CHAPTER I
INTRODUCTION

In the past few decades there is a rapid growth in the applications of Data mining in various fields of engineering, technology and computer science. Data mining is a multidisciplinary field, which covers the areas such as high-performance database technology, artificial intelligence, machine learning, neural networks, knowledge-based systems, knowledge acquisition, data visualization, etc. Data mining [37] refers to extracting or “mining” knowledge from large amount of data. Data mining activities use statistical techniques especially stochastic modeling, since there are uncertainties inherently present in the pattern evaluation. The data mining emerged in early 1990s to extract the valuable knowledge, embedded in the vast amount of data. In late 1990s, the field started receiving the intensive focus from the researchers [58]. The data mining task includes various activities, viz. data cleaning, data integration, data selection, data transformation, data reduction, pattern extraction, pattern evaluation, and knowledge representation.

The scope of the applications of stochastic modeling which are the function of time or space or both is ever increasing. Incorporation of methods such as Markov chain [46][47] and time series analysis [16][74][113], etc made this field to gather more momentum. The
stochastic modeling is applied in stock market prediction, survival analysis, and digital image processing.

1.1 STOCHASTIC PROCESSES

Over the last few decades, it has been increasingly realized that probabilistic models are more realistic than deterministic models. There are many well known common and non-trivial areas of application for probabilistic models. Queueing and inventory problems are two such areas where probabilistic models can effectively be used. In the study of any system, the main work is to derive mathematical models for the responses in terms of input variables as well as time. Almost all of these models are stochastic models, because most of the variables are subject to random variations and most of the measurements of the responses are subject to random measurement errors.

Definitions

A family of random variables \( \{X(t), t \in T\} \) is called Stochastic processes for each \( t \in T \), where \( T \) is the index set of the process, \( X(t) \) is a random variable. An element of \( T \) is usually referred as a time parameter. The state space of the process is the set of all possible values that random variables \( X(t) \) can assume. Each of these values is called the state of the process.
1.1.1 SPECIFICATION OF STOCHASTIC PROCESS

The set of possible values of a single random variable \( X_n \) of a stochastic process \( \{X_n, n \leq 1\} \) is known as its state space. The state space is discrete if it contains a finite or a denumerable infinity of points; otherwise, it is continuous. For example, if \( X_n \) is the total number of sixes appearing in the first \( n \) throws of a die, the set of possible values of \( X_n \) is the finite set of non-negative integers \( 0, 1, \ldots, n \). Here, the state space of \( X_n \) is discrete. We can write \( X_n = Y_1 + \ldots + Y_n \), where \( Y_i \) is a discrete random variables denoting the outcome of the \( i \)th throw and \( Y_i = 1 \) or 0 according as the \( i \)th throw shows six or not. Secondly, consider \( X_n = Z_1 + \ldots + Z_n \), where \( Z_i \) is a continuous random variables assuming values in \( (0, \infty) \). Here, the set of possible values of \( X_n \) is the interval \( (0, \infty) \) and so the state space of \( X_n \) is continuous.

In the above two examples we assume that the parameter \( n \) of \( X_n \) is restricted to the non-negative integers \( n = 0, 1, 2, \ldots \). We consider the state of the system at distinct time points \( n = 0, 1, 2, \ldots \), only. Here the word time is used in a wide sense. We note that in the first case considered above the “time \( n \)” implies throw number \( n \).

On the other hand, one can visualize a family of random variables \( \{X_n, t \in T\} \) (or \( \{X(t), t \in T\} \)) such that the state of the system is characterized at every instant over a finite or infinite interval. The system is then
defined for a continuous range of time and we say that we have a family of random variables in continuous time. A stochastic process in continuous time may have either a discrete or a continuous state space. For example, suppose that $X(t)$ gives the number of incoming calls at a switchboard in an interval $(0, t)$. Here the state space of $X(t)$ is discrete though $X(t)$ is defined for a continuous range of time. We have a process in continuous time having a discrete state space. Suppose that $X(t)$ represents the maximum temperature at a particular place in $(0, t)$, then the set of possible values of $X(t)$ is continuous. Here we have a system in continuous time having a continuous state space.

