

R Package for the Two-Parameters Crack Distribution

Wikanda Phaphan

Department of Applied Statistics
Faculty of Applied Science
King Mongkuts University of Technology North Bangkok
Bangkok, Thailand

email: wikanda.p@sci.kmutnb.ac.th

(Received March 24, 2021, Revised April 26, 2021, Accepted May 31, 2021)

Abstract

The two-parameter crack distribution was introduced by Saengthong and Bodhisuwan [4] and this distribution was a combination between the inverse Gaussian and the length biased inverse Gaussian distribution. This distribution consists of two parameters: (1) a scale parameter, and (2) a shape parameter. In this paper, we are interested in creating a new R language package for a two-parameter crack distribution, called the **twoCrack** package which has many applications in engineering and science. The new R-language package consists of a probability density function, distribution function, and random generation procedure for Two-parameter Crack distribution. Additionally, it includes a probability density function and random numbers generation procedure of sub-distribution such as Inverse Gaussian, Length Biased Inverse Gaussian, and Birnbaum-Saunders distribution. The use of the package is illustrated using a vertical roller mill vibration data set.

1 Introduction

The measured data is in the form of a period of time such as calculating the useful life of equipment or the period of time of equipment or machines

Key words and phrases: Inverse Gaussian distribution, length biased inverse Gaussian distribution, Birnbaum-Saunders distribution, crack distribution, twoCrack.

AMS (MOS) Subject Classifications: 62-04

ISSN 1814-0432, 2021, <http://ijmcs.future-in-tech.net>

expired, the period of time of the damaged machines. These data are important in the fields of engineering and science. The obtained values from measurement might be in the set of the number of times or rounding numbers of a failure of equipment and machines or it might be in the set of a unit of time-second, minute, or hour. The mentioned data have a particular name: lifetime data. In statistics, the obtained values from the calculations are the values of random variables derived from a random experiment.

In statistics, the rounding numbers or days of the failure of equipment and machines are discrete random variables. Therefore, the probability model consistent and used to describe the occurrence of random variables is the probability mass function like Bernoulli, Binomial, and Poisson distributions. Also, the failure period of time with a unit of time (second, minute, or hour) is a continuous random variable which the probability density function, used to explain the occurrence of random variables, often use an exponential distribution, Weibull distribution, Log-normal distribution, Inverse Gaussian distribution, Log-Logistic distribution, Length Biased Inverse Gaussian distribution, Birnbaum-Saunders distribution, Crack distribution, and so forth. For related fields with the widely known above-distribution have particular names which are called lifetime or failure distribution.

A crack distribution was first introduced by Jorgensen et al. [1] and it was called “A mixed distribution between an inverse Gaussian distribution and its complementary reciprocal.” Then, in 2011, Bowonrattanasat and Budsaba [2] changed the new distribution’s parameters and they calculated its various properties as well and called it “The crack distribution”. This distribution is a continuous distribution with random variables in value greater than or equal 0 and so it is classified as the group set of the lifetime distribution. Later on, Saengthong and Bodhisuwan [4] reduced the number of parameters of the crack distribution and they called it “The two-parameter crack distribution” and estimated the parameter value by using the Newton-Raphson method, including calculating various properties of two-parameter crack distribution.

In addition, nowadays, crack distribution is a well-known and widespread distribution. In this paper, we present the new R package **twoCrack** for the two-parameter crack distribution in order to easily apply it to engineering and science. The **twoCrack** package is composed of a probability density, distribution function, and random generation for two-parameter crack distribution. Also included is a probability density and random numbers generation of sub-distribution such as inverse Gaussian, length biased inverse Gaussian, and Birnbaum-Saunders distribution.

