Weibull-H-NHE distribution with properties and applications on COVID-19 dataset

Lal Babu Sah Telee
Department of Management Science, Nepal Commerce Campus, Tribhuvan University, Nepal
Email: lalbabu3131@gmail.com

Cite this paper

Sah Telle, L.B. (2020). Weibull-H-NHE distribution with properties and applications on COVID-19 dataset. *NCC Journal, 5*(1), 63-73.

https://doi.org/10.3126/nccj.v5i1.56952

Abstract

A new three parameters univariate continuous distribution called Weibull-H-NHE distribution on COVID-19 data with Properties and Applications on COVID-19 data using Weibull-Hclass of distribution is introduced here. Some major properties of model with application on covid-19 data are presented in study. Different curves like probability density, hazard rate, and empirical distribution etc. are included in the study. Parameters are estimated by using method of least square. Model validation is tested by calculating different information criteria values as well as using Q-Q and P-P plots. The applicability of the model is tested on three sets of real data sets. Model application is checked by considering their real data set. All the calculations are done using the R language programming.

Keywords: Weibull–H class. Probability density function, COVID-19, Methods of least square, R language programming

Introduction

Data analysis is one of the important aspects of the research. There are different methods and tools available for analyzing the data. Probability model is one of the tools for analyzing and interpreting the data. There are many probability distributions are available in literature that are very useful in analyzing all types of the data. There are some types of recent real data on classical models do not give desired results. To get the more precise result, many new probability models are introduced in literature during last decades. Different techniques of introducing new models like modifying the existing models, adding some extra parameters, and use of family of probability distributions are available in theory. A new modified distribution (Laie, et al., 2003), A new modified Weibull distribution (Almalki &Yuan, 2013) and the beta modified distribution (Silva et al., 2010) etc are modified distribution of existing distributions. The exponentiated generalized class of distributions (Cordeiro and Ortega, 2013), Exponentiated distributions (Al-Hussani & Ahsanullah, 2015) and the exponentiated Weibull distribution (Nadarajah, et al, 2013) are some exponentiated models. In this study, Weibull-H class (Cordeiro,et al., 2017) and the exponential model's extension (Nadarajah & Haghighi, 2011) which is also named as NHE distribution (Chaudhary & Kumar, 2020)is used for formulation of the Weibull-H-NHE (WHNHE) distribution. This class is found to be better for fitting the covid-19 data. Let x is continuous random variable then CDF and PDF of the Weibull – H class is given by

Different section of the study is divided as follows. Introduction section of the article contains literature and some related probability models. Model formulation section contains the Cumulative distribution and probability density function with graphs. Statistical properties section contains some properties like quantile function and skewness and kurtosis etc. In Applications to real data sets section, model is applied on three real data sets. In Conclusion section, summary of the study is mentioned and last section contains the reference of the study.

$$F(x:\quad ,^{-}) = 1-\exp\left[-\quad\left(\frac{H(x;¶)}{\overline{H}(x;¶)}\right)^{-}\right] \quad (1.1)$$

$$f(x;\quad ,^{-}) = \quad ^{-}h(x:¶)\left(\frac{H(x;¶)^{-\,-1}}{\overline{H}(x,¶)^{-\,+1}}\right)\exp\left[-\quad\left(\frac{H(x;¶)}{\overline{H}(x,¶)}\right)^{-}\right] \quad (1.2)$$

Where, $H(x;¶)$, $\overline{H}(x,¶)^{-\,+1}$ and $h(x:¶)$ are the CDF, reliability function and PDF of the base line distribution respectively. NHE distribution is extension of the exponential distribution which is very useful in real data analysis.

Model formulation

The cumulative distribution function (cdf) and probability density function (pdf) of the NHE distribution is given as,

$$H(x;\sim,ª) = 1-e^{1-(1+ªx)^{\sim}}\;;\quad x \ge 0,(\sim,ª)>0, \quad (2.1)$$

$$h(x:\sim,ª) = \sim ª\,(1+ªx)^{\sim -1}\,e^{1-(1+ªx)^{\sim}} \quad x \ge 0,(\sim,ª)>0 \quad (2.2)$$

And, $\overline{H}(x;\sim,ª) = e^{1-(1+ªx)^{\sim}}\;;\quad x \ge 0, \sim > 0, ª > 0$ (2.3)

Substituting equations (2.1), (2.2), and (2.3) and the cdf & pdf of the proposed model WHNHE can be obtained as,

$$F(x;\sim,\quad ,ª,^{-}) = 1-\exp\left[-\quad\left\{e^{(1+ªx)^{\sim}-1}-1\right\}^{-}\right]; x \ge 0,(\sim,\quad ,ª,ª)>0 \quad (2.4)$$

$$f(x;\sim,\quad ,ª,^{-}) = \sim\quad ª^{-}(1+ªx)^{\sim -1}\left\{e^{(1+ªx)^{\sim}-1}-1\right\}^{-\,-1} e^{\left\{(1+ªx)^{\sim}-1\right\}}$$

$$\exp\left[-\quad\left\{e^{(1+ªx)^{\sim}-1}-1\right\}^{-}\right]; x \ge 0,(\sim,\quad ,ª,ª)>0 \quad (2.5)$$

Reliability function

The reliability function R of the survival of any component beyond the specified time x and is given as, $R(x,\sim,\quad ,ª,^{-}) = 1-F(x,\sim,\quad ,ª,^{-}) = \exp\left[-\quad\left\{e^{(1+ªx)^{\sim}-1}-1\right\}^{-}\right]$ (2.6)

