Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/050110

A Generalized Discrete Uniform Distribution

Prasanth C B[1,*] *and Sandhya E*[2]

[1] Suprabhath, Nanminda, Calicut, Kerala, India-673613.
[2] Department of Statistics, Prajyothi Niketan College, Pudukkad, Thrissur-680 301, India

Received: 28 Sep. 2015, Revised: 10 Jan. 2016, Accepted: 13 Jan. 2016
Published online: 1 Mar. 2016

Abstract: A new family of distributions, viz, Harris Discrete Uniform distribution is introduced. The various characteristics such as of hazard rate, entropy, distribution of minimum of sequence of i.i.d random variables and the relation with the Marshall Olkin Discrete Uniform Distribution are derived. An AR (1) model with this distribution for marginals is considered. The goodness of the distribution is tested with a real data.

Keywords: Discrete uniform distribution, Marshall-Olkin family, Harris family, Failure rate, AR (1) model.

1 Introduction

Adding one or more parameters to a distribution makes it richer and more flexible for modeling data. Marshall and Olkin [3] introduced a new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Jose and Krishna [2] have developed Marshall-Olkin extended uniform distribution. A similar set up in the case of discrete uniform distribution- Marshall-Olkin discrete uniform (MODU) distribution has been developed by Sandhya and Prasanth [5] and this is found to be suitable for discrete data exhibiting clear positive or negative skewness. Sandhya and Prasanth [4] also considered Marshall-Olkin Geometric (MOG) distribution and some characterizations of it. Satheesh *et al.* [10] discussed a generalization of the Marshall-Olkin (MO) Scheme. A new method for adding two extra parameters for discrete uniform distribution with Harris distribution (Sandhya *et al.* [7] and [6]) is introduced here. We call the new distribution as Harris Discrete Uniform (HDU) distribution and study its properties. A problem with discrete models in some cases is that it is difficult to get compact mathematical expressions for even simple descriptive statistics like expectation, variance and maximum likelihood estimates and so on. Here we overcome this difficulty by making use of the numerical analysis based on the software package Mathematica.

If $\overline{F}(x)$ is the survival function (s.f) of a distribution with $F(x)$ as the distribution function (d.f), then by MO method we get another s.f $\overline{G}(x)$, by adding a new parameter θ to it. That is,

$$\overline{G}(x,\theta) = \theta\overline{F}(x)/(1-(1-\theta)\overline{F}(x)), \quad -\infty < x < \infty, \theta > 0. \tag{1.1}$$

Then the corresponding d.f is,

$$G(x,\theta) = 1-\overline{G}(x,\theta) = F(x)/(1-(1-\theta)\overline{F}(x)).$$

Let the variable X be discrete. Now from (1.1) consider the new p.m.f $g(x,\theta)$ as,

$$g(x,\theta) = G(x,\theta) - G(x-1,\theta) = f(x)/\Big[(1-(1-\theta)\overline{F}(x))(1-(1-\theta)\overline{F}(x-1))\Big] \tag{1.2}$$

where $f(x)$ is the p.m.f. corresponding to $F(x)$.
Let $\gamma_G(x)$ be the hazard rate of X, i.e.,

$$\gamma_G(x) = g(x)/\overline{G}(x). \tag{1.3}$$

* Corresponding author e-mail: cbpwarrier@yahoo.co.in

Let X be a discrete random variable (r.v) following uniform distribution with p.m.f $f(x) = 1/a$, $x = 1,2,3\ldots a$, then the distribution function is $F(x) = x/a$ and the s.f is $\overline{F}(x) = (a-x)/a$. By MO method we can form another s.f of MODU distribution by substituting $\overline{F}(x) = (a-x)/a$ in $\overline{G}(x,\theta)$, in (1.1) and we get,

$$\overline{G}(x,\theta) = \theta(a-x)/[a\theta + (1-\theta)x] \quad 0 < \theta < \infty. \tag{1.4}$$

We write $X \sim MODU(a,\theta)$ for a r.v. with $\overline{G}(x,\theta)$ given in (1.4).

Harris [1] considered a probability generating function (p.g.f.) defined by,

$$P(s) = s/(m-(m-1)s^k)^{1/k},\ k \in N$$

where N is the set of positive integers and $m > 1$. (1.5)

He introduced this p.g.f. while considering a simple mathematical model for a branching process. From this p.g.f. we can see that this is a generalization of the geometric distribution on $\{1,2,3\ldots\}$ to which it reduces when $k = 1$ and its probability carrying integers are k integers apart or the probabilities are concentrated on the points $1, 1+k, 1+2k\ldots$. It is also true that these probabilities coincide with that of the negative binomial distribution on $\{0,1,2,\ldots\}$ with parameters $1/m$ and $1/k$. The distribution corresponding to the p.g.f. is denoted by $H_1(m,k,1/k)$. In the notation the suffix 1 suggests that the support of the distribution starts from unity, m determines the probabilities, k implies that the probability carrying integers of the distribution are k integers apart and $1/k$ is the exponent. The role played by this distribution in schemes with random (N) sample sizes (random-sums or N-sums and random-extremes or N-extremes) in general and time series models in particular where N is a non-negative integer-valued r.v is discussed by Satheesh *et al.* [9]. Satheesh and Sandhya [8] have discussed this p.g.f. in the context of N-sums and N-extremes. The p.g.f. of Harris distribution has been widely used in summation schemes. Satheesh *et al.* [9] have considered a generalization of geometric sums and its stability by studying distributions that are stable under summation with respect to $H_1(m,k,1/k)$ law. They discussed the stability of N-sums of r.v.s when N is Harris.

