

Stochastic Modeling and Prediction of the COVID-19 Spread in Kenya

ABSTRACT

Since the inception of the novel Covid-19 in December in China, the spread has been massive leading World Health Organization to declare it a world pandemic. While epicenter of COVID-19 was Wuhan city in China mainland, Italy has been affected most due to the high number of recorded deaths as at 1st April, 2020 at the same time USA recording the highest number of virus reported cases. In addition, the spread has been experienced in many developing African countries including Kenya. While the Kenyan government have had plans for those who have tested positive through self-quarantine beds at Mbagathi Hospital, lack of a proper mathematical model that can be used to model and predict the spread of Covid-19 for adequate response security has been one of the main concerns for the government. Many mathematical models has been proposed for proper modeling and forecasting, but this paper will focus on using a generalized linear regression that can detect linear relationship between the risk factors. The paper intents to model and forecast the confirmed Covid-19 cases in Kenya as a Compound Poisson process where the parameter follows a generalized linear regression that is influenced by the number of daily contact persons and daily flights with the already confirmed cases of the virus. Ultimately, this paper should assist the government in proper resource allocation to deal with pandemic in terms of available of bed capacities, public awareness campaigns and virus testing kits not only in the virus hotbed within Nairobi but also in the other 47 Kenyan counties.

Keywords: Covid-19, Stochastic Modeling, Compound Poison Process, Generalized Linear Regression, Contact Persons



1. Introduction

COVID-19 infection is an infectious disease that is caused by the severe acute respiratory syndrome known as corona virus. While the virus was identified first in December 2019 in the city Wuhan, which is the capital of Hubei province in China [9], it now a global pandemic and has affected close to a million people. Many of the countries had made mild preparations knowing that the diseases would ultimately catch up with them [3] due to inadequate information on its short and long term effects. Some of the common signs of COVID-19 includes fever, shortness of breath and dry coughs [10]. Other uncommon symptoms includes muscle pains, mild diarrhea, abdominal pain, sputum production, loss of smell, as well as sore throat.

According to WHO statistics as at March 30, 2020, [6] the mortality rate of persons who had been diagnosed cases was on average of 4.6 percent and it ranges from 0.2 percent to higher level of 15 percent depending on the age group, which also depends on the health status of the predisposed person, location of residence as well as immunity system among other factors.

The diseases is mostly spread through via respiratory droplets, which are produced whenever a person coughs or sneezes. In addition, an individual can also contract the virus through touching a contaminated place or surface before touching their face[5]. This novel virus can easily survive on the ground or surfaces for up to a maximum of 72 hours; when a person is exposed, symptoms can easily show up from two to a maximum of fourteen days, with the expected mean number of days as five days.

Some of the recommended measures that can be taken as prevention of infection includes social distancing and frequent hand washing, covering sneezes and coughs with inner elbow, a tissue and keeping away unwashed hands from your face [6]. One can also use **mask** especially when one is a health practitioner, the public needs to maintain just general personal hygiene to avoid contamination with other people who are either symptomatic or asymptomatic stages.

As at 2nd April 2020, only seven African countries had not reported a case of corona virus (COVID-19). These countries includes Lesotho, Sierra Leone, South Sudan, Burundi, Malawi, São Tomé and Príncipe, and Comoros. This means that Kenya is among the most affected countries in Kenya with confirmed cases totaling to over 80 cases. On 12th March 2020, Kenya had recorded the first case, which was confirmed by the President. It was a young woman who had arrived in the country from USA via **UK in London**. While many measures had been taken by the government of Kenya to combat the spread of the virus among the citizens; some of these measures includes banning of traveling of undocumented foreigners from any country in the world with recorded cases of Corona virus. In addition, any Kenyan and permitted foreigner who

travels back to the country from the hard-hit countries to proceed to mandatory self-quarantine at a designated quarantine facility for 14 days before being released to the public.

All schools in Kenya as well as higher learning institutions including universities were closed as at 20th March, 2020. The government also advised its citizens to start working from their homes except for those offering essential services to the public. Most businesses are also going cashless transactions as opposed to cash after the massive reduction of the mobile money transaction's costs. Effective from 20th March, 2020, the government has also outlawed congressional meetings, which include attending weddings, bars and night clubs, visiting malls, attending churches, and visiting hospitals limitations among many other measures.

While these measures are taken to ensure that the government contains the spread the severity of the virus especially to the vulnerable citizens especially from the limited hospital capacities within the country, it is important to model the virus spread. This is because lack of proper modeling of the disease spread put a huge challenged especially when making appropriate mitigation measures in case Kenya becomes the "*Italy of Africa*" in terms of deaths recorded from the Covid-19 virus.

