Chapter -2

MATHEMATICAL AND STATISTICAL MODELS
2.1 Introduction

The user can access the data and resources have communicated with the help of computer network along with other interconnected computers. On the contrary we can see the continuous spate in number of crime in world of web. Researchers have developed models to characterise the problem face by computer networks. For the best possible solution, the mathematical modelling is used to help in better understanding the problems and identifying the crime. The models provide theoretical results supports by simulated results, which include the underlying assumptions variable and parameter, involved using system of ordinary differential equations. Mathematical modelling plays an important role in providing this understanding. Once a model that captures the main features of the progression and transmission of a particular cyber crime (virus) in a population has been formulated, it can be used to predict the effects of different strategies for crime eradication or control. The world wide eradication of computer virus through a carefully developed vaccination campaign initiated by the ICT. Infectious viruses modelling, though often inexact, have enormous potential to help improve human lives as models aids in the understanding and prediction of the phenomena. The spread of virus (cyber crime) has always posed serious concerns and challenges to public as well as economic and social development of the society. Thus its prevention and control becomes extremely important. The aim of epidemic dynamics is to investigate the transmission dynamics of infectious virus.

Epidemic dynamic research is an innovative strategy to fight against the cyber crimes. It designs mathematical modelling to explain the mechanism of transmission as well as dynamics of contagious agents. The mathematical model gives an insight to study the transmission feature of contagious agent by taking the social factor in to consideration. Through this process we can understand the dynamic behaviour of the infectious nodes by
statistical, numerical and simulation approaches. This study provides a depth understanding of dynamic transmission in global prospective.

2.2 Basic Epidemic Models

The epidemiological modelling means the dynamic modelling process where the total population segregated into specific number of compartment or class. It can be represented through system of differential equations by moving between the compartments. The reasons like infection, susceptible, migrations, birth, deaths, recovery, etc.

To design a dynamic model for transmission of epidemic disease, the population fragmented into several class and compartments. Hence this type of model is known as compartmental models.

The following steps are presented in differential equation in compartmental ways.

1. Let’s take the n number of infectious nodes can categorise in class
2. In every stages the vector population is $X(t)$.
3. Let’s assume that the infected nodes are very less as compare to total population, so that the number of susceptible nodes remains constant.

$$\frac{dX}{dt} = F(x(t)) - V(x(t))$$

Where $V$ be the transform matrix, which can transform the $j$ to $i$ compartments of the entry $(i, j)$ and also $F$ be the new in infection rate which can appearance in the $i$ class cause by $j$ class in the entry $(i, j)$.

Kermack and McKendrick [49-51] had initially proposed by dynamic model for infectious diseases, based on compartmental structures. After that other mathematician developed the model. Transmission of disease are epidemic in nature these analogy to computer virus, worm etc.

In the field of epidemiology numerous model have been used. Among them some basic epidemic models like SIS, SIR and SEIR is explained here.
2.2.1 SIS model

\[ \frac{dS}{dt} = \mu - \beta SI - \mu S + \gamma I \]
\[ \frac{dI}{dt} = \beta SI - (\mu + \gamma)I \]  \hspace{1cm} (2.1)

Where \( \beta \) be the contact rate, \( \mu \) be the natural death, birth due to attack and \( \gamma \) be the rate of susceptible from infectious node to susceptible nodes. The basic reproduction number is \( R_0 = \frac{\beta}{\mu + \gamma} \)

The significance of this model is single infected nodes with temporary immunity.

2.2.2 SIR Model
\[
\frac{dS}{dt} = \mu - \beta SI - \mu S \\
\frac{dI}{dt} = \beta SI - (\mu + \gamma)I \\
\frac{dR}{dt} = \gamma I - \mu R
\]

(2.2)

Where \( \beta \) be the contact rates, \( \mu \) be the new born as well as natural death due to attack. Here the basic reproduction number can be calculated as

\[ R_0 = \frac{\beta}{\mu + \gamma} \]

The physical significance of above model is single infected class with permanent immunity.