Proceeding as the MO set up mentioned in (1.1), let $F(x)$ be the d.f of a discrete random variable X, $f(x)$ be its p.m.f. and let $\overline{F}(x)$ be the s.f of a distribution, then we can write the s.f of a new distribution by substituting this $\overline{F}(x)$ in (1.5) and we get a new Harris family of s.f by adding a new parameter θ as,

$$\overline{H}(x,\theta,k) = \left\{\theta\overline{F}^k(x)/[1-(1-\theta)(\overline{F}^k(x))]\right\}^{1/k} .0 < \theta < \infty, k \in N \tag{1.6}$$

where N is the set of positive integers.

2 Harris Discrete Uniform Distribution

Let X follows the discrete uniform distribution with d.f, $F(x) = x/a$ and $\overline{F}(x) = 1-(x/a), x = 1,2\ldots a$. Then, by adding new parameters θ and k, by (1.6) we get the s.f of the new distribution as,

$$\overline{H}(x,\theta,k) = \theta^{1/k}(a-x)/[a^k-(1-\theta)(a-x)^k]^{1/k},$$

$$x = 1,2,3\ldots a,\ 0 < \theta < \infty, k \in N \tag{2.1}$$

where N is the set of positive integers.

We write $X \sim HDU(a,\theta,k)$, a,k integer ≥ 1, $\theta > 0$ for the discrete r.v. with $\overline{H}(x)$ given in (2.1).

Remark 2.1*We see that when we put $k = 1$, $HDU(a,\theta,k)$ distribution reduces to $MODU(a,\theta)$ distribution (Sandhya and Prasanth [5]) with s.f,*

$$\overline{G}(x,\theta) = \theta(a-x)/[a\theta + (1-\theta)x], x = 1,2\ldots a, \theta > 0.$$

From (2.1) it is clear that

Remark 2.2*HDU distribution does not possess additive property.*

Remark 2.3*HDU is not infinitely divisible (i.d) since their support is $\{1,2,3\ldots a\}$ a finite. It is not log convex, since the class of log convex distribution forms a sub class of the class of i.d. distributions Satheesh, S and Sandhya, E [8].*

Note: Throughout this paper we have, $\theta > 0$ and $k \in N$ where N is the set of positive integers.

2.1 The p.m.f. of HDU Distribution

The p.m.f of HDU distribution is,

$$h(x,\theta,k) = \overline{H}(x-1,\theta,k) - \overline{H}(x,\theta,k).$$
$$= \theta^{1/k}\{[(a-x+1)/[a^k-(1-\theta)(a-x+1)^k]^{1/k}]$$
$$-[(a-x)/[a^k-(1-\theta)(a-x)^k]^{1/k}]\}, x = 1,2,3,\ldots a. \quad (2.2)$$

Since it is difficult to obtain compact mathematical expressions for reliability characteristics and moments for this discrete distribution, we numerically evaluate these.

For different values for the parameters (a,θ,k), we plot the graph of the p.m.f. of X (Figure 1 and Figure 2). Some are tabulated (Table 1) here.

Table 1: The p.m.f. of $X \sim HDU(a,\theta,k)$ with $a = 20$, $k = 2$.

X \ θ	2	0.5
1	0.02596117	0.09317948
2	0.02797983	0.08179188
3	0.03014582	0.07299292
4	0.03246097	0.06604138
5	0.03492407	0.0604511
6	0.03752991	0.05589132
7	0.04026831	0.05213009
8	0.04312304	0.04900054
9	0.04607084	0.04638003
10	0.0490805	0.04417678
11	0.05211224	0.04232111
12	0.05511756	0.04075945
13	0.05803968	0.03945028
14	0.06081477	0.03836125
15	0.06337411	0.03746708
16	0.06564707	0.03674816
17	0.06756502	0.03618942
18	0.06906557	0.0357796
19	0.07009705	0.03551066
20	0.07062246	0.03537746

As 'a' tends to infinity, the probabilities are becoming infinitesimally small as clear from the third figure here.
Also we have

$$\underset{(a\to\infty)}{\text{Lt}} \overline{H}(x,\theta,k) = \underset{(a\to\infty)}{\text{Lt}} \theta^{1/k}(a-x)/[a^k-(1-\theta)(a-x)^k]^{1/k}$$
$$= \underset{(a\to\infty)}{\text{Lt}} \theta^{1/k}(1-(x/a))/[1-(1-\theta)(1-(x/a))^k]^{1/k}$$
$$= \theta^{1/k}/[1-(1-\theta)]^{1/k}.$$
$$= 1$$

i.e., $\overline{H}(x,\theta,k)$ tends to 1 as 'a' tends to ∞.

For $\theta > 1$ the p.m.f. increases with x and for $\theta < 1$ the p.m.f. decreases with x.

