Journal of Statistical Modelling: Theory and Applications
Vol. 1, No. 2, 2020, pp. 153-166
Yazd University Press 2020

JSMTA

Research Paper

A new discrete distribution based on geometric odds ratio

Vahid Nekoukhou*[1], Hamid Bidram[2]
[1]Department of Statistics, University of Khansar, Khansar, Iran.
[2]Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan, Iran.

Received: March 4, 2020/ Revised: June 12, 2020 / Accepted: June 12, 2020

Abstract: In this paper, a new discrete distribution is studied based on geometric odds ratio. This new distribution has three parameters and can be a unimodal or a bimodal discrete distribution. Some important distributional properties are studied. For example, moments, the behaviour of the hazard rate function, stochastic orders, mixing processes, infinite divisibility, Rényi and Shannon entropies and the distributions of order statistics are investigated. We will see that the hazard rate function of the new discrete distribution can be monotonically increasing and decreasing and bathtub-shaped. The parameters of the distribution are estimated by the maximum likelihood method, and a real data set is analyzed in order to show the effectiveness of the model.

Keywords: Discrete odds ratio; Entropy; Geometric distribution; Infinite divisibility; Order statistics; Stochastic orders; Stress-strength parameter.

Mathematics Subject Classification (2010): 62F10; 62H10

1 Introduction

Data may be discrete by nature and discrete data occur frequently in practice in various fields. For example, the stress pattern in a step-stress accelerated life test can be treated as a discrete random variable and the number of attempts needed to crack a password are discrete in nature. In these cases it is better to deal with these data by a discrete probability model.

Several attempts have been made to introduce different discrete probability distributions and to develop and study their properties. For example, Gómez-Déniz (2010) studied the generalized geometric (GG) distribution whose hazard rate function is

*Corresponding author: v.nekoukhou@gmail.com

monotone. Nekoukhou et al. (2013) introduced originally the discrete generalized exponential (DGE) distribution, which can be considered as the discrete analog of the well-known absolute continuous generalized exponential distribution of Gupta and Kundu (1999). The hazard rate function of the DGE distribution can be increasing, decreasing, or constant. So, the geometric distribution is a special case of the DGE distribution. Nekoukhou and Bidram (2015) studied the exponentiated discrete Weibull distribution, which is a generalization of the two-parameter discrete Weibull distribution of Nakagawa and Osaki (1975).

The odds function is an important quantity in different aspects of Statistics. Specially, serious discussions are propounded in Distribution Theory and Reliability. For example, Sankaran and Jayakumar (2008) gave some physical interpretations of the Marshall-Olkin family of distributions using odds function. Gupta (2011) considered a class of bivariate distributions by forming the odds of failure of a two-component system. Unnikrishnan Nair and Sankaran (2015) considered the odds function in a discrete setup.

Suppose that $S(x) = P(X \geq x)$ and $F(x) = P(X \leq x)$ are the survival and cumulative distribution functions of a non-negative discrete random variable X, respectively. The odds ratio of X, in the discrete setup, is defined as

$$\phi(x) = \frac{F(x)}{S(x+1)}, \quad x \in \mathbb{N}_0 = \{0, 1, 2 \ldots\}, \tag{1}$$

see Unnikrishnan Nair and Sankaran (2015).

In this paper, a new discrete distribution is introduced based on the odds function of a geometric distribution. This distribution has some interesting properties and can be adjusted to suit most discrete data sets. The survival and cumulative distribution functions have analytical forms and the probability mass function can be unimodal or bimodal. The rest of the paper is organized as follows. Section 2 introduces the geometric odds ratio distribution, for the first time. Some important features and properties of the new discrete distribution such as the cumulative distribution and survival functions are studied. In addition, the mean and variance of the distribution will be obtained. The behavior of the hazard rate function is illustrated. We will see that the new distribution exhibits bathtub, and monotonically increasing and decreasing hazard rates. Rényi and Shannon entropies are obtained and the infinite divisibility of the distribution in question is discussed. The distributions of order statistics will be attained and some stochastic orders are also discussed. The estimation process of the parameters is provided in Section 3. Additionally, the stress-strength parameter is illustrated and the new model with a real data set is also examined in this section. Some concluding remarks are given in Section 4.

