

A stationary distribution markov chain for prediction of Covid-19 pandemic in Nigeria

Abstract

The global world irrespective of whether been developed, developing and under-developed is at this time experiencing the corona virus (COVID-19) pandemic challenges which have affected the economy drastically. Mathematical models, combined with the prevalence data, have been used to study the pandemic spreading caused by the virus. This study proposes a stationary distribution Markov model by considering three States namely as active cases, recoveries cases and deaths cases to help in curtailing the spread of the COVID-19 pandemic. The proposed estimation procedure based upon the three states using transition patterns and the long-run stable transition probability obtained from the Markov model with Monte Carlo Simulation for generating random number and assigning numbers to the cases provides a convenient approach to estimate the case fatality rate of the COVID-19. By analyzing the data from the NCDC official website of the daily cases of Covid-19, result shows that active cases increases more at the initial stage compare to when adequate precautions, sensitization awareness and covid-19 vaccine was administered, after which the number of recoveries increases and it leads to decreases in the number of deaths cases. The stationary transition probability shows that the COVID-19 has 25.38% chances of being active while recoveries are 35.79% and Deaths is 38.82%. Government should intensify measures and policies in order to ensure that active cases reduce to the minimum.

Keywords: Markov chain, transition probability, stationary distribution, COVID-19, Monte Carlo simulation

Volume 11 Issue 2 - 2022

Nkemnole EB, Osubu I

Department of Mathematics, University of Lagos, Nigeria

Correspondence: Dr Bridget Nkemnole, Department of Mathematics) Statistics, Faculty Of Science. University Of Lagos Nigeria, Tel 07033835855, Email enkemnole@unilag.edu.ng, idowuosub@gmail.com

Received: February 23, 2022 | **Published:** April 27, 2022

Introduction

Coronavirus disease, or also known as Covid-19, is a new disease found in the late 2019. It caused by a virus called SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). It is a respiratory virus that is transferred via contact with an infected person through droplets when a person coughs or sneezes, or through saliva droplets. The main clinical manifestations of the infection are fever, fatigue, respiratory symptoms (mainly dry cough), and emergence of dyspnea. In some cases, it can cause pneumonia, severe acute respiratory syndrome, kidney failure, or even death. The World Health Organization (WHO) declared COVID-19 to be a pandemic on 11th March, 2020 when it became clear that the illness was severe and that it was spreading quickly among humans, mainly in China, since December 31, 2019. It started as an epidemic in mainland China being firstly discovered in Wuhan, the capital city of the Hubei province in February 2020. The virus afterwards spread to other provinces in China. Cases of infection have also appeared in other countries.

There are 206,900 cases confirmed by March 18, 2020 causing 8,272 death. It was predicted that the number of confirmed cases will continue to increase. On January 30, 2020, WHO declared this as Public Health Emergency of International Concern (PHEIC) [<https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>]. The European Centre for Disease Prevention and Control stated that as of 17th of June 2020; 8,142,129 cases of COVID-19 and 443,488 deaths have been reported worldwide since 31st December 2019. American continent was among those with utmost number of cases (3,987,543) with United States and Brazil the leading countries (2,137,731 and 923,189 respectively).

The outbreak of this virus infection has certainly affected people's life, economy and health. How long this situation will last and when

the disease will be controlled is a great concern to everyone. In the last few decades, there were several major outbreaks of infectious diseases, such as atypical pneumonia (SARS) in 2003, H1N1 influenza in 2009, and H7N9 influenza in 2013. It is important to develop predictive and warning capabilities for the disease pandemic. Mathematical models, combined with the prevalence data, have been used to study the dynamics, analyze the causes and key factors of the outbreaks, forecast the trend of disease spread, and provide optimal control strategies and measures.

Related literature

A lot of researches have studied pandemic spreading caused by virus with mathematical modeling. Aldila et al.,¹ proposed a SIR model for MERS-CoV. Yong² discussed a model about HIV. Sianturi³ applied stochastic processes to analyze the spread of dengue hemorrhagic fever virus. Wang et al.,⁴ used the stationary distribution Markov chain to optimize combinatory drugs. The prediction of electron condition on atomic orbitals and human-pose based on humans' body kinetic movement were given by Szczepanik and Mrozek⁵ and Chan et al.⁶ respectively.