3 Some Relations between MO Scheme and Harris Scheme

Consider the p.g.f.,

$$P_1(s) = \theta s^k/(1-(1-\theta)s^k). \quad (3.1)$$

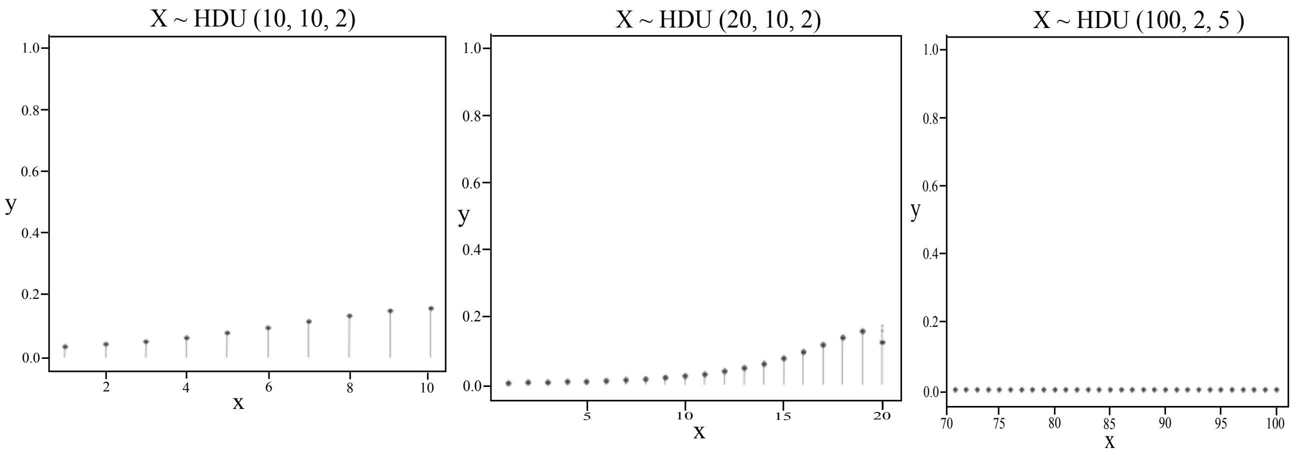

Fig. 1: The p.m.f. of $X \sim HDU(a,\theta,k)$ with $\theta > 1$

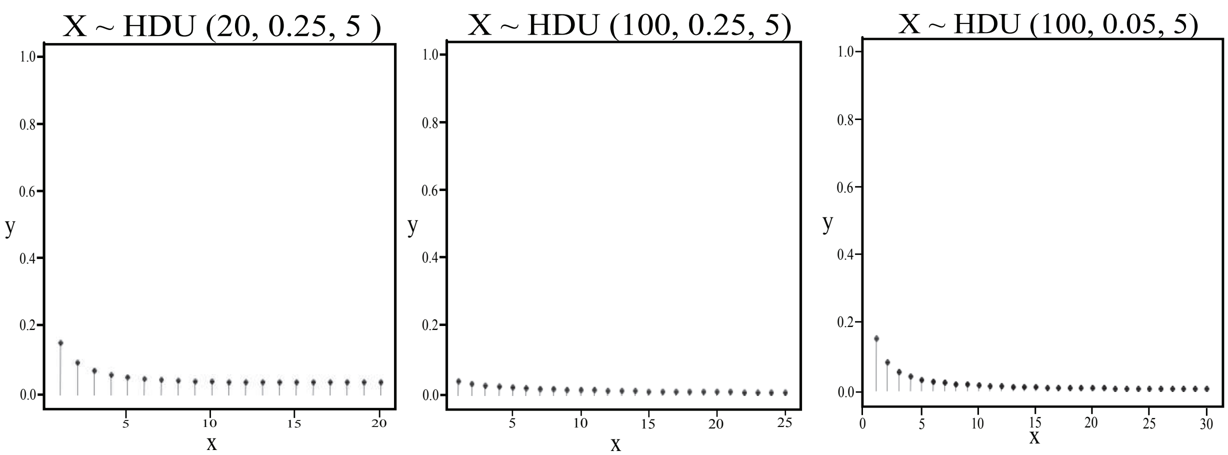

Fig. 2: The p.m.f. of $X \sim HDU(a,\theta,k)$ with $\theta < 1$

This corresponds to the extended geometric r.v Y on $\{k,2k,3k\ldots\}$ which has p.m.f.

$$P(Y=y)=\theta(1-\theta)^y, y=k,2k,3k\ldots, \text{ when } 0<\theta<1. \tag{3.2}$$

Now as in the case of defining a s.f in the MO scheme, we can define a new s.f,

$$J(x,\theta)=\theta\overline{F}^k(x)/[1-\overline{\theta}(\overline{F}^k(x))], k\geq 1 \text{ integer.} \tag{3.3}$$

Note that when $k=1$, this reduces to the MO scheme.

Now introduce MO extended geometric uniform (MOEGU) r.v on $\{1,2,3\ldots a\}$, which has s.f,

$$\begin{aligned} J(x,a,\theta) &= \theta((a-x)/a)^k/(1-(1-\theta)((a-x)/a)^k) \\ &= \theta(a-x)^k/((a)^k-(1-\theta)(a-x)^k) \end{aligned} \tag{3.4}$$

Now consider X following HDU distribution.
Then from (2.1) its s.f. is,
$\overline{H}(x,\theta,k)=\theta^{1/k}(a-x)/[a^k-(1-\theta)(a-x)^k]^{1/k}$,

$$x=1,2,3\ldots a,\ 0<\theta<\infty, k>0.$$

Let $Y = \min(X_1, \dots X_k)$, where x_i, $i = 1,2,3,\dots k$ are independent copies of $X_i \sim HDU(a,\theta,k)$.
Then the s.f of Y is,

$$\begin{aligned} P(Y > y) &= [P(Yi > y)]^k \\ &= \{\theta^{1/k}(a-x)/[a^k - (1-\theta)(a-x)^k]^{1/k}\}^k \\ &= \theta(a-x)^k/[a^k - (1-\theta)(a-x)^k] \end{aligned}$$

which is the s.f of MOEGU. Thus we have,

Theorem 3.1 $\min(X_1, \dots X_k)$ *follows* $MOGEU(\theta,a,k)$, *if and only if* X_i, $i = 1,2,\dots k$ *are i.i.d* $HDU(\theta,a,k)$ *on* $\{1,2,3\dots a\}$.

The following theorem connects MODU and HDU distributions.

Theorem 3.2 $\min(X_1^k, \dots X_k^k)$ *follows* $MODU(a,\theta)$ *if and only if* $X_i, i = 1,2,\dots k$ *are i.i.d* $HDU(\theta,a,k)$ *on* $\{1,2,3\dots a\}$.