2 The geometric odds ratio distribution

2.1 Definition and interpretations

Let X denote the number of failures before the first success in a Bernoulli experiment whose probability of success is p. In this case, the survival function (SF) of X is

$$S(x) = P(X \geq x) = q^x, \quad x \in \mathbb{N}_0, \tag{2}$$

which is known as the geometric SF, and $q = 1 - p$. The cumulative distribution function (CDF) of the above geometric random variable $X(\sim \mathrm{G}(p))$, is

$$F(x) = P(X \leq x) = 1 - S(x+1) = 1 - q^{x+1}, \quad x \in \mathbb{N}_0. \tag{3}$$

Based on (1), the odds ratio of the $\mathrm{G}(p)$ distribution is given by

$$\phi(x) = \frac{1 - q^{x+1}}{q^{x+1}}, \quad x \in \mathbb{N}_0. \tag{4}$$

Definition 2.1. *A discrete random variable X is said to be geometric odds ratio (GOR) distributed, if its CDF has the following form*

$$F(x; \theta, q, \alpha) = 1 - \theta^{\left(\frac{1-q^{x+1}}{q^{x+1}}\right)^{\alpha}}, \quad x \in \mathbb{N}_0, \tag{5}$$

where $0 < \theta < 1$, $0 < q < 1$ and $\alpha > 0$ are the model parameters. A GOR distribution with parameters θ, q and α will be denoted by $GOR(\theta, q, \alpha)$ in the sequel.

It is easy to investigate that F in (5) is a bona fide CDF whose corresponding SF is

$$S(x; \theta, q, \alpha) = \theta^{\left(\frac{1-q^{x}}{q^{x}}\right)^{\alpha}}, \quad x \in \mathbb{N}_0. \tag{6}$$

In addition, the probability mass function (PMF) of a $\mathrm{GOR}(\theta, q, \alpha)$ distribution, for $x \in \mathbb{N}_0$, is

$$f(x; \theta, q, \alpha) = p_x = P(X = x) = S(x) - S(x+1) = \theta^{\left(\frac{1-q^{x}}{q^{x}}\right)^{\alpha}} - \theta^{\left(\frac{1-q^{x+1}}{q^{x+1}}\right)^{\alpha}}. \tag{7}$$

A GOR distribution, depending on its parameters values, can have a unimodal or a bimodal PMF. Figure 1, illustrates the PMF plots of GOR distributions for some possible values of the parameters. Figure 1 shows that the GOR distribution can be unimodal, right-skewed, left-skewed, almost symmetric and even bimodal. So, the parameters of the GOR distribution can be adjusted to suit most discrete data sets.

The r-th moment of the $\mathrm{GOR}(\theta, q, \alpha)$ distribution is given by

$$E(X^r) = \sum_{x=1}^{\infty} \{x^r - (x-1)^r\} P(X \geq x) = \sum_{x=1}^{\infty} \{x^r - (x-1)^r\} \theta^{\left(\frac{1-q^{x}}{q^{x}}\right)^{\alpha}}. \tag{8}$$

Specially, the first and second moments are

$$E(X) = \sum_{x=1}^{\infty} \theta^{((1-q^x)/q^x)^{\alpha}} \tag{9}$$

$$E(X^2) = \sum_{x=1}^{\infty} (2x-1) \theta^{((1-q^x)/q^x)^{\alpha}}, \tag{10}$$

For different values of the parameters, the mean and variance of the GOR distribution have been calculated in Table 1.

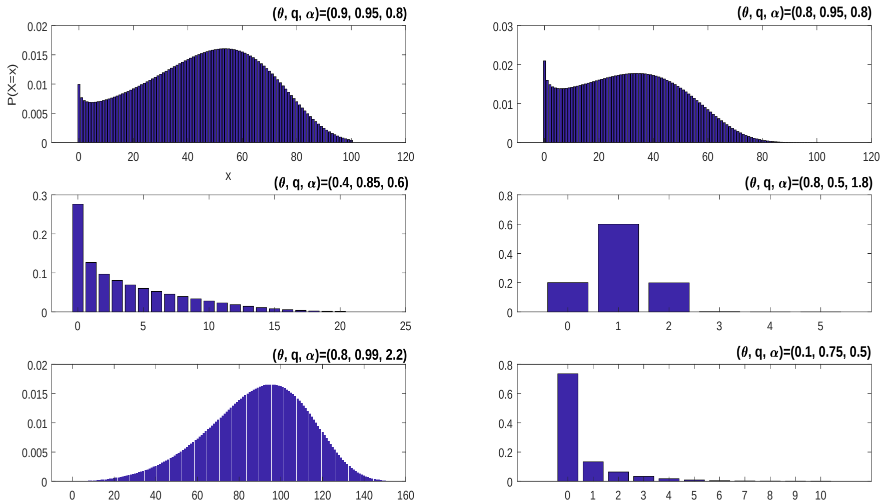

Figure 1: PMF plots of a GOR(θ, q, α) distribution for some parameters values.

As we see, the mean increases when θ increases, or q increases. In addition, the mean decreases when α increases. Because

$$
\begin{aligned}
\frac{\partial}{\partial \alpha} E(X) &= \sum_{x=1}^{\infty} \theta^{((1-q^x)/q^x)^\alpha} \log \left(\theta^{(1-q^x)/q^x}\right) < 0, \\
\frac{\partial}{\partial \theta} E(X) &= \sum_{x=1}^{\infty} \left(\frac{1-q^x}{q^x}\right)^\alpha \theta^{((1-q^x)/q^x)^\alpha - 1} > 0 \\
\frac{\partial}{\partial q} E(X) &= \sum_{x=1}^{\infty} -\alpha \log(\theta) \theta^{((1-q^x)/q^x)^\alpha} \left(\frac{1-q^x}{q^x}\right)^{\alpha-1} \\
&\quad \times \frac{xq^{2(x-1)} + (x-1)q^{x-2}(1-q^x)}{q^{2(x-1)}} > 0.
\end{aligned}
$$

The variance has the same manner, too. In addition, the variance can be larger, equal, or greater than the mean. Therefore, the parameters of a GOR distribution can be adjusted to suit over- and under-dispersed data sets. In addition, like the Poisson distribution, it may have equal mean and variance.