A discrete mathematical model to study the transmission of SARS was formulated by Zhou and Ma.⁷ Their results conform to the data and show that timely quarantine and a high quarantine rate are important to the control of SARS. Chowell et al.⁸ and Lekone et al.⁹ developed ordinary differential equations and stochastic SEIR models to study the dynamics of infectious disease and the effect of control interventions, respectively. Both models used the outbreak of Ebola in the Democratic Republic of Congo in 1995 as a case study. Lately, the spread of COVID-19 have been studied by some researchers.¹⁰⁻¹⁶ Most of them estimated the basic reproductive number R_0 , an important parameter to evaluate the potential of viral transmission.

Consequently, this research consider an extension from several different approaches to proposing a stationary distribution Markov chain for long-term prediction of Covid-19 spread and to examine the prediction of infected people in long-term by analyzing the Covid-19 daily cases. The method provides an advantage in computational simplicity compared to simulation experiments. Important probabilistic properties of the model are deduced. We make use of the existing reported data and carry out parameter estimation and numerical simulation based on the data.

Methodology

Markov chain

A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. A countably infinite sequence, in which the chain moves at discrete time steps, gives a discrete-time Markov chain (DTMC). In this study, discrete-time Markov chain is used to analyze the dataset. The dataset is checked whether it is a positive recurrent, aperiodic and irreducible or not. This study aims to obtain the stationary distribution of the given data, which is obtained if the dataset satisfies all of the above conditions. Markov model have many applications as statistical models of real-world processes, such as studying cruise control systems in motor vehicles, queues or lines of customers arriving at an airport, currency exchange rates and animal population dynamics.

The principles of Markov are described as follows:

A Markov Chain consists of a countable set S (called the state space) together with a countable family of random variables X_0, X_1, X_2, \dots with values in S such that

$$P[X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \dots, X_0 = s_0] = P[X_{t+1} = s \mid X_t = s_t]. \quad (1)$$

Discrete-time markov chain: Let $(X_t, t=0,1,2,\dots)$ be a discrete-time stochastic process with time parameter $t=0,1,2,\dots$ and state space $s=0,1,2,\dots$, $X_t = i$ defines that the process is in state i at time t . If the probability in the future time $(t+1)$ in a state j is dependent only to the present condition in state i at the present time t , then the process is called a discrete-time Markov chain Azizah et al.¹⁷ denoted by p_{ij} , which define the transition probability from state i to state j .

The transition matrix P is obtained by collecting all the transition probability p_{ij} from every possible state i and j into matrix. The transition matrix P is expressed as:

$$P = p_{ij} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \dots \\ p_{10} & p_{11} & p_{12} & \dots \\ p_{20} & p_{21} & p_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (2)$$

where $p_{ij} \geq 0$ and $\sum p_{ij} = 1, \forall i, j = 0,1,2,\dots$. The process is called stationary distribution Markov chain, if the transition probability in the future is independent from the present.

The stationary distribution of a Markov chain describes the distribution of X_t after a sufficiently long time that the distribution of X_t does not change any longer. To put this notion in equation form, let π be a column vector of probabilities on the states that a Markov chain can visit. Then, π is the stationary distribution if it has the property.

$$\begin{aligned} \pi &= \pi p = \pi p^2 = \pi p^3 = \dots \\ \pi^T &= \pi^T p \end{aligned} \quad (3)$$

A Markov chain is **irreducible** if it is possible to reach any state from any other state.

A state has period k if, when leaving it, any return to that state requires a multiple of k time steps (k is the greatest common divisor of all the possible return path length). If $k = 1$, then the state is said to be aperiodic and a whole Markov chain is **aperiodic** if all its states are aperiodic. For an irreducible Markov chain, we can also mention the fact that if one state is aperiodic then all states are aperiodic.

Privault¹⁸ stated that, a state i is called aperiodic if and only if $d(i) = 1$, where $d(i)$ is given by:

$$d(i) = \gcd\{n \mid n \geq 1, p_{ii}^n > 0\} \quad (4)$$

$d(i)$ is the great common divisor (gcd) of all possible n which makes the process in state i going back to the same state i by n step.

A state is **recurrent** if we know that we will return to that state, in the future, with probability 1 after leaving it. That is, a state is recurrent if and only if

$$\sum_{n=1}^{\infty} p_{ii}^n = \infty$$

For a recurrent state, we can compute the mean recurrence time when leaving the state. Note that even if the probability of return is equal to 1, it doesn't mean that the expected return time is finite. So, among the recurrent states, we can make a difference between **positive recurrent state** (finite expected return time) and **null recurrent state** (infinite expected return time).

A state i is called positive recurrent if $\mu_i < \infty$, where μ_i denotes the average recurrent time of state i .