Proof. Let $Y = \min(X_1^k, X_2^k, \dots, X_k^k)$. Then

$$\begin{aligned} P(Y > y) &= P\left(X_1^k > y, X_2^k > y, \dots, X_k^k > y\right) \\ &= [P\left(X_i^k > y\right)]^k \\ &= [P(X_i > y^{1/k})]^k \\ &= \frac{\theta(a-x)}{[a-(1-\theta)(a-x)]}, \text{ which is the s.f of MODU } (a,\theta). \end{aligned}$$

From the s.f of $HDU(a,\theta,k)$ in (2.1), the s.f of X_i (evaluated at $y^{1/k}$) implies that, the s.f $P(Y > y) = \theta(a-x)/[a-(1-\theta)(a-x)]$ which is the s.f of $MODU(a,\theta)$. Hence the proof. Retracing the steps we have *only if* part.

Generalizing this result we get,

Theorem 3.3 *A r.v X has s.f of the form* (2.1) *if and only if* $\min(X_1^k \dots X_k^k)$ *has s.f of the form* (1.1), $X_i, i = 1,2,\dots k$ *are i.i.d r.v.*

Proof. We have the s.f, $\overline{H}(x,\theta,k) = \theta^{1/k}\,\overline{F}(x)/[1-(1-\theta)(\overline{F}^k(x))]^{1/k}$ for Harris family.
Let $Z = \min\left(X_1^k, X_2^k, \dots, X_k^k\right)$. Then

$$\begin{aligned} \overline{H}_z(z,\theta,k) &= [P\left(X_i^k > z\right)]^k \\ &= [P\left(X_i > z^{1/k}\right)]^k \\ &= \left\{\theta^{1/k}\overline{F}^{1/k}(x)/[1-(1-\theta)(\overline{F}^k(x))]^{1/k}\right\}^k. \\ &= \theta\overline{F}(x)/[1-(1-\theta)(\overline{F}(x))] \end{aligned}$$

from (1.1) this s.f of MO scheme of distributions.

4 HDU Distribution as the Distribution of Minimum of a Sequence of i.i.d. Random Variables

The following theorem gives a characterization of minimum of a sequence of i.i.d. r.vs following discrete uniform distribution.

Theorem 4.1 *Let* $\{X_i, i \geq 1\}$ *be a sequence of i.i.d. r.vs with common s.f,* $\overline{F}(x)$. *Let N be a* $H_1(\theta,k,1/k)$ *r.v independent of* $\{X_i, i \geq 1\}$ *such that* $P(N = n) = \binom{\frac{n}{k}-1}{(n-1)/k}(\theta)^{1/k}(1-\theta)^{(n-1)/k}$, $n = 1, 1+k,\ 1+2k, \dots k \in N$ *where N is the set of positive integers,* $0 < \theta < \infty$. *Let* $U_N = \min_{(1 \leq i \leq N)}(X_i)$. *Then* $\{U_N\}$ *is distributed as* $HDU(a,\theta,k)$ *distribution if and only if* $\{X_i\}$ *follows discrete uniform distribution.* $i = 1,2,3\dots a$.

*Proof.*The s.f of U_N,

$$\begin{aligned}\overline{J}(x) &= P(U_N > x)\\ &= \sum_{1}^{\infty}[\overline{F}(x)]^n P(N=n)\\ &= \theta^{1/k}\overline{F}(x)/[1-(1-\theta)\overline{F}^k(x)]^{1/k} \text{ with } \overline{F}(x) = (a-x)/a.\\ &= \theta^{1/k}[(a-x)/a]/[1-(1-\theta)((a-x)/a)^k]^{1/k}.\\ &= \theta^{1/k}(a-x)/[a^k-(1-\theta)(a-x)^k]^{1/k}\end{aligned}$$

which is the s.f of HDU (a,θ,k).
By retracing the steps we have the converse.

5 Hazard Function

Let $\gamma_H(x)$ be the hazard function of $X \sim HDU(a,\theta,k)$, then,

$$\begin{aligned}\gamma_H(x) &= h(x)/\overline{H}(x).\\ &= \Big\{[\theta^{1/k}(a-x+1)/(a^k-(1-\theta)(a-x+1)^k)^{1/k}]/\\ &\quad [\theta^{1/k}(a-x)/(a^k-(1-\theta)(a-x)^k)^{1/k}]\Big\} - 1.\\ &= [(a-x+1)/(a-x)]\{(a^k-(1-\theta)(a-x)^k)/\\ &\quad (a^k-(1-\theta)(a-x+1)^k)\}^{1/k}.\end{aligned} \tag{5.1}$$

5.1 Increasing /Decreasing Failure Rate (IFR/DFR)

From (5.1), Comparing hazard function at x and $x+1$, we have HDU is IFR when $\theta > \dfrac{a-2x}{2a-2x}$ and DFR when $\theta < \dfrac{a-2x}{2a-2x}$.

We numerically (Table 2) and graphically (Figure 3) evaluate this.

From Table 2 we have,

Remark 5.1*The value of the Hazard function is decreasing when the value of θ increasing.*

From the graph also we can see that, the failure rate of the distribution is, increasing (IFR) when $\theta > (a-2x)/(2a-2x)$, decreasing (DFR) when $\theta < (a-2x)/(2a-2x)$.

