Predicting Pandemic Curve Distribution Using Statistical Models

Eman Almuhur[1], Mona Khandaqji[2], Manal Al-labadi[3], Anwar Alboustanji[4]

[1]Assistant Professor, [2]Associate Professor, Department of Basic Science and Humanities, Faculty of Arts and Science, Applied Science Private University, Amman Jordan, [3]Assistant Professor, Mathematics Department, Faculty of Science, University of Petra, Amman, Jordan, [4]Lecturer, Department of Mathematics, Faculty of Basic Science and Humanities, German Jordanian University, Amman Jordan

Abstract

This article discusses the distribution of pandemic in the world and pandemic curve in Jordan and how the science of probability and statistics predict when active cases tend to zero by determining the shape of epidemic curve and relating it to a special probability distribution that has specific measures and properties. At the beginning of the outbreak of any virus in a society, reliable data describing it and its spread will be rare, hence researchers set up statistical models that have the ability to predict the spreads' shape, where the prospected people hosting such viruses will go to and the likelihood of transmitting it to places they travel. Those models use known statistical measures that estimate the probability of disease transmission from infected people to others. In addition, the factors related to roads and people's movement, taking into consideration, public health interventions, such as wearing masks, closing places of people's aggregations like schools, universities mosques and churches and quarantine make difference in numbers of infected people. The fundamental differences between the "Spanish flu" that attacked the world a hundred years ago and "Coronavirus" the world facing since the beginning of the current year 2020 is the amount of huge data concluded from scientific studies and reports related to virology and epidemiology.

Keywords*: Coronavirus, Pandemic Curve Distribution, Exponential Distribution.*

Introduction

Nowadays, it is possible to track the increasing number of scientific searches day after day using available search engines that may serve as indicators of the imminent emergence of a specific epidemic in a certain region.

Data Science mainly has three packages of skills, science and knowledge, starting with algebra, statistics and probability, where such data related to the nature of the field we are studying are displayed, analyzed and decisions are taken based on obtainable data. Lots of researches working on collecting, showing and analyzing genomic data for the evolution of "Coronaviruses" helped in predicting the next mutation of the virus. Such researches considered as sources of people where such researches facilitate data and raise up levels of understanding of the epidemic.

Typically, the lack of data is worse than having some inaccurate data. Ali Abdel-Hadi, a professor of mathematics and founder of the Doctoral Sciences Program at the American University in Cairo, says: "The main data must be available first, for data science to play its role, specifically in two spatio-temporal scales, and then the basic variables, such as injuries, recoveries and deaths, are added" [1]. Statistical models can be used to gauge the effectiveness of government actions and measures. It can also be used to explore "what if?" Scenarios.

Material and Methods

A statistical function which describes likelihoods and possible values that any random variable can take within a given interval (era) is a probability distribution that we depend on in our study. Such interval is bounded by the minimum and maximum values, but precisely number of factors (measures) force possible value to be plotted on the probability distribution. Such measures are: mean, median, mode, standard deviation, interquartile range, skewness and kurtosis. The most common distribution is the "Gaussian-normal" distribution which has a bell shaped curve. Now to interpret data generated by some phenomenon, we dictate its probability distribution which is called the probability density function or simply (pdf). Our methodology in this article depends on the probability distribution the pandemic curve has which leads to determine the passage of the pandemic and track when it increases and decreases and hence, when it tends to zero.

Findings

The notion of data science unifies each of statistics, informatics, data analysis.

Probability and statistics besides data, scientific methods and processes used to get extract knowledge about the spread patterns of viruses help in giving alarm if viruses spread reached its maximum, hence such methods are necessity in each country. Phenomena like viruses spread have to be understood and analyzed. Everything concerning science changes since the impact of information technology is becoming huge.

Spanish Flu

In 1918, the Spanish flu which is the most severe disease outbreak in human history, reached remote areas days before global air travel, and caused between 20 and 50 million deaths worldwide. The epidemic began in Spain after huge religious celebrations in Madrid in May and after few days, the disease spread an unprecedented spread. The second epidemic wave came in September during harvest celebrations and it was more mortal than the first one where governments at that time were easing prevention procedures. The end of the epidemic happened because the societies began to form general immunity that help each one to resist such epidemic and that generates the so-called herd immunity. "According to the results of seasonal influenza, we see that influenza the is the second most studied virus in the world, after HIV" [2]. If we take a look at the figure below, we notice that The curve of the epidemic displays, with its two parts, ascending and descending, is the movement of community members in which the epidemic is spreading between three different sections:

Susceptible, Infected and Recovered.

The Stages of Epidemic:

1) This stage starts with the entry of the epidemic virus into a community of healthy susceptible individuals and since it is a virus with an epidemic spread, each infected person will transmit the infection to more than one case per day.

