International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 5, Issue 6, 2026
DOI: https://doi.org/10.38124/ijsrmt.v5i6.1542

Nonparametric Kernel Density Estimation of COVID-19 Incidence

Calisto Munongi[1]

[1]Department of Mathematical Sciences, Zimbabwe Open University, Harare, Zimbabwe

Publication Date: 2026/08/21

Abstract

Several studies have been done to predict the course of the COVID-19 pandemic. However, it is critical to infer from observed data, the characteristics of COVID-19 incidence to improve both predictions, and interventional strategies by policy makers. We estimate the probability density of the daily COVID-19 incidence data for the country Zimbabwe in the first 150 days since the first case was recorded. We apply nonparametric kernel density estimation to obtain a suitable smoothed distribution fit for the observed COVID-19 incidence data from 20 March to 16 August 2020. The density of COVID-19 daily incidence in Zimbabwe is characterised by a sharp peak and a positive fatter tail with several jumps (shocks). The probability mass is concentrated on the tails, and the density is greatly influenced by the few jumps. The findings suggest a distribution of the Brownian motion type. Daily incidence, among other factors, is important in understanding the dynamics of COVID-19. A kernel density estimate of COVID-19 incidence has been obtained. In the study of COVID-19 dynamics, outliers (jumps or shocks) should not be excluded in models since they help explain the density of the observed data.

Keywords: *COVID-19, Kernel Density Estimation, Probability Density Function, Kernel Smoothing Function, Incidence.*

I. INTRODUCTION

COVID-19 is an infectious disease caused by the recently discovered novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (WHO 2020). As of 01 October 2020, more than thirty-three million cases have been reported worldwide with more than a million disease-related deaths. Daily incidence (number of new cases) has been generally rising since the disease was first identified in December 2019 in Wuhan, China. For example, worldwide, there were 253,801 new cases on 30 July 2020 while there were 281,436 new cases on 27 September of the same year (WHO 2020). Since the onset of COVID-19, many research papers modelling the disease have been published.

To the best of our knowledge, most published papers are predictive models while some deal with the optimal control of the disease, age, and population structure (Colombo et al 2020, Fanelli 2020, Fosu et al 2020, Sengar 2020,). However, it is important to understand from the observed data, the pattern of occurrence of the cases so that predictive and optimal control models are equipped with more accurate assumptions. The aim of this descriptive study is to estimate the probability density of the observed COVID-19 incidence data.

II. METHODS AND MATERIAL

There are several nonparametric methods used to smooth data and these include: kernel smoothers, splines, the histogram, and locally weighted scatter plot smoothing (Bowman 1997, Rosenblatt 1956, Sheather 2004). A kernel distribution is a nonparametric presentation of a probability density function (pdf) of a random variable. Like a histogram, the kernel distribution creates a function to represent the pdf of the observed data. However, A kernel distribution sums the component smoothing functions for each data value to produce a smooth, continuous probability curve as opposed to a histogram which places the values into discrete bins.

Kernel density estimation is used to estimate an unknown probability density function of some given observed data. A suitable kernel function is generally expected to be non-negative, symmetric, and decreasing. We use a kernel distribution when we want to avoid making presumptions about the distribution of the data, or when a parametric distribution cannot properly describe the data (Scott 2001, Silverman 1986). The kernel density estimator is the estimated pdf of a random variable. Given any "$x \in \mathbb{R}$", the kernel density estimator is given by:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\frac{x - x_i}{h} \qquad (1)$$

Where $x_1, \ldots, x_n$ are random samples from an unknown distribution, $K(.)$ is the kernel smoothing function, and h is the bandwidth. $K(.)$ determines the shape of the curve used to generate the pdf while h controls the smoothness of the resulting density curve. Examples of kernel smoothing functions include normal, box, triangle, and Epanechnikov.

