Weibull-Fréchet distribution: A new lifetime distribution with application to gastric cancer data

Marko Chindranata, *Siti* Nurrohmah*, and *Ida* Fithriani

Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Abstract. Lifetime data is a type of data that consists of waiting time until an event occurs. Some of the events of lifetime data are deaths, occurrence of a disease, or failure of a machine. The distribution usually used for modeling lifetime data is the Weibull distribution. However, Weibull distribution has a limitation in its application: it can only model data with a monotonic hazard function. Therefore, a method for generalizing the Weibull distribution is needed so it can model data with a non-monotonic hazard function. One of those generalizations is the Weibull-Fréchet distribution (WFr) which was introduced by Afify in 2016. The WFr distribution has an advantage over the Weibull distribution, due to its capability in modeling data with unimodal hazard function. The method used in generating the WFr distribution is the Weibull-*G* (WG) that were introduced by Bourguignon in 2014. The WG method combines the distribution of a Weibull distribution with an arbitrary distribution with a cumulative distribution function (cdf) $G(x)$ using a function $W[G(x)]$. The characteristics of WFr distribution discussed include probability density function (pdf), cumulative distribution function, survival function, hazard function, and the moment. The hazard function of WFr can be monotonic or unimodal. The maximum likelihood estimation method is used in estimating the parameters of the distribution. Finally, lifetime data of gastric cancer patients is given for illustration purposes. The data is modeled using the WFr distribution, and both the Weibull and Fréchet distribution for comparison. The model result shows that the WFr distribution is the best distribution for modeling the lifetime data of gastric cancer patients.

Keywords. Hazard function, maximum likelihood method, unimodal, Weibull-*G*

1 Introduction

Lifetime data consists of waiting time until an event occurs. Some of the events of lifetime data are deaths, occurrence of a disease, or failure of a machine [1]. Lifetime data have

* Corresponding author: snurrohmah@sci.ui.ac.id

positive real value and a skewed distribution pattern. There are two function that are usually used in modeling lifetime data: survival function and hazard function. Hazard function usually gives more information of the data than survival function due to its variety of hazard shapes. The shape of hazard function can be constant, monotonic, bathtub, or unimodal (inverse-bathtub).

Weibull distribution is one of the most popular distribution in modeling lifetime data, due to its skewed distribution pattern and its monotonic hazard function. However, Weibull distribution is unable to model data with a non-monotonic hazard function. We consider Fréchet distribution which is also able to model lifetime data [2] but has a different hazard function shape that is unimodal. Both Weibull and Fréchet distribution are combined using a method called Weibull-*G* (WG) [3]. In generating the cumulative distribution function of a WG distribution, the cdf of Weibull distribution will be modified using $W[G(x)]$, which is a function of the cumulative distribution function with an arbitrary distribution $G(x)$. Some of other generalizations of Weibull distribution using this method are the Weibull-Uniform, Weibull-Weibull, Weibull-Burr XII, and Weibull-Normal [3].

Using the WG method, Afify [4] generate the new distribution by using the cdf of Fréchet distribution for *G(x)*, resulting in the Weibull-Fréchet (WFr) distribution. The WFr distribution is a generalization of Weibull distribution. This generalization can describe lifetime data better than the Weibull distribution and its generalizations with WG method, due to its capability in modeling data with a unimodal hazard shape.

This paper explains how to generate the WFr distribution using the WG method, along with its mathematical properties. Mathematical properties include the shapes of the density and hazard functions, the moments, and the quantile. The maximum likelihood estimation (mle) method is used in estimating the parameters of this distribution. At the end of this paper, real data modeled using this distribution is presented along with the Weibull and Fréchet distribution for comparison. For applicative purposes, we use a different data used by Afify, and use a lifetime data of gastric cancer patient instead.