So for we have assumed that the values assumed by the random variables $X_n$ (or $X(t)$) are one-dimensional, but process $\{X_n\}$ (or $\{X(t)\}$ may be multi-dimensional. Consider $X(t) = (X_1(t), X_2(t))$, where $X_1$ represents the maximum and $X_2$ the minimum temperature at a place in an interval of time $(0, t)$. We have here a two-dimensional stochastic process in continuous time having continuous state space. One can similarly have multi-dimensional processes. One-dimensional processes can be classified, in general, into the following four types of processes:

- Discrete time, Discrete state space
- Discrete time, Continuous state space
- Continuous time, Discrete state space
Continuous time, Continuous state space.

All the four types may be represented by \( \{X(t), t \in T\} \). In case of discrete time, the parameter generally used is \( n \), i.e. the family is represented by \( \{X_n, n = 0, 1, 2, \ldots\} \). In case of continuous time both the symbols \( \{X_t, t \in T\} \) and \( \{X(t), t \in T\} \) (where \( T \) is a finite or infinite interval) are used. The parameter \( t \) is usually interpreted as time, though it may represent such characters as distance, length, thickness and so on. Some authors call the discrete parameter family a stochastic sequence, and the continuous parameter family a stochastic process.

1.1.2 TRANSITION MATRIX

The transition probabilities \( P_{jk} \) satisfy

\[
P_{jk} \geq 0, \sum_k P_{jk} = 1 \text{ for all } j.
\]

These probabilities may be written in the matrix form

\[
p = \begin{pmatrix}
P_{11} & P_{12} & P_{13} & \cdots \\
P_{21} & P_{22} & P_{23} & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\vdots & \vdots & \vdots & \ddots \\
\end{pmatrix}
\]

This is called the transition probability matrix or matrix of transition probability (t.p.m.) of the Markov chain. \( P \) is a stochastic matrix i.e. a square matrix with non-negative elements and unit row sums.
MARKOV CHAIN

Definition

A Markov chain \( \{X_n\} \) is said to be of order \( s (s=1,2,3,...) \), if, for all \( n \),

\[
\Pr\{X_n = k | X_{n-1} = j, X_{n-2} = j_{i-1}, ..., X_{n-s} = j_{i-1}\} = \Pr\{X_n = k | X_{n-1} = j\}
\]

whenever the l.h.s. is defined.

A Markov chain \( \{X_n\} \) is said to be of order one (or simply a Markov chain) if

\[
\Pr\{X_n = k | X_{n-1} = j, X_{n-2} = j_{i-1}\} = \Pr\{X_n = k | X_{n-1} = j\}
\]

whenever \( \Pr\{X_{n-1} = j, X_{n-2} = j, ..., X\} > 0. \)

CHAPMAN-KOLMOGOROV EQUATION:

We have so far considered unit-step or one-step outcome at the nth step or trial given the outcome at the previous step; \( p_{jk} \) gives the probability of unit-step transition from the state \( j \) at a trial to the state \( k \) at the next following trial. The m-step transition probability is denoted by

\[
\Pr\{X_{m+n} = k | X_n = j\} = p_{jk}^{(m)}; \quad \text{...}(1.3)
\]

\( p_{jk}^{(m)} \) gives the probability that from the state \( j \) at nth trial, the state \( k \) is reached at (m + n)th trial in m steps, i.e. the probability of transition from the state \( j \) to th estate \( k \) in exactly m steps. The number \( n \) does not
occur in the r.h.s of the relation (1.3) and the chain is homogeneous. The one-step transition probabilities $p_{jk}^{(l)}$ are denoted by $p_{jk}$ for simplicity.

### 1.1.3 CLASSIFICATION OF STATES

The states $j$, $j=0,1,2,...$ of a Markov chain $\{X_n, n \geq 0\}$ can often be classified in a distinctive manner according to some fundamental properties of the system. By means of such classification it is possible to identify certain types of chains.

