American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Modelling Infectious Diseases Using Markov Chain

Kehinde Adigun[a]*, Ayan Adeleke[b], Oluwasesan Adewusi[c], Oluwafunmilola Olubiyi[d], Omobolaji Halid[e], Bayowa Babalola[f]

[a,b,c,d,e,f]*Department of Statistics, Ekiti State University, Ado-Ekiti, Ekiti State, P.M.B. 5363*

[a]*Email: kabimbola@gmail.com*

Abstract

This study is one of the few that has dwell on the application of Markov Chain in modeling infectious diseases in Nigeria. The study takes a look at the environmental factors that leads to the spread of infectious diseases. This research estimates the transition pattern of the diseases, Testing the Markovian property, and how stationary the process is over the study period, the study concluded that the past history of infectious diseases will affect the future through the present state and recommended that government still need to do more in the area of sensitization and fight against infectious diseases.

Keywords: Infectious Diseases; Markov Chain.

1. Introduction

A Stochastic process is a mathematical model that develops over time in a probabilistic manner. Also, is a family of random variables where is a time parameter running over a suitable index set in a given (common) situation, Stochastic processes for which are particularly of important applications in most fields of study? Different situations are frequently arise where the time parameter t, could stand for something else. For example, may represent distance from an arbitrary origin, and may count the number of defects in the interval along a thread, or the number of cars in the interval along a highway [1]. A Markov chain is a mathematical model of a random phenomenon evolving with time in a way that the past affects the future only through the present with some level of probability. The "time" can be discrete (integers), continuous (real numbers), or, more generally, a totally ordered set. The term Markov chain is used to describe a process observed at discrete intervals. It describes the situations where the result of an experiment depends only on the outcome of the former experiment.

---

* Corresponding author

The next state of the system depends only on the current state, not on the preceding states. Applications of Markov Chains in medicine are quite common and have become a standard tool of medical decision making. Markov Chains is named after Russian mathematician, A.A. Markov (1856-1922), who started the theory of stochastic processes. This study is of utmost importance because the model adopted which is Markov Chain will give a better results than other previous models often adopted like SIR model and SIAD model. More so that the study will cover more than one disease so federal disease surveillance, state ministry of health will be adequately guided in their policies. In [2], a multistate models based on Markov processes method of estimating rates of transition between stages of disease, and the diagnoses of the stages of diseases are sometimes subject to error. They used a general hidden Markov model for simultaneously estimating transition rates and probabilities of stage misclassification. Covariates were fitted to both the transition rates and the misclassification probabilities. They analyzed how Multistate Markov Models in continuous time are often used to model the course of diseases. We found out failure of the Markov assumption will probably appear as an inflation of the misclassification probabilities since different people typically have different aneurysm growth rates, so the transition probabilities may depend on their history as well as the most recent stage. It was also discovered that in developing our understanding of disease screening programmes, modeling the progress of the disease through time is important. The authors in [3] analyzed Hospital Infections data, using Markov Models. They presented a general approach to estimating parameters of continuous-time Markov Chains from discrete sampled data. Their methodology was combined with a new stochastic model for transmission of hospital-acquired infections one which accounts for dynamic bed occupancy providing a method for estimating the parameters of such system. Their study provides a new method for incorporating partial observability and comparing the results from this method to the commonly used existing approach. They provided a model and accompanying methodology for addressing an important problem often encountered when analyzing hospital infection data, namely dynamic bed occupancy. They also investigated when it is necessary to incorporate dynamic bed occupancy in estimation procedures and they provide clear methodology for how this may be practiced effectively. Their results are anticipated to have wide application in studying nosocomial infections, and for assessing the efficacy of possible management strategies designed to decrease the prevalence of such infections. In their research we found out that large amount of variation exists in their estimations and hidden Markov model approach needs to be carefully reviewed for further research. Alyssa in [4] proposed a newer method involving Bayesian inference and then Markov Chain Monte Carlo to estimate the parameters involved in analyzing infectious diseases. He built a model that specifies the mechanism of the spread of the disease using certain variables which include latent periods, variable infectivity rates or natural immunity, immunity upon recovery, time of infection and more. He formulated the problem using Bayesian inference and used Markov Chain Monte Carlo techniques to estimate the solution because of its flexibility and allowing for missing data. They also discovered that one could use this method to determine the unknown infection times instead of the times at which symptoms appeared.

