On New Generalized Logistic Distributions and Applications

Barreto FHS, Mota JMA and Rathie PN*

Department of Statistics and Applied Mathematics, Federal University of Ceara Fortaleza, Brazil

Research Article

Received date: 06/03/2017

Accepted date: 18/04/2017

Published date: 07/04/2017

***For Correspondence**

Rathie PN, Department of Statistics and Applied Mathematics, Federal University of Ceara Fortaleza, Brazil.

E-mail: pushpanrathie@yahoo.com

Keywords: Rathie-Swamee distribution, Azzalini method, multimodal data set analysis, Akaike criterion information, Bayesian criterion information, maximum likelihood method, Kolmogorov-Smirnov test

ABSTRACT

In 2006, Rathie and Swamee had proposed a generalization of the logistic distribution which is more flexible and multimodal. This work presents an addition of a new parameter to increase the flexibilization of the distribution as well as an asymmetric distribution using the Azzalini method, adding another parameter of asymmetry. Five data sets (Human Body Fat Index, HIV, Precipitation, pH Concentration, Relative Humidity) are analysed by applying the new distributions. The estimation of the parameters of the new distributions and mixture of the normals was accomplished by the automaximum likelihood method. Due to complex mathematical resources required to calculate the estimates of the new distributions, we use interactive numerical methods such as L-BFGS-B, BFGS, SANN etc. using an adaptive barrier algorithm added to enforce the constraint and an adapted function that searches for global maximum of a very complex non-linear objective function to initial values of the algorithm of estimation. All computational work was implemented in software R. In most cases, we use the Hartigan's test to reject unimodality. Using the Kolmogorov-Smirnov test at significance level of 5% and applying various criteria, such as Mean Square Error, Mean Absolute Deviation and Maximum Deviation, to indicate the best fit. The classical and general method for multimodal adjustment is a mixture of distributions, in particular, the mixture of the normal distributions because the normal distribution presents good mathematical properties. In the case of mixture of the normals, we use EM algorithm to calculate the estimates. We also use Akaike Information Criterion and Bayesian Information Criterion as selection criteria to highlight the best distribution, in both cases, comparing them with the mixture of normal distributions to illustrate the applicability of the results derived in this paper.

INTRODUCTION

There are several classical models, such as normal, exponential, binomial, Poisson, logistic etc. to analyze different data sets. As there is not a single unified model, we have to construct new models suitable for the data sets under consideration. The logistic model is very useful in many areas in statistics and physics. This article is divided as follows: Section 2 deals with symmetric generalized logistic distribution whereas in Section 3 the skew form is studied. Section 4 presents applications to analyze five real data sets using the results of earlier sections and comparing them with the mixture of two normal distributions where possible. The article ends with a short conclusion and a list of references. Rathie et al.[1] defined a multimodal symmetric distribution function G(x) for a random variable X~RS (a, b, p) as

$$G(x) = \frac{1}{1+e^{-x(a+b|x|^{p})}}, x \in R, p > -1, a, b \geq 0 \quad (1)$$

With a and b not zeros simultaneously. For b=0 or when p=0, (1) is written as a logistic distribution

$$F(x)=\frac{1}{1+e^{-cx}} \quad 2$$

Where c=a or c=a + b. The density function corresponding to (1) is

$$g(x)=\frac{[a+b(p+1)\,|\,x\,|^{p}]e^{-x(a+b|x|^{p})}}{[1+e^{-x(a+b|x|^{p})2}]} \quad 3$$

GENERALIZED SYMMETRIC LOGISTIC DISTRIBUTION

A symmetric distribution can be generated by using the method proposed by Jones in 2004 [2]. Let U ~Beta (α,α), and X=G-1(U), where G(x) is a distribution function of g(x). Then, the distribution function H(x) of X is given as

$$H(x)=\frac{1}{B(\alpha,\alpha)}\int_{0}^{G(x)} u^{\alpha-1}(1-u)^{1-\alpha}\,du \quad 4$$

Differentiating H(x) yields the corresponding density function as

$$h(x)=\frac{g(x)[G(x)]^{\alpha-1}[1-G(x)]^{\alpha-1}}{B(\alpha,\alpha)} \quad 5$$

Using (1) and (3) in (5), the generalized symmetric logistic density function for X ~ RSG (a, b, p, α) is given by

$$h(x)=\frac{[a+b(p+1)\,|\,x\,|^{p}]e^{-\alpha x(a+b|x|^{p})}}{B(\alpha,\alpha)\{1+e^{-x(a+b|x|^{p})}\}^{2\alpha}},\ x\in\mathbb{R},\ a,b\geq\sim 0,\ \alpha>0,\ p>-1 \quad 6$$

Where both a and b not zeros simultaneously and B (., .) is the beta function. For α=1, reduce to (3). We may introduce the location parameter μ in the model (6). There is no need to introduce the scale parameter, otherwise the density function will become non-identifiable. The density function (6) takes the following form on introducing the location parameter μ ϵ R:

$$h(x)=\frac{[a+b(p+1)\,|\,x-\mu\,|^{p}]e^{-\alpha(x-\mu)(a+b|x-\mu|^{p})}]}{B(\alpha,\alpha)\{1+e^{-(x-\mu)(a+b|x-\mu|^{p})}\}^{2\alpha}} \quad 7$$

The **Figures 1 to 4** show graphs for (6) and (7) respectively for various values of the parameters μ, a, b, p and α.

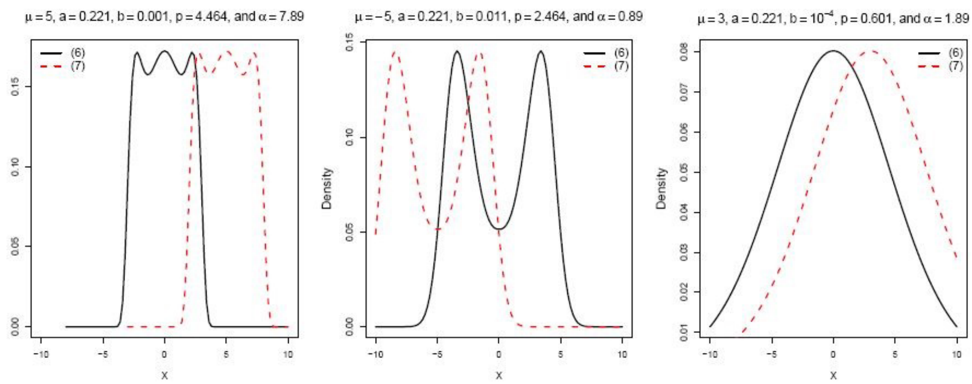


Figure 1. Graphs of (6) and (7) for Fixed a.

