


Article

Applications of the Sine Modified Lindley Distribution to Biomedical Data

Lishamol Tomy ¹ , Veena G ² and Christophe Chesneau ^{3,*}

¹ Department of Statistics, Deva Matha College, Kuravilangad 686633, Kerala, India; lishatomy@gmail.com

² Department of Statistics, St. Thomas College, Palai 686574, Kerala, India; veenagpillai@hotmail.com

³ Laboratoire de Mathématiques Nicolas Oresme (LMNO), Université de Caen Normandie, Campus II, Science 3, 14032 Caen, France

* Correspondence: christophe.chesneau@gmail.com

Abstract: In this paper, the applicability of the sine modified Lindley distribution, recently introduced in the statistical literature, is highlighted via the goodness-of-fit approach on biological data. In particular, it is shown to be beneficial in estimating and modeling the life periods of growth hormone guinea pigs given tubercle bacilli, growth hormone treatment for children, and the size of tumors in cancer patients. We anticipate that our model will be effective in modeling the survival times of diseases related to cancer. The R codes for the figures, as well as information on how the data are processed, are provided.

Keywords: goodness-of-fit; biomedical data; Lindley distribution; trigonometric function; continuous distribution



Citation: Tomy, L.; G, V.; Chesneau, C. Applications of the Sine Modified Lindley Distribution to Biomedical Data. *Math. Comput. Appl.* **2022**, *27*, 43. <https://doi.org/10.3390/mca27030043>

Received: 5 April 2022

Accepted: 10 May 2022

Published: 16 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When people are diagnosed with cancer, COVID-19, or any other severe condition, or when clinical trials of a new treatment are conducted, survival is a major concern. Carcinoma is a general term that refers to a variety of diseases that can affect any part of the body. One of the risk factors is the abrupt development of aberrant cells that grow beyond their usual bounds, allowing them to infect nearby portions of the body and travel to other organs, which is one of the risk factors; this is known as metastasis. Widespread metastases are the leading cause of cancer death.

Doctors seek to regulate the growth and size of these tumors in the places where they arise in order to protect human lives. The tumor stage and survival times are two aspects, as they aid doctors in determining the best treatment for their patients. As a result, determining the probability distribution of tumor size and survival durations is crucial for selecting the best treatment option.

In order to analyze the numbers of people who are diagnosed with and die from severe diseases each year, the number of people who are currently living after the diagnosis of a disease, the mean age at which a disease was diagnosed, and the number of people who are still alive at a given time after diagnosis, statistics can be used. It also gives an idea of the differences among groups defined by age, sex, racial/ethnic group, geographic location, and other categories.

One such way of analyzing the properties of the survival data or the size of the tumor is by modeling the data. Data modelling related to biological science is of utmost importance to understanding the data statistically. Over the years, many researchers have developed discrete as well as continuous distributions that help in modelling biological data. Ref. [1] developed the Marshall–Olkin Inverse Lomax distribution (MO-ILD), which is used in modeling cancer stem cells. Ref. [2] studied the weighted generalized Quasi Lindley distribution, which was studied to model COVID-19 data from Algeria and Saudi Arabia, and Ref. [3] modeled the survival times of guinea pigs infected with virulent tubercle

bacilli using the Sine Half-Logistic Inverse Rayleigh distribution. With this motivation in mind, we use the existing sine-modified Lindley (S-ML) distribution developed by [4] in modelling data related to different types of cancer. We also provide optimized open source S-ML distribution codes for practitioners to use.

This paper is structured as follows. Section 2 covers a review of the existing S-ML distribution. Section 3 includes the application of the distribution to cancer data, as well as various visual presentations to back up the numbers, and Section 4 concludes the study.

2. The S-ML Distribution

In this section, a brief review of the definitions and properties of the sine generated (S-G or Sin-G in some references) family of distributions, the modified Lindley distribution, and the S-ML distribution is implemented. Due to their application and operating capability in a range of contexts, the families defined by “trigonometric transformations” have sparked a lot of interest in recent years. The sinusoidal transformation that contributes to the S-G family was initially studied by [5].

2.1. S-G Family of Distributions

The corresponding basic definitions of the associated distribution function (DF) and PDF given, respectively, by

$$F_{S-G}(y; \gamma) = \sin \left[\frac{\pi}{2} G(y; \gamma) \right], \quad y \in \mathbb{R}$$

and

$$f_{S-G}(y; \eta) = \frac{\pi}{2} g(y; \gamma) \cos \left[\frac{\pi}{2} G(y; \gamma) \right], \quad y \in \mathbb{R}$$

where $G(y; \gamma)$ and $g(y; \gamma)$ are the DF and PDF of a certain continuous distribution with parameter vector denoted by γ , respectively. These functions are linked to a reference or parent distribution that the practitioner determines ahead of time based on the study’s goals. The S-G family is well-known as a potential parent family alternative. Without introducing extra parameters, the following stochastic ordering holds: $G(y; \gamma) \leq F_{S-G}(y; \gamma)$ for every $y \in \mathbb{R}$. The S-G family provides the capability to develop flexible statistical models that can handle a variety of data. The recent works on the S-G family include the sine Lindley and the sine exponential distribution introduced by [6], the transformed S-G family studied by [7], the sine Topp Leone-G family of distributions developed by [8], sine Kumaraswamy-G family introduced by [9], the sine extended odd Fréchet-G family of distributions studied by [10], and the sine power Lomax model by [11].

Ref. [4] improved the S-G family’s performance by applying it to a specific one-parameter distribution established by [12]: the modified Lindley (ML) distribution. The S-ML distribution was developed as a result.

