DISCRETIZATION OF THE METHOD OF GENERATING AN EXPANDED FAMILY OF DISTRIBUTIONS BASED UPON TRUNCATED DISTRIBUTIONS

by

Muhammad FAROOQ [a], ***Muhammad MOHSIN*** [a], ***Muhammad NAEEM*** [b], ***Muhammad FARMAN*** [c], ***Ali AKGUL*** [d*], ***and Muhammad Umar SALEEM*** [e]

[a] Department of Statistics, COMSATS University, Islamabad Lahore Campus, Lahore, Pakistan
[b] Department of Economics and Business, Administration, University of Education, Lahore, Pakistan
[c] Department of Mathematics and Statistics, University of Lahore, Lahore, Pakistan
[d] Department of Mathematics, Art and Science Faculty, SIIRT University, Siirt, Turkey
[e] Department of Mathematics, Division of Science and Technology, University of Education, Lahore, Pakistan

Original scientific paper
https://doi.org/10.2298/TSCI200605003F

Discretization translates the continuous functions into discrete version making them more adaptable for numerical computation and application in applied mathematics and computer sciences. In this article, discrete analogues of a generalization method of generating a new family of distributions is provided. Several new discrete distributions are derived using the proposed methodology. A discrete Weibull-Geometric distribution is considered and various of its significant characteristics including moment, survival function, reliability function, quantile function, and order statistics are discussed. The method of maximum likelihood and the method of moments are used to estimate the model parameters. The performance of the proposed model is probed through a real data set. A comparison of our model with some existing models is also given to demonstrate its efficiency.

Key words: *beta generated distributions, discrete analogue, method of moment, truncated distributio, T-X family of distributions*

Introduction

Generally, the data obtained from the real world phenomena are always discrete in nature even for the continuous variables *e.g.* computer databases. *All actual sample spaces* are *discrete, and all observable random variables have discrete distributions. The continuous distribution is a mathematical construction, suitable for mathematical treatment, but not practically observable* [1]. Indeed, the continuous data can be modeled by using continuous distributions but it is difficult to derive their properties especially in case of complex systems. In this situation the discrete version of the underlying distribution is a more realistic approach to model and analyze the data. Out of several methods of discretization present in literature, some prominent work is referenced. Brasquemond and Gaudoin [2] give a detailed survey of discrete

[*] Corresponding author, e-mail: aliakgul00727@gmail.com

distributions used in reliability of non-repairable discrete lifetimes systems. They segregate the discrete distributions into two families, one derived from the continuous distributions, and the other as Polya urn distributions. They also describe some norms for the selection of appropriate distribution for application among several given distributions. Dilip [3] follows the approach of discrete concentration set up discrete version of normal distribution. He proposes the expression of probability mass function (PMF) for discrete normal distribution:

$$P(x) = F\left(\frac{x+1-\mu}{\sigma}\right) - F\left(\frac{x-\mu}{\sigma}\right), \quad x = ...,-1,0,1,..., \tag{1}$$

where $F(.)$ is the cumulative distribution function (CDF) of normal variate. Dilip [4] defines the same technique to propose discrete Rayleigh distribution with PMF:

$$P(x) = \begin{cases} \theta^{x^2} - \theta^{(x+1)^2} & x = 0,1,... \\ 0 & \text{otherwise} \end{cases}$$

Inusah and Kozubowski [5] follow the same method as Kemp [6] to propose discrete analogues of Laplace distribution. As compared to discrete normal distribution proposed by [6] the expression for discrete Laplace distribution has closed-form for PMF, distribution function, mean and variance. Alzaatreh *et al*. [7, 8] extend the technique of generating continuous distribution and propose its discrete versions. In recent years, a lot of work is done on discrete random variable. For instance, Al-Masoud [9] uses difference of survival functions to construct discrete distribution and introduces discrete modified Weibull extension, discrete modified Weibull Type-I and discrete modified Weibull Type-II distribution study the survival of lungs cancer patients. Almalki and Nadarajah [10] develop discrete analogue of reduced modified Weibull distribution and study its various characteristics. distributions. Nekoukhou and Bidram [11] present discrete exponentiated Weibull distribution and discuss its special cases. Hossain [12] establishes a generalized family of discrete distribution get a general expression of probability mass function which yields PMF for different factors. Alamatsaz *et al*. [13] propose a new discrete generalized Rayleigh distribution and use difference of survival functions to develop a discrete analogue of continuous distribution. Chakraborty [14] provides a detailed study on discretization of continuous distributions. Chakraborty and Chakravarty [15] give a distribution having integer domain over (– ∞, ∞) as discrete analogue of logistic distribution. They discuss several statistical properties and derive discrete analogue of characterization result of continuous logistic distribution. Jayakumar and Babu [16] propose Weibull-Geometric distribution using conditional CDF and study its properties. They discuss discrete Weibull, discrete Rayleigh, and geometric distributions as special cases of their proposed distribution. Mazucheli *et al*. [17] establish two discrete anlogues of Shanker distribution, one by the method based os survival function and other by using the method of infinite series to model the over dispersed data. In literature, usually a continuous approximation of discrete distributions is used to fit the models, but in the last decade the researchers have preferred discrete analogue models to fit the discrete data.