2 Materials and method

2.1 Weibull-*G* method

The Weibull-*G* method introduced by Bourguignon [3] is one of the methods used to generalize The Weibull distribution. This method modifies the cumulative distribution function of Weibull Distribution with a function $W[G(x)]$. Let $G(x)$ be the cdf of arbitrary distribution, the cdf of WG is defined as follows:

$$F(x) = \int_0^{W[G(x)]} abt^{b-1}e^{-at^b}\,dt, \qquad x > 0, a > 0, b > 0. \tag{1}$$

According to Alzaatreh [5], there can be several options in choosing the $W[G(x)]$ function for Equation (1). By choosing the $W[G(x)]$ such as:

$$W[G(x)] = \frac{G(x)}{1 - G(x)}, \tag{2}$$

the cdf of WG distribution can be generated by substituting the $W[G(x)]$ in Equation (2) to Equation (1), and then by solving the integration part of Equation (1):

$$F(x) = 1 - \exp\left[-a\left(\frac{G(x)}{1-G(x)}\right)^b\right], \qquad x > 0, a > 0, b > 0. \tag{3}$$

2.2 Generating Weibull-Fréchet distribution

Generating the cdf of The Weibull-Fréchet distribution can be done by substituting *G(x)* of Equation (3) with the Fréchet distribution's cdf. The cdf of WFr is defined as follows:

$$F(x) = 1 - \exp\left[-a\left(\exp\left[\left(\tfrac{\alpha}{x}\right)^\beta - 1\right]\right)^{-b}\right].$$

The pdf of WFr can be found by finding the derivative of its cdf. Some illustrations on the pdf of WFr distribution are given on Fig. 1.

$$f(x) = ab\beta\alpha^\beta x^{-\beta-1}\exp\left[-b\left(\frac{\alpha}{x}\right)^\beta\right]\left(1-\exp\left[-\left(\frac{\alpha}{x}\right)^\beta\right]\right)^{-b-1} \exp\left[-a\left(\exp\left[\left(\tfrac{\alpha}{x}\right)^\beta - 1\right]\right)^{-b}\right] \tag{4}$$

Survival function of WFr can be formulated by finding the complement of its cdf. Some illustrations on the survival function of WFr distribution are given in Fig. 2.

$$S(x) = \exp\left[-a\left(\exp\left[\left(\tfrac{\alpha}{x}\right)^\beta - 1\right]\right)^{-b}\right].$$

The hazard function of WFr distribution can be found by finding the ratio of its pdf and survival function. Some illustrations on the survival function of WFr distribution are given on Fig. 3.

$$h(x) = ab\beta\alpha^\beta x^{-\beta-1}\exp\left[-b\left(\tfrac{\alpha}{x}\right)^\beta\right]\left(1-\exp\left[-\left(\tfrac{\alpha}{x}\right)^\beta\right]\right)^{-b-1}.$$

The pdf, cdf, survival function, and hazard function of WFr distribution given in the equations above will be valid for $x > 0$, and its parameters $a > 0, b > 0, \alpha > 0$, and $\beta > 0$.

From Fig. 1, we can see that increasing value of parameter a will result in a pdf graph with a steeper curve, while increasing value of parameter α will result in a pdf graph with a smoother curve. Altering the values of parameter b or β will result in a pdf graph with differing skewness. In Fig. 2, increasing the value of parameter a will result in a faster decrease of survival function graph, while increasing the value of parameter α will have a slower decrease. Different values of b or β will result in a survival function graph that is slowly decrease in the start, followed by a steep decrease after certain point of time. For the hazard function, altering the values of a or α only generate a monotonically increasing hazard shape, but with a different steepness. Different value of b or β will generate other shapes of hazard function such as monotonically decreasing or unimodal.