2 A Brief Review of the Birnbaum-Saunders Distribution

The Birnbaum-Saunders distribution arises as a fatigue lifetime model. This distribution used to explain how material failure occurs in the development and growth of a dominant crack. The classical two-parameter Birnbaum-Saunders distribution was introduced by Birnbaum-Saunders [5, 6]. If a random variable X has the Birnbaum-Saunders distribution, then its probability density function (p.d.f.) can be written as:

$$f_{BS}(x; \alpha, \beta) = \begin{cases} \frac{x+\beta}{2\alpha(2\pi\beta)^{\frac{1}{2}}x^{\frac{3}{2}}} \exp\left[-\frac{1}{2\alpha^2}\left(\frac{x}{\beta} + \frac{\beta}{x} - 2\right)\right] & ; x > 0 \\ 0 & ; otherwise, \end{cases} \quad (2.1)$$

where $\alpha > 0$ and $\beta > 0$ are the shape and the scale parameter or median. Ahmed et al. [3] introduced the new parameterization of the Birnbaum-Saunders distribution. The re-parameterized distribution is important in fitting the physical phenomena of fatigue cracks. The parameter $\lambda > 0$ corresponds to the thickness of the machine element and $\theta > 0$ corresponds to the nominal treatment pressure on the machine element. The relation between classical parameters α, β and the adjusted parameters λ, θ in physical interpretation can be expressed as:

$$\begin{aligned} \lambda &= \frac{1}{\alpha^2} \quad \text{and} \quad \theta = \alpha^2\beta; \\ \alpha &= \frac{1}{\sqrt{\lambda}} \quad \text{and} \quad \beta = \alpha\theta. \end{aligned}$$

The probability density function of a re-parameterized Birnbaum-Saunders distribution, denoted as $f_{BS}(\lambda, \theta)$, is

$$f_{BS}(x; \lambda, \theta) = \begin{cases} \frac{1}{2\theta\sqrt{2\pi}} \left[\lambda \left(\frac{\theta}{x}\right)^{\frac{3}{2}} + \left(\frac{\theta}{x}\right)^{\frac{1}{2}} \right] \exp\left[-\frac{1}{2}\left(\sqrt{\frac{x}{\theta}} - \lambda\sqrt{\frac{\theta}{x}}\right)^2\right] & ; x > 0 \\ 0 & ; otherwise. \end{cases} \quad (2.2)$$

3 A Brief Review of the Inverse Gaussian Distribution

Chikara and Folk [7] published the classical parameter of the inverse Gaussian distribution which is a two-parameter family of continuous probability

distributions with support on $(0, \infty)$. Suppose X is a random variable with an inverse Gaussian distribution. Then, a probability density function can be written in this formula:

$$f_{IG}(x; \mu, \beta) = \begin{cases} \sqrt{\frac{\beta}{2\pi}} x^{-\frac{3}{2}} \exp\left(-\frac{\beta(x-\mu)^2}{2\mu^2 x}\right) & ; x > 0 \\ 0 & ; \textit{otherwise.} \end{cases} \quad (3.3)$$

where parameter $\mu > 0$ is a location parameter or a mean of the distribution and $\beta > 0$ is a scale parameter. The probability density function of re-parameterized inverse Gaussian based on Ahmed et al. [3], denoted as $f_{IG}(\lambda, \theta)$, can be written in the following form:

$$f_{IG}(x; \lambda, \theta) = \begin{cases} \lambda \sqrt{\frac{\theta}{2\pi}} x^{-\frac{3}{2}} \exp\left(-\frac{(x-\lambda\theta)^2}{2\theta x}\right) & ; x > 0 \\ 0 & ; \textit{otherwise.} \end{cases} \quad (3.4)$$

$$f_{IG}(x; \lambda, \theta) = \begin{cases} \frac{\lambda}{\theta\sqrt{2\pi}} \left(\frac{\theta}{x}\right)^{\frac{3}{2}} \exp\left[-\frac{1}{2} \left(\sqrt{\frac{x}{\theta}} - \lambda\sqrt{\frac{\theta}{x}}\right)^2\right] & ; x > 0 \\ 0 & ; \textit{otherwise.} \end{cases} \quad (3.5)$$