**RECURRENT STATE**

A state $i$ said to be recurrent or persistent if and only if eventual return is certain. i.e., $f_{ii}^* = 1$

A recurrent state is further classified as (i) Positive recurrent and (ii) Null recurrent

**POSITIVE RECURRENT**

A recurrent state $i$ said to be positive recurrent if and only if mean recurrence time is finite. i.e., $\mu_{ii} < \infty$

**NULL RECURRENT**

A recurrent state is said to be null recurrent if and only if the mean recurrence time is infinite. i.e., $\mu_{ii} = \infty$

**TRANSIENT STATE**

A state $i$ said to be non-recurrent or transient if and only if eventual return is uncertain. i.e., $f_{ii}^* < 1$
ABSORBING STATE

A state $i$ is said to be an absorbing state if $P_{ii}=1$ and $P_{ij}=0$ for $j \neq i$. A relation reaching $i$ will remain there forever.

REDUCIBLE AND IRREDUCIBLE MARKOV CHAIN

An equivalence class is a class of state in which all the states communicates with each other in closed set. A Markov Chain whose state space consisting of one equivalence class is called an irreducible Markov Chain, otherwise it is reducible.

A Set $C$ is said to be closed if transition from a state in $C$ to a state lying out of $C$ is not possible. i.e., $P_{ij}^{(n)} = 0$ for $i \in C, j \notin C$ $P_{ij} = 0$ for all $n$.

PERIODICITY OF STATES

The period of a state $i$ denoted by $d_i$, is defined as the greatest common deviser of all integers $n \geq 0$ which $P_{ii}^{(n)} \geq 0$.

i.e., $d_i = G.C.D\{n : P_{ii}^{(n)} \geq 0\}$

State $i$ is said to be aperiodic if $d_i=1$ and periodic if $d_i>1$. It states Clearly that $i$ is aperiodic if $P_{ii} \neq 0$.

FIRST PASSAGE TIME DISTRIBUTION

Let $F_{ij}$ denote the probability that starting with state $i$ the system will reach state $j$, let $f_{ij}^{(n)}$ be the probability that it reaches the state $j$ for first time at the $n^{th}$ step

$$F_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \quad \cdots (1.4)$$
we have

$$\sup_{n \geq 1} P_{ij}^{(n)} \leq F_{ij} \leq \sum_{m \geq 1} P_{ij}^{(m)} \quad \text{for all } n \geq 1$$

We have to consider the two cases $F_{ij} = 1$ and $F_{ij} < 1$. When $F_{ij} = 1$, it is certain that the system starting with state $i$ will reach state $j$; in this case

$\{f_{ij}^{(n)}, n = 1, 2, \ldots\}$ is a proper probability distribution and this gives the first passage time distribution for $j$ given that the system starts with $i$.

**MEAN RECURRENCE TIME**

The mean time from state $i$ to state $j$ is given by

$$\mu_{ij} = \sum_{n=1}^{\infty} \eta f_{ij}^{(n)} \quad \ldots \quad (1.5)$$

In particular, when $i=j$, $\{f_{ij}^{(n)}, n = 1, 2, \ldots\}$ will represent the distribution of the recurrence times of $i$ and $F_{ij} = 1$ will imply that the return to the state $j$ is certain. In this case

$$\mu_{ii} = \sum_{n=1}^{\infty} \eta f_{ii}^{(n)} \quad \ldots \quad (1.6)$$

The equation (1.6) is called mean recurrence time for the state $i$. Since $\mu_{ii}$ gives the probabilities of first returns and therefore their sum

$$f_{ii}^{*} = \sum_{n=1}^{\infty} \eta f_{ii}^{(n)}$$

gives the probability of ultimate return to the state $i$. The following are some of the properties of states, which can be studied under a Markov chain.
ERGODIC MARKOV CHAIN

A Positive recurrent, aperiodic state is defined as ergodic state. The Markov Chain with ergodic state is called an Ergodic Markov Chain. When the Markov Chain is recurrent ergodic, the probability for the stochastic process to have transient from a state to other states for a long term can be calculated using the stationary distribution.