### 2.2.3 SEIR model

![SEIR Model](image)

\[
\frac{dS}{dt} = \mu - \beta SI - \mu S \\
\frac{dE}{dt} = \beta SI - (\mu + \varepsilon)E \\
\frac{dI}{dt} = \varepsilon E - (\mu + \gamma)I \\
\frac{dR}{dt} = \gamma I - \mu R
\]

(2.3)

Where \( \beta \) be the contact rate, \( \varepsilon \) rate of transmission from E class to I class, \( \gamma \) is the rate of recovery after treatment of antivirus. The basic reproduction number is

\[ R_0 = \frac{\beta \varepsilon}{(\mu + \gamma)(\varepsilon + \mu)} \]
The physical interpretation of the model is presence of latently infected nodes with permanent immunity.

2.3 Analysis of Variance

It is a parametric statistical practice. The sole concept is installed in order to compare a set of data. Designed and developed by Sir R.A. Fisher [32], thus seldom considered as Fisher’s ANOVA. It takes care of methods like t-test and z-test, which are applied to assess means and the qualified variance among them. It also can be nicely executed in the place which consist more than 2 populations or samples supposed to be evaluated.

The vitality of observation in ANOVA is not about analysing the population variance, but similarity of regression. Generally the regression analysis, look into and model the connection amid a dependent and one or more independent changeable parameter. Nevertheless ANOVA diverges from regression in two paths: one, the self-governing varying parameters are valued (categorical), and second, the characteristic of correlation is supposition free (implies model avoid coefficients for the changeable characters). It allows one to settle on whether the differences between the samples are simply due to random error (sampling errors) or not. Frequently it is used to compare the equality of three or more means, however when the means from two samples are compared using such technique it is alike to exercise the t-test for the assessment of very means of independent samples. The concept relies upon the variance association among the data samples in order to reflect the individual samples. The means of various samples will differ from each other if variance within is much smaller than the between aspect. However both the variation approximately of same sizes then there will be absence between the values of sample means.

Assumption of ANOVA

1. All associated population follows a normal distribution
2. Entire population possess the same variance (or standard deviation)
3. The samples are independently chosen and they distinct to each other.
As ANOVA adopts the normal distribution, so it comes into a class of hypothesis test named as parametric test

2.3.1 Types of ANOVA

**One way analysis:** - It is that sort of approach where more than 3 set of data based on a single aspect is judged. For example, average output of 3 employees evaluated on effective period basis.

**Two way analysis:** - Similar to that of above address, here 2 aspects are considered instead of one. For example, prevailing atmosphere and effective period are 2 aspects in order to express the average output of 3 numbers of employees.

**K-way analysis:** - As the name suggest, here the number of aspect will be k.

Particularly the evaluation goes around null hypothesis:

\[ H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k \]

Where \( \mu \) suggests cluster average and \( k \) indicates cluster figure. Though the one-way ANOVA precedes a noteworthy consequence statistically in general, but there another proposition has to be adopted where the duo average diverges from one another.

![Normal Distribution Curve](image)

**Fig. 2.4: Normal Distribution Curve**

2.3.2 Testing of Hypothesis

Hypothesis is explicitly expressed as the set of plans capable of being tested by scientific approach which relating the independent variables to some dependent variables.
This can be put to a test in order to realise its validity. In common proposition evaluation goes on the foundation of null hypothesis holding the other hypothesis there.

The hypothesis where it’s assumes, there is no statistical relationship between two variables is called as null hypothesis. It is generally symbolised as $H_0$. Alternately hypothesis is the reverse of null hypothesis. It says that there is a statistically noteworthy relationship between two variables. It is noted as $H_1$.

### 2.3.3 Steps in Hypothesis Testing

- Create a null and alternative hypothesis.
- Assign the level of significance (ex. 0.01 or 0.05).
- Recover the sample data and compute the test statistic.
- Utilise the values of test statistic to evaluate the calculated values.
- Discard the null hypothesis if calculated values reflected less than the tabulated values.