The γ-th percentile point of a GOR(θ, q, α) distribution is also given by $\xi_\gamma = -\log\left(q\left\{1 + (\log(1-\gamma)/\log\theta)^{1/\alpha}\right\}\right)/\log q$.

Now, we show that the GOR distributions are closed under minimum. That is, the minimum of a random sample of a GOR distribution, is itself a GOR variable. More precisely, we have the following result.

Theorem 2.2. *If $X_1, \ldots, X_n$ is a random sample from a GOR(θ, q, α) distribution, then $\min\{X_1, \ldots, X_n\}$ follows a GOR(θ^n, q, α) distribution.*

Table 1: Mean (Variance) of the GOR(θ, q, α) distributions.

$\theta = 0.25$				
α/q	0.10	0.25	0.50	0.75
0.05	0.9109 (4.8134)	1.5952 (14.0195)	3.3209 (58.4249)	8.1961 (347.3344)
0.2	0.1515 (0.2074)	0.3316 (0.7022)	0.8103 (3.2834)	2.2045 (20.8394)
0.5	0.0156 (0.0154)	0.0953 (0.0956)	0.3713 (0.5485)	1.2635 (3.7993)
0.7	0.0016 (0.0016)	0.0503 (0.0480)	0.3048 (0.3308)	1.1777 (2.3037)
$\theta = 0.50$				
α/q	0.10	0.25	0.50	0.75
0.05	3.0445 (18.6848)	5.2217 (53.1816)	10.6990 (217.7728)	26.1488 (1281.4)
0.2	0.5940 (0.9301)	1.1507 (2.8690)	2.5740 (12.4456)	6.6244 (75.8395)
0.5	0.1260 (0.1122)	0.3734 (0.3869)	1.0546 (1.8076)	3.0549 (11.3436)
0.7	0.0397 (0.0381)	0.2340 (0.1991)	0.8013 (0.9378)	2.4949 (5.8340)
$\theta = 0.75$				
α/q	0.10	0.25	0.50	0.75
0.05	7.7665 (47.7174)	13.1476 (133.6933)	26.6736 (541.0240)	64.8076 (3161.8)
0.2	1.6746 (2.6564)	3.0293 (7.7121)	6.4477 (32.0562)	16.1055 (190.0473)
0.5	0.4791 (0.3643)	1.0479 (1.0754)	2.5072 (4.5422)	6.6601 (27.1394)
0.7	0.2628 (0.1953)	0.6903 (0.5299)	1.8071 (2.1890)	4.9083 (12.9507)

Proof. The proof is straightforward and the details are avoided. □

The hazard rate function of the GOR(θ, q, α) distribution, for $x \in \mathbb{N}_0$, is given by

$$h(x) = \frac{f(x;\theta,q,\alpha)}{S(x;\theta,q,\alpha)} = \frac{\theta^{((1-q^x)/q^x)^\alpha} - \theta^{\left((1-q^{x+1})/q^{x+1}\right)^\alpha}}{\theta^{((1-q^x)/q^x)^\alpha}}. \tag{11}$$

Figure 2 illustrates the behaviour of the hazard rate function of the GOR distribution. This figure indicates that the hazard rate function of the GOR distribution is monotonically increasing and decreasing, and bathtub-shaped. Hence, the GOR distributions can analyze more failure rate data with respect to the geometric and DGE (or GG) distributions whose hazard rate functions are constant and monotone, respectively.

2.2 Rényi and Shannon entropies

The entropy of a random variable X is a measure of uncertainty variation. The Rényi and Shannon entropies are important in Statistics, Reliability and Quantum information theory. The Rényi and Shannon entropies of a discrete random variable X, whose PMF is $P(X = x)$, in general, are given by

$$R_\eta = \frac{1}{1-\eta} \log \sum_{x=0}^{\infty} \{P(X=x)\}^\eta, \quad \eta \neq 1$$

$$S = -\sum_{x=0}^{\infty} P(X=x) \log P(X=x),$$

For the GOR distribution, using the binomial expansion, we find that

$$\{P(X=x)\}^\eta = \left\{\theta^{((1-q^x)/q^x)^\alpha} - \theta^{\left((1-q^{x+1})/q^{x+1}\right)^\alpha}\right\}^\eta$$