Hazard rate function

The hazard rate function (hrf) of the WHNHE is given as,

$$h(x) = \frac{f(x)}{R(x)} = \sim \; a^{-} (1 + a\,x)^{\sim -1} e^{(1+a\,x)^{\sim} -1} \left\{ e^{(1+a\,x)^{\sim} -1} - 1 \right\}^{- -1} , x > 0 \quad (2.7)$$

Reversed hazard rate function

$$h_{rf} = \frac{f(x)}{R(x)} = \sim \; a^{-} (1 + a\,x)^{\sim -1} e^{(1+a\,x)^{\sim} -1} \left\{ e^{(1+a\,x)^{\sim} -1} - 1 \right\}^{- -1}$$

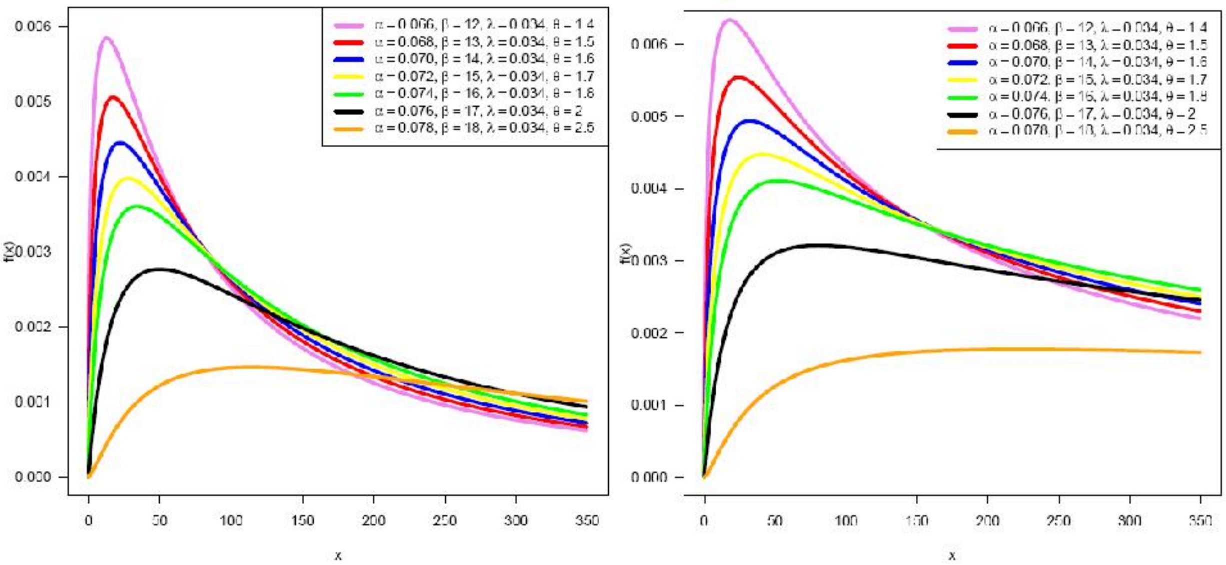
$$\exp\left[- \left\{ e^{(1+a\,x)^{\sim} -1} - 1 \right\}^{-} \right] . \left[1 - \exp\left[- \left\{ e^{(1+a\,x)^{\sim} -1} - 1 \right\}^{-} \right] \right]^{-1} ; \; x > 0 \quad (2.8)$$

Cumulative hazard rate function

$$H(x) = -\log\, H(x) = -\log(1 - F(x)) = \left\{ e^{(1+a\,x)^{\sim} -1} - 1 \right\}^{-} \quad (2.9)$$

Figure 1 displays the pdf and hazard rate function of WHNHE.

Figure 1
Probability density function (left panel) and hazard rate function (right panel) of WHNHE



Statistical properties

Here, some important characteristics of the model is discussed.

Quantile function

The Quantile function is $Q(u) = F^{-1}(u)$, u being uniform variable. This function can be alternative to CDF & PDFand helps to find summary of the data. Let Q is quantile function of the WHNHE,

$$Q(u) = \frac{1}{a}\left[1 + \ln\left\{ 1 + \left\{ \left(\frac{-1}{\;} \right) \ln(1-u) \right\}^{1/-} \right\} \right]^{1/\sim} \; ; 0 < u < 1 \quad (3.1)$$

Median of WHNHE

Median can be calculated taking u = ½ (3.1), That is,

$$\text{Median} = \frac{1}{a}\left[1+\ln\left\{1+\left\{\left(\frac{-1}{\ }\right)\ln\left(\frac{1}{2}\right)\right\}^{1/\bar{\ }}\right\}\right]^{1/\sim} \tag{3.2}$$

Random deviate generation

The random deviate generation of the model is

$$x = \frac{1}{a}\left[1+\ln\left\{1+\left\{\left(\frac{-1}{\ }\right)\ln\left(1-u\right)\right\}^{1/\bar{\ }}\right\}\right]^{1/\sim} \quad ;0<u<1 \tag{3.3}$$

The random deviate generation helps to generate random numbers (observations).

Asymptotic properties

Asymptotic behavior of the WHNHE can be determined by checking $\lim_{x\to 0} f(x) = \lim_{x\to\infty} f(x)$.

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} \sim \; a^{-}\left(1+ax\right)^{\sim -1}\left\{e^{(1+ax)^{\sim}-1}-1\right\}^{- -1} e^{\left\{(1+ax)^{\sim}-1\right\}}$$

$$\exp\left[-\;\left\{e^{(1+ax)^{\sim}-1}-1\right\}^{-}\right]=0$$

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \sim \; a^{-}\left(1+ax\right)^{\sim -1}\left\{e^{(1+ax)^{\sim}-1}-1\right\}^{- -1} e^{\left\{(1+ax)^{\sim}-1\right\}}$$

$$\exp\left[-\;\left\{e^{(1+ax)^{\sim}-1}-1\right\}^{-}\right]=0$$

Since $\lim_{x\to 0} f(x) = \lim_{x\to\infty} f(x)$, so proposed distribution satisfies the asymptotic properties showing that the distribution has modal value.

