STATISTICA, anno LXXVI, n. 1, 2016

ON A GENERALISATION OF UNIFORM DISTRIBUTION AND ITS PROPERTIES

K. Jayakumar
Department of Statistics, University of Calicut, Kerala - 673 635, India
Kothamangalth Krishnan Sankaran [1]
Department of Statistics, S.N.College, Alathur, Kerala - 678 682, India

1. Introduction

Many researchers are interested in search that introduces new families of distributions or generalized some of the presented distributions which can be used to describe the lifetimes of some devices or to describe sets of real data. Exponential, Rayleigh, Weibull and linear failure rate are some of the important distributions widely used in reliability theory and survival analysis. However, these distributions have a limited range of applicability and cannot represent all situations found in application. For example, although the exponential distribution is often described as flexible, its hazard function is constant. The limitations of standard distributions often arouse the interest of researchers in finding new distributions by extending ones. The procedure of expanding a family of distributions for added flexibility or constructing covariates models is a well known technique in the literature.

Uniform distribution is regarded to the simplest probability model and is related to all distributions by the fact that the cumulative distribution function, taken as a random variable, follows Uniform distribution over (0,1) and this result is basic to the inverse method of random variable generation. This distribution is also applied to determine power functions of tests of randomness. It is also applied in a power comparison of tests of non random clustering. There are also numerous applications in nonparametric inference, such as Kolmogrov-Smirnov test for goodness of fit. The uses of Uniform distribution to represent the distribution of roundoff errors and in connection with the probability integral transformations are also well known.

1.1. Marshall-Olkin family of distribution

Marshall and Olkin(1997) introduced a new family of distribution by adding a parameter to a family of distribution. They started with a survival function $\bar{F}(x)$

[1] Corresponding Author e-mail: snsankaran08@gmail.com

and considered a family of survival functions given by

$$\bar{G}(x) = \frac{\alpha \bar{F}(x)}{F(x) + \alpha \bar{F}(x)}.$$

Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables with survival function $\bar{F}(x)$. Let N be a geometric random variable with probability mass function $P(N = n) = \alpha(1-\alpha)^{n-1}$, for n = 1,2,... and $0 < \alpha < 1$. Then $U_N = min(X_1, X_2, ..., X_N)$ has the survival function given by above equation. If $\alpha > 1$ and N is a geometric random variable with probability mass function $P(N = n) = \frac{1}{\alpha}(1-\frac{1}{\alpha})^{n-1}$, for n = 1,2,... then $V_N = max(X_1, X_2, ..., X_N)$ also has the survival function as above.
In the past, many authors have proposed various univariate distributions belonging to the family of Marshall-Olkin distributions. Also Jayakumar and Thomas (2008) proposed a new generalization of the family of Marshall-Olkin distribution as

$$\bar{G}(x;\alpha,\gamma) = \left[\frac{\alpha \bar{F}(x)}{1-(1-\alpha)\bar{F}(x)}\right]^{\gamma}, for \quad \alpha > 0, \gamma > 0, x \in \Re$$

1.2. Truncated Negative Binomial distribution

Nadarajah,et,al. (2013) introduced a new family of distributions as follows:
Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables with survival function $\bar{F}(x)$. Let N be a truncated negative binomial random variable with parameters $\alpha \in (0,1)$ and $\theta > 0$. That is,

$$P(N = n) = \frac{\alpha^{\theta}}{1-\alpha^{\theta}} \begin{pmatrix} \theta + n - 1 \\ \theta - 1 \end{pmatrix} (1-\alpha)^n \quad for \quad n = 1, 2, ...$$

Consider $U_N = min(X_1, X_2, ..., X_N)$. We have

$$\bar{G}_{U_N}(x) = \frac{\alpha^{\theta}}{1-\alpha^{\theta}} \sum_{n=1}^{\infty} \begin{pmatrix} \theta + n - 1 \\ \theta - 1 \end{pmatrix} ((1-\alpha)\bar{F}(x))^n$$

$$\bar{G}_{U_N}(x) = \frac{\alpha^{\theta}}{1-\alpha^{\theta}} [(F(x) + \alpha \bar{F}(x))^{-\theta} - 1] \qquad (1)$$