1.2 STATISTICAL DATA MINING

Statistical data mining, also known as knowledge or data discovery, is a computerized method of collecting and analyzing information. The data-mining tool takes data and categorizes the information to discover patterns or correlations that can be used in important applications, such as medicine, computer programming, business promotion, and robotic design. Statistical data mining techniques use complex mathematics and complicated statistical processes to create an analysis.

Data mining involves five major steps. The first data mining application collects statistical data and places the information in a warehouse-type program. Next, the data in the warehouse is organized and creates a management system. The next step creates a way to access the managed data. Then, the fourth step develops software to analyze the data, also known as data mining regression, while the final step facilitates using or interpreting the statistical data in a practical way.
Ten years ago *data mining* was a pejorative phrase amongst statisticians, but the English language evolves and that sense is now encapsulated in the phrase *data dredging*. In its current sense *data mining* means finding structure in large-scale databases. It is one of many newly-popular terms for this activity, another being *KDD* (Knowledge Discovery in Databases), and is a subject at the boundaries of statistics, engineering, machine learning and computer science.

Statistical data mining is concerned with querying very large databases, although building efficient database interfaces to statistical software is becoming a very important area in statistical computing. Statistics has been concerned with detecting structure in data under uncertainty for many years: that is what the design of experiments developed in the inter-war years had as its aims. Generally that gave a single outcome (‘yield’) on a hundred or so experimental points. *Multivariate analysis* was concerned with multiple (usually more than two and often fewer than twenty) measurements on different subjects. In engineering, very similar (often identical) methods were being developed under the heading of *pattern recognition*.

### 1.3 OUTLIER DETECTION

Outlier detection encompasses aspects of a broad spectrum of techniques. Many techniques employed for detecting outliers are fundamentally identical but with different names chosen by the authors.
For example, different terminologies described outlier detection as novelty detection, anomaly detection, noise detection, deviation detection or exception mining.

A more exhaustive list of applications that utilise outlier detection are:

- Fraud detection-detecting fraudulent applications for credit cards, state benefits or detecting fraudulent usage of credit cards or mobile phones.

- Activity monitoring- detecting mobile phone fraud by monitoring phone activity or suspicious trades in the equity markets.

- Satellite image analysis - identifying novel features or misclassified features.

- Time-series monitoring - monitoring safety critical applications such as drilling or high-speed milling.

- Detecting unexpected entries in databases - for data mining to detect errors, frauds or valid but unexpected entries.

- Detecting mislabelled data in a training data set.

### 1.4 APPLICATIONS OF FUZZY STOCHASTIC MODELS

In day to day life random experiments are conducted where outcomes are not numbers but are expressed in inexact linguistic variables that change with the time parameter $t$. Consider the group of sampling points chosen at random which are evaluated about the degree of atmospheric
pollution in a particular city. Some possible outcomes would be 'extreme pollution', 'serious pollution', 'light pollution', 'no pollution' and so on, and these results change with the time parameter t. The problem randomness occurs because it is not known in which the outcome may be expected from any given sampling point. Once the outcome is available, then also the concept of analyses uncertainty about the precise meaning of the outcome. The latter uncertainty will be characterized by fuzziness, in the sense that each of the outcomes extreme pollution, serious pollution, light pollution and no pollution will be represented by a dynamic fuzzy set.

Fuzzy random variables are mathematical descriptions for fuzzy stochastic phenomena, but only one time descriptions. For phenomena that are similar to this kind of dynamic fuzzy stochastic phenomena, it is not enough to describe and observe them only once, but it should be done repeatedly and even continuously to describe and observe their evolitional procedures. For these requirements, study fuzzy stochastic processes have to be analyzed. Actually, the objects described with fuzzy stochastic processes are the regularities of the evolitional procedures of the dynamic fuzzy stochastic phenomena.

Applications of fuzzy random variables are analyzed in engineering by researchers. The concept of a fuzzy random variable as a function $F: \Omega \rightarrow F(R)$ (subject to certain measurability conditions) where $(\Omega, A, P)$
is a probability space, and $F(R)$ denotes all piecewise continuous functions $U: R \to [0,1]$. The notion of a fuzzy random variable as a function (subject to certain measurability requirements) $F: \Omega \to F(R^n)$, where $(\Omega, A, P)$ is a probability space, and $F(R^n)$ denotes all functions (fuzzy subsets of $R^n$) $u: R^n \to [0,1]$, such that $\{x \in R^n; u(x) \geq \alpha\}$ is nonempty and compact for each $\alpha \in (0,1]$.