Skewness and kurtosis

Skewness describes about the consistency of the data. Here we have used Bowley's coefficient of skewness (Al-saiary et al., 2019) based on quantiles as,

$$\text{SK}(\text{B}) = \frac{\text{Q}(0.75)+\text{Q}(0.25)-2*\text{Q}(0.50)}{\text{Q}(0.75)-\text{Q}(0.25)}$$

Coefficient of Octiles Kurtosis by (Moors, 1998) and (Al-saiary et al., 2019) can be calculated using relation,

$$\text{K}_\text{u} = \frac{\text{Q}(0.875)-\text{Q}(0.625)+\text{Q}(0.375)-\text{Q}(0.125)}{\text{Q}(0.75)-\text{Q}(0.25)}$$

Application to real data sets

In this section, application of the proposed model WHNHE taking three set of Cpovid-19 data sets. The first two stet of data are the cases of COVID -19 dataset from January 21 to march 27, 2020 reported WHO (2020) and Worldometer (2020). The third set of the data is the total daily infected cases in China from January 21 to march 27, 2020 as reported by the Worldometer.

Data set 1

Daily COVID-19 cases all over the world from January 21 to March 27 as reported by WHO (2020).

60, 32, 265, 472, 698, 785, 1781, 1477, 1755, 2010, 2127, 2603, 2838, 3239, 3915, 3721, 3173, 3437, 2676, 3001, 2546, 2035, 14153, 5151, 2662, 2097, 2132, 2003, 1852, 516, 977, 996, 978, 554, 882, 741, 992, 1292, 1503, 1989, 1981, 1858, 2573, 2298, 3111, 3625, 4049, 3892, 4390, 4567, 7266, 8295, 10907, 11059, 13042, 12897, 15745, 20585, 26158, 30648, 29429, 32480, 41371, 43744, 48461, 60830, 64501

Exploratory analysis of data

Exploratory analyses of data explain the structure the data. The *boxplot* and the *Total Time Test* (TTT) plot are displayed in figure 1.. The TTT checks that considered data set can be applied for a specific probability model. Following expression is Empirical version of TTT plot

$$T\left(\frac{r}{n}\right)=\sum_{i=1}^{n} y_{(i:n)}+(n-r)y_{i:n}\left(\sum_{i=1}^{n} y_{(i:n)}\right)^{-1}$$

Where,

r = 1, 2... n and $y_{(i:n)}\,(i=1,2,\ldots,r)$ be sample order statistics .

Convexity of TTT plot shows that hazard rate shape is increasing.

Figure 2
Box plot and TTT plot of data1

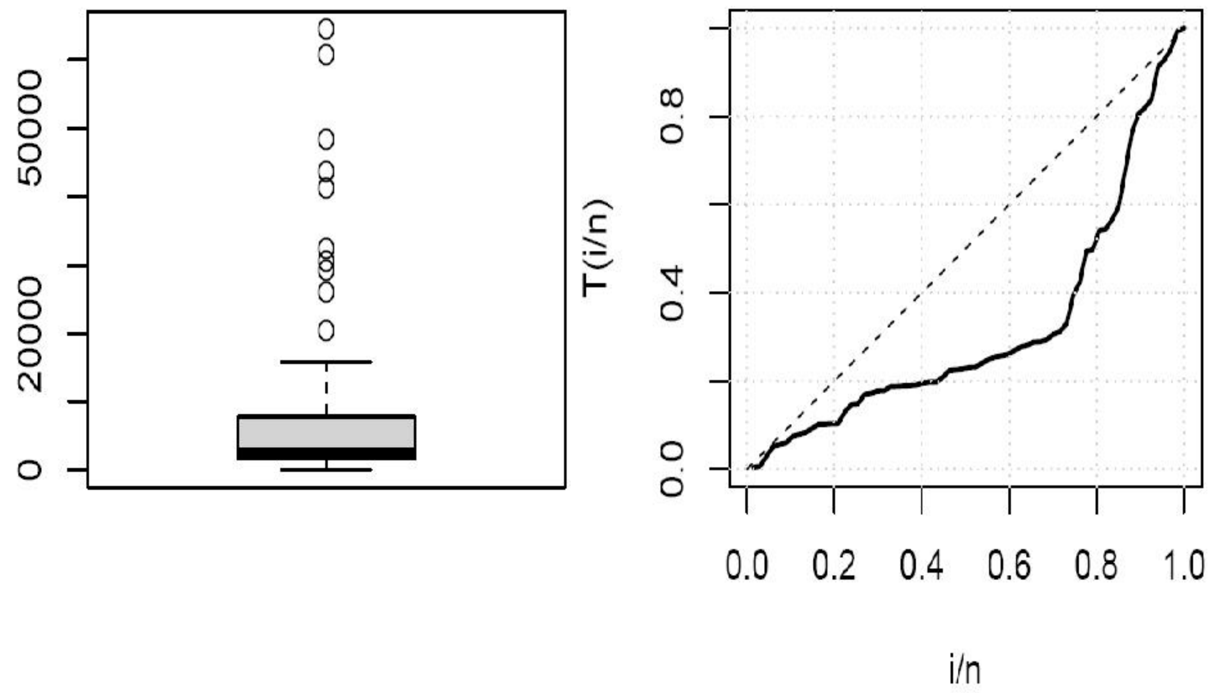


Table 1
Statistical summaries

Least	Q_1	Md.	Mean	Q_3	SD	Sk.	K	Max.
32	1629	2662	8894	7780	14538.08	2.351476	7.857866	64501

Shape of the curve of data is non normal wit right skewed.