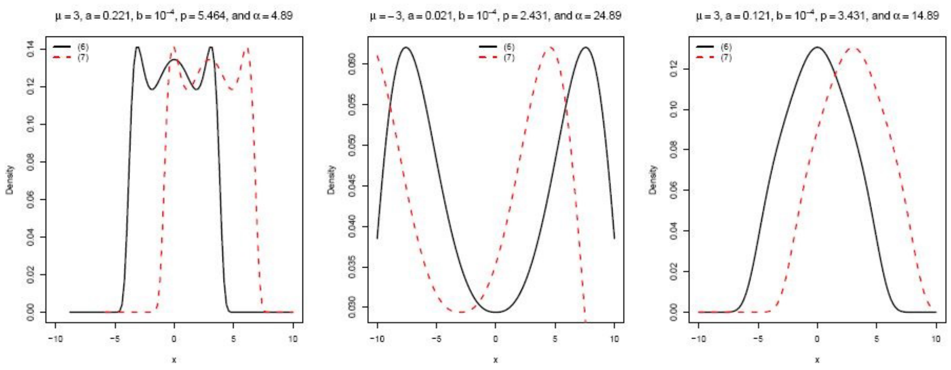


Figure 2. Graphs of (6) and (7) for fixed b.

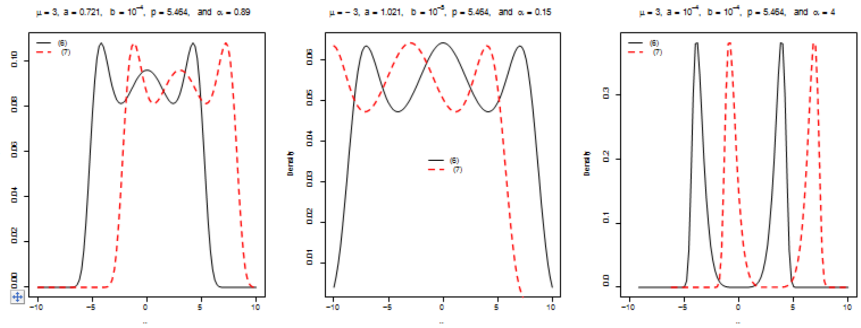


Figure 3. Graphs to (6) and (7) for fixed p.

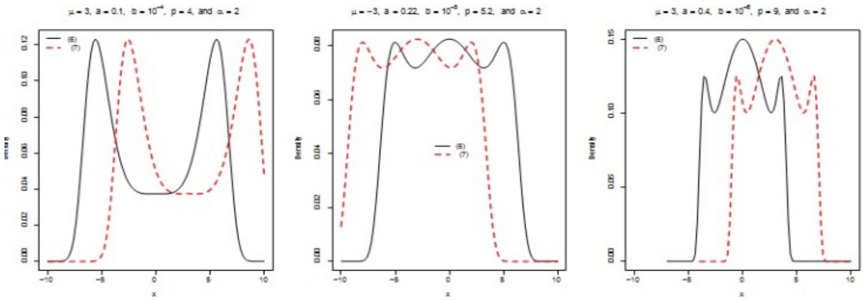


Figure 4. Graphs of (6) and (7) for fixed α.

Distribution function

In this subsection, we prove that the distribution function corresponding to (6) is given by

$$H(x)=\begin{cases}\frac{1}{2}+\frac{1}{\Gamma^2(\alpha)}[G^{1,2}_{2,2}(^{1-2\alpha,1-\alpha}_{0,-\alpha}|1)-e^{-\alpha x(a+bx^p)}G^{1,2}_{2,2}(^{1-2\alpha,1-\alpha}_{0,-\alpha}|e^{-x(a+bx^p)})], if x>0\\ \frac{1}{2}+\frac{1}{\Gamma^2(\alpha)}[G^{1,2}_{2,2}(^{1-2\alpha,1-\alpha}_{0,-\alpha}|1)-e^{-\alpha x(a+bx^p)}G^{1,2}_{2,2}(^{1-2\alpha,1-\alpha}_{0,-\alpha}|e^{-x(a+bx^p)})], if x<0\end{cases} \quad 8$$

Proof. For x > 0, we have

$$H(x)=\frac{1}{2}+\int_0^x\frac{[a+b(p+1)y^p]e^{-\alpha y(a+by^p)}}{B(\alpha,\alpha)\{1+e^{-y(a+by^p)}\}^{2\alpha}}dy \quad 9$$

Substuting $z=e^{-y(a+by^p)}$, *we get*

$$H(x)=\frac{1}{2}+\int_{e^{[-x(a+bx^p)]}}^1\frac{z^{\alpha-1}}{B(\alpha,\alpha)(1+z)^{2\alpha}}dz$$