2.2. Modified Lindley Distribution

The ML distribution proposed by [12] is made possible by applying the tuning function $e^{-\beta y}$, $\beta > 0$ to the Lindley distribution with the goal of boosting its capabilities in a variety of domains. As a result, the ML distribution is defined by the DF expressed as follows:

$$G_{ML}(y; \beta) = 1 - \left[1 + \frac{\beta y}{1 + \beta} e^{-\beta y} \right] e^{-\beta y}, \quad y > 0.$$

The PDF is given by

$$g_{ML}(y; \beta) = \frac{\beta}{1 + \beta} e^{-2\beta y} \left[(1 + \beta) e^{\beta y} + 2\beta y - 1 \right], \quad y > 0,$$

respectively, with $\beta > 0$, and $G_{ML}(y; \beta) = g_{ML}(y; \beta) = 0$ for $y \leq 0$.

The ML distribution adapts to rising, reverse bathtub, and constant hazard rates and is a mixture of the exponential and gamma distributions with parameters β and $(2, 2\beta)$.

The practical benefit is very significant; for the three data sets shown in [12], the ML model outperforms the Lindley and exponential models. The wrapped modified Lindley distribution proposed by [13] and the inverted modified Lindley distribution proposed by [14] are two examples of improvements to the ML distribution.

2.3. S-ML Distribution

The corresponding DF and PDF of the S-ML distribution, respectively,

$$F_{S-ML}(y; \beta) = \cos \left[\frac{\pi}{2} \left(1 + e^{-\beta y} \frac{y\beta}{1+\beta} \right) e^{-\beta y} \right], \quad y > 0$$

and

$$f_{S-ML}(y; \beta) = \frac{\pi}{2} \frac{\beta}{1+\beta} e^{-2\beta y} \left[(1+\beta)e^{\beta y} + 2y\beta - 1 \right] \sin \left[\frac{\pi}{2} \left(1 + e^{-\beta y} \frac{y\beta}{1+\beta} \right) e^{-\beta y} \right], \quad y > 0,$$

with $\beta > 0$, and $F_{S-ML}(y; \beta) = f_{S-ML}(y; \beta) = 0$ for $y \leq 0$.

By varying the value of β , different variants of $f_{S-ML}(y; \beta)$ can be obtained. Figure 1 depicts the most representative of them.

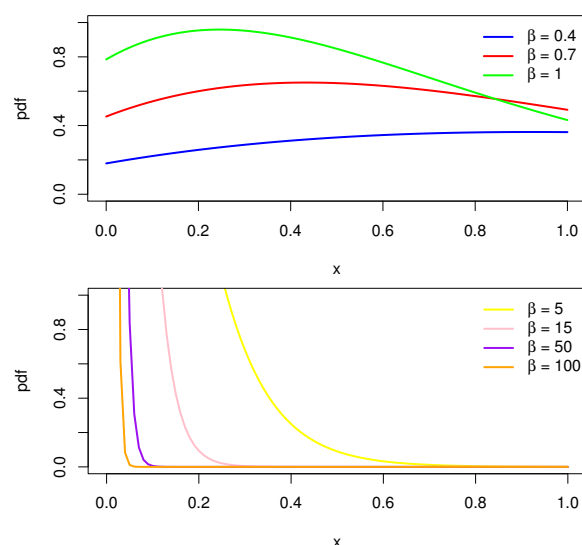


Figure 1. Illustration of $f_{S-ML}(y; \beta)$ with selected values of β .

We can imply from Figure 1, that for

smaller values of β , local increasing shape are seen; the distribution is unimodal, larger values of β , the plot of $f_{S-ML}(y; \beta)$ decreases and is leptokurtic in shape.

The shapes of the S-ML probability density function (PDF) are found to be adaptable to different shapes, being unimodal, decreasing, and right-skewed.

The S-ML distribution has also been shown to exhibit a non-monotonic hazard rate function (HRF), depicting an increasing-reverse bathtub-constant shape. The distribution's applicability and adaptability make it very appealing for modeling data from various fields and [4] has proved that the model stands strong against twelve other competent distributions, such as the generalized beta type 2 distribution introduced by [15], the Lomax distribution studied by [16], and the lognormal distribution developed by [17] in modelling data related to weather and engineering.

3. Applications

In the statistical literature on life-testing experiments, numerous distributions have been developed. Some of which can be used to model the increase or decrease in failure rates, while others can model bathtub and upside-down bathtub failure rates, and still others can do both. We have examined a few distributions in this case, which include the S-ML distribution against the sine-Lindley distribution (S-Lindley) defined by [18], the sine-exponential (S-Expo) distribution studied by [6], the inverse Lindley distribution (IL) introduced by [19], and the exponential (Expo) distribution as seen in [20].

The PDF and DF of the competing models used against the S-ML model are displayed in Table 1.

Table 1. DF and PDF of the competitive models used against the S-ML model.

Model	DF	PDF
S-Lindley	$\cos\left[\frac{\pi}{2}\left(1 + \frac{y\beta}{1+\beta}\right)e^{-\beta y}\right]$	$\frac{\pi}{2} \frac{\beta^2}{1+\beta} (1+y)e^{-\beta y} \sin\left[\frac{\pi}{2}\left(1 + \frac{\beta y}{1+\beta}\right)e^{-\beta y}\right]$
S-Expo	$\cos\left(\frac{\pi}{2}e^{-\beta y}\right)$	$\frac{\pi}{2} \beta \sin\left(\frac{\pi}{2}e^{-\beta y}\right)e^{-\beta y}$
IL	$\left(1 + \frac{\beta}{y(1+\beta)}\right)e^{-\beta/y}$	$\frac{\beta^2}{1+\beta} \left(\frac{1+y}{y^3}\right)e^{-\beta/y}$
Expo	$1 - e^{-\beta y}$	$\beta e^{-\beta y}$

