Original Research Article

Some Area-Biased Distributions

Abstract

The preference of the proper distribution for modelling data is often a substantial concern to researchers and practitioners. For this reason, new statistical distributions or the generalizations of well-known distributions have been proposed for flexible modeling. Weighted distributions are one of the generalization methods for these distributions. In this article, area biased beta, Rayleigh and log-normal distributions are introduced. Some main statistical properties, like probability density functions, cumulative distribution functions, moments and the estimation of the parameters of these distributions are obtained. Real data examples are used for illustration of these distributions.

Keywords: Weighted distribution; Area-Biased distributions; Log-Normal; Rayleigh; Beta

2010 Mathematics Subject Classification: 62E10; 62P99

1 Introduction

In recent years, it has been seen that, traditional statistical distributions have limited application areas. For this reason, some authors have proposed some generalization methods of these distributions or some other new distributions. Rao (1965) and Patil and Rao (1977, 1978) have discussed and introduced the weighted distributions and applied these distributions to real life data. The weighted distribution is defined as

$$f(x;\theta) = \frac{w(x)f_0(x;\theta)}{E(w(x))} \tag{1.1}$$

where $w(x)$ is any function of random variable X. Size-biased distributions are the special cases of the weighted distributions and has the form

$$f(x;\theta) = \frac{x^k f_0(x;\theta)}{\mu_k} \tag{1.2}$$

where $f_0(x;\theta)$ is the original underlying distribution and $\mu_k = E(X^k)$ is the k^{th} moment. We get the length-biased and area-biased distributions for $k = 1$ and $k = 2$ respectively.

Size-biased observations occur in many research areas and its fields of applications includes econometrics, environmental science, medical science, sociology, psychology, ecology, geological sciences etc. There are many papers dealing with size-biased distributions methodology. Dennis and Patil (1984) have introduced weighted multimodal gamma distributions as models of population abidance, Patil and Rao (1978) also discussed weighted binomial distribution to model the human families and estimation of the wildlife family size, Das and Roy (2011) have studied the applicability of length-biased weighted generalized Rayleigh Distribution, Ducey and Gove (2015) have proposed size-biased distributions in the generalized beta distribution family and applied this distribution to the forestry, Perveen et al. (2016) proposed size-biased double weighted exponential distribution and Al-Omari et al (2019) have studied the size-biased Ishita distribution. On the other hand, area-biased distributions have been studied but there is not wide literature as length-biased distributions.

In this paper, some area-biased distributions have been studied. Statistical properties like expected value, variance, cumulative distribution function, parameter estimations of these distributions have been obtained. Some real-life data examples have been applied to these area-biased distributions. The rest of the paper organizes as follows: In Section 2, area-biased beta (ABB), area-biased Rayleigh (ABR) and area-biased log-normal (ABLN) distributions are introduced and main statistical properties of them are obtained. In section 3, real life data examples are given for illustration of these introduced distributions. Conclusion is given at the end of this study.

2 Some Area-Biased Distributions

In this section, we introduce ABB, ABR and ABLN distributions.

2.1 Area-Biased Beta Distribution

Beta distribution is one of the well-known statistical distributions. Since the range of Beta distribution is between the interval (0,1), it has a various application area especially modelling the data about proportions. Beta distribution is also used in Bayesian estimation method as a prior distribution. The ABR distribution has the following pdf

$$f(x) = \frac{(\alpha+\beta+1)(\alpha+\beta)x^{\alpha+1}(1-x)^{\beta-1}}{\alpha(\alpha+1)B(\alpha,\beta)},\ 0 < x < 1 \tag{2.1}$$

where $B(\alpha,\beta)$ is the Beta function, and the corresponding cdf is

$$F(x) = \frac{(\alpha+\beta)(\alpha+\beta+1)I_x(k,\beta)}{\alpha(\alpha+1)},\ 0 < x < 1 \tag{2.2}$$

where $I_x(k,\beta)$ is the regularized incomplete beta function and $k = \alpha+2$. Figure 1 shows the possible shapes of the ABB distribution with different α and β values.