$$G^{1,1}_{1,1}(^{1-2\alpha}_{0}|z)=\frac{\Gamma(2\alpha)}{(1+z)^{2\alpha}}$$

we have

$$\begin{aligned}H(x)&=\frac{1}{2}+\int_{e^{-x(a+bx^p)}}^1\frac{z^{\alpha-1}}{B(\alpha,\alpha)\Gamma(2\alpha)}G^{1,1}_{1,1}\left(\begin{matrix}1-2\alpha\\0\end{matrix}\middle|z\right)dz\\&=\frac{1}{2}+\frac{(2\pi i)^{-1}}{B(\alpha,\alpha)\Gamma(2\alpha)}\int_L\Gamma(s)+\Gamma(2\alpha+s)\int_{e^{-x(a+bx^p)}}^1 z^{\alpha+s-1}dzds\\&=\frac{1}{2}+\frac{(2\pi i)^{-1}}{\Gamma^2(\alpha)}\int_L\Gamma(s)\Gamma(2\alpha+s)\left[\frac{1}{\alpha+s}-\frac{e-(\alpha+s)x(a+bx^p)}{\alpha+s}\right]ds\\&=\frac{1}{2}+\frac{(2\pi i)^{-1}}{\Gamma^2(\alpha)}\int_L\frac{\Gamma(s)\Gamma(2\alpha+s)\Gamma(\alpha+s)}{\Gamma(\alpha+s+1)}\{1-e^{-(\alpha+s)x(a+bx^p)}\}ds\\&=\frac{1}{2}+\frac{1}{\Gamma^2(\alpha)}\left[\frac{1}{2\pi i}\int_L\frac{\Gamma(s)\Gamma(2\alpha+s)\Gamma(\alpha+s)}{\Gamma(\alpha+s+1)}1^s ds\right]-\frac{e^{-\alpha x(a+bx^p)}}{\Gamma^2(\alpha)}\times\\&\times\int_L\frac{\Gamma(s)\Gamma(2\alpha+s)\Gamma(\alpha+s)}{\Gamma(\alpha+s+1)}\left\{e^{-x(a+bx^p)}\right\}^s ds\\&=\frac{1}{2}+\frac{1}{\Gamma^2(\alpha)}\left[G^{1,2}_{2,2}\left(\begin{matrix}1-2\alpha,1-\alpha\\0,-\alpha\end{matrix}\middle|1\right)-e^{-\alpha x(a+bx^p)}G^{1,2}_{2,2}\left(\begin{matrix}1-2\alpha,1-\alpha\\0,-\alpha\end{matrix}\middle|-e^{-\alpha x(a+bx^p)}\right)\right]\end{aligned} \quad 10$$

By symmetry, we easily write the result for x < 0.

Moments

In this subsection, we obtain the n-th moments about the origin. By definition,

when n is an even integer

$$\mathbb{E}[X^n] = \int_{-\infty}^{+\infty} x^n h(x) dx$$

$$= \{2\int_{0}^{+\infty} x^n h(x)\, dx$$

Let $I_{n,\alpha} = \int_0^{\infty} x^n h(x)\, dx$ 11

Then, by expanding the denominator by binomial theorem, we have

$$\mathbb{E}[X^n] = \frac{2}{B(\alpha,\alpha)} \sum_{r=0}^{+\infty}\sum_{s=0}^{+\infty} \frac{(-1)^r (2\alpha)_r (-(\alpha+r) b^s)}{r!s![(\alpha+r)a]^{n+(p+1)s+1}} \times [a\,\Gamma(n+(p+1)s+1) + \frac{b(p+1)\Gamma(p+n+(p+1)s+1)}{[(\alpha+r)a]^p}] \quad 12$$

when n is an even integer.

The variance of X ~ RSG (a, b, p, α) is given by

$$Var[X] = \frac{2}{B(\alpha,\alpha)} \sum_{r=0}^{+\infty}\sum_{s=0}^{+\infty} \frac{(-1)^r (2\alpha)_r (-(\alpha+r) b^s)}{r!s![(\alpha+r)a]^{3+(p+1)s}} \times [a\,\Gamma(3+(p+1)s+1) + \frac{b(p+1)\Gamma(p+3+(p+1)s+1)}{[(\alpha+r)a]^p}] \quad 13$$

GENERALIZED SKEW LOGISTIC DISTRIBUTION

In Azzalini density [3]

$$s(x) = 2\, v(x)\, V[w(x)], x \in R \quad 14$$

With w(x)=kx; k 2 R, take v(x) as the density function of X ~ RSG(a; b; p;) and V (x) as the distribution function of X ~ RS(a; b; p). Then, the density function of generalized skew logistic model X ~ RSGA (a; b; p;α; k) is given by

$$s_1(x) = \frac{2[a+b(p+1)\,|\,x\,|^p] e^{-\alpha x(a+b)|x|^p}}{B(\alpha,\alpha)[1+e^{-x(a+b|x|^p)}]^{2\alpha}[1+e^{-k x(a+b|kx|^p)}]} \quad 15$$

Introducing the location parameter μ∈ᴿ, the density function of X ~ RSGA (a, b, p, α, k) is given by

$$s_1(x) = \frac{2[a+b(p+1)\,|\,x-\mu\,|^p] e^{-\alpha(x-\mu)(a+b)|x-\mu|^p})}{B(\alpha,\alpha)[1+e^{-x(a+b|x-\mu|^p)}]^{2\alpha}[1+e^{-k(x-\mu)(a+b|k(x-\mu)|^p)}]} \quad 16$$

For certain values of the parameters, s(x) and s1(x) are plotted in **Figure 5** for k =±0.7 and in **Figures 6 and 7** for a=0 and b=0 respectively.

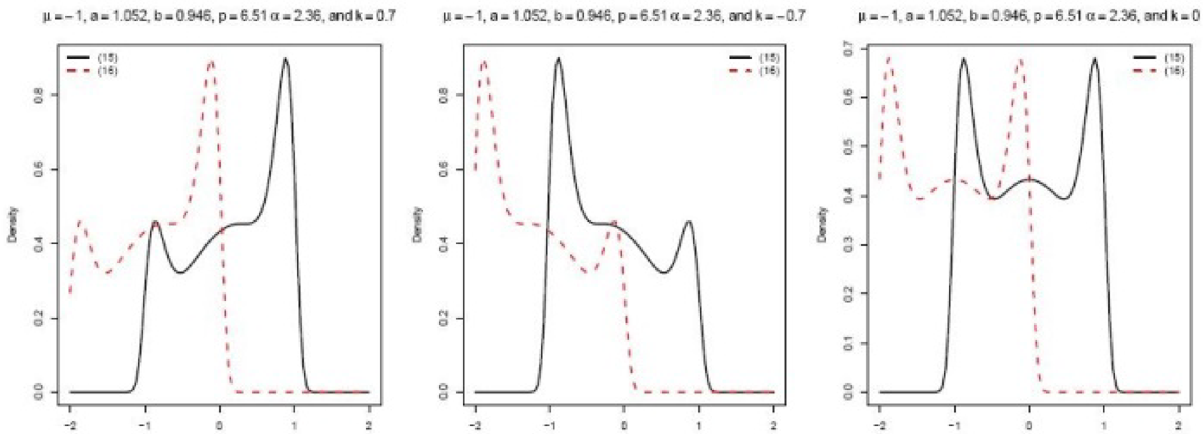


Figure 5. Graphs of (15) and (16) for certain values of the parameters.