The adjusted parameters are $\lambda > 0$ and $\theta > 0$ corresponding to the thickness of the machine element and nominal treatment pressure on the machine element, respectively. The relations between classical parameters α, β and adjusted parameters λ, θ is

$$\begin{aligned} \lambda &= \frac{\beta}{\mu} \quad \textit{and} \quad \theta = \frac{\mu^2}{\beta}; \\ \mu &= \lambda\theta \quad \textit{and} \quad \beta = \lambda^2\theta. \end{aligned}$$

4 A Brief Review of the Length Biased Inverse Gaussian Distribution

Define X as a non-negative random variable which has the length biased inverse Gaussian distribution. Then, the probability density function of the length biased inverse Gaussian distribution based on distribution as [3, 8], denoted as $f_{LBIG}(\lambda, \theta)$, is given by the following formula:

$$f_{LBIG}(x; \lambda, \theta) = \begin{cases} \frac{1}{\theta\sqrt{2\pi}} \left(\frac{\theta}{x}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2} \left(\sqrt{\frac{x}{\theta}} - \lambda\sqrt{\frac{\theta}{x}}\right)^2\right] & ; x > 0 \\ 0 & ; \textit{otherwise,} \end{cases} \quad (4.6)$$

where parameters $\lambda > 0$ and $\theta > 0$ are the shape and scale parameter, respectively.

5 A Brief Review of the Two-parameter Crack Distribution

The inverse Gaussian mixture distribution, which was introduced by Jorgensen et al. [1], was a combination of two distributions. They are inverse Gaussian and length biased inverse Gaussian distribution in the form of original parameters. Later, a combination of inverse Gaussian and length biased inverse Gaussian distributions in the form of adjusted parameters based on the distribution as Ahmed et al. [3] was introduced by Bowonrattanaset and Budsaba [2]. Those parameters consisted of three parameters: (1) a scale parameter (2) a shape parameter, and (3) a weighted parameter. Also, this distribution is called a crack distribution. Moreover, the various properties of this distribution were calculated. Then, in 2014, Saengthong and Bodhisuwan [4] reduced the number of parameters of the crack distribution to two parameters for easier estimate parameter. In addition, they estimated a parameter value by using the Newton-Raphson method, including calculating various properties of distribution. This research received attention for studying the crack distribution based on the crack distribution of Saengthong and Bodhisuwan [4]. Consequently, the probability density function of the crack distribution, denoted as $f_{TCR}(\lambda, \theta)$, can be written in this form:

$$f_{TCR}(x; \lambda, \theta) = \left(\frac{\theta}{\theta + 1}\right) f_{IG}(x; \lambda, \theta) + \left(\frac{1}{\theta + 1}\right) f_{LBIG}(x; \lambda, \theta), \quad (5.7)$$

where $\lambda > 0, \theta > 0$ and $x > 0$. Eq. (5.7) can be expressed as

$$f_{TCR}(x; \lambda, \theta) = \begin{cases} \frac{1}{\theta(\theta+1)\sqrt{2\pi}} \left[\lambda\theta \left(\frac{\theta}{x}\right)^{\frac{3}{2}} + \left(\frac{\theta}{x}\right)^{\frac{1}{2}} \right] \exp \left[-\frac{1}{2} \left(\sqrt{\frac{x}{\theta}} - \lambda\sqrt{\frac{\theta}{x}} \right)^2 \right] & ; x > 0 \\ 0 & ; otherwise. \end{cases} \quad (5.8)$$

Here, $\lambda > 0, \theta > 0$ and this distribution will be denoted by $TCR(\lambda, \theta)$. The cumulative distribution function of $X \sim TCR(\lambda, \theta)$ is

$$F_{TCR}(\lambda, \theta, p) = \begin{cases} \Phi \left(\sqrt{\frac{x}{\theta}} - \lambda\sqrt{\frac{\theta}{x}} \right) - \left(\frac{1-\theta}{\theta+1}\right) e^{2\lambda} \left[1 - \Phi \left(\sqrt{\frac{x}{\theta}} + \lambda\sqrt{\frac{\theta}{x}} \right) \right] & ; x > 0 \\ 0 & ; x \leq 0. \end{cases} \quad (5.9)$$

