Annals of Data Science (2022) 9(1):141–152
https://doi.org/10.1007/s40745-021-00338-9

# Modeling of COVID-19 Cases in Pakistan Using Lifetime Probability Distributions

**Muhammad Ahsan-ul-Haq**[1] · **Mukhtar Ahmed**[2] · **Javeria Zafar**[1] · **Pedro Luiz Ramos**[3]

Received: 28 November 2020 / Revised: 5 March 2021 / Accepted: 17 April 2021 /
Published online: 4 May 2021


## Abstract
The Coronavirus Disease (COVID-19) is a respiratory disease that caused a large number of deaths all over the world since its outbreak. The World Health Organization (WHO) has declared the outbreak a global pandemic. The understanding of the random process related to the behavior infection of COVID-19 is an important health and economic problem. In the proposed study, we analyze the frequency of daily confirmed cases of COVID-19 using different two-parameter lifetime probability distributions. We consider the data from the period of March 11, 2020, to July 25, 2020, of Pakistan. We consider nine lifetime probability distributions for the analysis purpose and the selection of best fit was carried out using log-likelihood, AIC, BIC, RMSE, and $R^2$ goodness-of-fit measures. Results indicate that Weibull distribution provides generally the best-fit probability distribution.

**Keywords** Coronavirus · Daily confirmed cases · Data analysis · Lifetime distributions · Goodness-of-fit

## Abbreviations
Covid-19 Coronavirus disease

✉ Muhammad Ahsan-ul-Haq
ahsanshani36@gmail.com

Mukhtar Ahmed
mukhtarah7@gmail.com

Javeria Zafar
javeriazafar38@gmail.com

Pedro Luiz Ramos
pedrolramos@usp.br

1 College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

2 School of Statistics, Minhaj University Lahore, Lahore, Pakistan

3 Institute of Mathematical Science and Computing, University of São Paulo, São Carlos, SP, Brazil

| | |
|---|---|
| WHO | World Health Organization |
| SARS-COV-2 | Severe acute respiratory syndrome |
| $R^2$ | Coefficient of determination |
| RMSE | Root mean square error |
| WD | Weibull distribution |
| PFD | Power function distribution |
| LLD | Log-logistics distribution |
| LND | Log-normal distribution |
| IWD | Inverse Weibull distribution |
| GuD | Gumbel distribution |
| BIIID | Burr III distribution |
| BXIID | Burr XII distribution |
| BSD | Birnbaum Saunders distribution |
| NIH | National Institute of Health |
| KPK | Khyber Pakhtunkhwa |
| AIC | Akaike information criterion |
| BIC | Bayesian information criterion |
| MLE | Maximum likelihood estimation |

## 1 Introduction

A viral infectious disease named coronavirus 2019 (COVID-19) was initially reported in the mid of December in Wuhan City of China [1]. COVID-19 spread worldwide and it affected more than 213 countries including Pakistan [2]. It is an infectious disease caused by Severe Acute Respiratory Syndrome (SARS-COV-2). The COVID-19 infection leads to respiratory illness and has the most common symptoms like fever, dry cough, tiredness, other symptoms are also widely reported such as sore throat, diarrhea, and loss of taste or smell, aches, and pains [3]. It is an exceptionally infectious and spreads utilizing real contacts and a respirational globule from the tainted ones, which is presently the principal wellspring of transmission of the malady. The infection can be active as long as 12 h or even two days on a reached surface [4].

In Pakistan, the first report of COVID-19 emerged on 26th February 2020 with two positive cases, within 2 days three new cases were reported in different cities without a connection between these patients [5]. Further, reported cases increased constantly until 12th June, where 139,230, positive cases were reported, later there was a decreasing trend of total cases. The total number of confirmed cases until 25th July was 273,113. The province wise detail of COVID-19 positive cases of Punjab, Sindh, KPK, and Baluchistan was 91,901, 117,598, 33,220, and 11,578 respectively.