The aim of this paper is model the spread of Covid-19 virus thus enabling the government to make ready and proper preparations thus reducing Kenya becoming another "**hotbed**" of Covid-19 virus deaths in Africa. In addition, the results from paper after stochastically modeling of the virus spread [7]help in making predictions on the potential effects thus assisting the response agency deal with inadequate hospital and infrastructure development for safety of Kenyans at large. This is to prevent any form unnecessary death that is likely to be experienced during the period.

2. Stochastic Modeling Methods

2.1 An Overview of Covid-19 Conceptual Framework

In an overview of the Covid-19 Conceptual framework subsection, it is important to know how Covid-19 infections move from one stochastic state to another during the period of infection. In any given population, it comprises of three states of nature namely; Susceptibles (S), Infectious (I) and Recoveries (R) [1]. The states can be summarized in a conceptual framework as described the figure below;

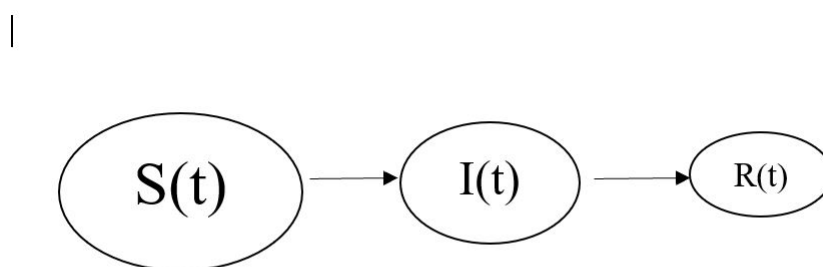


Figure 1: Conceptual Framework of Covid-19 Infections

From the above conceptual framework (Figure (1)), the Susceptible ($S(t)$) consists of people who have susceptible to the Covid-19 virus [2], the Infectious ($I(t)$) are people who have been infected with the virus while Recoveries ($R(t)$) are those who have either died or those who have recovered from the virus thus having a high level of improved immunity from the infections. It is important to make an assumption that an individual can move from state to another during the entire period of the virus existence.

2.2 The Compound Poisson Model

In this paper, we will model the number of confirmed cases of infections of Covid-19 in Kenya as a Compound Poisson Process. The model follows as;

$$Z_k \sim \text{Poisson}\left(\sum_{k=1}^{\infty} \vartheta_k \theta\right) \tag{0.0.1}$$

where the parameter $\vartheta_k \theta$ is the parameter mean of the number of confirmed Covid-19 cases. However, also follows a generalized linear regression distribution or a multiple linear regression as;

$$\sum_{k=1}^{\infty} (\vartheta_k \theta) = Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i \tag{0.0.2}$$

From the above equation (0.0.2) model, we can note that Z_k is the confirmed cases of Covid-19 in Kenya, where x_1 is the number of persons who made contact persons who have confirmed cases of the virus, x_2 is the number of daily flights to Kenya from infected countries and e_i is the error term, which is assumed to be *i.i.d.* and follows a Gaussian distribution as $(0, \sigma^2)$, i.e. $e_i \sim N(0, \sigma^2)$. The values of β_0, β_1 and β_2 are the parameters to be estimated from the available data according to Kenyan Covid-19 confirmed cases as at 3rd April, 2020.

2.3 Estimation of the Multiple Linear Regression Parameters

From the above equation of (0.0.2), it is easy to estimated the values of the parameters used on the model [8]. From the systems of two equations from the previous model,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i \tag{0.0.3}$$

$$Y = X\beta + e \tag{0.0.4}$$

where the values of Y, X, β and e are vectors of equation (0.0.3).

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ and } \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where X is called the **design matrix**, vector β contains all regression coefficients, which should be known through estimation of the ordinary least squares estimation method . Summing all the equations above, to obtain;

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

We obtain the vector β as;

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Making Y the subject of the equation;

$$Y' = X\beta'$$

The estimated regression equation is called a fitted model. It is important to estimate the values of the fitted model (Y') from observed values (Y) as e where $e = (Y - Y')$. This means that our model becomes;

$$Y' = X\beta' \tag{0.0.5}$$

The model can also be fitted using a *hat matrix* denoted by H using;

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$Y' = X\hat{\beta}$$

$$Y' = (X'X)^{-1}X'Y$$

$$Y' = HY \tag{0.0.6}$$

where $H = (X'X)^{-1}X'$ and can transform the values of observed values of Y to estimated values of Y i.e. Y' . Once the values of $\hat{\beta}$ vector are estimated, they can be used to estimate the values of Compound Lambda $\sum_{k=1}^{\infty} (\vartheta_k \theta)$ which is used in equation (0.0.1) above to model the value of Z_k which follows a Compound Poisson process [4] during valuation process.