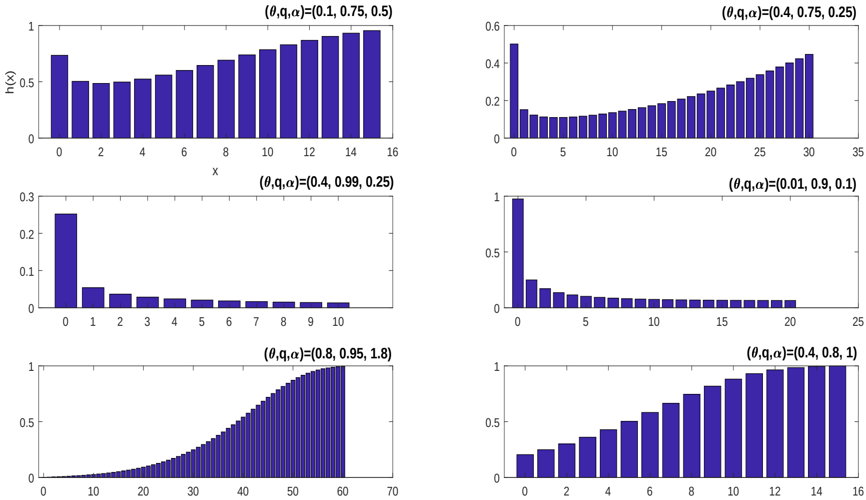

Figure 2: Hazard rate function plots of a GOR(θ, q, α) distribution for some parameters values.

$$= \sum_{j=0}^{\infty}(-1)^j \binom{\eta}{j} \theta^{j\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}} \theta^{(\eta-j)((1-q^x)/q^x)^{\alpha}}. \tag{12}$$

It must be mentioned that for an integer value of η, $\sum_{j=0}^{\infty}$ should be replaced by $\sum_{j=0}^{\eta}$ in (12). Therefore, the Rényi entropy of a GOR(θ, q, α) distribution is rewritten as

$$R_\eta = \frac{1}{1-\eta} \log \sum_{x=0}^{\infty}\sum_{j=0}^{\infty}(-1)^j \binom{\eta}{j} \theta^{j\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}+(\eta-j)((1-q^x)/q^x)^{\alpha}}. \tag{13}$$

For an integer η, the interior summation stops at η in the above relation.

The Shannon entropy of the GOR distribution is also given by

$$S = -\sum_{x=0}^{\infty}\left(\theta^{((1-q^x)/q^x)^{\alpha}} - \theta^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}}\right)\log\left(\theta^{((1-q^x)/q^x)^{\alpha}} - \theta^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}}\right). \tag{14}$$

The Rényi and Shannon entropies of the GOR distribution have been calculated in Table 2 for some values of the parameters. According to the results of Table 2, we see that when $\eta \to 1$, the values of Rényi entropy are approximately equivalent to that of Shannon entropy. Therefore, the known relation between these two entropies, i.e. $S = \lim_{\eta\to 1} R_\eta$, is investigated in a GOR distribution, too.

2.3 Stochastic orders

Numerous stochastic orders between random variables X and Y have been introduced in the literature. For details, see, e.g., Shaked and Shanthikumar (2007). We first need

Table 2: Rényi and Shannon entropies of the GOR distribution.

η	(θ, q, α)	Rényi	Shannon
0.50	(0.9, 0.95, 0.8)	4.5564	4.4873
0.995	(0.8, 0.95, 0.7)	4.3944	4.3939
1.005	(0.75, 0.95, 0.7)	4.2847	4.2853
2.2	(0.1, 0.95, 0.8)	2.3495	2.7253
0.50	(0.8, 0.75, 0.2)	3.9312	3.7225
0.995	(0.85, 0.85, 0.1)	5.0025	4.9989
1.005	(0.85, 0.99, 0.5)	6.4387	6.4394
2.2	(0.4, 0.65, 0.005)	0.9398	2.7398
0.50	(0.95, 0.85, 0.1)	5.5987	5.4663
0.995	(0.99, 0.8, 0.1)	5.3368	5.3356
1.005	(0.85, 0.95, 0.1)	6.0530	6.0584
2.2	(0.9, 0.9, 0.005)	4.2405	0.5131

to review some notions of stochastic orders, in the discrete setup, which are relevant in the context of the present paper.

Simple stochastic order: X is said to be stochastically smaller than Y (written as $X \leq_{st} Y$) if for all integer values of x, $S_X(x+1) \leq S_Y(x+1)$, where S_X and S_Y are the SFs of X and Y, respectively.

Hazard rate order: X is smaller than Y in the hazard rate order (written as $X \leq_{hr} Y$), if $h_X(x) \geq h_Y(x)$. h_X and h_Y are the hazard rate functions of X and Y, respectively.

Odds ratio order: The discrete random variable X is smaller than Y in the odds ratio order, denoted by $X \leq_{odd} Y$, if and only if for all integer values of x, $\phi_X(x) \leq \phi_Y(x)$, where ϕ is given by (1). In addition, $X \leq_{st} Y \Leftrightarrow X \leq_{odd} Y$; see, Unnikrishnan Nair and Sankaran (2015).