3.1. Methodology

- We begin by investigating the descriptive measures of the modeled data-sets, which include the mean (μ), median (M), standard deviation (σ), skewness (γ_1) and the kurtosis (γ_2).
- A statistical analysis is conducted on the data-sets with the help of the statistical software [21]. The statistical analysis includes evaluating the estimate ($\hat{\beta}$) of the data by the method of maximum likelihood estimation, the related standard error (SE), and other statistical measures such as the goodness-of-fit (GOF) test statistics including Akaike Information criterion (AIC), Bayesian information criterion (BIC) along with Anderson Darling statistic (A^*), Cramér-von Mises statistic (w^*) and Kolmogorov–Smirnov statistic (D_n) with its corresponding p -value. The AIC is defined to be

$$AIC = 2k - 2ll,$$

the BIC is given by

$$BIC = k \log(n) - 2ll,$$

where ll denotes the log-likelihood function taken at the maximum likelihood estimate, n denotes the number of data and k represents the number of model parameters.

The model with the highest p -value and the lowest values for D_n , w^* , and A^* , as well as the AIC and BIC values, is the best fit for the data. It will be highlighted in the coming numerical tables with the blue color. The software R is used to conduct the estimation.

- Finally, for a visual representation, the empirical probability density function (EPDF) plots and the empirical cumulative density function (ECDF) plots, accompanied by the box plot and total time on test (TTT) plot, are displayed. The box plot gives a visual representation of the descriptive measures of the data and the TTT plot, proved useful for gaining information about the hazard form of the data. In many real-world situations, there is qualitative information about the shape of the failure rate function that might help in the selection of a particular distribution. The TTT plot has a convex shape for decreasing HRF and a concave shape for increasing HRF.

3.2. Survival Times of Growth Hormone Medication

The first data set consists of the estimated time from growth hormone medication until the children reached the target age in the Programa Hormonal de Secretaria de Saude de Minas Gerais in 2009, as reported in [22].

A summary of the measures of descriptive statistics is provided in Table 2 with the box and TTT plots plotted in Figure 2.

Table 2. Descriptive statistics of survival times of growth hormone medication.

μ	M	σ	γ_1	γ_2
5.979	5.260	2.810	0.851	3.119

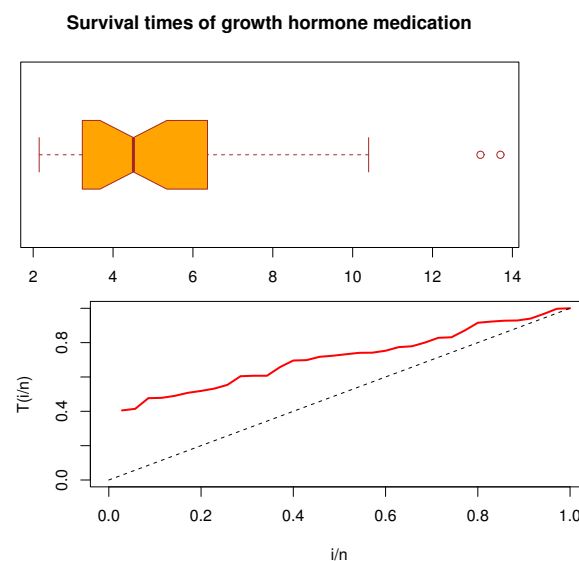


Figure 2. Box plot and TTT plot for the survival times of growth hormone medication.

Table 3 provides $\hat{\beta}$, the SE and the GOF metrics of the survival times of growth hormone medication.

Table 3. $\hat{\beta}$, SE and GOF metrics of the survival times of growth hormone medication.

Distribution	$\hat{\beta}$	SE	AIC	BIC	A^*	w^*	D_n	p -Value
S-ML	0.14870	0.01672	170.8874	172.4428	1.5424	0.23195	0.1978	0.1291
S-Lindley	0.22660	0.02426	173.38	174.9353	1.866	0.29474	0.22065	0.06620
S-Expo	0.10780	0.01691	185.8088	187.364	4.0824	0.77177	0.31926	0.00159
IL	4.9096	0.7251	186.918	188.4733	4.6262	0.8702	0.2905	0.00542
Exp	0.18848	0.03186	188.8149	190.3703	4.4891	0.85831	0.33317	0.00084

Statistical Analysis—Based on the information in Table 2, we can conclude that the data are positively skewed and mesokurtic, as evidenced by the box plot in Figure 2. The TTT plot of the survival times of the data set is displayed in Figure 2. It shows an increasing HRF plot. In addition, analysis of the data set shows that the evaluated model (S-ML) is the best model throughout all elements of the model selection criteria, such as the increasing hazard function. The S-ML model has a higher p -value and minimum values for the test statistics including the AIC, BIC, A^* , w^* , and D_n values, as shown in Table 3. The EPDF and ECDF plots are given in Figure 3.

The plots in Figure 3 display that the S-ML and S-Lindley models give a better fit to the data set than the S-Lindley, S-Expo, IL, and Expo models.