The notion of a fuzzy random variable is introduced. The measurable fuzzy set-valued function is defined as $X: \Omega \to F_0(R)$, where $R$ is the real line, $(\Omega, A, P)$ is a probability space, $F_0(R) = \{A: R \to [0,1]\}$ and $\{x \in R; A(x) \geq \alpha\}$ is a bounded closed interval for each $\alpha \in (0,1]$.

**FUZZY RANDOM VARIABLES**

A fuzzy random variable has been conceptualized as a vague perception of a crisp but un-observable random variable and also as a random fuzzy set. Since actuaries seem receptive to these notions of fuzzy random variables, and there are sources of uncertainty that random variables cannot accommodate, but fuzzy random variables can, one would expect fuzzy random variables to be a component of the actuarial arsenal, but generally this is not the case.

Consider the following, an urn contains approximately 100 balls of various sizes. Several are large. It is easy to identify the probability that a ball drawn at random is large by fuzzy random variable. If the descriptive
variables, namely “approximately”, “several” and “large”, were crisp numbers, the answer to the questions would be a numerical probability. However, since these terms are fuzzy, rather than crisp values, the solution, like the data upon which it is based, is a fuzzy number. Situations of this sort, which involve a function from a probability space to the set of fuzzy variables, give rise to the notion of a fuzzy random variable. When the notion of fuzzy random variables is explained to actuaries, they seem receptive to the idea. They generally recognize that random variables do not capture sources of uncertainty such as modeling choices, parameter choices, application of expertise, boundary conditions and lack of knowledge, while fuzzy random variables can accommodate these things. Consequently, the fuzzy random variables are considered to be a component of the actuarial arsenal, but this is not the case.

A plausible explanation of why fuzzy random variables are not being implemented more often is that potential users are not sufficiently familiar with fuzzy random variables methodology and consequently, they forego opportunities for implementation. Assuming this to be so, the purpose of this is to help alleviate this situation by presenting an overview of fuzzy random variables. The topics addressed include: differentiating between fuzziness and randomness, fuzzy random variables, and conceptualizing fuzzy random variables. For the most part, the discussion is conceptual rather than technical. The goal is to
illustrate how naturally compatible and complementary probability theory and fuzzy logic are and to illustrate how the two can be combined.

In our daily life, forecasting techniques are often used to predict the population growth, economy and stock prices. In recent years many researchers used fuzzy time series to handle forecasting problems. Song and Chissom [97] have introduced the definitions of fuzzy time series and its modeling by using fuzzy relation equations and approximate reasoning (Zadeh [116]). Song and Chissom [98] have outlined the modeling procedures and implemented the time invariant and time variant models to forecast the enrollments of University of Alabama. Chen [21] has developed a basic or simplified method for time series forecasting using the arithmetic operations rather than complicated max–min composition operations. The time series data have been mostly used for forecasting methods.

The time series is a sequence of observations ordered in time. Mostly these observations are collected at equally spaced, discrete time interval. The basic assumption of time series analysis is that some aspects of the past pattern will continue to remain in the future. Melike Sah et al. [64] have proposed forecasting method using first order fuzzy time series. Ruey-Chyn Tsaur et al. [83] have proposed forecasting method using Fuzzy relation analysis in fuzzy time series model. Sun Xihao et al. [102] have proposed the forecasting method used to average
based fuzzy time series model for forecasting Shanghai compound index. Fuzzy time series is used to deal with forecasting problems in which historical data are linguistic values. People can plan or prevent beforehand by forecasting activities. It is impossible to make a hundred percent forecast but the accuracy of forecasts can be increased. The conventional forecasting methods can deal with many forecasting cases but they cannot solve forecasting problem in which the historical data are linguistic values.