Estimation of Parameter and validity testing

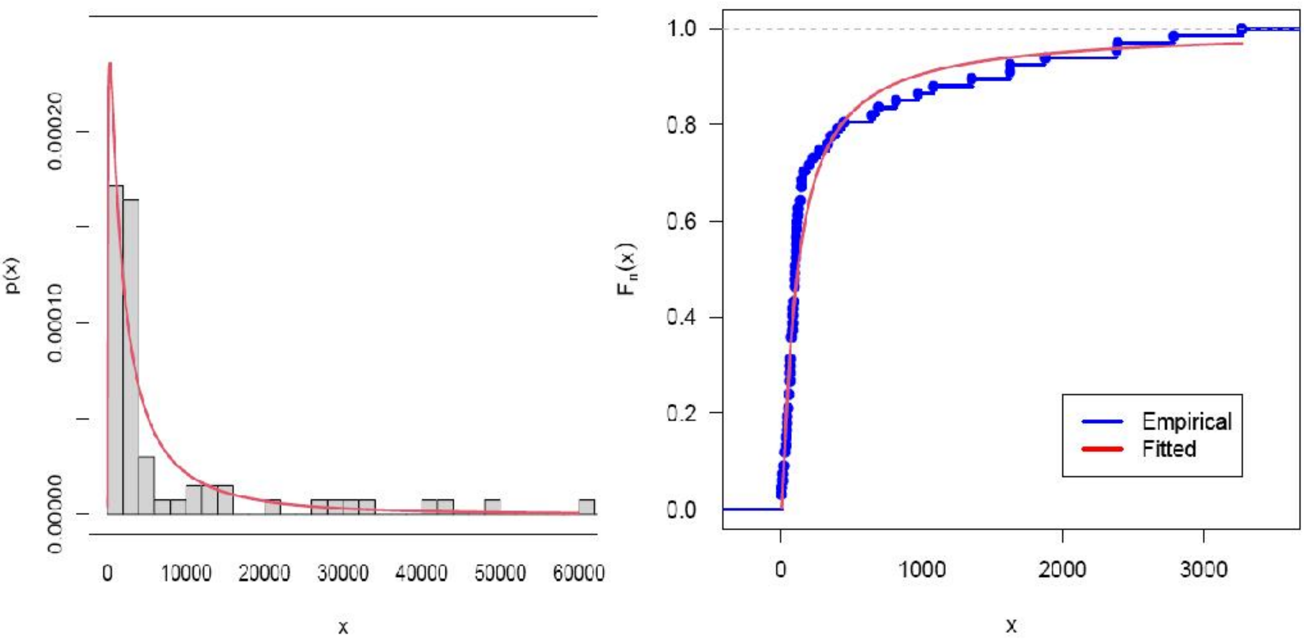
Estimation of parameters using analytical method is impossible due to the non linearity of the partial derivatives. For the estimation, optim () function of R software is used (R Core Team, 2020). MLE and the corresponding standard error of estimate SE of the parameters are displayed in table 2.

Table 2
MLE of parameters and standard errors

Parameters	MLE	Standard Error
Alpha	0.0922	0.0387
Beta	5.5071	6.1018
Lambda	0.0027	0.0005
Theta	1.5982	0.2495

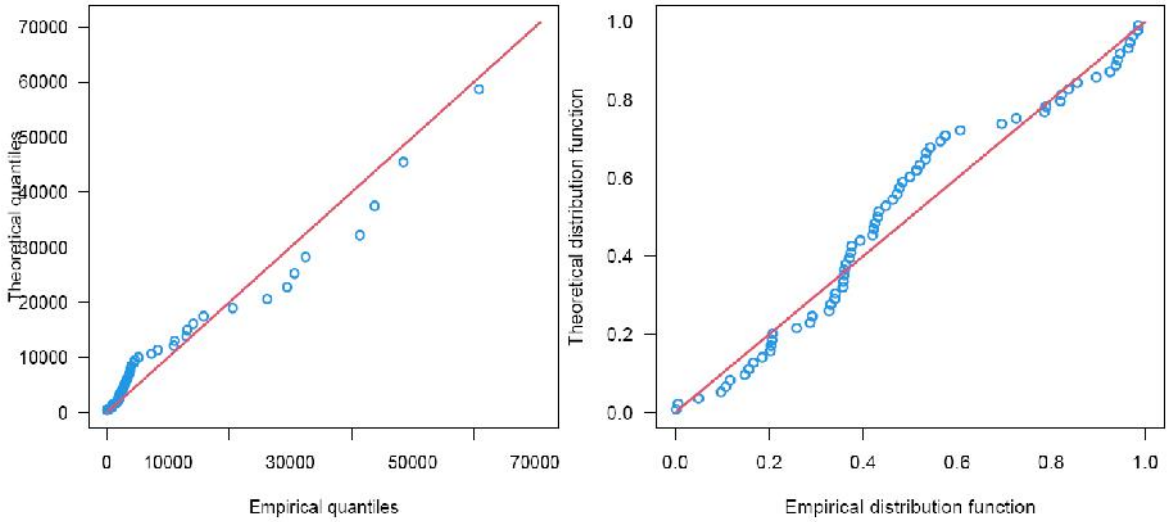
Figure 2 displays the histogram and density fitted model. It also displays the ecdfto theoretical distribution function.