Similarly if $\alpha > 1$ and N is a truncated negative binomial random variable with parameters $\frac{1}{\alpha}$ and $\theta > 0$, then $V_N = max(X_1, X_2, ..., X_N)$ also has the survival function given by (1). This implies that we can consider a new family of distributions given by the survival function

$$\bar{G}(x;\alpha,\theta) = \frac{\alpha^{\theta}}{1-\alpha^{\theta}} [(F(x) + \alpha \bar{F}(x))^{-\theta} - 1]$$

for $\alpha > 0, \theta > 0$ and $x \in \Re$. Note that if $\alpha \to 1$ then $\bar{G}(x;\alpha,\theta) \to \bar{F}(x)$. The family of distributions (1) is a generalization of the family of Marshall-Olkin distributions. Namely, if $\theta = 1$, then (1) reduces to the family of Marshall-Olkin distributions.

Suppose the failure times of a device are observed. Every time a failure occurs the device is repaired to resume function. Suppose also that the device is deemed no longer usable when a failure occurs that exceeds a certain level of severity. Let $X_1, X_2, ...$ denote the failure time and let N denote the number of failures. Then U_N will represent the time to the first failure of the device and V_N will represent the lifetime of the device. So (1) could be used to model both the time to the first failure and the lifetime.

In Section 2 we introduce the Generalised Uniform distribution and study its properties.Concluding remarks are presented in Section 3.

2. A new family of Uniform distribution

In this section, we set $\bar{F}(x) = 1 - x, 0 < x < 1$ and introduce a new family of distributions given by the survival function

$$\bar{G}(x;\alpha,\theta) = \frac{\alpha^\theta}{1-\alpha^\theta}[[x(1-\alpha)+\alpha]^{-\theta} - 1], \quad \theta > 0, \alpha > 0. \tag{2}$$

Therefore, the distribution function is given by

$$G(x;\alpha,\theta) = \frac{1-\alpha^\theta[x(1-\alpha)+\alpha]^{-\theta}}{1-\alpha^\theta} \tag{3}$$

The corresponding probability density function is given by

$$g(x;\alpha,\theta) = \frac{(1-\alpha)\theta\alpha^\theta}{(1-\alpha^\theta)[x(1-\alpha)+\alpha]^{\theta+1}} \tag{4}$$

for $0 < x < 1, \alpha > 0$, and $\theta > 0$. We refer to this new distribution as the Genralised Uniform distribution with parameters α and θ . We write it as GUD(α, θ).
Remark I: If $\theta = 1$, we obtain Marshall-Olkin Extended Uniform distribution introduced by Jose and Krsihna (2011).
Remark II: When $\theta = 1$ and $\alpha \to 1$, GUD reduces to standard Unifrom distribution.
GUD(α, θ) random variable can be simulated using

$$X = \bar{F}^{-1}\left(\frac{1-\alpha[(1-\alpha^\theta)Y+\alpha^\theta]^{\frac{-1}{\theta}}}{1-\alpha}\right)$$

for $Y \sim$ Uniform (0,1). Therefore,

$$X = \frac{\alpha}{1-\alpha}\left([(1-\alpha^\theta)Y+\alpha^\theta]^{\frac{-1}{\theta}} - 1\right).$$

2.1. Shapes of the distribution and density function

If X follows $GUD(\alpha, \theta)$, then

$$G(x;\alpha,\theta) = \frac{1-\alpha^\theta[x(1-\alpha)+\alpha]^{-\theta}}{1-\alpha^\theta}$$

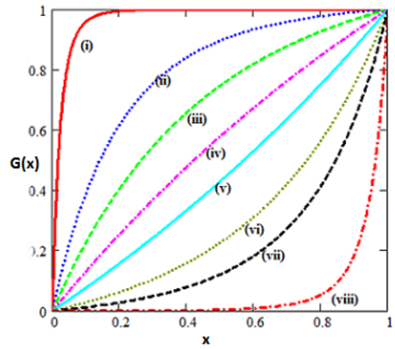

Figure 1 – The d.f of the GUD for different values of α when $\theta = 5$;(i) $\alpha = 0.1$ (ii)$\alpha = 0.5$ (iii) $\alpha = 0.7$ (iv) $\alpha = 0.9$ (v) $\alpha = 1.1$ (vi) $\alpha = 1.5$ (vii) $\alpha = 1.9$(viii) $\alpha = 5$.