In our real life we come across many situations where we need to translate continuous outcomes into the discrete values. The continuous phenomena are sometimes difficult to be interpreted, limited and concluded therefore, we discretize them into meaningful groups or levels. For instance, the data of GPS location can be better understood if it is discretized into cities, countries or states. Similarly the continuous features of neural networks, data miming and information theory based methods, *etc.* cannot be managed by machine learning until or unless they are discretized. This discretization not only helps to improve the performance of the considered method but also enhances its accuracy. In literature we see a considerable work on the discretization of the continuous distributions [14]. We have develop a new method to generate families of continuous distribution, due to availability of closed form CDF we use it to develop the discrete analogues of the new family of distribution. By using this methodology the resulting discrete distribution is obtained as the combination of discrete distributions which certainly helps to model a variety of real life situations. We also explore various mathematical properties of the new proposed method which eases to study the characteristics of any mixed distributions. For illustrative purpose we develop a new discrete Weibull-Geometric distribution which is useful in lifetime data as well as medical and environmental data. The new composed discrete distribution has a closed CDF which makes it preferred to the other discrete distributions which do not have closed CDF. We expect that the new method of developing discrete distributions will prove a new dimension in research.

Methodology

If X is a discrete random variable with CDF $F(x)$ and t be a continuous random variable with PDF $h(t)$ defined on any domain, then the CDF of truncated T-X family of distributions is given:

$$G(x)=\frac{1}{H(1)-H(a)}\int_a^{F(x)} h(t)\mathrm{d}t=\frac{H\left[F(x)\right]-H(a)}{H(1)-H(a)} \quad -\infty\le a\le x\le b\le\infty \tag{2}$$

where $h(.)$ and $H(.)$ are PDF and CDF of r.v. t, respectively, while $F(x)$ is CDF of input distribution. The corresponding PDF is obtained:

$$g(x)=\frac{1}{H(1)-H(a)}h\left[F(x)\right]f(x) \tag{3}$$

For several choices of $h(t)$ and $F(x)$, one can easily generate new truncated continuous distributions. On the parallel lines the discrete analogue of eq. (3) can be obtained using technique given:

$$g(x)=S(x)-S(x+1) \tag{4}$$

The survival function:

$$S(x)=\frac{H(1)-H\left[F(x)\right]}{H(1)-H(a)}$$

Substituting the value of survival function in eq. (4) the PMF of discrete analogue of truncated distribution is given:

$$g(x)=\frac{H\left[F(x+1)\right]-H\left[F(x)\right]}{H(1)-H(a)} \tag{5}$$

The hazard rate of proposed model:

$$r(x)=\frac{H\left[F(x+1)\right]-H\left[F(x)\right]}{H(1)-H\left[F(x)\right]}$$

Several continuous truncated distribution can be generated by using the new method of generating generalized distribution. In tab. 1, some new discrete distributions are derived using the proposed methodology.

