ProbStat Forum, Volume 08, April 2015, Pages 83–94
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

A new discrete family of distributions

S.R. Supanekar[a], D.T. Shirke[b]

[a]*Department of Statistics, Balasaheb Desai College, Patan, Satara, (MS) India* 415206
[b]*Department of Statistics, Shivaji University, Kolhapur, (MS) India* 416004

Abstract. In this paper we introduce a new discrete family of distributions. A scale family of distributions with scale parameter σ is generalized by introducing an additional parameter $\alpha > 0$ using an approach due to Marshall and Olkin (1997). Based on this generalized scale family of distributions a new discrete family of distributions is introduced. The new discrete family of distributions includes discrete exponential, discrete half-logistic, discrete Rayleigh, discrete Weibull and many other discrete distributions. Some distributional properties of the proposed family are reported. Generalized discrete exponential distribution is discussed in detail.

1. Introduction

Most of the lifetime models that have been proposed to represent the lifetime data assume random variables to be continuous. However, sometimes it is impossible or inconvenient to measure the life length of a device on a continuous scale. In practice, we come across situations, where lifetimes are available on a discrete scale. In the recent past, special roles of discrete distributions are getting recognition from survival analysts. In fact, for many equipments and components, lives are being measured by number of completed cycles of operation. Even for a continuous life measure, records made at periodic time points result in a situation where a discrete model may become more appropriate. It is therefore required to develop suitable discrete life distributions to model such situations. The present paper deals with construction of a discrete family of distributions using a suitable family of continuous distributions. In the past, many authors have derived construction of discrete distribution from continuous distribution. Nakagawa and Osaki (1975) obtained discrete Weibull distribution, Roy (2004) analyzed the discrete Rayleigh distribution, Kemp (2008) examined the discrete half-normal distribution, Krishna and Singh (2009) obtained the Burr discrete distribution, Gomez-Deniz (2010) reported a new generalization of the geometric distribution and Gomez-Deniz and Calderin (2011) have reported the discrete Lindley distribution. Other procedures to generate discrete distributions are presented in Rodriguez-Avi et al. (2003, 2004). Nekoukhou et al. (2011) obtained discrete generalized exponential distribution of a second type. Sandhya and Prasanth (2013, 2014) have considered generalisations of geometric and discrete uniform distributions invoking the approach of Marshal and Olkin (1997) while Sandhya and Prasanth (2012) has developed another generalisation of the discrete uniform distribution by adding two parameters to it, generalizing the Marshal-Olkin scheme itself.

In this paper we introduce a new discrete family of distributions, which contains two parameters σ and an additional parameter $\alpha > 0$. A new discrete family of distributions includes discrete exponential, discrete half-normal, discrete half-logistic, discrete Rayleigh, discrete Weibull and many other discrete distributions.

2010 *Mathematics Subject Classification.* Primary 06E05; Secondary 62E99.
Keywords. Continuous scale family, Discrete family of ditributions.
Received: 05 July 2014; Revised: 10 January 2015; Accepted: 23 April 2015
Email addresses: `sup_sanjay@yahoo.com` (S.R. Supanekar), `dts_stats@unishivaji.ac.in` (D.T. Shirke)

In Section 2, we introduce a new discrete family of distributions. Some probabilistic properties and other results are also studied. In Section 3, maximum likelihood estimators of the parameters involved in new discrete family of distributions are discussed. In Section 4, we obtain some discrete distributions, which are members of newly introduced discrete family of distributions. Section 5 is devoted to a study of generalized discrete exponential distribution which is a member of new discrete family of distributions. We also state some properties of generalized discrete exponential distribution using properties of discrete family of distributions in the same section. Maximum likelihood estimators of the parameters involved in this distribution are discussed. Two data sets are used to illustrate the goodness of the proposed model. In the last section conclusions are given.

2. A new discrete family of distributions

A methodology to add a parameter to obtain a new family of distributions is introduced by Marshall and Olkin (1997). They derived a new family of distributions with survival function given by

$$S(t) \equiv S(t;\alpha) = \frac{\alpha \overline{F}(t)}{1-\overline{\alpha}\overline{F}(t)}, \tag{1}$$

where $\alpha > 0$, $\overline{\alpha} = 1-\alpha$ and $\overline{F}(t)$ is a survival function. In particular, when $\alpha = 1$, $S(t) = \overline{F}(t)$. Using this methodology, Marshall and Olkin (1997) constructed a generalization of the exponential and Weibull distributions. Recently, Gomez-Deniz (2010) obtained a new generalization of the geometric distribution by discretizing the generalized exponential distribution of Marshall and Olkin (1997) and Gomez-Deniz et al. (2014) obtained a discrete version of the half-normal distribution and reported its generalization with applications.

Here, we consider the class of continuous scale family of distributions with scale parameter $\sigma > 0$ and we will denote probability density function and the cumulative distribution function (cdf) of the same by $g(\cdot|\sigma)$ and $G(\cdot|\sigma)$ respectively.