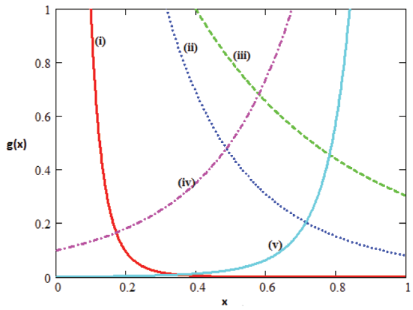

Figure 2 – The probability density function of the GUD for different values of α when $\theta = 5$ (i) $\alpha = 0.1$ (ii)$\alpha = 0.5$ (iii) $\alpha = 0.7$ (iv) $\alpha = 1.9$(v) $\alpha = 5$.

For different values of α when $\theta = 5$, the structure of the distribution function is given in Figure 1.

In order to derive the shape properties of the probability density function, we consider the function

$$(log\ g)' = \frac{g'(x)}{g(x)} = -\frac{(1-\alpha)(\theta+1)}{x(1-\alpha)+\alpha}$$

Let $s(x) = \frac{(1-\alpha)(\theta+1)}{x(1-\alpha)+\alpha}$
i)If $\alpha \in (0,1)$, then the function $s(x)$ is positive and this implies that g is a decreasing function with $g(0) = \frac{(1-\alpha)\theta}{\alpha(1-\alpha^\theta)}$ and $g(1) = \frac{\alpha^\theta(1-\alpha)\theta}{1-\alpha^\theta}$.
ii) If $\alpha > 1$, then the function $s(x)$ is negative and this implies that g is an increasing function with $g(0) = \frac{(\alpha-1)\theta}{\alpha(\alpha-1)}$ and $g(1) = \frac{\alpha^\theta(\alpha-1)\theta}{\alpha^\theta-1}$.

Some possible shapes of the probability density function $g(x;\alpha,5)$ are given in Figure 2.

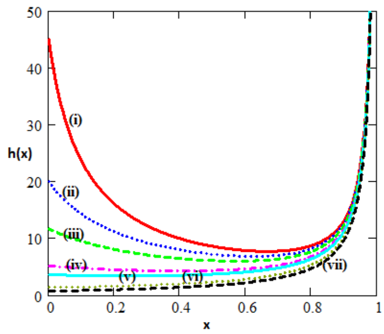

Figure 3 – The hazard function of the GUD for different values of α when $\theta = 5$ (i) $\alpha = 0.1$ (ii)$\alpha = 0.2$ (iii) $\alpha = 0.3$ (iv) $\alpha = 0.5$ (v) $\alpha = 0.6$ (vi) $\alpha = 0.9$ (vii) $\alpha = 1.1$

.

2.2. Hazard rate function

The hazard function of a random variable X with density g(x) and a cumulative distribution function G(x) is given by

$$h(x) = \frac{g(x)}{\bar{G}(x)} = \frac{\theta(1-\alpha)}{[x(1-\alpha)+\alpha][1-(x(1-\alpha)+\alpha)^{\theta}]}$$

Shapes of the hazard function $h(x;\alpha)$ for $\theta = 5$ are presented in Figure 3.

For $\alpha \leq 0.6$, hazard rate is initially decreasing and there exists an interval where it changes to be IFR. For $\alpha > 0.6$, the hazard function is evidently IFR. Let us consider the reverse hazard rate function. The reverse hazard rate function is useful in constructing the information matrix and in estimating the survival function for censored data. The reverse hazard function of Generalised Uniform distribution is given by

$$r(x) = \frac{g(x)}{G(x)} = \frac{(1-\alpha)\theta\alpha^{\theta}}{[x(1-\alpha)+\alpha][(x(1-\alpha)+\alpha)^{\theta}-\alpha^{\theta}]}$$

The reverse hazard rate function decreases on $(0,1)$ with r(0)=∞ and $r(1) = \frac{(1-\alpha)\theta\alpha^{\theta}}{1-\alpha^{\theta}}$.