We compare density plots of the observed data using four different smoothing functions to understand the effect of different kernel smoothing functions. We also compare the density estimates produced by different bandwidths when using a particular chosen kernel function. As a result, an optimal bandwidth is chosen. Finally, we study the resulting density estimates to explain the potential features of the observed data. For the computer coding, we adapted the live scripts on kernel distribution available in the Statistics and Machine Learning Toolbox of MATLAB 2019a.

Incidence is taken to mean the number of new cases. Daily COVID-19 incidence data was accessed from the website of the Ministry of Health and Child Care for the country Zimbabwe. The observed data is shown in Table 1.

Table 1 Daily Incidence of COVID-19 in Zimbabwe from 20 March to 16 August 2020

1	**1**	**0**	**0**	**1**	**0**	**2**	**2**	**0**	**0**
1	0	0	1	0	0	0	1	1	0
0	2	1	0	3	1	5	1	0	1
0	0	3	0	0	0	3	0	1	0
2	0	0	0	0	0	0	0	0	1
1	0	1	0	0	1	2	4	2	0
0	2	3	5	12	0	64	0	0	17
11	14	4	26	3	16	15	28	14	3
5	27	6	12	11	13	27	4	4	10
62	16	7	3	23	13	5	21	10	6
0	7	17	14	12	8	73	18	18	53
98	41	16	40	3	49	30	273	58	58
133	102	107	214	90	172	138	78	192	113
62	213	77	490	262	154	146	118	56	56
124	74	99	70	75	97	82	104	85	

III. RESULTS AND DISCUSSION

A frequency histogram of the observed data is shown in Figure 1. Our intention is to create a smooth, and continuous probability density for the observed data that explains the data.