Table 1. Some new discrete distributions

			$F(x)$		
			Geometric	Discrete Weibull	Discrete Burr
			$1-p^{x+1}$	$1-\mathrm{e}^{-\left(\frac{x+1}{\alpha}\right)^{\beta}}$	$1-\theta^{\log\left[1+(x+1)^{\alpha}\right]}$
$H(t)$	Exponential	$1-\mathrm{e}^{-\theta t}$	$\dfrac{1-\mathrm{e}^{-\theta\left(1-p^{x+1}\right)}}{1-\mathrm{e}^{-\theta}}$	$\dfrac{1-\mathrm{e}^{-\theta\left(1-\mathrm{e}^{-\left(\frac{x+1}{\alpha}\right)^{\beta}}\right)}}{1-\mathrm{e}^{-\theta}}$	$\dfrac{1-\mathrm{e}^{-\theta\left(1-\theta^{\log\left[1+(x+1)^{\alpha}\right]}\right)}}{1-\mathrm{e}^{-\theta}}$
	Weibull	$1-\mathrm{e}^{-\left(\frac{t}{\gamma}\right)^{c}}$	$\dfrac{1-\mathrm{e}^{-\left[\frac{\left(1-p^{x+1}\right)}{\gamma}\right]^{c}}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^{c}}}$	$\dfrac{1-\mathrm{e}^{-\left\{\frac{\left(1-\mathrm{e}^{-\left(\frac{x+1}{\alpha}\right)^{\beta}}\right)}{\gamma}\right]^{c}}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^{c}}}$	$\dfrac{1-\mathrm{e}^{-\left\{\frac{\left(1-\theta^{\log\left[1+(x+1)^{\alpha}\right]}\right)}{\gamma}\right\}^{c}}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^{c}}}$
	Logistic	$\dfrac{1}{1+\mathrm{e}^{-t/\theta}}$	$\dfrac{2\left(1+\mathrm{e}^{-1/\theta}\right)}{\left(1-\mathrm{e}^{-1/\theta}\right)}\dfrac{1}{\left[1+\mathrm{e}^{-\left(1-p^{x+1}\right)/\theta}\right]}$	$\dfrac{2\left(1+\mathrm{e}^{-1/\theta}\right)}{\left(1-\mathrm{e}^{-1/\theta}\right)}\dfrac{1}{\left\{1+\mathrm{e}^{-\left[1-\mathrm{e}^{-\left(\frac{x+1}{\alpha}\right)^{\beta}}\right]/\theta}\right\}}$	$\dfrac{2\left(1+\mathrm{e}^{-1/\theta}\right)}{\left(1-\mathrm{e}^{-1/\theta}\right)}\dfrac{1}{\left(1+\mathrm{e}^{-\left\{1-\theta^{\log\left[1+(x+1)^{\alpha}\right]}\right\}/\theta}\right)}$
	Rayleigh	$1-\mathrm{e}^{-\frac{t^{2}}{2\sigma^{2}}}$	$\dfrac{1-\mathrm{e}^{-\frac{(1-p^{x+1})^{2}}{2\sigma^{2}}}}{1-\mathrm{e}^{-\frac{1}{2\sigma^{2}}}}$	$\dfrac{1-\mathrm{e}^{-\frac{\left[1-e^{-\left(\frac{x+1}{\alpha}\right)^{\beta}}\right]^{2}}{2\sigma^{2}}}}{1-\mathrm{e}^{-\frac{1}{2\sigma^{2}}}}$	$\dfrac{1-\mathrm{e}^{-\frac{\left\{1-\theta^{\log\left[1+(x+1)^{\alpha}\right]}\right\}^{2}}{2\sigma^{2}}}}{1-\mathrm{e}^{-\frac{1}{2\sigma^{2}}}}$
	Lomax	$1-\left[1+\dfrac{x}{\lambda}\right]^{\alpha}$	$\dfrac{1-\left[1+\frac{1-p^{x+1}}{\lambda}\right]^{\alpha}}{1-\left[1+\frac{1}{\lambda}\right]^{\alpha}}$	$\dfrac{1-\left[1+\frac{1-\mathrm{e}^{-\left(\frac{x+1}{\alpha}\right)^{\beta}}}{\lambda}\right]^{\alpha}}{1-\left[1+\frac{1}{\lambda}\right]^{\alpha}}$	$\dfrac{1-\left[1+\frac{1-\theta^{\log\left[1+(x+1)^{\alpha}\right]}}{\lambda}\right]^{\alpha}}{1-\left[1+\frac{1}{\lambda}\right]^{\alpha}}$

Truncated Weibull-Geometric distribution (TWG)

Let T be a continuous Weibull random variable with PDF:

$$h(t)=\frac{c}{\gamma}\left(\frac{t}{\gamma}\right)^{c-1}\mathrm{e}^{-\left(\frac{t}{\gamma}\right)^{c}}$$

and X be a geometric random variable with PMF $1 - p^{x+1}$, then the CDF of Weibull-Geometric distribution:

$$G(x) = \frac{1 - e^{-\left[\frac{\left(1-p^{x+1}\right)}{\gamma}\right]^c}}{1 - e^{-\left(\frac{1}{\gamma}\right)^c}} \tag{6}$$

and PMF:

$$g(x) = \frac{e^{-\left[\frac{\left(1-p^{x}\right)}{\gamma}\right]^c} - e^{-\left[\frac{\left(1-p^{x+1}\right)}{\gamma}\right]^c}}{1 - e^{-\left(\frac{1}{\gamma}\right)^c}} \quad x = 0, 1, 2, ... \tag{7}$$

Let X be a discrete random variable distributed as TWG then the survival function and failure rate are given:

$$S(x) = 1 - G(x-1) = \frac{e^{-\left[\frac{\left(1-p^{x}\right)}{\gamma}\right]^c} - e^{-\left(\frac{1}{\gamma}\right)^c}}{1 - e^{-\left(\frac{1}{\gamma}\right)^c}} \tag{8}$$

and

$$R(x) = \frac{g(x)}{S(x)} = \frac{e^{-\left[\frac{\left(1-p^{x}\right)}{\gamma}\right]^c} - e^{-\left[\frac{\left(1-p^{x+1}\right)}{\gamma}\right]^c}}{e^{-\left[\frac{\left(1-p^{x}\right)}{\gamma}\right]^c} - e^{-\left(\frac{1}{\gamma}\right)^c}} \tag{9}$$

respectively.