The survival function of our new scale family of continuous distributions on the positive real line using (1) is given by

$$S(t;\alpha,\sigma) = \frac{\alpha \overline{G}(t|\sigma)}{1-\overline{\alpha}\overline{G}(t|\sigma)}, \tag{2}$$

where $\overline{G}(t|\sigma) = 1 - G(t|\sigma)$. The survival function (2) can be considered as a generalization of the scale family of distributions and we will write the corresponding family of distributions by $GSF(\alpha,\sigma)$.

Let X be a discrete random variable associated to a continuous random variable belonging to $GSF(\alpha,\sigma)$. The probability mass function (pmf) is given by

$$P(X = x;\alpha,\sigma) = p_x = S(x;\alpha,\sigma) - S(x+1;\alpha,\sigma) = \frac{\alpha[\overline{G}(x|\sigma) - \overline{G}((x+1)|\sigma)]}{(1-\overline{\alpha}\overline{G}(x|\sigma))(1-\overline{\alpha}\overline{G}((x+1)|\sigma))}, \quad x = 0,1,2,\ldots \tag{3}$$

where $\alpha > 0$, $\overline{\alpha} = 1-\alpha$ and $\sigma > 0$. We denote this distribution by $X \sim DF(\alpha,\sigma)$.

Lemma 2.1. *The cdf of a discrete random variable having the pmf* (3) *is given by,*

$$F(x;\alpha,\sigma) = \frac{1-\overline{G}((x+1)|\sigma)}{1-\overline{\alpha}\overline{G}((x+1)|\sigma)}, \quad x = 0,1,2,\ldots$$

and

$$P(X \geq x;\alpha,\sigma) = \frac{\alpha \overline{G}(x|\sigma)}{1-\overline{\alpha}\overline{G}(x|\sigma)}, \quad x = 0,1,2,\ldots$$

Proof. It is straightforward. □

Remark 2.2. *Hazard rate function corresponding to $DF(\alpha, \sigma)$ is given by*

$$r(x; \alpha, \sigma) = \frac{\overline{G}(x|\sigma) - \overline{G}((x+1)|\sigma)}{\overline{G}(x|\sigma))(1 - \overline{\alpha}\overline{G}((x+1)|\sigma))}.$$

Lemma 2.3. *The probability generating function (pgf) of a discrete random variable having the DF(α, σ) in equation (3) is given by,*

$$P_x(s) = 1 + \alpha(s-1)\sum_{x=1}^{\infty} s^{x-1}\frac{\overline{G}(x|\sigma)}{1 - \overline{\alpha}\overline{G}(x|\sigma)}.$$

Proof. Omitted for brevity. □

Lemma 2.4. *Mean and variance of random variable having DF(α, σ) are respectively*

$$\mu(\alpha, \sigma) = E(X) = \alpha\sum_{x=1}^{\infty}\frac{\overline{G}(x|\sigma)}{1 - \overline{\alpha}\overline{G}(x|\sigma)},$$

$$V(X) = \alpha\sum_{x=1}^{\infty}(2x-1)\frac{\overline{G}(x|\sigma)}{1 - \overline{\alpha}\overline{G}(x|\sigma)} - \left(\alpha\sum_{x=1}^{\infty}\frac{\overline{G}(x|\sigma)}{1 - \overline{\alpha}\overline{G}(x|\sigma)}\right)^2.$$

Proof. It is easy to obtain using Lemma 2.3. □

Lemma 2.5. *The recurrence relation for generating probabilities of DF(α, σ) is given by*

$$p_{x+1} = \frac{\overline{G}((x+1)|\sigma) - \overline{G}((x+2)|\sigma)}{\overline{G}(x|\sigma) - \overline{G}((x+1)|\sigma)} \cdot \frac{1 - \overline{\alpha}\overline{G}(x|\sigma)}{1 - \overline{\alpha}\overline{G}((x+2)|\sigma)}p_x, \quad x = 0, 1, 2, \ldots$$

where $p_0 = \frac{1 - \overline{G}(1|\sigma)}{1 - \overline{\alpha}\overline{G}(1|\sigma)}$.

Proof. It is straightforward. □

Lemma 2.6. *Quantile x_v and median $x_{0.5}$ of DF(α, σ) are respectively given by*

$$x_v = \left[\sigma G^{-1}\left(\frac{\alpha v}{1 - \overline{\alpha}v}\right) + 1\right],$$

$$x_{0.5} = \left[\sigma G^{-1}\left(\frac{\alpha}{1+\alpha}\right) - 1\right],$$

where $[\cdot]$ denotes the integer part.

Proof. The above expressions can be obtained easily using Lemma 2.1. □

Theorem 2.7. *If $\frac{\partial\overline{G}(x|\sigma)}{\partial\sigma} > 0$ then the mean is monotonic increasing function of both parameters α and σ.*

Proof. We note that

$$\frac{\partial\mu(\alpha, \sigma)}{\partial\alpha} = \sum_{x=1}^{\infty}\frac{\overline{G}(x|\sigma)(1 - \overline{G}(x|\sigma))}{(1 - \overline{\alpha}\overline{G}(x|\sigma))^2} > 0,$$

and

$$\frac{\partial\mu(\alpha, \sigma)}{\partial\sigma} = \sum_{x=1}^{\infty}\frac{\alpha\overline{G}'(x/\sigma)}{(1 - \overline{\alpha}\overline{G}(x|\sigma))^2} > 0,$$

if $\overline{G}'(x|\sigma) = \frac{\partial\overline{G}(x|\sigma)}{\partial\sigma} > 0$. Hence the proof. □

In the following we report some results which are consequence of methodology due to Marshall and Olkin (1997).