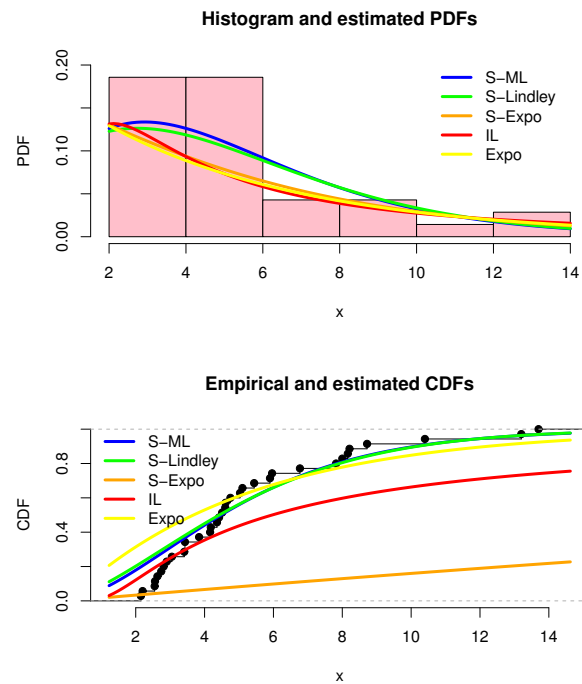


Figure 3. The EPDF and ECDF plot for the survival times of growth hormone medication.

3.3. Survival Times of Guinea Pigs Data

This data set was originally studied by [23], which has also been analyzed previously by [24]. The data set represents the survival times of $n = 72$ guinea pigs injected with different doses of tuberculosis bacilli. The main concern of this data set is to predict the survival times of the guinea pigs because they have a high susceptibility to human tuberculosis.

A summary of measures of descriptive statistics is provided in Table 4 with the box and TTT plots displayed in Figure 4.

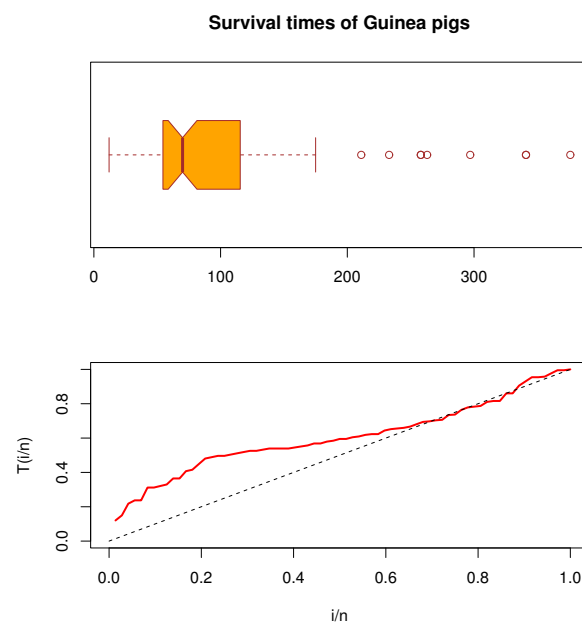


Figure 4. Box plot and TTT plot for the survival times of guinea pigs.

Table 4. Descriptive statistics of the survival times of guinea pigs.

μ	M	σ	γ_1	γ_2
99.82	70.00	81.11	1.796245	5.614

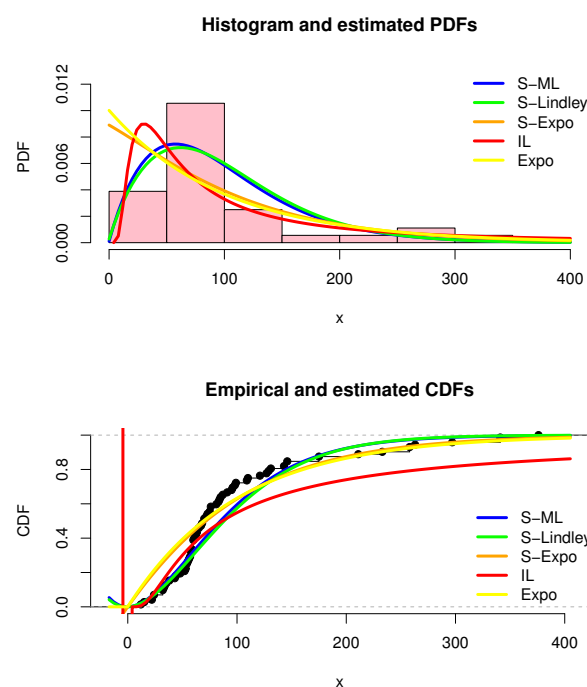
Table 5 displays $\hat{\beta}$, the SE and the GOF metrics for the survival times of guinea pigs.

Table 5. $\hat{\beta}$, SE and GOF metrics for the survival times of guinea pigs.

Distribution	$\hat{\beta}$	SE	AIC	BIC	A^*	w^*	D_n	p -Value
S-ML	0.0082780	0.0006544	791.9648	794.2415	2.022	0.3638	0.14389	0.1014
S-Lindley	0.013329	0.001003	795.0701	797.3468	2.6594	0.5020	0.16629	0.03728
S-Expo	0.00567	0.0006084	805.7007	807.9773	3.8741	0.68284	0.19629	0.0077
IL	61.066	7.084	807.3371	809.6137	4.590	0.8316	0.1845	0.01479
Expo	0.010018	0.001169	808.8843	811.1609	4.472	0.8059	0.2115	0.00317

Statistical Analysis—Table 4 informs us that the data are right-skewed and leptokurtic, as demonstrated by a graphical representation of the box plot in Figure 4. Figure 4 also illustrates the TTT plot of this data set. It displays an increasing HRF plot. Moreover, analysis of the data set implies that the S-ML distribution is the best model among the other competitive models, when statistical GOF criteria and the increasing HRF are considered. We can observe from Table 5, that the S-ML distribution has minimum values for the test statistics with a higher p -value and least values for GOF metrics. The EPDF and ECDF plots are displayed in Figure 5.

From Figure 5, we can also confirm this suitability behavior, as the plots of S-ML and S-Lindley distribution trace the shape of the data very well. We can conclude from Table 5 and Figure 5 that the S-ML model perfectly describes the survival times of guinea pigs.