2.3. Moments

If X has the GUD(α,θ) distribution, then the r^{th} order moment is given by

$$\begin{aligned} E(X^r) &= \int_0^1 x^r \frac{(1-\alpha)\theta\alpha^{\theta}}{(1-\alpha^{\theta})[x(1-\alpha)+\alpha]^{\theta+1}}dx \\ &= \frac{(1-\alpha)\theta\alpha^{\theta}}{(1-\alpha^{\theta})}\int_0^1 \frac{x^r}{[x(1-\alpha)+\alpha]^{\theta+1}}dx. \end{aligned}$$

By equation 2.2.5.2 Prudinikov et,al. (1986),

$$\int_a^b \frac{(x-a)^{(\alpha-1)}}{(cx+d)^{\alpha+n+1}}dx = \frac{(b-a)^{\alpha}}{(ac+d)(bc+d)^{\alpha}} \sum_{k=0}^{n} \binom{n}{k} \frac{B(\alpha+k, n-k+1)}{(bc+d)^k (ac+d)^{n-k}}$$

where $B(a,b) = \frac{\Gamma a \Gamma b}{\Gamma(a+b)}$

$$\begin{aligned} E(X^r) &= \frac{(1-\alpha)\theta\alpha^{\theta})}{\alpha(1-\alpha^{\theta}} \sum_{k=0}^{\theta-r-1} \binom{\theta-r-1}{k} \frac{B(r+k+1, \theta-r-k)}{\alpha^{n-k}} \\ &= \frac{(1-\alpha)\theta\alpha^{\theta-1}}{(1-\alpha^{\theta})} \sum_{k=0}^{\theta-r-1} \binom{\theta-r-1}{k} \frac{\Gamma(r+k+1)\Gamma(\theta-r-k)}{\Gamma(\theta+1)\alpha^{n-k}}. \quad (5) \end{aligned}$$

Using (5), we get the mean and the variance of a random variable X with $GUD(\alpha, \theta)$ are respectively as

$$\mu_1^{'} = \frac{\alpha^{\theta}}{1-\alpha^{\theta}} \left[\frac{\alpha(1+\alpha^{-\theta}) + \theta(1-\alpha)}{(1-\alpha)(1-\theta)} \right], \quad and$$

$$\mu_2 = \frac{\alpha^{\theta}}{(1-\alpha^{\theta})} \left[-1 + 2\left(\frac{1}{(1-\alpha)(1-\theta)} - \frac{(1-\alpha^{-\theta+2})}{(1-\alpha)^2(1-\theta)(2-\theta)} \right) - \frac{\alpha^{\theta}}{1-\alpha^{\theta}} \left(\frac{\alpha(1+\alpha^{-\theta}) + \theta(1-\alpha)}{(1-\alpha)(1-\theta)} \right)^2 \right].$$

The q^{th} quantile of a random variable X with $GUD(\alpha, \theta)$ is given by

$$x_q = G^{-1}(q) = \frac{\alpha}{1-\alpha} \left[\frac{1}{[1-q(1-\alpha)^{\theta}]^{\frac{1}{\theta}}} - 1 \right], 0 \le q \le 1,$$

where $G^{-1}(.)$ is the inverse distribution function. In particular, the median of $GUD(\alpha, \theta)$ is given by

$$Median(X) = \frac{\alpha}{1-\alpha} \left[\frac{1}{[1-\frac{1}{2}(1-\alpha)^{\theta}]^{\frac{1}{\theta}}} - 1 \right].$$

2.4. Order Statistics

Assume $X_1, X_2, ..., X_n$ are independent random variables having the $GUD(\alpha, \theta)$ distribution. Let $X_{i:n}$ denote the i^{th}order statistic. The probability density function of $X_{i:n}$ is

$$\begin{aligned} g_{i:n}(x; \alpha, \theta) &= \frac{n!}{(i-1)!(n-1)!} g(x; \alpha, \theta) G^{i-1}(x; \alpha, \theta) \bar{G}^{n-i}(x; \alpha, \theta) \\ &= \frac{n!}{(i-1)!(n-i)!} \left[\frac{(1-\alpha)\theta\alpha^{\theta}}{(1-\alpha^{\theta})[x(1-\alpha)+\alpha]^{\theta+1}} \right] \\ &\quad \left[1 - \frac{\alpha^{\theta}}{(1-\alpha^{\theta})} [(x(1-\alpha)+\alpha)^{-\theta} - 1] \right]^{i-1} \\ &\quad \left[\frac{\alpha^{\theta}}{(1-\alpha^{\theta})} [(x(1-\alpha)+\alpha)^{-\theta} - 1] \right]^{n-i}. \end{aligned}$$

2.5. Renyi and Shannon entropies

An entropy is a measure of variation or uncertainty. The Renyi entropy of a random variable with probability density function g(.) is defined as

$$I_R(\gamma) = \frac{1}{1-\gamma} log \int_0^\infty g^\gamma(x)dx, \gamma > 0, \gamma \neq 1.$$