Lemma 2.8. *Let* $S_1(t) = \frac{\alpha\overline{F}(t)}{1-\overline{\alpha}\ \overline{F}(t)}$ *and* $S_2(t) = \frac{\beta\overline{G}(t)}{1-\overline{\beta}\ \overline{G}(t)}$. *Then* $S_1(t) = S_2(t)$ *if and only if* $\overline{G}(t) = \frac{k\overline{F}(t)}{1-\overline{k}\ \overline{F}(t)}$, *where* $k = \frac{\alpha}{\beta}$, $\overline{\beta} = 1 - \beta$ *and* $\overline{k} = 1 - k$.

Proof. Suppose, $S_1(t) = S_2(t)$ then

$$\frac{\overline{G}(t)F(t)}{\overline{F}(t)G(t)} = \frac{\alpha}{\beta} = k. \ (say)$$

Hence we get,

$$\overline{G}(t) = \frac{k\overline{F}(t)}{1-\overline{k}\ \overline{F}(t)}.$$

Now suppose

$$\overline{G}(t) = \frac{k\overline{F}(t)}{1-\overline{k}\ \overline{F}(t)} \quad then \quad \overline{G}(t)F(t) = k\overline{F}(t)G(t).$$

Letting $k = \alpha/\beta$, we have that

$$\begin{aligned}
\alpha\overline{F}(t) - \beta\overline{G}(t) &= (\alpha-\beta)\overline{F}(t)\overline{G}(t), \\
\alpha\overline{F}(t) - \beta\overline{G}(t) &= (\alpha-\alpha\beta+\alpha\beta-\beta)\overline{F}(t)\overline{G}(t), \\
\frac{\alpha\overline{F}(t)}{1-\overline{\alpha}\overline{F}(t)} &= \frac{\beta\overline{G}(t)}{1-\overline{\beta}\ \overline{G}(t)}.
\end{aligned}$$

Hence the converse holds. □

Lemma 2.9.

$$S_i(t) = \frac{\left(\prod_{j=1}^{i}\alpha_j\right)S_0(t)}{1-(1-\prod_{j=1}^{i}\alpha_j)S_0(t)}, \quad i = 1, 2, 3, \ldots$$

where $S_0(t)$ *is a survival function and* $\alpha_j > 0$.

Proof. We prove this theorem by method of induction. Let survival function of a new family of distributions by Marshall and Olkin (1997) be given by

$$S_1(t) = \frac{\alpha_1 S_0(t)}{1-(1-\alpha_1)S_0(t)}, \tag{4}$$

where $S_0(t)$ is a survival function.

Using equation (4) we can obtain a new survival function $S_2(t)$ by introducing parameter α_2 as

$$S_2(t) = \frac{\alpha_2 S_1(t)}{1-(1-\alpha_2)S_1(t)}.$$

Using equation (4) in the RHS of above equation we get,

$$S_2(t) = \frac{\alpha_1\alpha_2 S_0(t)}{1-(1-\alpha_1\alpha_2)S_0(t)}.$$

Hence result holds for $i = 2$.

Now suppose result holds for $i = k$. Hence,

$$S_k(t) = \frac{\left(\prod_{j=1}^{k} \alpha_j\right) S_0(t)}{1 - (1 - \prod_{j=1}^{k} \alpha_j) S_0(t)}. \tag{5}$$

Here $S_k(t)$ is a survival function and hence again using equation (4) we can obtain,

$$S_{k+1}(t) = \frac{\alpha_{k+1} S_k(t)}{1 - (1 - \alpha_{k+1}) S_k(t)},$$

where α_{k+1} is newly introduced parameter.

Using an equation (5) in $S_{k+1}(t)$ we get,

$$S_{k+1}(t) = \frac{\left(\prod_{j=1}^{k+1} \alpha_j\right) S_0(t)}{1 - (1 - \prod_{j=1}^{k+1} \alpha_j) S_0(t)}.$$

This shows that the result also holds for $i = k + 1$. Hence the proof. $\square$

Lemma 2.10. *Let $\overline{F}(\cdot)$ and $\overline{G}(\cdot)$ are the survival functions of two distributions. If $\overline{G} = \frac{2\overline{F}}{1+\overline{F}}$ then a new distribution obtained using $\overline{G}(\cdot)$ with adding parameter α by method of Marshall and Olkin* (1997) *is the same as that obtained by using $\overline{F}(\cdot)$ with additional parameter β, where $\beta = 2\alpha$.*