Quantile function

The discrete TWG distribution is unimodal. Here we use the condition for unimodality of a probability function defined by [18]. Since the PMF in eq. (7) satisfies the log-concave inequality $p^2(x) \geq p(x-1)\, p(x+1)$ for $x = 0, 1, 2, ...$, so the TWG distribution is a unimodal which can also be observed in fig. 1 as well. Failure function for several values of the parameters can be seen in fig. 2. Let X be discrete Weibull-Geometric random variable then the quantile function is given:

$$Q(u) = \frac{\ln\left(1 - \gamma\left\{\ln\left[\frac{1}{1 - u\left(1 - e^{-1/\gamma^c}\right)}\right]\right\}^{1/c}\right)}{\ln(p)} - 1 \tag{10}$$

where $0 \leq u \leq 1$. The median of TWG distribution can be obtained by substituting.

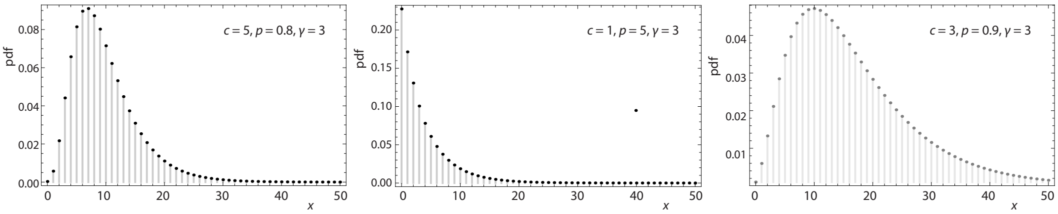

Figure 1. The PDF of the TWG distribution for several values of the parameters

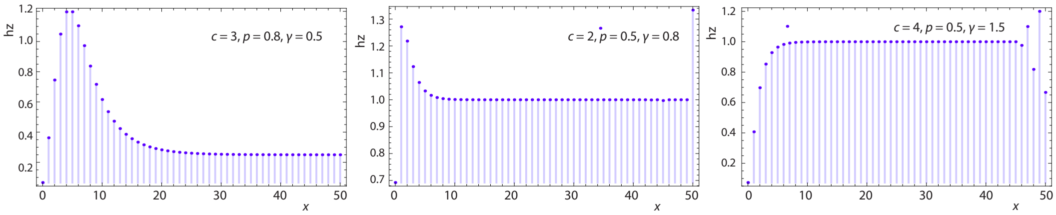

Figure 2. Plots of failure function for several values of the parameters

Moment

If X is distributed as TWG distribution then the r^{th} moment:

$$\mu_r = E\left(x^r\right) = \sum_{r=0}^{\infty} x^r \frac{e^{-\left[\frac{\left(1-p^x\right)}{\gamma}\right]^c} - e^{-\left[\frac{\left(1-p^{x+1}\right)}{\gamma}\right]^c}}{1-e^{-\left(\frac{1}{\gamma}\right)^c}} \tag{11}$$

Algebraic calculation of eq. (11) is arduous so we evaluate mean, variance, skewness and kurtosis numerically. Results are given in tabs. 2 and 3.

Table 2. Mean and variance of TWG distribution for different values of parameters

	$p = 0.5$											
	$c = 2$				$c = 5$				$c = 8$			
	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
Mean	1.5631	1.4846	1.2716	0.4954	2.7796	2.7331	2.3505	0.3684	3.4191	3.4024	2.9645	0.3679
Variance	2.5276	2.4165	2.0915	0.7125	3.1131	3.0596	2.5582	0.2337	3.2587	3.2404	2.6909	0.2325
	$p = 0.75$											
	$c = 2$				$c = 5$				$c = 8$			
	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
Mean	4.3634	4.1829	3.6906	1.8443	7.1222	7.0239	6.2018	1.7033	8.4906	8.4580	7.5814	1.7463
Variance	12.8172	12.3028	10.7698	3.9248	14.8792	14.6494	12.4644	0.5587	15.3242	15.2411	12.8154	0.2791
	$p = 0.90$											
	$c = 2$				$c = 5$				$c = 8$			
	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
Mean	7.4186	7.3617	7.1398	5.2403	7.5284	7.6529	8.6017	5.5155	5.9839	6.0417	7.6003	5.6224
Variance	35.1816	33.9612	30.6275	16.6256	54.2988	53.7978	47.8636	3.6299	59.7183	59.8724	60.8869	1.6105

Table 3. Skewness and kurtosis for of TWG distribution for different values of parameters