From the Compound Poisson process, we can be able to model number of confirmed cases of infections of Covid-19 in Kenya before making predictions on what is likely to take place at daily discrete-time intervals when the announcement are made the ministry of health for proper planning in case the country needs a lock-down or not as a way of containing the virus from spreading to many rural parts of the country. The government can make necessary mitigation measures to deal with it in the long run.

2.4 Discrete Compound Poisson Distribution

From equation (0.0.1), Let X_1, X_2, \dots be non-negative integers which are identically and independently distributed random variables with $P[X_1 = k] = \vartheta_k$, for $k = 1, 2, 3, \dots$) A discrete random variable Z will satisfy a *pgf* characterization [11];

$$P[Z = z] = \sum_{k=1}^{\infty} P[Z = z] * z^i = e^{\sum_{k=1}^{\infty} \vartheta_k \theta (z^k - 1)} \quad (0.0.7)$$

where the above equation (0.0.7) varies for for the absolute varies of $z \leq 1$. The equation has a Discrete-Compound Poisson distribution with parameters $(\vartheta_1 \theta, \vartheta_2 \theta, \vartheta_3 \theta, \dots)$. This can also be defined by Z_k follows a Discrete-compound Poisson distribution with $(\sum_{k=1}^{\infty} \vartheta_k \theta)$. From the distribution, the mean and variance will be similar as $\sum_{k=1}^{\infty} \vartheta_k \theta$ i.e. $Mean = Variance = \sum_{k=1}^{\infty} \vartheta_k \theta$.

The unique feature about the distribution is that both the mean and variance is the same thus we will find it easy to use to model and forecast the potential Covid-19 infections in Kenya using the readily available data.

3. The Data

The data available at the ministry of health website(<http://www.health.go.ke/covid-19/>). From the available data, the first incidence of Covid-19 in Kenya was noted on 13th March, 2020. Let $Day(i)$ be the daily discrete announcement on confirmation of the Covid-19, $Y(t)$ is the confirmed cases in Kenya, $X_1(t)$ is the contact persons with the confirmed cases of virus, $X_2(t)$ is the number of daily flights to Kenya and $Q(t)$ is the confirmed cases of Covid-19 in Italy in Europe. The data is summarized below;

$Day(i)$	$Y(t)$	$X_1(t)$	$X_2(t)$	$Q(t)$
1	1	10	1	3
2	1	24	2	3
3	3	36	3	3
4	3	23	3	3
5	4	368	2	3
6	7	456	2	3
7	7	1065	3	3
8	7	345	1	3
9	15	356	3	3
10	16	276	2	3
11	31	1125	3	3
12	31	898	1	3
13	38	689	2	3
14	42	4545	3	3
15	50	8976	2	3
16	58	12349	2	3
17	81	13784	2	3
...

Table 1: Experienced number of contact Persons Vs Daily Flights in Kenya at as 1st April 2020

From the above daily extract sample (*Table (1)*), we can use R program to do an analysis thus estimating the values of the parameters for modeling and forecasting purposes. This should enable us communicate the results that has been modeled before forecasting is done.

4. Results

From the data analysis using R statistical program, we can tabulate the generalized linear regression equation as follows;


```

> glm(`Cases in Kenya` ~ `Contact Persons in Kenya` + `Flights to Kenya`)
Call:  glm(formula = `Cases in Kenya` ~ `Contact Persons in Kenya` +
`Flights to Kenya`)

Coefficients:
(Intercept)      `Contact Persons in Kenya`      `Flights to Kenya`
-8899.8822                0.2783                3308.5984

Degrees of Freedom: 39 Total (i.e. Null); 37 Residual
Null Deviance:      1.81e+10
Residual Deviance: 2.038e+09      AIC: 831.4
> glm(`Cases in Kenya` ~ `Contact Persons in Kenya`)
Call:  glm(formula = `Cases in Kenya` ~ `Contact Persons in Kenya`)

Coefficients:
(Intercept)      `Contact Persons in Kenya` |
-2338.0013                0.2652

Degrees of Freedom: 39 Total (i.e. Null); 38 Residual
Null Deviance:      1.81e+10
Residual Deviance: 2.218e+09      AIC: 832.8