**DEFUZZIFIER**

The function of the defuzzifier is to convert the levels of belief in output fuzzy sets to a crisp decision variable of some kind. The result of fuzzy logic operations with fuzzy sets is invariably a conclusion in the form of a fuzzy set. In practice, the output of the defuzzifier process is a single value from the set. There are several built-in defuzzifier methods. The centre of gravity method is the most commonly used methods for extracting a crisp value from a fuzzy set. This method calculates the weighted average of the elements in the support set. The bisector method focuses on the axis of the vertical line which divides the area under the diagram into two equal parts. The mean of maxima method chooses the point by taking the mean of the maximal memberships. The smallest maximum and largest maximum methods choose either the lower or
upper boundary of the maximal membership. The illustration of the description for all methods is as follows:

![Defuzzifier Methods](image)

**Figure 1.1: Defuzzifier Methods**

**UNCERTAINTY**

In many problems, the outcomes of a random event can also be identified through the values of a function, which is called a random variable. This function maps the events of the outcome space of random experiments onto the set of real numbers. For example, if the value of $X$ represents the number of vehicles passing an intersection in a day, $X>1000$ represents the set of events that more than 1000 cars pass an intersection in a day. Instead of using sets to represent random events in a universe, we use values of random variables to represent different events.
Random variables can either be discrete or continuous. The range of a discrete random variable is a set of isolated points, whereas the range of a continuous random variable is a continuum. Probability theory models a discrete random variable in terms of a probability mass function and models a continuous random variable in terms of a probability density function. The probability distribution of a random variable $X$ is:

$$F_X(x) = P(X \leq x) \quad \text{... (1.7)}$$

For a discrete random variable $X$, the above definition can be written as follows:

$$F_X(x) = P(X \leq x) = \sum_{a_i \leq x} P_X(x_i) \quad \text{... (1.8)}$$

where $P_X(x_i)$ is the probability mass function. The probability mass function of a discrete variable, $X$, assigning to each real number, $x_i$, is the probability of the random variable $X$ being equal to $x_i$. Probability distribution functions must also satisfy the same axioms satisfied by probability measures.

The counterparts of random variables in fuzzy set theory are fuzzy variables. Fuzzy set theory uses the membership function (possibility distribution function), $\Pi_X(x)$ (possibility of $X$ being equal to $x$) for both discrete and continuous variables. Possibility distribution functions must satisfy the same axioms as possibility measures. Following are some
differences in the properties of a probability density function and a possibility distribution function of a continuous variable are explained in figure 1.2

The area under the probability density function of a variable in an interval is equal to the probability of the variable assuming any value in that interval, whereas the area under the possibility distribution function has no meaning.

The total area under any probability density function is one whereas the area below a possibility distribution can be less or greater than one.

The probability of a variable, whose probability density function is a continuous function, assuming a specific value is zero, whereas the possibility of the same event can be any value between zero and one.

Figure 1.2: Comparison of probability density function and possibility distribution function

- The area under the probability density function of a variable in an interval is equal to the probability of the variable assuming any value in that interval, whereas the area under the possibility distribution function has no meaning.
- The total area under any probability density function is one whereas the area below a possibility distribution can be less or greater than one.
- The probability of a variable, whose probability density function is a continuous function, assuming a specific value is zero, whereas the possibility of the same event can be any value between zero and one.
Both the probability density function of a random variable and the possibility distribution of a fuzzy variable must be nonnegative. In the case of the possibility distribution function, we have the additional restriction that its maximum value must be one.

The horizontal interval between two points with possibility $\alpha$ on the possibility distribution curve represents an $\alpha$-cut, which is a crisp set that contains all the variables with possibility greater or equal than $\alpha$. For example, the interval $[a, b]$ in figure 1.3 represents a set for which the possibility of each variable is greater or equal to $\alpha$. For a probability density function, there is no analogy for this representation.