The COVID-19 became a worldwide pandemic and its spread could be controlled by taking preventive measures. For the patients, all symptoms above should be ceaselessly checked with essential signs and to maintain a strategic distance from additionally spread, they ought to be hatched with severe clinical measures under preventive rules. The administration needs to discover a system to fight this war in an opportune manner, for example, specialists took further proportions of shutting fringes, suspending network administrations and schools, limiting both local and universal goes until further notification [6]. The reason for these measures is to constrain the odds of physical contact among individuals with the goal of controlling the transmission of COVID-19, especially because the brooding time frame for this infection is moderately longer than different infections.

Because of the novel nature of the virus, there is more prominent vulnerability around the choice on the ideal season of the vanishing of this sickness. In this manner, transient determining is critical even in the smallest insight for anticipating the up and coming month for the better administration of the cultural, financial, social, and general medical problems [7]. Data science techniques have been used to describe the behavior of pandemies, crop harvesting, business data mining, e-commerce fraud as well as others applied problems [8–19]. In the previous, not many months' scientists have created or utilized existing scientific and measurable strategies to anticipate the quantity of COVID-19 cases and related results. The summed up strategic model shows that pestilence development was exponential in china [20]. In view of the forecast, the circumstance will be exacerbated in whole Europe and the USA will turn into the focal point of new cases during the mid of April 2020 [21]. Around 115 million individuals are already tainted worldwide by March, 05, 2021 with more than 2,570,000 deaths. Expectations/gauges help to reinforce the procedures to keep the pandemic from compounding. Soltani-Kermanshahi et al. [22] worked on the statistical distribution of novel coronavirus in Iran. The study compared three types of parametric distributions known as normal, log-normal, and Weibull distribution of COVID-19 cases based on daily reported data of Iran. Yousaf et al. [5] conducted statistical Analysis of forecasting COVID-19 for the upcoming month in Pakistan.

Due to a lack of epidemiological analyses, there are many uncertainties in assessing the risk of this disease in the population. In Pakistan, it will take at least a year for any future treatment or vaccination of COVID-19. In the meantime, the only way to avoid contact with this virus is through precautionary measures and Lockdowns. It causes economic problems and it is not easy to implement without economic losses. So, effective decisions by policymakers or SOPS need to be implemented. In short, the proper modeling of a pandemic can reduce the exponential spread of this infection. Researchers are needed to fully explain its pathways and mechanisms and to identify potential curative targets, which can be effective in developing common preventive and therapeutic targets. This Global Problem has attracted the interest of researchers, giving rise to several proposals to analyze and predict the evolution of pandemic. The first importance is to check the behavior of the number of cases of COVID-19. For this, we considered different parametric distributions to describe the number of daily reported COVID-19 cases in Pakistan.

This paper aimed to identify the best fit model for the analysis of daily confirmed COVID-19 cases in Pakistan, as well as province wise. It is considered the most common two-parameter lifetime model to fit the data. To the best of our knowledge, for the first time, these probability distributions are used for modeling the number of occurrence of COVID-19 cases. The daily confirmed cases are taken from four provinces of Pakistan (Punjab, Sindh, KPK, and Balochistan). The parameters are estimated using the maximum likelihood approach. The best fit model selection was carried out using AIC, BIC, Coefficient of determination ($R^2$) and root mean square error (RMSE) criteria.

The rest of the paper is as follows; Sect. 2 is based on information on Covid-19 data of selected regions. In Sect. 3 description of statistical models, Sect. 4 is presented by information about model evaluation measures. In Sect. 5, Data is analyzed by Parameter estimates and goodness of fit measures. Finally, conclusions, discussions, and future research are given in Sect. 6.

## 2 Materials and Methods

### 2.1 Lifetime Probability Distributions

Lifetimes models are mathematical functions that return the probability of observing the event of interest given a specific time. Usually referred to as probability density function (pdf), this function is used to achieve the probability that the event takes values in a given time interval. Here, the event of interest is the daily occurrence of COVID-19 in the Pakistan population.