2. Materials And Methods

2.1 Environmental factors influencing the spread of infectious diseases

Many factors have power to affect the spread of infectious diseases that are prone to cause epidemics. The most

important of these are; Water Supply, Sanitation facilities, Food, Climate. Lack of safe water, inadequate excreta disposal facilities, poor hygiene, poor living conditions and unsafe food can all cause diarrheal/ communicable diseases. These diseases are a major cause of suffering and death in an emergency situation. Climate can affect disease transmission in a variety of ways. The distribution and population size of disease vectors can be heavily affected by local climate. Flooding after heavy rains can result in sewage overflow and widespread water contamination. There is some evidence to suggest that pathogens can be spread from one region to another along air streams or by wind [5]. Natural methods, Medical methods and Preventive measure are major methods of fighting diseases. The study will start with epidemiology, which is the study and analysis of the distribution who, when, and where and determinants of health and disease conditions in defined populations. Working out why certain people are getting ill. Epidemiology has helped develop methodology used in clinical research, public health studies, and, to a lesser extent, basic research in the biological sciences. Major areas of epidemiological study include disease causation, transmission, outbreak investigation, disease surveillance, forensic epidemiology, occupational epidemiology, screening, bio monitoring, and comparisons of treatment effects such as in clinical trials. Epidemiologists rely on other scientific disciplines like biology to better understand disease processes, statistics to make efficient use of the data and draw appropriate conclusions, social sciences to better understand proximate and distal causes, and engineering for exposure assessment [6]. Epidemiology is concerned with the incidence of disease in populations and does not address the question of the cause of an individual's disease. This question, sometimes referred to as specific causation, is beyond the domain of the science of epidemiology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiology addresses whether an agent can cause a disease, not whether an agent did cause a specific plaintiff's disease. Applied epidemiology is the practice of using epidemiological methods to protect or improve the health of a population. Applied field epidemiology can include investigating communicable and non-communicable disease outbreaks, mortality and morbidity rates, and nutritional status, among other indicators of health, with the purpose of communicating the results to those who can implement appropriate policies or disease control measures. The second major diseases is epidemic, an epidemic is the rapid spread of infectious disease to a large number of people in a given population within a short period of time, usually two weeks or less. For example, in meningococcal infections, an attack rate in excess of 15 cases per 100,000 people for two consecutive weeks is considered an epidemic. Epidemics of infectious disease are generally caused by several factors including a change in the ecology of the host population (e.g. increased stress or increase in the density of a vector species), a genetic change in the pathogen reservoir or the introduction of an emerging pathogen to a host population (by movement of pathogen or host). Generally, an epidemic occurs when host immunity to either an established pathogen or newly emerging novel pathogen is suddenly reduced below that found in the endemic equilibrium and the transmission threshold is exceeded [7]. The virus is present in the blood and can be transmitted when infected blood comes in contact with mucous membranes or cuts in the skin. Children usually acquire HIV from their mothers while in the womb or at birth. Other modes of transmission include unprotected sexual intercourse, sharing Intravenous needles and direct blood to blood contact. HIV is not spread by casual contact (routine classroom activity, toilet seats, etc.), including contact with saliva or tears. Infants who acquire HIV infection before or during birth from infected mothers typically develop symptoms between 12-18 months of age, although some remain symptom-

free for more than 5 years.