2) If the infected person transmits the infection to two other people, then on the next day the number of cases will be two, then on the following day it will be four, then eight, the next will be 16, then 32, and so on according to the "geometric sequence" $2,4,8,16,32,\ldots = 2^n 2,4,8,16,32,\ldots = 2^n$ that has a base (ratio) 2 as shown in figure 1 below [11], where "the exponential function with unending succession of numbers is increasing".

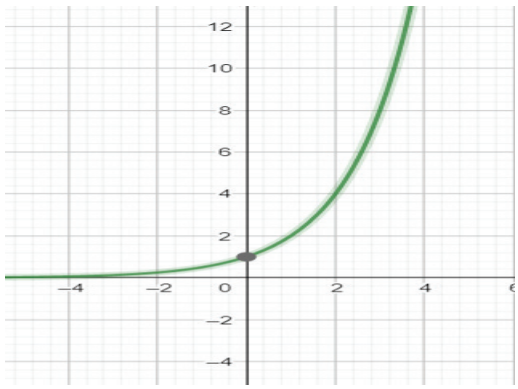


Figure 1

The geometric series of this sequence is: $\sum_{n=0}^{\infty} 2^n = 2 + 4 + 8 + \cdots + .. = \infty$ because it is a divergent (endless) series not convergent.

"When numbers of infected people start to descend, we realize that we left the summit of the curve and we are going gradually to the final stage of the epidemic wave where the geometric series became convergent" [5].

Increasing numbers of recovered people who passed the disease successfully or received vaccinations forces the epidemic curve to decrease gradually. According to probability distributions, we are talking about "log-normal distribution" which is "a continuous probability distribution with curve" shown in figure 2 below [4] where this curve represents a single epidemic wave. Such distributions have symmetrical or forms a bell shaped and 68% of data falls within one deviation about the mean and 95% of data falls within two deviations.

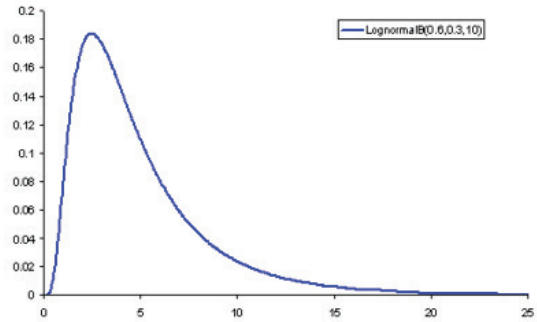


Figure 2

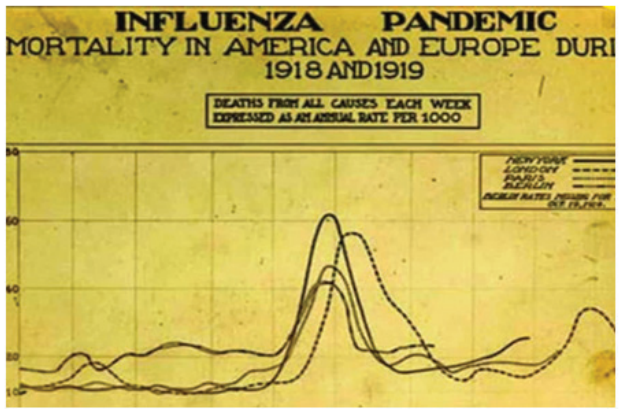
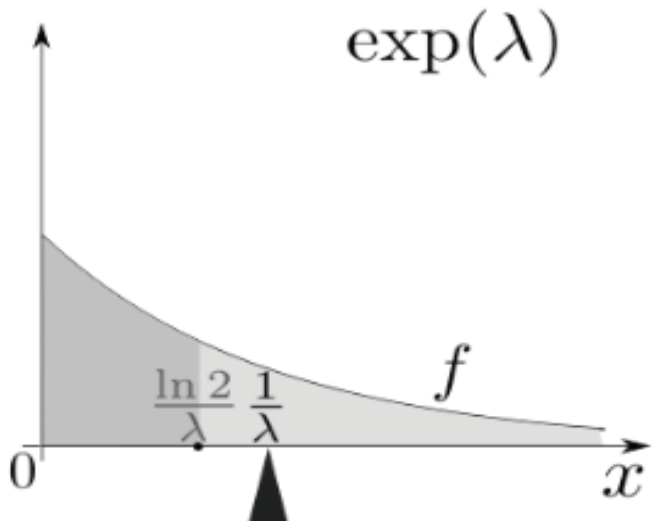


Figure 3

Figure 3 [12] shows the "spread of Spanish Flu in 1918-1920".

