

RESEARCH ARTICLE

Gendist: An R Package for Generated Probability Distribution Models

Shaiful Anuar Abu Bakar^{1*}, Saralees Nadarajah², Zahrul Azmir ABSL Kamarul Adzhar¹, Ibrahim Mohamed¹

¹ Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia, ² School of Mathematics, University of Manchester, Manchester, United Kingdom

* saab@um.edu.my



Abstract

In this paper, we introduce the R package gendist that computes the probability density function, the cumulative distribution function, the quantile function and generates random values for several generated probability distribution models including the mixture model, the composite model, the folded model, the skewed symmetric model and the arc tan model. These models are extensively used in the literature and the R functions provided here are flexible enough to accommodate various univariate distributions found in other R packages. We also show its applications in graphing, estimation, simulation and risk measurements.

OPEN ACCESS

Citation: Abu Bakar SA, Nadarajah S, ABSL Kamarul Adzhar ZA, Mohamed I (2016) Gendist: An R Package for Generated Probability Distribution Models. PLoS ONE 11(6): e0156537. doi:10.1371/journal.pone.0156537

Editor: Duccio Rocchini, Fondazione Edmund Mach, Research and Innovation Centre, ITALY

Received: September 18, 2015

Accepted: May 16, 2016

Published: June 7, 2016

Copyright: © 2016 Abu Bakar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from the stated R packages.

Funding: SAAB is funded by Fundamental Research Grant Scheme (FRGS) with grant number FP050-2014A.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Various probability distribution models have been proposed in the past and the number increases with time. Recently, in the area of actuarial loss modeling, several new models were found to provide good fits to the loss data. For instance, mixture of Erlang distribution has been proposed to model catastrophic loss data in the United States [1]. Mixture of exponential with peaks-over-threshold has been considered to fit Danish fire losses and medical claims data found in Society of Actuaries (SOA) Group Medical Large Claims Database [2]. Mixture of lognormal and inverse Gaussian distribution were used to model fire insurance portfolio in Serbia [3]. In the actuarial context, loss data are monetary losses claimed by insureds purchasing general insurance policies such as fire and catastrophic insurances. Many reasons in which the data is found to be crucial and recorded by insurance companies or insurance service agencies, among others, to model their future financial obligations. In addition to the above, various applications of mixture model can be found in the literature for other area of studies. To name a few: a logistic mixture distribution model has been applied on polychotomous item responses [4]; mixture of logistic has been proposed to fit long tail distributions in analyzing network performance [5]; Rayleigh mixture model has been studied for plaque characterization in intravascular ultrasound [6]; a finite mixture of two Weibull distributions has been suggested to model the diameter distributions of rotated-sigmoid, uneven-aged stands [7]; two-component mixture Weibull statistics has been used to

estimate wind speed distributions [8]; gamma mixture models have been proposed for target recognition [9]; Gaussian mixture model has been applied for human skin color and its applications in image and video databases [10].

Another important model receiving increasing applications in actuarial loss modeling is the composite model. Composite model is constructed by piecing two weighted distributions together. Such a model has been introduced initially with constant mixing weight [11]. It was then improved by allowing flexible mixing weight [12]. Several other related composite models have been proposed by various authors, see [13–15] and [16]. All these authors employed the well known Danish fire insurance data to measure their model performance. Composite models have also been applied to some simulated data belong to a particular class of distribution. Among others, a comparison study have been made on Weibull-Pareto and Lognormal-Pareto composite models [17]. Besides, some properties, inferences and numerical illustration using simulated data have been provided for composite exponential-Pareto models [18].

Another model that provides useful application in loss modeling is the folded model. It has been used to model the Norwegian fire claims data [19]. An extension to this study proposed three new folded models, namely, the folded generalized t distribution, folded Gumbel distribution and folded exponential power distribution [20]. An interesting feature of this model is the folding mechanism of a real value defined distribution into positive value distribution. Several existing folded models found in the literature are the folded normal distribution [21], folded t distribution [22], folded Cauchy distribution [23] and folded logistic distribution [24].

In addition to the above, skewed symmetric models also featured attractive properties with respect to loss data. The skewed normal and skewed t distributions have been studied in fitting insurance claims data [25]. Later, the same author applied the two models to asset returns of insurance companies [26]. Skewed t distribution is found to provide promising results for both data. A variety of other skew models have been considered, among them, skewed Cauchy [27], skewed Laplace [28], skewed logistic [29], skew reflected gamma, skew double Weibull and skew beta-prime [30] and several skew inverse reflected distributions [31].

More recently, the arc tan model has been introduced to model Norwegian fire claims data [32]. This new methodology has been proposed for an underlying Pareto distribution and found to provide good fit compared to several classical distributions. Its statistical properties also have been derived.

Beside heavily used in the actuarial field, most of the aforementioned models received vast applications in many other areas. Motivated by this, we compile several important models into an R package **gendist** for academics and public use. R is a free statistical computing and graphics software downloadable from <http://www.r-project.org>, see [33].

Distribution models provided in the R package **gendist** include the mixture, the composite, the folded, the skewed symmetric and the arc tan models. Computation functions of these models are given for probability density function (pdf), cumulative distribution function (cdf), quantile function (qf) and random generated values (rg).

The conventional R prefixes d , p , q and r define the pdf, cdf, qf and rg of an arbitrary distribution function. For instance `dexp`, `pexp`, `qexp` and `rexp` of the **stats** package [33] in R gives the pdf, cdf, qf and rg for an exponential distribution. The **gendist** package follows similar rule to define all the functions related to the generated probability distribution models.

In preparing the **gendist** package, we have thoroughly studied the Distribution CRAN Task View of R [34] which lists a number of available packages related to probability distributions. None of the packages on this page has so far provide tools to work with the folded and arc tan models discussed in this paper. Several packages are found for mixture models: **mixtool** [35] provides d and r functions for finite mixture models; **norlmix** [36] provides d , p and r functions for a specific underlying distribution, Gaussian; **GSM** [37] provides d , p and r functions

with underlying Gamma shape distribution; **AdMit** [38] provides d and r functions for mixture of student distribution.

Only the **CompLognormal** package [39] is available for composite models. However, the model is restricted to head of a lognormal distribution. In addition, the function is bound below at zero. For skewed symmetric model, several packages are available: **skewt** [40], **sn** [41], **gamlss.dist** [42] and **sgt** [43] provides functions d , p , q and r for variety of either skew normal or skew student distributions; **Newdistns** [44] provides functions related to skew symmetric G distribution. **SkewHyperbolic** [45] provides functions related to skew hyperbolic student distribution. All the distributions mentioned above plus others (including newly proposed distributions) can be easily managed by the functions developed in the proposed **gendist** package.

The paper is structured as follows. First, we discuss in detail all the above models along with examples on using related functions developed in the **gendist** package. Then, we provide descriptions on the program structure of the package and some discussion on appropriate usage with respect to the support of the models. We finally conclude the works carried out. Further details of the **gendist** package can be found at <http://CRAN.R-project.org/package=gendist>.

Generated Probability Distribution Models

Models presented here are generated with underlying parent distributions. These models are specified by a predefined rule or structure. In what follows, we describe in detail each generated model giving particular emphasis on related functions encompassed in the **gendist** package.

The mixture models

The mixture model was first considered with an underlying normal distribution to address the decomposition issue with respect to non-normal forehead to body length attributes in a population of female shore crabs, see [46]. In general, the pdf of the mixture model is given by

$$f(x) = \sum_{i=1}^n w_i g_i(x) \quad (1)$$

where $0 \leq w_i \leq 1$ for $i = 1, 2, \dots, n$ are the mixing weights such that the sum of all w_i equals to one. Note that, there are n components involved in Eq (1).

The proposed mixture model in this paper is restricted to a two component distribution. Hence, the pdf can be represented as follows:

$$f(x) = w g_1(x) + (1 - w) g_2(x) \quad (2)$$

We choose mixing weight of the form $w = \frac{1}{1+\phi}$ so that $\phi > 0$. $g_1(x)$ and $g_2(x)$ represent the densities of arbitrary parent distributions. Note, it is not necessary that the two arbitrary distributions in Eq (2) are identical. However, both must be defined on the same range dimension.

Finding cdf of the mixture model is straightforward. By direct integration, the cdf for the two component mixture model in Eq (2) can be written as

$$F(x) = \frac{1}{1+\phi} G_1(x) + \frac{\phi}{1+\phi} G_2(x) \quad (3)$$

with $G_i(x)$ for $i = 1, 2$ are arbitrary cdfs of the parent distribution.

Explicit general expression for qf of the mixture model is not available. However, it can be obtained numerically by solving the root, $Q(u)$, of the following equation

$$wG_1(Q(u)) + (1 - w)G_2(Q(u)) - u = 0 \quad (4)$$

Finally, random generated numbers for the two component mixture model can be obtained as

$$x_i = Q(u_i) \quad (5)$$

where u_i for $i = 1, 2, \dots, n$ are random numbers from a uniform $[0, 1]$ distribution.