6 AR (1) Model with HDU Distribution as Marginal Distribution

Consider k independent AR(1) sequences $\{X_{i,n}\}$, $i = 1,2,\ldots,k$, for $n > 0$ integer and some $b > 0$, and the structure

$$\bigwedge_i X_{i,n} = \begin{cases} b\{\bigwedge_i X_{i,n-1}\} & \text{with probability } p\\ b\{\bigwedge_i X_{i,n-1}\}\bigwedge\{\bigwedge_i \varepsilon_{i,n-1}\} & \text{with probability } (1-p).\end{cases} \tag{6.1}$$

The structure (6.1), was defined, by Satheesh *et al.* [11]. We now have

Theorem 6.1*In an AR (1) process with structure (6.1), $\{X_{i,n}\}$ is stationary with HDU(θ,a,k) marginal if and only if $\{\varepsilon_n\}$ is distributed as a discrete uniform r.v on $\{1,2,\ldots a\}$.*

Table 2: Hazard function $\gamma_H(x)$ for $X \sim HDU(a,\theta,k)$ with $a = 20$, $k = 2$

θ / X	3	1/3	2	0.5	5	0.2	10	0.1
1	0.017846	0.150695	0.026653	0.102754	0.010745	0.241034	0.005387	0.442335
2	0.020152	0.134322	0.029575	0.099138	0.012310	0.187763	0.006239	0.268165
3	0.022850	0.123955	0.032913	0.097060	0.014180	0.159333	0.007277	0.202867
4	0.026022	0.117449	0.036743	0.096271	0.016433	0.142569	0.008554	0.169857
5	0.029771	0.113692	0.041158	0.096637	0.019167	0.132400	0.010139	0.151063
6	0.034228	0.112085	0.046276	0.098114	0.022509	0.126506	0.012129	0.140028
7	0.039561	0.112327	0.052246	0.100730	0.026632	0.123729	0.014658	0.133932
8	0.045991	0.114309	0.059266	0.104585	0.031765	0.123500	0.017914	0.131428
9	0.053808	0.118082	0.067598	0.109868	0.038220	0.125597	0.022169	0.131895
10	0.063410	0.123854	0.077603	0.116880	0.046433	0.130064	0.027819	0.135147
11	0.075350	0.132025	0.089795	0.126089	0.057016	0.137193	0.035458	0.141344
12	0.090438	0.143276	0.104940	0.138222	0.070866	0.147595	0.045997	0.151010
13	0.109914	0.158738	0.124232	0.154444	0.089339	0.162350	0.060877	0.165169
14	0.135803	0.180349	0.149653	0.176721	0.114610	0.183361	0.082475	0.185688
15	0.171700	0.211645	0.184765	0.208607	0.150447	0.214140	0.114941	0.216051
16	0.224744	0.259744	0.236693	0.257237	0.204159	0.261786	0.166190	0.263339
17	0.311550	0.341300	0.322067	0.339272	0.292477	0.342940	0.253752	0.344181
18	0.481948	0.506331	0.490803	0.504733	0.465192	0.507618	0.428111	0.508587
19	0.985239	1.005027	0.992560	1.003765	0.970942	1.006039	0.937083	1.0068

*Proof.*We have,

$$\overline{H}(t) = \{\theta\overline{F}^k(bt)/[1-(1-\theta)\overline{F}^k(bt)\}^{1/k}.$$

$$\text{Let} \quad \overline{F}(bt) = (a-x)/a,$$

$$\text{then,} \quad \overline{H}(t) = \{\theta\overline{F}^k(bt)/[1-(1-\theta)\overline{F}^k(bt)\}^{1/k}.$$

$$= \left[\theta^{1/k}(a-x)/a\right]/[1-(1-\theta)((a-x)/a)^k]^{1/k}.$$

$$= \theta^{1/k}(a-x)/[a^k-(1-\theta)(a-x)^k]^{1/k}.$$

So $X_{i,1}$ is $HDU(a,\theta,k)$. We can show the converse by retracing the steps.

7 Expectation, Standard Deviation and Entropy of HDU Random Variable

We numerically compute expectation, standard deviation (sd) (Table 3, Table 4) and entropy (Table 6) of the HDU r.v with different a,k and θ. We have,