This section presents a brief description of the two-parameter models that will be considered in this study. Exploring the literature, some common probability distributions are used as lifetime distributions. For instance, Weibull distribution (WD), Power function distribution (PFD), Log-Logistic distribution (LLD), Log-Normal distribution (LND), inverse Weibull distribution (IWD), Gumbel distribution (GuD), Burr III distribution (BIIID), Burr XII distribution (BXIID), and Birnbaum Saunders distribution (BSD). The probability density function and range of parameters, range of pdf are given in Table 1.

The two-parameters models considered here are standard in statistical analysis and their properties, applicability, and inferential procedures are presented in the statistical literature. Our aim here is not proposed new distributions but to verify if some of the well-established distributions can be used to describe the frequency numbers of Covid-19 cases.

### 2.2 Data Set

We collect the data for daily positive cases of COVID-19, the time period was from March 11, 2020, to July 25, 2020, which were obtained from the public reports of the National Institute of Health (NIH)—Islamabad, Pakistan. It is also considered the confirmed daily case data from four provinces, Punjab, Sindh, Khyber Pakhtunkhwa (KPK), and Balochistan. Table 2 presents an exploratory analysis related to the COVID-19.

### 2.3 Model Selection and Inference

Here, it is considered the following goodness-of-fit measures for the selection of best-fitted probability distribution. The measures are Akaike information criterion (AIC),

**Table 1** Investigated PDFs and their parameters

| Model | PDF | Range/values | Parameters |
|---|---|---|---|
| WD | $f(x) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$ | $x > 0$ and $\alpha, \beta > 0$ | α: Scale<br>β: Shape |
| PFD | $f(x) = \frac{\beta x^{\beta-1}}{\alpha^{\beta}}$ | $0 < X < \alpha$ and $\alpha, \beta > 0$ | α: Scale<br>β: Shape |
| LLD | $f(x) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^{2}}$ | $x > 0$ and $\alpha, \beta > 0$ | α: Scale<br>β: Shape |
| LND | $f(x) = \frac{1}{x\beta\sqrt{2\pi}} \exp\left\{-\frac{(\ln x-\alpha)^2}{2\beta^2}\right\}$ | $x > 0$ and $\alpha \in \mathbb{R}, \beta > 0$ | α: Scale<br>β: Shape |
| IWD | $f(x) = \alpha\beta x^{-\beta-1} \exp\left(-\alpha x^{-\beta}\right)$ | $x > 0$ and $\alpha, \beta > 0$ | α: Scale<br>β: Shape |
| GuD | $f(x) = \frac{1}{\sigma} \exp\left\{-\left(\frac{x-\mu}{\sigma}\right) + \exp\left(-\frac{x-\mu}{\sigma}\right)\right\}$ | $x > 0$ and $\sigma, \mu > 0$ | σ: Scale<br>μ: Shape |
| BIIID | $f(x) = \alpha\beta x^{-\beta-1}\left(1 + x^{-\beta}\right)^{-\alpha-1}$ | $x > 0$ and $\alpha, \beta > 0$ | α: Shape<br>β: Shape |
| BXIID | $f(x) = \alpha\beta x^{\beta-1}\left(1 + x^{\beta}\right)^{-\alpha-1}$ | $x > 0$ and $\alpha, \beta > 0$ | α: Shape<br>β: Shape |
| BSD | $f(x) = \frac{1}{2\alpha\beta\sqrt{2\pi}}\left[\left(\frac{\beta}{x}\right)^{0.5} + \left(\frac{\beta}{x}\right)^{1.5}\right] e^{-\left\{\frac{1}{2\alpha^2}\left(\frac{x}{\beta}+\frac{\beta}{x}-2\right)\right\}}$ | $x > 0$ and $\alpha, \beta > 0$ | α: Scale<br>β: Shape |

**Table 2** Descriptive statistics of Covid-19 Daily cases data

| Regions | Mean | Var | Skewness | Kurtosis | n |
|---|---|---|---|---|---|
| Pakistan | 1979 | 3,183,864 | 0.85320 | 2.78342 | 138 |
| Punjab | 695.9 | 447,935.2 | 1.19635 | 3.60396 | 132 |
| Sindh | 1250.6 | 462,046.3 | 0.59938 | 2.51550 | 90 |
| KPK | 253.6 | 42,014.58 | 0.87237 | 3.45107 | 131 |
| Baluchistan | 87.05 | 8934.262 | 1.89294 | 7.16524 | 133 |