In what follows, we show an application of the cdf function of the mixture model, `pmixt` to measure the closeness of two different set of measurements using P-P plot. Data used for this purpose is the US indemnity loss data scaled by 1000 found in the **copula** package [47]. We also use the **actuar** package [48] to illustrate several heavy tailed distributions as theoretical measurements. Assuming pre-installation of both packages the data ([S1 Data](#)) can be obtained and sorted as follows:

```
R> library(actuar)
R> library(copula)
R> data(loss)
R> x <- loss$loss/1000
R> n <- length(x)
R> xx <- seq(1, n) / (n+1)
R> x <- sort(x)
```

For each distribution involved, values of 0.2, 0.7 and 20 are chosen for ϕ (the mixing weight component), shape and scale parameters, respectively. These values are only for illustration purposes. The P-P plots are produced by the following command:

```
R> par(mfrow=c(2, 2))
R> arg1 <- c(shape=0.7, scale=20)
R> arg2 <- c(shape=0.7, scale=20)
R> plot(xx, pmixt(x, phi=0.2, spec1="weibull", arg1,
+ spec2="llogis", arg2), xlim=c(0,1), ylim=c(0,1), col=2,
+ ylab="Expected", xlab="Observed",
+ main="Mixture of Weibull-Loglogistic")
R> abline(0,1)
R> plot(xx, pmixt(x, phi=0.2, spec1="weibull", arg1,
+ spec2="pareto", arg2), xlim=c(0,1), ylim=c(0,1), col=2,
+ ylab="Expected", xlab="Observed",
+ main="Mixture of Weibull-Pareto")
R> abline(0,1)
R> plot(xx, pmixt(x, phi=0.2, spec1="weibull", arg1,
+ spec2="invpareto", arg2), xlim=c(0,1), ylim=c(0,1), col=2,
+ ylab="Expected", xlab="Observed",
+ main="Mixture of Weibull-Inverse Pareto")
R> abline(0,1)
R> plot(xx, pmixt(x, phi=0.2, spec1="weibull", arg1,
+ spec2="paralogis", arg2), xlim=c(0,1), ylim=c(0,1), col=2,
+ ylab="Expected", xlab="Observed",
+ main="Mixture of Weibull-Paralogistic")
R> abline(0,1)
```

Analysis of Application on Mixture Model. P-P plots are used to compare the degree of agreement between two sets of measurements based on their cdf. [Fig 1](#) shows the P-P plot of the empirical data against the theoretical values of the mixture of Weibull-Loglogistic,

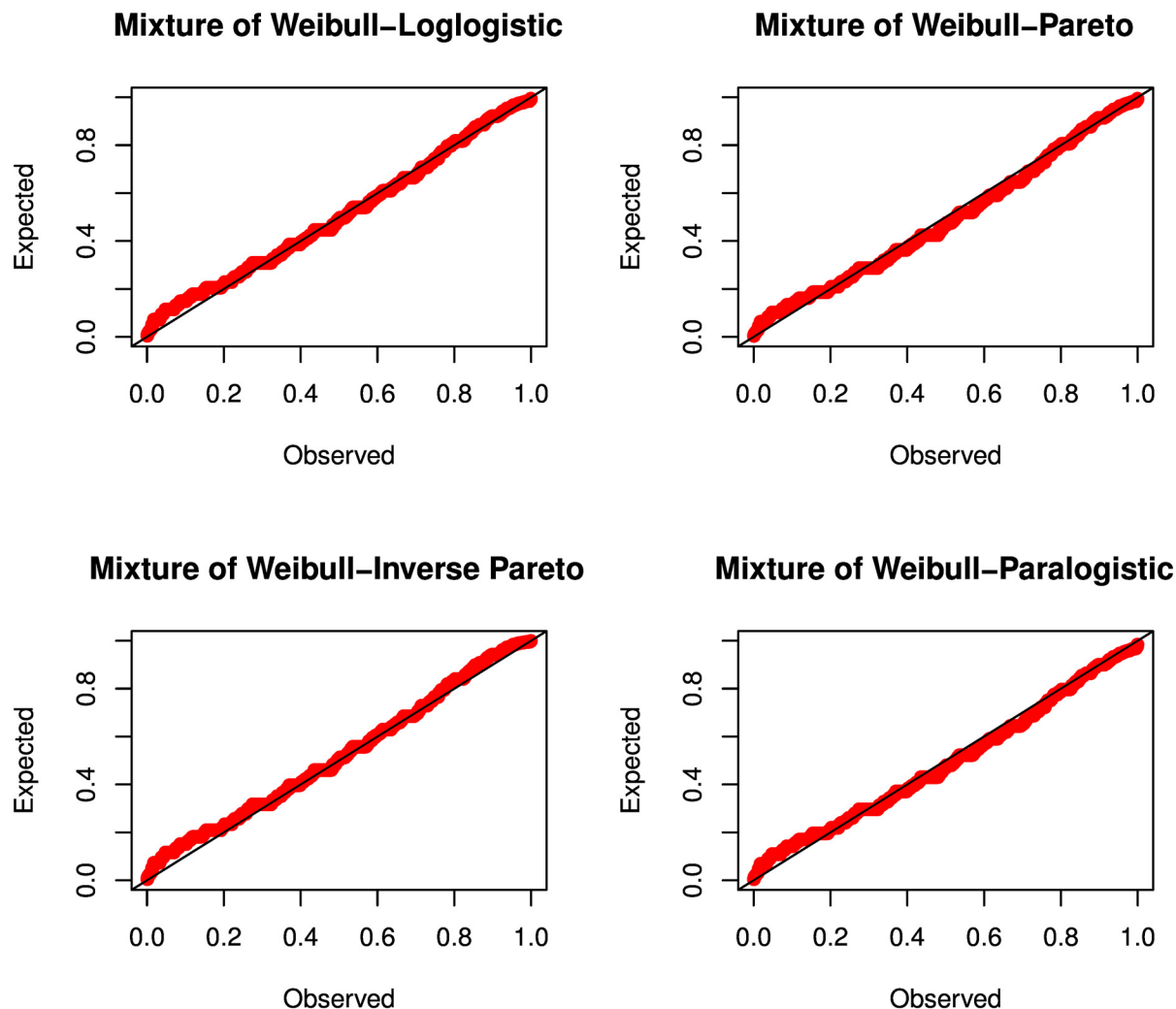


Fig 1. Probability-probability plots (P-P plots) for the US indemnity loss data for several mixture models. Expected and observed denote the theoretical and empirical data, respectively. Four models are shown including the mixture of Weibull-Logistic, mixture of Weibull-Pareto, mixture of Weibull-Inverse Pareto and mixture of Weibull-Paralogistic models. Values of 0.2, 0.7 and 20 are chosen for the mixing weight component, shape and scale parameters. All the models show a good fit to the empirical data.

doi:10.1371/journal.pone.0156537.g001

Weibull-Pareto, Weibull-Inverse Pareto and Weibull-Paralogistic models. A small deviation of the plots to the 45° line indicate the closeness of the empirical and theoretical probability and vice versa. It is apparent from the P-P plots that the fits are good for all the theoretical models proposed.

The composite models

In general, the pdf of the composite model is given by

$$f(x) = \begin{cases} wf_1^*(x), & \text{if } 0 < x \leq \theta \\ (1-w)f_2^*(x), & \text{if } \theta < x < \infty \end{cases} \quad (6)$$

whereby w is the mixing weight, θ denote the threshold and $f_i^*(x)$ for $i = 1, 2$ are the truncated

pdfs of the parent distribution defined by

$$f_1^*(x) = \frac{f_1(x)}{F_1(x)} \quad (7)$$

and

$$f_2^*(x) = \frac{f_2(x)}{1 - F_2(x)} \quad (8)$$

respectively. Eq (6) can be made continuous and smooth by applying the continuity and differentiability conditions and thus provide a closed form for w , see [12]. Furthermore, it has been suggested that w take the form of $w = \frac{1}{1+\phi}$ for $\phi > 0$ for convenience [13]. We find this useful to ease program writing in R for the composite model. A more comprehensive approach to implement Eq (6) which we adopt in this paper specifies the component of mixing weight, ϕ , and the threshold, θ in term of other parameters of the model, see [15]. Further to this, we do not restrict the composite model to bound below at zero and thus allow parent distribution defined on real line as well as that defined for positive values to be considered for $f_1^*(x)$. All other functions including the cdf, qf, and rg for the composite model are developed in a similar manner.

