Journal of Statistics: Advances in Theory and Applications
Volume 13, Number 2, 2015, Pages 167-178

ESTIMATION OF $R = P[Y < X\}$ FOR THE THREE PARAMETER DAGUM DISTRIBUTION

ALAA AHMED

Department of Applied Statistics and Econometrics
Institute of Statistical Studies and Research
Cairo University
Egypt
e-mail: Alaa_mnn@yahoo.com

Abstract

In this paper, we estimate probability $P\{X < Y\}$ when X and Y are two independent random variables from Dagum distribution. We obtain maximum likelihood estimator and its asymptotic distribution. We also perform a simulation study.

1. Introduction

The reliability of a system is the probability that when operating under the state of the environment, the system will perform its intended function adequately. For stress-strength models, both the strength of the system, X, and the stress, Y, imposed on it by its operating environments are considered random variables. In addition, R, can be defined as the probability that the system is strong enough to overcome the stress imposed on it. This model is also known as the load-capacity model in the context of solid mechanics or structural engineering. $P(Y < X)$ is an important tool in other fields too, for example, biometry, physics, and

2010 Mathematics Subject Classification: 62G05, 62H12.
Keywords and phrases: Dagum distribution, maximum likelihood estimation.
Received June 28, 2015

engineering. Inference regarding $P(Y < X)$, defining the reliability of the system, has been widely discussed in literature, when X and Y are assumed to be independent random variables. See, for example, Amin [6], Basu [7], Downtown [17], Tong [31, 32], Kelley et al. [23], Beg [8], Iwase [22], McCool [25], Ivshin [21], Ali et al. [5], Ali et al. [3, 4], Ali and Woo [1, 2], Pal et al. [26], Raqab and Kundu [27], Raqab et al. [28], Rezaei et al. [29], Masoom et al. [24], Wong [33], and Francisco et al. [20].

In this article, we consider the reliability, R, when X and Y are independent but not identically distributed three parameter Dagum random variables. In the 1970s, Camilo Dagum embarked on a quest for a statistical distribution closely fitting empirical income and wealth distributions. Not satisfied with the classical distributions, he looked for a model accommodating the heavy tails present in empirical income and wealth distributions as well as permitting an interior mode. He end up with the Dagum type I distribution, a three-parameter distribution, and two fourparameter generalizations (Dagum [14, 15, 16]). The Dagum distribution is also called the inverse Burr, especially in the actuarial literature, as it is the reciprocal transformation of the Burr XII. Nevertheless, unlike the Burr XII, which is widely known in various fields of science, the Dagum distribution is not much popular, perhaps, because of its difficult mathematical tractability. Since Dagum proposed his model as income distribution, its properties have been appreciated in economics and financial fields and its features have been extensively discussed in the studies of income and wealth. Kleiber and Kotz [9] and Kleiber [10] provided an exhaustive review on the origin of the Dagum model and its applications. Recent contributions from Quintano and D'Agostino [11] adjusted the Dagum model for income distribution to account for individual characteristics, while Domma et al. [18, 19] studied the Fisher information matrix in doubly censored data from the Dagum distribution and reliability studies of the Dagum distribution.

The Dagum distribution has the following distribution function for $X > 0$:

$$F(x) = (1 + \lambda x^{-\beta})^{-\theta}, \quad x > 0, \tag{1.1}$$

with probability density function

$$f(x) = \lambda\beta\theta x^{-(\beta+1)}(1 + \lambda x^{-\beta})^{-(\theta+1)}, \quad x > 0, \tag{1.2}$$

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

The rest of paper is organized as follows. In Section 2, we discuss the maximum likelihood estimator of R. In Section 3, we assume the parameters λ and β are assumed known. In Section 4, Bayesian estimation of R is discussed. Simulation studies is devoted in Section 5.

