1}{6}$
m	68	$\frac{2}{12} = \frac{1}{6}$
m	70	$\frac{2}{12} = \frac{1}{6}$
f	64	$\frac{4}{12} = \frac{1}{3}$
f	68	$\frac{2}{12} = \frac{1}{6}$
f	70	0

Now summing up the corresponding rows we should find the following distribution:

H	$P(H)$
64	$\frac{1}{6} + \frac{1}{3} = 0.5$
68	$\frac{1}{6} + \frac{1}{6} = 0.334$
70	$\frac{1}{6} + 0 = 0.166$

In the similar way, any combination of variables and their joint distribution can be derived from a larger join distribution table.

2.1.2 Conditional Probability

In case of conditional probability, the idea is a little tricky. It follows the same procedure. For any given condition, we only consider the rows that matches with the condition and then use marginal probability law to find the distributions. However, the remaining rows might not sum to 1. That is why we need to normalize. Here is an example. Lets suppose we wish to find $P(H|G = m)$ from $P(G, H)$

G	H	$P(G,H)$
m	64	$\frac{2}{12} = \frac{1}{6}$
m	68	$\frac{2}{12} = \frac{1}{6}$
m	70	$\frac{2}{12} = \frac{1}{6}$
f	64	$\frac{4}{12} = \frac{1}{3}$
f	68	$\frac{2}{12} = \frac{1}{6}$
f	70	0

H	$P(H|G = m)$
64	$\frac{1}{6}$ after normalization $= \frac{1}{3}$
68	$\frac{1}{6}$ after normalization $= \frac{1}{3}$
70	$\frac{1}{6}$ after normalization $= \frac{1}{3}$

2.2 Chain Rule

From conditional probability, we derive the following:

$$P(A|B) = \frac{P(A,B)}{P(B)} \tag{2.1}$$

Hence,

$$P(A,B) = P(A|B) \times P(B) \tag{2.2}$$

This is the form of chain rule for two variables. Lets extend that to three variables, A, B, C.
$P(A,B,C) = P(X,C) = P(X|C)P(C) = P(A,B|C)P(C) = P(A|B,C)P(B|C)P(C)$

$$P(A,B,C) = P(A|B,C)P(B|C)P(C) \tag{2.3}$$

In the same way, chain rule is applicable to any number of variables:

$$P(X_n, X_{n-1}, \cdots, X_2, X_1) = P(X_n|X_{n-1}, \cdots, X_1) \cdots P(X_2|X_1)P(X_1) \tag{2.4}$$

Now the order that the list of variables are arbitrary. However, independence and conditional independence will give us freedom to choose the order.
Suppose, two variables are independent of each other, $A, B \implies A \perp B$
Then Eq. 2.1. becomes

$$P(A,B) = P(A)P(B) \tag{2.5}$$

Similar way, if $A \perp B|C$, Eq. 2.3 becomes:

$$P(A,B,C) = P(A|C)P(B|C)P(C) \tag{2.6}$$

2.3 Bayes Theorem

Now, we are ready to apply Bayes theorem! First lets get the theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.7}$$

lets illustrate this with one example. Consider the following data:

Stock Price	Interest Rates			
		Decline	Increase	Unit Frequency
	Decline	200	950	1150
	Increase	800	50	850
		1000	1000	2000

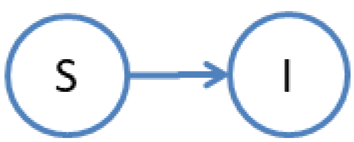
Here lets define two variables, S, I to denote stock price increment and interest rate increment. Now,

$$\begin{aligned} P(S|I) &= \frac{P(I|S)P(S)}{P(I)} \\ &= \frac{\frac{50}{850} \times \frac{850}{2000}}{\frac{1000}{2000}} \\ &= \frac{50}{1000} \\ &= 0.05 \end{aligned}$$