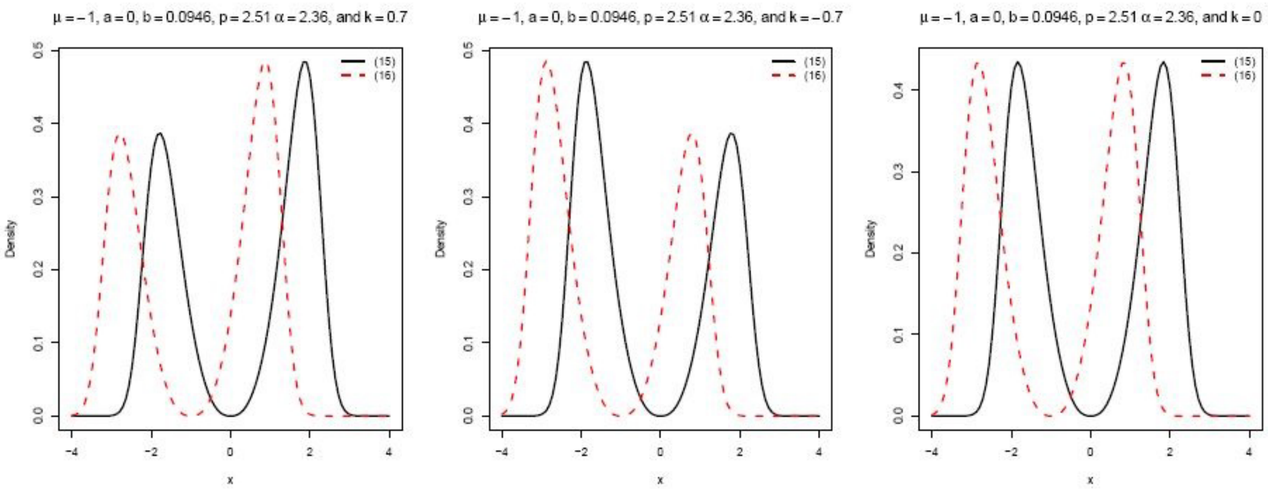


Figure 6. Graphs of (15) and (16) for certain values of the parameters with a=0.

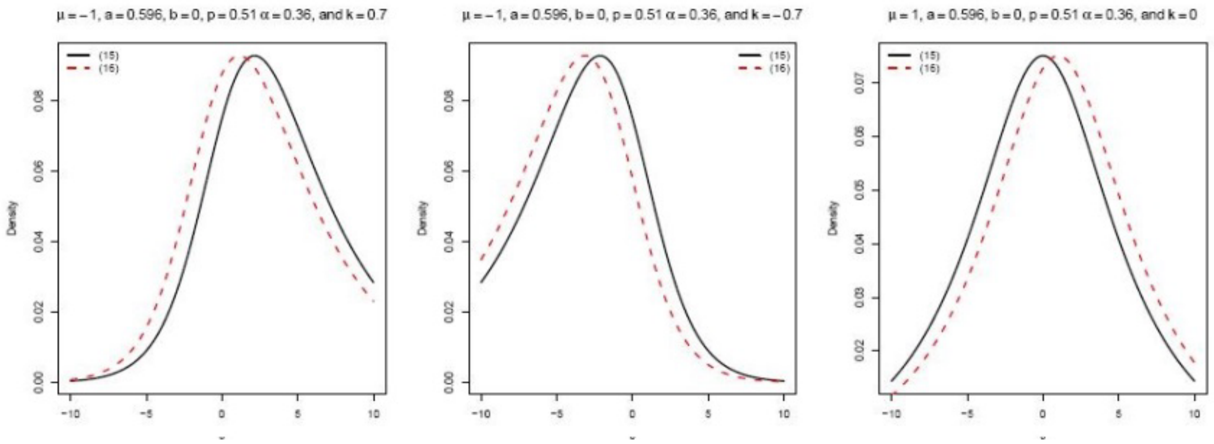


Figure 7. Graphs of (15) and (16) for certain values of the parameters with b=0.

APPLICATIONS INVOLVING REAL DATA

In the present section, five data sets are analyzed by using the distributions defined in earlier sections as well as the mixture of two normals for bimodal data. The estimation of parameters is done by utilizing the method of maximum likelihood estimation. Akaike Cri- terion Information [4], Bayesian Information Criterion, Mean Square Error, Absolute Mean Deviation and Maximum Absolute Deviation are calculated to judge the fit of RSG, RSGA and mixture of two normals. The goodness of fit test of Kolmogorov-Smirnov is used with significance level of 5%. Some packages of sotfware R are used. The GenSA package [5] is used to obtain initial values to interactive algorithm. For interactive algorithm, we use the bbmle::mle2 package [6], in most cases, using BFGS method and optimizer constrOptim to guarantee that the estimated parameters are consistent within their respective parametric space. For more details to adaptive barrier algorithm, see stats::constrOptim into soft- ware R. We obtain the estimates of the parameters, approximate the standard errors of the estimates based on quadratic approximation to the curvature at the maximum likelihood estimate, and a test (z test) of the parameter difference from zero based on this standard error and on an assumption that the sampling distribution of the estimated parameters is normal.

The AIC and BIC for the classification of the model-fit on data sets in various applications will be used. These are defined below

$$AIC = -2l(\hat{\theta}; \mathrm{x}) + 2n_{par} \qquad 17$$

where n_{par} is the number of parameters to be estimated and l(.;.) is the logarithm of the estimated likelihood function.

$$BIC = -2l(\hat{\theta}; \mathrm{x}) + n_{par} \log n \qquad 18$$

where η is the number of observations. Mean Square Error (MSE), Mean Absolute Deviation (MAD) and Maximum Absolute Deviation (MD) are defined below:

$$\mathrm{MSE} = \sum_{i=1}^{n} \frac{[\overline{F}(x_i) - \hat{F}(x_i)]^2}{n}$$

$$\mathrm{MAD} = \sum_{i=1}^{n} \frac{[\overline{F}(x_i) - \hat{F}(x_i)]^2}{n}$$

$$\mathrm{MD} = \max | \overline{F}(x_i) - \hat{F}(x_i) |, i = 1, 2, 3....n,$$

where $\hat{F}$ is the empirical cumulative distribution and $\hat{F}$ is the fitted cumulative distribution of the data. Of course, the smallest value obtained will indicate that there is a good fit.