**Figure 5.** EPDF and ECDF plots for the survival times of Guinea pigs.

3.4. Size of Tumors in Lung Cancer Patients

A swelling or tumor arises when the cells in the lungs expand at an abnormally fast rate, which can lead to lung cancer. It is possible to identify that and see if its spread to

other organs based on a variety of indicators. One of these characteristics is tumor stage, which aids doctors in determining the best treatment for their patients. The tumor size is used to determine the staging system. The data show the tumor size of 76 lung cancer patients at Tanta University's chest hospital, sixty of whom are in stage I, seven in stage II, and the rest in stage III.

A summary of measures of descriptive statistics is provided in Table 6 with the box and TTT plots plotted in Figure 6.

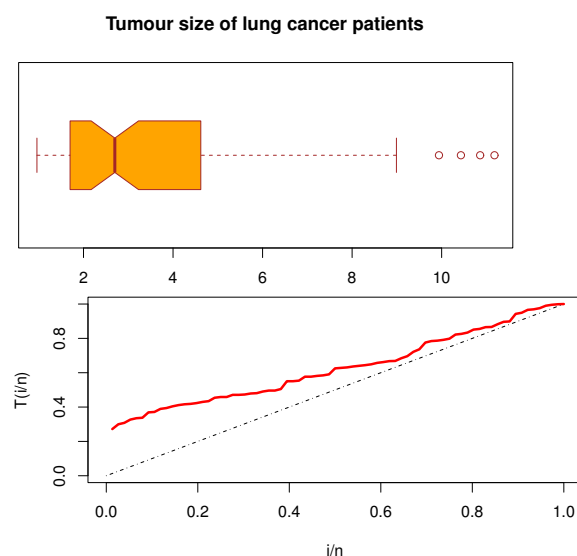


Figure 6. Box plot and TTT plot of the tumor size of the lung cancer patients.

Table 6. Descriptive statistics of the tumor size of the lung cancer patients.

μ	M	σ	γ_1	γ_2
3.531	2.700	2.570	1.442	4.269

Table 7 displays $\hat{\beta}$, the SE and the GOF metrics of the tumor size of the lung cancer patients.

Table 7. $\hat{\beta}$, SE and GOF metrics of the tumor size of the lung cancer patients.

Distribution	$\hat{\beta}$	SE	AIC	BIC	A^*	w^*	D_n	p -Value
S-ML	0.21916	0.02031	324.8085	327.1393	1.9164	0.2733	0.1239	0.1934
S-Lindley	0.3201	0.02373	329.8057	332.1365	2.3303	0.3321	0.15054	0.06381
S-Expo	0.16063	0.01724	341.6504	343.9812	4.5324	0.7384	0.2303	0.00063
IL	2.9640	0.2832	337.1246	339.4554	5.3844	0.9086	0.1823	0.01278
Expo	0.28313	0.03248	345.8022	348.133	5.2671	0.8884	0.2461	0.0002

Statistical Analysis—From Table 6, we see that the data are right-skewed and leptokurtic. This is proved in a graphical display of the box plot in Figure 6. Figure 6 shows the TTT plot of this data set. It illustrates an increasing HRF plot. From Table 7, the S-ML model has minimum values for D_n and higher p -value with least values for AIC and BIC. The EPDF and ECDF plots are illustrated in Figure 7.

The plots in Figure 7 show that S-ML distribution captures the shape of the histogram of the data set. We can conclude from Table 7 and Figure 7 that the S-ML distribution can be used to model this data set related to the size of tumors in lung cancer patients.

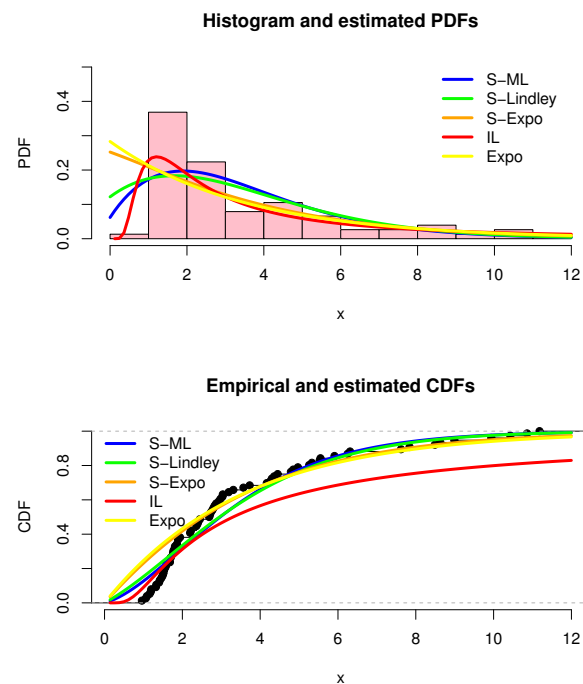


Figure 7. The EPDF and ECDF plot for the size of the tumor of the lung cancer patients.

4. Conclusions

In this paper, we have extended the applications of the sine-modified Lindley (S-ML) distribution developed by [4] to model biomedical data. The distribution yields the benefits of both the modified Lindley and S-G distributional functionalities. It was used to investigate the distribution of tumor size, patients diagnosed with cancer's survival durations, and medications provided. The AIC, BIC, and test statistics such as A^* , w^* , and D_n with their associated p -values are used to select the best-fitting model. These metrics are supported by a visual representation of how well the S-ML model fits the data, such as a box plot or a TTT plot. We believe the findings are superior to other competing distributions for modeling biomedical data and can be used to model a range of other biological data. We have also included the data sets and R codes for all of the figures in the paper, as well as all of the estimations, and the tests carried out. We refer readers to the Appendix A for these R codes.