Here, some new findings regards to the stochastic orders of the GOR distributions are proved which are useful for comparing them.

Theorem 2.3. *Let $X \sim GOR(\theta_1, q, \alpha)$ and $Y \sim GOR(\theta_2, q, \alpha)$. If $\theta_1 \leq \theta_2$, then $X \leq_{hr} Y$.*

Proof. If $\theta_1 \leq \theta_2$, it is obvious that

$$\theta_1^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha} - ((1-q^x)/q^x)^{\alpha}} \leq \theta_2^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha} - ((1-q^x)/q^x)^{\alpha}},$$

or equivalently,

$$1 - \theta_1^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha} - ((1-q^x)/q^x)^{\alpha}} \geq 1 - \theta_2^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha} - ((1-q^x)/q^x)^{\alpha}},$$

which means that $h_X(x) \geq h_Y(x)$. □

Theorem 2.4. *Let $X \sim GOR(\theta, q_1, \alpha)$ and $Y \sim GOR(\theta, q_2, \alpha)$. If $q_1 \leq q_2$, then $X \geq_{st} Y$.*

Proof. If $q_1 \leq q_2$, it is easy to show that $\theta^{\left((1-q_1^{x+1})/q_1^{x+1}\right)^{\alpha}} \geq \theta^{\left((1-q_2^{x+1})/q_2^{x+1}\right)^{\alpha}}$, or equivalently, $S_X(x+1) \geq S_Y(x+1)$. □

Remark 2.5. *Under the conditions of Theorem 3, we see that* $X \geq_{odd} Y$.

Now, we want to state that the CDF of a GOR distribution, given by (3), can be expanded to a general family of discrete odds ratio distributions. More precisely, for non-negative integer values of x, suppose that $G(x;\boldsymbol{\xi})$ is an arbitrary discrete CDF and $S(x;\boldsymbol{\xi})$ is its corresponding SF. Then,

$$F(x;\boldsymbol{\xi},\theta,\alpha) = 1 - \theta^{\left(\frac{G(x;\boldsymbol{\xi})}{S(x+1;\boldsymbol{\xi})}\right)^{\alpha}}, \quad x \in \mathbb{N}_0, 0 < \theta < 1, \alpha > 0, \tag{15}$$

defines a general family of discrete odds ratio (DOR) distributions, in which contains the GOR distribution as a special case. Let us consider the notation $\mathrm{DOR}(\theta, G, \alpha)$ to represent the CDF (15), which is introduced for the first time here.

Theorem 2.6. *Let* $X \sim DOR(\theta, G_1, \alpha)$ *and* $Y \sim DOR(\theta, G_2, \alpha)$*. If* $G_1 \leq_{st} G_2$*, then* $F_X \geq_{st} F_Y$*, where* F_X *and* F_Y *are the CDFs of* X *and* Y*, respectively.*

Proof. $G_1 \leq_{st} G_2$ means that $G_1(x) \geq G_2(x)$ and $S_1(x+1) \leq S_2(x+1)$, where S_i, $i = 1, 2$, corresponds to G_i. Hence, we see that $\frac{G_1(x)}{S_1(x+1)} \geq \frac{G_2(x)}{S_2(x+1)}$. Therefore, we conclude that $\theta^{\left(\frac{G_1(x)}{S_1(x+1)}\right)^{\alpha}} \geq \theta^{\left(\frac{G_2(x)}{S_2(x+1)}\right)^{\alpha}}$. The above relation, yields that $F_X(x) \leq F_Y(x)$. $\square$

2.4 Order statistics

Order statistics are among the most fundamental tools in Non-parametric statistics and Inference. They usually enter the problems of estimation and hypothesis testing. Here, we want to establish some general relations regarding the GOR distributions. More precisely, let $F_{i:n}(x;\theta,q,\alpha)$ and $f_{i:n}(x;\theta,q,\alpha)$ denote the CDF and PMF of the i-th order statistic of a random sample of size n from the $\mathrm{GOR}(\theta, q, \alpha)$ distribution.

In general, the CDF of the i-th order statistic, is

$$\begin{aligned} F_{i:n}(x;\boldsymbol{\xi}) &= \sum_{k=i}^{n} \binom{n}{k} \{F(x;\boldsymbol{\xi})\}^k \{1 - F(x;\boldsymbol{\xi})\}^{n-k} \\ &= \sum_{k=i}^{n} \binom{n}{k} \{F(x;\boldsymbol{\xi})\}^k \{S(x+1;\boldsymbol{\xi})\}^{n-k}, \end{aligned} \tag{16}$$

where $\boldsymbol{\xi}$ is the parameters vector of F. If F corresponds to the GOR distribution, the CDF of the i-th order statistic is rewritten as

$$\begin{aligned} F_{i:n}(x;\theta,q,\alpha) &= \sum_{k=i}^{n} \binom{n}{k} \{F(x;\theta,q,\alpha)\}^k \{S(x+1;\theta,q,\alpha)\}^{n-k} \\ &= \sum_{k=i}^{n} \binom{n}{k} \theta^{(n-k)\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}} \left\{1 - \theta^{\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}}\right\}^k. \end{aligned} \tag{17}$$