The expected value and the variance of the ABB distribution are obtained

$$E(X) = \frac{\alpha+2}{\alpha+\beta+2} \tag{2.3}$$

and

$$V(X) = \frac{\beta(\alpha+2)}{(\alpha+\beta+3)(\alpha+\beta+2)^2} \tag{2.4}$$

respectively. The general formula the moments of the ABB distribution is

$$E(X^n) = \frac{\alpha+n+1}{\alpha+\beta+n+1}E(X^{n-1}) \tag{2.5}$$

respectively.

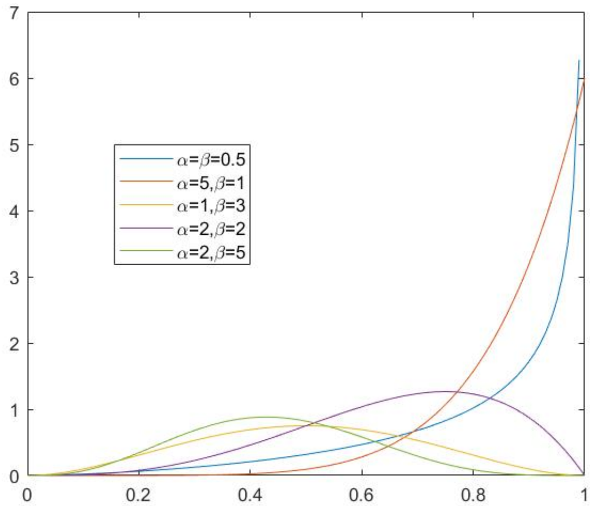

Figure 1: Possible shapes of ABB distribution.

Table 1: The expected values, the standard deviations, the skewness and the kurtosis values of ABB and Beta distributions

(α,β)	$(0.5, 0.5)$		$(5, 1)$		$(1, 3)$		$(2, 2)$		(2,5)	
	ABB	B	ABB	B	ABB	B	ABB	B	ABB	B
$E(X)$	0.83	0.5	0.87	0.83	0.50	0.25	0.66	0.50	0.44	0.28
$S(X)$	0.18	0.35	0.11	0.14	0.18	0.19	0.17	0.22	0.15	0.16
γ_1	-1.43	0.00	-1.36	-1.18	0.00	0.86	-0.46	0.00	0.13	0.59
γ_2	4.56	1.51	4.97	4.20	2.33	3.09	2.62	2.14	2.52	2.88

Table 1 shows the expected value, variance, skewness and kurtosis values of the ABB distribution and original beta distribution as a comparison.

To obtain the maximum likelihood estimators of the unknown parameters, we use the following log-likelihood function.

$$\begin{aligned} lnL = nln(\alpha+\beta+1) + nln(\alpha+\beta) + (\alpha+1)\sum_{i=1}^{n} ln(x_i) + (\beta-1)\sum_{i=1}^{n} ln(1-x_i) \\ -nln(\alpha) - nln(\alpha+1) - nlnB(\alpha,\beta) \end{aligned} \tag{2.6}$$

The partial derivatives of the corresponding likelihood function are obtained as

$$\begin{aligned} \frac{\partial \ln L}{\partial \alpha} &= \frac{n}{\alpha+\beta+1} + \frac{n}{\alpha+\beta} - \frac{n}{\alpha} - \frac{n}{\alpha+1} + \sum_{i=1}^{n} ln(x_i) + \frac{n\psi(\alpha+\beta)}{\Gamma(+\beta)} - \frac{n\psi(\alpha)}{\Gamma(\alpha)} = 0 \\ \frac{\partial \ln L}{\partial \beta} &= \frac{n}{\alpha+\beta+1} + \frac{n}{\alpha+\beta} + \sum_{i=1}^{n} ln(1-x_i) + \frac{n\psi(\alpha+\beta)}{\Gamma(+\beta)} - \frac{n\psi(\beta)}{\Gamma(\beta)} = 0 \end{aligned} \tag{2.7}$$

where $\psi(.)$ is the derivative of gamma function. It can be seen from the likelihood equations, there is no explicit solutions. Therefore, iterative methods can be used to solve the equations.