Exponential Distribution



The exponential distribution model is vastly used for products with constant failure or arrival rates. The most

featured thing when taking about it is the exponential probability density function (pdf) in which independent events occur at a constant rate. "Such distribution is used to model the time between independent events that happen at the average with a constant rate" [6].

"A random variable (denoted by X) is a role that associates to each outcome ω of an experiment a real number (denoted by $X(\omega)$), and "the probability density function (denoted by pdf $f(x)$) is a role that associates to each real number $X(\omega)$ a probability $P[X = x]$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & if\ x > 0 \\ 0 & if\ x \leq 0 \end{cases}$$

Where $\lambda > 0$ is a rate parameter" [7].

The expected value of a random variable X that represents number of infected people is given by $E(X) = \frac{1}{\lambda}$ and the standard deviation from the average $E(X)$ is given by $\sigma = \frac{1}{\lambda^2}$ " [8]. These two measures give us an idea that the pandemic curve is decreasing considering λ a constant that takes different values .

Explaining the Curve

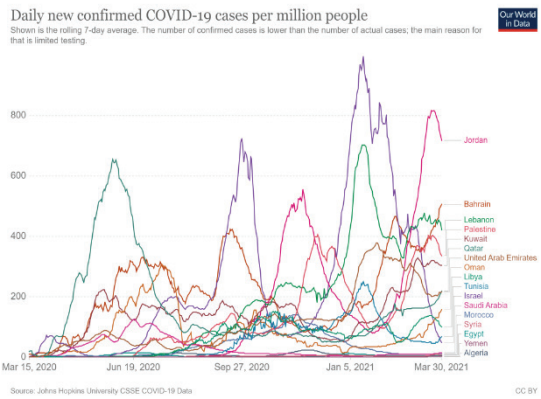


Figure 5

"The equation of the pandemic curve might not be familiar; its parameters are understood. The best description of it is given by: $f(x) = ke^{-\frac{(t-m)}{2w^2}}$ $f(x) = ke^{-\frac{(t-m)}{2w^2}}$

where $f(t)f(t)$ is the number of active cases per day, tt is the time, mm represents the day on which maximum kk occurs where a constant $k\,k$ is the maximum number of active cases recorded per day, w is the width of the curve at half k, and is given by:

Full Width at Half Maximum (FWHk) divided by $2\sqrt{2ln2}$ " [2].

We obtain "pandemic curves that display and analyze pandemic path from WHO statistics" [10] Such curves imply at different stages of the pandemic comparisons across countries and display different policies in many regions.

According to Jordanian Ministry of Health, "the pandemic curve after two waves till April, 2021 is shown in figure 6 "[13]

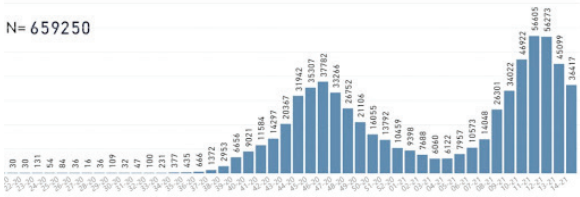


Figure 6

We notice that an extreme decline in the epidemiological curve in Jordan after witnessing a second wave of pandemic and The government has taken severe procedures to limit the spread of the epidemic and to treat the infected people and isolate them till they get well.

End of Epidemic

Typically, we found that the shape of the epidemic curve takes the exponential distribution and the spread of the epidemic is according a geometric sequence, limit of this distribution as time approaches infinity is zero, mathematically translation of this is:

$$\lim_{t \to \infty} ke^{-\frac{(t-m)}{2w^2}} = 0$$

According to a writer Gina Colata in the "New York Times" [3], historians noticed that epidemics have two types of endings: the first one is the medical end that occurs when infection and death rates decrease. The second one is the social end that occurs when the fear of disease disappears.

Before the Spanish flu vanished, it killed 20 million to 50 million people around the world, old and young and soldiers, in the midst of World War I. After it swept the world, the Spanish flu virus faded away and evolved into a different form of the milder flu that occurs every year. The pandemic is also socially over. After World War I ended, people were ready for a new era and they were eager to put the disease and war behind them.

Comparison with other Diseases

Deaths of Corona virus exceeded 3 million people around the world and 11381 in Jordan. These numbers are fatal compared to other viruses, nevertheless number of its victims till this moment is much less than the victims of Spanish flu. As the pandemic continues to spread, the death toll provides a "reference point" for comparing Corona with other viruses in the past and present. Twelve years ago, there were more between 151,700 and 575,400 people who are victims of the SARS virus that was the first coronavirus to cause panic in the world, but the total number of its victims did not exceed 774 people. Corona is compared with the deadly seasonal flu, even though the such flu scarcely appears in the headlines. "Seasonal influenza kills more than half million people annually", according to the WHO [PERDUE, NGUYEN]. Death toll caused by Corona exceeds that which is caused by the Ebola hemorrhagic fever, Ebola appeared in 1976 its last outbreak between 2018 and 2020 killed around 2300 people. Over forty years, a seasonal spread of Ebola killed 15,000 people throughout Africa. AIDS that was detected at the beginning of eighties of the last century is the deadliest disease among modern epidemics, as 33 million people have died around the world as a result of the disease that strikes the immune system, but till now, no effective vaccine has been discovered to prevent it. The death toll caused by hepatitis B and C viruses is so high, exceeding one million deaths every year, the majority of which are in poor countries.