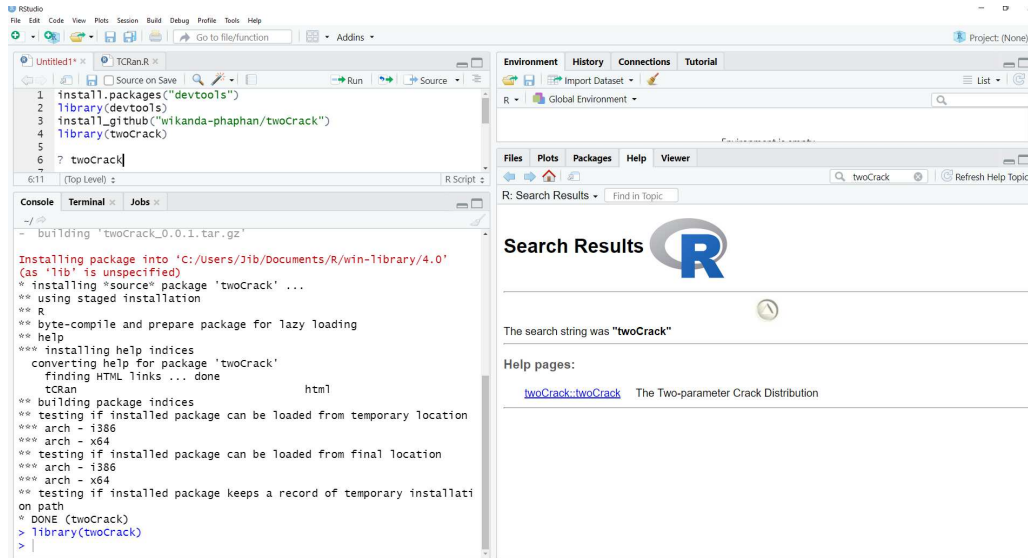
6 A New R Package for the Two-parameters Crack Distribution

We present a new contributed package for R [9] that is the **twoCrack** package. The first objective in developing the **twoCrack** package was to provide R users a set of functions that making it easy to use the probability density function such as fitting the two-parameter crack distribution or sub-distribution as an inverse Gaussian, length biased inverse Gaussian, and Birnbaum-Saunders distribution to real data set or to determine the probabilities of events, when a random variable follows above distribution. We also provide R users a set of functions for generating random numbers that follow the two-parameter crack distribution and sub-distribution. The **twoCrack** package requires a version of R $\geq 3.6.1$ and is published only on GitHub at <https://github.com/wikanda-phaphan/twoCrack>, there is a function that allows users to install it directly-`devtools::install_github` from the **devtools** package. Therefore, the step to install the **twoCrack** package on R [9] or RStudio [10] is as follows:

```
install.packages("devtools")
library(devtools)
install_github("wikanda-phaphan/twoCrack")
library(twoCrack)
```

After installing the **twoCrack** package, users can use “`? twoCrack`” command for getting an R package manual, see Figures (1) and (2) . This new package includes the four distributions and the corresponding calling sequences to compute the probability density function, random numbers, and distribution function for some distribution. The **twoCrack** package consists of nine functions:

1. `dTwoCrack(X,lambda,theta)`, the `dTwoCrack` function gives the probability density of the two-parameter crack distribution in Eq. (5.8),
2. `pTwoCrack(X,lambda,theta)`, the `pTwoCrack` function gives the distribution function of the two-parameter crack distribution in Eq. (5.9),
3. `rTwoCrack(n,lambda,theta)`, the `rTwoCrack` function generates random numbers of the two-parameter crack distribution by using the acceptance-rejection method proposed by Phaphan and Pongsart [11],
4. `dBS(X,lambda,theta)`, the `dBS` function gives the probability density of the Birnbaum-Saunders distribution in Eq. (2.2),
5. `rBS(n,lambda,theta)`, the `rBS` function generates random numbers of the Birnbaum-Saunders distribution with the composition method [12],