2.2 Area-Biased Rayleigh Distribution

Rayleigh distribution is one of the continuous distributions with non-negative range. It has an application areas like magnetic resonance imaging, ballistics and physical oceanography. The ABR distribution has the following pdf

$$f(x) = \frac{x^3}{2\sigma^4} e^{-\left(\frac{x^2}{2\sigma^2}\right)},\ x > 0 \tag{2.8}$$

The corresponding cdf is

$$F(x) = 1 - e^{-\left(\frac{x^2}{\sigma^2}\right)} \left(\frac{x^2}{\sigma^2} + 1\right),\ x > 0. \tag{2.9}$$

The expected value and the variance of the ABR distribution are obtained

$$E(X) = \sigma\sqrt{2}\Gamma\left(1 + \frac{3}{2}\right) = 1.88\sigma \tag{2.10}$$

and

$$V(X) = 2\sigma\left(2\sigma - \Gamma\left(1 + \frac{3}{2}\right)^2\right) \tag{2.11}$$

respectively. The general formula the moments of the ABR distribution is

$$E(X^n) = \sigma^n 2^{n/2}\Gamma\left(1 + \frac{n+2}{2}\right). \tag{2.12}$$

Figure 2 shows the possible shapes of the ABR distribution for different σ parameter.

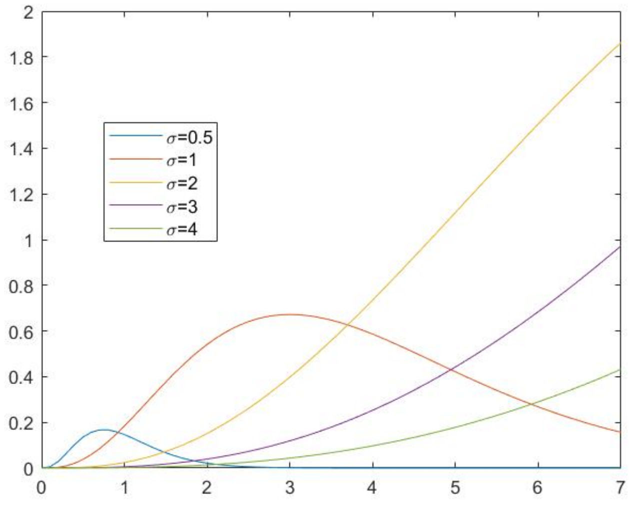

Figure 2: Possible shapes of ABR distribution.

Table 2 shows the expected value, standard deviation, skewness and kurtosis values of the ABR and Rayleigh distribution.

Table 2: The expected values, the standard deviations, the skewness and the kurtosis values of ABR and Raylegih distributions

σ	0.5		1.0		2.0		3.0		4.0	
	ABR	R	ABR	R	ABR	R	ABR	R	ABR	R
$E(X)$	0.94	0.62	1.88	1.25	3.76	2.51	5.64	3.76	7.52	5.01
$S(X)$	0.34	0.32	0.68	0.65	1.36	1.31	1.95	2.04	2.72	2.62
γ_1	0.41	0.63	0.41	0.63	0.41	0.63	0.41	0.63	0.41	0.63
γ_2	3.05	3.24	3.05	3.24	3.05	3.24	3.05	3.24	3.05	3.24

To obtain the ML estimator of the unknown model parameter the following log-likelihood function is maximized.

$$lnL = 3\sum_{i=1}^{n} ln(x_i) - nln2 - 4nln(\sigma) - \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} \tag{2.13}$$

The likelihhod equaiton is obtained as

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{4n}{\sigma} + \frac{\sum_{i=1}^{n} x_i^2}{\sigma^3} \tag{2.14}$$

and the solution is called ML estimator of σ.

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{4n}} \tag{2.15}$$

2.3 Area-Biased Log-Normal Distribution

In probability theory, if the random variable X is distributed log-normal, then $Y = ln(X)$ is distributed normally. It has a wide application area such as human behaviors, medicine, biology, engineering and economics. The ABLN distribution has the following pdf.

$$f(x) = \frac{x}{e^{2\mu+2\sigma^2}\sigma\sqrt{2\pi}} exp\left(-\frac{(ln(x)-\mu)^2}{2\sigma^2}\right),\ x > 0 \tag{2.16}$$

Figure 3 shows the possible shapes of the ABLN for different μ and σ paramters.