https://doi.org/10.1051/itmconf/20246101012

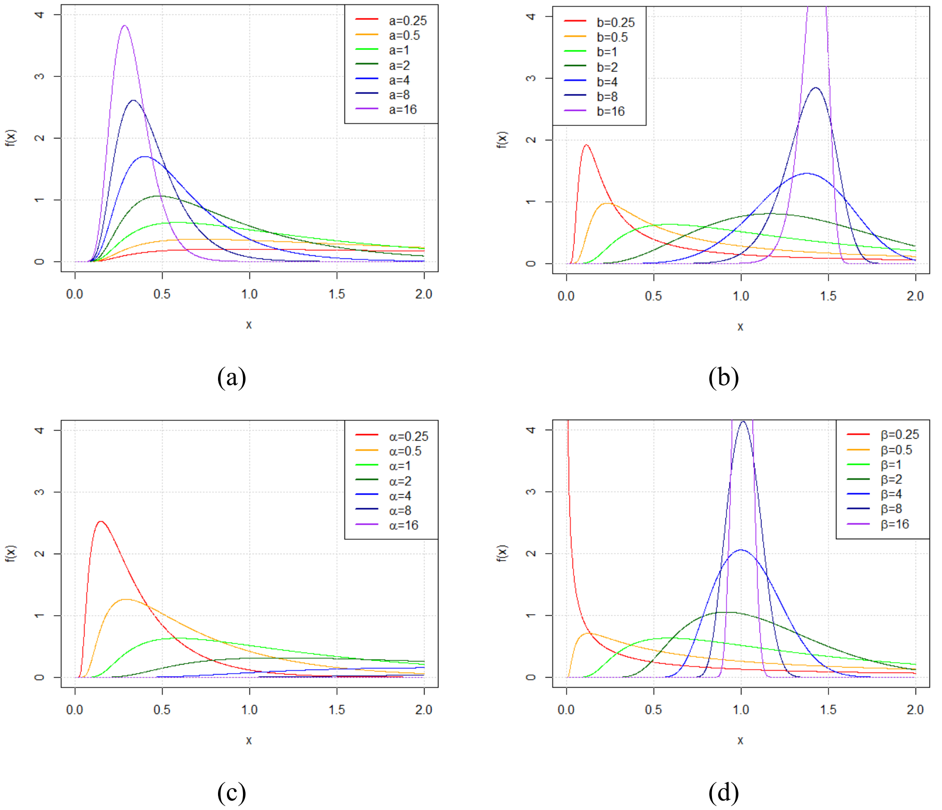


Fig. 1. Pdf graph of WFr distribution with (a) varying values of a while other parameters $b = \alpha = \beta = 1$, (b) varying values of b while other parameters a$= \alpha = \beta = 1$, (c) varying values of α while other parameters $a = b = \beta = 1$, and (d) varying values of β while other parameters $a = b = \alpha = 1$.