Human body fat index

The data consist of 252 observations on 17 variables about human body fat. For details, see Jonhson [7], Penrose et al. [8], and Ambler et al. [9]. **Figure 8** demonstrates that the data is unimodal which is also confirmed by test [10,11] with statistics D=0.014114 and p-value near 1. The estimates of the parameters using RSG and RSGA models are given in **Table 1**.

Table 1. Estimates associated with RSG and RSGA models.

RSG Parameter	Estimate	Error	z-value	P (z)
μ	19.26	2.1087×10^{-5}	9.1336×10^{5}	<0.0001
a	0.15401	1.127×10^{-2}	13.662	<0.0001
b	10^{-4}	3.3937×10^{-5}	2.9467	<0.004
p	2.1986	1.0742×10^{-4}	2.0468×10^{4}	<0.0001
α	1.2338	8.1766×10^{-4}	1.5089×10^{3}	<0.0001
log L	−890.9885			
RSGA Parameter	**Estimate**	**Error**	**z-value**	**P (z)**
μ	7.8768	1.0392×10^{-2}	757.9289	<0.0001
a	0.18403	2.8006×10^{-2}	6.5712	<0.0001
b	10^{-4}	3.0035×10^{-5}	3.3294	<0.0001
p	2.2996	1.9703×10^{-3}	1167.086	<0.0001
α	0.35062	7.2149×10^{-2}	4.8597	<0.0001
k	1.7177	2.8455×10^{-2}	60.3678	<0.0001
log L−889.786				

Table 2 shows the comparison of the models used. **Figure 8** presents the histogram with adjusted models. The empirical and theoretical distributions are shows in **Figure 9**.

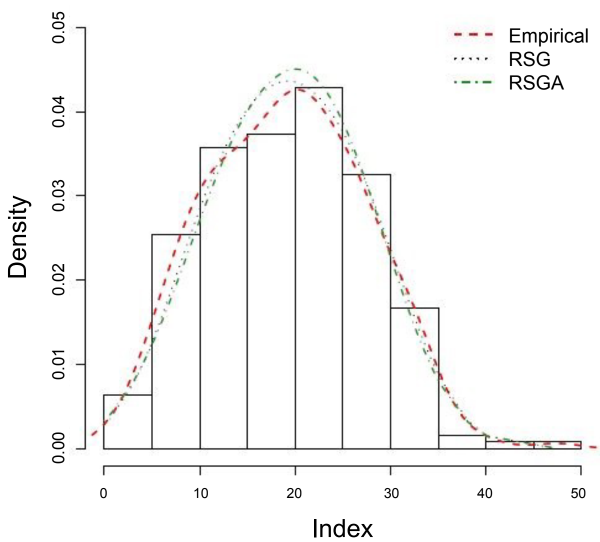
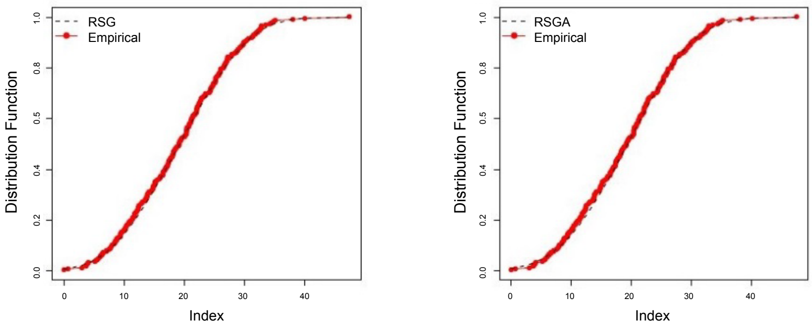


Figure 8. Adjustments of two new distributions to Body Fat Index.

Table 2. The comparison of adjusted models used.

Model	K-S	p-value	MSE (10−4)	MAD	MD	AIC	BIC
RSG	0.047619	0.9375	1.315639	0.009163	0.033421	1791.977	1809.624
RSGA	0.06746	0.615	1.189378	0.008951	0.030355	1791.572	1812.749



(a) RSG distribution (b) RSGA distribution

Figure 9. Graphs of empirical and theoretical distributions.

For AIC, it may be observed that the RSGA fit is better than RSG fit for this data set. The Bayesian criterion indicates a better fit for RSG distribution.

Precipitation

The data consist of 121 observations about annual precipitation (rain) between 1978 and 1998 at the center of the city of Los Angeles. These data were obtained from the site [12]. **Figure 10** demonstrates that the data is unimodal which is also confirmed by Hartigan's test with statistics D=0.027273 and p-value equal to 0.7971. The estimates of the parameters, using RSGA distribution, are given in **Table 3**.

Table 3. Estimates associated with RSGA model.

Parameter	Estimate	Error	z-value	P(z)
μ	4.0393	4.4968×10^{-2}	89.825	<0.0001
a	49.999	$2:6007 \times 10^{-4}$	1.9225×10^{5}	<0.0001
b	34.113	3.9072×10^{-4}	8.7308×10^{4}	<0.0001
p	0.7582	0.1095	6.9239	<0.0001
α	2.9333×10^{-4}	8.0064×10^{-5}	3.6638	<0.0003
	3.838	2.6556×10^{-4}	$1.4452.10^{-4}$	<0.0001
Log L-393.2849				

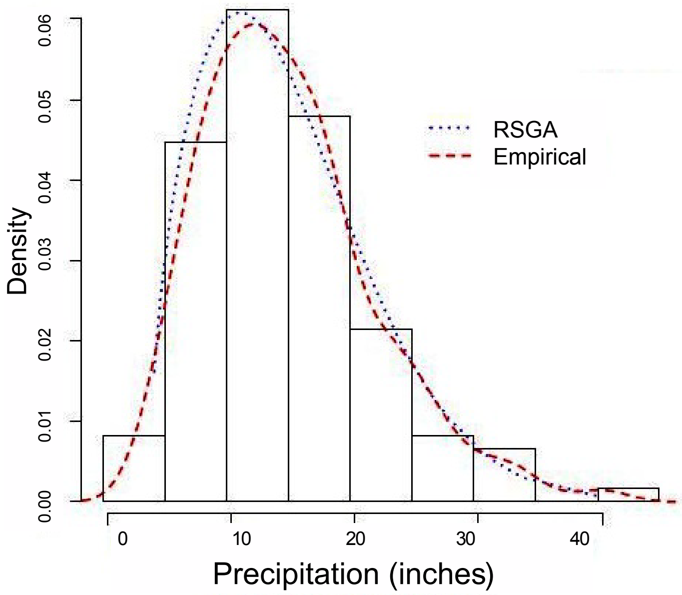


Figure 10. RSGA distribution fitted to histogram.