Hence, the cdf, qf and n random generated numbers provided in **gendist** package can be written as

$$F(x) = \begin{cases} \frac{1}{1+\phi} \frac{F_1(x)}{F_1(\theta)} & \text{if } x \leq \theta \\ \frac{1}{1+\phi} \left(1 + \phi \frac{F_2(x) - F_2(\theta)}{1 - F_2(\theta)} \right) & \text{if } x > \theta \end{cases} \quad (9)$$

$$Q(u) = \begin{cases} Q_1^*(u(1+\phi)F_1(\theta)) & \text{if } u \leq \frac{1}{1+\phi} \\ Q_2 \left(F_2(\theta) + (1 - F_2(\theta)) \left(\frac{u(1+\phi) - 1}{\phi} \right) \right) & \text{if } u > \frac{1}{1+\phi} \end{cases} \quad (10)$$

and

$$x_i = Q(u_i) \quad (11)$$

where u_i with $i = 1, 2, \dots, n$ are n uniform $[0, 1]$ random numbers.

As a result, the number of parameters of the composite models equal the total number of parameters of the two underlying parent distributions selected. For instance, if truncated exponential distribution with scale parameter is chosen for $f_1^*(x)$ and Weibull distribution with scale and shape parameters are chosen for $f_2^*(x)$, the Exponential-Weibull composite model will consist of three parameters.

We now show the implementation of the pdf function `dcomposite` of the composite model. All other functions related to the composite model follow similarly. Let $f_1(x)$ and $f_2(x)$ denote the pdfs of Weibull and Gamma distributions, respectively. From Eq (6) (with $w = \frac{1}{1+\phi}$),

the composite Weibull-Gamma model can be written as

$$f(x) = \begin{cases} \frac{1}{1+\phi} \frac{\beta e^{-\left(\frac{x}{\lambda}\right)^\beta} \left(\frac{x}{\lambda}\right)^\beta}{x - x e^{-\left(\frac{x}{\lambda}\right)^\beta}}, & \text{if } 0 < x \leq \theta \\ \frac{\phi}{1+\phi} \frac{\sigma^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\sigma}}}{\Gamma\left(\alpha, \frac{\theta}{\sigma}\right)}, & \text{if } \theta < x < \infty \end{cases} \quad (12)$$

where $\Gamma(a, z)$ is the incomplete gamma function defined by $\int_z^\infty t^{a-1} e^{-t} dt$. Suppose, $\alpha = \beta = 1$ and $\sigma = \lambda = 2$, then, the pdf values of the composite model for some values of x are computed as follows:

```
R> dcomposite(1:3, spec1="weibull", arg1=list(shape=1, scale=2),
              spec2="gamma", arg2=list(shape=1, scale=2))
```

Output:

```
0.3032653 0.1839397 0.1115651
```

Density plots of this model can be done using curve function as below.

```
R> curve(dcomposite(x, spec1="weibull", arg1=list(shape=2.5, scale=1),
+          spec2="gamma", arg2=list(shape=1, scale=1), initial=c(0.1, 1)),
+       xlim=c(0, 10), xlab="x", ylab="f(x)")
R> curve(dcomposite(x, spec1="weibull", arg1=list(shape=1.5, scale=2),
+          spec2="gamma", arg2=list(shape=2, scale=1), initial=c(0.1, 2)),
+       add=TRUE, col=2)
R> curve(dcomposite(x, spec1="weibull", arg1=list(shape=2.3, scale=2),
+          spec2="gamma", arg2=list(shape=2, scale=0.5),
+          initial=c(0.1, 10)), add=TRUE, col=3)
R> curve(dcomposite(x, spec1="weibull", arg1=list(shape=0.1, scale=2),
+          spec2="gamma", arg2=list(shape=0.5, scale=1),
+          initial=c(0.1, 10)), add=TRUE, col=4)
```

Analysis of Application on Composite Model. Fig 2 shows the pdf plots of the composite Weibull-Gamma models with varying parameter values. The density shapes depend on the values of the scale and shape parameters. For the composite Weibull-Gamma model, all the four densities exhibit unimodal distribution, that is, they have a single mode. It is also observed that choosing lower shape parameters for both parent distributions will result in a thin and steep distribution while having a larger value result in a thick and gradual slope. Note that the observation made here applies for composite Weibull-Gamma. Other composite models may have different result depending on the features of the pair of the parent distributions chosen.

The folded models

Existence of folded models can be traced back to 1960s when methods were described to estimate mean and variance of normal distribution based on its folded form and an application was shown to real camber data [21]. The folded model is obtained from a transformation of random variable by taking its absolute value. Suppose that Y is a real valued random variable with cdf $G(\cdot)$ and $X = |Y|$ a positive random variable, then the cdf for X is

$$F(x) = G(x) - G(-x) \quad x > 0 \quad (13)$$

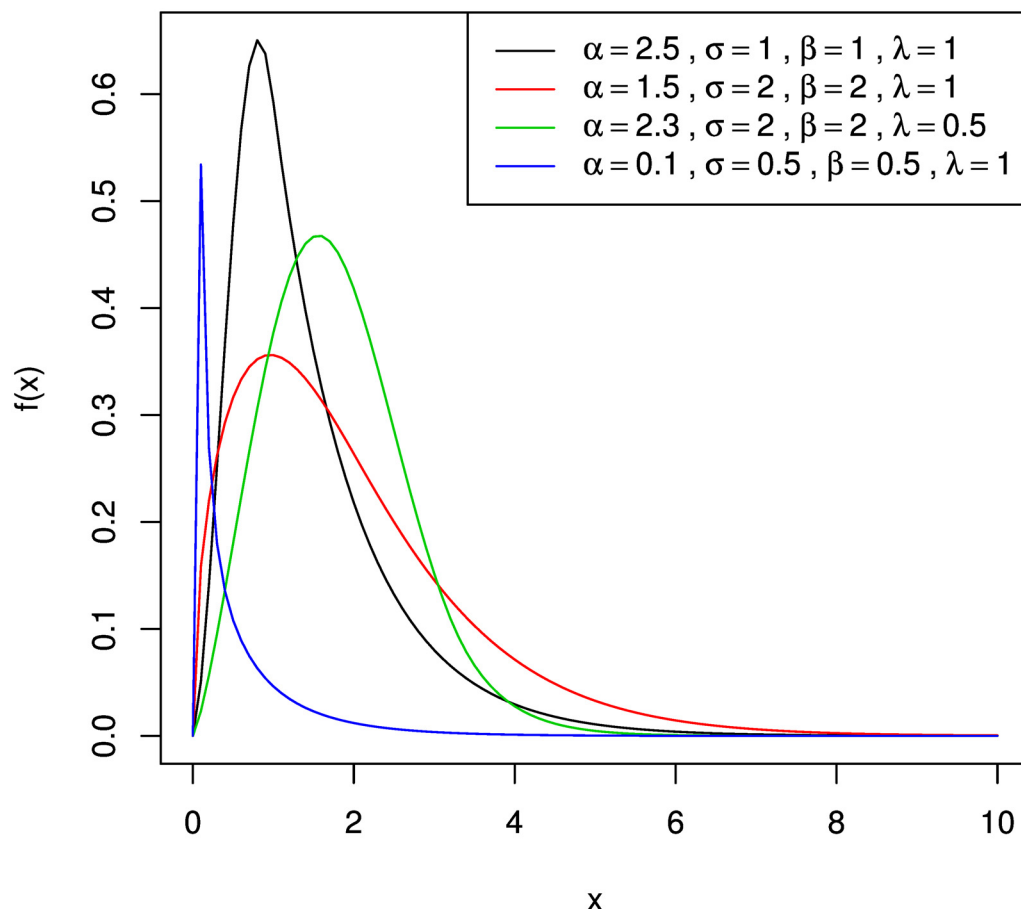


Fig 2. Probability density function curves for composite Weibull-Gamma model with varying parameters. Four composite Weibull-Gamma models with varying parameters are shown. The shapes depend on the value of parameters chosen.

doi:10.1371/journal.pone.0156537.g002

We can then write its pdf as

$$f(x) = g(x) + g(-x) \quad x > 0 \quad (14)$$

In this case the parent distribution has a support of real values and the resulted folded distribution has $x > 0$.

It is obvious from Eq (13) that the quantile function may not have an explicit form. However, this may not be the case if the underlying distribution is symmetric around 0. The quantile function for a folded model of this case can be obtained as follows:

$$Q(u) = F^{-1}(u) = G^{-1}\left(\frac{1+u}{2}\right) \quad (15)$$

where $G^{-1}(\cdot)$ is the inverse function of the underlying symmetric distribution. For an underlying non-symmetric distribution (also applies to symmetric case), we can solve this numerically using computer programs by finding root of the cdf function, that is, $Q(u)$ is the solution of the following equation

$$G(Q(u)) - G(-Q(u)) - u = 0 \quad (16)$$

Next, rg is obtained as follows

$$x_i = Q(u_i) \quad (17)$$

where u_i with $i = 1, 2, \dots, n$ are n uniform $[0, 1]$ random numbers. Note that $g(\cdot)$ and $G(\cdot)$ correspond to the pdf and cdf of the parent distribution of the folded model. We implement Eqs (13), (14), (16) and (17) in the **gendist** package for cdf, pdf, qf and rg of the folded model.