The expected value and the variance of ABLN distribution are obtained as

$$E(X) = e^{\mu+5/2\sigma^2} \tag{2.17}$$

and

$$V(X) = e^{2\mu+5\sigma^2}(e^{\sigma^2} - 1) \tag{2.18}$$

respectively. The general formula the moments of the ABLN distribution is

$$E(X^n) = e^{n\mu+(1/2(n+2)^2-2)\sigma^2} \tag{2.19}$$

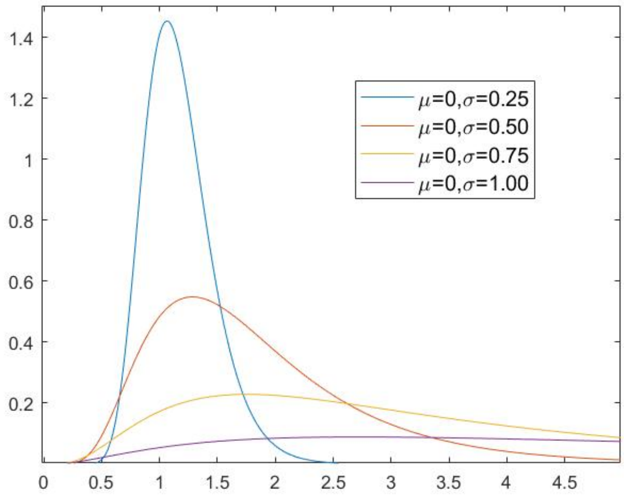

Figure 3: Possible shapes of ABLN distribution.

Table 3 shows the expected values, standard deviations, skewness and the kurtosis values of the ABLN and log-normal distributions.

Table 3: The expected values, the standard deviations, the skewness and the kurtosis values of ABL and Log-Normal distributions

(μ,σ)	$(0,0.25)$		$(0,0.50)$		$(0,0.75)$		$(0,1.00)$	
	ABLN	LN	ABLN	LN	ABLN	LN	ABLN	LN
$E(X)$	1.16	1.03	1.86	1.13	4.08	1.32	12.1	1.64
$S(X)$	0.29	0.26	0.99	0.61	3.54	1.15	15.9	2.16
γ_1	0.77	0.77	1.75	1.75	3.26	3.26	6.18	6.18
γ_2	4.09	4.09	8.89	8.89	26.5	26.5	113.9	113.9

To obtain the ML estimators of the unknown parameters, the loglikelihood function

$$lnL = \sum_{i=1}^{n} ln(x_i) - n(2\mu + 2\sigma^2) - nln(\sigma) - n/2ln(2\pi) - \sum_{i=1}^{n} \frac{(ln(x_i) - \mu)^2}{2\sigma^2} \tag{2.20}$$

is maximized with respect to the unknown parameters. By differentiating the log-likelihood function with respect to the unknown parameters and equating them to zero we obtain the following likelihood equations

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu} &= -2n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(ln(x_i) - \mu) = 0 \\ \frac{\partial \ln L}{\partial \sigma} &= 4n\sigma + \frac{n}{\sigma} - \sum_{i=1}^{n}\frac{(ln(x_i) - \mu)^2}{\sigma^3} = 0 \end{aligned} \tag{2.21}$$

Solutions of these equations are the ML estimators.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} ln(x_i) - 2n\sigma^2}{n} \tag{2.22}$$

and

$$\hat{\sigma^2} = \frac{-n + \sqrt{n^2 + 16n \sum_{i=1}^{n} (ln(x_i) - \mu)^2}}{8n} \tag{2.23}$$

3 Numerical Data

The first data is about U.S family nominal income taken from the Census Population Report and the data is used to fit weighted generalized beta distribution at Ye (2012). The data is: 0.124, 0.114, 0.105, 0.148, 0.179, 0.126, 0.122, 0.043, 0.039. We fit the data with traditional beta, SBB and ABB distributions. The table shows the ML estimations of the unknown parameters α and β, log-likelihood values and AIC values.

Table 4: ML Estimations, lnL and AIC values of Beta, SBB and ABB Distributions

	B	SBB	ABB
$\hat{\alpha}$	4.58	3.25	2.06
$\hat{\beta}$	36.82	37.56	32.25
lnL	14.75	14.87	14.96
AIC	-25.51	-25.75	-25.93

It can be seen from the table that the data fits ABB distribution with respect to SBB and traditional beta distribution with maximum log-likelihood and minimum AIC values.