		$p = 0.5$		$p = 0.75$		$p = 0.9$	
γ	c	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
0.5	2	2.0064	5.1654	1.7293	3.6029	1.4825	2.3921
	5	1.5274	2.8238	1.3642	2.1182	1.4950	2.3362
	8	1.3936	2.2861	1.2650	1.7740	1.7007	2.9779
1	2	2.0527	5.4408	1.7625	3.7659	1.4853	2.4089
	5	1.5366	2.8657	1.3706	2.1434	1.4813	2.2950
	8	1.3955	2.2944	1.2662	1.7788	1.6922	2.9483
1.5	2	2.1959	6.3492	1.8630	4.2905	1.4987	2.4775
	5	1.6154	3.2496	1.4240	2.3692	1.3838	2.0148
	8	1.4452	2.5161	1.2984	1.9086	1.4983	2.3177
2	2	3.1067	14.6030	2.3814	8.1105	1.6503	3.2137
	5	1.6543	1.7569	1.2266	1.4078	1.1567	1.4161
	8	1.6487	1.7182	1.1097	1.0152	1.0684	1.1543

Distribution of range

The distribution of range given:

$$P\left(X_{1:n} = X_{n:n} = y\right) = \left[G\left(y\right) - G\left(y-1\right)\right]$$

for $y = 0, 1...$ If $R = X_{1:n} - X_{n:n}$ is the range of order statistics:

$$P\left(R=0\right) = \frac{1}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^{c}}} \sum_{y=0}^{\infty} \left\{ \mathrm{e}^{-\left[\frac{\left(1-p^{y}\right)}{\gamma}\right]^{c}} - \mathrm{e}^{-\left[\frac{\left(1-p^{y+1}\right)}{\gamma}\right]^{c}} \right\}^{n} \text{ž} \tag{12}$$

Parameter estimation

In this section, point estimation of the parameters is discussed on the basis of observed random sample $X = (x_1, x_2,... x_n)$ of size n. The method of moments and maximum likelihood estimation are used for this purpose.

Method of moments

In this method, the parameter estimates are achieved by solving $E(x) = \mu_1 = m_1$, $E(x^2) = \mu_2 = m_2$, and $E(x^3) = \mu_3 = m_3$, for c, γ, and p. Where m_1, m_2, and m_3 are first, second, and third sample moments:

$$m_r = \frac{\sum_{i=1}^{n} x_i^r}{n}.$$

Since it is not possible to solve them analytically therefore, we solve them numerically by minimizing the quadratic loss function with respect to c, γ, and p:

$$\mathcal{L}(c,\gamma,p)=\sum_{r=1}^{3}(\mu-m_r)^2=\sum_{r=1}^{3}\left\{\sum_{x=0}^{\infty}x^r\,\frac{\mathrm{e}^{-\left[\frac{(1-p^x)}{\gamma}\right]^c}-\mathrm{e}^{-\left[\frac{(1-p^{x+1})}{\gamma}\right]^c}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^c}}-\sum_{x=0}^{n}\frac{x^r}{n}\right\}^2 \tag{13}$$

We utilize the functions *nlm*, *optim*, and *solnp* in *R* environment for the computation.

Method of maximum likelihood estimation

Let a random sample of size n is selected from TWG distribution and the observed frequencies are denoted by n_x, $x = 1, 2,..., k$ where $\sum_{x=1}^{k} n_x = n$, then log-likelihood function is given:

$$l(x\,|\,p,c,\gamma)=\sum_{x=1}^{k}\ln\left[g(x)\right]=\sum_{x=1}^{k}n_x\ln\left(\frac{\mathrm{e}^{-\left[\frac{(1-p^x)}{\gamma}\right]^c}-\mathrm{e}^{-\left[\frac{(1-p^{x+1})}{\gamma}\right]^c}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^c}}\right)=$$

$$=\ln\left(\frac{1-\mathrm{e}^{-\left[\frac{(1-p)}{\gamma}\right]^c}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^c}}\right)+\sum_{x=1}^{k}n_x\ln\left(\frac{\mathrm{e}^{-\left[\frac{(1-p^x)}{\gamma}\right]^c}-\mathrm{e}^{-\left[\frac{(1-p^{x+1})}{\gamma}\right]^c}}{1-\mathrm{e}^{-\left(\frac{1}{\gamma}\right)^c}}\right) \tag{14}$$