The Shannon entropy of a random variable X is defined by E[-log g(X)]. It is a particular case of the Renyi entropy for $\gamma \uparrow 1$. We have

$$\begin{aligned}\int_0^\infty g^\gamma(x)dx &= \int_0^1 \left[\frac{(1-\alpha)\theta\alpha^\theta}{(1-\alpha^\theta)(x(1-\alpha)+\alpha)^{\theta+1}}\right]^\gamma dx \\ &= \left[\frac{(1-\alpha)\theta\alpha^\theta}{1-\alpha^\theta}\right]^\gamma \left[\frac{\alpha^{\gamma(\theta+1)+1}-1}{(1-\gamma(\theta+1))(1-\alpha)\alpha^{\gamma(\theta+1)+1}}\right]\end{aligned}$$

So the Renyi entropy is

$$I_R(\gamma) = \frac{1}{1-\gamma} log \left(\left[\frac{(1-\alpha)\theta\alpha^\theta}{1-\alpha^\theta}\right]^\gamma \left[\frac{\alpha^{\gamma(\theta+1)+1}-1}{(1-\gamma(\theta+1))(1-\alpha)\alpha^{\gamma(\theta+1)+1}}\right]\right)$$

Now consider the Shannon entropy. From its definition, we have

$$\begin{aligned}E[-log\ g(X)] &= E\left[log\frac{1-\alpha^\theta}{(1-\alpha)\theta\alpha^\theta} + (\theta+1)log[x(1-\alpha)+\alpha)]\right] \\ &= log\frac{1-\alpha^\theta}{(1-\alpha)\theta\alpha^\theta} + (\theta+1)E(log[x(1-\alpha)+\alpha]) \\ E(log[X(1-\alpha)+\alpha]) &= \frac{(1-\alpha)\theta\alpha^\theta}{1-\alpha^\theta}\int_0^1 \frac{log[x(1-\alpha)+\alpha]}{[x(1-\alpha)+\alpha]^{\theta+1}}dx \\ &= \frac{\theta\alpha^\theta}{1-\alpha^\theta}\left[\alpha^{-\theta}log\alpha + \frac{(\alpha^{-\theta}-1)}{\theta^2}\right]\end{aligned}$$

Hence

$$E[-log\ g(X)] = log\frac{1-\alpha^\theta}{(1-\alpha)\theta\alpha^\theta} + \frac{\theta(\theta+1)\alpha^\theta}{1-\alpha^\theta}\left[\alpha^{-\theta}log\alpha + \frac{(\alpha^{-\theta}-1)}{\theta^2}\right]$$

2.6. Estimation

Since the moments of an GUD random variable cannot be obtained in closed form, we consider estimation of the unknown parameters by the method of maximum likelihood. For a given sample $(x_1, x_2, ..., x_n)$, the log-likelihood function is given by

$$\begin{aligned}log\ L(x;\alpha,\theta) &= nlog\frac{(1-\alpha)\theta\alpha^\theta}{1-\alpha^\theta} + \sum_{i=1}^n log\frac{1}{[x_i(1-\alpha)+\alpha]^{\theta+1}} \\ &= nlog(1-\alpha) + nlog\theta + n\theta log\alpha - nlog(1-\alpha^\theta) \\ &\quad -(\theta+1)\sum_{i=1}^n log[x_i(1-\alpha)+\alpha]\end{aligned}$$

The partial derivatives of the log likelihood function with respect to the parameters are

$$\frac{\partial logL}{\partial \alpha} = \frac{-n}{1-\alpha} + \frac{n\theta}{\alpha} + \frac{n\theta\alpha^{\theta-1}}{1-\alpha^{\theta}} - (\theta+1)\sum_{i=1}^{n} \frac{(1-x_i)}{[x_i(1-\alpha)+\alpha]},$$

$$\frac{\partial logL}{\partial \theta} = \frac{n}{\theta} + \frac{nlog\alpha}{1-\alpha^{\theta}} - \sum_{i=1}^{n} log[x_i(1-\alpha)+\alpha].$$

The maximum likelihood estimates can be obtained numerically solving the equation $\frac{\partial logL}{\partial \alpha} = 0$ and $\frac{\partial logL}{\partial \theta} = 0$. We can use, for example, the function *nlm* from the programming language **R**.