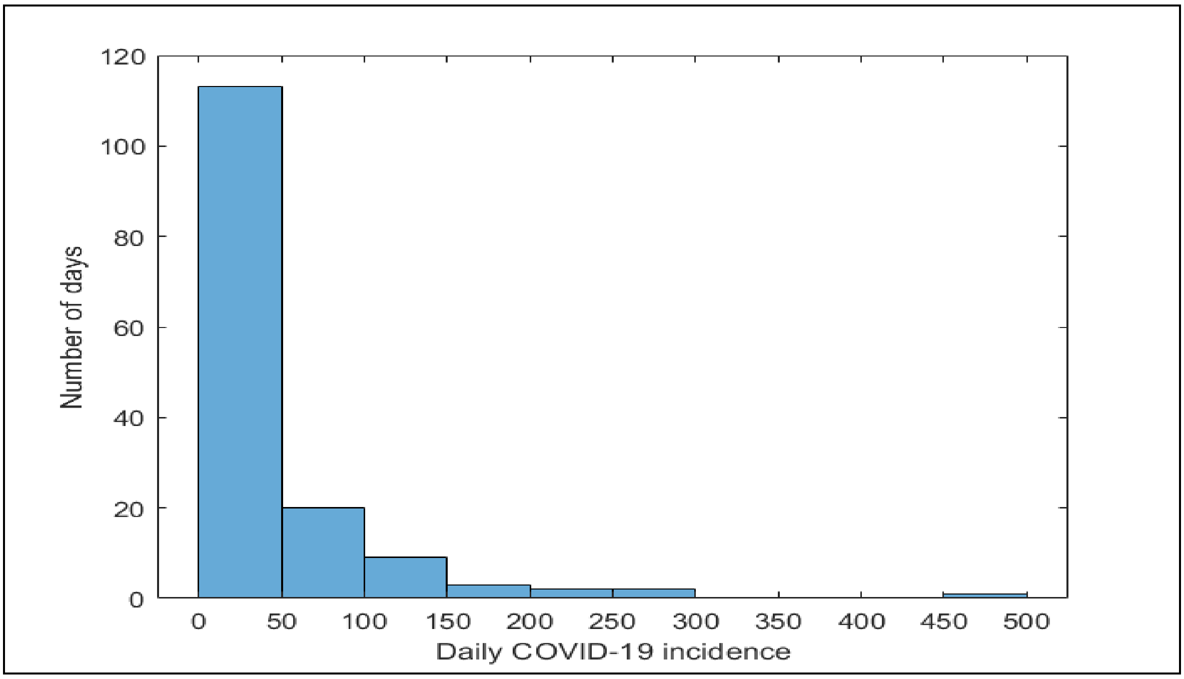

Fig 1 Histogram of the Daily COVID-19 Incidence

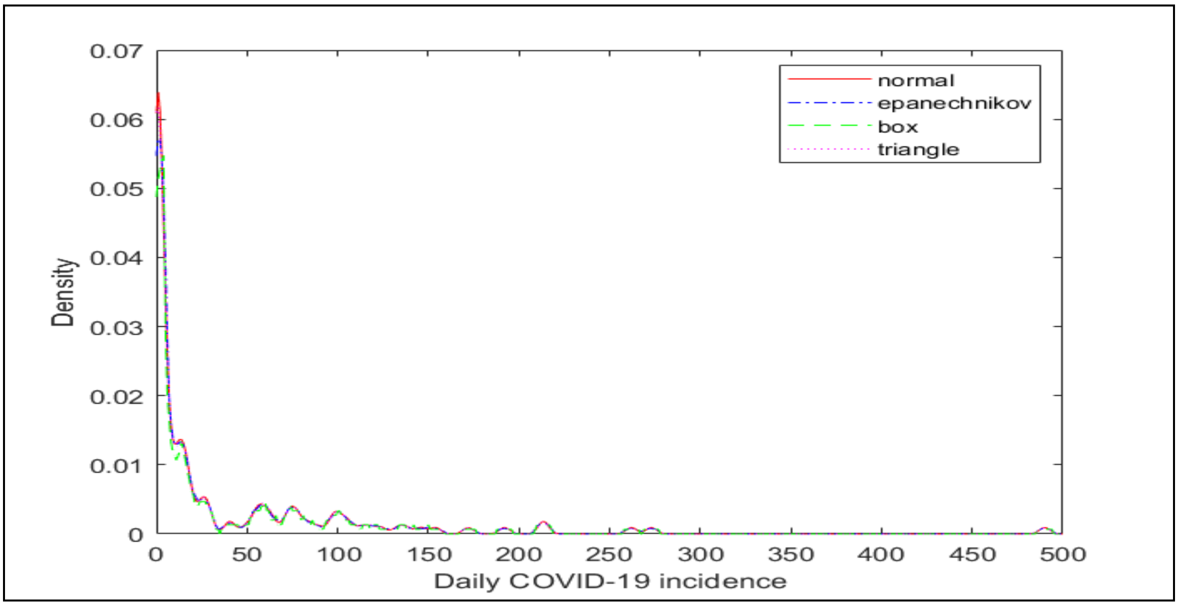

Fig 2 Density Estimates of the Observed Daily COVID-19 Incidence Using Different Kernel Smoothing Functions

In Figure 1, each density curve uses the same input data but applies a different kernel smoothing function to generate the pdf. The density estimates for the four different kernel functions are roughly comparable, but there is a slight variation in the shapes of the curves. In particular, the box kernel has a density curve with the least smoothness while the normal kernel has the sharpest peak, followed by the triangle kernel.

Density estimates obtained using the standard normal kernel function for different bandwidths are shown in Figure 3. The standard normal kernel function is given by:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, -\infty < x < \infty \tag{2}$$