Figure 1: Installing the `twoCrack` package in RStudio

6. $dIG(X, \lambda, \theta)$, the `dIG` function gives the probability density of the inverse Gaussian distribution in Eq. (3.5),

7. $rIG(n, \lambda, \theta)$, the `rIG` function generates random numbers of the inverse Gaussian distribution employing the composition method [12],

8. $dLBIG(X, \lambda, \theta)$, the `dLBIG` function gives the probability density of the length biased inverse Gaussian distribution in Eq. (4.6),

9. $rLBIG(n, \lambda, \theta)$, the `rLBIG` function generates random numbers of the length biased inverse Gaussian distribution employing the composition method [12].

In the calling sequence for using the functions X must be a vector of data values, λ must be a value of the parameter λ , θ must be a value of the parameter θ , and n must be a number of observations.

For illustrating the use of various functions of the `twoCrack` package with real data, we will use the vibration of the vertical roller mill in 60 minutes. This data set contains values of mill vibration (μm) and day-month-year-time (a.m.), collected from the Siam Cement (Ta Luang) company limited, Khao Wong plant in February 2019, see Table 1. Before fitting one or more distributions to a data set, it is generally necessary to choose good candidates among a predefined set of distributions by considering the histogram and distribution function plots of an empirical distribution. From Figure (3), it clear that the data is right-skewed data. Hence, we choose the two-parameter crack,