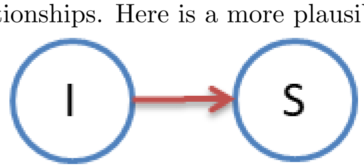
Exercise: Suppose you feel there is a :4 probability the NASDAQ will go up at least 1% today. This is based on your knowledge that, after trading closed yesterday, excellent earnings were reported by several big companies in the technology sector, and that U.S. crude oil supplies unexpectedly increased. Furthermore, if the NASDAQ does go up at least 1% today, you feel there is a :1 probability that your favorite stock NTPA will go up at least 10% today. If the NASDAQ does not go up at least 1% today, you feel there is only a :02 probability NTPA will go up at least 10% today. You have these beliefs because you know from the past that NTPA's performance is linked to overall performance in the technology sector. You checked NTPA after the close of trading, and you noticed it went up over 10%. What is the probability that the NASDAQ went up at least 1%?

2.4 A simple network

The previous problem can be expressed using a simple Bayesian net with two variables.



This is a graph with two variables denoted by the two nodes and the edge represents the relationship between them. There are four possible graphs possible with two variables. Such graphical models usually repsents the casual relationships. Here is a more plausible relationship:



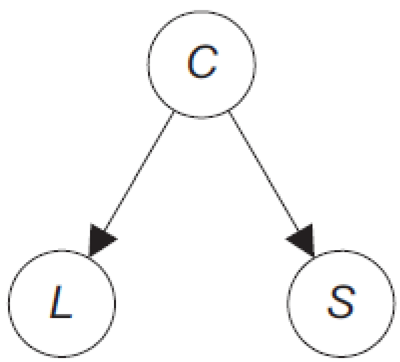
Note that for a Joint Distribution Table with these two variables would require 3 entries (actually 4, however, 3 is sufficient). Here, in this case too we need only 3. For the variable I, $P(I)$ and for S, $P(S|I)$. In the case, where both are independent of each other only 2 values are required. Now, it is a question of how we model the system or which model best fits to our observations.

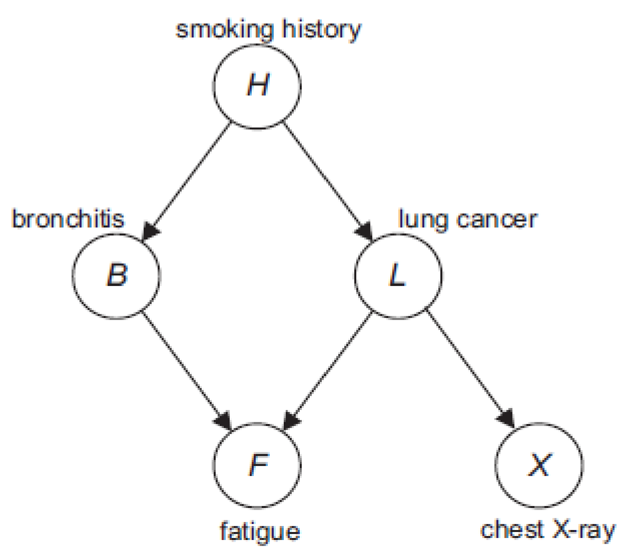
Exercise: How many graphs are possible with 3 variables?