Proof. Let $S_1(t)$ and $S_2(t)$ are the survival functions of new distributions obtained using method of Marshall and Olkin (1997) by adding parameters α and β respectively to $\overline{F}(\cdot)$ and $\overline{G}(\cdot)$ we get,

$$S_1(t) = \frac{\alpha \overline{G}(\cdot)}{1 - \overline{\alpha}\overline{G}(\cdot)} \quad and \quad S_2(t) = \frac{\beta \overline{F}(\cdot)}{1 - \overline{\beta}\ \overline{F}(\cdot)}.$$

Now putting $\overline{G} = \frac{2\overline{F}}{1+\overline{F}}$ in $S_1(t)$ we get,

$$S_1(t) = \frac{2\alpha \overline{F}(\cdot)}{1 - (1 - 2\alpha)\overline{F}(\cdot)}.$$

The proof follows by replacing 2α by β in above expression. $\square$

Using above Lemma 2.10 we get some interesting situation, which is given in the following.

Let $\overline{G}(t) = \frac{2e^{-t/\sigma}}{1+e^{-t/\sigma}}$ and $\overline{F}(t) = e^{-t/\sigma}$. Here $\overline{G}$ and $\overline{F}$ are the survival functions of half-logistic distribution and exponential distribution respectively and $\overline{G} = \frac{2\overline{F}}{1+\overline{F}}$. Hence the survival functions of new distributions by adding parameters α and β respectively to $\overline{G}$ and $\overline{F}$ are given by,

$$S_1(t) = \frac{\alpha \overline{G}(t)}{1 - \overline{\alpha}\overline{G}(t)} = \frac{2\alpha e^{-t/\sigma}}{1 - (1 - 2\alpha)e^{-t/\sigma}},$$

and

$$S_2(t) = \frac{\beta \overline{F}(t)}{1 - \overline{\beta}\ \overline{F}(t)} = \frac{\beta e^{-t/\sigma}}{1 - \overline{\beta} e^{-t/\sigma}}.$$

This shows that $S_1(t) = S_2(t)$, if $\beta = 2\alpha$. Thus half-logistic distribution and exponential distribution lead to similar distributions by following method of Marshall and Olkin (1997).

In the following we discuss estimation of α and β for $DF(\alpha, \sigma)$.

3. Estimation

Population moments for $DF(\alpha,\sigma)$ are in the form of an infinite series. So these equations cannot be solved easily to get estimators based on method of moments.

So here we consider method of maximum likelihood estimation (MLE). Suppose $x_1, x_2, \ldots, x_n$ is a random sample of size n obtained from $DF(\alpha,\sigma)$. The log-likelihood function is given by

$$\log L = n\log\alpha + \sum_{i=1}^{n}\log(\overline{G}(x_i|\sigma) - \overline{G}((x_i+1)|\sigma)) - \sum_{i=1}^{n}\log(1-\overline{\alpha}\overline{G}(x_i|\sigma)) - \sum_{i=1}^{n}\log(1-\overline{\alpha}\overline{G}((x_i+1)|\sigma)).$$

Hence the likelihood equations are

$$\frac{\partial \log L}{\partial\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n}\frac{\overline{G}_x}{1-\overline{\alpha}\overline{G}_x} - \sum_{i=1}^{n}\frac{\overline{G}_{x+1}}{1-\overline{\alpha}\overline{G}_{x+1}}, \tag{6}$$

$$\frac{\partial \log L}{\partial\sigma} = \sum_{i=1}^{n}\frac{\overline{G}'_x - \overline{G}'_{x+1}}{\overline{G}_x - \overline{G}_{x+1}} + \overline{\alpha}\sum_{i=1}^{n}\frac{\overline{G}'_x}{1-\overline{\alpha}\overline{G}_x} - \overline{\alpha}\sum_{i=1}^{n}\frac{\overline{G}'_{x+1}}{1-\overline{\alpha}\overline{G}_{x+1}}, \tag{7}$$

where $\overline{G}_x = \overline{G}(x_i|\sigma)$, $\overline{G}'_x = \frac{\partial\overline{G}(x_i|\sigma)}{\partial\sigma}$.

Solving equations (6) and (7) simultaneously, we get maximum likelihood estimators $\hat{\alpha}$ and $\hat{\sigma}$ of α and σ respectively. MLE's can be obtained by numerical method.

The second order partial derivatives are given below.

$$\frac{\partial^2 \log L}{\partial\alpha^2} = -\frac{n}{\alpha^2} + \sum_{i=1}^{n}\frac{(\overline{G}_x)^2}{(1-\overline{\alpha}\overline{G}_x)^2} + \sum_{i=1}^{n}\frac{(\overline{G}_{x+1})^2}{(1-\overline{\alpha}\overline{G}_{x+1})^2},$$

$$\frac{\partial \log L}{\partial\alpha\partial\sigma} = -\sum_{i=1}^{n}\frac{\overline{G}'_x}{(1-\overline{\alpha}\overline{G}_x)^2} - \sum_{i=1}^{n}\frac{\overline{G}'_{x+1}}{(1-\overline{\alpha}\overline{G}_{x+1})^2},$$