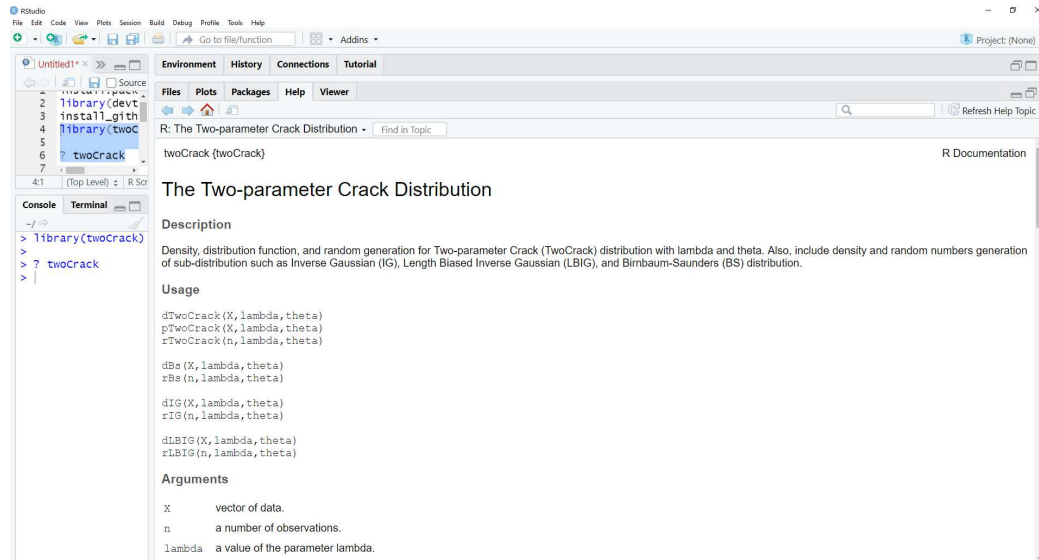


Figure 2: The manual of the **twoCrack** package in RStudio

inverse Gaussian, length biased inverse Gaussian, and Birnbaum-Saunders distribution to be candidate distribution for the mill vibration data. We estimate the distribution parameters of each distribution by maximizing the likelihood function using the `nlminb` function and consider the Akaike's information criterion (AIC) statistics to compare the fitting of distribution, the following R commands will produce the result in Table 2

```
#Example for find MLE by using nlminb function and twoCrack package
data1 <- read.csv("Vibration.csv")
X<-c(data1[1:60,1])
library(twoCrack)
lik.BS <- function(theta){-sum(dBS(X,theta[1],theta[2]))}
lik.IG <- function(theta){-sum(dIG(X,theta[1],theta[2]))}
lik.LBIG <- function(theta){-sum(dLBIG(X,theta[1],theta[2]))}
lik.TCR <- function(theta){-sum(dTwoCrack(X,theta[1],theta[2]))}
MLETCR<-nlminb(c(1,1), lik.TCR, lower =0.1, upper = Inf)$par
MLEBS<-nlminb(c(0.001, 0.1), lik.BS, lower = 0.0001, upper = Inf)$par
MLEIG<-nlminb(c(3,0.01), lik.IG, lower = 0.0001, upper = Inf)$par
MLELBIG<-nlminb(c(1,0.01), lik.LBIG, lower = 0.0001, upper = Inf)$par

#Example for find AIC of TCR, IG, LBIG, and BS distribution
library(twoCrack)
```


Table 1: A vibration of of the vertical roller mill data set in 60 minutes (um) and Day-Month-Year-Time (a.m.) where data is collected.

Date & Time	Vibration	Date & Time	Vibration	Date & Time	Vibration
10.02.2019 05:00	77.01	10.02.2019 05:20	93.72	10.02.2019 05:40	78.9
10.02.2019 05:01	77.01	10.02.2019 05:21	82.32	10.02.2019 05:41	74.88
10.02.2019 05:02	68.72	10.02.2019 05:22	71.93	10.02.2019 05:42	77.37
10.02.2019 05:03	71.38	10.02.2019 05:23	77.75	10.02.2019 05:43	85.67
10.02.2019 05:04	73.99	10.02.2019 05:24	92.91	10.02.2019 05:44	66.05
10.02.2019 05:05	66.87	10.02.2019 05:25	76.27	10.02.2019 05:45	66.06
10.02.2019 05:06	71.43	10.02.2019 05:26	91.95	10.02.2019 05:46	72.03
10.02.2019 05:07	81.3	10.02.2019 05:27	73.65	10.02.2019 05:47	74.17
10.02.2019 05:08	82.33	10.02.2019 05:28	71.42	10.02.2019 05:48	70.65
10.02.2019 05:09	76.78	10.02.2019 05:29	76.77	10.02.2019 05:49	75.99
10.02.2019 05:10	70.31	10.02.2019 05:30	66.5	10.02.2019 05:50	72.54
10.02.2019 05:11	74.88	10.02.2019 05:31	75.61	10.02.2019 05:51	85.45
10.02.2019 05:12	69.25	10.02.2019 05:32	71.14	10.02.2019 05:52	74.6
10.02.2019 05:13	66.57	10.02.2019 05:33	69.7	10.02.2019 05:53	76.37
10.02.2019 05:14	68.97	10.02.2019 05:34	72.48	10.02.2019 05:54	70.02
10.02.2019 05:15	67.45	10.02.2019 05:35	71.53	10.02.2019 05:55	77.66
10.02.2019 05:16	72.09	10.02.2019 05:36	74.37	10.02.2019 05:56	88.71
10.02.2019 05:17	72.09	10.02.2019 05:37	77.88	10.02.2019 05:57	86.86
10.02.2019 05:18	71.19	10.02.2019 05:38	69.67	10.02.2019 05:58	76.82
10.02.2019 05:19	63.63	10.02.2019 05:39	66.85	10.02.2019 05:59	68.56

Table 2: The MLE of the model parameters for the mill vibration, and AIC measure

Fitting Dist.	Estimate parameters		AIC
	λ	θ	
TCR	19.73067	0.1	7.013233
BS	0.001	0.1	255.989
IG	3	0.01	96.78106
LBIG	1	0.01	81.35225

```

LLTCR <- log(sum(dTwoCrack(X, MLETCR[1], MLETCR[2])))
AICTCR = -2*LLTCR+2*2
LLIG <- log(sum(dIG(X, MLEIG[1], MLEIG[2])))
AICIG = -2*LLIG+2*2
LLLBIG <- log(sum(dLBIG(X, MLELBIG[1], MLELBIG[2])))
AICLBIG = -2*LLLBIG+2*2
LLBS <- log(sum(dBS(X, MLEBS[1], MLEBS[2])))
AICBS = -2*LLBS+2*2