Figure 3
Fitted density and histogram (Left panel) and ecdf & theoretical distribution function (Right panel)



The figure displays the Q-Q plot and the P-P plot of the WHNHE

Figure 4
The Q-Q (Left) and P-P (Right) of WHNHE



Negative log - likelihood value, Akaike (AIC), Bayesian (BIC), Corrected Akaike (CAIC), and Hannan-Quinn (HQIC) informationcriteria of WHNHE are tabulated in table3.Table also contains Kolmogrove-Smirnov (KS), Anderson-Darling (A^2), along with Cramer's-Von Mises (W^2) statistics It also contains respective p-values the proposed model.

Table 3
Negative Log-likelihood, AIC, BIC, CAIC, HQIC, KS, A^2 and W^2 and p-values

-LL	AIC	BIC	CAIC	HQIC	KS(p-values)	A^2(p-values)	W^2(p-values)
662.108	1332.216	1341.035	1332.862	1335.706	0.1439(0.1130)	1.3017(0.2317)	0.2528 (0.1848)

Different statistics and corresponding p- values shows that model fits the data set 1 well

Data set 2

Daily COVID-19 cases all over the world from January 21 to March 27 as reported by Worldometer (2020)

3, 3, 8, 16, 15, 24, 26, 26, 38, 43, 46, 45, 58, 64, 66, 73, 73, 86, 89, 97, 108, 97,146, 122, 143, 143, 106, 98, 136, 117, 121, 113, 100, 158, 81, 64, 37, 58, 65, 54, 73, 67, 85, 83, 102, 107, 105, 228, 198, 271, 332, 353, 447, 405, 687, 642, 817, 972, 1079, 1356, 1625, 1629, 1873, 2381, 2388, 2791, 3271

Exploratory data analysis

Figure 5 shows boxplot and the *Total Time Test* (TTT) plot. Convexity of TTT plot indicates increasing hazard rate curve.

Figure 5
Box plot (Left) and TTT plot (Right) of data 2

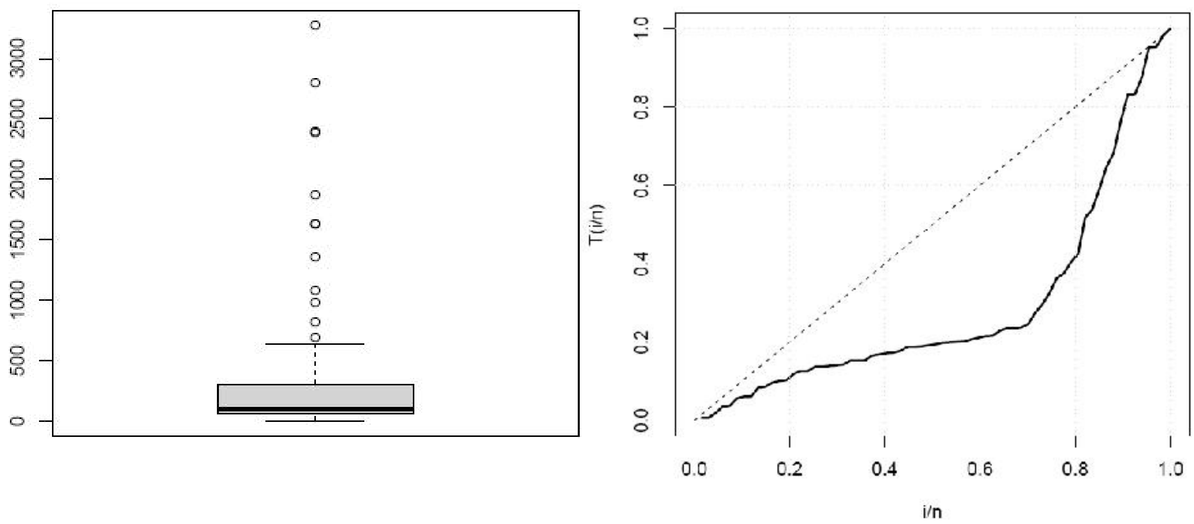


Table 4

Statistical Summaries

Least	Q_1	Md.	Mean	Q_3	SD	Sk	K	Max.
3.0	64.0	102.0	408.0	301.5	719.0619	2.416066	8.098535	3271.0

Data set 2 is non normal with longer tail in right.

Estimation of Parameter and testing of validity

MLE and the corresponding standard error of estimate SE of the parameters are displayed in table 5.

Table 5

MLE of parameters and standard errors

Parameters	MLE	Standard Error
Alpha	0.0977	0.0424
Beta	5.7133	5.0203
Lambda	0.0318	0.0325
Theta	0.0317	0.3117

Figure 6 displays the histogram and density fitted model. It also displays ecdf to the theoretical distribution function.

Figure 6

Fitted density and histogram (Left panel) and ecdf & theoretical distribution function (Right panel)