$$\begin{aligned}\frac{\partial^2 \log L}{\partial\sigma^2} &= \sum_{i=1}^{n}\frac{(\overline{G}_x - \overline{G}_{x+1})(\overline{G}''_x - \overline{G}''_{x+1}) - (\overline{G}'_x - \overline{G}'_{x+1})^2}{(\overline{G}_x - \overline{G}_{x+1})^2}\\ &\quad + \overline{\alpha}\sum_{i=1}^{n}\frac{\overline{G}''_x - \overline{\alpha}\overline{G}_x\overline{G}''_x + \overline{\alpha}(\overline{G}'_x)^2}{(1-\overline{\alpha}\overline{G}_x)^2} + \overline{\alpha}\sum_{i=1}^{n}\frac{\overline{G}''_{x+1} - \overline{\alpha}\overline{G}_{x+1}\overline{G}''_{x+1} + \overline{\alpha}(\overline{G}'_{x+1})^2}{(1-\overline{\alpha}\overline{G}_{x+1})^2},\end{aligned}$$

where $\overline{G}''_x = \frac{\partial^2\overline{G}(x_i|\sigma)}{\partial\sigma^2}$.

The Fisher information matrix can be estimated by using the following approximations

$$E\left(-\frac{\partial^2 \log L}{\partial\alpha^2}\right) \approx -\left.\frac{\partial^2 \log L}{\partial\alpha^2}\right|_{(\hat{\alpha},\hat{\sigma})},$$

$$E\left(-\frac{\partial^2 \log L}{\partial\alpha\partial\sigma}\right) \approx -\left.\frac{\partial^2 \log L}{\partial\alpha\partial\sigma}\right|_{(\hat{\alpha},\hat{\sigma})},$$

$$E\left(-\frac{\partial^2 \log L}{\partial\sigma^2}\right) \approx -\left.\frac{\partial^2 \log L}{\partial\sigma^2}\right|_{(\hat{\alpha},\hat{\sigma})}.$$

4. Some members of new discrete family of distributions

4.1. Discrete exponential distribution

The cdf and survival function of exponential distribution with scale parameter σ are respectively as follows

$$G(x|\sigma) = 1 - e^{-x/\sigma} \quad \text{and} \quad \overline{G}(x|\sigma) = e^{-x/\sigma}.$$

Now the pmf of new distribution using equation (3) is given by

$$p_x = \frac{\alpha[e^{-x/\sigma} - e^{-(x+1)/\sigma}]}{(1-\overline{\alpha}e^{-x/\sigma})(1-\overline{\alpha}e^{-(x+1)/\sigma})}, \quad x = 0, 1, 2, \dots$$

This distribution is obtained and studied by Gomez-Deniz (2010).

4.2. Discrete half-normal distribution

Let $\Phi_\sigma(x)$ be the cdf of normal distribution with mean zero and standard deviation $\sigma > 0$. Then survival function of generalized half-normal distribution is

$$S(x; \alpha, \sigma) = \frac{(1+\alpha)\overline{\Phi}_\sigma(x)}{1-\overline{\alpha}\overline{\Phi}_\sigma(x)}.$$

The pmf of generalized discrete half-normal distribution GDHN(α, σ) using equation (3) is given by

$$p_x = \frac{(1+\alpha)[\Phi_\sigma(x+1) - \Phi_\sigma(x)]}{(1-\overline{\alpha}\overline{\Phi}_\sigma(x))(1-\overline{\alpha}\overline{\Phi}_\sigma(x+1))}, \quad x = 0, 1, 2, \dots$$

This distribution is obtained and studied by Gomez-Deniz et al. (2014).

We can obtain discrete half-logistic, discrete Rayleigh, discrete Weibull distributions as members of newly defined discrete family of distributions. In the following we study a discrete generalized exponential distribution in detail.

5. A generalized discrete exponential distribution:

Let $Y_1, Y_2, \dots, Y_\beta$ be independent and identically distributed exponential random variables with mean σ. Define $X = \max(Y_i)$ then survival function of X is

$$\overline{G}(x; \sigma) = 1 - (1 - e^{-x/\sigma})^\beta, \quad x > 0.$$

Let $q = e^{-1/\sigma}$ and $\beta = 2$ then the survival function can be written as

$$\overline{G}(x) = 2q^x - q^{2x}, \quad x > 0.$$

Then the pmf of generalized discrete exponential distribution using (3) is

$$p_x = \frac{2\alpha p q^x - \alpha p(1+q)q^{2x}}{(1-\overline{\alpha}(2q^x - q^{2x}))(1-\overline{\alpha}(2q^{x+1} - q^{2(x+1)})}, \quad x = 0, 1, 2, \dots \tag{8}$$

where $p = 1 - q$.

A distribution having the pmf given in (8) is denoted by GDE(α, q), which is a member of $DF(\alpha, \sigma)$.