Applying the non-parametric Kolmogorov-Smirnov test, the K-S value obtained is 0.07438 with p-value 0.8914, thus not reject the hypothesis that the data satisfies RSGA distribu- tion. In 2014, Eirado et al. [13] proposed an asymmetric model and applied to this data set. The MSE obtained is equal to 0.001058396, the mean absolute deviation (MAD) is 0.02785116 and the maximum absolute deviation (MD) is 0.06496284. Also, we obtained MSE equal to 0.0002414233, MAD equal to 0.01185483 and MD equal to 0.04669135.

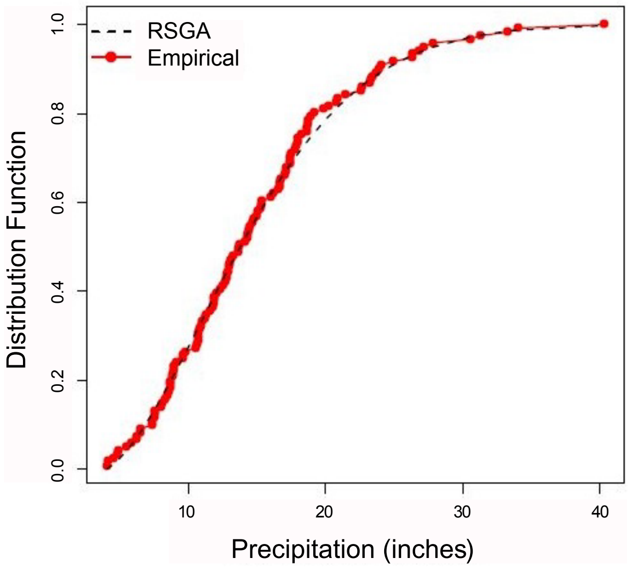


Figure 11. Empirical and theoretical distributions of precipitation.

AIC and BIC of the fits of the two models are given in **Table 3,4**. The empirical and theoretical distributions are shows in **Figure 11**. Clearly, the RSGA distribution gave better fit to the precipitation data.

Table 4. The comparison of the models.

Model	log-likelihood	AIC	BIC
RSGA	−393.2849	798.5697	815.3444
Eirado-Rathie	−551.6425	1113.285	1127.264

HIV Data

The HIV data with 2843 observations is available in fitdistrplus: Aids2 package of software R, giving the age when a patient is diagnosed with AIDS in Australia in 1991. **Table 5** presents the estimates of the parameters of RSG and RSGA models.

Table 5. Estimates associated with RSGA and RSG models.

RSG	Estimate	Error	z-value	P (z)
μ	36.931	0.18698	197.51	<0.0001
a	0.16731	0.017989	9.3006	<0.0001
p	8.9282	5.2278×10^{-17}	1.7078×10^{17}	<0.0001
α	1.1148	0.017463	6.3838	<0.0001
log L	−10552.23			
RSGA	**Estimate**	**Error**	**z-value**	**P (z)**
μ	27.477	0.031826	86.336	<0.0001
a	0.05717	0.001222	46.779	<0.0001
p	9.7371	1.0564×10^{-15}	9.2174×10^{15}	<0.0001
α	3.5391	0.20708	17.091	<0.0001
k	4.5317	0.17192	26.359	<0.0001
log L−10508.95				

In **Table 6**, the estimates of the parameters of the skew normal (NORSKEW) and normal distributions are given.

Table 6. Estimates associated with asymmetric normal and normal distributions.

NORSKEW	Estimate	Error	z-value	P (z)
μ	37.5304	0.187355	200.317	<0.0001
σ	10.01696	0.13529	74.041	<0.0001
ξ	1.273675	0.031561	40.355	<0.0001
log L −10549.26				
μ	37.40907	0.1887	198.245<0.0001	
σ	10.06149	0.13343	75.406<0.0001	
log L−10597.72				

The values of AIC, BIC, MSE, MAD and MD given in **Table 7** indicate that the RSGA model fits well the HIV data.

Table 7. Comparison of the models used.

Model	K-S	p-value	MSE(10−4)	MAD	MD	AIC	BIC
RSG	0.289524	0.0014	4.376593	0.017052	0.04955	21112.47	21136.28
RSGA	0.063492	0.69	1.450326	0.009691	0.032249	21027.9	21057.66
NORSKEW	0.041857	0.01373	3.033533	0.014451	0.040824	21104.53	21122.39
NORMAL	0.059796	7.696×10^{-5}	8.539093	0.025396	0.058367	21199.44	21211.35

Histogram and RSGA distributions to HIV data are shown in **Figure 12** while Empirical and RSGA distributions in **Figure 13.** In **Table 7**, the Kolmogorov-Smirnov test rejects almost all adjusted distributions except RSGA distribution.

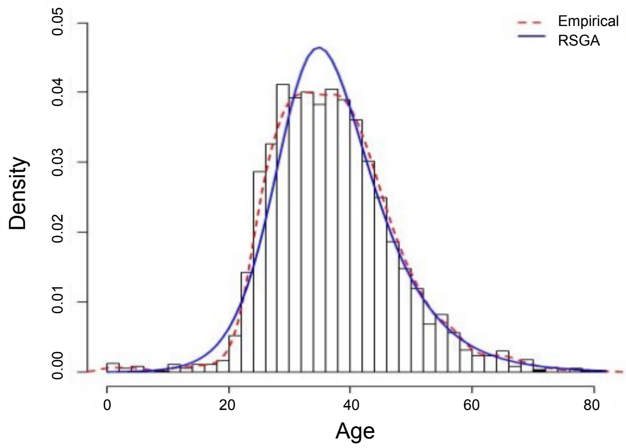


Figure 12. Adjustments of RSGA distribution to HIV data.