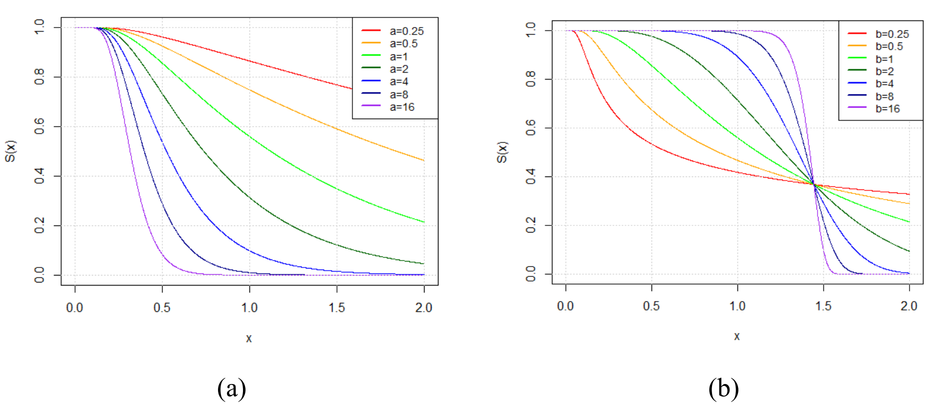
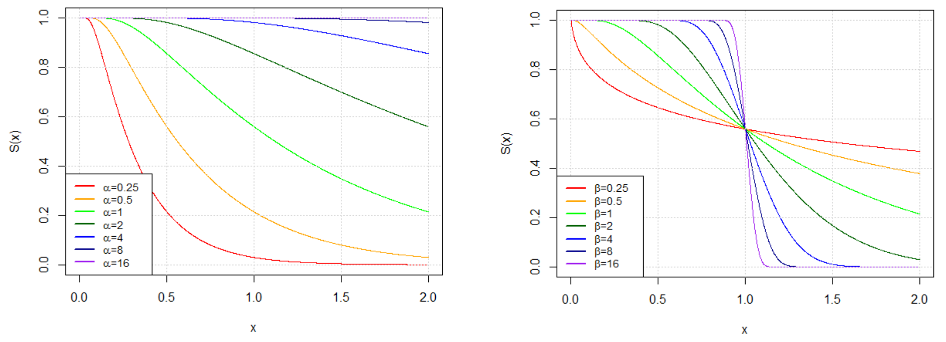
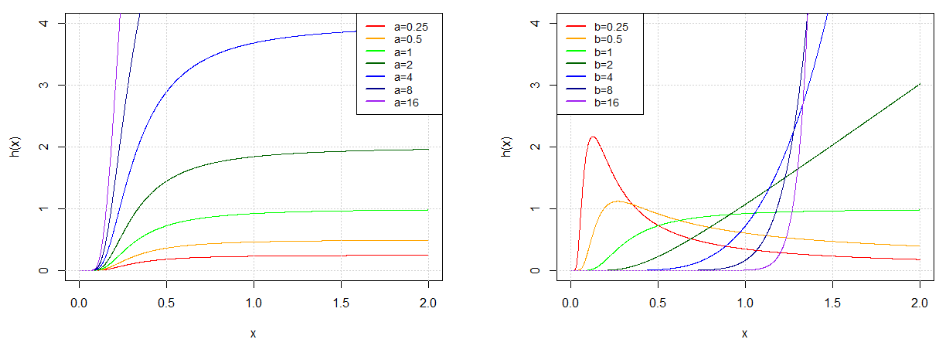


Fig. 2. Survival function graph of the WFr distribution with (a) varying values of a while other parameters $b = \alpha = \beta = 1$, (b) varying values of b while other parameters a$= \alpha = \beta = 1$, (c) varying values of α while other parameters $a = b = \beta = 1$, and (d) varying values of β while other parameters $a = b = \alpha = 1$.

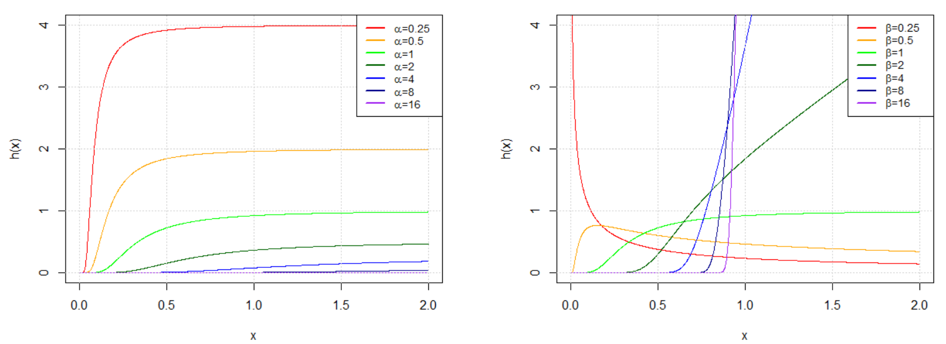


(c) (d)

Fig. 2 (continued). Survival function graph of the WFr distribution with (a) varying values of a while other parameters $b = \alpha = \beta = 1$, (b) varying values of b while other parameters a= $\alpha = \beta = 1$, (c) varying values of α while other parameters $a = b = \beta = 1$, and (d) varying values of β while other parameters $a = b = \alpha = 1$.