Now we study some properties of GDE(α, q). Using Lemmas 2.1, 2.3, 2.4, 2.5 and 2.6 some results related to GDE(α, q) are given in the following.

a) The cdf and survival function:

$$F(x; \alpha, q) = \frac{1 - q^{x+1}(2 - q^{x+1})}{1 - \overline{\alpha}q^{x+1}(2 - q^{x+1})}, \quad x = 0, 1, 2, \dots$$

$$P(X \geq x; \alpha, q) = \frac{\alpha q^x(2 - q^x)}{1 - \overline{\alpha}q^x(2 - q^x)}, \quad x = 0, 1, 2, \dots$$

b) The probability generating function (pgf):

$$P_x(s) = 1 + \alpha(s-1)\sum_{x=1}^{\infty}(s^{x-1})\frac{2q^x - q^{2x}}{1-\overline{\alpha}(2q^x - q^{2x})}.$$

c) Mean and variance:

$$E(X) = \alpha\sum_{x=1}^{\infty}\frac{2q^x - q^{2x}}{1-\overline{\alpha}(2q^x - q^{2x})}$$

$$V(X) = \alpha\sum_{x=1}^{\infty}(2x-1)\frac{2q^x - q^{2x}}{1-\overline{\alpha}(2q^x - q^{2x})} - \left[\alpha\sum_{x=1}^{\infty}\frac{2q^x - q^{2x}}{1-\overline{\alpha}(2q^x - q^{2x})}\right]^2.$$

d) The recurrence relation for probabilities:

$$p_{x+1} = \frac{\left[(1-q^{x+2})^2 - (1-q^{x+1})^2\right]\left[1-\overline{\alpha}(2q^x - q^{2x})\right]}{\left[(1-q^{x+1})^2 - (1-q^x)^2\right]\left[1-\overline{\alpha}(2q^{x+2} - q^{2x+4})\right]}p_x\ , \quad x = 0,1,2,\ldots$$

where $p_0 = \frac{1-(2q-q^2)}{1-\overline{\alpha}(2q-q^2)}$.

e) Quantile (x_v) and median ($x_{0.5}$)

$$x_v = \left[\frac{\log\left(1-\sqrt{\frac{\alpha v}{1-\overline{\alpha}v}}\right)}{\log q} - 1\right],$$

$$x_{0.5} = \left[\frac{\log\left(1-\sqrt{(\frac{\alpha}{1+\alpha}}\right)}{\log q} - 1\right],$$

where $[\cdot]$ denotes the integer part.

Here $\overline{G}(x) = 2q^x - q^{2x}$ and

$$\frac{\partial\overline{G}(x)}{\partial q} = 2xq^{x-1} - 2xq^{2x-1} = 2xq^{x-1}(1-q^x) > 0.$$

Hence by Theorem 2.7, The mean increases with respect to each of parameters α and q.

Mean and the variance of the GDE(α, q) are calculated for various values of α and q. These are shown in Table 1. It is observed that, for $0 < q < 0.5$ and $\alpha > 1$, the distribution is underdispersed and the mean increases faster than the variance. Hence, the parameter can be adjusted to suit most of the data sets.

Pmf of GDE(α, q) for various values of α and q is plotted and shown in Figure 1. In the Figure 1, we see that if α is large then the mode moves to the right, indicating that for large α, distribution attains symmetry and that for the small values of α, the distribution is positively skewed.

5.1. Maximum likelihood estimation of α and q

Suppose $x_1, x_2, \ldots, x_n$ is a random sample of size n obtained from GDE(α, q). The log-likelihood function is given by

$$\log L = n\log\alpha + \sum_{i=1}^{n}\log(2q^{x_i} - q^{2x_i} - 2q^{(x_i+1)} + q^{2(x_i+1)})$$
$$- \sum_{i=1}^{n}\log(1-\overline{\alpha}(2q^{x_i} - q^{2x_i})) - \sum_{i=1}^{n}\log(1-\overline{\alpha}(2q^{(x_i+1)} - q^{2(x_i+1)})).$$