Figure 1.3: $\alpha$-cut representation in a possibility distribution function
1.5 STATISTICS IN DIGITAL IMAGE PROCESSING

An image may be defined as a two dimensional function, $f(x,y)$, where $x$ and $y$ are spatial (plane) coordinates, and the amplitude of $f$ at any pair of coordinates $(x,y)$ is called as the intensity or gray level of the image at that point. When $(x,y)$ and the amplitude values of $f$ are all finite, discrete quantities, the image is a digital image. The field of digital image processing refers to processing digital images by means of a digital computer. A digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements, and pixels. Pixel is the term most widely used to denote the elements of a digital image.

Images are produced by a variety of physical devices, including still and video cameras, x-ray devices, electron microscopes, radar, and ultrasound, and used for a variety of purposes, including entertainment, medical, business (e.g. documents), industrial, military, civil (e.g. traffic), security, and scientific. The goal in each case is for an observer, human or machine, to extract useful information about the scene being imaged. Often the raw image is not directly suitable for this purpose, and must be processed in some way. Such processing is called image enhancement; processing by an observer to extract information is called image analysis. Enhancement and analysis are distinguished by their output, images vs. scene information, and by the challenges faced and methods employed.
Image enhancement has been done by chemical, optical, and electronic means, while analysis has been done mostly by humans and electronically.

Digital image processing is a subset of the electronic domain wherein the image is converted to an array of small integers, called \textit{pixels}, representing a physical quantity such as scene radiance, stored in a digital memory, and processed by computer or other digital hardware. Digital image processing, either as enhancement for human observers or performing autonomous analysis, offers advantages in cost, speed, and flexibility, and with the rapidly falling price and rising performance of personal computers and it has become the dominant method in use.

\textbf{1.5.1 PROPERTIES OF DIGITIZED IMAGE}

The properties of digital images can be viewed as follows:

The distance between two pixels in a digital image is a significant quantitative measure. The distance between points with co-ordinates \((i, j)\) and \((h, k)\) may be defined in several different ways;

The Euclidean distance is defined by equation (1.9)

\[
D_{E} \left( (i, j), (h, k) \right) = \sqrt{(i - h)^2 + (j - k)^2} \quad \text{... (1.9)}
\]

City block distance is as in equation (1.10),

\[
D_{4} \left( (i, j), (h, k) \right) = \left| i - h \right| + \left| j + k \right| \quad \text{...(1.10)}
\]
and the chessboard distance is given in equation (1.11),

\[ D_8 ((i, j), (h, k)) = \max \left\{ |i - h|, |j - k| \right\} \quad \ldots(1.11) \]

Pixel adjacency is another important concept in digital images. A pixel is said to possess a 4-neighbourhood or an 8-neighborhood as depicted in the Figure 1.4

Figure 1.4 Pixel Neighborhoods

Border is the set of pixels within the region that have one or more neighbors outside inner borders, if outer borders exist. Edge is a local property of a pixel and its immediate neighborhood is a vector given by a magnitude and direction. The edge direction is perpendicular to the gradient direction which points in the direction of image function growth. The border is a global concept related to a region, while edge expresses local properties of an image function. Four crack edges are attached to each pixel, which are defined by its relation to its 4-neighbors. The direction of the crack edge is that of increasing brightness, and is a multiple of 90 degrees, while its magnitude is the absolute difference between the brightness of the relevant pair of pixels.
The histogram of a digital image with gray levels in the range $[0, L-1]$ is a discrete function $h(r_k) = n_k$ where $r_k$ is the $k$th gray level and $n_k$ is the number of pixels in the image having gray level $r_k$. It is a common practice to normalize a histogram by dividing each of its values by the total number of pixels in the image, denoted by $n$. Thus a normalized histogram is given by $p(r_k) = n_k / n$ for $k = 0, 1, ..., L-1$. Loosely speaking, $p(r_k)$ gives an estimate of the probability of occurrence of gray level $r_k$. The sum of all components of a normalized histogram is equal to 1.

Noise can occur during image capture, transmission or processing, and may be dependent on or independent of image content. Images are often degraded by random noise. Noise is usually described by its probabilistic characteristics. White noise is a constant power spectrum (its intensity does not decrease with increasing frequency) with very crude approximation of image noise whereas Gaussian noise is an approximation of noise that occurs in many practical cases.