(a) (b)



(c) (d)

Fig. 3. Hazard function graph of the WFr distribution with (a) varying values of a while other parameters $b = \alpha = \beta = 1$, (b) varying values of b while other parameters $a = \alpha = \beta = 1$, (c) varying values of α while other parameters $a = b = \beta = 1$, and (d) varying values of β while other parameters $a = b = \alpha = 1$.

By applying some Taylor series expansion on the pdf of WFr in Equation (4), the mixture representation of the WFr pdf is written as:

$$f(x) = \sum_{j,k=0}^{\infty} v_{j,k} h_{(k+1)b+j}(x), \tag{5}$$

with $v_{j,k}$:

$$v_{j,k} = \frac{(-1)^k b a^{k+1}[(k+1)b+1]^j}{j!\,k!\,[(k+1)b+j]}, \tag{6}$$

$$[(k+1)b+1]^j = \frac{\Gamma\big((k+1)b+1+j\big)}{\Gamma\big((k+1)b+1\big)}, \tag{7}$$

and $h_{(k+1)b+j}(x)$ the pdf of Fréchet distribution with parameters $\alpha[(k+1)b+j]^{\frac{1}{\beta}}$ and β [4]. Using Equation (5), (6), and (7), we can solve the r-th moment of WFr distribution:

$$E[X^r] = \sum_{j,k=0}^{\infty} v_{j,k}\, \alpha^r [(k+1)b+j]^{\frac{r}{\beta}} \Gamma\left(1-\frac{r}{\beta}\right). \tag{8}$$

The moments (first and second) WFr distribution are derived from Equation (8) by substituting $r = 1$ and $r = 2$:

$$E[X] = \sum_{j,k=0}^{\infty} v_{j,k}\, \alpha [(k+1)b+j]^{\frac{1}{\beta}} \Gamma\left(1-\frac{1}{\beta}\right), \tag{9}$$

$$E[X^2] = \sum_{j,k=0}^{\infty} v_{j,k}\, \alpha^2 [(k+1)b+j]^{\frac{2}{\beta}} \Gamma\left(1-\frac{2}{\beta}\right). \tag{10}$$

Using Equation (9) and (10), we can get variance of WFr distribution as follows:

$$Var(X) = \sum_{j,k=0}^{\infty} v_{j,k}\, \alpha^2 [(k+1)b+j]^{\frac{2}{\beta}} \Gamma\left(1-\frac{2}{\beta}\right) - \left\{\sum_{j,k=0}^{\infty} v_{j,k}\, \alpha [(k+1)b+j]^{\frac{1}{\beta}} \Gamma\left(1-\frac{1}{\beta}\right)\right\}^2.$$

The quantile of WFr distribution is given by:

$$\pi_p = \frac{\alpha}{\left[\ln\left[1+\left[-\frac{a}{\ln[1-p]}\right]^{\frac{1}{b}}\right]\right]^{\frac{1}{\beta}}}.$$

2.3 Parameter estimation of Weibull-Fréchet distribution

The maximum likelihood estimation (mle) is used in estimating the parameters of WFr distribution. Let $X_1, X_2, \dots, X_n$ be a random sample from WFr distribution with parameters $a > 0, b > 0, \alpha > 0,$ and $\beta > 0$. The likelihood function is given by:

$$L(\boldsymbol{\theta}) = a^n b^n \beta^n \alpha^{n\beta}\left(\prod_{i=1}^{n} x_i^{-\beta-1}\right) e^{-b\sum_{i=1}^{n}\left(\frac{\alpha}{x_i}\right)^{\beta}} \prod_{i=1}^{n}\left(1 - e^{-\left(\frac{\alpha}{x_i}\right)^{\beta}}\right)^{-b-1} \times \sum_{i=1}^{n} \exp\left[-a\left(\frac{e^{-\left(\frac{\alpha}{x_i}\right)^{\beta}}}{1 - e^{-\left(\frac{\alpha}{x_i}\right)^{\beta}}}\right)^{b}\right].$$