2.5 Representation of a Bayesian Net

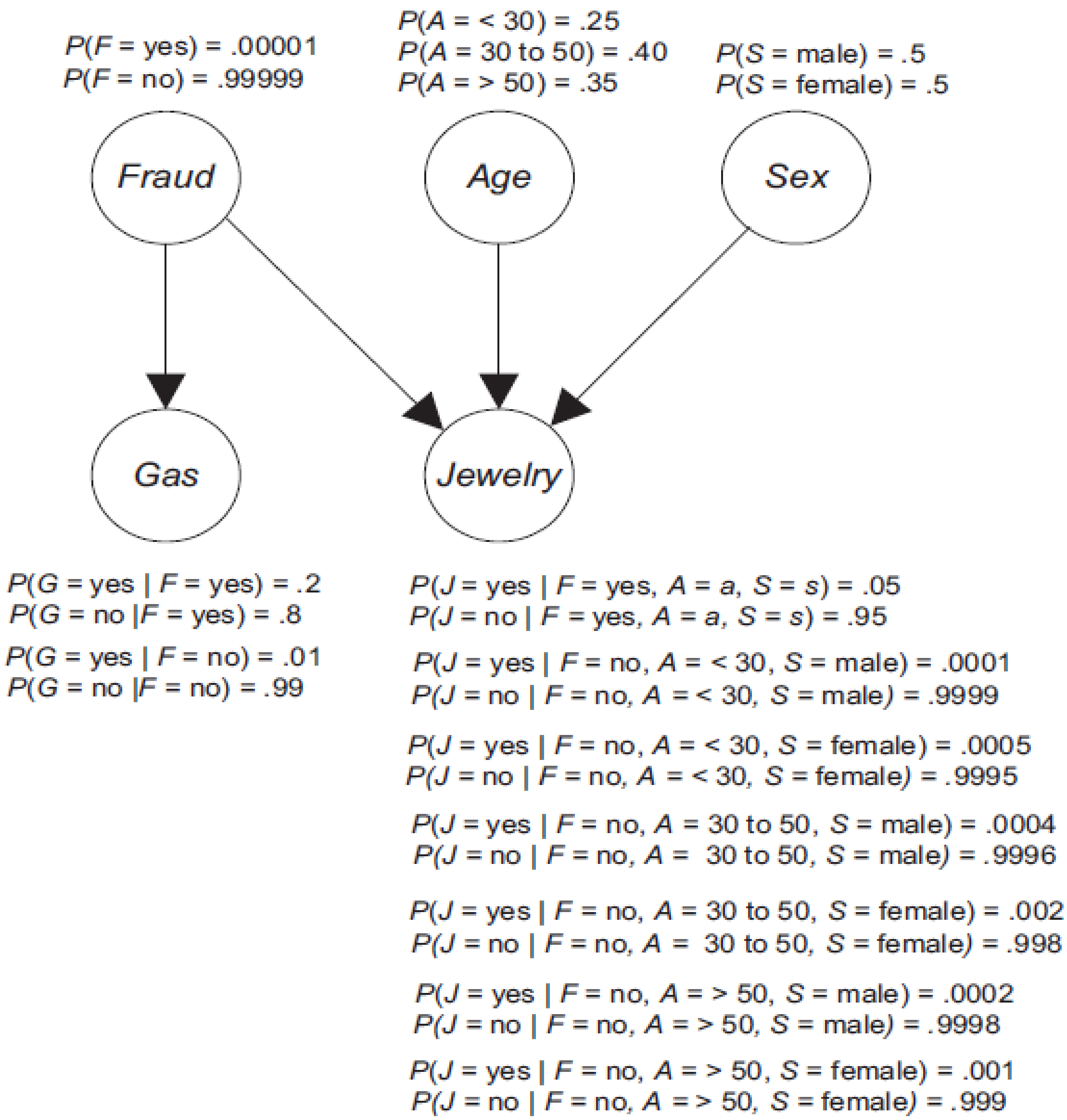
Bayesian network is a directed acyclic graph. Arrows of the directed edges represent the casual flow of reasoning. In a graphical model, each variable is presented as a node and arrows represent the casual relationship. Generally, an arrow is placed from parent variable to the child. Thus the casual flow is always from parent to child. **Each node must be represented with a distribution conditioned by its parents**.

Try to find representations for the following Bayesian Nets.