2. Likelihood Estimation of *R*

Let X and Y are two independent Dagum random variables with parameters $(\lambda, \beta, \theta_1)$ and $(\lambda, \beta, \theta_2)$, respectively. Therefore, the reliability of the system

$$R = P(Y < X)$$

$$= \int_0^\infty \int_0^x \lambda\beta\theta_1 x^{-(\beta+1)}(1 + \lambda x^{-\beta})^{-(\theta_1+1)} \lambda\beta\theta_2 y^{-(\beta+1)}(1 + \lambda y^{-\beta})^{-(\theta_2+1)} dy dx$$

$$= \int_0^\infty \lambda\beta\theta_1 x^{-(\beta+1)}(1 + \lambda x^{-\beta})^{-(\theta_1+1)}[(1 + \lambda y^{-\beta})^{-\theta_2}]_0^x dx$$

$$= \int_0^\infty \lambda\beta\theta_1 x^{-(\beta+1)}(1 + \lambda x^{-\beta})^{-(\theta_1+\theta_2+1)} dx$$

$$= \theta_1[\theta_1 + \theta_2]^{-1}, \tag{2.1}$$

R can be computed using the following definition as well

$$R = \int_0^\infty P(Y < X \setminus X = x) P(X = x)$$

$$= \int_0^\infty \int_0^x \lambda\beta\theta_1 x^{-(\beta+1)}(1 + \lambda x^{-\beta})^{-(\theta_1+1)}(1 + \lambda y^{-\beta})^{-\theta_2} dx$$

$$= \theta_1[\theta_1 + \theta_2]^{-1}.$$

To find the MLE estimator of R, we need first to estimate the shape parameters θ_1 and θ_2. Let $(X_1, X_2, \dots, X_n)$ and $(Y_1, Y_2, \dots, Y_m)$ be two independent random samples from Dagum $(\lambda, \beta, \theta_1)$ and Dagum $(\lambda, \beta, \theta_2)$, respectively. The likelihood function of $\lambda, \beta, \theta_1,$ and θ_2 for the observed samples is

$$L(\lambda, \beta, \theta_1, \theta_2) = \lambda^n \beta^n \theta_1^n \prod_{i=1}^{n} [x_i^{-(\beta+1)}(1 + \lambda x_i^{-\beta})^{-(\theta_1+1)}]$$

$$\times \lambda^m \beta^m \theta_2^m \prod_{i=1}^{m} [y_j^{-(\beta+1)}(1 + \lambda y_j^{-\beta})^{-(\theta_2+1)}].$$

Therefore, the log-likelihood function of $\lambda, \beta, \theta_1,$ and θ_2 as follows:

$$\log L = n \log \lambda + n \log \beta + n \log \theta_1 - (\beta + 1)\sum_{i=1}^{n} \log x_i - (\theta_1 + 1)\sum_{i=1}^{n} \log(1 + \lambda x_i^{-\beta})$$

$$+ m \log \lambda + m \log \beta + m \log \theta_2 - (\beta + 1)\sum_{j=1}^{m} \log y_j - (\theta_2 + 1)\sum_{i=1}^{n} \log(1 + \lambda y_j^{-\beta})$$

$$= (n + m) \log \lambda + (n + m) \log \beta + n \log \theta_1 + m \log \theta_2$$

$$- (\beta + 1)\left[\sum_{i=1}^{n} \log x_i - \sum_{j=1}^{m} \log y_j\right] - (\theta_1 + 1)\sum_{i=1}^{n} \log(1 + \lambda x_i^{-\beta})$$

$$- (\theta_2 + 1)\sum_{j=1}^{n} \log(1 + \lambda y_j^{-\beta}). \tag{2.2}$$

The estimators $\hat{\lambda}, \hat{\beta}, \hat{\theta}_1,$ and $\hat{\theta}_2$ of the parameters $\lambda, \beta, \theta_1,$ and θ_2, respectively, can be then obtained as the solution likelihood equations

$$\frac{\partial \log L}{\partial \lambda} = \frac{(n+m)}{\lambda} - (\theta_1 + 1)\sum_{i=1}^{n} \frac{x_i^{-\beta}}{(1 + \lambda x_i^{-\beta})} - (\theta_2 + 1)\sum_{j=1}^{m} \frac{y_j^{-\beta}}{(1 + \lambda y_j^{-\beta})} = 0, \tag{2.3}$$

$$\frac{\partial \log L}{\partial \beta} = \frac{(n+m)}{\beta} - \left[\sum_{i=1}^{n} \log x_i - \sum_{j=1}^{m} \log y_j\right] + (\theta_1 + 1)\sum_{i=1}^{n} \frac{\lambda x_i^{-\beta} \log x_i}{(1 + \lambda x_i^{-\beta})}$$

$$- (\theta_2 + 1)\sum_{j=1}^{m} \frac{\lambda y_j^{-\beta} \log y_j}{(1 + \lambda y_j^{-\beta})} = 0, \tag{2.4}$$

$$\frac{\partial \log L}{\partial \theta_1} = \frac{n}{\theta_1} - \sum_{i=1}^{n} \log(1 + \lambda x_i^{-\beta}) = 0, \tag{2.5}$$

$$\frac{\partial \log L}{\partial \theta_2} = \frac{m}{\theta_2} - \sum_{j=1}^{m} \log(1 + \lambda y_j^{-\beta}) = 0. \tag{2.6}$$