```

Regarding the results in Table 2, we observe that the two-parameter crack distribution has an AIC value less than another distribution and so the two-parameter crack distribution fits the data better than other distributions. Hence, we fit the two-parameter crack distribution to the mill vibration data set by maximum likelihood estimation for finding the likelihood that a vertical roller mill has a vibration greater than 85, R command as follows.

```

#Example for find probability when data follow TCR distribution
data1 <- read.csv("Vibration.csv")
library(twoCrack)
sample <- data1[,1]
V.dist <- dTwoCrack(sample, 19.73067, 0.10000)
V.df <- data.frame("Vibration" = sample, "Density" = V.dist)
library(ggplot2)
ggplot(V.df, aes(x = Vibration, y = Density)) + geom_point()
# likelihood of Vibration >= 85?
sum(V.df$Density[V.df$Vibration >= 85])*100

```

From Figure (4), we can find the percentage that a vertical roller mill has a vibration greater than 85 by calculating the sum of the likelihood of

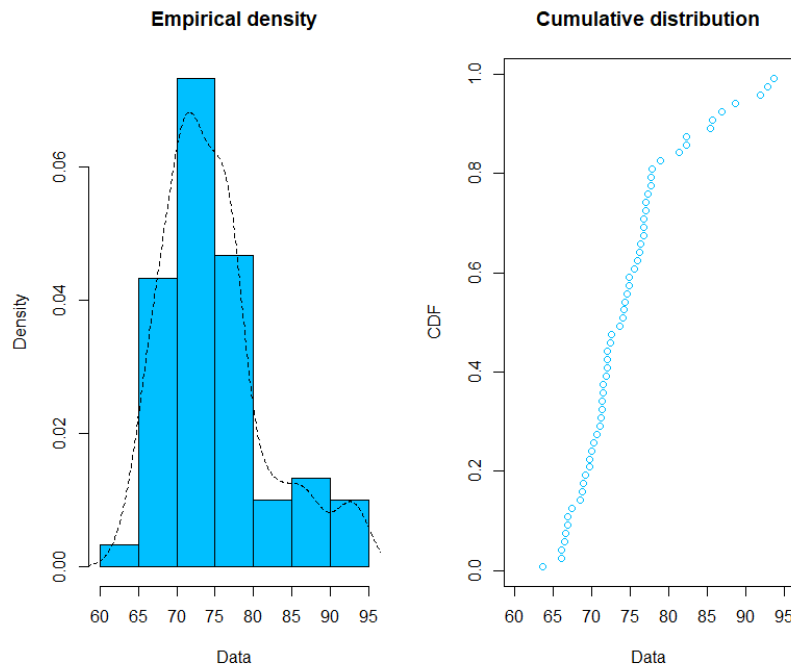


Figure 3: Histogram and distribution function plots of an empirical distribution

vibration greater than or equal to 85 and multiply by 100. The result from the R command is 2.144049%.

7 Conclusion

We have developed a new R package for the two-parameter crack distribution. The R package **twoCrack** allows to easily use the two-parameter crack distribution and sub-distribution. Our main objective, while developing this package, was to provide tools for helping R users to use the probability density function of the two-parameter crack distribution and sub-distribution such as fitting the two-parameter crack, Birnbaum-Saunders, inverse Gaussian and length-biased inverse Gaussian distribution to a real data set or determining the probabilities of events when a random sample from the two-parameter crack, and sub-distribution. The second objective was to provide R users a set of functions for generating random numbers that follow the two-parameter crack distribution and sub-distribution. We have shown an