The second data is about the mole fractions given at Bennett and Filliben (2000). The data set is 3.051, 2.779, 2.604, 2.371, 2.214, 2.045, 1.715, 1.525, 1.296, 1.154, 1.016, 0.7948, 0.7007, 0.6292, 0.6175, 0.6449, 0.8881, 1.115, 1.397, 1.506, 1.528. Ajami and Jahanshahi (2017) used this data by modelling size-biased Rayleigh (SBR) distribution. We fit the data with traditional Rayleigh, SBR and ABR distributions. The table shows the ML estimations of the unknown parameter σ, log-likelihood values and AIC values.

Table 5: ML Estimations, lnL and AIC values of R, SBR and ABR Distributions

	R	SBR	ABR
$\hat{\sigma}$	1.118	0.945	0.819
lnL	-26.96	-26.26	-15.94
AIC	55.94	54.52	33.89

It can be seen from the table that the data fits ABR distribution with respect to SBR and traditional Rayleigh distribution with maximum log-likelihood and minimum AIC values.

The last data is about the incomes given at Aitchison and Brown (1963). The data set is 52.53, 73.30, 61.96, 47.36, 46.23, 52.86, 58.49, 67.95, 72.27, 66.11, 70.74, 76.22, 91.91, 114.03. The authors used this data by modelling log-normal (LN) distribution. We fit the data with LN, Size-biased log-normal (SBLN) and ABLN distributions. The table shows the ML estimations of the unknown parameters μ and σ, log-likelihood values and AIC values.

Table 6: ML Estimations, lnL and AIC values of LN, SBLN and ABLN Distributions

	L	SBLN	ABLN
$\hat{\mu}$	4.22	4.16	4.10
$\hat{\sigma}$	0.056	0.056	0.055
lnL	-38.70	-38.45	-38.30
AIC	81.41	80.91	80.60

It can be seen from the table that the data fits ABLN distribution with respect to SBLN and traditional LN distribution with maximum log-likelihood and minimum AIC values.

4 Conclusion

Area-biased distributions are the specific version of the weighted distributions which have been used in some modelling applications. In this paper, we introduce the area-biased beta, Rayleigh and log-Normal distributions. We obtain some main characteristics of these introduced distributions. Real data sets are given for illustrating and the introduced distributions give better fits than the corresponding underlying and size-biased versions of them.

References

Aitchison, J. and Brown, J.A.C. (1963). The Log-Normal Distribution, Cambridge at the university press.

Ajami, M. and Jahanshahi, S.M.A. (2017). Parameter Estimation In Weighted Rayleigh Distribution. Journal of Modern Applied Statistical Methods. 16(2), 256-276.

Al-Omari, A.I. Al-Nasser, A.D. and Ciavolinio, E. (2019). A size-biased Ishita distribution and application to real data. Qual Quant. 53, 493–512.

Bennet, H. S. and Fiiliben, J. J. A. (2000). A systematic approach for multidimensional, closed form analytic modeling: minority electron mobilities in Ga1-xAlxAs heterostructures. Journal of Research of the National Institute of Standards and Technology. 105(3), 441–452.

Das, K.K. and Roy, T.D. (2011). Applicability of Length Biased Weighted Generalized Rayleigh Distribution. Advances in Applied Science Research. 2, 320-327.

Dennis, B. and Patil, G. (1984). The gamma distribution and weighted multimodal gamma distributions as models of population abtidance. Mathematical Biosciences. 68(2), 187-212.

Ducey, M.J. and Gove, H.G. (2015). Size-biased distributions in the generalized beta distribution family, with applications to forestry. Forestry. 88(1), 143-151.

Patil, G. P. and Rao, C. R. (1977). Weighted distributions: a survey of their applications. In Applications of Statistics; Krishnaiah, P. R., Eds.; North Holland Publishing Company.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics. 34(2), 179-189.

Perveen, Z. Ahmed, Z. and Ahmad, M. (2016). On size-biased double weighted exponential distribution (SDWED). Open J. Stat. 6, 917–930.

Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment. In Classical and Contagious Discrete Distributions; Patil, G.P., Eds.; Pergamon Press

Ye, Y. (2012) Properties of Weighted Generalized Beta Distribution of the Second Kind, Electronic Theses and Dissertations, Georgia Southern University