From (2.5) and (2.6), we have

$$\hat{\theta}_1 = n[\sum_{i=1}^{n} \log(1 + \lambda x_i^{-\beta})]^{-1},$$

and

$$\hat{\theta}_2 = m[\sum_{j=1}^{m} \log(1 + \lambda y_j^{-\beta})]^{-1}. \tag{2.7}$$

Substituting (2.7) into (2.3) and (2.4), we have

$$\frac{(n+m)}{\lambda} - [1 + \frac{n}{\sum_{i=1}^{n} \log(1+\lambda x_i^{-\beta})}] \sum_{i=1}^{n} \frac{x_i^{-\beta}}{(1+\lambda x_i^{-\beta})}$$

$$- [1 + \frac{m}{\sum_{j=1}^{m} \log(1+\lambda y_j^{-\beta})}] \sum_{j=1}^{m} \frac{y_j^{-\beta}}{(1+\lambda y_j^{-\beta})} = 0, \tag{2.8}$$

$$\frac{(n+m)}{\beta} - [\sum_{i=1}^{n} \log x_i - \sum_{j=1}^{m} \log y_j] + [1 + \frac{n}{\sum_{i=1}^{n} \log(1+\lambda x_i^{-\beta})}] \sum_{i=1}^{n} \frac{\lambda x_i^{-\beta} \log x_i}{(1+\lambda x_i^{-\beta})}$$

$$- [1 + \frac{m}{\sum_{j=1}^{m} \log(1+\lambda y_j^{-\beta})}] \sum_{j=1}^{m} \frac{\lambda y_j^{-\beta} \log y_j}{(1+\lambda y_j^{-\beta})} = 0. \tag{2.9}$$

Thus, to derive the MLE estimators of λ, β, θ_1 and θ_2 we have to solve the two nonlinear Equations (2.8) and (2.9) with respect to λ and β, and then we can use (2.7) with respect to $\hat{\lambda}$ and $\hat{\beta}$ to obtain the MLE of θ_1 and θ_2 as

$$\hat{\theta}_1 = n[\sum_{i=1}^{n} \log(1+\hat{\lambda} x_i^{-\hat{\beta}})]^{-1}, \text{ and } \hat{\theta}_2 = m[\sum_{j=1}^{m} \log(1+\hat{\lambda} y_j^{-\hat{\beta}})]^{-1}. \tag{2.10}$$

Once we have obtained the estimators of θ_1 and θ_2, the MLE of R becomes

$$\hat{R} = \hat{\theta}_1 [\hat{\theta}_1 + \hat{\theta}_2]^{-1}. \tag{2.11}$$

3. The Maximum Likelihood Estimator of R if λ and β are Known

In this section, we consider the problem of estimation of R when both λ and β are known. Without loss of generality, we can assume that $\lambda = \beta = 1$, i.e., we assume that the independent samples $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_m)$ are drawn from Dagum $(1, 1, \theta_1)$ and Dagum $(1, 1, \theta_2)$, respectively. Based on that the MLE of R and its distributional properties are

$$\hat{\theta}_1 = n[\sum_{i=1}^{n} \log(1 + x_i^{-1})]^{-1}, \quad \text{and} \quad \hat{\theta}_2 = m[\sum_{j=1}^{m} \log(1 + y_j^{-1})]^{-1}. \quad (3.1)$$

$$\begin{aligned}\hat{\theta}_1 + \hat{\theta}_2 &= n[\sum_{i=1}^{n} \log(1 + x_i^{-1})]^{-1} + m[\sum_{j=1}^{m} \log(1 + y_j^{-1})]^{-1} \\ &= [n\sum_{j=1}^{m} \log(1 + y_j^{-1}) + m\sum_{i=1}^{n} \log(1 + x_i^{-1})] \\ &\quad \times [\sum_{i=1}^{n} \log(1 + x_i^{-1}) \sum_{j=1}^{m} \log(1 + y_j^{-1})]^{-1}, \end{aligned} \quad (3.2)$$

and

$$\hat{R} = [n\sum_{j=1}^{m} \log(1 + y_j^{-1})][n\sum_{j=1}^{m} \log(1 + y_j^{-1}) + m\sum_{i=1}^{n} \log(1 + x_i^{-1})]^{-1}. \quad (3.3)$$