In what follows, we show applications of two risk measures by implementation of qf for the folded model, `qfolded`. It involves computation of Value-at-Risk (VaR) and Conditional Tail Expectation (CTE). VaR is related to the quantile function of a random variable with a specified probability, say, α . Often it describes the risk that a loss random variable exceeds a certain amount although it is also used outside the financial area. Let $G^{-1}(\alpha)$ denote the inverse function of $G(x)$ for a random variable Y , then VaR at a given level of confidence, $1 - \alpha$, is given by

$$VaR(\alpha) = G^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (18)$$

For a range of α values, this is illustrated for folded normal and folded t distributions in Fig 3. Some arbitrary values are chosen for the parameters of both models. The following commands describe the plotting of VaR with respect to α :

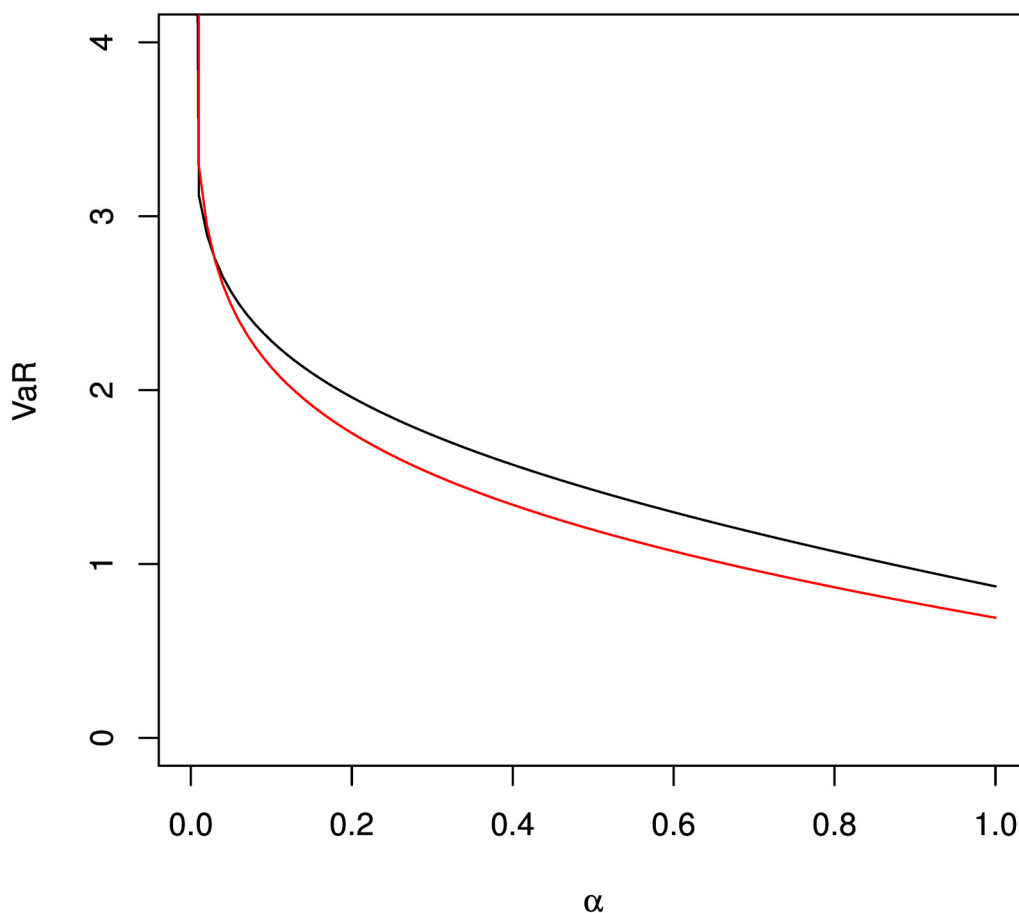


Fig 3. Value-at-Risk for folded normal (black) and folded t (red) distributions.

doi:10.1371/journal.pone.0156537.g003

```
R> curve(qfolded(1-x/2, spec = "norm", arg = c(mean = 0.8, sd = 0.9),
+ interval = c(0,100)), xlim = c(0,1), ylim = c(0,4),
+ xlab = expression(alpha), ylab = "VaR")
R> curve(qfolded(1-x/2, spec = "t", arg = c(df = 15), interval = c(0,100)),
+ add = TRUE, col = 2)
```

The CTE measures the expected value of risk beyond VaR at a given probability level, α . Mathematically,

$$CTE(\alpha) = \frac{1}{\alpha} \int_{\alpha}^1 G_Y^{-1}\left(1 - \frac{\alpha}{2}\right) d\alpha \quad (19)$$

For a range of α values, this is illustrated for folded normal and folded t distributions in Fig 4. The following commands describe the plotting of CTE with respect to α :

```
R> curve((1/(x))*integrate(function(x) qfolded(1-x/2, spec = "norm",
+ arg = c(mean = 0.8, sd = 0.9), interval = c(0,100)), x, 1)$value,
+ xlim = c(0,1), ylim = c(0,20), xlab = expression(alpha), ylab = "CTE")
R> curve((1/(x))*integrate(function(x) qfolded(1-x/2, spec = "t",
+ arg = c(df = 15), interval = c(0,100)), x, 1)$value, add = TRUE, col = 2)
```

Analysis of Application on Folded Model. Fig 3 shows that for both distributions VaR appears to be a decreasing function of the probability level α . This is a natural behavior of risk,

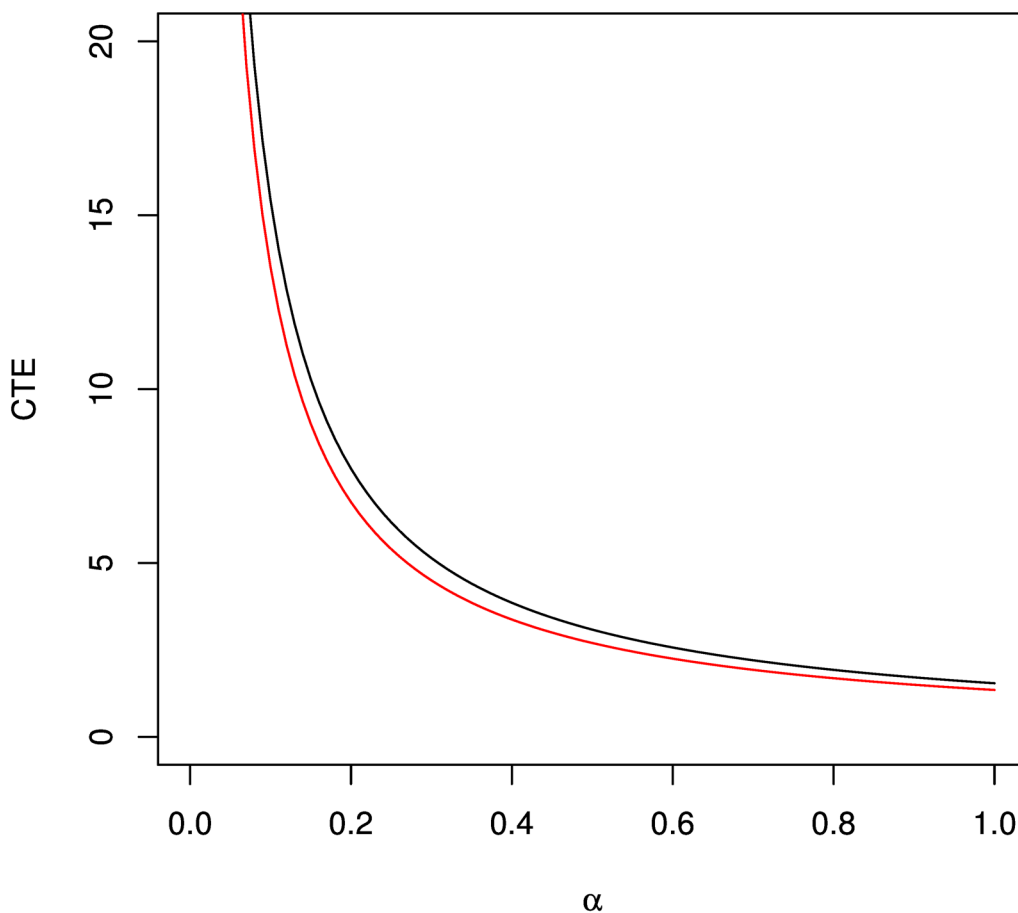


Fig 4. Conditional Tail Expectation for folded normal (black) and folded t (red) distributions.

doi:10.1371/journal.pone.0156537.g004

that is, higher VaR is found for higher level of confidence. In practice, α are commonly chosen as 1% or 5%. In this example, folded t distribution consistently show a lower VaR value than the normal distribution. However, the actual results will depend on the parameter values for each model. Similar to VaR, CTE for both distributions are decreasing function of the confidence level, α . This is depicted in Fig 4.