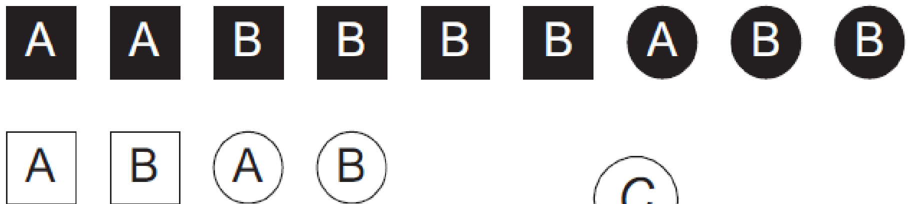
2.5.1 A Fraud Detection Problem

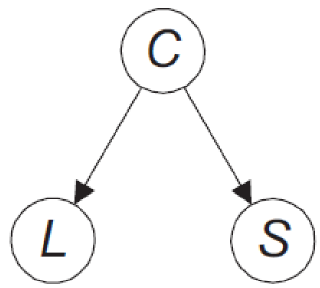


Variable	What the Variable Represents
Fraud (F)	Whether the current purchase is fraudulent
Gas (G)	Whether gas has been purchased in the past 24 hours
Jewelry (J)	Whether jewelry has been purchased in the past 24 hours
Age (A)	Age of the card holder
Sex (S)	Sex of the card holder

2.6 Markov Condition

Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $\mathbb{G} = (V, E)$. We say that $(\mathbb{G}, P)$ satisfies the Markov condition if for each variable $X \in V$, X is conditionally independent of the set of all its nondescendants given the set of all its parents.

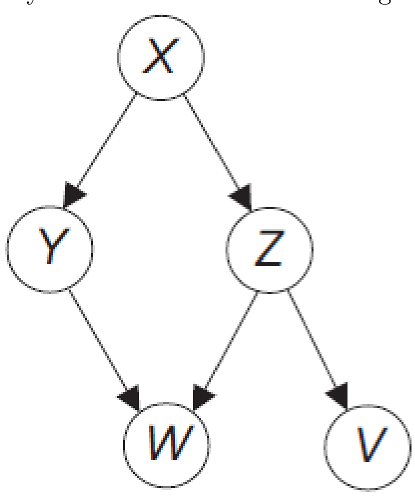




Find all the nodes.
Find all the parents.
Find the descendants.
Now prove the following:

$$L \perp S | C$$

Try the similar with the following DAG.



2.7 Bayesian Inference Using Enumeration

Chapter 3

Naive Bayes Classifier

RID	*age*	*income*	*student*	*credit_rating*	*Class: buys_computer*
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Chapter 4

Markov Models

Chapter 5

Exploratory Data Analysis

5.1 Histograms

Histograms are not bar diagrams. Bar graphs usually relates two variables, histogram deals with a single variable.

5.1.1 Number of Bins and width

Using wider bins where the density of the underlying data points is low reduces noise due to sampling randomness; using narrower bins where the density is high (so the signal drowns the noise) gives greater precision to the density estimation. Thus varying the bin-width within a histogram can be beneficial. Nonetheless, equal-width bins are widely used.

5.1.2 Variable Bin Widths

Rather than using equal sized bins, in some cases we might be interested about equiprobable bins. Suppose, your data is sparser in some areas compared to others.

Creating a Simple Histogram

Different types of data

unimodal, bimodal, multimodal, skewed

Creating a Multiple Histograms

show pak-ind

show overlaps

Creating a 2D Histogram

show hex

5.2 Kernel Density Estimator

Example from handbook

`https://jakevdp.github.io/PythonDataScienceHandbook/04.05-histograms-and-binnings.html` [3]

also need to show the effect of bandwidth and applications in finance

Chapter 6

Time Series Analysis

Chapter 7

Linear Regression

Chapter 8

Multiple Linear Regression

Chapter 9

Logistic Regression

Chapter 10

Artificial Neural Network

Chapter 11

Recurrent Neural Network

11.1 Example Papers to Study

1. [4]

Chapter 12

Conclusion

Bibliography

[1] CFA Walters et al. Towards a discipline of financial informatics. *Available at SSRN 2122059*, 2014.

[2] Richard E Neapolitan and Xia Jiang. *Probabilistic methods for financial and marketing informatics*. Elsevier, 2010.

[3] Jake VanderPlas. *Python data science handbook: essential tools for working with data*. " O'Reilly Media, Inc.", 2016.

[4] C Okan Sakar, S Olcay Polat, Mete Katircioglu, and Yomi Kastro. Real-time prediction of online shoppers purchasing intention using multilayer perceptron and lstm recurrent neural networks. *Neural Computing and Applications*, 31(10):6893–6908, 2019.