

A linear study of the spread of COVID-19 in China and Iran

Mohammad Javad Sayadi^{1*}, Fateme Moghbeli¹, Hafez Mehrjoo², Mohammadreza Mahaki³

¹Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

²Doornama Company, Data Science Lab, Ilam, Iran

³Faculty of Human Movement Sciences, Vrije University of Amsterdam, Amsterdam, Netherlands

Article Info

Article type:
Research

Article History:

Received: 2020-03-12

Accepted: 2020-05-13

Published: 2020-05-17

* Corresponding author:

Mohammad Javad Sayadi

Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Email: sayadi.javad@gmail.com

Keywords:

Corona Virus Disease

Fast Fourier Transformation

Linear Regression

Prediction

ABSTRACT

Introduction: Studying trends in observed rates provides valuable information in terms of need assessment, planning of programs and development indicators of each country. The purpose of the present study was to apply the regression model and the Fourier series in terms of predicting the trends in growth and mortality rate of coronavirus disease.

Material and Methods: In this study, two linear analysis methods were used to predict the incidence and mortality rate of coronavirus disease in Iran and China. The methods used are linear regression and Fourier transform. The data used were collected by referring to the official media of the mentioned countries, the general form of which is a time series of the incidence and mortality rate in recent days and the model implemented to estimate the incidence and mortality rate for the coming days. Python programming language version 3.7 is used to implement models.

Results: The results of this study show that the rates of coronavirus disease incidence and mortality are still increasing. Meanwhile, the Fourier transform-based analytical method is more accurate than the linear regression method and on the other hand, the accuracy of both algorithms for predicting mortality was much higher than the prediction rate. This indicates that the mortality rate is higher than that of its linearity over time. The other point is that based on the results of this study, however, linear methods are very suitable for future prediction, due to the nature of epidemic diseases whose growth chart is nonlinear, linear methods cannot be used to predict the rate and mortality used in distant times.

Conclusion: The accuracy of the mathematics-based methods for predicting the trajectory of COVID-19 was really high. We predicted that the epidemics of COVID-19 will be high during 10 days. If the data are reliable and there are no second transmissions, we can accurately predict the transmission dynamics of the COVID-19 across the cities in China and Iran. The mathematics-inspired methods are a powerful tool for helping public health planning and policymaking.

Cite this paper as:

Sayadi M, Moghbeli F, Mehrjoo H, Mahaki M. A Linear Study of the Spread of COVID-19 in China and Iran. Front Health Inform. 2020; 9: 32. DOI: [10.30699/fhi.v9i1.221](https://doi.org/10.30699/fhi.v9i1.221)

INTRODUCTION

One of the most important components in any community health planning is to determine the incidence of diseases in that community. Knowing the pattern of changes in disease incidence in each country can be of great importance for national planning.

Public health organizations believe that monitoring the process of disease incidence, mortality and health risk factors may contribute to adverse health events.

Tracking changes in observed incidence or prevalence rates provides valuable information for need assessment, planning, review of programs and development indicators of each country. Studying data over time can also play a role in predicting the magnitude and frequency of future events.

In December, 2019, COVID-19 virus which slipped from animals to humans in China caused an outbreak of respiratory illness [1].

Actually, Iran reported its first confirmed cases

of CoV-2 infections on 19 February 2020 in Qom. The kind of pneumonia caused by the novel coronavirus disease (COVID-19) is a highly infectious disease, and the ongoing outbreak has been declared by WHO as a global public health emergency [2]. Coronaviruses are a group of viruses that make diseases in mammals and birds. In humans, coronaviruses cause respiratory tract infections that are mild, like the common cold, though rarer forms such as SARS, MERS, and COVID-19 can be lethal [3].

According to ARCGIS, 105427 confirmed cases and 3583 deaths in around 101 countries have been registered. Among these countries China, South Korea, Italy, Iran and Germany have the most confirmed cases [4].

Linear regression, known as piecewise regression, least squares regression, two regressions, two- or multiple regressions, 3-step, and broken-line regression, is one of the methods of regression analysis in which the independent variable is divided at intervals and for each interval one. The regression line is fitted separately and is called the boundary between the fracture points. Linear regression is used to express continuous breakpoints in disease mortality and incidence rates [5].