Bayesian information criterion (BIC), Root mean square error (RMSE), and Coefficient of determination ($R^2$). The test statistics are;

$$AIC = 2k - 2\ln L(\boldsymbol{\theta}),$$

$$BIC = k\ln n - 2\ln L(\boldsymbol{\theta}),$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{v}_i - v_i\right)}{n}},$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - x_i\right)^2}{\sum_{i=1}^{n} \left(y_i - z_i\right)^2},$$

where $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f\left(x_i;\boldsymbol{\theta}\right)$ is log-likelihood function evaluated at the MLEs and k refers to the number of parameters in the model. For each parameter $\theta_i$, MLE involves maximizing the likelihood function by solving the following:

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} = 0, i = 1, 2.$$

We apply such approach to obtain the likelihood functions for the parameters of the selected models, in this case, numerical techniques were used to obtain such parameter estimates. Interested readers can use statistical softwares such as R with packages that contains some of the cited models implemented, see for instance, Delignette-Muller and Dutang [23]. The codes and routines to obtain the parametes estimates can be obtained upon request.

## 3 Results

The parameters of the probability distributions are estimated using the maximum likelihood estimation method. Table 3 presents the estimates for the parameters of all probability models. Table 4 provides the results related to the goodness of fit measures. For Pakistan COVID-19 daily cases, W, Gu, PF, and LL distributions seem to have maximum $R^2$ and minimum AIC, BIC, and RMSE. Hence, among the selected distributions, we conclude that these four distributions can be utilized for describe the distributions of the diary number cases. For Punjab, we observed that W, LL, LN, and Gu distributions returned better fit than the other distributions with smaller RMSE, AIC, and BIC and higher $R^2$ values. Similar conclusions with the Weibull, LL, LN, and Gu distributions are observed for Sind, KPK, and Balochistan provinces.

Overall, it is evident from Table 4 that the best suitable model to describe the data of the different provinces of Pakistan is Weibull distribution. Figures 1, 2 presents a box-plot of $R^2$, RMSE, AIC, and BIC with the results obtained from the different models. As can be seen in the figures, we can easily identify the Weibull distribution performed better than the other models.

Figure 3 provides the adjusted Weibull distribution with the empirical distributions for Pakistan, and Punjab, Sindh, KPK, and Balochistan provinces. It can be seen the figures that Weibull distribution has a good fit for all the considered datasets, which confirms the goodness of fit tests. Hence, the findings indicate that using Weibull distribution for analysis of COVID-19 daily cases returns more accurate probabilities than using the competitor distributions.

From the adjusted results we can compute the expected number of cases assuming different levels of probability. The values can be computed from

**Table 3** The parameter estimates of all fitted probability distributions

| Model | Parameters | Pakistan | Punjab | Sindh | KPK | Baluchistan |
|---|---|---|---|---|---|---|
| WD | $\hat{\alpha}$ | 0.8837 | 0.9771 | 1.9304 | 1.0961 | 0.8656 |
| | $\hat{\beta}$ | 1880.7 | 702.95 | 1388.8 | 260.36 | 81.066 |
| PFD | $\hat{\alpha}$ | 6825.0 | 2705.0 | 3038.0 | 1035.0 | 501.00 |
| | $\hat{\beta}$ | 0.4967 | 0.5065 | 0.9474 | 0.5146 | 0.3997 |
| LLD | $\hat{\alpha}$ | 1168.8 | 414.92 | 925.00 | 176.57 | 47.856 |
| | $\hat{\beta}$ | 1.1194 | 1.3511 | 2.7450 | 1.4216 | 1.1856 |
| LND | $\hat{\mu}$ | 6.8150 | 5.9284 | 6.9634 | 4.9988 | 3.7146 |
| | $\hat{\sigma}$ | 1.7532 | 1.3389 | 0.6219 | 1.3109 | 1.4871 |
| IWD | $\hat{\alpha}$ | 10.806 | 19.718 | 346.64 | 13.803 | 5.9025 |
| | $\hat{\beta}$ | 0.4094 | 0.5731 | 0.8829 | 0.6141 | 0.6077 |
| GuD | $\hat{\mu}$ | 1167.9 | 407.59 | 930.80 | 160.75 | 49.097 |
| | $\hat{\sigma}$ | 1301.6 | 443.99 | 536.88 | 154.15 | 57.524 |
| BIIID | $\hat{\alpha}$ | 14.485 | 25.116 | 443.65 | 16.869 | 7.8600 |
| | $\hat{\beta}$ | 0.4471 | 0.6113 | 0.9185 | 0.6471 | 0.6673 |
| BXIID | $\hat{\alpha}$ | 0.0241 | 0.0269 | 0.0248 | 0.0269 | 0.0237 |
| | $\hat{\beta}$ | 6.0873 | 6.2693 | 5.8018 | 7.4420 | 11.359 |
| BSD | $\hat{\alpha}$ | 3.8585 | 2.0913 | 0.7238 | 1.7599 | 1.9144 |
| | $\hat{\beta}$ | 156.47 | 180.59 | 771.95 | 86.398 | 27.486 |