By using the binomial expansion, $F_{i:n}$ can be written as

$$F_{i:n}(x;\theta,q,\alpha) = \sum_{k=i}^{n} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \binom{n}{k} \theta^{(n-k+j)\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}}. \tag{18}$$

The PMF of the i-th order statistic for non-negative integer values of x, $f_{i:n}(x) = F_{i:n}(x) - F_{i:n}(x-1)$, is then given by

$$\begin{aligned} f_{i:n}(x;\theta,q,\alpha) &= \sum_{k=i}^{n}\sum_{j=0}^{k}(-1)^{j}\binom{k}{j}\binom{n}{k}\Big\{\theta^{(n-k+j)\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}} \\ &\quad -\theta^{(n-k+j)((1-q^{x})/q^{x})^{\alpha}}\Big\} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{k}(-1)^{j+1}\binom{k}{j}\binom{n}{k}\Big\{\theta^{(n-k+j)((1-q^{x})/q^{x})^{\alpha}} \\ &\quad -\theta^{(n-k+j)\left((1-q^{x+1})/q^{x+1}\right)^{\alpha}}\Big\} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{k}(-1)^{j+1}\binom{k}{j}\binom{n}{k}f_{GOR}(x;\theta^{n-k+j},q,\alpha). \end{aligned} \tag{19}$$

Therefore, the PMFs of different order statistics can be written as linear combinations of the GOR PMFs. This is a useful advantage. Some characteristics of the order statistics' PMFs, can be obtained from those of GOR PMFs, immediately. For example, different moments of the order statistics, which is widely used in L-moments, can be obtained by means of the GOR moments. For instance, the mean of the i-the order statistic is obtained as

$$\mu_{i:n} = \sum_{x=1}^{\infty}\sum_{k=i}^{n}\sum_{j=0}^{k}(-1)^{j+1}\binom{k}{j}\binom{n}{k}\theta^{(n-k+j)((1-q^{x})/q^{x})^{\alpha}}. \tag{20}$$

2.5 Infinite divisibility

Here we make the following note in regards to the famous structural property of infinite divisibility of the GOR distribution. Infinite divisibility has a close relation to the central limit theorem (CLT) and waiting time distributions. Hence, it is an important question in modelling to know whether a given distribution is infinitely divisible or not. First note that in a GOR(θ, q, p) distribution,

$$\begin{aligned} p_0 &= P(X=0) = 1-\theta^{((1-q)/q)^{\alpha}} > 0, \\ p_1 &= P(X=1) = \theta^{((1-q)/q)^{\alpha}} - \theta^{((1-q^2)/q^2)^{\alpha}} > 0. \end{aligned}$$

In addition,

$$\frac{p_{x+1}}{p_x} = \frac{\theta^{((1-q^{x+1})/q^{x+1})^{\alpha}} - \theta^{((1-q^{x+2})/q^{x+2})^{\alpha}}}{\theta^{((1-q^{x})/q^{x})^{\alpha}} - \theta^{((1-q^{x+1})/q^{x+1})^{\alpha}}}.$$

In Table 3, p_{x+1}/p_x has been calculated for $x = 0, 1, 2, 3, 4$ and some different values of the parameters. In general, increasing trends are not seen in the sequences. So, it seems that the GOR distributions are not infinitely divisible in general. Remember that according to Warde and Katti (1971), a PMF is infinitely divisible if p_{x+1}/p_x forms a monotone increasing sequence, for all $x \in \mathbb{N}_0$.

Since the classes of self-decomposable and stable distributions, in their discrete concepts, are subclasses of infinitely divisible distributions, one can conclude that the GOR distributions can be neither self-decomposable nor stable in general.

Table 3: The behaviour of p_{x+1}/p_x.

(θ, q, α)	p_1/p_0	p_2/p_1	p_3/p_2	p_4/p_3	p_5/p_4
(0.85, 0.99, 0.5)	0.4130	0.7685	0.8453	0.8839	0.9074
(0.1, 0.3, 0.6)	0.0221	0.0045	2.6×10^{-5}	4.9×10^{-10}	8.2×10^{-20}
(0.6, 0.3, 0.2)	0.2241	0.9137	0.9510	0.9106	0.8409

2.6 Mixing process

Sometimes it is important to consider that one or more parameters of a distribution vary according to the certain given probability distribution, called the mixing distribution. For instance, such situations occur in problems associated with accident proneness and entomological field data.