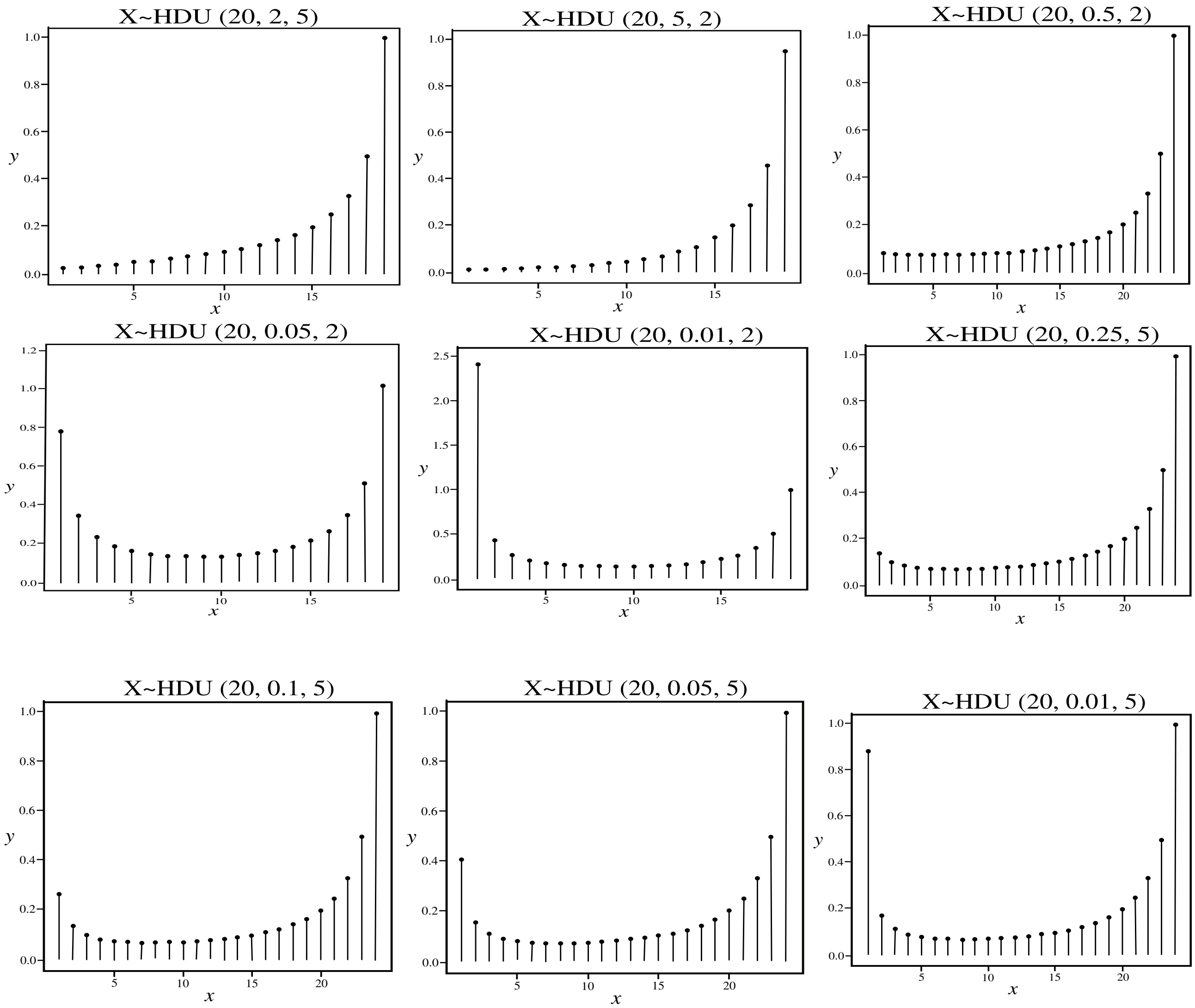

Fig. 3: Hazard function for $X \sim HDU(a,\theta,k)$

$$\begin{aligned}\text{Shannon's Entropy} &= -\sum_i p(x_i)\log p(x_i), i = 1,2,\ldots,\\ &= -\sum_i \theta^{1/k}\{[(a-x)/(a^k-(1-\theta)(a-x)^k)^{1/k}]\\ &\quad -[(a-x-1)/(a^k-(1-\theta)(a-x-1)^k)^{1/k}]\}\\ &\quad \log(\theta^{1/k}\{[(a-x)/(a^k-(1-\theta)(a-x)^k)^{1/k}]\\ &\quad -[(a-x-1)/(a^k-(1-\theta)(a-x-1)^k)^{1/k}]\}).\end{aligned}$$

From Table 3 and Table 4, we have,

Remark 7.1*If $X \sim HDU(a,\theta,k)$, then for all $\theta > 0$, $E(X)$ increases when θ increases.*

Remark 7.2*In the case of $HDU(a,\theta,k)$, the sd is decreasing with increasing value of θ, when $\theta > 1$ and increasing with increasing value of θ, when $\theta < 1$.*

Table 3: Expectation and sd of $X \sim HDU(a,\theta,k)$, $\theta < 1$, $k = 2$ & with different values of 'a'

a	$\theta = 0.25$		$\theta = 0.5$		$\theta = 0.75$	
	$E(X)$	$Sd(X)$	$E(X)$	$Sd(X)$	$E(X)$	$Sd(X)$
10	3.861796	2.843892	4.652862	2.912528	5.144906	2.902485
12	4.523884	3.42449	5.479514	3.502731	6.072462	3.489097
15	5.519222	4.292872	6.720372	4.386299	7.464119	4.367621
20	7.181154	5.736697	8.789652	5.856586	9.783979	5.829971
25	8.84495	7.178379	10.85965	7.325473	12.1041	7.29121
30	10.50969	10.50969	12.93	8.793659	14.42435	8.751893
50	17.17249	14.37691	21.21283	14.6636	23.70586	14.59241
75	25.50389	21.57098	31.56745	21.99891	35.30814	21.89138
100	33.83625	28.76394	41.92243	29.33352	46.91055	29.18981

Table 4: Expectation and sd of $X \sim HDU(a,\theta,k)$, $\theta > 1$, $k = 2$ & with different values of 'a'

A	$\theta = 2$		$\theta = 5$		$\theta = 7.5$	
	$E(X)$	$Sd(X)$	$E(X)$	$Sd(X)$	$E(X)$	$Sd(X)$
10	6.350238	2.724793	7.392823	2.397142	7.803423	2.219902
12	7.523084	3.274012	8.777634	2.880833	9.272117	2.668547
15	9.281715	4.096862	10.85342	3.605371	11.47332	3.340406
20	12.21192	5.466982	14.31117	4.811627	15.13956	4.45873
25	15.14161	6.836327	17.76779	6.017109	18.80434	5.576207
30	18.07105	8.205286	21.22383	7.222207	22.46839	6.693266
50	29.7878	13.67958	35.04576	12.04107	37.12171	11.15984
75	44.43297	20.52129	52.32146	18.0635	55.43619	16.74181
100	59.07788	27.36262	69.5966	24.08555	73.74994	22.32338

Remark 7.3*The $E(X)$ and $sd(X)$ for $X \sim HDU(a,\theta,k)$ is greater than or equal to the $E(X)$ and $sd(X)$ of $X \sim MODU(a,\theta)$ for $\theta < 1$ and is less than or equal to the $E(X)$ and $sd(X)$ of $X \sim MODU(a,\theta)$ for $\theta > 1$.*

Let $X \sim HDU(a,\theta,k)$, then the mean, median and mode of the distribution from (Figure 1, Figure 2, Table 1, Table 3 and Table 4) for different a and θ are computed below (Table 5).