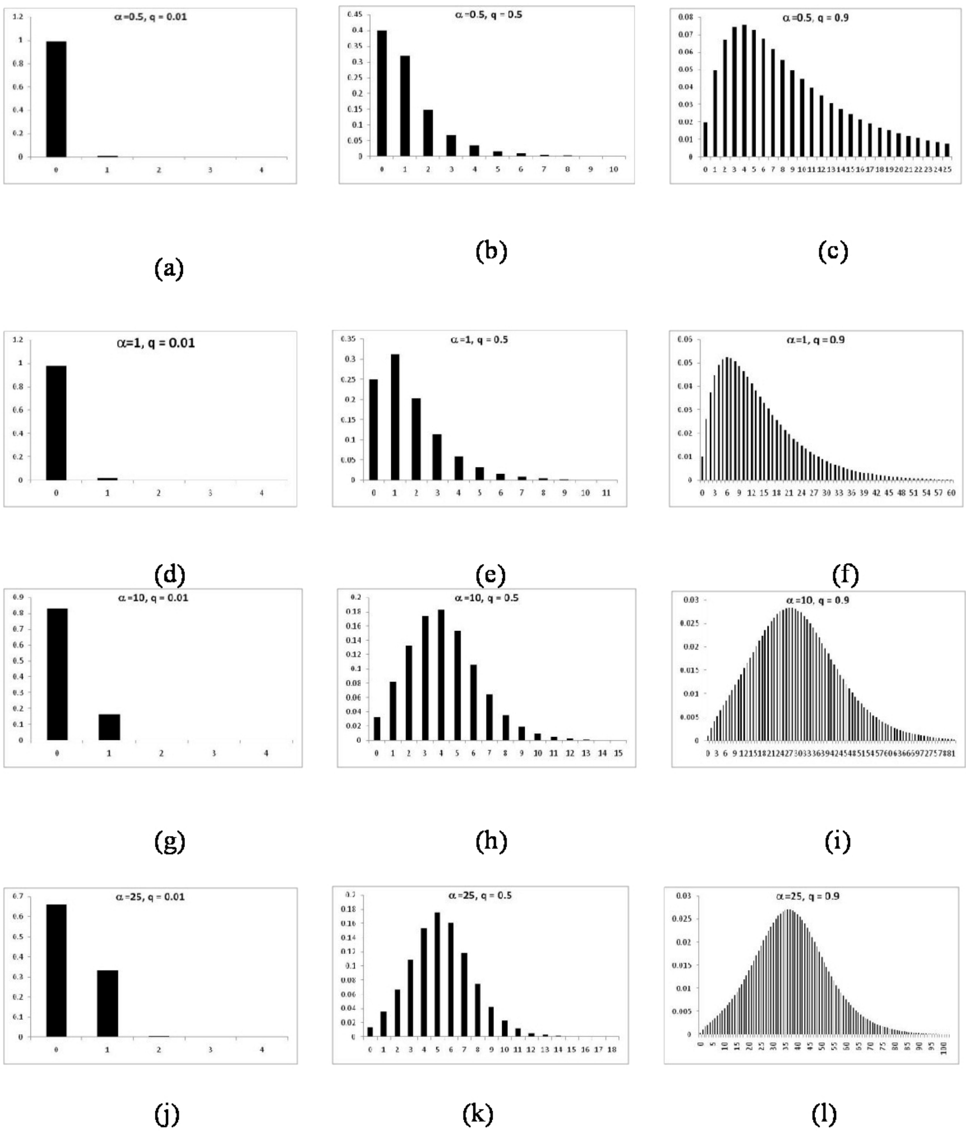

Figure 1: Probability mass function of GDE(α, q) for different values of the α and q.

Table 1: Mean and variance (in parentheses) of the generalized discrete exponential distribution for varying α and q.

α/q	0.1	0.25	0.5	0.75	0.9
0.1	0.0252 (0.0295)	0.09 (0.1288)	0.3591 (0.7028)	1.397 (4.4067)	4.6657 (32.5559)
1	0.2121 (0.2163)	0.6 (0.6756)	1.6667 (2.6666)	4.7143 (15.1835)	13.7353 (112.5444)
2	0.3628 (0.3279)	0.9055 (0.8962)	2.2994 (3.4383)	6.2423 (19.6143)	17.9062 (145.4489)
5	0.643 (0.4604)	1.392 (1.1608)	3.2802 (4.4367)	8.6071 (25.3811)	24.3589 (188.1366)
10	0.8917 (0.5265)	1.8065 (1.3314)	4.1111 (5.0998)	10.6096 (29.2197)	29.8192 (216.3557)

Hence the likelihood equations are

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \frac{2q^{x_i} - q^{2x_i}}{1-\overline{\alpha}(2q^{x_i} - q^{2x_i})} - \sum_{i=1}^{n} \frac{2q^{(x_i+1)} - q^{2(x_i+1)}}{1-\overline{\alpha}(2q^{(x_i+1)} - q^{2(x_i+1)})}, \tag{9}$$

$$\begin{aligned}\frac{\partial \log L}{\partial q} &= 2\sum_{i=1}^{n} \frac{x_i q^{x_i-1}(1-q^{x_i}) - (x_i+1)q^{x_i}(1-q^{x_i+1})}{(1-q^{x_i+1})^2 - (1-q^{x_i})^2} \\ &\quad + \overline{\alpha}\sum_{i=1}^{n} \frac{x_i q^{x_i-1}(1-q^{x_i})}{1-\overline{\alpha}(2q^{x_i} - q^{2x_i})} + \overline{\alpha}\sum_{i=1}^{n} \frac{(x_i+1)q^{x_i}(1-q^{x_i+1})}{1-\overline{\alpha}(2q^{x_i+1} - q^{2x_i+2})}.\end{aligned} \tag{10}$$

Solving equations (9) and (10) simultaneously, we get maximum likelihood estimators of α and q.

5.2. Examples

In this section, we present data sets to examine the fitting of a newly proposed generalized discrete exponential distribution. The data set presented in Table 2 gives fish catch data (Kemp, 1992) and in Table 3 gives results of ten shots fired from a rifle at each of 100 targets (Consul and Jain, 1973).

Table 2 and Table 3 provide the expected frequencies and corresponding chi-square values of newly introduced generalized discrete exponential distribution and compared with values of generalized geometric distribution (GG) calculated by Gomez-Deniz et al. (2010).

From these Table 2 and Table 3 we observe that GDE(α, q) is a good fit for data sets considered here.