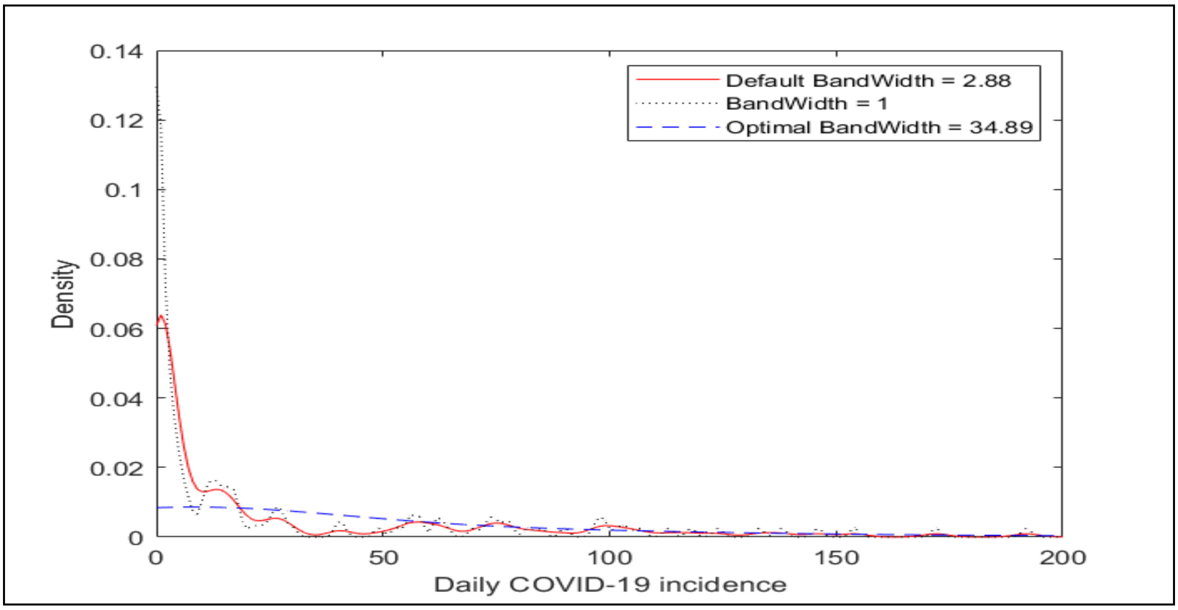

Fig 3 Density Estimate of the Observed Daily COVID-19 Incidence Using the Normal Kernel Function for Different Bandwidths

The smoothness of the probability density curves is a consequence of the bandwidth choices. The optimal bandwidth (34.89) is obtained using Silverman's adaptive measure of spread. Figure 3 illustrates that the optimal bandwidth of 34.89 produces a density curve which is too smooth, with a flattened peak, and obscuring potential characteristics of the data. A smaller bandwidth (e.g., bandwidth=1) produces a very rough curve which reveals the sharp peak and the several jumps in the data. The default bandwidth for the standard normal kernel of 2.88 produces a curve less rough than that of a smaller bandwidth of 1 and the curve reveals several moderate jumps in the data.

The resulting pdf of the estimated kernel density to the daily COVID-19 incidence is shown in Figure 4. The distribution has a sharp peak, jumps, and a positive fatter tail.