Monte carlo simulation

Monte Carlo simulation is a specialized probability application where the variables have been replaced with a random number generator. It is a statistical simulation method for iteratively evaluating a deterministic model using sets of random numbers as inputs.

Generating randomly the sequence and movement of states of the corona virus (COVID-19), state "i" is selected arbitrarily. The random values between 1 and 3 were produced by way of the use of random number generator which is also referred to as Monte Carlo Simulation. We established the cumulative probability transition matrix by using successive multiplication of P matrix. Thus, given the transition from the i^{th} row to j^{th} state as P_{ij} , then the cumulative probabilities is:

$$P_{ij} = \sum_{j=1}^i P_{ij} \quad (5)$$

Experimental result and discussion

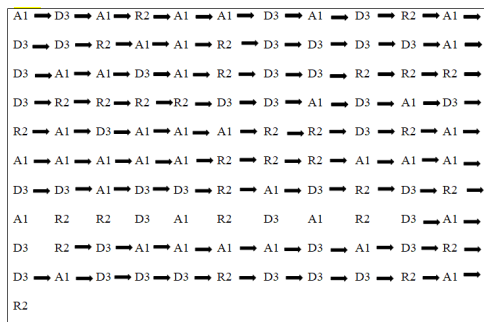
The proposed Markov model is applied to predict the number of daily total infected cases of COVID-19 in Nigeria respectively. It presents the analysis of data collected and interpretation of results. Data were analyzed using frequency counts and percentages to describe the characteristics of the cases. In addition; tables, charts, graphs, and trend analysis was used to explore the moving pattern of the cases, also Markov chain was adopted in obtaining the transition and absorbing state probabilities for the various cases.

Data

In order for reliability, validity and accuracy of results to be obtained from this study, secondary data were collected and extracted from Nigeria Centre for Disease Control (NCDC) official website on the 18th of June, 2021 such as total number of Corona virus patients due to the considered cases until absorbing states is reached for all the states in Nigeria and her capital.

Formulation of Transition (Movement) of the Covid-19 considered cases: (A=Active cases, R= Recoveries and D = Deaths)

Using the Monte Simulation approach by generating random number between 1 and 3 in excel spreadsheet, where 1 = A (Active cases), 2 = R (Recoveries) and 3 = D (Deaths), we have;



The possible states are Active cases (A), Recoveries (R), and Deaths (D); Active cases (A): Person that has been infected. Recoveries (R): Who recover from the disease and Deaths (D): person who will likely die of the disease.

$$P_{ij} = \begin{matrix} A \\ R \\ D \end{matrix} \begin{pmatrix} ARD \\ P_{AA} P_{AR} P_{AD} \\ P_{RA} P_{RR} P_{RD} \\ P_{DA} P_{DR} P_{DD} \end{pmatrix}$$

The probability transition matrix P describes the Markov chain representing three states of Corona virus (COVID-19) pandemic was obtained as:

$$P = \begin{pmatrix} 0.400 & 0.225 & 0.375 \\ 0.300 & 0.300 & 0.400 \\ 0.375 & 0.325 & 0.300 \end{pmatrix} \quad (6)$$

The cumulative probability transition matrix P_{ij} was obtained as;

$$P = \begin{pmatrix} 0.400 & 0.625 & 1.000 \\ 0.300 & 0.600 & 1.000 \\ 0.375 & 0.700 & 1.000 \end{pmatrix} \quad (7)$$

Monte Carlo simulation states are;

$$\text{State} = \begin{cases} A & \text{if } v > 0.400 \\ R & \text{if } 0.300 < v < 0.375 \\ D & \text{if } v < 0.300 \end{cases} \quad (8)$$

From equation (7), the transition chances shows that, there is 40% chance in the first level (Active stage) of the corona virus pandemic if $N = 166982$. In the second stage (Recoveries) we have 30% chances. Therefore, Corona virus recovered possibilities in Nigeria as at 18th June, 2021 is 30%. In the third stage (Death level), there is 37.5% of probabilities of occurring for the given statistics if $N = 166982$.

Equation (8) shows the result of the Monte Carlo simulation, the alternative of the corona virus (COVID-19), generated as the random range adjustments values and assumes the distinct states of corona infection; v is a random wide variety of uniform distribution.