```

Figure 2: Relationship between Covid-19 in Kenya and Contact Cases & Flights in Kenya

Summary of the trends of Covid-19 in Kenya are follows;

```

> summary(glm(`Cases in Kenya` ~ `Contact Persons in Kenya` + `Flights to Kenya`))
Call:
glm(formula = `Cases in Kenya` ~ `Contact Persons in Kenya` +
`Flights to Kenya`)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-16373.2  -3914.1   -603.2    5400.4   13218.2

Coefficients:
(Intercept)      Estimate Std. Error t value Pr(>|t|)
`Contact Persons in Kenya`  -8.900e+03  3.949e+03  -2.253  0.0302 *
`Flights to Kenya`         2.783e-01  1.722e-02  16.164 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 55084339)

Null deviance: 1.8103e+10 on 39 degrees of freedom
Residual deviance: 2.0381e+09 on 37 degrees of freedom
AIC: 831.37

Number of Fisher Scoring iterations: 2

```

Figure 3: Summary of Covid-19 in Kenya Results

From the above results, we can rewrite the equation (0.0.3) as;

$$Y = -8.9 + 2.783x_1 + 3.31x_2 \tag{0.0.8}$$

In terms of the degree of relationship between between the number of people who

made contact with other persons in Kenya.

```
cor.test(`Cases in Kenya`, `Contact Persons in Kenya`)
      Pearson's product-moment correlation
data:  Cases in Kenya and Contact Persons in Kenya
t = 16.497, df = 38, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8828412 0.9662821
sample estimates:
      cor
0.9367357
|
cor.test(`Cases in Kenya`, `Flights to Kenya`)
      Pearson's product-moment correlation
data:  Cases in Kenya and Flights to Kenya
t = -1.9665, df = 38, p-value = 0.05658
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.562200892 0.008386598
sample estimates:
      cor
-0.3039167
```

Figure 4: Relationship between Covid-19 Confirmed Cases and Contact Persons

From equation (0.0.8), we had used the equation to find the rates of cases of Covid-19 in Kenya. Once the values of the data is obtained from the analysis, the projected number of Covid-19 can tabulated as follows;

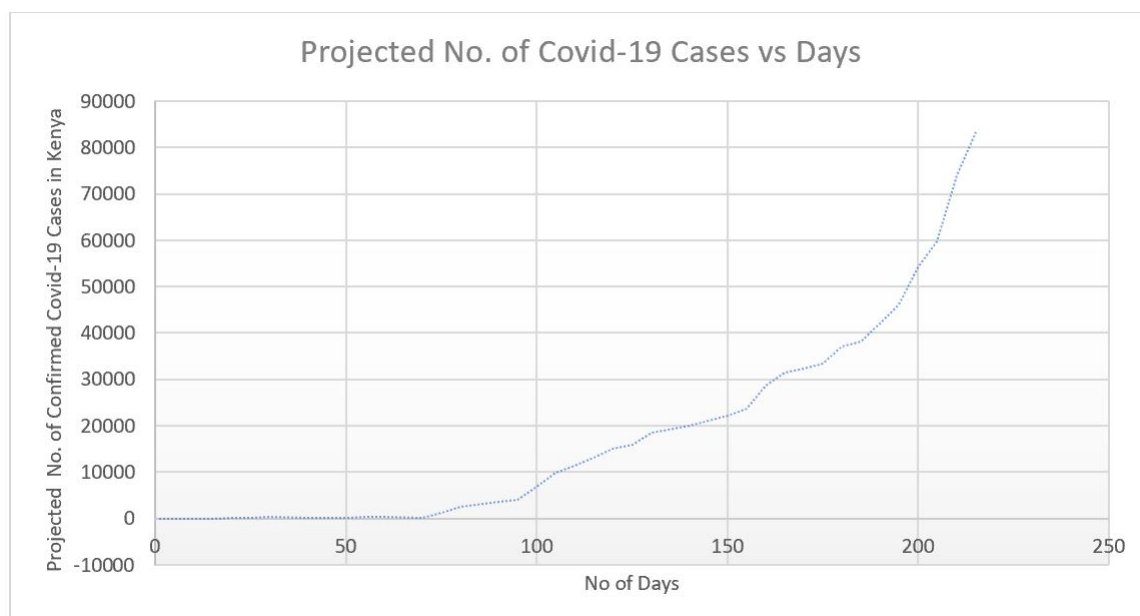


Figure 5: Projected Cases of Confirmed Covid-19 in Kenya

5. Conclusion

From the available data, we were able to find the correlation relationship between Covid-19 virus and contact persons made by the confirmed status. From figure 0.0.2, we are to obtain a positive relationship of Covid-19 cases and contact persons as well as the number of flights from the foreign countries that were getting to our country.