There are several predictive methods [6, 7], one of which is linear regression, which is a useful tool for describing trends in data changes, especially data on disease incidence or mortality. As the trend of disease pattern changes in the country changes on an annual basis, therefore, to study these trends, the use of mathematical models can often be problematic and provide new insights into medical issues [8].

The application of intelligent technologies in clinical decision making have started playing a vital role in improving the quality of patients' life and helping in decreasing cost and workload involved in their daily healthcare [6].

A study by Lau and Gange et al in 2003 examined the trend of decreasing lymphocyte counts and hemoglobin concentrations as a precursor to AIDS in men with HIV infection using single regression to evaluate the above trend [9]. Cutoff regression has had many applications in the medical sciences, biology, animal and plant sciences and genetics, and the problem of estimating the number of change points in a piecewise regression model was expressed by Kim in 2000. In 2000, Kim applied a series of permutation tests to determine the number of unknown change points in a single regression [10].

In this paper, a fast Fourier transform-coupled machine learning based ensemble model is adopted for predicting patients with coronavirus disease based on the analysis of their medical data during ten days. The purpose of this article is to predict the growth rate and mortality rate of patients with coronavirus disease using linear and time series

regression.

MATERIAL AND METHODS

The model consists of two popular models: linear regression (LR) and fast Fourier transform (FFT). In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. A FFT is an algorithm that computes the Discrete Fourier transform (DFT) of a sequence, or its Inverse (IDFT). Fourier analysis converts a signal from its original domain (often time or space) to a representation in the frequency domain and vice versa [8, 11, 12].

The model of predicting the incidence of coronavirus disease indifferent countries are studied using regression model and Fourier series. The dataset is a series of times the number of officially confirmed deaths is reported on the WHO web site (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>). In this paper, we use linear regression and Fourier analysis and make predictions for the ten days.

In this paper, linear regression is not applied to data whose growth is exponential, thus using the fast Fourier series. The main purpose of the study is to investigate the coupling of the fast Fourier transformation with a linear regression in predicting COVID-19. In this section, an overview is first presented on the architecture of the recommender system followed by the detailed discussion on the fast Fourier transformation and a linear regression, two major technical components of the system.

Linear regression (LR)

Linear regression is the first kind of regression analysis to be studied rigorously, and to be used extensively in useful applications. This is because models which depend linearly on their unknown parameters are better to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to characterized [13].

Fast Fourier Transform (FFT)

The FFT is an efficient technique to compute the Discrete Fourier Transform (DFT) and its inverse. It is like the wavelet transformation as a windowing technique. The DFT decomposes the input data sequence in order to extract frequency information for the purpose of predicting the patient's condition one day in advance [12, 14]. In this paper FFT formulas were used as following (Eq1, Eq2) [15]:

$$a_n = \frac{2}{p} \int_p^\infty s(x) \cdot \cos\left(2\pi x \frac{n}{p}\right) dx \quad (1)$$

$$b_n = \frac{2}{p} \int_p^\infty s(x) \cdot \sin\left(2\pi x \frac{n}{p}\right) dx \quad (2)$$

If $s(x)$ is P -periodic; then any interval of that length is sufficient.

a_0 and b_0 can be reduced to $a_0 = \frac{2}{p} \int_p^\infty s(x) dx$ and $b_0 = 0$.

Many texts choose $P = 2\pi$ to simplify the argument of the sinusoid functions (Eq3):

$$s_N(x) = \frac{a_0}{2} + \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nx}{p}\right) + b_n \sin\left(\frac{2\pi nx}{p}\right) \right) \quad (3)$$

RESULTS

The data used for analysis in this article were collected through official statistics of two countries, China and Iran, up to the date of 7th of March 2020, and therefore based on the properties of the data used in the form of a time series, are the basis of all results and discussions. This article deals with this hypothesis. Fig 1 to 4 show the prediction rate of coronavirus disease in the next 10 days.