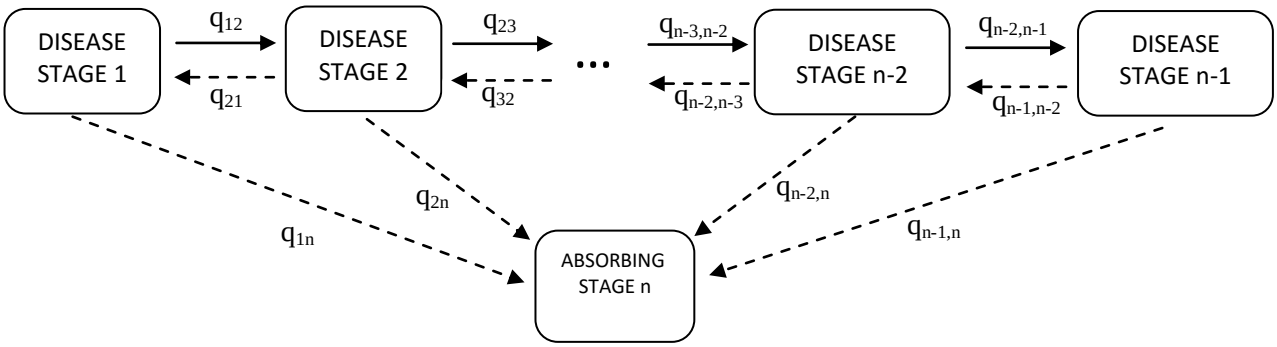

Figure 1

General model for disease progression

Represents a series of successively more severe stages of disease and an "absorbing" state, often death. The patient may advance into or recover from adjacent stages of disease or die at any stage of disease. Observations of the stage are made on a number of individuals at arbitrary time, which may be between individuals.

The Application of Markov Chain has been extensively reported in numerous literatures. The study will adopt the work of Arnoldo and Bernd in [8] who modeled Infectious diseases using Markov chain with four possible states. Their work has been modified into five possible states to suit the purpose of this study. Arnoldo and Bernd (2015) modeled infectious diseases with four possible states:

Susceptible (S), Infected (I), Immune (A), Dead (R). Possible transitions are from S to I, S or R; from I to A or R; from A to A or R; from R to R only. The transitions probabilities, from S to I, S to R and the loop S to S, must sum to one and can depend on characteristics of the individuals modeled, like age, gender, life style, etc. All individuals start in S, and move at each time unit (say a day). Given observations of the sequence of visited states (called trajectory) for a sample of individuals, with their personal characteristics, one can estimate the transition probabilities, by logistic regression, for example. This model assumes that the transition probability at time t from one state A to state B only depends on the state A and not on the trajectory that leads to A. This might not be realistic, as for example, remaining in the diseased state I over many days could increase the probability of transition to R. It is possible to model a system with longer memory, and thus leave the simplest setting of a Markov Chain (though one can formulate such a model still as a Markov Chain over a more complex state space which includes the length of stay in the current state).

After a careful study of their model, Arnoldo and Bernd (2015), this study will assume that modeling infectious diseases involves five possible states in order to get more efficient result. The possible states are Susceptible (S), Infected (I), Immune (A), Recovered (R) and Dead (D); Rather than the traditional SIAR model of four states.

2.2 Model Formulation

In this work, infectious disease will be modeled using Markov Chain by assuming five possible states. The possible states are Exposed (E), Infected (I), Immune (A) Recovered (R) and Dead;

Susceptible (S): Person that can be infected.

Infected (I): Transmittable stage of the infection.

Immune (A) Who have Immunity

Recovered (R) Who recover from the disease

Dead (R): Who will die of the disease

The possible transition from one state to another are as follow; from S to I, S to A or R or D, S to R, from S to I or R or D, S to A or R, from I to A or R , From A to A or R, R to R only, and lot more.

$$S = \frac{\sum Ps}{\sum PS + \sum PI + \sum PA + \sum PR + \sum PD}$$

$$I = \frac{\sum PI}{\sum PS + \sum PI + \sum PA + \sum PR + \sum PD}$$

$$A = \frac{\sum PA}{\sum PS + \sum PI + \sum PA + \sum PR + \sum PD}$$

$$R = \frac{\sum PR}{\sum PS + \sum PI + \sum PA + \sum PR + \sum PD}$$

$$SS = \frac{\sum PSS}{\sum PSS + \sum PSI + \sum PSA + \sum PSR + \sum PSD}$$

$$DD = 1 - P(S + E + I + A + \cdots \ldots \ldots \ldots \ldots)$$

$$P_{ij} = \begin{bmatrix} P_{SS} & P_{SI} & P_{SA} & P_{SR} & P_{SD} \\ P_{IS} & P_{II} & P_{IA} & P_{IR} & P_{ID} \\ P_{AS} & P_{AI} & P_{AA} & P_{AR} & P_{AD} \\ P_{RS} & P_{RI} & P_{RA} & P_{RR} & P_{RD} \\ P_{DS} & P_{DI} & P_{DA} & P_{DR} & P_{DD} \end{bmatrix}$$

The methodologies that this research work adopted include testing for the Markovian properties, the stationarity, and estimating the transition probability of the diseases using matrix multiplication method.