The skewed symmetric models

Azzalini proposed a new class of distributions with an underlying normal distribution [49]. In its original form, the pdf is given by

$$f(x) = 2g(x)G(\lambda x), \quad -\infty < x < \infty \quad (20)$$

where $g(\cdot)$ and $G(\cdot)$ are the pdf and cdf of the normal distribution and λ is a shape parameter. The distribution reduces to normal distribution when $\lambda = 0$.

Further study leads to a general form of the skew symmetric distribution [50]. The new skewed symmetric distribution has the following pdf

$$f(x) = 2h(x)G(x), \quad -\infty < x < \infty \quad (21)$$

where $h(x)$ is a pdf symmetric at 0 and $G(x)$ is a Lebesgue measurable function that satisfies $0 \leq G(x) \leq 1$ and $G(x) + G(-x) = 1$ for $z \in \mathbb{R}$. In this new form, the parent distributions $h(x)$ and $G(x)$ may be of different type satisfying the conditions above. In the **gendist** package, we adopt Eq (21) for computer implementation of the skewed symmetric model. Several functions for $h(x)$ and $G(x)$ include the normal, logistic, students t and Cauchy distributions.

General form for the cdf, qf and rg functions of the skewed symmetric models are not available. Thus, implementation of these functions in **gendist** are done via numerical methods. In particular, `integrate` and `uniroot` functions are used. The cdf of skewed symmetric model is found by

$$F(x) = \int_{-\infty}^x 2h(y)G(y)dy, \quad -\infty < x < \infty \quad (22)$$

Solving $Q(u)$ of the following equation leads to its qf

$$\int_{-\infty}^{Q(u)} 2h(y)G(y)dy - u = 0 \quad (23)$$

and rg is obtained as follows

$$x_i = Q(u_i) \quad (24)$$

where u_i with $i = 1, 2, \dots, n$ are n uniform $[0, 1]$ random numbers.

Some skewed symmetric models may have an explicit form. For instance, consider a skewed Cauchy-Cauchy distribution. Its pdf, cdf and qf are given by

$$f(x) = \frac{\lambda \left(2 \tan^{-1} \left(\frac{x}{\lambda} \right) + \pi \right)}{\pi^2 (\lambda^2 + x^2)} \quad (25)$$

$$F(x) = \frac{\left(2 \tan^{-1} \left(\frac{x}{\lambda} \right) + \pi \right)^2}{4\pi^2} \quad (26)$$

and

$$Q(u) = \lambda \tan\left(\frac{1}{2}(2\pi\sqrt{u} - \pi)\right) \quad (27)$$

respectively.

In the following illustration, we check goodness-of-fit of two theoretical distributions, the skew Logistic-Logistic and the skew Normal-Normal distributions, to the length of rivers scaled by 1000. The data consist of 141 observations related to major North American rivers. The data ([S2 Data](#)) can be obtained as follows:

```
R> data(rivers)
R> x<- rivers/1000
```

Initial step involves parameter estimation using the maximum likelihood method.

```
R> nloglik<- function(p, spec1, arg1, spec2, arg2){
+ tt <- 1.0e20
+ if(all(p>0)){
+   tt <- sum(log(dskew(x, spec1, arg1, spec2, arg2)))
+ }
+ return(tt)
+ }
R> par<-nlm(function(p){ nloglik(p, spec1="logis", arg1=list(scale=p[1]),
+   spec2="logis", arg2=list(scale=p[2])), p=c(1,1)) $estimate
R> par2<- nlm(function(p){ nloglik(p, spec1="norm", arg1=list(sd=p[1]),
+   spec2="norm", arg2=list(sd=p[2]))}, p=c(1,1)) $estimate
```

Finally, the Q-Q plots are produced as presented in [Fig 5](#) with the following commands:

```
R> qqplot(x, qskew(u, spec1="logis", arg1=list(scale=par[1]),
+   spec2="logis", arg2=list(scale=par[2]), interval=c(-10,10)),
+   xlim=c(0,4), ylim=c(0,4), ylab="Theoretical", xlab="Empirical",
+   main="Skew Logistic-Logistic distribution")
R> abline(0,1)
R> qqplot(x, qskew(u, spec1="norm", arg1=list(sd=par2[1]), spec2="norm",
+   arg2=list(sd=par2[2]), interval=c(-10,10)), xlim=c(0,4),
+   ylim=c(0,4), ylab="Theoretical", xlab="Empirical",
+   main="Skew Normal-Normal distribution")
R> abline(0,1)
```

Analysis of Application on Skewed Symmetric Model. Parameter estimation is performed using the maximum likelihood method which is known to have the following properties: sufficient, invariance, consistent, efficient and asymptotically normal. Estimated parameters are then used to find the theoretical values of the quantile function. These values are matched to the empirical values and plotted to the Q-Q plot. Goodness-of-fit is then measured by relative distance of the plots to the 45° line. Closer theoretical versus empirical plots to the line indicate a good fit and vice versa. Majority of the plots in [Fig 5](#) are close to the line except for some values at the extreme ends. These are the five longest river in the North America. Between the two models, the skew Logistic-Logistic distribution gives a lower negative log likelihood value, that is, a value of 57.5903. Thus it has a better goodness-of-fit to the river data as compared to the skew Normal-Normal distribution.

The arc tan models

Arc tan model has been proposed to model a specific Norwegian insurance loss data [[32](#)]. Consider parent distribution with pdf and cdf of $g(x)$ and $G(x)$, respectively. Then, the pdf, cdf, qf

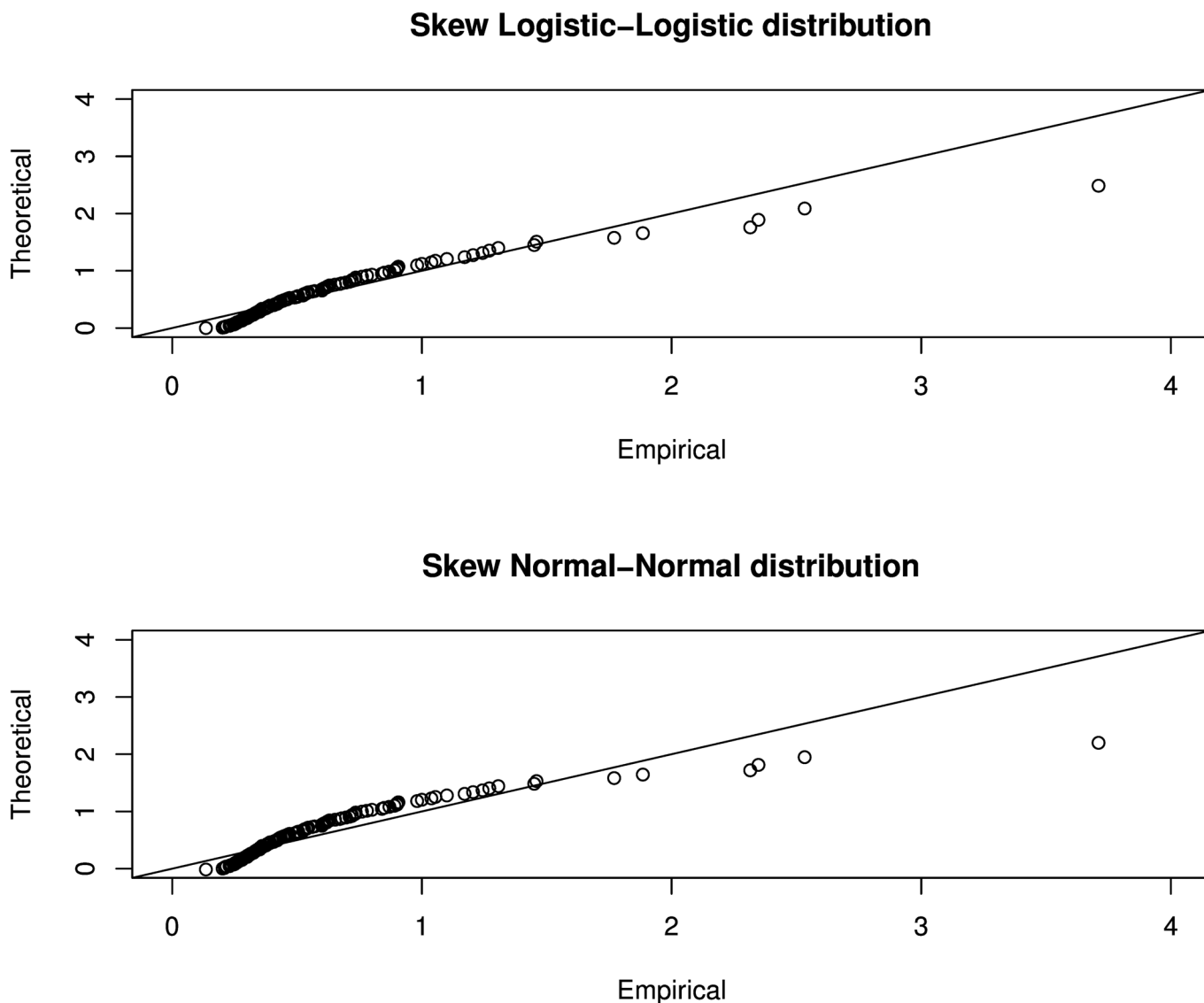


Fig 5. Q-Q plots of rivers data for the fitted skew logistic-logistic and skew normal-normal distributions. Empirical values are plotted against two theoretical distributions, that is, the skew Logistic-Logistic distribution and the skew Normal-Normal distribution. Both show a close plot to the 45° line except for several extreme values.