Then the log-likelihood function of WFr distribution:

$$l(\boldsymbol{\theta}) = n\big(\ln(a) + \ln(b) + \ln(\beta) + \beta\ln(\alpha)\big) - (\beta+1)\sum_{i=1}^{n}\ln(x_i) - b\sum_{i=1}^{n}\left(\frac{\alpha}{x_i}\right)^{\beta} - (b+1)\sum_{i=1}^{n}\ln\left(1 - e^{-\left(\frac{\alpha}{x_i}\right)^{\beta}}\right) - a\sum_{i=1}^{n}\left(\frac{e^{-\left(\frac{\alpha}{x_i}\right)^{\beta}}}{1-e^{-\left(\frac{\alpha}{x_i}\right)^{\beta}}}\right)^{b}.$$

By letting $s_i = \exp\left[-\left(\frac{\alpha}{x_i}\right)^{\beta}\right]$ and $z_i = \left(\frac{\alpha}{x_i}\right)^{\beta}\ln\left[\frac{\alpha}{x_i}\right]$, the partial derivatives of the log-likelihood function with respect to the parameters are given below:

$$\frac{\partial}{\partial a} l(\boldsymbol{\theta}) = \frac{n}{a} - \sum_{i=1}^{n}\left(\frac{s_i}{1-s_i}\right)^{b},$$

$$\frac{\partial}{\partial b} l(\boldsymbol{\theta}) = \frac{n}{b} - \sum_{i=1}^{n}\ln(s_i) - \sum_{i=1}^{n}\ln(1-s_i) - a\sum_{i=1}^{n}\left(\frac{s_i}{1-s_i}\right)^{b}\ln\left(\frac{s_i}{1-s_i}\right),$$

$$\frac{\partial}{\partial \alpha} l(\boldsymbol{\theta}) = \frac{n\beta}{\alpha} - \frac{b\beta}{\alpha^{1-\beta}}\sum_{i=1}^{n} x_i^{-\beta} - \frac{(b+1)\beta}{\alpha^{1-\beta}}\sum_{i=1}^{n}\frac{s_i x_i^{-\beta}}{1-s_i} + \frac{ab\beta}{\alpha^{1-\beta}}\sum_{i=1}^{n}\frac{s_i^{-1}x_i^{-\beta}}{(s_i^{-1}-1)^{1+b}},$$

$$\frac{\partial}{\partial \beta} l(\boldsymbol{\theta}) = \frac{n}{\beta} + n\ln(\alpha) - \sum_{i=1}^{n}\ln(x_i) - b\sum_{i=1}^{n} z_i + (b+1)\sum_{i=1}^{n}\frac{s_i z_i}{1-s_i} + \sum_{i=1}^{n}\frac{ab s_i^{-1} z_i}{(s_i^{-1}-1)^{1+b}}.$$

The maximum likelihood estimates of $a, b, \alpha,$ and β are obtained by solving the nonlinear equations: $\frac{\partial}{\partial a} l(\boldsymbol{\theta}) = 0, \frac{\partial}{\partial b} l(\boldsymbol{\theta}) = 0, \frac{\partial}{\partial \alpha} l(\boldsymbol{\theta}) = 0,$ and $\frac{\partial}{\partial \beta} l(\boldsymbol{\theta}) = 0$. It's relatively hard to find the result analytically, therefore the estimates will be solved using Newton method in by RStudio version 4.2.2 with nlm() function.