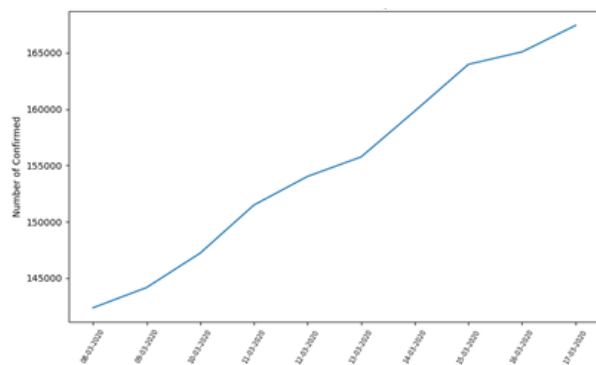


Fig 1: Prediction of coronavirus disease incidence in China in the next ten days using the Fourier transform method

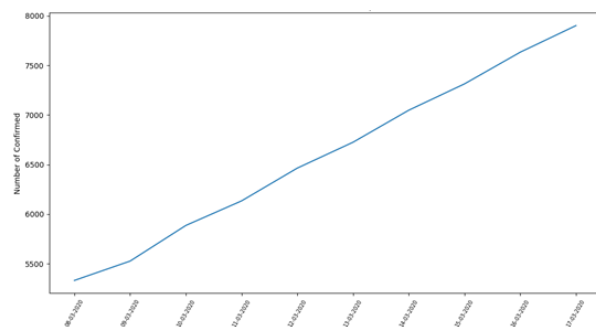


Fig 2: Prediction of coronavirus disease incidence in Iran in the next ten days using Fourier transform method

Fig 1 to 4 show the growth of COVID-19 in both Iran and China. At the same time intervals, China's rate of growth is much higher than Iran, which could be due to China's large population or the time difference between the two countries when the outbreak began. The outbreak of this disease in Iran started two months later and we can say that we are at the beginning of its growth chart. On the other hand, as

the outbreak of the disease has begun in China, awareness has been very effective in reducing the rate of infection in other countries.

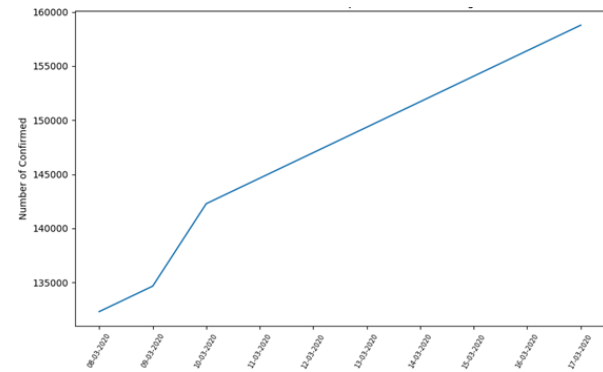


Fig 3: Linear regression prediction for China in the next ten days

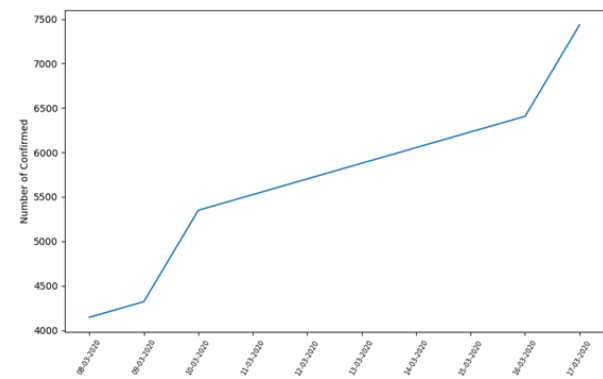


Fig 4: Linear regression prediction of coronavirus disease in Iran in the next ten days

As it is clear from these two graphs, the rate is increasing linearly over the next ten days. Given the high prevalence of infection in China in recent days, the implemented model predicts a higher incidence of infection in China. What is evident from the linear model test is that the prediction performs very well for the near days and the error rate increases rapidly for the third day onwards. The reason for this is due to the non-linear nature of the growth of epidemic diseases. Both the Fourier transform method and the linear regression predict the day ahead with high precision data analysis, and they can be advanced day by day. Meanwhile, the Fourier transform method was more accurate than the linear regression method and its error rate was relatively lower for the following days.

Fig 5 to 8 show the prediction of mortality rates for coronavirus disease in the next ten days using either Fourier transform or linear regression.