doi:10.1371/journal.pone.0156537.g005

and rg of the arc tan model with support $[a, b]$ are given by

$$f(x) = \frac{1}{\arctan(\alpha)} \frac{\alpha g(x)}{1 + (\alpha(1 - G(x)))^2} \quad (28)$$

$$F(x) = 1 - \frac{\arctan(\alpha(1 - G(x)))}{\arctan(\alpha)} \quad (29)$$

$$Q(u) = G^{-1} \left(1 - \frac{1}{\alpha} \tan((1 - u) \arctan(\alpha)) \right) \quad (30)$$

and

$$x_i = Q(u_i) \quad (31)$$

respectively, where u_i with $i = 1, 2, \dots, n$ are n uniform $[0, 1]$ random numbers, \arctan denote the inverse function of trigonometric tangent and $G^{-1}(\cdot)$ is the inverse of $G(\cdot)$. Note that the model consist of additional parameter $\alpha > 0$ and the support is $a \leq x \leq b$ where a and b take the support of parent distribution.

Consider a Weibull distribution with pdf

$$g(x) = \frac{\beta e^{-\left(\frac{x}{\lambda}\right)^\beta} \left(\frac{x}{\lambda}\right)^{\beta-1}}{\lambda}, \quad x > 0 \quad (32)$$

In what follows, we illustrate simulation of the arc tan distribution with parent distribution [Eq \(32\)](#). For simplicity, we let $\beta = 2$ and $\lambda = 0.5$. The empirical biases and mean square errors of α for the Weibull arc tan distribution can be obtained as follows:

1. Ten thousand sample sizes are generated by inversion of [Eq \(29\)](#). Each sample size is n . The variates of the Weibull arc tan distribution can be written as

$$X = \log \left(\frac{1}{2} \sqrt{-\log \left(\frac{\tan((1-u) \arctan(\alpha))}{\alpha} \right)} \right) \quad (33)$$

where $U \sim U(0,1)$ is a variates of the uniform distribution.

2. Estimate the parameter, $\hat{\alpha}_i$ for $i = 1, 2, \dots, 10000$.
3. Bias and mean squared error are obtained using

$$bias_x(n) = \frac{1}{10000} \sum_{i=1}^{10000} (\hat{\alpha}_i - \alpha) \quad (34)$$

and

$$MSE_x(n) = \frac{1}{10000} \sum_{i=1}^{10000} (\hat{\alpha}_i - \alpha)^2 \quad (35)$$

The above process is repeated for $n = 10, 20, \dots, 1000$ with $\alpha = 1.5$. The following commands implement the process:

```
R> nsim <- 10000
R> nsiz <- 100
R> est1 <- matrix(0, nsiz, nsim)
R> mml <- matrix(0, nsiz, nsim)
R> bias <- rep(0, nsiz)
R> mse <- rep(0, nsiz)
R> ss <- rep(0, nsiz)
R> alpha <- 1.5
R> for(j in 1:nsiz){
+   for(i in 1:nsim){
+     x <- rarctan(j*10, alpha = alpha, spec = "weibull",
+               arg = c(shape = 2, scale = 0.5))
+     nlogl <- function(p, alpha, spec, arg){
```

```
+ tt <- 1.0e20
+ if(all(p>0)){
+   tt <- -sum(log(darctan(x, alpha, spec, arg)))
+ }
+ return(tt)
+ }
+ est <- nlm(function(p) nlogl(p, alpha=p, spec="weibull",
+   arg=list(shape=2, scale=0.5)), p=1, hessian=T)
+ est1[j, i] <- est$estimate[1]
+ mm1[j, 1] <- solve(est$hessian)[1, 1]
+ bias[j] <- mean(est1[j,] - alpha)
+ mse[j] <- mean((est1[j,] - alpha) ^ 2)
+ ss[j] <- mean(mm1[j,] ^ (1/2))
+ }
```

Analysis of Application on Arc Tan Model. Fig 6 shows the variation of biases and mean squared error with respect to n . The dashed line corresponds to zero biases or theoretical mean squared errors, whichever applicable. Several observations can be made:

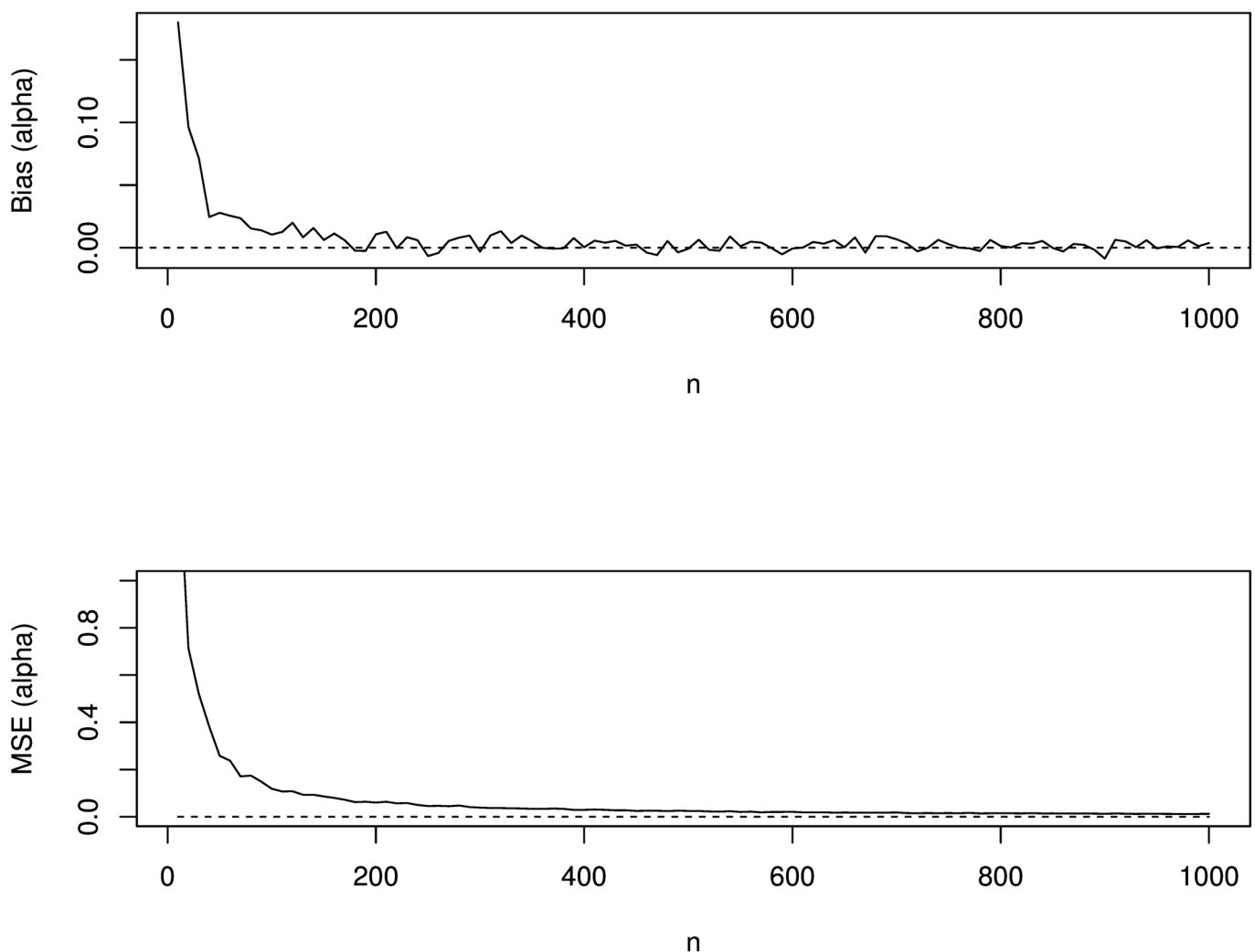


Fig 6. Empirical and theoretical biases and mean square errors of $\hat{\alpha}$ versus $n = 10, 20, \dots, 1000$. Biases and mean squared errors for Weibull arc tan models are presented for simulated data. The biases are generally positive and decreases to zero as n approaches infinity. Similarly, mean squared errors decreases to zero as n approaches infinity.

doi:10.1371/journal.pone.0156537.g006

1. the biases are generally positive
2. as expected, the magnitude of bias always decreases to zero as $n \rightarrow \infty$
3. the biases are small for $\hat{\alpha}$
4. as expected, the mean squared errors always decrease to zero as $n \rightarrow \infty$
5. the theoretical mean squared errors are always larger than empirical ones

Results are presented only for $\alpha = 1.5$ due to reason of space. However, similar results can be obtained for other choices of α .