From the Summary of the trends of Covid-19 in Kenya as in figure 0.0.3, it is important to note that we have enough evidence to reject the null hypothesis that we stated that there is no correlations between the Covid-19 and number of flights from international countries that has been affected by the virus. This means that we can state that there is relationship between virus and number of flights that traveled back to Kenya. In addition, with the ban of the international flights from landing in Kenya unless those of Kenyans as well as those foreign citizens with documentation. With the correlation coefficient of 0.936736 from (figure 0.0.4), it means that the relationship is fully positive when looking at how the contact persons affected the confirmed cases of the virus in Kenya. From the multiple regression equation from equation (0.0.8), we can model the parameter, $\sum_{k=1}^{\infty} (\vartheta_k \theta)$ before modeling it to find the projected number of confirmed cases of Covid-19 in the next 250 days in case the virus still exists in Kenya from figure 0.0.5. We expect cases of confirmed cases of 83417.542 by the 235 day in Kenya when the current state is modeled using a Poisson distribution whose parameter is a generalized linear regression.

The above information should help the government make plans on how to deal with pandemic especially when dealing with the current situation in Kenya. From the data of Kenya when compared to Italy, Kenya might be the next “Italy of Africa” if necessary measures are not taken in consideration when dealing with the pandemic. The government must come up with more isolation beds, do more mass educations and campaigns with an aim of ensuring that the public have information about how to stop the spread of the virus.

When we consider that a huge population of the Kenyan population resides in rural areas, more education as well as infrastructure must be done in the rural to make preparations in case the Covid-19 finds its way to the villages. All other remaining 46 Counties of Kenya must ensure that they come with specific measures that would ensure that people do not spread the virus among themselves whenever they live in the villages.

The government should also open more testing places in the country to ensure that the general public can test themselves must earlier thus taking measures that may combat the spread of the virus in the villages of Kenya when compared to rates that have been experienced in Nairobi, which is the “hotbed” of the virus. With these preventive and

curative measures, the severity of Covid-19 will be limited when compared to other countries such as USA that are now leading in the number of infections in the world.

References

- [1] Leon Danon, Ellen Brooks-Pollock, Mick Bailey, and Matt J Keeling. A spatial model of covid-19 transmission in england and wales: early spread and peak timing. *medRxiv*, 2020.
- [2] Diego Giuliani, Maria Michela Dickson, Giuseppe Espa, and Flavio Santi. Modelling and predicting the spread of coronavirus (covid-19) infection in nuts-3 italian regions. *arXiv preprint arXiv:2003.06664*, 2020.
- [3] Zixin Hu, Qiyang Ge, Li Jin, and Momiao Xiong. Artificial intelligence forecasting of covid-19 in china. *arXiv preprint arXiv:2002.07112*, 2020.
- [4] Dimitris Karlis and Evdokia Xekalaki. Mixed poisson distributions. *International Statistical Review*, 73(1):35–58, 2005.
- [5] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 2020.
- [6] World Health Organization et al. Global surveillance for covid-19 disease caused by human infection with the 2019 novel coronavirus, interim guidance, 27 february 2020. 2020.
- [7] Jomar F Rabajante. Insights from early mathematical models of 2019-ncov acute respiratory disease (covid-19) dynamics. *arXiv preprint arXiv:2002.05296*, 2020.
- [8] Gülden Kaya Uyanık and Neşe Güler. A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, 106(1):234–240, 2013.
- [9] Yunlu Wang, Menghan Hu, Qingli Li, Xiao-Ping Zhang, Guangtao Zhai, and Nan Yao. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with covid-19 in an accurate and unobtrusive manner. *arXiv preprint arXiv:2002.05534*, 2020.

- [10] Fan Wu, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, Zhao-Wu Tao, Jun-Hua Tian, Yuan-Yuan Pei, et al. A new coronavirus associated with human respiratory disease in china. *Nature*, 579(7798):265–269, 2020.
- [11] Huiming Zhang and Bo Li. Characterizations of discrete compound poisson distributions. *Communications in Statistics-Theory and Methods*, 45(22):6789–6802, 2016.