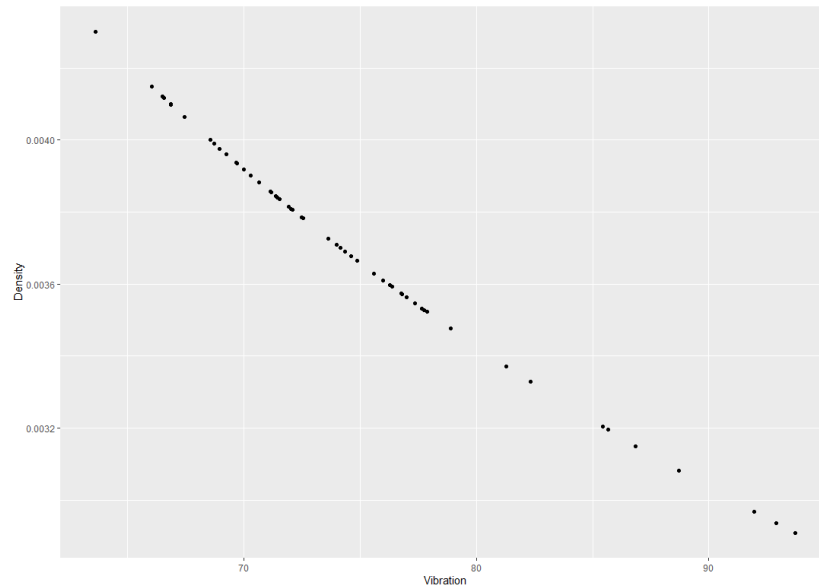


Figure 4: The probability density of the mill vibration distribution

example of applying this package with the vertical roller mill vibration data set. The results indicate that we can easily fit the two-parameter crack distribution and sub-distribution to the real data set by using R language.

Acknowledgments. The author would like to express her sincere thanks to the Siam Cement (Ta Luang) Co., Ltd. Khao Wong Plant for support data. Moreover, she expresses her appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper. Furthermore, this research was funded by King Mongkut's University of Technology North Bangkok. Contract no. KMUTNB-63-DRIVE-2

References

- [1] B. Jorgensen, V. Seshadri, G. A. Whitmore, On the mixture of the inverse Gaussian distribution with its complementary reciprocal, *Scand. J. Statist.*, **18**, (1991), 77–89.
- [2] P. Bowonrattanaset, K. Budsaba, Some properties of the three-parameter crack distribution, *Thailand Statistician*, **9**, no. 2, (2011), 195–203.
- [3] S. E. Ahmed, K. Budsaba, S. Lisawadi, A. Volodin, Parametric estimation for the Birnbaum-Saunders lifetime distribution based on new parametrization, *Thailand Statistician*, **6**, no. 2, (2008), 213–240.
- [4] P. Saengthong, W. Bodhisuwan, A new two-parameter crack distribution, *Journal of Applied Sciences*, **14**, no. 8, (2014), 758–766.
- [5] Z. W. Birnbaum, S. C. Saunders, A new family of lifetime distribution, *Journal of Applied Probability*, **6**, no. 2, (1969a), 319–327.
- [6] Z. W. Birnbaum, S. C. Saunders, Estimation for a family of life distributions with applications to fatigue, *Journal of Applied Probability*, **6**, no. 2, (1969b), 328–347.
- [7] R. S. Chhikara, J. L. Folks, The inverse Gaussian distribution as a lifetime model, *Technometrics*, **19**, no. 4, (1977), 461–468.
- [8] W. Phaphan , Asymptotic confidence ellipses for the re-parameterized inverse Gaussian distribution, *The Journal of Applied Science*, **19**, no. 1, (2020), 36–50.
- [9] R Core Team, R: A language and environment for statistical computing, The R project for statistical computing, 2021-03-16, <https://www.r-project.org/>
- [10] RStudio Team, RStudio: Integrated development for R, 2021-03-16, <https://rstudio.com/>
- [11] W. Phaphan, T. Pongsart, Estimating the parameters of a two-parameter crack distribution, **12**, no. 1, (2019), 29–36.
- [12] T. Ngamkham, On the crack random numbers generation procedure, *Lobachevskii Journal of Mathematics*, **40**, no. 8, (2019), 1204–1217.