An advanced example featuring empirical data related to air quality is given in the Appendix (S1 Appendix). The subroutines provided there emphasize on calling functions from the **gendist** package and thus avoid creating separate object running similar computation. The underlying concept pertaining to this idea make use of `do.call` which is discussed next.

Results and Discussion

In this paper, we specify the generated probability distribution models by its parent distribution, that is, the underlying function which generates new models. Codes for writing functions in **gendist** package use a general construct of parent distribution of the form

```
do.call(paste(d, spec, sep = ""), c(list(z), arg)),
do.call(paste(d, spec1, sep = ""), c(list(z), arg1))
```

or

```
do.call(paste(d, spec2, sep = ""), c(list(z), arg2)).
```

As such, `spec = "Weibull"`, for instance, corresponds to Weibull parent distribution and `arg = list(shape, scale)` list its parameters. Similarly, for two component distributions `spec1` and `spec2` specify the parent distributions with corresponding parameters `arg1` and `arg2`.

The parent distribution may assume any value on real line or positive real numbers. The suitability of the parent distributions for the models must be checked by the user. No warnings are given for choosing inappropriate distributions. Table 1 describes a proper selection for the parent distribution corresponding to each model with respect to its support.

Functions to compute pdf, cdf, qf and rg for all five models produced in **gendist** are summarised in Table 2. As mentioned earlier, the input argument `spec`, `spec1` and `spec2` specify the parent distribution and their corresponding parameters `arg`, `arg1` and `arg2`. The distribution can be one that is implemented in R **base** package, contributed R packages or one written by a user. In any case, the parent functions must be defined with prefix `d`, `p`, `q` and `r` attached to `spec`, `spec1` and `spec2`.

Table 1. Support for models in gendist package.

Models	Support of parent distribution	Support of generated models
Mixture	Real numbers and positive real numbers	Real numbers and positive real numbers
Composite	Real numbers and positive real numbers	Real numbers and positive real numbers
Folded	Real numbers	Positive real numbers
Skewed	Real numbers	Real numbers
Arc tan	Real numbers and positive real numbers	Real numbers and positive real numbers

doi:10.1371/journal.pone.0156537.t001

Table 2. Calling sequences for mixture, composite, folded, skewed symmetric and arc tan models.

Models	Functions	Calling sequence
Mixture	$f(x)$ in Eq (2)	<code>dmixt(x, phi, spec1, arg1, spec2, arg2, log = FALSE)</code>
	$F(x)$ in Eq (3)	<code>pmixt(q, phi, spec1, arg1, spec2, arg2, lower.tail = TRUE, log.p = FALSE)</code>
	$Q(x)$ in Eq (4)	<code>qmixt(p, phi, spec1, arg1, spec2, arg2, interval = c(0,100), lower.tail = TRUE, log.p = FALSE)</code>
	x_i in Eq (5)	<code>rmixt(n, phi, spec1, arg1, spec2, arg2, interval = c(0,100))</code>
Composite	$f(x)$ in Eq (6)	<code>dcomposite(x, spec1, arg1, spec2, arg2, initial = 1, log = FALSE)</code>
	$F(x)$ in Eq (9)	<code>pcomposite(q, spec1, arg1, spec2, arg2, initial = 1, lower.tail = TRUE, log.p = FALSE)</code>
	$Q(x)$ in Eq (10)	<code>qcomposite(p, spec1, arg1, spec2, arg2, initial = 1, lower.tail = TRUE, log.p = FALSE)</code>
	x_i in Eq (11)	<code>rcomposite(n, spec1, arg1, spec2, arg2, initial = 1)</code>
Folded	$f(x)$ in Eq (14)	<code>dfolded(x, spec, arg, log = FALSE)</code>
	$F(x)$ in Eq (13)	<code>pfolded(q, spec, arg, lower.tail = TRUE, log.p = FALSE)</code>
	$Q(x)$ in in Eq (16)	<code>qfolded(p, spec, arg, interval = c(0,100), lower.tail = TRUE, log.p = FALSE)</code>
	x_i in Eq (17)	<code>rfolded(n, spec, arg, interval = c(0,100))</code>
Skew symmetric	$f(x)$ in Eq (21)	<code>dskew(x, spec1, arg1, spec2, arg2, log = FALSE)</code>
	$F(x)$ in Eq (22)	<code>pskew(q, spec1, arg1, spec2, arg2, lower.tail = TRUE, log.p = FALSE)</code>
	$Q(x)$ in Eq (22)	<code>qskew(p, spec1, arg1, spec2, arg2, interval = c(1,10), lower.tail = TRUE, log.p = FALSE)</code>
	x_i in Eq (24)	<code>rskew(n, spec1, arg1, spec2, arg2, interval = c(1,10))</code>
Arc tan	$f(x)$ in Eq (28)	<code>darctan(x, alpha, spec, arg, log = FALSE)</code>
	$F(x)$ in Eq (29)	<code>parctan(q, alpha, spec, arg, lower.tail = TRUE, log.p = FALSE)</code>
	$Q(x)$ in Eq (30)	<code>qarctan(p, alpha, spec, arg, lower.tail = TRUE, log.p = FALSE)</code>
	x_i in Eq (31)	<code>rarctan(n, alpha, spec, arg)</code>

doi:10.1371/journal.pone.0156537.t002

It is important to note that some models may not have explicit general form for its probability functions. Whenever this is the case, numerical methods are used. In **gendist**, we utilise `integrate`, `nlm` and `uniroot` functions. Some of these functions, in particular, `uniroot` require interval values to search for roots and they are specified by `interval`. `nlm` require a starting value specified by `initial`.

Conclusions

In this paper, we discuss five models provided in the R package **gendist**, namely, the mixture model, the composite model, the folded model, the skewed symmetric model and the arc tan model. The functions related to these models are written in R environment and are freely downloadable from <http://www.r-project.org>, see [33]. Users can easily create any specific model by specifying the parent distributions along with their parameters. Thus, uncountable number of distributions can be produced from the tools in the **gendist** package. Another advantage is that the users have the option to write their own function to serve as the parent distributions.

Next, we also show at least five important applications of the tools given in the **gendist** package. Usefulness of these tools in graphing are shown for Q-Q plot, P-P plot, producing pdf curves as well as to show output for some risk measures. Simulation study is provided for a specific Weibull arc tan model and produced favorable result with respect to biases and mean squared error of the parameter estimated. Both measures decrease to zero when the number of observations approaches infinity.

Nevertheless, we only provide pdf, cdf, qf and rg for five common models that recently receive applications in actuarial loss modeling. Some of the models have also been considered in other areas while others such as arc tan model are potentially useful in areas involving parametric framework. Many other available models can be designed in similar general form as the **gendist** package and thus avoid the problem of offering separate R packages for distributional models of the same class into production of a single package that incorporates all. Therefore, we encourage other scientists to furnish the package with additional models in their respective field or send suggestion for further improvement. The **gendist** package provides a great flexibility to work with many distributions and hopes to assist users in their respective fields.

Supporting Information

S1 Script. Script to replicate examples. The script to replicate examples in this paper is provided in (.R) form.

(R)

S1 Appendix. Appendix.

(TXT)

S1 Data. Loss data.

(TXT)

S2 Data. Rivers Data.

(TXT)

S3 Data. Air Quality Data.

(TXT)

Acknowledgments

The authors greatly thank the Academic Editor and the anonymous reviewer for carefully reading the paper and their valuable comments. This research was supported by Fundamental Research Grant Scheme (FRGS FP050-2014A), Ministry of Higher Education Malaysia.

Author Contributions

Conceived and designed the experiments: SAAB. Performed the experiments: SAAB. Analyzed the data: SN ZA IM. Contributed reagents/materials/analysis tools: SN ZA IM. Wrote the paper: SAAB SN ZA IM.