2.3 Tests for the Markovian Property

A discrete time and discrete state space stochastic process is *Markovian* if and only if the conditional

probabilities$P(X_{n+1}|X_0,\dots,X_n) = \frac{P(X_0,\dots,X_n,\ X_{n+1})}{P(X_0,\dots,X_n)}$ do not dependon $(X_0,\dots,X_n)$ in full, but only on the most recent state X_n:

$$P(X_{n+1}|X_0,\dots,X_n) = P(X_{n+1}|X_n) \quad (1)$$

The likelihood of going to any next state at time *n* + 1 depends only on the state we

Find ourselves in at time *n*. The system is said to have *no memory*, that is whether the transition probabilities are independent of the past, testing whether the movement of patients into a particular state depends only on the previous state occupied by the patient.

$$P(X_0,\dots\dots,X_T) = P\prod_{n=1}^{T} P(X_n\,|X_{n-1}) \quad (2)$$

Proof

$P(X_0,\dots\dots,X_T) = P(X_T|X_0,\dots\dots,X_{T-1})\ \mathrm{P}(X_0,\dots.,|X_{T-1})$

$$= P(X_T|X_0,\dots,X_{T-1}X_{T-2})P(X_T|X_0,\dots\dots,X_{T-2})\mathrm{P}(X_0,\dots.,|X_{T-2})=$$

$P(X_T|X_0,\dots\dots,X_{T-1})\mathrm{P}(X_{T-1}|X_0,\dots,X_{T-2})\dots\ P(X_2|X_0,X_1)\mathrm{P}(X_1|X_0)\ \mathrm{P}(X_0) = P(X_T|X_{T-1})$
$\mathrm{P}(X_{T-1}|X_{T-2})\ \dots\dots P(X_2|X_1)P(X_1|X_0)P(X_0)$

$$= P(X_0)\prod_{n=1}^{T} P(X_n\,|X_{n-1}).$$

We can also use the approach below to test for the Markovian property

$$H_0: P_{ij} = P_1 \text{ VS}$$

$H_1: P_{ij} \neq P_{0,}\ i = 1,2,3,\dots\dots m+1; j = 1,2,3,\dots\dots., m+1.$

The hypothesis will be tested by using Likelihood Ratio Criterion (Anderson and Goodman 1957) the criterion is given by

$\square = \pi\,(\frac{P}{P})^n$

$P = \frac{n_i}{n..}$

$P = \frac{n_i}{\sum n}$

$= \frac{n_{ij}}{n}$

n_{ij} Denotes the number of observed transition from ith state to jth state

The test statistic is given by

$$-2\ln\lambda = 2\ \sum_{i,j=1}^{6} n_{ij}\ln\left(\frac{n_{ij}n}{n_i n_j}\right)$$

$$= 2\ \sum_{i,j=1}^{6} \ln\left(\frac{p_{ij}}{p_j}\right) \sim \chi_{ij}^2\ ,\ \ \propto$$

Again, consider a sequence of random variables,$\{x_{n,} n \in T\}$ where T is a subset of the integers. We say that this sequence has the Markov property if,

For any n ∈T,

The future process$(X_{m,} m\ > n, m\epsilon\ T)$ is independent of the past process,

$(X_{m,} m\ < n, m\epsilon\ T)$ Conditionally on the present process.

2.4 Test for Stationarity

Definition: A (discrete-time) stochastic process $X_n: n \geq 0$ is stationary if for any time points $(i_{1,} \ldots., i_{n,})$and any$m\ \geq 0,$ the joint distribution of $(X_{i1,\ldots,}X_{in})$is the same as the joint distribution of. $(X_{i1+m,\ldots\ldots}X_{in+m})$So "stationary" refers to "stationary in time". In particular, for a stationary process, the distribution of $X_n$is the same for all.