From the Figure 1, Figure 2, Table 5 we have,

Remark 7.4*The HDU distribution is positively skewed when $\theta < 1$ since mode $<$ median $<$ mean and the distribution is negatively skewed when $\theta > 1$ since mode $>$ median $>$ mean. Also it is to be noted that the distribution is unimodal, i.e., when $\theta < 1$ the mode $= 1$ and when $\theta > 1$ mode $= a$.*

The HDU distribution is therefore a suitable model for, the data showing positive skewness when $\theta > 1$ and the data showing negative skewness when $\theta < 1$.

Remark 7.5*For $X \sim HDU(a.\theta,k)$ the mean, median and mode increase with k for $\theta < 1$ and decrease with k for $\theta > 1$.*

Remark 7.6*For $X \sim HDU(a.\theta,k)$ the mean and median are greater than that of $X \sim MODU(a,\theta)$ when $\theta < 1$ and less than that of $X \sim MODU(a,\theta)$ when $\theta > 1$ (mode is equal in both $MODU(a,\theta)$ and $HDU(a,\theta,k)$ distributions).*

Table 5: Mean, median, mode, s.d. and skewness of the HDU distribution with different a,θ and $k=2$

k	a	θ	mean	median	mode	Sd	skewness
2	10	0.2	3.636	3	1	2.80063	0.9378863
		0.5	4.652	4	1	2.912528	1.25419
		2	6.350	7	10	2.724793	-1.339464
		5	7.392	8	10	2.397142	-1.087619
	20	0.2	6.699	5	1	5.657518	1.007351
		0.5	8.789	8	1	5.856586	1.330067
		2	12.212	13	20	5.466982	-1.424567
		5	14.311	15	20	4.811627	-1.182309
	100	0.2	31.405	21	1	28.38168	1.071307
		0.5	41.92	37	1	29.33352	1.395074
		2	59.077	63	100	27.36262	-1.495548
		5	69.596	75	100	24.08555	-1.262308

Now we have the computed values for entropy.

Table 6: Entropy of $X \sim HDU(a,\theta,k)$ with different "a" & "θ", with $k=2$

a	$\theta=0.1$	$\theta=0.2$	$\theta=0.25$	$\theta=0.5$	$\theta=2$	$\theta=4$	$\theta=5$	$\theta=10$
10	1.797096	2.04458	2.109813	2.254229	2.256818	2.128346	2.071995	1.859359
15	2.180142	2.443605	2.511336	2.65915	2.66204	2.532955	2.476311	2.262265
20	2.458405	2.728861	2.797568	2.94664	2.949637	2.820337	2.763589	2.549047
30	2.856328	3.132525	3.201971	3.351967	3.355042	3.225588	3.168766	2.953867
50	3.362921	3.642403	3.712243	3.862721	3.865836	3.736303	3.679444	3.464362
100	4.054183	4.335145	4.405155	4.555839	4.55897	4.429404	4.372528	4.157369

From table 6 it is observed that,

Remark 7.7*The entropy of $X \sim HDU(a,\theta,k)$ is approximately equal to the entropy of $X \sim HDU(a,1/\theta,k)$.*

Remark 7.8*The entropy of $X \sim HDU(a,\theta,k)$ is always greater than or equal to the entropy of $X \sim MODU(a,\theta)$ for all θ.*

8 Maximum Likelihood Estimates (MLE) of the Parameters of HDU (a,θ,k)

Let $x_1,x_2,\ldots,x_n$ be a random sample from $HDU(a,\theta,k)$. Then from (2.2) we write the likelihood function L of the distribution. By partial differentiation of L w.r.to a,θ and k and equating them to zero we formulate 3 non-linear equations. We can numerically find (with the help of mathematica) the MLE of a,θ and k as the solution of these non-linear equations. But here the maximum of the range of observations is a. So MLE of the parameter a, $(\widehat{a})$ is the largest value of the observations. Then substituting the MLE of a in the remaining two equations we find MLE of θ and k.

9 An Application of HDU

Example 91*A travel agency in Palakkad city (Kerala, India) arrange mini luxury trips (19 seats) to Cochin (another town about 150 Kms from Palakkad) on all days excluding Sundays. The booking status (vacancies) was observed from their database on the day before the scheduled date (since they are making arrangement and announcing their trip status before 12 hours so that if the booking status is very less, they will pool their trips.) (data collected in each 50 days).*

Table 7: Observed frequencies: O_i – the number of vacancies in each day

O_i / Year 2010-11	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Apr-May	9	6	6	1	2	3	2	3	3	2	1	1	1	1	1	2	1	2	2	1
June-July	10	5	3	3	4	3	3	2	1	4	1	1	1	0	1	0	1	3	2	2
Aug-Sep	9	7	2	6	1	1	2	2	1	2	3	2	2	3	0	1	4	0	1	1
Oct-Nov	9	10	5	1	3	2	0	1	2	3	0	1	2	1	2	3	3	2	0	0
Dec-Jan 2011	10	6	4	4	2	1	3	2	1	5	2	1	2	1	2	0	2	1	1	0
Feb-Mar	10	6	2	5	2	2	3	1	2	2	4	0	2	0	0	2	3	1	0	3
Apr-May	10	4	5	4	3	3	3	1	1	2	0	2	1	1	1	0	3	1	1	4
June-July	9	6	4	2	5	3	1	2	2	1	0	4	2	3	0	2	0	3	0	0
Aug-Sep	10	9	1	2	3	3	5	1	1	2	1	2	3	1	0	1	1	2	1	1

Let X be the number of vacancies in each day plus one unit and assume X is $HDU(a,\theta,k)$. i.e., $x = 1,2,3\dots 20$. We arbitrarily fix 4 terms, and the MLE's are computed.