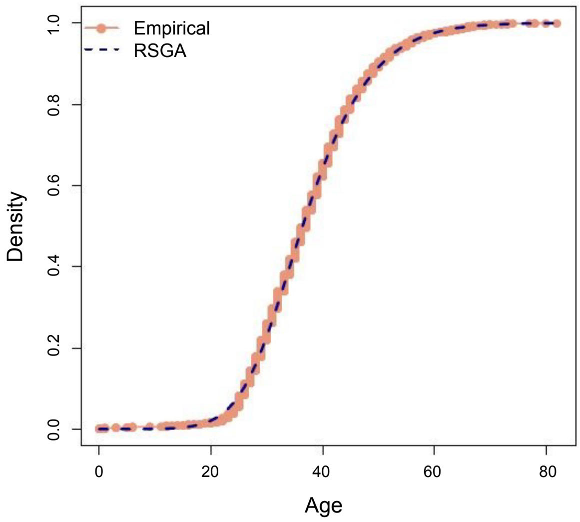


Figure 13. Empirical and theoretical distributions to HIV data.

pH Concentration data

The pH concentration data [14] with 252 observations show bimodality which is also demonstrated by Hartigan's test with statistics of the test equal to 0.046498 and p-value of 0.00045. The estimates of the parameters are given in **Table 8**.

Silva et al. [15] proposed two new asymmetric models by Azzalini's method h1(x) and h2(x) where the pH concentration data was fitted by these two models. **Table 10** shows the performance of the fitted distributions.

Table 8. Estimates associated with RSGA and RSG models.

RSGA	Estimate	Error	z-value	P (z)
μ	3.094726	0.071289	43.4109	<0.0001
a	8.242063	2.241954	3.6763	<0.0003
b	0.003	0.001066	2.8153	0.004874
p	6.244648	0.344886	18.1064	<0.0001
α	0.045077	0.011673	3.8616	<0.0002
k	0.86603	0.335523	2.5811	0.009848
log L −364.2				

μ	4.918676	0.042907	114.6364	<0.0001
a	6.027683	0.616692	9.7742	<0.0001
b	2.906972	1.071397	2.7133	<0.007
p	2.711035	0.459798	5.8961	<0.0001
α	0.068114	0.006893	9.8812	<0.0001
log L −363.7172				

Using package of Benaglia et al. [16], the estimates of mixture of normals are given in **Table 9** with parametric bootstrap performed for standard error approximation.

Table 9. Estimates of mixture of two normal.

Parameters	Component 1	Component 2	Error of Component 1	Error of Component 2
λ	0.50439	0.49561	0.041677	0.0416768
μ	3.892103	5.961384	0.076694	0.07539492
σ	0.575443	0.568638	0.056243	0.05409495
log L	−366.8661			

Histogram of pH values along with the distributions adjusted are shown in **Figures 14 and 15**

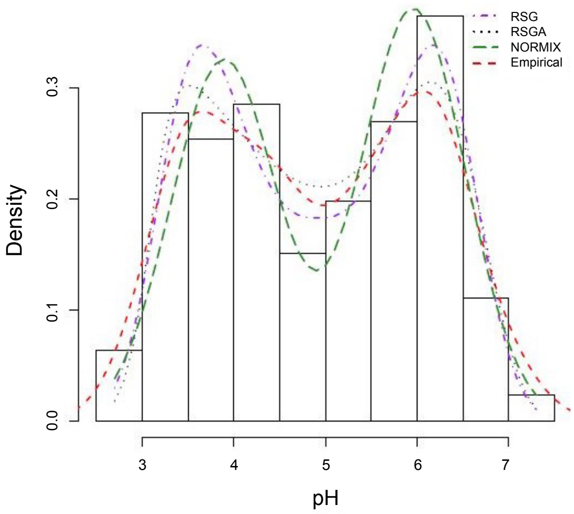
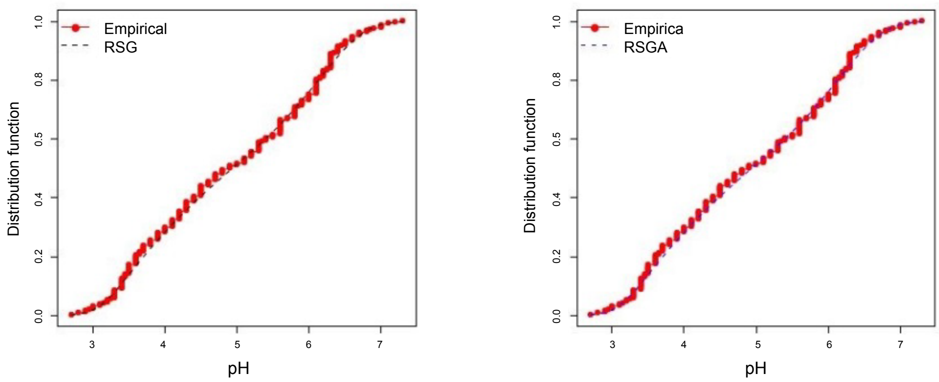


Figure 14. pH histogram and the fitted models.



(a) RSG distribution

(b) RSGA distribution

Figure 15. Graphs of empirical and theoretical distributions.

Table 10 gives the accuracy values of AIC, BIC, MSE etc, for various models. The RSG model adjusted well the bimodal data.

Table 10. Comparison of the models used.

Model	K-S	p-value	MSE (10^{-4})	MAD	MD	AIC	BIC
RSG	0.06746	0.61	1.814886	0.01067501	0.039083	737.4343	755.0871
RSGA	0.075397	0.4709	2.546568	0.01283771	0.038684	740.4067	761.5833

NORMIX	0.083333	0.3457	7.407901	0.02202145	0.064505	743.7322	761.3793
h1(x)	–	0.8316	3	0.0152	0.0373	744.6913	776.4561
h2(x)	–	0.09438	96	0.0912	0.1454	857.387	889.1519

Relative Humidity (RH)

The RH observations data are taken from Nychka et al. [17]. The estimates of the parameters for RH data using the RSGA model are given in **Table 11**.