**NOISE**

Noise is any undesirable signal. Noise is everywhere and thus has to be learned to live with it. The two common types of noise in images are impulse (or salt and pepper) noise, and random (or Gaussian) noise. Random noise can be expressed in terms of its mean and variance values. In this work impulse noise removal is mainly focused. Impulse
noise randomly and sparsely corrupts pixels to two intensity levels relative high or relative low, when compared to its neighboring pixels.

### 1.5.2 LINEAR AND NON LINEAR FILTERS

In image processing, various linear and non linear filtering methods have been proposed for the removal of impulse noise. Linear filtering techniques used for noise reduction in images are simply given by the average of the pixels contained in the neighborhood of the filter mask. However, linear filters cannot effectively reduce impulse noise and have a tendency to blur the edges of an image. In such situations, median filters, which are non linear filters, provide an effective solution. Compared with convolution filters, the median filter is more robust in that a single very unrepresentative pixel in the filter window will not affect the median value significantly. Also, since the median must actually be one of the pixels in the filter window, the median filter does not create new pixel values when the filter crosses an edge. For this reason, the median filter is better in preserving sharp discontinuities than linear filters. Unfortunately, the median filter is prone to alter pixels undisturbed by noise, thereby causing a number of artifacts including edge jitter and streaking. Modified forms of the median filter which still retain the rank order structure have been proposed to overcome these shortcomings. Basically, the task is to decide when to apply the median filter and when to keep the pixels unchanged. Among those are the
Center-Weighted Median filters, which give current pixel a large weight and the final output is chosen between the median and the current pixel value, detail-preserving median filters and rank ordered mean filter excludes the current pixel itself from the median filter, progressive switching median filter, soft-decision-based filter and prediction-based-filter.

1.6 ORGANISATION OF THE THESIS

This thesis consists of seven chapters, Chapter-I introduces the basic concepts of Stochastic Processes, Statistical Data Mining, and outlier detection, applications of Fuzzy Stochastic Models, Statistics in Digital Image processing, properties of Digitized Images, and some other related relevant concepts to this topic.


Chapter-III discusses the applications of Fuzzy Stochastic Models such as time series using fuzzy logic relationship and membership function, Average Based Lengths in Fuzzy Time Series and Computational Algorithm. These methods are applied and verified with student enrollment data.
Chapter-IV deals with Stochastic Models using Markov Chain Monte Carlo Methods. It also discusses the application of Markov Chain Monte Carlo, Cox model and Bayesian survival analysis with medical data.

Chapter-V explores various time series techniques in data mining which are able to predict with future closing stock prices. In addition to that various global events and their issues predicting on stock market are analysed with five methods such as Typical Price(TP), Bollinger Bands, Relative Strength Index(RSI), CMI and Moving Average(MA) and the results are compared with the developed BSRCTP algorithm. The BSRCTB algorithm performs better than all other algorithms. This shows that the result of our algorithm will provide a good response in the stock market prediction.

Chapter-VI deals with Fuzzy stochastic systems in data mining. Fuzzy Stochastic Impulse noise Detection and Reduction Method (FSIDRM) is developed with appropriate algorithm for reducing Salt and Pepper noise. The result is an image quasi without (or with very little) impulse noise so that other filters can be used afterwards. This nonlinear filtering technique contains two separated steps namely, an impulse noise detection and a reduction it preserves edge sharpness. Adopting the concept of fuzzy gradient values, the detection method constructs a fuzzy set impulse noise. This fuzzy set is represented by a membership function. It will be used by the filtering method, which is a fuzzy method.
adopting appropriate statistical modeling on neighboring pixels. Experimental results reveals that FSIDRM provides a significant improvement on other existing filters.

A new two step filter (FSIDRM), which uses a fuzzy detection and reduction algorithm, has been presented. Its main feature is that it leaves the pixels which are noise-free unchanged. Experimental results show the feasibility of the new filter. A numerical measure, such as the PSNR and visual observations show convincing results for grayscale images. This new method is easy to implement and it has a optimum execution time.

Chapter-VII gives the overall summary and salient findings of the research. It also proposes some new topics for further research in this field.