3 Results and discussion

3.1 Data illustration

For applicative purposes, real lifetime data is given as an illustration for modeling with the Weibull-Fréchet distribution. The data in Table 1 is taken from Klein and Moeschberger [1], which consists of times until gastric cancer patients die (in days). Descriptive Statistics of the data can be viewed in Table 2. Histogram of the data is given in Fig. 4. It can be seen graphically that the data is right skewed. Figure 5 gives the Total Time on Test (TTT) plot of the data. The Total Time on Test is a method to identify the hazard shape of a data, and by the plot given in Fig. 5 we can see that the data have a unimodal hazard shape. Therefore, it is possible to assume that Weibull-Fréchet distribution can model gastric cancer data. For comparison purposes, we also model the gastric cancer data with both Weibull and Fréchet distribution. Parameters generated by the mle method for all the distribution is given in Table 3. In Fig. 6, some comparison between the pdf and histogram, also for cdf and survival function to its respective empirical function. Graphically, we can see that the that the distributions are relatively close with the real data with Weibull-Fréchet distribution is the closest.

Table 1. Time until gastric cancer patients die (in days).

1	63	105	129	182	216	250	262	301
301	342	354	356	358	380	383	388	394
408	460	489	499	523	524	535	562	569
675	676	748	778	786	797	955	968	1000
1245	1271	1420	1551	1694	2363			

Table 2. Descriptive statistics of the gastric cancer data.

Mean	619.6279
Median	489
Mode	301
Variance	231323.1
Skewness	1.546212
Kurtosis	2.495586

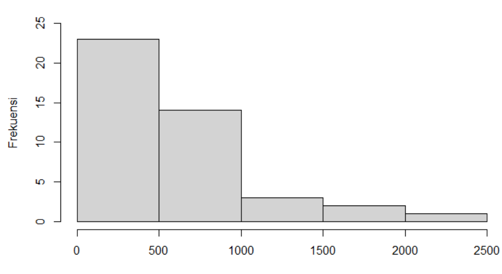


Fig. 4. Histogram of gastric cancer data.

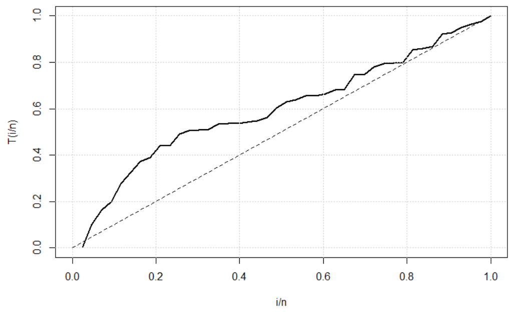
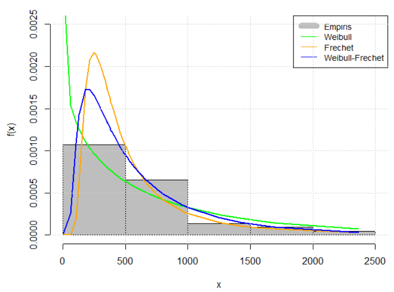


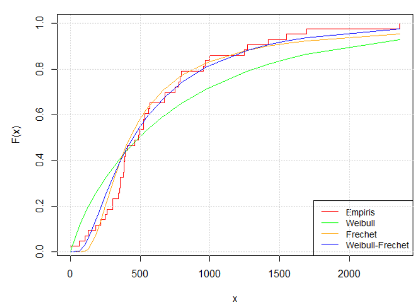
Fig. 5. TTT plot of gastric cancer data.

Table 3. Estimated Parameter for gastric cancer data.