4. Simulation Study

In this section, we present a simulation study to see the performance of an estimator R at different values for parameters.

n	m	R	R(est)	MSE
5	5	0.5	0.508	0.024
		0.33	0.348	0.02
		0.25	0.269	0.016
		0.571	0.5655	0.023
		0.625	0.6144	0.022
		0.667	0.6528	0.021
		0.75	0.7314	0.017
5	10	0.5	0.508	0.018
		0.33	0.352	0.016
		0.25	0.271	0.013
		0.571	0.574	0.017
		0.625	0.624	0.016
		0.667	0.663	0.015
		0.75	0.743	0.011
10	5	0.5	0.495	0.018
		0.33	0.333	0.014
		0.25	0.254	0.01
		0.571	0.565	0.17
		0.625	0.615	0.017
		0.667	0.648	0.016
		0.75	0.735	0.013
10	10	0.5	0.504	0.011
		0.33	0.336	0.009
		0.25	0.259	0.007
		0.571	0.564	0.012
		0.625	0.619	0.011
		0.667	0.658	0.009
		0.75	0.741	0.006

n	m	R	R(est)	MSE
10	15	0.5	0.4988	0.011
		0.33	0.348	0.009
		0.25	0.258	0.006
		0.571	0.574	0.01
		0.625	0.623	0.009
		0.667	0.664	0.008
		0.75	0.742	0.006
15	10	0.5	0.499	0.01
		0.33	0.334	0.008
		0.25	0.255	0.006
		0.571	0.566	0.009
		0.625	0.62	0.009
		0.667	0.652	0.009
		0.75	0.74	0.009
15	15	0.5	0.501	0.008
		0.33	0.336	0.006
		0.25	0.255	0.006
		0.571	0.571	0.008
		0.625	0.621	0.007
		0.667	0.666	0.006
		0.75	0.746	0.005
25	25	0.5	0.496	0.004
		0.33	0.331	0.003
		0.25	0.253	0.002
		0.571	0.568	0.004
		0.625	0.624	0.004
		0.667	0.665	0.003
		0.75	0.747	0.002

n	m	R	R(est)	MSE
50	50	0.5	0.5001	0.001
		0.33	0.33	0.001
		0.25	0.25	0.002
		0.571	0.568	0.002
		0.625	0.622	0.002
		0.667	0.668	0.001
		0.75	0.746	0.001

5. Conclusion

In this paper, we have considered the estimation of the probability $P\{X < Y\}$ when X and Y are two independent random variables from Dagum distribution. We found maximum likelihood estimator and used its asymptotic distribution to construct confidence intervals. We performed a simulation study to show the consistency property of the MLE estimators of R.

References

[1] M. Ali and J. Woo, Inference on reliability $P(Y < X)$ in a p-dimensional Rayleigh distribution, Mathematical and Computer Modeling 42 (2005a), 367-373.

[2] M. Ali and J. Woo, Inference on $P(Y < X)$ in a Pareto distribution, Journal of Modern Applied Statistical Methods 4 (2005b), 583-586.

[3] M. Ali, M. Pal and J. Woo, Inference on $P(Y < X)$ in generalized uniform distributions, Calcutta Statistical Association Bulletin 57 (2005), 35-48.

[4] M. Ali, M. Pal and J. Woo, Estimation of $\Pr(Y < X)$ when X and Y belong to different distribution families, Journal of Probability and Statistical Science 8 (2010), 35-48.

[5] M. Ali, J. Woo and M. Pal, Inference on reliability $P(Y < X)$ in two-parameter exponential distributions, International Journal of Statistical Sciences 3 (2004), 119-125.

[6] E. A. Amin, Bayesian and non-Bayesian estimation of $P(Y < X)$ from type I generalized logistic distribution based on lower record values, Journal of Applied Sciences Research 8(1) (2012), 118-125.

[7] A. P. Basu, Estimates of reliability for some distributions useful in reliability, Technometrics 6 (1964), 215-219.

[8] M. A. Beg, On the estimation of $P(Y < X)$ for two-parameter exponential distribution, Metrika 27 (1980), 29-34.

[9] C. Kleiber and S. Kotz, Statistical Size Distribution in Economics and Actuarial Sciences, John Wiley & Sons, Inc., Hoboken, NJ, 2003.