The diagrams in Fig 5 to 8 show an increase in mortality from coronavirus disease. In comparison, it can be said that the death rate in China is higher than that of Iran and this can be influenced by several factors. The important point in these graphs is that

the models implemented for the first and second days had a much higher prediction accuracy than the next days, but the accuracy of the linear models was far better than the predicted mortality.

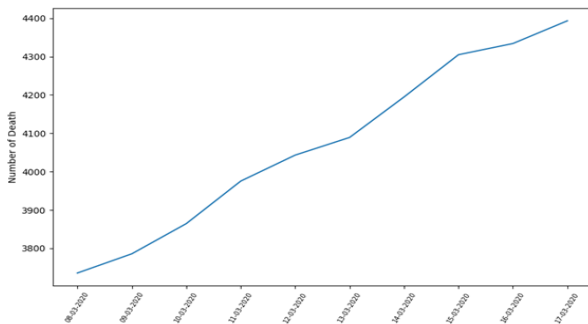


Fig 5: Prediction of coronavirus disease death in China in the next ten days using the Fourier transform method

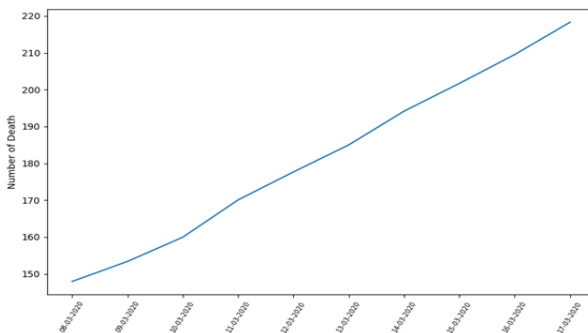


Fig 6: Prediction of coronavirus disease death in Iran in the next ten days using Fourier transform method

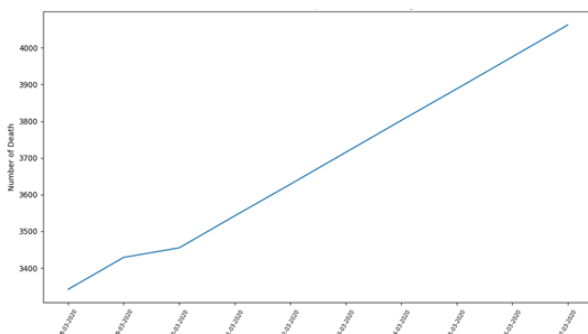


Fig 7: Linear regression prediction for China in the next ten days

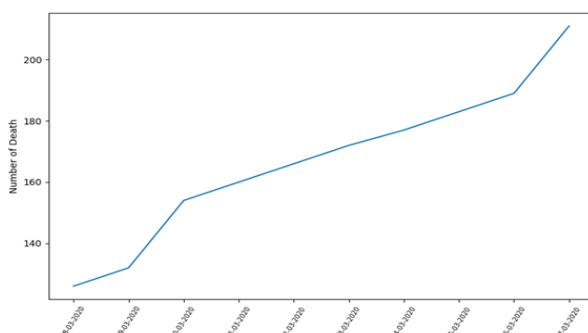


Fig 8: Linear regression prediction of coronavirus disease in Iran in the next ten days

Examination of mortality charts shows that unfortunately, mortality has increased linearly over time and, despite the awareness of many communities, its growth rate has not yet slowed down. Forecasts show that China has seen a slowdown in its growth rate through a period of exponential growth in the incidence of deaths from COVID-19, but Iran has for several reasons that may be due to its relative readiness to cope with the disease. There was a beginning of the outbreak, a linear behavior from the beginning, and this linear increase continues. The mortality rate of the disease in Iran is expected to be much lower in the same period than in China.

DISCUSSION

Results showed that linear models can predict incidence and mortality of coronavirus for a near future. When we talk about a prediction over a long period of time, these models are unable to predict in an acceptable distance to the real values. FFT predicts better than linear regression has a better accuracy. Both FFT and Linear regression have predicted that the growth rate of incidence and mortality in China is more than Iran. This can be resulted from several reasons. Coronavirus started in Iran two months after China so Iranian had more readiness than Chinese when they encounter the first infected case.

Population density, public health, public access to medical centers and the power of government in facing with coronavirus effect on the results. This means that some factors have linear behavior and some of them change the results in nonlinear way.