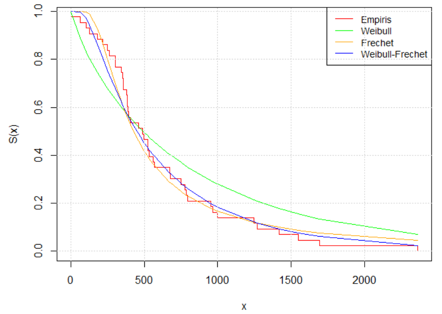
Distribution	**Estimates**			
	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$
Weibull	1.27359	0,85806	-	-
Fréchet	-	-	0.33844	1.56958
Weibull-Fréchet	1.27588	0.78096	0.52215	0.98389



(a)



(b)



(c)

Fig. 6. Graphical comparison between data and estimated distribution for (a) pdf with histogram, (b) cdf with empirical cdf, and (c) survival function with empirical survival function.

Table 4. Kolmogorov-Smirnov statistics and AIC for the distributions.

Distribution	Kolmogorov-Smirnov Test Statistics	AIC
Weibull	0.18181	521.26625
Fréchet	0.14136	431.93568
Weibull-Fréchet	0.14478	312.10688

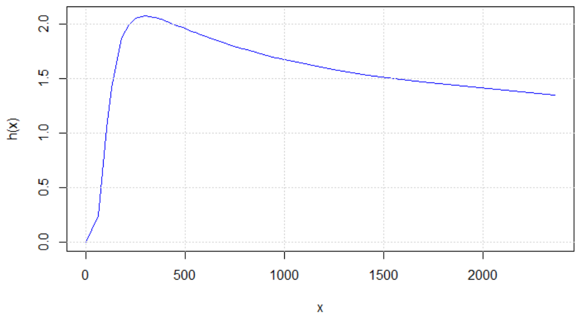


Fig. 7. Hazard function of WFr distribution based on gastric cancer data.

To get a better conclusion, we will compare The WFr distribution with both Weibull and Fréchet distribution using Kolmogorov-Smirnov test statistics and the Akaike Information Criterion (AIC). The AIC is quantified by $-2\ln L + 2q$, when $\ln L$ is the maximum value of log-likelihood function and q is the number of parameters of the distribution. The best model for the data would have the lowest value of AIC. The Kolmogorov-Smirnov statistics and AIC for all the distributions can be viewed in Table 4. From the results, we can conclude that the best distribution to model the gastric cancer data is the WFr distribution. It can be seen in Fig. 7 that the WFr distribution based on gastric cancer data has a unimodal shaped hazard function.

4 Conclusion

This paper introduces a new lifetime distribution called Weibull-Fréchet (WFr) distribution which is a generalization of Weibull distribution. WFr distribution is generated using the Weibull-G (WG) method and using cumulative distribution function of Fréchet distribution for G(x). The shape of hazard function of WFr distribution can be unimodal, which is an improvement from Weibull distribution. The characteristics of WFr distribution such as pdf, cdf, survival function, hazard function, moment, and quantile can be given explicitly. The mle method is used in estimating the parameters of WFr distribution. Modeling the illustration data results in WFr distribution to be the best-fit model for gastric cancer data, compared with both Weibull and Fréchet distribution.

Acknowledgements

The authors would like to acknowledge the assistance of all those who contributed to this paper. Our thanks go to the lecturers for their academic support and consultation to the research, and to our families and friends for their unwavering support and encouragement throughout this endeavour. We would also like to express our sincere gratitude to the Faculty of Mathematics and Natural Sciences of Universitas Indonesia for providing an opportunity and making this research possible.

References

1. J. P. Klein and M. L. Moeschberger, Survival analysis: techniques for censored and truncated data (Springer, New York, 2003).
2. D. H. Kutal and L. Qian, Stats **1**, 176–188 (2018).
3. M. Bourguignon, R. B. Silva, and G. M. Cordeiro, J. Data Sci. **12**, 53–68 (2014).
4. A. Z. Afify, H. M. Yousof, G. M. Cordeiro, E. M. M. Ortega, and Z. M. Nofal, J. Appl. Stat. **43**, 2608–2626 (2016).
5. A. Alzaatreh, C. Lee, and F. Famoye, Metron **71**, 63–79 (2013).