Conclusion

According to the distribution of the curve, we notice that after a steady increasing of active cases, the numbers started to decrees for the main reasons:

"Strict government procedures in dealing with the epidemic, people commit to social distance, people are becoming aware of the importance of vaccines and all of us try to adapt to this epidemic" [9].

Decision

At first, the spread pattern of Coronavirus was according a geometric series that is strictly increasing, and we noticed that the curve's distribution of the attacking virus is according log-normal distribution as implied in figures, then after a steady increasing of active cases of infected people and reaching the peak of a wave, the distribution of the curve representing active cases became exponential which means that the curve starts to decrease with a limit zero after a long time. The end of this pandemic is either socially or taking a long time and giving a huge number of vaccines.

Acknowledgement: All thanks and appreciation to the administrative and academic staffs of Applied Science Private University and everyone who contributed to the completion of this scientific study that we hope it will give a clear view about how probability and statistics beside data science help in protecting people from attacking viruses by striking lots of applications.

Declaration of Competing Interest

Authors declare that they have no

conflicts of interest to disclose.

Funding

There is no financial support.

Ethical Clearance: Taken from Applied

Science Privet University

References

[1] Taha A, Hadi A, Anomaly Detection Methods for Categorical Data, Association for Computing Machinery, May 2019; (52) 2-38. Available from https://dl.acm.org/doi/10.1145/3312739#sec-terms

[2] Leipnik B, On Lognormal Random Variables: I – The Characteristic Function, Journal of the Australian Mathematical Society, January 1991; 32 (3), 327–347.

[3] Shabrawi N, The history of influenza and how epidemics end, Tebb Alyawm, January 2021;17. Available from https://tebelyoum.net/research

[4] Lukacs E, Probability and Mathematical Statistics: An Introduction, Academic Press, May, 2014. https://www.theguardian.com/commentisfree/2020/apr/29/us-responses-1918-flu-pandemic-offer-stark-lessons-coronavirus-now

[5] Almuhur E, Al-labadi M, Shatarah A, Khan N, Bashir R, Studying structure of Coronaviridae, analyzing their geometry and focusing on the role of electronic applications in health awareness of viruses, International Journal of Human Rights in Healthcare, 2021, 10,1-10. Available from: https://www.emerald.com/insight/content/doi/10.1108/IJHRH-01-2021-0003/full/html?skipTracking=true https://www.emerald.com/insight/content/doi/10.1108/IJHRH-01-2021-0003/full/html

[6] Azzalini A, A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 1985, 12, pp. 171-178.

[7] Lahcene B, On recent modifications of extended Rayleigh distribution and its applications, JP Journal of Fundamental and Applied Statistics, 2018, 12, pp.1-13.

[8] Bartoszewicz J, Mixtures of exponential distributions and stochastic orders, Journal of Statistics Probability Letters, 2002, 57, pp. 23-31.

[9] Almuhur E, Comparison between cells and Lindelof spaces in the mathematical structure The relationship between Coronavirus Spread pattern in Jordan and the compact, International Medical Journal, 2020, 25(5), pp. 2301-2309. Available from: https://www.researchgate.net/publication/342411582_Comparison_between_cells_and_Lindelof_spaces_in_the_mathematical_structure_The_relationship_between_Coronavirus_Spread_pattern_in_Jordan_and_the_compact_topological_space

[10] Perdue M L, Nguyen T, World Health Organization Bulletin, April 2012; (90) 4, 245-320. https://www.who.int/bulletin/volumes/90/4/12-020412/ar/

[11] Groza, Shaw V, Shelley, Susanne M, Precalculus mathematics, New York: Holt, Rinehart and Winston,1972, pp. 182 Available from https://books.google.jo/books?id=yM_lSq1eJv8C&pg=PA182&redir_esc=y#v=onepage&q&f=false

[12] Bristow N, What the 1918 flu pandemic tells us about whether social distancing works, The Guardian, 2020 April, Available from https://www.theguardian.com/commentisfree/2020/apr/29/us-responses-1918-flu-pandemic-offer-stark-lessons-coronavirus-now

[13] https://corona.moh.gov.jo/en