Author Contributions: Conceptualization, L.T., V.G. and C.C.; methodology, L.T., V.G. and C.C.; software, L.T., V.G. and C.C.; validation, L.T., V.G. and C.C.; formal analysis, L.T., V.G. and C.C.; investigation, L.T., V.G. and C.C.; resources, L.T., V.G. and C.C.; data curation, L.T., V.G. and C.C.; writing—original draft preparation, L.T., V.G. and C.C.; writing—review and editing, L.T., V.G. and C.C.; visualization, L.T., V.G. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to thank the two referees for the constructive comments on the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this section, we have included the code to analyze data set 1, using the software R. The codes for the graphs in the data analysis are also plotted.

*Appendix A.1. Data Sets***Data set 1**

(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43, 3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10, 5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

Data set 2

(12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48, 52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 60, 61, 62, 63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131, 143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376)

Data set 3

(0.96, 1.06, 1.09, 1.16, 1.19, 1.20, 1.32, 1.33, 1.40, 1.42, 1.46, 1.49, 1.51, 1.52, 1.54, 1.57, 1.59, 1.68, 1.70, 1.70, 1.76, 1.76, 1.77, 1.80, 1.81, 1.86, 1.89, 1.89, 1.94, 2.20, 2.20, 2.22, 2.36, 2.36, 2.39, 2.41, 2.45, 2.69, 2.71, 2.73, 2.77, 2.80, 2.83, 2.87, 2.94, 2.98, 3.03, 3.04, 3.19, 3.31, 3.57, 3.73, 4.17, 4.27, 4.30, 4.36, 4.45, 4.79, 4.85, 4.97, 5.26, 5.33, 5.53, 5.55, 5.91, 6.25, 6.31, 7.62, 7.84, 8.49, 8.63, 8.99, 9.94, 10.43, 10.86, 11.18)

Appendix A.2. Graphics for the PDF of S-ML Distribution

```
x= 0:10
f= function(x,p) #defining the pdf of S-ML model
{
  (((pi/2)*(p/(1+p))*exp(-2*p*x))*(((1+p)*exp(p*x))+(2*p*x)-1)*sin((pi/2)*
  (1+(exp(-p*x)*((p*x)/(1+p))))*exp(-p*x)))
}

curve(f(x,p= 5),col="yellow",xlab="x", ylim=c(0,1),ylab="pdf",lwd=2 )
curve(f(x,p= 15),col="pink", lwd=2, add= TRUE)
curve(f(x,p= 50),col="purple", lwd=2, add= TRUE)
curve(f(x,p=100),col="orange", lwd=2, add= TRUE)

legend("topright",legend=c(expression(paste(beta," = ",5)),
expression(paste(beta," = ",15)),
expression(paste(beta," = ",50)),
expression(paste(beta," = ",100))),
ncol=1, col=c("yellow","pink","purple", "orange"),
lwd=c(2,2,2,2), cex=c(1,1,1,1),text.width = 0.1, inset=0.011, bty ="n")
### In the same way, we can plot the pdf for other beta~values.
```

Appendix A.3. Parameter Estimate along with GOF Metrics

```
install.packages (c("EstimationTools", "MASS", "plyr" ))
library(EstimationTools)
library(MASS)
library(plyr)

# Data set 1
st = c(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43,
      3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10,
      5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

# S-ML distribution
dSml = function(x, p, log = FALSE) #log of pdf of S-ML model
{
  n=count(x)
  loglik <- (log(pi/2)+log(p)-log(1+p)-(2*p*x)+log((1+p)*exp(x*p)*(2*p*x)-1)+
  log(sin((pi/2)*(1+exp(-p*x)*(p*x)/(1+p))*exp(-x*p))))
  if ( log == FALSE)
```

```

density <- exp(loglik)
else density <- loglik
return(density)
}

theta <- maxlogL(x =st, dist = "dSml",start = 0.59)

summary(theta)

# S-Lindley~distribution

dSL = function(x, p, log = FALSE) #log of pdf of S-Lindley model
{
n=count(x)
loglik <- (log(pi/2)+log(p^2)+log(1+x)-log(1+p)-(p*x)+
log(sin((pi/2)*(1+((x*p)/(1+p))))*exp(-x*p))))
if ( log == FALSE)
density <- exp(loglik)
else density <- loglik
return(density)
}

theta <- maxlogL(x =st, dist = "dSL",start = 0.59)
summary(theta)

# SE distribution
dSE = function(x, p, log = FALSE) #log of pdf of SE model
{
n=count(x)
loglik <- (log(pi/2)+log(p)-(p*x)+log(sin((pi/2)*exp(-x*p))))
if ( log == FALSE)
density <- exp(loglik)
else density <- loglik
return(density)
}
theta <- maxlogL(x =st, dist = "dSE",start = 0.59)

summary(theta)

#IL distribution
dIL = function(x,p, log = FALSE)
{
n=count(x)
loglik <- (2*log(p))-log(1+p)-(p/x)+log(1+x)-(3*log(x))
if ( log == FALSE)
density <- exp(loglik)
else density <- loglik
return(density)
}

theta <- maxlogL(x =st, dist = "dIL",start = 0.65)

summary(theta)

```

```

#Exp distribution
dE = function(x,p, log = FALSE) #log of pdf of Expo model
{
  n=count(x)
  loglik <- log(dexp(x, p, log = FALSE))
  if ( log == FALSE)
    density <- exp(loglik)
  else density <- loglik
  return(density)
}

st = c(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43,
      3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10,
      5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

theta <- maxlogL(x =st, dist = "dE",start = 0.6)

summary(theta)