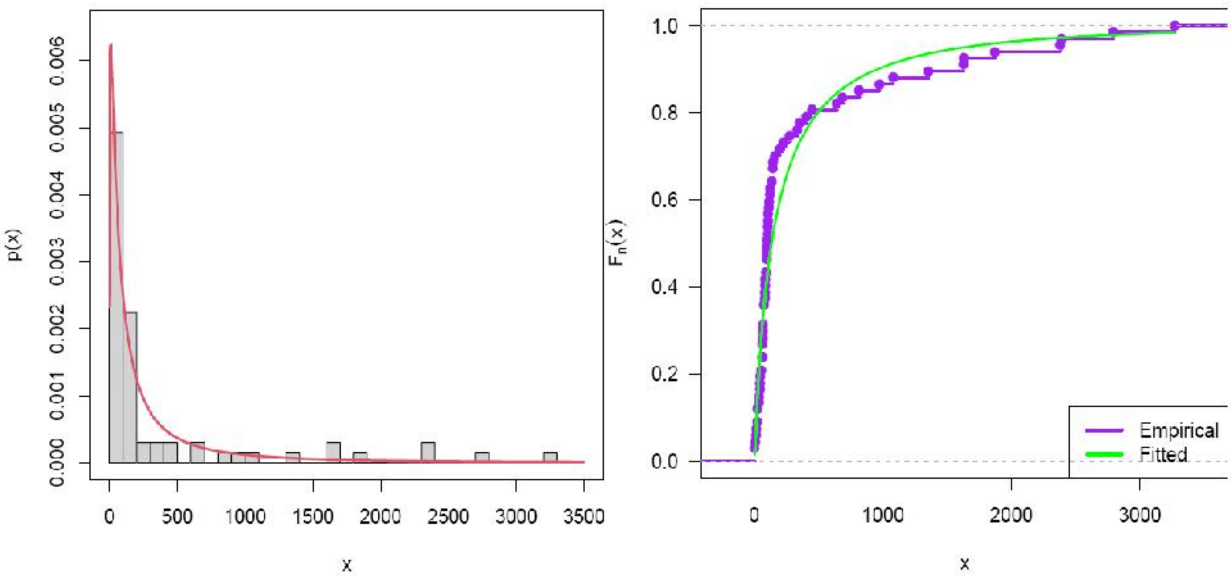
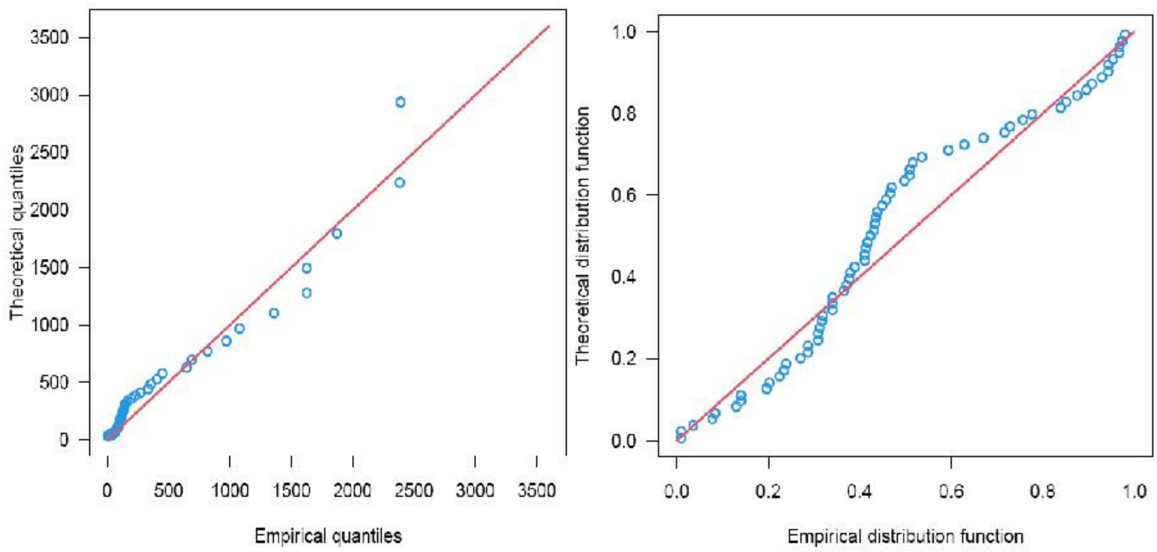


Figure 7

The Q-Q (Left) and P-P plots (Right) of WHNHE



The Q-Q and P-P plots of the WHNHE are displayed in figure 7.

Negative log - likelihood value, Akaike (AIC), Bayesian (BIC), Corrected Akaike (CAIC), and Hannan-Quinn information criteria (HQIC) of WHNHE are tabulated in table6.Table also contains Kolmogorov-Smirnov (KS), Anderson-Darling (A^2), along with Cramer's-Von Mises (W^2) statistics It also contains respective p-values the proposed model.

Table 6
NegativeLog-likelihood, AIC, BIC, CAIC, HQIC, KS, A^2 and W^2 and p-values

-LL	AIC	BIC	CAIC	HQIC	KS(p-values)	A^2(p-values)	W^2(p-values)
449.112	906.224	915.042	906.869	909.713	0.1660 (0.0510)	1.6751 (0.1398)	0.3368 (0.1072)

Different statistics and corresponding p- values shows that model also fits the data set 2 well

Data set 3

Daily COVID-19 cases in China from January 21 to March 27 as reported by Worldometer (2020)

92, 277, 483, 663, 801, 2631, 576, 2054, 1659, 2088, 4736, 3086, 3987, 3729, 3144, 3522, 2703, 3012, 2516, 2021, 372, 15133, 6460, 2055, 2099, 1918, 1775, 407, 453, 473, 1450, 16, 214, 508, 405, 433, 326, 427, 575, 200, 125, 120, 151, 151, 79, 47, 36, 22, 28, 8, 8, 26, 21, 19, 17, 22, 25, 46, 37, 80, 24, 62, 41, 77, 47

Exploratory data analysis

Figure 8 shows boxplot and *Total Time Test* (TTT) plotsfor data set 3. Convexity of TTT plot indicates that hazard rate curve is increasing - decreasing