To find maximum likelihood estimates of parameters in model, we differentiate eq. (14) with respect to p, c, and γ and obtain:

$$\frac{\partial l}{\partial p}=-\frac{cn_0\mathrm{e}^{-(w)^c}(w)^{c-1}}{\gamma\left[1-\mathrm{e}^{-(w)^c}\right]}+\sum_{x=1}^{k}\frac{n_x\left(\frac{cxp^{x-1}\mathrm{e}^{-(v)^c}(v)^{c-1}}{\gamma}-\frac{c(x+1)p^x\mathrm{e}^{-(z)^c}(z)^{c-1}}{\gamma}\right)}{\mathrm{e}^{-(v)^c}-\mathrm{e}^{-(z)^c}}$$

$$\frac{\partial l}{\partial c}=\frac{n_0\left(1-\mathrm{e}^{-\gamma^{-c}}\right)\left(\frac{\mathrm{e}^{-\gamma^{-c}}\gamma^{-c}\ln(\gamma)\left[1-\mathrm{e}^{-(w)^c}\right]}{\left(1-\mathrm{e}^{-\gamma^{-c}}\right)^2}+\frac{\mathrm{e}^{-(w)^c}(w)^c\ln(w)}{1-\mathrm{e}^{-\gamma^{-c}}}\right)}{1-\mathrm{e}^{-(w)^c}}\cdot$$

$$\cdot\sum_{x=1}^{k}\frac{\left(1-\mathrm{e}^{-\gamma^{-c}}\right)n_x}{\mathrm{e}^{-(v)^c}-\mathrm{e}^{-(z)^c}}\left\{\frac{\mathrm{e}^{-\gamma^{-c}}\gamma^{-c}\ln(\gamma)\left[\mathrm{e}^{-(v)^c}-\mathrm{e}^{-(z)^c}\right]}{\left(1-\mathrm{e}^{-\gamma^{-c}}\right)^2}+\frac{e^{-(z)^c}(z)^c\ln(z)-\mathrm{e}^{-(v)^c}(v)^c\ln(v)}{1-\mathrm{e}^{-\gamma^{-c}}}\right\}$$

$$\frac{\partial l}{\partial \gamma}=\frac{n_0\left(1-\mathrm{e}^{-\gamma^{-c}}\right)\left[\frac{c\mathrm{e}^{-\gamma^{-c}}\gamma^{-c-1}\left(1-\mathrm{e}^{-(w)^c}\right)}{\left(1-\mathrm{e}^{-\gamma^{-c}}\right)^2}-\frac{c(1-p)\mathrm{e}^{-(w)^c}(w)^{c-1}}{\gamma^2\left(1-\mathrm{e}^{-\gamma^{-c}}\right)}\right]}{1-\mathrm{e}^{-(w)^c}}+$$

$$+\sum_{x=1}^{k}\frac{cn_x\left[\left(1-p^x\right)\mathrm{e}^{-(v)^c}(v)^{c-1}-\left(1-p^{x+1}\right)\mathrm{e}^{-(z)^c}(z)^{c-1}\right]}{\gamma^2\left[\mathrm{e}^{-(v)^c}-\mathrm{e}^{-(z)^c}\right]}+\sum_{x=1}^{k}\frac{cn_x\left\{\frac{\mathrm{e}^{-\gamma^{-c}}\gamma^{-c-1}\left[\mathrm{e}^{-(v)^c}-\mathrm{e}^{-(z)^c}\right]}{\left(1-e^{-\gamma^{-c}}\right)}\right\}}{\mathrm{e}^{-(v)^c}-\mathrm{e}^{-(z)^c}}$$

where $w = (1 - p)/\gamma$, $v = (1 - p^x)/\gamma$, and $z = (1 - p^{x+1})/\gamma$, the aforementioned equations should be set to zero and solved simultaneously. Since these equations are non-linear equations therefor some numerical methods are required to solve them.

Simulation study

In this section, we discuss a simulation study to analyze the effectiveness of the method of maximum likelihood for the estimation of model parameters. For this, we use the inversion method to 10000 random samples. Different samples of sizes 50, 250, 500, and 1000 are selected randomly. Table 4 provides means and MSE for TWG distribution for different sample sizes and several combinations of the parameters. It is observed from tab. 4 that estimated values of the parameters become closer to their actual values as the sample size increases. This behavior indicates that the maximum likelihood method is suitable for the estimation of model parameters.

Application

In this section, a numerical study is conducted using real-life data from the field of medical sciences. The data set, considered in this example, comprises of tallies of cysts of kidneys utilizing steroids. This data, also used by [19], was initially collected by [20] during an investigation study the impact of a corticosteroid on cyst formation in mice fetuses in the department of nephro-urology at the institute of child health of University College London. In their study embryonic mouse kidneys were observed, and a random sample of size 110 was exposed to steroids. We apply our proposed model on this data to check the goodness of fit and compare the results with those of discrete Lomax, discrete Burr distribution, zero inflated poisson distribution, poisson distribution, discrete Rayleigh distribution, and geometric distributions. Table 5 represents the observed frequencies of our proposed model along with those of the models in comparison.