```

Appendix A.4. KS Test Statistic, p-Value and Other Test Statistics

```

install.packages ("gofest")
library(gofest)
y = c(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43,
      3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10,
      5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

# S-ML~CDF

pSml = function(x,p)
{
  p = 0.14870
  cos((pi/2)*(1+(exp(-p*x)*(x*p)/(1+p)))*exp(-p*x))
}

ks1=ad1=cvm1=NULL
ks1=ks.test(y,pSml)
ad1=ad.test(y,pSml)
cvm1=cvm.test(y,pSml)
result1=c(ks1$statistic,ks1$p.value,ad1$statistic,ad1$p.value,
cvm1$statistic,cvm1$p.value)

# S-Lindley~CDF

pSL = function(x,p)
{
  p = 0.22660
  cos((pi/2)*(1+((x*p)/(1+p)))*exp(-p*x))
}

ks1=ad1=cvm1=NULL
ks1=ks.test(y,pSL)
ad1=ad.test(y,pSL)
cvm1=cvm.test(y,pSL)
result1=c(ks1$statistic,ks1$p.value,ad1$statistic,ad1$p.value,

```

```

cvm1$statistic,cvm1$p.value)

# S-Expo CDF
pSe = function(x,p)
{
  p = 0.10780
  cos((pi/2)*exp(-p*x))
}

ks1=ad1=cvm1=NULL
ks1=ks.test(y,pSe)
ad1=ad.test(y,pSe)
cvm1=cvm.test(y,pSe)
result1=c(ks1$statistic,ks1$p.value,ad1$statistic,ad1$p.value,
cvm1$statistic,cvm1$p.value)

#IL distribution
pIL = function(x,p)
{
  p = 4.9096
  (1 + (p/((1+p)*x)))*exp(-p/x)
}
ks1=ad1=cvm1=NULL
ks1=ks.test(y,pIL)
ad1=ad.test(y,pIL)
cvm1=cvm.test(y,pIL)
result1=c(ks1$statistic,ks1$p.value,ad1$statistic,ad1$p.value,
cvm1$statistic,cvm1$p.value)

# Expo~CDF

pEx = function(x,p)
{
  p = 0.18848
  pexp(x,p)
}

ks1=ad1=cvm1=NULL
ks1=ks.test(y,pEx)
ad1=ad.test(y,pEx)
cvm1=cvm.test(y,pEx)
result1=c(ks1$statistic,ks1$p.value,ad1$statistic,ad1$p.value,
cvm1$statistic,cvm1$p.value)

```

Appendix A.5. Graphics—To Plot the EPDF for the First Data Set

```

x = c(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43,
3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10,
5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

hist(x,prob=T,main="Histogram and estimated PDFs",
col="pink", ylab = "PDF",ylim=c(0,0.05), bty ="n")

p = 0.14870 ## parameter estimate of S-ML model

```

```

curve(((pi/2)*(p/(1+p))*(exp(-2*p*x))*((1+p)*exp(p*x)+(2*x*p)-1)*
(sin((pi/2)*(1+exp(-p*x))*((x*p)/(1+p)))*exp(-x*p)))),col="blue",lwd=3,
add=T)

p = 0.22660 # parameter estimate of S-Lindley
curve(((pi/2)*((p^2)/(1+p))*(1+x)*exp(-p*x)*(sin((pi/2)*
(1+((x*p)/(1+p)))*exp(-x*p))))), col="green",lwd = 3, add=T)

p = 0.10780 # parameter estimate of S-Expo
curve(((pi/2)*(p*exp(-p*x))*(sin((pi/2)*exp(-x*p))))), col="orange",
lwd = 3, add=T)

p = 4.9096 # parameter estimate of IL
curve((((p^2)/(1+p))*((1+x)/x^3)*exp(-p/x)),col="red",lwd = 3, add=T)

p = 0.18848 # parameter estimate of Expo
curve(dexp(x,p), col="yellow", lwd = 3, add = T)

legend("topright",legend = c("S-ML","S-Lindley","S-Expo","IL","Expo"),
ncol = 1,
col= c("blue","green","orange","red","yellow"),lty =1,lwd=3,
text.width = 2.5 , inset= 0.00005, bty ="n")

```

Appendix A.6. Graphics—To Plot the ECDF for the First Data Set

```

y = c(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43,
3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10,
5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

plot(ecdf(y) , verticals=TRUE, main="Empirical and estimated CDFs",
ylab="CDF", xlab="x", bty ="n")

p = 0.14870 # parameter estimate of S-ML
curve((cos((pi/2)*(1+(exp(-p*x))*((x*p)/(1+p)))*exp(-p*x))),col="blue",
lwd=3, add=T)

p = 0.22660 # parameter estimate of S-Lindley
curve((cos((pi/2)*(1+((x*p)/(1+p)))*exp(-p*x))), col="green",lwd = 3,
add=T)

p = 0.010780 # parameter estimate of S-Expo
curve((cos((pi/2)*exp(-p*x))),col="orange",lwd = 3, add=T)

p = 4.9096 # parameter estimate of IL
curve(((1 + (p/((1+p)*x)))*exp(-p/x)),col="red",lwd = 3, add=T)

p = 0.18848 # parameter estimate of Expo
curve(pexp(x,p), col="yellow", lwd = 3, add = T)

legend("topleft",legend = c("S-ML","S-Lindley","S-Expo","IL","Expo"),
ncol = 1,
col= c("blue","green","orange","red","yellow"),lty =1,lwd=3,
text.width = 2.5 , inset= 0.00005, bty ="n")
#####