Figure 8
Box plot (Left) and TTT plots (Right) of data3

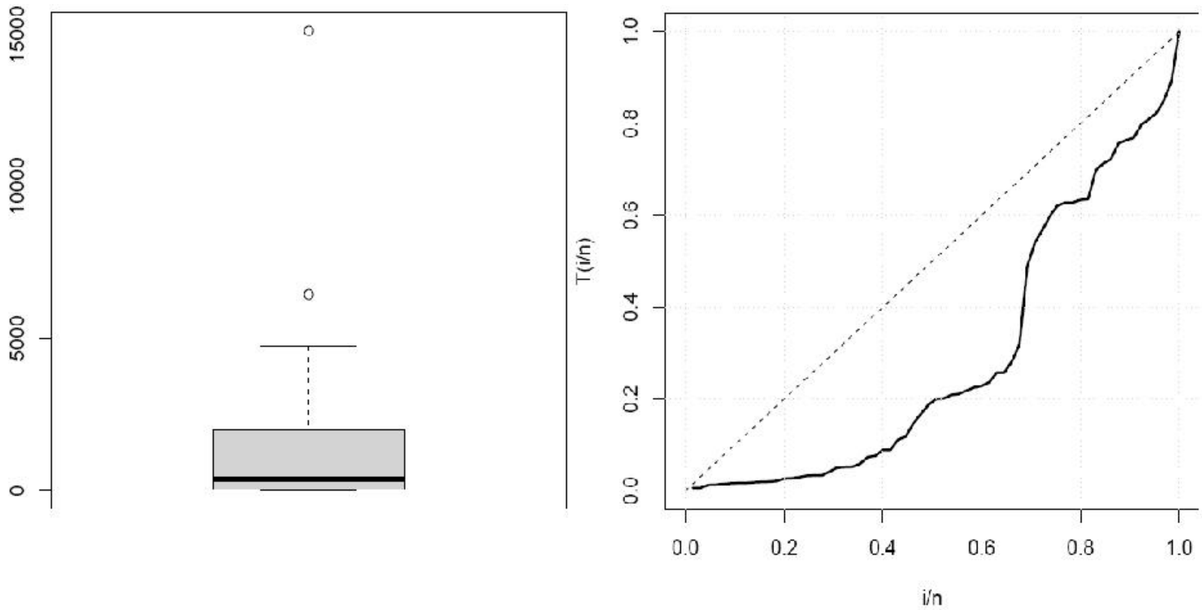


Table 7
Statistical summaries

Least	Q_1	Md	Mean	Q_3	SD	Sk	K	Max.
8	47	405	1243	2021	2239.027	4.0499	24.2113	15133

Summary shows that data is non normal with right longer tail.

Estimation of Parameter and validity testing

MLE and corresponding standard error of estimate SE of the parameters are displayed in table 8.

Table 8
MLE of parameters and standard errors

Parameters	MLE	Standard Error
Alpha	0.1185	0.0794
Beta	0.9540	1.1016
Lambda	0.1389	0.2343
Theta	1.1016	0.7115

Figure 9 displays the histogram and density fit model. It displays the ecdf to theoretical distribution function.

Figure 9
Fitted density and histogram (Left panel) and ecdf & theoretical distribution function (Right panel)

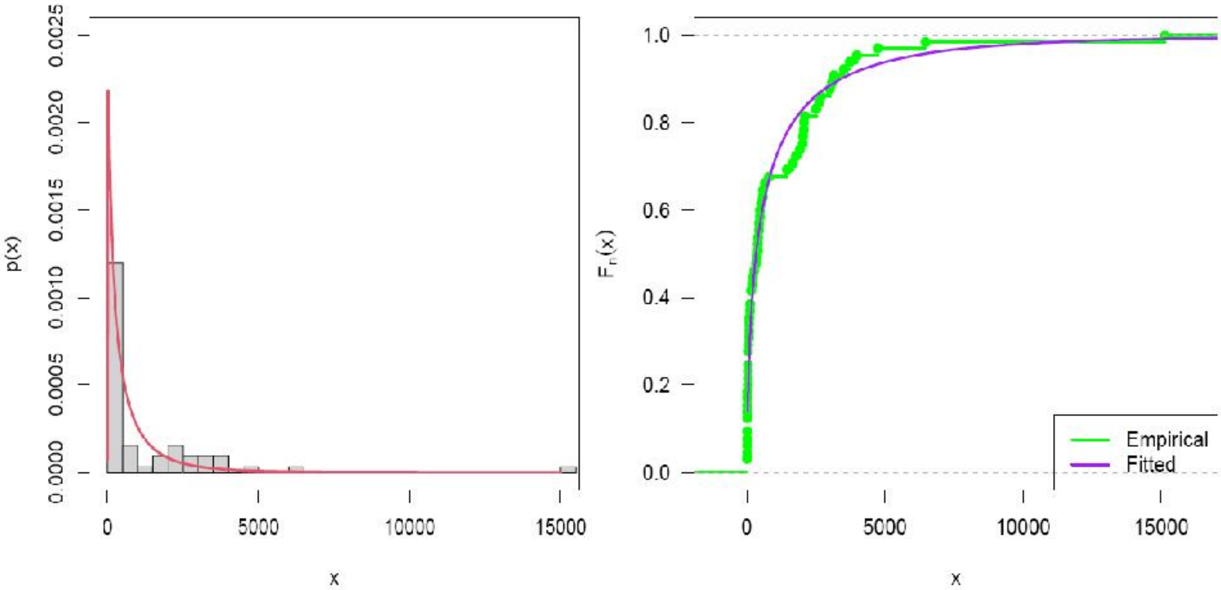
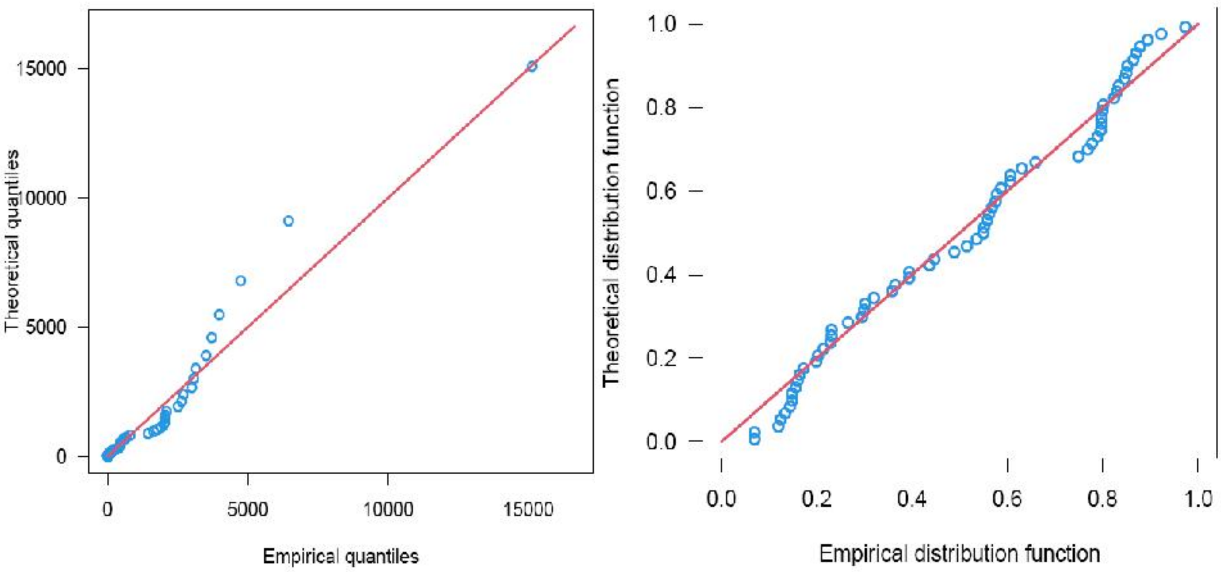