### 2.3.4 Level of Significance

This is a subsidiary part of hypothesis testing. The impact stage at 5% indicates null hypothesis certainly discarded. The sampling consequences less than 0.05 probabilities of happening indicates null hypothesis valid. Alternatively the 5% stage of impact indicates the 5% risk of rejecting the null hypothesis.

### 2.4 Basic Terminology

**Asymptomatic** (individual): Infected nodes have no any recognized sign of infection.

**Basic reproduction number** ($R_0$): Expected number of secondary infectious nodes created during the infection of wholly susceptible nodes.

**Beta:** - Technically it signifies per capita rate where 2 particular nodes have significant contact with one another on time frame. It is also referred to using other names such as transmission coefficient, transmission rate, contact parameter etc.
**Closed Population Model:** a model in which it is assumed that there is no migration into or out of the population, and in which there are no births or deaths.

**Compartmental Model:** A model in which individual in the nodes are sub divided in to broad subgroup (compartments) and the model tracks individuals collectively.

**Deterministic model:** A model which explains that what happen on average in nodes and does not incorporate the effects of chance.

**Dynamic Model:** Models which describe changes in given quantities over time. The term is frequently used to describe models which describe contact between individual and therefore for which change in the prevalence of infectious persons are feedback into change in the force of infection.

**Effective Contact:** The rate contact at which transmission of nodes occurs between infectious and susceptible nodes.

**Endemic Infection:** In the compartmental model the infection present at a long stretch in the computer unit.

**Epidemic:** The very rise and subsequent declining of incidence follow the reintroduction of an infection in a compartment/computer system. Also refers to the occurrence of case in a given locality at a frequency which greatly exceeds that expected.

**Epidemic Threshold:** - Minimum proportion of the nodes that needs to be susceptible for the infection incidence to increase, calculate as $1/R_0$.

**Eradication:** Reduction in the incidence of an infection to zero.

**Growth Rate of an Epidemic:** Rate at which the prevalence of infectious computer nodes increases typically calculated during the early stages of an epidemic.

**Immune (individual):** The node that has complete protection to an infection which results from either quarantine or previous infection. Computer nodes are said to be partially immune, if they are not fully protected.

**Incidence Rate:** The number of new events such as infections or cases of susceptible per unit time.
**Infectious Period:** The time period during which individual are infectious.

**Next Generation Matrix:** It the matrix or Jacobian number of secondary infectious nodes produced by infectious nodes in each subgroup in the model.

**Stable Population Model:** A model in which it is assumed that the population size does not change.

**Static Model:** - A model which does not describe change in given quantities in the model over time.

**Stochastic Model:** A model which incorporate the effects of chance, for example in determining the number of individual that infected, become infectious per unit time, recover etc per unit time.

**Transmission Dynamics:** Change in the transmission of infection over time.

**Typical Infectious Nodes:** The theoretical average of all infectious nodes in a population, averaging over, for example infectious individuals in different contact group.

**Susceptible:** An individual who is at risk of infection but has not yet been infected.

**Exposed:** Exposed is a class of population who have been in contact with malicious codes but do not show any infective effect to the infection free populations

**Infectious:** The section of nodes in which all the individual are infected by malicious objects.

**Quarantine Nodes:** The nodes which can isolate from the computer network, affected by the malicious object or infected by the virus.

**Recovered:** The nodes which are eliminated from the infection.

**Incidence:** The absolute fresh infection contaminated through single identity upon the system considering time as the pivotal is called incidence of the node.

**Simple Mass Action:** It is defined as $\beta SI$, where $\beta$ is called transmission coefficient.

**Latent period:** The period of time taken to transmit the agent from infection another host.
Pre Infectious Period: Time period between infection and onset of infectiousness. Some time it refers as latent period.

Death of Malicious Object: When an antivirus run in a computer system for a particular time period the infected files eliminated from the infections and called as death of malicious objects.

Equilibrium: It is a situation of balance among infection, birth, death and recovered rate etc., there by signifying a stable state.