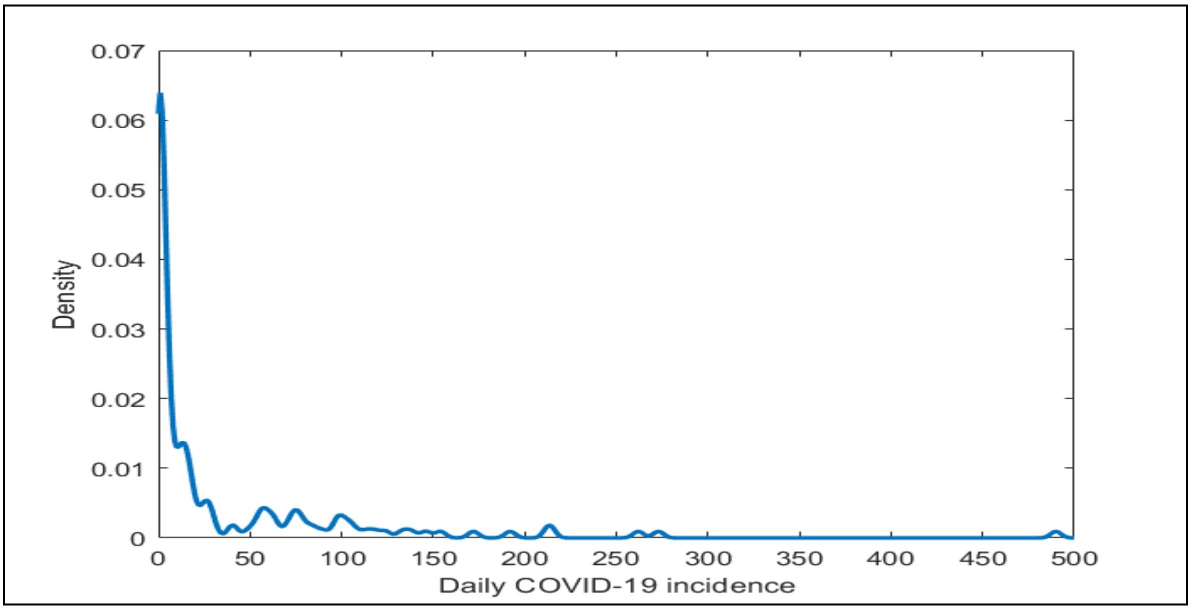

Fig 4 PDF of the Estimated Kernel Distribution to the Daily COVID-19 Incidence

IV. CONCLUSION

Our findings suggest that the daily incidence of COVID-19 has a probability density that is characterised by a sharp peak and a fatter tail with several jumps. Moreover, the distribution has a positive fatter tail. The density of COVID-19 incidence is greatly influenced by the few jumps (shocks). The only assumption made was that the data has a pdf. Further statistical modelling of the observed data is possible. However, the density estimate gives important conclusions to be obtained from the data. Our results, to a larger extent, depend on the quality of the observed data. More data and studies are required to better understand the distribution of COVID-19 incidence. We understand that, apart from incidence there are other factors affecting the dynamics of COVID-19 pandemic. It is beyond this study to obtain the actual processes or events causing the jump property. Considering these properties in models of COVID-19 may help further understanding of the pandemic.

- *Conflict of Interest*

The author declares that there is not any conflict of interests.

REFERENCES

[1]. Bowman, A. W., and Azzalini, A., Applied Smoothing Techniques for Data Analysis. New York: Oxford University Press Inc., 1997.

[2]. Colombo, R.M., Garavello, M., Marcellini, F., Rossi, E., An age and space structured SIR model describing the Covid-19 pandemic. *Journal of Mathematics in Industry*, 10:22, 2020.

[3]. Fanelli, D., and Piazza, F., Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons and Fractals*, 134: 109761, 2020.

[4]. Gabriel Obed Fosu1, G.O., Akweittey, E., and Adu-Sackey, A., Next-generation matrices and basic reproductive numbers for all phases of the Coronavirus disease. *Open Journal of Mathematical Sciences* Vol 4, 261-272

[5]. Rosenblatt, M., Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27 (3): 832-837, 1956. doi: 10.1214/aoms/1177728190.

[6]. Scott, D. W., Probability density estimation, 18: 12095-12099, Elsevier Ltd, 2001.

[7]. Sengar, D., Digital Technology for predicting, preventing and controlling COVID-19. *International Journal of Scientific Research in Science, Engineering and Technology*, 7 (3): 07-13.

[8]. Sheather, S. J., Density estimation. Statistical *Science*, 19(4):588-597, 2004.

[9]. Silverman, B. W., Density Estimation for Statistics and Data Analysis. Chapman and Hall, 1986.

[10]. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports Last accessed June 2026

[11]. https://www.mohcc.gov.zw/?option=com_phocadownload&view=category&id=15&Itemid=741 Last accessed February 2026

[12]. https://ww2.mathworks.cn/en/products.html?s_tid=nav_products Last accessed May 2026