```

Appendix A.7. Graphics: Bar Plot and TTT Plot for First Data Set

```
###Bar plot
x = c(2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43,
      3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51, 4.60, 4.61, 4.75, 5.03, 5.10,
      5.44, 5.90, 5.96, 6.77, 7.82, 8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70)

boxplot(x,main = "Tumour size of lung cancer patients",
        col = "orange", border="brown",horizontal = TRUE,notch = TRUE)

###TTT plot
install.packages ("AdequacyModel")
library(AdequacyModel)

TTT(x, lwd = 2, lty 2, col = "red", grid=FALSE)

#####
```

References

- Maxwell, O.; Chukwu, A.U.; Oyamakin, O.S.; Khaleel, M.A. The Marshall–Olkin inverse Lomax distribution (MO-ILD) with application on cancer stem cell. *J. Adv. Math. Comput. Sci.* **2019**, *33*, 1–12. [\[CrossRef\]](#)
- Benchiha, S.; Al-Omari, A.I.; Alotaibi, N.; Shrahili, M. Weighted generalized quasi Lindley distribution: Different methods of estimation, applications for COVID-19 and engineering data. *AIMS Math.* **2021**, *6*, 11850–11878. [\[CrossRef\]](#)
- Shrahili, M.; Elbatal, I.; Elgarhy, M. Sine Half-Logistic Inverse Rayleigh Distribution: Properties, Estimation, and Applications in Biomedical Data. *J. Math.* **2021**, *2021*, 4220479. [\[CrossRef\]](#)
- Tomy, L.; G, V.; Chesneau, C. The sine modified Lindley distribution. *Math. Comput. Appl.* **2021**, *26*, 81. [\[CrossRef\]](#)
- Souza, L.; Junior, W.R.O.; de Brito, C.C.R.; Chesneau, C.; Ferreira, T.A.E.; Soares, L. On the Sin-G class of distributions: Theory, model and application. *J. Math. Model.* **2019**, *7*, 357–379.
- Kumar, D.; Singh, U.; Singh, S.K. A new distribution using sine function—Its application to bladder cancer patients data. *J. Stat. Appl. Probab.* **2015**, *4*, 417.
- Jamal, F.; Chesneau, C.; Bouali, D.L.; Ul Hassan, M. Beyond the Sin-G family: The transformed Sin-G family. *PLoS ONE* **2021**, *16*, e0250790. [\[CrossRef\]](#) [\[PubMed\]](#)
- Al-Babtain, A.A.; Elbatal, I.; Chesneau, C.; Elgarhy, M. Sine Topp-Leone-G family of distributions: Theory and applications. *Open Phys.* **2020**, *18*, 574–593. [\[CrossRef\]](#)
- Chesneau, C.; Jamal, F. The sine Kumaraswamy-G family of distributions. *J. Math. Ext.* **2020**, *15*, 1–33.
- Jamal, F.; Chesneau, C.; Aidi, K. The sine extended odd Fréchet-G family of distribution with applications to complete and censored data. *Math. Slovaca* **2021**, *71*, 961–982. [\[CrossRef\]](#)
- Nagarjuna, V.B.; Vardhan, R.V.; Chesneau, C. On the accuracy of the sine power Lomax model for data fitting. *Modelling* **2021**, *2*, 78–104. [\[CrossRef\]](#)
- Chesneau, C.; Tomy, L.; Gillariose, J. A new modified Lindley distribution with properties and applications. *J. Stat. Manag. Syst.* **2021**, *24*, 1383–1403. [\[CrossRef\]](#)
- Chesneau, C.; Tomy, L.; Jose, M. Wrapped modified Lindley distribution. *J. Stat. Manag. Syst.* **2021**, *24*, 1025–1040. [\[CrossRef\]](#)
- Chesneau, C.; Tomy, L.; Gillariose, J.; Jamal, F. The inverted modified Lindley distribution. *J. Stat. Theory Pract.* **2020**, *14*, 46. [\[CrossRef\]](#)
- Kalbfleisch, J.D.; Prentice, R.L. *The Statistical Analysis of Failure Time Data*; Wiley: New York, NY, USA, 1980.
- Lomax, K.S. Business Failures: Another Example of the Analysis of Failure Data. *J. Am. Stat. Assoc.* **1954**, *49*, 847–852. [\[CrossRef\]](#)
- Aitchison, J.; Brown, J.A.C. *The Lognormal Distribution*; Cambridge University Press: Cambridge, UK, 1957.
- Kumar, D.; Singh, U.; Singh, S.K.; Chaurasia, P.K. Statistical properties and application of a lifetime model using sine function. *Int. J. Creat. Res. Thoughts (IJCRT)* **2018**, *6*, 993–1002.
- Sharma, V.K.; Singh, S.K.; Singh, U.; Agiwal, V. The inverse Lindley distribution: A stress-strength reliability model with application to head and neck cancer data. *J. Ind. Prod. Eng.* **2015**, *32*, 162–173. [\[CrossRef\]](#)
- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; John Wiley & Sons: Hoboken, NJ, USA, 1995; Volume 2.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2005; ISBN 3-900051-07-0.
- Alizadeh, M.; Bagheri, S.; Bahrami, S.E.; Ghobadi, S.; Nadarajah, S. Exponentiated power Lindley power series class of distributions: Theory and applications. *Commun. Stat. Simul. Comput.* **2018**, *47*, 2499–2531. [\[CrossRef\]](#)

-
23. Bjerkedal, T. Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *Am. J. Hyg.* **1960**, *72*, 130–148. [[PubMed](#)]
 24. Gupta, R.C.; Kannan, N.; RayChoudhuri, A. Analysis of lognormal survival data. *Math. Biosci.* **1997**, *139*, 103–115. [[CrossRef](#)]