$$x_p = \lambda(-\log p)^{\frac{1}{k}},$$

where $\lambda$ and $k$ are the MLEs available in Table 3, $x$ is the integer part of x and p is the probability level. As an example, assuming a probability level ofof 0.5 and using the estimates from Pakistan, we have that $x_{0.5} = 1241$.

It is important to point out that computing estimates in real-time play a key role as a tool for decision making during pandemic periods.In this way, we have provided the necessary codes in R (available in Supplemental Material) to update the estimates and compute the expected values according to different levels.

## 4 Discussion

The current study is conducted to analyze COVID-19 daily case data of the Pakistan region, as well as also analyze province wise. Our focus was also to identify the appropriate two-parametric models that can be used to describe the distribution of the daily number of positive COVID-19 cases. It is concluded that the Weibull distribution returned better results when compared with other well-known distributions with two parameters. This conclusion is based on widely used metrics to discriminate models such as R2, AIC, BIC, and RMSE. Visual confirmation was also observed comparing the empirical distributions with the adjusted by the Weibull

**Table 4** The goodness-of-fit measures of fitted distributions

| Model | −2L | AIC | BIC | RMSE | $R^2$ |
|---|---|---|---|---|---|
| *Pakistan* | | | | | |
| WD | 2367.75 | 2371.75 | 2377.60 | 0.0427 | 0.9804 |
| PFD | 2350.08 | 2354.08 | 2359.94 | 0.0436 | 0.9747 |
| LLD | 2407.84 | 2411.84 | 2417.70 | 0.0554 | 0.9628 |
| LND | 2427.53 | 2431.53 | 2437.39 | 0.0794 | 0.9165 |
| IWD | 2516.58 | 2520.58 | 2526.44 | 0.1239 | 0.7161 |
| GuD | 2425.54 | 2429.54 | 2435.40 | 0.0516 | 0.9739 |
| BIIID | 2501.32 | 2505.32 | 2511.18 | 0.1164 | 0.7553 |
| BXIID | 2690.87 | 2694.87 | 2700.73 | 0.2376 | 0.2216 |
| BSD | 2485.50 | 2489.50 | 2495.36 | 0.2502 | 0.3964 |
| *Punjab* | | | | | |
| WD | 1991.84 | 1995.84 | 2001.61 | 0.0205 | 0.9953 |
| PFD | 2008.69 | 2012.69 | 2018.46 | 0.0865 | 0.8852 |