Table 6 exhibits the estimated values of the model parameters by using the method of maximum likelihood estimation and the method of moments. The standard errors of the maximum likelihood estimates are also given in parentheses. Moreover negative log-likelihood (ℓ), Akaike information criterion (AIC), and Bayesian information criterion (BIC) are used for comparison purpose. Table 7 shows that TWG distribution has the lowest negative log-likelihood, AIC, and BIC which indicates better performance of the proposed model than that of the other distributions. Figure 3 is the the graphical demonstration of the fitting of the density and distribution function of the proposed model on the data set. From this figure, we can see that the TWG distribution fits adequately to this data set.

Table 4. Simulation study for the TWG distribution

		Mean			MSE	
	Actual values			Actual values		
	c	p	γ	c	p	γ
n	0.5	2	2.5	0.5	2	2.5
	$\hat{c}$	$\hat{p}$	$\hat{\gamma}$	$\hat{c}$	$\hat{p}$	$\hat{\gamma}$
	0.5149	2.0259	2.5371	0.0175	0.0448	0.0587
	0.5033	2.0080	2.5090	0.0056	0.0356	0.0284
	0.5022	2.0046	2.5062	0.0059	0.0247	0.0307
	0.5014	2.0024	2.5025	0.0071	0.0189	0.0105
	Actual values			Actual values		
	c	p	γ	c	p	γ
n	1	1.5	2	1	1.5	2
	$\hat{c}$	$\hat{p}$	$\hat{\gamma}$	$\hat{c}$	$\hat{p}$	$\hat{\gamma}$
	1.0198	1.5193	2.0369	0.0361	0.0420	0.0578
	1.0053	1.5075	2.0095	0.0207	0.0248	0.0289
	1.0024	1.5041	2.0059	0.0093	0.0151	0.0320
	1.0013	1.5020	2.0033	0.0059	0.0125	0.0183
	Actual Values			Actual Values		
	c	p	γ	c	p	γ
n	2	0.5	1	2	0.5	1
	$\hat{c}$	$\hat{p}$	$\hat{\gamma}$	$\hat{c}$	$\hat{p}$	$\hat{\gamma}$
	2.0165	0.5371	1.0405	0.0197	0.0850	0.0688
	2.0052	0.5063	1.0098	0.0186	0.0141	0.0497
	2.0032	0.5041	1.0061	0.0182	0.0169	0.0362
	2.0014	0.5024	1.0029	0.0076	0.0146	0.0162

Table 5. Goodness of fit

X	Observed	Discrete TWG	DBD-XII	Discrete Lomax	ZIP	Poisson	Discrete Raleigh	Geometric
0	65	64.76	63.32	61.89	64.92	27.4	11	45.98
1	14	15.46	18.19	21.01	5.82	38.08	26.83	26.76
2	10	9.02	9.29	9.65	9.52	26.47	29.55	15.57
≤4	10	9.96	9.01	8.41	18.86	16.52	34.72	14.34
>4	11	10.78	10.19	9.03	10.87	1.51	7.9	7.36

Table 6. Estimated parameters of the TWG distribution by the method of maximum likelihood and the method of moments

Parameter	ML estimates (std. error)	MM estimate
γ	3.7558(12.4355)	3.1094
c	0.4589(0.13409)	0.1545
p	0.75896(0.04869)	0.8106

Table 7. Estimated values of ℓ, AIC, and BIC for TWG distribution and for the models in comparison

Criterion	Discrete TWG	DBD-XII	Discrete Lomax	ZIP	Poisson	Discrete Raleigh	Geometric
ℓ	167.2239	170.4806	168.7708	182.2449	246.21	277.778	178.7667
AIC	340.4479	344.9612	343.5415	368.4897	494.42	557.556	359.5333
BIC	348.5493	350.3622	351.6429	373.8907	497.1205	560.2565	362.2338

Conclusion

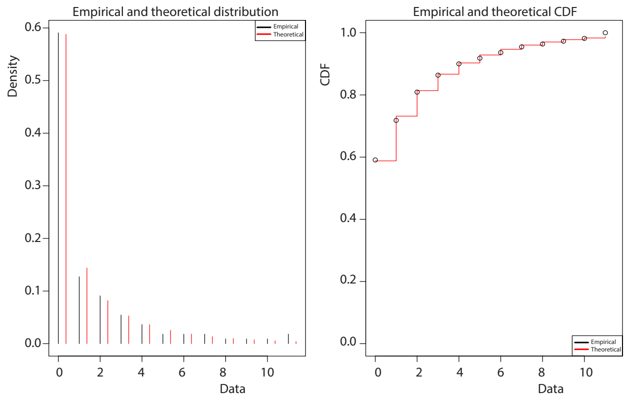

Figure 3. The TWG distribution fitted to counts of cysts of kidneys utilizing steroids