Table 11. Estimation of the parameters of the RSGA model.

Parameter	Estimate	Error	z-value	P (z)
μ	59.72236	0.008989	6643.879	<0.0001
a	0.034228	0.016025	2.1359	<0.04
b	0.002588	0.001281	2.0199	<0.05
p	1.227392	0.151744	8.0886	<0.0001
α	0.266291	0.115667	2.3022	<0.03
k	−0.4621166	0.095596	−4.8341	<0.0001

The estimation for a mixture of two normal s are given in **Table 12**. The values of AIC, BIC etc. measuring the quality of fit are given in **Table 13**.

Table 12. Estimation of the parameters of the mixture of two normal.

NORMIX	Component 1	Component 2	Error Component 1	Error Component 2
λ	0.6975	0.3025	0.025634	0.02563423
μ	36.8122	77.08626	0.865337	1.139471
σ	11.835	9.28641	0.648855	0.8474422
log L	−1958.626			

Table 13. Comparison of the models used,

Model	K-S	p-value	EQM (10−4)	MAD	MD	AIC	BIC
RSGA	0.080178	0.1115	8.497387	0.0217099	0.076005	3926.544	3951.182
NORMIX	0.073497	0.1768	9.316102	0.02176488	0.066207	3927.252	3947.787

In **Figure 16**, the histogram and the fit using Empirical, RSGA and the mixture of two normals distributions are shown. In **Figure 17**, the empirical and theoretical distributions are shown.

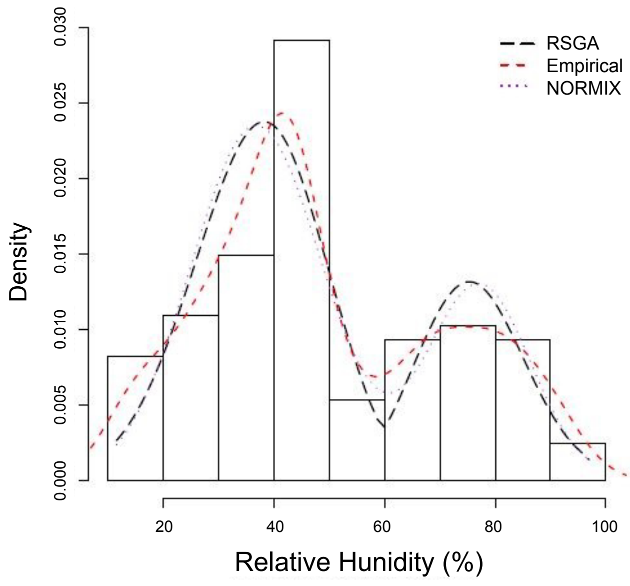


Figure 16. Relative Humidity and adjusted model.

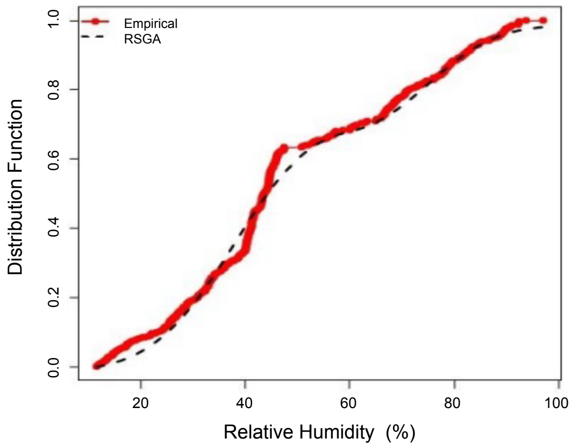


Figure 17. The empirical and theoretical distributions.

CONCLUSION

The Rathie-Swamee generalized distribution (RSG) and its skew form (RSGA) proved useful to five data sets analyzed, thus demonstrating their applicabilities over the mixture of two normals, in case of bimodal sets (pH concentration and relative humidity).

ACKNOWLEDGEMENT

P. N. Rathie thanks the Coordination for the Improvement of Higher Level Personnel (CAPES) for supporting his Senior National Visiting Professorship.

REFERENCES

1. Rathie PN, et al. On a new invertible generalized logistic distribution approximation to normal distribution, technical research report in statistics. Appli scie phy. 2006;1:120-125.
2. Jones M, et al. Families of distributions arising from distributions of order statistics. Test. 2004;13:1-43.
3. Azzalini A. A class of distributions which includes the normal ones. Scand J Stat Theory. 1985;12:171-178.
4. Akaike RW. A new look at the statistical model identi cation. IEEE Trans Automat Contr. 1974;19:716-723.
5. Xiang Y, et al. Generalized simulated annealing for ecient global optimization: the GenSA package. R J. 2013;5:13-28.
6. https://rdrr.io/cran/bbmle/
7. Johnson RW and Carleton C. Fitting percentage of body fat to simple body measurements. J Stat Educ. 1996;4:265-266.
8. Penrose KW, et al. Generalized body composition prediction equation for men using simple measurement techniques, Med Scie Spo Exe. 1985;17:1-8.
9. https://cran.r-project.org/web/packages/mfp/index.html
10. Hartigan JA and Hartigan PM. The dip test of unimodality. Ann Stat.1985;13:70-84.
11. https://cran.r-project.org/web/packages/diptest/index.html
12. http://www.noaa.gov/
13. Eirado CRDR and Rathie PN. On new invertible skew and symmetric distributions. Stat Rese Lett.2014;3:17-22.
14. Growitz DJ, et al. Reconnaissance of mine drainage in the coal elds of eastern Pennsylvania. Tech Rep US Geological Survey. 1985.
15. Silva RO, et al. On the new skew distributions using azzalinis formula. Mathematica Aeterna. 2016;6:649-673.
16. Benaglia T, et al. Mixtools: An R package for analyzing nite mixture models. J Stat Softw. 2009;32:1-29.
17. Nychka D, et al. Tools for plotting skew-T diagrams and wind profiles, R package version. 2014;1.4.