Table 8: The MLE of θ and k (Sample size = 50, MLE of a, $\widehat{a} = \max(x) = 20$) 4 months

Month	June-July	Feb-March	Apr-May	Aug-Sep	Mean	SE
MLE $\widehat{\theta}$	0.1993	0.1733	0.186	0.1854	0.186	0.01738
MLE $\widehat{k}$	2.688	3.270	2.970	2.9113	2.9598	0.04314
$\widehat{k}$ rounded	3	3	3	3	3	

Table 9: Mean, median and mode of the data in the selected months.

a	Month	MLE of $\theta = \widehat{\theta}$	mean	median	mode	Sd	skewness
20	June-July	0.1993	7.3	5	1	6.060528	1.036695
	Feb-March	0.1733	7.36	5	1	6.0721	1.047414
	Apr-May	0.186	7.68	5	1	5.124217	1.303614
	Aug-Sep	0.1854	6.54	5	1	5.978997	1.093829

Here mode $<$ median $<$ mean the distribution is is positively skewed. From remark 7.4, then HDU distribution is supposed to apply for, the data showing positive skewness when $\theta < 1$. Assume that the data follows HDU distribution. Sample size = 50, MLE of a, $\widehat{a} = \max(x) = 20$ and estimate of k rounded, $\widehat{k} = 3$. Initially we fit the MODU distribution to the data.

The Chi-square test is used to test the goodness of fit of the distribution with level of significance $\alpha = 0.05$. The expected frequencies are calculated ($E_i = N * h(x_i, a, \theta, k) = 50 * h(x_i, 20, \widehat{\theta}, \widehat{k})$ for $i = 1, 2, \dots 20$). Then found the value of χ^2 statistics and observed the degrees of freedom (d.f). The results are tabulated below.

Result: From the p-values it is seen that HDU distribution is a better fit than MODU distribution in this situation.

Table 10: Goodness of fit of MODU and HDU distribution with $\alpha = 0.05$

Distribution		June-July	Feb-Mar	Apr-May	Aug-Sep
MODU	χ^2 statistics	12.807	17.157	19.449	11.1659
	d.f	6	6	6	6
	p-value	0.0462	0.0087	0.0035	0.0834
HDU	χ^2 statistics	5.688	5.2	6.2715	5.98572
	d.f	7	7	7	7
	p-value	0.5766	0.6356	0.5084	0.5414

References

[1] Harris, T E (1948); Branching Processes, *Annals of Mathematical Statistics*, 19, 474–494.
[2] Jose, K K and Krishna, E (2011); Marshall-Olkin Extended Uniform Distribution, *ProbStat Forum*, Volume 04, Pages 78–88 ISSN 0974-3235.
[3] Marshall, A.W and Olkin, I (1997); A New Method for Adding a Parameter to a Family of Distribution with Application to the Exponential and Wiebull Families, *Biometrika*, 84.3 pp 641–652.
[4] Sandhya, E and Prasanth, C B (2012); A generalized geometric distribution, *Proceedings of international conference on frontiers of statistics and application & 32nd annual conference of Indian society for probability and statistics*, Department of Statistics, Podichery University, Dec-2012; Pages 261–269, ISBN 978-93-82338-78-9, Bonfring publication.
[5] Sandhya, E and Prasanth, C B (2013); Marshall-Olkin Discrete uniform distribution, *Journal of probability*, Volume 2014, 10 pages, Article ID 979312, Hindawi Publishing Corporation.
[6] Sandhya, E, Sherly, S, Jose, K K and Raju, N (2006b); Characterizations of the Extended Geometric, Harris, Negative Binomial and Gamma Distributions, *Stars Int.Journal.*
[7] Sandhya, E, Sherly, S and Raju, N (2006a); Harris family of Discrete Distributions, *Proceedings of National Seminar Conducted by Kerala Statistical Association and The Dept.of Statistics, University of Kerala*, March 2006.
[8] Satheesh, S and Sandhya, E (2003); Infinite divisibility and max-infinite divisibility with random sample size, *Statist.Meth.*, 5, 126–139.
[9] Satheesh, S, Sandhya, E and Nair, N U (2002); Stability of Random Sums, *Stoch. Modelling and Appl.*, 5, 17–26.
[10] Satheesh, S, Sandhya, E and Prasanth, C B (2012); A Generalization of the Marshal-Olkin Scheme and Related Processes, *Collection of Recent Statistical Methods and Applications*, 2012, pp.12–17. Publication of Department of Statistics, University of Kerala.
[11] Satheesh,S, Sandhya, E and Sherly, S (2006); A generalization of Stationary AR(1) Schemes, *Statistical Methods*, 8(2), 213–225.

Prasanth C B received his PhD degree in Statistics from Mahatma Gandhi University, Kerala, India under the supervision and guidance of Dr.E.Sandhya. His main research interests are: distribution theory, applied statistics, probability theory and study on urbanization. He has published research articles in reputed international and national journals.

E. Sandhya took her Ph.D from the Department of Statistics,University of Kerala in 1992 under the supervision and guidance of Dr.R.N.Pillai. Her research interest includes distribution theory,probability theory in general and infinite divisibility and geometric infinite divisibility and its extensions in particular. She has more than 35 published papers in various international and national publications. She is the reviewer for Mathematical Reviews of American Mathematical Society, USA .