References

1. Lee SC, Lin XS. Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*. 2010; 14(1):107–130. doi: [10.1080/10920277.2010.10597580](https://doi.org/10.1080/10920277.2010.10597580)
2. Lee D, Li WK, Wong TST. Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics*. 2012; 51(3):538–550.
3. Kočović J, Rajić VĆ, Jovanović M. Estimating a tail of the mixture of log-normal and inverse Gaussian distribution. *Scandinavian Actuarial Journal*. 2015; 2015(1):49–58. doi: [10.1080/03461238.2013.775665](https://doi.org/10.1080/03461238.2013.775665)
4. Rost J. A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology*. 1991; 44(1):75–92. doi: [10.1111/j.2044-8317.1991.tb00951.x](https://doi.org/10.1111/j.2044-8317.1991.tb00951.x)
5. Feldmann A, Whitt W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. In: *INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and*

- Communications Societies. Driving the Information Revolution., Proceedings IEEE. vol. 3. IEEE; 1997. p. 1096–1104.
6. Seabra J, Ciompi F, Pujol O, Mauri J, Radeva P, Sanches J. Rayleigh mixture model for plaque characterization in intravascular ultrasound. *Biomedical Engineering, IEEE Transactions on*. 2011; 58(5):1314–1324. doi: [10.1109/TBME.2011.2106498](https://doi.org/10.1109/TBME.2011.2106498)
7. Zhang L, Gove JH, Liu C, Leak WB. A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. *Canadian Journal of Forest Research*. 2001; 31(9):1654–1659. doi: [10.1139/x01-086](https://doi.org/10.1139/x01-086)
8. Carta J, Ramirez P. Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. *Renewable Energy*. 2007; 32(3):518–531. doi: [10.1016/j.renene.2006.05.005](https://doi.org/10.1016/j.renene.2006.05.005)
9. Webb AR. Gamma mixture models for target recognition. *Pattern Recognition*. 2000; 33(12):2045–2054. doi: [10.1016/S0031-3203\(99\)00195-8](https://doi.org/10.1016/S0031-3203(99)00195-8)
10. Yang MH, Ahuja N. Gaussian mixture model for human skin color and its applications in image and video databases. In: *Electronic Imaging'99*. International Society for Optics and Photonics; 1998. p. 458–466.
11. Cooray K, Ananda MM. Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*. 2005; 2005(5):321–334. doi: [10.1080/03461230510009763](https://doi.org/10.1080/03461230510009763)
12. Scollnik DP. On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*. 2007; 2007(1):20–33. doi: [10.1080/03461230601110447](https://doi.org/10.1080/03461230601110447)
13. Nadarajah S, Bakar SAA. New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*. 2014; 2014(2):180–187. doi: [10.1080/03461238.2012.695748](https://doi.org/10.1080/03461238.2012.695748)
14. Scollnik DP, Sun C. Modeling with Weibull-Pareto Models. *North American Actuarial Journal*. 2012; 16(2):260–272. doi: [10.1080/10920277.2012.10590640](https://doi.org/10.1080/10920277.2012.10590640)
15. Bakar SAA, Hamzah N, Maghsoudi M, Nadarajah S. Modeling loss data using composite models. *Insurance: Mathematics and Economics*. 2015; 61:146–154.
16. Calderin-Ojeda E, Kwok CF. Modeling claims data with composite Stoppa models. *Scandinavian Actuarial Journal*. 2015;(ahead-of-print):1–20. doi: [10.1080/03461238.2015.1034763](https://doi.org/10.1080/03461238.2015.1034763)
17. Preda V, Ciunara R. On composite models: Weibull-Pareto and Lognormal-Pareto. A comparative study. *Romanian Journal of Economic Forecasting*. 2006; 3(2).
18. Teodorescu S. Some composite Exponential-Pareto models for actuarial prediction. *Romanian Journal of Economic Forecasting*. 2009; 12(4):82–100.
19. Brazauskas V, Kleefeld A. Folded and log-folded-t distributions as models for insurance loss data. *Scandinavian Actuarial Journal*. 2011; 2011(1):59–74. doi: [10.1080/03461230903424199](https://doi.org/10.1080/03461230903424199)
20. Nadarajah S. and Bakar S. A. A.. New folded models for the log-transformed Norwegian fire claim data. *Communications in Statistics-Theory and Methods* 44, 20 (2015): 4408–4440.
21. Leone F, Nelson L, Nottingham R. The folded normal distribution. *Technometrics*. 1961; 3(4):543–550. doi: [10.1080/00401706.1961.10489974](https://doi.org/10.1080/00401706.1961.10489974)
22. Psarakis S, Panaretos J. The folded t distribution. *Communications in Statistics-Theory and Methods*. 1990; 19(7):2717–2734. doi: [10.1080/03610929008830342](https://doi.org/10.1080/03610929008830342)
23. Johnson NL, Kotz S, Balakrishnan N. Continuous univariate distributions, vol. 1–2. New York: John Wiley & Sons; 1994.
24. Cooray K, Gunasekera S, Ananda MM. The folded logistic distribution. *Communications in Statistics—Theory and Methods*. 2006; 35(3):385–393. doi: [10.1080/03610920500476234](https://doi.org/10.1080/03610920500476234)
25. Eling M. Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*. 2012; 51(2):239–248.
26. Eling M. Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*. 2014; 59:45–56.
27. Arnold BC, Beaver RJ. The skew-Cauchy distribution. *Statistics & probability letters*. 2000; 49(3):285–290. doi: [10.1016/S0167-7152\(00\)00059-6](https://doi.org/10.1016/S0167-7152(00)00059-6)
28. Aryal G, Nadarajah S. On the skew Laplace distribution. *Journal of information and optimization sciences*. 2005; 26(1):205–217. doi: [10.1080/02522667.2005.10699644](https://doi.org/10.1080/02522667.2005.10699644)
29. Wahed A, Ali MM. The skew-logistic distribution. *J Statist Res*. 2001; 35:71–80.
30. Ali MM, Woo J. Skew-symmetric reflected distributions. *Soochow Journal of Mathematics*. 2006; 32(2):233.
31. Ali MM, Woo J, Nadarajah S. Some skew symmetric inverse reflected distributions. *Brazilian Journal of Probability and Statistics*. 2010; 24(1):1–23. doi: [10.1214/08-BJPS100](https://doi.org/10.1214/08-BJPS100)

32. Gómez-Déniz E, Calderín-Ojeda E. Modelling Insurance Data With The Pareto Arctan Distribution. *ASTIN Bulletin*. 2015; 45(03): 639–660. doi: [10.1017/asb.2015.9](https://doi.org/10.1017/asb.2015.9)
33. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2014. Available from: <http://www.R-project.org/>.
34. Dutang C. CRAN Task View: Probability Distributions; 2014. <https://cran.r-project.org/web/views/Distributions.html>.
35. Benaglia T, Chauveau D, Hunter D, Young D. mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*. 2009; 32(6): 1–29. doi: [10.18637/jss.v032.i06](https://doi.org/10.18637/jss.v032.i06)
36. Mächler M. nor1mix: Normal (1-d) Mixture Models (S3 Classes and Methods); 2015. R package version 1.2.1. Available from: <http://CRAN.R-project.org/package=nor1mix>.
37. Venturini S. GSM: Gamma Shape Mixture; 2015. R package version 1.3.2. Available from: <http://CRAN.R-project.org/package=GSM>.
38. Ardia D, Hoogerheide LF, Van Dijk HK. Adaptive mixture of student-t distributions as a flexible candidate distribution for efficient simulation: The r package admit. *Journal of Statistical Software*. 2009; 29(3):1–32. doi: [10.18637/jss.v029.i03](https://doi.org/10.18637/jss.v029.i03)
39. Nadarajah S, Bakar S. CompLognormal: An R Package for Composite Lognormal Distributions. *R Journal*. 2013; 5(2): 98–104.
40. King R, with contributions from Emily Anderson. skewt: The Skewed Student-t Distribution; 2012. R package version 0.1. Available from: <http://CRAN.R-project.org/package=skewt>.
41. Azzalini A. The R package sn: The skew-normal and skew-t distributions (version 1.2-3). Università di Padova, Italia; 2015. Available from: <http://azzalini.stat.unipd.it/SN>.
42. Stasinopoulos M, with contributions from Calliope Akantziliotou BR, Heller G, Ospina R, Motpan N, McElduff F, et al. gamlss.dist: Distributions to be Used for GAMLSS Modelling; 2015. R package version 4.3-5. Available from: <http://CRAN.R-project.org/package=gamlss.dist>.
43. Davis C. sgt: Skewed Generalized T Distribution; 2015. R package version 1.1. Available from: <http://CRAN.R-project.org/package=sgt>.
44. Nadarajah S, Rocha R. Newdistns: Computes Pdf, Cdf, Quantile and Random Numbers, Measures of Inference for 19 General Families of Distributions; 2015. R package version 2.0. Available from: <http://CRAN.R-project.org/package=Newdistns>.
45. Scott D, Grimson F. SkewHyperbolic: The Skew Hyperbolic Student t-Distribution; 2015. R package version 0.3–2. Available from: <http://CRAN.R-project.org/package=SkewHyperbolic>.
46. Pearson K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A*. 1894;p. 71–110. doi: [10.1098/rsta.1894.0003](https://doi.org/10.1098/rsta.1894.0003)
47. Yan J. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*. 2007; 21(4):1–21. doi: [10.18637/jss.v021.i04](https://doi.org/10.18637/jss.v021.i04)
48. Dutang C, Goulet V, Pigeon M. actuar: An R Package for Actuarial Science. *Journal of Statistical Software*. 2008; 25(7):38. Available from: <http://www.jstatsoft.org/v25/i07>.
49. Azzalini A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*. 1985;p. 171–178.
50. Huang WJ, Chen YH. Generalized skew-Cauchy distribution. *Statistics & probability letters*. 2007; 77(11):1137–1147. doi: [10.1016/j.spl.2007.02.006](https://doi.org/10.1016/j.spl.2007.02.006)