| LLD | 2012.27 | 2016.27 | 2022.04 | 0.0320 | 0.9881 |
| LND | 2016.73 | 2020.73 | 2026.50 | 0.0442 | 0.9762 |
| IWD | 2085.88 | 2089.88 | 2095.64 | 0.0897 | 0.8580 |
| GuD | 2044.69 | 2048.69 | 2054.46 | 0.0619 | 0.9617 |
| BIIID | 2075.22 | 2079.22 | 2084.98 | 0.0836 | 0.8813 |
| BXIID | 2300.36 | 2304.36 | 2310.13 | 0.2408 | 0.1445 |
| BSD | 2069.77 | 2073.77 | 2079.54 | 0.1628 | 0.6675 |
| *Sindh* | | | | | |
| WD | 1415.28 | 1419.28 | 1424.28 | 0.0271 | 0.9917 |
| PFD | 1443.14 | 1447.14 | 1452.14 | 0.1108 | 0.7982 |
| LLD | 1429.62 | 1433.62 | 1438.62 | 0.0877 | 0.9166 |
| LND | 1423.32 | 1427.32 | 1432.32 | 0.0299 | 0.9894 |
| IWD | 1489.02 | 1493.02 | 1498.02 | 0.1221 | 0.6703 |
| GuD | 1417.87 | 1421.87 | 1426.87 | 0.0328 | 0.9885 |
| BIIID | 1484.58 | 1488.58 | 1493.58 | 0.1168 | 0.7023 |
| BXIID | 1782.73 | 1786.73 | 1791.73 | 0.2852 | 0.0142 |
| BSD | 1443.33 | 1447.33 | 1452.33 | 0.1558 | 0.7356 |
| *Khyber Pakhtunkhwa* | | | | | |
| WD | 1710.76 | 1714.76 | 1720.51 | 0.0399 | 0.9827 |
| PFD | 1745.76 | 1749.76 | 1755.51 | 0.1005 | 0.8491 |
| LLD | 1744.55 | 1748.55 | 1754.30 | 0.0538 | 0.9655 |
| LND | 1752.35 | 1756.35 | 1762.10 | 0.0750 | 0.9309 |
| IWD | 1816.05 | 1820.05 | 1825.80 | 0.1080 | 0.8134 |
| GuD | 1740.05 | 1744.05 | 1749.80 | 0.0414 | 0.9825 |
| BIIID | 1808.43 | 1812.43 | 1818.18 | 0.1035 | 0.8304 |
| BXIID | 1993.31 | 1997.28 | 2003.03 | 0.2295 | 0.2153 |
| BSD | 1772.83 | 1776.80 | 1782.58 | 0.1580 | 0.7199 |
| *Balochistan* | | | | | |
| WD | 1449.43 | 1453.43 | 1459.21 | 0.0311 | 0.9891 |
| PFD | 1498.03 | 1502.03 | 1507.81 | 0.1157 | 0.7659 |

**Table 4** (continued)

| Model | − 2L | AIC | BIC | RMSE | $R^2$ |
|---|---|---|---|---|---|
| LLD | 1473.02 | 1477.02 | 1482.80 | 0.0503 | 0.9710 |
| LND | 1471.06 | 1475.06 | 1480.84 | 0.0651 | 0.9518 |
| IWD | 1514.77 | 1518.77 | 1524.55 | 0.0910 | 0.8883 |
| GuD | 1519.66 | 1523.66 | 1529.44 | 0.0563 | 0.9666 |
| BIIID | 1506.41 | 1510.41 | 1516.19 | 0.0853 | 0.9044 |
| BXIID | 1610.24 | 1614.24 | 1620.02 | 0.1805 | 0.5161 |
| BSD | 1471.07 | 1475.07 | 1480.85 | 0.1135 | 0.8552 |