[10] C. Kleiber, A guide to the Dagum distribution, modeling income distributions and Lorenz curves economic studies in equality, Social Exclusion and Well-Being 5 (2008), 97-117.

[11] C. Quintano and A. D'Agostino, Studying inequality in income distribution of single-person households in four developed countries, Rev. Income Wealth 52(4) (2006), 525-546.

[12] Dagum ditribution, Comm. Statist. Simulation Comput. 36(6) (2007), 1187-1199.

[13] C. Dagum, A model of income distribution and the conditions of existence of moments of finite order, Bulletin of the International Statistical Institute, 46 (Proceedings of the 40th Session of the ISI, Warsaw, Contributed Papers), (1975), 199-205.

[14] C. Dagum, A new model of personal income distribution: Specification and estimation, Economie Appliqufiee 30 (1977), 413-437.

[15] C. Dagum, The generation and distribution of income, the Lorenz curve and the Gini ratio, Economie Appliqufiee 33 (1980a), 327-367.

[16] C. Dagum, Generating systems and properties of income distribution models, Metron. 38 (1980b), 3-26.

[17] F. Downtown, The estimation of $\mathrm{Pr}fY < Xg$ in the normal case, Technometrics 15 (1973), 551-558.

[18] F. Domma, S. Giordano and M. Zenga, The Fisher information matrix in doubly censored data from the Dagum distribution, Working Paper No. 8, Department of Economics and Statistics, University of Calabria, Italy, 2009.

[19] F. Domma, G. Latorre and M. Zenga, Reliability studies of the Dagum distribution, Working Paper No. 207, Department of Quantitative Methods for Economics and Business, University of Milan - Bicocca, Italy, 2011.

[20] Francisco J. Rubio and Mark F. J. Steel, Bayesian inference for $P(X < Y)$ using asymmetric dependent distributions, Journal of Bayesian Analysis 7(3) (2012), 771-792.

[21] V. V. Ivshin, Unbiased estimators of $P(X < Y)$ and their variances in the case of uniform and two-parameter exponential distributions, Journal of Mathematical Sciences 81 (1996), 2790-2793.

[22] K. Iwase, On UMVU estimators of $\Pr(Y < X)$ in the 2-parameter exponential case, Memoirs of the Faculty of Engineering, Hiroshima University, 9 (1987), 21-24.

[23] G. D. Kelley, J. A. Kelley and W. R. Suchany, Efficient estimation of $P(Y < X)$ in the exponential case, Technimetrics 18 (1976), 359-360.

[24] M. Masoom Ali, Manisha Pal and Jungsoo Woo, Estimation of $P(Y < X)$ in a four-parameter generalized gamma distribution, Austrian Journal of Statistics 41(3) (2012), 197-210.

[25] J. I. McCool, Inference on $P(X < Y)$ in the Weibull case, Communications in Statistics: Simulation and Computation 20 (1991), 129-148.

[26] M. Pal, M. Ali and J. Woo, Estimation and testing of $P(Y < X)$ in two parameter exponential distributions, Statistics 39 (2005), 415-428.

[27] M. Z. Raqab and D. Kundu, Comparison of different estimators of $P(Y < X)$ for a scaled Burr Type X distribution, Communications in Statistics: Simulation and Computation 34 (2005), 465-483.

[28] M. Z. Raqab, M. T. Madi and D. Kundu, Estimation of $P(Y < X)$ for a 3-parameter generalized exponential distribution, Communications in Statistics: Theory and Methods 37 (2008), 2854-2864.

[29] S. Rezaei, R. Tahmasbi and M. Mahmoodi, Estimation of $P(Y < X)$ for generalized Pareto distribution, Journal of Statistical Planning and Inference 140 (2010), 480-494.

[30] Studies in Inequality, Social Exclusion and Well-Being, Vol. 5, C. Duangkamon, Editor, Springer, New York, NY, 2008.

[31] H. Tong, A note on the estimation of $\Pr(Y < X)$ in the exponential case, Technometrics 16 (1974), 625.

[32] H. Tong, On the estimation of $\Pr(Y < X)$ for exponential families, IEEE Transactions on Reliability R-26 (1977), 54-56.

[33] A. Wong, Interval estimation of $P(Y < X)$ for generalized Pareto distribution, Journal of Statistical Planning and Inference 142 (2012), 601-607.

■