According to the results, the number of incidence and mortality will exceed grow rapidly based on the past reports. A very rapid and effective way is needed to reduce or stop the growth rate of incidence and mortality of coronavirus. Staying at home and strict rules must be applied beside the diagnosis and recovery of infected patients.

CONCLUSION

This paper presents two methods of linear modeling of Fourier transform and linear regression to predict the incidence and mortality rate of coronavirus disease in both China and Iran.

The results show that linear methods can predict coronavirus disease rate and mortality rate in near future with good accuracy. The overall growth chart for epidemic diseases shows that the rates and deaths from epidemic diseases such as corona are non-linear and that linear methods can work correctly for near-prediction times. When we shorten the time interval sufficiently, the graph behavior becomes closer to the linear model, and therefore the prediction is coupled with possible linear methods and with good accuracy. Nonlinear methods are proposed to predict the incidence and mortality of coronavirus disease over

longer periods.

AUTHOR'S CONTRIBUTION

The authors agree on this final form of the manuscript, and attested that all authors contributed in the final draft of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

- Hu Z, Ge Q, Jin L, Xiong M. Artificial intelligence forecasting of COVID-19 in China. *arXiv*. 2020; 2002: 07112.
- Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *The Lancet*. 2020; 395(10226): 809-15. PMID: 32151335 DOI: 10.1016/S0140-6736(20)30360-3 [[PubMed](#)]
- Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*. 2020; 295(3): 715-21. PMID: 32053470 DOI: 10.1148/radiol.2020200370 [[PubMed](#)]
- World Health Organization. WHO coronavirus disease (COVID-19) dashboard [Internet]. 2020 [cited: 1 Mar 2020]. Available at: <https://covid19.who.int>.
- Hope TM. *Linear regression: Machine learning*. Elsevier; 2020.
- Langarizadeh M, Moghbeli F. Applying naive bayesian networks to disease prediction: A systematic review. *Acta Inform Med*. 2016; 24(5): 364-9. PMID: 28077895 DOI: 10.5455/aim.2016.24.364-369 [[PubMed](#)]
- Moghbeli F, Langarizadeh M, Yoonesi A, Radmard AR, Rahmanian MS, Orooji A. A method for body fat composition analysis in abdominal magnetic resonance images via self-organizing map neural network. *Iranian Journal of Medical Physics*. 2018; 15(2): 108-16.
- Bonakdari H, Pelletier J-P, Martel-Pelletier J. A reliable time-series method for predicting arthritic disease outcomes: New step from regression toward a nonlinear artificial intelligence method. *Comput Methods Programs Biomed*. 2020; 189: 105315. PMID: 31972347 DOI: 10.1016/j.cmpb.2020.105315 [[PubMed](#)]
- Lau B, Gange SJ, Phair JP, Riddler SA, Detels R, Margolick JB. Rapid declines in total lymphocyte counts and hemoglobin concentration prior to AIDS among HIV-1-infected men. *AIDS*. 2003; 17(14): 2035-44. PMID: 14502006 DOI: 10.1097/00002030-200309260-00004 [[PubMed](#)]
- Kim H-J, Yu B, Feuer EJ. Selecting the number of change-points in segmented line regression. *Stat Sin*. 2009; 19(2): 597-609. PMID: 19738935 [[PubMed](#)]
- Zhang J, Lafta RL, Tao X, Li Y, Chen F, Luo Y, et al. Coupling a fast fourier transformation with a machine learning ensemble model to support recommendations for heart disease patients in a telehealth environment. *IEEE Access*. 2017; 5: 10674-85.
- Younce JR, Campbell MC, Perlmutter JS, Norris SA. Thalamic and ventricular volumes predict motor response to deep brain stimulation for Parkinson's disease. *Parkinsonism Relat Disord*. 2019; 61: 64-9. PMID: 30527905 DOI: 10.1016/j.parkreldis.2018.11.026 [[PubMed](#)]
- Yan X, Su X. *Linear regression analysis: Theory and computing*. World Scientific; 2009.
- Ozaktas HM, Kutay MA. The fractional Fourier transform. *European Control Conference*. 2001; IEEE.
- Boche H, Pohl V. On approximations for functions in the space of uniformly convergent Fourier series. *Journal of Approximation Theory*. 2020; 249: 105307.