6. Conclusions

In this paper we have introduced a new discrete family of distributions. This family includes discrete exponential, discrete half logistic, discrete Rayleigh, discrete Weibull and many other discrete distributions. We have studied distributional properties and estimation of parameters of the new discrete family of distributions. For illustration we studied generalized discrete exponential distribution. The additional parameter α plays an important role in this newly introduced discrete family of distributions. We can use these models for real life situations.

The newly introduced discrete family of distributions is suitable for modeling discrete data and is a better alternative to some other existing distributions.

Table 2: Fish catch data Kemp (1992) counts of the number of European red mites on apple Leaves.

Number of mites per leaf	Observed frequency	Expected frequency (GDE)	Expected frequency (GG)
0	1	0.74	0.86
1	2	2.93	2.89
2	11	8.95	8.84
3	20	21.22	21.14
4	29	30.15	30.24
5	23	22.01	22.08
6	10	9.5	9.47
7	3	3.16	3.12
8	1	0.95	0.93
≥ 9	0	0.4	0.36
$(\hat{\alpha}, \hat{q})$		(136.767, 0.29)	(283.51,0.28)
χ^2		0.83 (4 d.f.)	0.89 (4 d.f.)
$(p-value, L_{max})$		(0.933, -177.668)	(0.92, -177.722)

$(\bar{x} = 4.04; s^2 = 2.06)$

Table 3: Results of ten shots fired from a rifle at each of 100 targets.

Number of hits	Observed frequency	Expected frequency (GDE)	Expected frequency (GG)
0	0	1.20	0.49
1	2	2.33	1.45
2	4	4.73	4.07
3	10	9.68	10.33
4	22	18.15	20.61
5	26	25.67	26.65
6	18	22.21	20.26
7	12	11.15	10.04
8	4	3.69	3.94
9	2	0.93	1.40
10	0	0.19	0.48
$(\hat{\alpha}, \hat{q})$		(191.381, 0.338)	(394.75, 0.33)
χ^2		0.73 (4 d.f.)	0.76 (4 d.f.)
$(p-value, L_{max})$		(0.946, -191.771)	(0.94, -190.868)

$(\bar{x} = 5; s^2 = 2.64)$

Acknowledgement

Authors thank to referee for valuable suggestions. The first author would like to acknowledge the financial support of University Grants Commission, New Delhi, India through a minor research project. The second author wishes to thank Department of Science and Technology, New Delhi, India for financial support through DST-PURSE scheme to Shivaji University, Kolhapur, India.

References

[1] Consul, P.C., Jain, G.C. (1973) A generalization of the Poisson distribution, Technometrics 15(4), 791–799.

[2] Gomez-Deniz, E. (2010) Another generalization of the geometric distribution, Test 19(2), 399–415.

[3] Gomez-Deniz, E., Calderin, E. (2011) The discrete lindley distribution: properties and applications, J. Stat. Comput. Simul. 81(11), 1405–1416.

[4] Gomez-Deniz, E., Vazquez-Polo, F.J., Garcia-Garcia, V. (2014) A discrete version of the half- normal distribution and its generalization with applications, Stat. Papers 55, 497–511.

[5] Kemp, A.W. (1992). Heine-Euler extensions of the Poisson distribution, Commun. Stat. Theory Methods 21(3), 571–588.

[6] Kemp, A.W. (2008) The discrete half-normal distribution. In: Advances in mathematical and statistical modeling, Birkhauser, Basel, 353–365.

[7] Krishna, H., Singh, P. (2009) Discrete Burr and discrete Pareto distributions, Stat. Methodol. 6, 177–188.

[8] Marshall, A.W., Olkin, I. (1997) A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika 84(3), 641–652.

[9] Nakagawa, T., Osaki, S. (1975) The discrete Weibull distribution, IEE Trans Reliab 24(5), 300–301.

[10] Nekoukhou, V., Alamatsaz, M.H., Bidram, H. (2011) Discrete generalized exponential distribution of a second type, Statistics 47(4), 876–887, DOI: 10.1080/02331888.2011.633707.

[11] Rodriguez-Avi, J., Conde-Sanchez, A., Saez-Castillo, A.J. (2003) A new class of discrete distributions with complex parameters, Stat. Papers 44, 67–88.

[12] Rodriguez-Avi, J., Conde-Sanchez, A., Saez-Castillo, A.J., Olmo-Jimanez, M.J. (2004) A triparametric discrete distributions with complex parameters, Stat. Papers 445, 81–95

[13] Roy, D. (2004) Discrete Rayleigh distribution, IEEE Trans Reliab 53(2), 255–260.

[14] Sandhya, E., Prasanth, C.B. (2012) Harris Discrete Uniform Distribution, Proceedings of World Congress on Research and Innovations, 17–20 December 2013, St. Josephs College, Thrissur. Excel India Publications.

[15] Sandhya, E., Prasanth, C.B. (2013) Marshall-Olkin Geometric distribution, Frontiers of Statistics and its Applications, Bonfring Publications, 261—269.

[16] Sandhya, E., Prasanth, C.B. (2014) Marshall-Olkin discrete uniform distribution, Journal of Probability, Article ID 979312.