By analyzing the state classification of this Markov chain, the result is that the Markov chain is irreducible, aperiodic, and positive recurrent. The stationary distribution Markov chain for daily cases of Covid-19 is:

$$\delta = \pi P$$

$$\text{Let } \delta = (\pi_1, \pi_2, \pi_3)$$

Therefore,

$$(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \begin{pmatrix} 0.400 & 0.225 & 0.375 \\ 0.300 & 0.300 & 0.400 \\ 0.375 & 0.325 & 0.300 \end{pmatrix}$$

$$\text{Recall that } \pi_1 + \pi_2 + \pi_3 = 1$$

Then, we have;

$$\pi_1 = 0.2538, \pi_2 = 0.3579 \text{ and } \pi_3 = 0.3882$$

$$\therefore \pi = (\pi_1, \pi_2, \pi_3) = (0.2538, 0.3579, 0.3882) v \quad \text{-----} \quad (9)$$

$$\therefore \pi = (\pi_1, \pi_2, \pi_3) = (25.38\%, 35.79\%, 38.82\%) \quad \text{-----} \quad \text{In percentage}$$

The result above reveals that based on the daily confirmed case from June 18, 2021, the long-term probability of Active cases will remain at 25.38%, The long term probability of Recoveries cases is 35.79%, while the long-term probability of Deaths cases will remain at 38.82%. This result may vary depends on the observation interval which can make different pattern of Covid-19 daily cases.

Table 1 shows the frequency and percentage of the Covid-19 Pandemic with respect to the cases of interest. This information is visualized for easy comprehension using some basic statistical charts such as pie chart and bar chart in the figure below: Figure 1.

Table 1 Showing the three considered Covid-19 Cases and their frequencies

Cases	Frequency	Percent	Valid Percent	Cumulative Percent
Active cases	163331	97.8	97.8	97.8
Recoveries	1534	.9	.9	98.7
Deaths	2126	1.3	1.3	100.0
Total	166991	100.0	100.0	

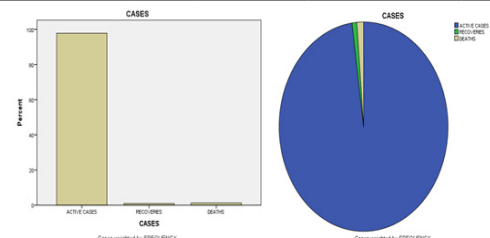


Figure 1 Using SPSS to get the Bar and pie chart output, it can be seen that the Active cases (A) of the Covid-19 is extremely high with about 98% while the other two states; Recoveries (R) and Deaths (D) has almost the same percentage, in which Deaths as 1.3% and Recoveries as 0.9%. These charts indeed conform and affirm the information gotten from the NCDC official website.

Figure 2, using SPSS trend output to see the pattern in which the Covid-19 pandemic cases in each state follows with respect to the considered cases of Covid-19 pandemic; Active cases (A), Recoveries (R) and Deaths (D). It can be deduced from the trend chart above that the cases could not be predicted as the trend line appears to be exponential in nature.

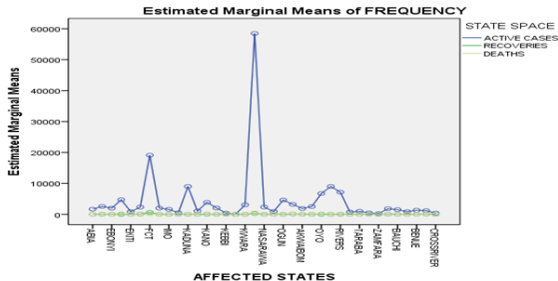


Figure 2 Trend chart representing the Covid-19 considered cases.

In Figure 3, we want to find out if we can predict and forecast the trend or pattern of the Covid-19 pandemics and to also see if the confirmed cases increases from months to months or otherwise.

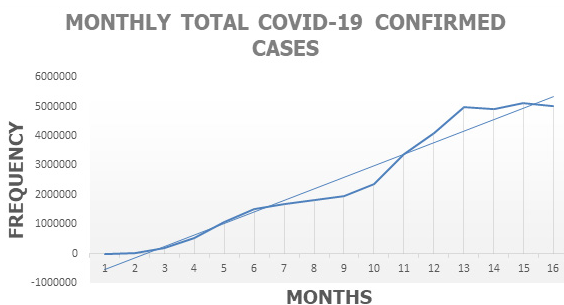


Figure 3 Trend chart of the Covid-19 monthly confirmed cases.

From the information gotten and visualized, it can be concluded that the total number of monthly confirmed cases increased systematically starting from the month of March, 2020 to the month of March, 2021. After which it dropped a bit and later increased before it later dropped in the month June, 2021 (the unit is month from March 2020 to June 2021). This could be caused by the fact that many Nigerians already taking the Covid-19 vaccines, it can be predicted that as time goes on, the total number of confirmed pandemic cases will reduce provided the masses take the vaccines and keep all NDDC laid down guidelines.