The second derivatives of the log likelihood function of GUD with respect to α and θ are given by

$$\frac{\partial^2 logL}{\partial \alpha^2} = \frac{n}{(1-\alpha)^2} - \frac{n\theta}{\alpha^2} + \frac{n\theta\alpha^{\theta-2}(1+\theta-\alpha^{\theta})}{(1-\alpha^{\theta})^2} + (\theta+1)\sum_{i=1}^{n} \frac{(1-x_i)^2}{[x_i(1-\alpha)+\alpha]^2},$$

$$\frac{\partial^2 logL}{\partial \theta^2} = \frac{-n}{\theta^2} + \frac{n\alpha^{\theta}log^2\alpha}{(1-\alpha^2)^2},$$

$$\frac{\partial^2 logL}{\partial \theta \partial \alpha} = \frac{n}{\alpha} + \frac{n\alpha^{\theta-1}(1-\alpha^{\theta}) + n\theta(\theta-1)\alpha^{\theta-2}(1-\alpha)^{\theta} + n\theta^2\alpha^{2(\theta-1)}}{(1-\alpha^{\theta})^2} - \sum_{i=1}^{n} \frac{(1-x_i)}{[x_i(1-\alpha)+\alpha]}.$$

If we denote the MLE of $\beta = (\alpha, \theta)$ by $\hat{\beta} = (\hat{\alpha}, \hat{\theta})$, then the observed information matrix is given by

$$I(\beta) = -E \begin{bmatrix} \frac{\partial^2 logL}{\partial \alpha^2} & \frac{\partial^2 logL}{\partial \theta \partial \alpha} \\ \frac{\partial^2 logL}{\partial \theta \partial \alpha} & \frac{\partial^2 logL}{\partial \theta^2} \end{bmatrix}$$

and hence the variance covariance matrix would be $I^{-1}(\beta)$. The approximate $(1-\delta)100\%$ confidence intervals for the parameters α and θ are $\hat{\alpha} \pm Z_{\frac{\delta}{2}}\sqrt{V(\hat{\alpha})}$ and $\hat{\theta} \pm Z_{\frac{\delta}{2}}\sqrt{V(\hat{\theta})}$ respectively, where $V(\hat{\alpha})$ and $V(\hat{\theta})$ are the variances of $\hat{\alpha}$ and $\hat{\theta}$, which are given by the diagonal elements of $I^{-1}(\beta)$, and $Z_{\frac{\delta}{2}}$ is the upper $(\frac{\delta}{2})$ percentile of standard normal distribution.

3. Conclusion

In this paper, we have introduced and studied a new family of distribution containing the Uniform as a generalization of the Marshall-Olkin extended uniform distribution studied in Jose and Krishna(2011). As a data analysis point of view GUD are more feasible and tractable. We have derived some properties of the Generalised Uniform distribution such as probability density function, hazard rate function , moments, distribution of order statistics, Renyi entropy and Shannon entropy. Estimation of parameters is done using maximum likelihood.

References

K. Jayakumar, M. Thomas (2008). *On a generalization to Marshall-Olkin scheme and its application to Burr type XII distribution.* Statist. Pap., 49, 421-439.

K. K. Jose, E. Krishna (2011). *Marshall-Olkin extended uniform distribution.* ProbStat Forum, 04, 78-88.

A. W. Marshall, I. Olkin (1997). *A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families.* Biometrika, 84, 641-652.

S. Nadarajah, K. Jayakumar, M. M. Ristic (2013). *A new family of lifetime models.* Journal of Statistical Computation and Simulation, 83, 1389-1404.

A. P. Prudnikov, Y. A. Brychkov, O. I. Marichev (1986). *Integrals and series vol. I.* Gordeon and Breach Sciences, Amsterdam, Netherlands.

Summary

Nadarajah et al.(2013) introduced a family life time models using truncated negative binomial distribution and derived some properties of the family of distributions. It is a generalization of Marshall-Olkin family of distributions. In this paper, we introduce Generalized Uniform Distribution (GUD) using the approach of Nadarajah et al.(2013). The shape properties of density function and hazard function are discussed. The expression for moments, order statistics, entropies are obtained. Estimation procedure is also discussed.The GDU introduced here is a generalization of the Marshall-Olkin extended uniform distribution studied in Jose and Krishna(2011).

Keywords: Distribution of order statistics; Entropy; Marshall-Olkin family of distributions; Maximum likelihood; Random variate generation; Truncated negative binomial distribution.