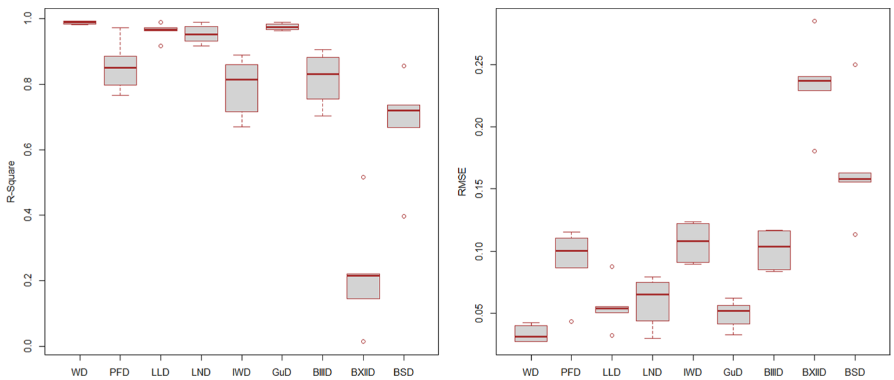


**Fig. 1** Combine box plots of $R^2$ and RMSE for all fitted probability distributions

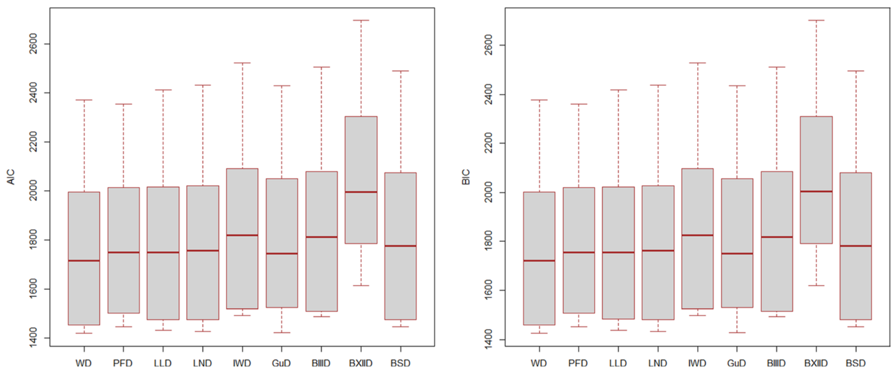


**Fig. 2** Combine box plots of AIC and BIC for all fitted probability distributions

distribution with different parameters. An interesting aspect of our findings is that while most of the analysis conducted with COVID-19 are aimed to flat the curve of the distributions due to the temporal observations ( the number of infected does not pass a threshold that could collapse the health system) here, we aim to obtain graphs with an exponential decay without a very long-tail, this would imply that there are

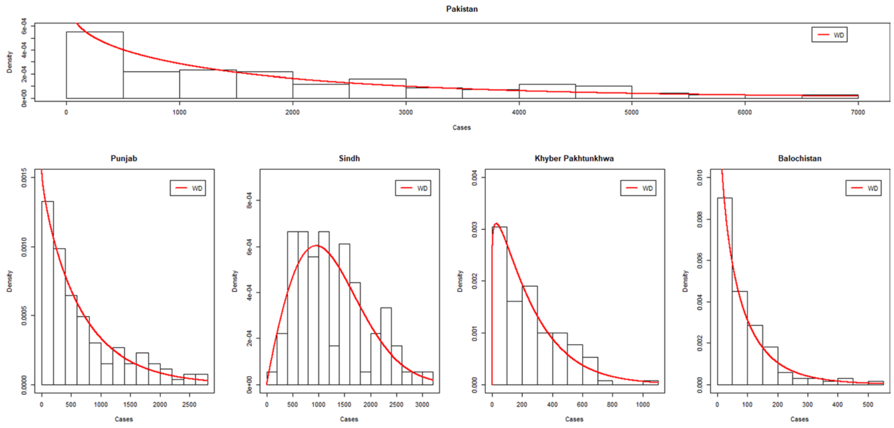


**Fig. 3** COVI-19 daily cases and fitted Weibull distribution

many days where the number of positives cases are decreasing with few positive cases. Additionally, with the adjusted parameters of the Weibull distribution, we can use the complementary of the cumulative distribution to estimate the probability that a number of cases could be greater or equal to a determinate number of positive cases of COVID-19 in Pakistan or its provinces. To the best of our knowledge, no comparison have been considered using the proposed lifetime models. To the best of our knowledge, no comparison has been considered using the proposed lifetime models. These results are of main interest during resource allocation planning or social isolation policies.

## Appendix

```
# R code for estimating the parameters of Weibull distribution.

x<-c() ##Data to be included

require(MASS)

fit<-fitdistr(x,"weibull")

AIC(fit)
```

**Acknowledgements** The authors are thankful Journal editor and reviewers for the improvent of this paper. Pedro L. Ramos acknowledges the support of the São Paulo State Research Foundation (FAPESP Proc. 2017/25971-0).