2.5 Stationary Distribution

Let us remember that a time-homogeneous Markov chain at time is characterized by its distribution and that

$\pi^{(n)} = (\pi^{(n)}, i\ \in S)$, Where $\pi^{(n)} = P(x_n = i)$, and that

$$\pi^{(n+1)} = \pi^{(n)}p, \text{ i.e. } \pi_j^{(n+1)} = \sum_{i\in S}\pi^{(n)}\,\pi^{(n)}p_{ij,}\ \forall\ j \in S \qquad (3)$$

A distribution$\pi^{*} = (\pi_{i,}^{*} i\ \in S)$ = is said to be a stationary distribution for the Markov Chain $(x_{n,} n \geq 0)$if

$$\pi^{*} =\ \pi^{*}P, \text{ i.e. } \pi_{j,}^{*}\ = \sum_{i\in S}\pi_i^{*}\ p_{ij,}\ \ \forall\ j \in S \qquad (4)$$

NOTE:

π^{*} does not necessarily exist, nor is it necessarily unique.

If π^{*}exists and is unique, then it can be interpreted as the average proportion of time by the chain. It implies

$$E(T_i|X_0 = i) = \frac{1}{\pi_i^*}$$

Where $T_i = \inf(n \geq 0: X_n = i)$ is the first time the chain comes back to state i.

If $\pi^{(0)} = \pi_{,}^{*}$then $\pi^{(1)} = \pi^* P = \pi^*_{;}$

Likewise, $\pi^{(n)} = \pi^* P^n = \cdots = \pi_{,}^{*} \forall n \geq 0,$

The initial distribution of the chain is stationary, and then it remains stationary over time.

Besides if $\left(X_{n,} n \geq 0\right)$ is a sequence of identical independent distribution (i.i.d) random variables, then $P_{ij} = P(X_{n+1} = j|X_n = i) = P(X_{n+1} = j)$ does actually depend neither i nor n, so $\pi_i^* = P(X_n = i)$ which is also independent of$n$ is a stationary distribution for the chain.

$$\sum_{i\in S} \pi_{i,}^{*} P_{ij} = (\sum_{i\in S} \pi_i^*) P(X_n = j) = 1. P(X_n = j) \quad (5)$$

$\pi_{i,}^{(0)} = \pi_{,}^{*}$The chain is in stationary state from the beginning.

3. Conclusion

From the study, we have established that the Markov chain can be used as a mathematical model for infectious diseases. The study reveals that the past history of infectious diseases will affect the future only through the present state of infectious in diseases which is a strong property of the Markov chain.

4. Recommendation

The study recommends that government at all level need to do more in the area of sensitization and fight against infectious diseases in Nigeria.

References

[1] D. C. Lay, S. R. Lay and J. J. McDonald**.** "Finite-State Markov Chains" (Download Only) for Linear Algebra and Its Applications" in Linear Algebra and its Applications, 5th Edition, Pearson, United States; 2016.

[2] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy and E. Couto. "Multistate Models for Disease progression with classification error. Based on Markov Processes Method of Estimating Rates of Transition between Stages of Diseases". Journal of the Royal Statistical Society Series, 52(2), pp.193-209, July, 2003.

[3] J. V. Ross and T. Taimre. "The Analysis of Hospital Infection Data using Markov Chain Models"

Kings College, Cambridge, U.K. and Department of Mathematics, University of Queensland, Queensland, Australia, 2006.

[4] A. Eisenberg. "The Application of Markov Chain Monte Carlo to Infectious Diseases". Journal of Statistical Planning and Inference, Vol.6 (7), March, 2011.

[5] W.H.O. Manual; The public health management of chemical incidents (2009).

[6] K. Arnand, Y. Kapil, K. Manmeet and K. Rajesh. "Epidemology to public health intervention for preventing cardiovascular diseases: the role of translational research" Indian Journal of Medical Research, 132(5): pp. 643-650, Nov., 2010.

[7] Epidemic (August 4, 2019). in Wikipedia. Retrieved September 11, 2019 from http://en.wiki/Epidemic

[8] Arnoldo and Bernd, Markov Chains: Department of Biostatistics, University of Oslo, Norway and Department of Econometrics, Vrije University, Amsterdam, The Netherlands, 2015.