Here, it is supposed that θ is itself a continuous random variable specified by the generalized beta (GB) distribution, introduced by McDonald (1984), as

$$\pi(\theta) = \frac{\zeta \theta^{a\zeta-1}(1-\theta^\zeta)^{b-1}}{B(a,b)}, \quad 0 < \theta < 1,$$

where $a > 0$, $b > 0$, $\zeta > 0$ and $B(a, b)$ is the known beta function.

Theorem 2.7. *Let $X \sim GOR(\theta, q, \alpha)$ and θ follows the GB distribution with $\zeta = j + 1; j = 0, 1, \ldots$. The marginal distribution of X is given by*

$$m(x) = \frac{B(\frac{k_x}{\zeta} + a, b) - B(\frac{k_{x+1}}{\zeta} + a, b)}{B(a,b)}, \quad x \in \mathbb{N}_0,$$

where $k_x = ((1 - q^x)/q^x)^\alpha$.

Proof. The proof is straightforward and the details are avoided. □

3 Statistical inference

3.1 Maximum likelihood estimation

Let $X \sim$ GOR(θ, q, α). In addition, let us consider $\mathbf{\Omega} = (\theta, q, \alpha)^T$. The likelihood function of a single observation x is given by

$$L(\mathbf{\Omega}) = \theta^{\left(\frac{1-q^x}{q^x}\right)^\alpha} - \theta^{\left(\frac{1-q^{x+1}}{q^{x+1}}\right)^\alpha}. \tag{21}$$

The derivatives of the likelihood function with respect to the parameters involved are given by

$$\begin{aligned}
\frac{\partial L}{\partial \theta} &= \theta^{((1-q^x)/q^x)^\alpha - 1}((1-q^x)/q^x)^\alpha - \theta^{((1-q^{x+1})/q^{x+1})^\alpha - 1}((1-q^{x+1})/q^{x+1})^\alpha, \\
\frac{\partial L}{\partial q} &= -\alpha \log(\theta)\theta^{((1-q^x)/q^x)^\alpha}((1-q^x)/q^x)^{\alpha-1} x\{q^{-1} + q^{-(x+1)}(1-q^x)\} \\
&\quad -\alpha \log(\theta)\theta^{((1-q^{x+1})/q^{x+1})^\alpha}((1-q^{x+1})/q^{x+1})^{\alpha-1}(x+1)
\end{aligned}$$

$$\times\{q^{-1}+q^{-(x+2)}(1-q^{x+1})\}$$

$$\frac{\partial L}{\partial \alpha} = \log\left(\theta^{((1-q^{x})/q^{x})}\right)\theta^{((1-q^{x})/q^{x})^{\alpha}} - \log\left(\theta^{((1-q^{x+1})/q^{x+1})}\right)\theta^{((1-q^{x+1})/q^{x+1})^{\alpha}}.$$

Now, let $x_1, x_2, \ldots, x_n$ be observations of a random sample drawn from a GOR(θ, q, α) distribution. In this case, the total likelihood function is

$$L_n(\boldsymbol{\Omega}) = \prod_{k=1}^{n} L_{[k]}(\boldsymbol{\Omega}),$$

where $L_{[k]}(\boldsymbol{\Omega})$; $k = 1, 2, \ldots, n$, is given by Eq. (21). The maximum likelihood estimate (MLE) of $\boldsymbol{\Omega}$, say $\hat{\boldsymbol{\Omega}}$, is obtained by solving the nonlinear equation $\boldsymbol{M}_n = (\partial L_n/\partial\theta, \partial L_n/\partial q, \partial L_n/\partial\alpha)^T = \mathbf{0}$. It is obvious that a numerical method must be used in order to solve the above equation.

The Fisher information matrix is also given by

$$\boldsymbol{I}(\boldsymbol{\Omega}) = [I_{\omega_i,\omega_j}]_{3\times 3}; \quad i,j = 1,2,3,$$

whose components can be calculated, numerically, by the relation $I_{\omega_i,\omega_j} = E(-\frac{\partial^2 L(\boldsymbol{\Omega})}{\partial\omega_i\partial\omega_j})$, $i,j = 1,2,3$. The total Fisher information matrix is given by $\boldsymbol{I}_n(\boldsymbol{\Omega}) = n\boldsymbol{I}(\boldsymbol{\Omega})$ which can be approximated by $\boldsymbol{I}_n(\hat{\boldsymbol{\Omega}}) \approx [-\frac{\partial^2 L_n(\boldsymbol{\Omega})}{\partial\omega_i\partial\omega_j}\mid_{\boldsymbol{\Omega}=\hat{\boldsymbol{\Omega}}}]_{3\times 3}$, $i,j = 1,2,3$.

Under some regularity conditions given, e.g., in Ferguson (1996), $\hat{\boldsymbol{\Omega}}$ has an asymptotic normal distribution as $N_3(\boldsymbol{\Omega}, \boldsymbol{I}_n(\hat{\boldsymbol{\Omega}})^{-1})$. Asymptotic normal distributions are usually used for constructing approximate confidence intervals, confidence regions, and testing hypotheses of the parameters.