**Funding** No funding received for this paper.

**Data Availability** Data sets are available on https://covid.gov.pk/.

**Declaration**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Code availability** Application code is given in appendix.

**Author contributions** All the authors equally contribute in this project.

## References

1. Paules CI, Marston HD, Fauci AS (2020) Coronavirus infections—more than just the common cold. JAMA 323(8):707–708
2. Noreen N, Dil S, Niazi S, Naveed I, Khan N, Khan F, Tabbasum S, Kumar D (2020) COVID 19 pandemic and Pakistan; limitations and gaps. Glob Biosecur 1(4)
3. Ceccarelli G, Scagnolari C, Pugliese F, Mastroianni CM, d'Ettorre G (2020) Probiotics and COVID-19. Lancet Gastroenterol Hepatol 5(8):721–722
4. Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E (2020) Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. arXiv preprint arXiv:2003.10776
5. Yousaf M, Zahir S, Riaz M, Hussain SM, Shah K (2020) Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. Chaos Solitons Fractals 138:109926
6. Raza S, Rasheed MA, Rashid MK (2020) Transmission potential and severity of COVID-19 in Pakistan
7. Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE 15(3):e0231236
8. Kumar S (2020) Monitoring novel corona virus (COVID-19) infections in India by cluster analysis. Ann Data Sci 7:417–425
9. Khakharia A, Shah V, Jain S et al (2021) Outbreak prediction of COVID-19 for dense and populated countries using machine learning. Ann Data Sci 8:1–19
10. Li J, Guo K, Herrera Viedma E, Lee H, Liu J, Zhong Z, Gomes L, Filip FG, Fang SC, Özdemir MS, Liu XH, Lu G, Sh Y (2020) Culture vs policy: more global collaboration to effectively combat COVID-19. The Innovation. https://doi.org/10.1016/j.xinn.2020.100023
11. Liu Y, Gu Z, Xia S, Shi B, Zhou X, Shi Y, Liu J (2020) What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization EClincialMedicine 22:100354
12. Ianishi P, Junior OA, Henriques MJ, do Nascimento DC, Mattar GK, Ramos PL, Ara A, Louzada F (2020) Probability on graphical structure: a knowledge-based agricultural case. Ann Data Sci 1–19
13. Olson DL, Shi Y (2007) Introduction to business data mining. McGraw-Hill/Irwin, New York
14. Shi Y, Tian YJ, Kou G, Peng Y, Li JP (2011) Optimization based data mining: theory and applications. Springer, Berlin
15. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. Ann Data Sci 4(2):149–217
16. Nascimento DC, Barbosa B, Perez AM, Caires DO, Hirama E, Ramos PL, Louzada F (2019) Risk management in e-commerce: a fraud study case using acoustic analysis through its complexity. Entropy 21(11):1087
17. Ramos PL, Nascimento DC, Ferreira PH, Weber KT, Santos TE, Louzada F (2019) Modeling traumatic brain injury lifetime data: improved estimators for the generalized gamma distribution under small samples. PLoS ONE 14(8):e0221332
18. Cao X (2020) COVID-19: immunopathology and its implications for therapy. Nat Rev Immunol 20(5):269–270
19. Jorge, P. R., & Nuno, T. (2020). Predicting the evolution and control of COVID-19 pandemic in Portugal. https://www.medrxiv.org/content/medrxiv/early/2020/03/31/2020.03.28.20046250.full.pdf
20. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G (2020) Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model 5:256–263
21. Kumar P, Kalita H, Patairiya S, Sharma YD, Nanda C, Rani M, Rahmani J, Bhagavathula AS (2020) Forecasting the dynamics of COVID-19 pandemic in top 15 countries in April 2020: ARIMA model with machine learning approach. MedRxiv

22. Soltani-Kermanshahi M, Gholami E, Mansori K (2020) Statistical distribution of novel coronavirus in Iran
23. Delignette-Muller ML, Dutang C (2015) fitdistrplus: an R package for fitting distributions. J Stat Softw 64(4):1–34

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.