Equilibrium Point: The other name of the point is called stationary point. Which has zero length.

Stable: If the equilibrium point is re-registered after disturbing the unit from its original neutral condition then the phenomena said to be stable equilibrium.

Unstable: The sole identity regarded as unstable, when the system once disturbed slightly from, it do not rehabilitate.

Global Stable: The exclusive identity is considered as globally stable if the same system registered stability after being disturbed irrespective to point of influence.

Asymptotical Stable: If the Lyponov function is negative definite in particular then the stability aim is considered as asymptotically stable.

Globally Asymptotical Stable: If a system is asymptotically confirmed for each starting clarification then the phenomena regarded as globally asymptotically stable.

Variational Matrix: The Variational matrix or Jacobian of the system of ordinary differential equation is defined as

\[ J = \begin{pmatrix} \frac{\partial F}{\partial N_1} & \frac{\partial G}{\partial N_1} \\ \frac{\partial G}{\partial N_2} & \frac{\partial G}{\partial N_2} \end{pmatrix}, \quad \frac{dN_1}{dt} = F(N_1, N_2) \]

\[ \frac{dN_2}{dt} = F(N_1, N_2) \]
**Lipunov Function:** A positive definite function $E(N_1,N_2)$ with the property that $\frac{\partial E}{\partial N_1} F + \frac{\partial E}{\partial N_2} G$ is negative semi define is called a Lipunov function.

**Rout-Hurwitz Criteria:** A necessary and sufficient condition for the Eigen values $\lambda$ is that everyone has unenthusiastic existent parts possess the positive values of every primary diagonal of minors deal to Hurwitz matrix.

$$H_n = \begin{pmatrix}
  a_1 & 1 & 0 & 0 & 0 & 0 & - & - & - & 0 \\
  a_2 & a_1 & 1 & 0 & 0 & - & - & - & 0 \\
  a_3 & a_4 & a_3 & a_2 & a_1 & 1 & 0 & - & - & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_n
\end{pmatrix}$$

As a matter of fact, the specific Hurwitz’s criteria has been induced, begin with 2nd to 4th degree polynomials.

(i) $\lambda^2 + a_1\lambda + a_2$, Hurwitz’s stipulation shrinks to, $a_1 > 0$ and $a_2 > 0$

(ii) $\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3$, Hurwitz clause diminishes to $\lambda > 0, a_2 > 0$ and $a_1a_2 - a_3 > 0$.

(iii) $\lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4$, Hurtwitz order lessen to $\lambda > 0, a_2 > 0, a_3 > 0, a_4 > 0$ and $a_1a_2a_3 - a_2^2 - a_1^2a_4 > 0$.

**2.5 Defence Mechanism for the Cyber Security**

Cyber security and infrastructure protection can be achieved only by understanding the behaviour and techniques of the attacker. Based on this knowledge the success of network security mechanism called defence mechanism can be made to protect the computer networking against the exploration opposite to worm action breakthrough.
software added with expert’s conventional potency. The defence mechanism quality expressed by a blend of shifting time begins with the communicable category to quarantine category beside with shifting time from quarantined category to improved category. Some of defence mechanism are virus scanner, antivirus software, firewall etc. The defence mechanism is cyclic model which is depicted in following figures.

Fig. 2.5: Defence Mechanisms
2.6 Control Techniques

The common control technique is quarantine. The installation of quarantine method in e-epidemic models generally observed at two circumstances. At the very first occurrence the underneath reproductions presume stable immunity, hence the quarantine nodes are supposed to enter straight to improve set of information. At the next occurrence the immunity considered provisional, henceforth the rehabilitated points not only miss their potency but also straight deficiency of possessed potency and added with vulnerability feature towards malicious influence. Almost in every time the inherent potency becomes falter responding to malware exploration, and also with the induction of the same through cyberspace. In a nutshell the stability of potency has certain range of practicality. By using both preventive and curative anti-virus technology in tandem, an individual project his own system more effectively and technically. This leads to reducing the chances of virus, worms, etc. spreading to other systems.