Figure 10 displays the Q-Q plot and the P-P plot of the WHNHE

Figure 10
The Q-Q plot (Left) and P-P plot (Right) of WHNHE



Negative log - likelihood value, Akaike (AIC), Bayesian (BIC), Corrected Akaike (CAIC), and Hannan-Quinn information criteria (HQIC) of WHNHE are tabulated in table 9 .Table also contains Kolmogorov-Smirnov (KS), Anderson-Darling (A^2), along with Cramer's-Von Mises (W^2) statistics It also contains respective p-values the proposed model.

Table 9
Log-likelihood (LL), AIC, BIC, CAIC, HQIC, KS, A2 and W2 and p-values

-LL	AIC	BIC	CAIC	HQIC	KS(p-values)	A^2(p-values)	W^2(p-values)
502.561	1013.121	1021.819	1013.788	1016.553	0.1072(0.4437)	0.9258(0.3981)	0.1389(0.4259)

Different statistics and corresponding p- values shows that model also fits the data set 3 well

Conclusion

This study is based on formulation of a new distribution called Weibull-H-NHE distribution. Some statistical properties such as survival, hazard rate as well as quantile functions are presented here. The density curve of the WHNHE has different shape based on different parameters values. Model application is tested using three sets of COVID-19 data. Method of maximum likelihood estimation (MLE) is applied for estimation of parameters. Validity is tested using Q-Q and P-P plot showing that WHNHE fits real data set well. Information criteria values and the validity testing show that the proposed model fit real-life data well. The graph of the hazard function isinverted bathtub and increasing –decreasing shaped. Application to real data sets shows that WHNHE is more flexible for different COVID-19 data.

References

Al-Hussaini, E. K., & Ahsanullah, M. (2015). Exponentiated distributions. *Atlantis Studies in Probability and Statistics, 21*.

Almalki, S. J., & Yuan, J. (2013). A new modified Weibull distribution. *Reliability Engineering & System Safety, 111,* 164-170.

Al-saiary, Z. A., Bakoban, R. A., & Al-zahrani, A. A. (2019). Characterizations of the Beta Kumaraswamy Exponential Distribution. *Mathematics*, *8*(1), 23.

Chaudhary, A. K., & Kumar, V. (2020). The logistic NHE distribution with properties and applications. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 8*(12), 591-603.

Cordeiro, G. M., Afify, A. Z., Yousof, H. M., Pescim, R. R., & Aryal, G. R. (2017). The exponentiated Weibull-H family of distributions: Theory and Applications. *Mediterranean Journal of Mathematics*, *14*(4), 1-22.

Cordeiro, G. M., Ortega, E. M., & da Cunha, D. C. (2013). The exponentiated generalized class of distributions. *Journal of data science*, *11*(1), 1-27.

Lai, C. D., Xie, M., & Murthy, D. N. P. (2003). A modified Weibull distribution. IEEE *Transactions on reliability, 52*(1), 33-37.

Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *37* (1), 25-32.

Nadarajah, S., & Haghighi, F. (2011). An extension of the exponential distribution. Statistics, *45*(6), 543-558.

Nadarajah, S., Cordeiro, G. M., & Ortega, E. M. (2013). The exponentiated Weibull distribution: A survey. *Statistical Papers*, *54*(3), 839-877.

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. URLhttps://www.R-project.org/.

Silva, G. O., Ortega, E. M., & Cordeiro, G. M. (2010). The beta modified Weibull distribution. *Lifetime data analysis, 16*(3), 409-430.

World Health Organization (WHO). Coronavirus disease (COVID-19) technical guidance: Surveillance and case definitions. Geneva: WHO; 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/surveillance-and-case-definitions

Worldometer (2020), Reported death cases 2020 worldwide. https://www.worldometers.info/coronavirus/about/.