Conclusion

The research revealed that there are more active cases than both recoveries and deaths cases, that is, almost 97.8% of the patients were in active cases, 0.9% recovered from the virus while about 1.3% were recorded dead. Also, it was recorded that Lagos State had the highest number of Corona virus reported cases of about 35.5%, FCT had about 11.9% while other states had almost the same percentage of reported cases except Kogi State which had approximately 0% of the reported cases as at June, 2021. (see [Appendix](#)).

The analysis revealed that there is no significant relationship between the affected states and the state space (Active cases, recoveries and deaths) at 95% level of significance.

The Markovian Analysis also showed that at steady state; Active cases will remain at 25.38%, followed by Recoveries with 35.79%, while Deaths cases will remain at 38.82%.

Based on our findings, we can conclude that most of the states that recorded the highest number of reported cases has larger human

population compare to the remaining states, using Lagos and Kogi States as reference. Also, increase in the number of reported cases could be as a result of not compliance to the NCDC COVID-19 protocols by residents of the state.

From the trend line analysis, it can also be noted that number of reported cases dropped in the month of May, 2021, this could be as a result of residents compliance to NCDC laid down guidelines and the immunization of vaccine.

Recommendation

Prior to the findings and conclusions made, we can therefore recommend that:

Residents are advised to strictly comply with the NCDC COVID-19 protocol, adequate and proper sensitization about the administration of the COVID-19 vaccine should be done by the Government and the limiting behaviour of the cases should be an applicable measure to curtail any future occurrence.

This study results revealed that Markov Chain is beneficial in simulating the corona in numerous stages. This type of simulation could be very much useful in generating the time Corona Virus infection. The evaluation of corona infection indicates that Markov Chain approach is one opportunity of modeling in future. However, the research design, findings, conclusions and recommendations can be extended to other sectors whose movement of occurrence is unpredictable.

Conflicts of interest

There are no conflicts of interest.

Acknowledgment

None.

References

1. Aldila D, Padma H, Khotimah K, et al. Analyzing the mers disease control strategy through an optimal control problem. *International Journal Applied Mathematics Computer Science*. 2018;28(1):0013.
2. Yong B. Model Matematika penyebaran HIV dalam penjara. *Mipa dan pembelajarannya*. 2007;36(1).
3. Sianturi P. Model epidemik stokastik penyebaran demam berdarah dengue di jawa barat. *Prosiding seminar hasil-hasil PPM IPB* 1. 2018.
4. Wang B, Wang W, Wang Y, et al. A novel approach to optimize combinatory drugs using Markov chain. *IEEE 11th Annual International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*. 2016.
5. Szczepanik D, Mrozek J. Stationarity of electron distribution in ground-state molecular systems. *Journal of Mathematical Chemistry*. 51.2013.
6. Chan KC, Koh CK, Lee CG. Using action classification for human-pose estimation. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013.
7. Zhou Y, Ma, Z, Brauer F. A discrete epidemic model for SARS transmission and control in china. *Math Comput Model*. 2004;40(13):1491–1506.
8. Chowell, G, Castillo-Chavez, C, Fenimore, P, et al. Model parameters and outbreak control for SARS. *Emerg Infect Dis*. 2004;10(7): 1258–1263.
9. Lekone P, Finkenstadt B. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*. 2006;62(4):1170–1177.
10. Wu J, Leung K, Leung G. Now casting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. *Lancet*. 2020;395(10225):689–697.

11. Prasse, B, Achterberg, M, Ma, L, et al. Network-based prediction of the 2019-nCoV epidemic outbreak in the chinese province hubei, arXiv preprint arXiv 2002.04482.
12. Zhao S, Musa, S, Lin, Q, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) vases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak. *J Clin Med.* 2020;9:388.
13. Anastassopoulou C, Russo L, Tsakris A, et al. Data-based analysis, modelling and forecasting of the novel coronavirus (2019-nCoV) outbreak. *medRxiv.* 2020.
14. Yang, Y, Lu, Q, Liu, M, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China, medRxiv. 2020.
15. You C, Deng Y, Hu W, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int J Hyg Environ Health.* 2020;228:113555.
16. Tang B, Wang, X, Li, Q, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med.* 2020;9:462.
17. Azizah A, Welastika R, Falah A N, et al. An application of markov chain for predicting rainfall data at west java using data mining approach. *IOP Conference Series: Earth And Environmental Science.* 2019;303 012026.
18. Privault N. Understanding markov chains. Examples and applications (singapore: springer-verlag). 2013.