Despite the fact that continuous distributions have been used for modelling continuous data quite effectively and it has been easy to study their distributional properties but for complex situations it is difficult to study the properties of continuous distributions. In this situation discrete analogues of the continuous distributions not only explain distributional properties but also help in developing discrete models. The contribution of the present research is to provide discretization of the method of generating an expanded family of distributions. We construct a new discret TWG distribution by using the proposed method and study several of its properties. The closed form of the TWG distribution discloses the scope of its application in different areas of mathematical statistics. The shapes of hazard rate function such as decreasing and inverse bathtub show its usefulness in different situations. The findings of the simulation study reveal that the parameters estimates of the TWG distribution become closer to the true values whereas the standard errors decrease as the sample sizes increase which show the stability of the model. In the illustrative example, the model parameters estimated by the method of moments and by the method of maximum likelihood estimation have close values which show the adequacy of the model. The values of negative log-likelihood, AIC and BIC provide enough evidence for the better performance of TWG distribution when compared with some contemporary models. We hope by using the proposed discretized method several useful discrete models can be developed to handle the complex situations.

References

[1] Pitman, E. J. G., *Some Basic Theory for Statistical Inference: Monographs on Applied Probability and Statistics*, CRC Press, Boka Raton, Fla., USA, 2018
[2] Bracquemond, C., Gaudoin, O., A Survey on Discrete Lifetime Distributions, *International Journal of Reliability*, *Quality and Safety Engineering*, *10* (2003), 03, pp. 69-98
[3] Dilip, R., The Discrete Normal Distribution, *Communications in Statistics-theory and Methods*, *32* (2003), 03, pp. 1871-1883
[4] Dilip, R., Discrete Rayleigh Distribution, *IEEE Transactions on Reliability*, *53* (2004), 02, pp. 255-260
[5] Inusah, S., Kozubowski, T. J., A Discrete Analogue of the Laplace Distribution, *Journal of Statistical Planning and Inference*, *136* (2006), 03, pp. 1090-1102
[6] Kemp, A. W., Characterizations of a Discrete Normal Distribution, *Journal of Statistical Planning and Inference*, *63* (1997), 02, pp. 223-229

[7] Alzaatreh, A., *et al*., On the Discrete Analogues of Continuous Distributions, *Statistical Methodology*, 9 (2012), 06, pp. 589-603
[8] Alzaatreh, A., *et al*., A New Method for Generating Families of Continuous Distributions, *Metron*, *71* (2013), 1, pp. 63-79
[9] Al-Masoud, T. A., *A Discrete General Class of Continuous Distributions*, M. Sc. thesis, King Abdul Aziz University, Jeddah, Saudi Arabia Kingdom, 2013
[10] Almalki, S. J., Nadarajah, S., A New Discrete Modified Weibull Distribution, *IEEE Transactions on Reliability*, *63* (2014), 1, pp. 68-80
[11] Nekoukhou, V., Bidram, H., The Exponentiated Discrete Weibull Distribution, *SORT*, *39* (2015), 1, pp. 127-146
[12] Hossain, M. S., On a Family of Discrete Probability Distributions (FDPD), *European Journal of Statistics and Probability*, *4* (2016), 5, pp. 1-10
[13] Alamatsaz, M. H., *et al*., Discrete Generalized Rayleigh Distribution, *Pakistan Journal of Statistics*, *32* (2016), 1, pp. 1-20
[14] Chakraborty, S., Generating Discrete Analogues of Continuous Probability Distributions – A Survey of Methods and Constructions, *Journal of Statistical Distributions and Applications*, *2* (2015), 1, 6
[15] Chakraborty, S., Chakravarty, D., A New Discrete Probability Distribution with Integer Support on, *Communications in Statistics-Theory and Methods*, *45* (2016), 2, pp. 492-505
[16] Jayakumar, K., Babu, M. G., Discrete Weibull Geometric Distribution and Its Properties, *Communications in Statistics-Theory and Methods*, *47* (2018), 7, pp. 1767-1783
[17] Mazucheli, J., *et al*., Two Useful Discrete Distributions to Model Overdispersed Count Data, *Revista Colombiana de Estadistica*, *43* (2020), 1, pp. 21-48
[18] Keilson, J., Gerber, H., Some Results for Discrete Unimodality, *Journal of the American Statistical Association*, *66* (1971), 334, pp. 386-389
[19] Para, B. A., Jan, T. R., On Discrete Three Parameter Burr Type xii and Discrete Lomax Distributions and Their Applications to Model Count Data from Medical Science, *Biometrics and Biostatistics International Journal*, *4* (2016), 2, pp. 71-82
[20] Chan, A.-K., *et al*., Corticosteroid-Induced Kidney Dysmorphogenesis is Associated with Deregulated Expression of Known Cystogenic Molecules, as Well as Indian Hedgehog, *American Journal of Physiology-Renal Physiology*, *298* (2009), 2, pp. F346-F356

Paper submitted: May 5, 2020
Paper revised: November 9, 2020
Paper accepted: November 12, 2020