3.2 Stress-strength parameter

The stress-strength parameter $R = P(X \geq Y)$ is a measure of component reliability and its estimation problem when X and Y are independent and follow a specified common distribution has been widely discussed in the literature. Suppose that the random variable X is the strength of a component which is subjected to a random stress Y. Estimation of R when X and Y are independent and identically distributed following a well-known distribution has been considered in the literature. A relatively small amount of work is devoted to discrete or categorical data. Hence, the estimation of R in regards to a GOR distribution is now considered.

The stress-strength parameter, in the discrete setup, is defined as

$$R = P(X \geq Y) = \sum_{x=0}^{\infty} f_X(x)F_Y(x),$$

where f_X and F_Y denote the PMF and CDF of the independent discrete random variables X and Y, respectively.

Now, let $X \sim$ GOR$(\boldsymbol{\Omega_1})$ and $Y \sim$ GOR$(\boldsymbol{\Omega_2})$, where $\boldsymbol{\Omega_1} = (\theta, q_1, \alpha_1)^T$ and $\boldsymbol{\Omega_2} = (\theta, q_2, \alpha_2)^T$. In this case, we find that

$$R = 1 - \sum_{x=0}^{\infty} \theta^{((1-q_1^{x})/q_1^{x})^{\alpha_1} + ((1-q_2^{x+1})/q_2^{x+1})^{\alpha_2}} + \sum_{x=0}^{\infty} \theta^{((1-q_1^{x+1})/q_1^{x+1})^{\alpha_1} + ((1-q_2^{x+1})/q_2^{x+1})^{\alpha_2}}. \tag{22}$$

Assume that $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ are independent observations from $X \sim \text{GOR}(\boldsymbol{\Omega_1})$ and $Y \sim \text{GOR}(\boldsymbol{\Omega_2})$, respectively. The total likelihood function is $L_R(\boldsymbol{\Omega^*}) = L_n(\boldsymbol{\Omega_1})L_m(\boldsymbol{\Omega_2})$, where $\boldsymbol{\Omega^*} = (\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$. The score vector is given by

$$M_R(\boldsymbol{\Omega^*}) = (\partial L_R/\partial\theta, \partial L_R/\partial q_1, \partial L_R/\partial\alpha_1, \partial L_R/\partial q_2, \partial L_R/\partial\alpha_2),$$

and the MLE of $\boldsymbol{\Omega^*}$, say $\hat{\boldsymbol{\Omega}}^*$, may be attained from that of nonlinear equation $M_R(\boldsymbol{\Omega^*}) = \mathbf{0}$. Hence, by substituting the MLEs in (22), the stress-strength parameter R will be estimated.

3.3 Data analysis

Here, the GOR distribution is examined for a real data set, given by Consul and Jain (1973). These data consider the results of ten shots fired from a rifle at each of 100 targets. Gómez-Déniz (2010) used these data in order to study its generalization of the geometric distribution.

Now, we want to compare the capacity of the GOR distribution with some of its rival models. The discrete Weibull (DW) distribution of Nakagawa and Osaki (1975), the two-parameter DGE distribution of Nekoukhou et al. (2013) which is a generalization of the geometric distribution, the exponentiated discrete Weibull (EDW) distribution studied by Nekoukhou and Bidram (2015), and also the generalized geometric (GG) distribution of Gómez-Déniz (2010) are the rival models. These distributions will be briefly introduced in the Appendix.

The parameters of the GOR distribution have been estimated by the maximum likelihood method when the Newton-Raphson procedure converges in MATLAB. Comparing the GOR distribution with its rival models is performed by using the Akaike information criterion (AIC) and kolmogorov-Smirnov (K-S) test statistic. Table 4 indicates the fitting computations.

Table 4: Summary of computations.

Models	MLEs	AIC	K-S statistic
GOR	$(\hat{\theta}, \hat{q}, \hat{\alpha}) = (0.46, 0.89, 1.47)$	48.2928	0.1088
EDW	$(\hat{\alpha}, \hat{\gamma}, \hat{p}) = (3.72, 1.05, 0.99)$	54.1788	0.2814
DGE	$(\hat{a}, \hat{p}) = (18.51, 0.53)$	53.9190	0.2424
DW	$(\hat{\alpha}, \hat{p}) = (3.80, 0.99)$	51.9968	0.2819
GG	$(\hat{\alpha}, \hat{\theta}) = (394.75, 0.33)$	52.3626	0.3004

According to the AICs and the values of K-S test statistics in Table 4, one can conclude that the GOR distribution gives a satisfactory fit to these data.

One can construct approximate confidence intervals for the parameters of the GOR distribution. Such confidence intervals are attained by means of asymptotic covariance matrix of the MLEs of the GOR parameters when the Newton-Raphson procedure converges. For instance, 95% asymptotic confidence intervals for the GOR parameters are calculated as $\theta \in (0.46 \mp 0.176)$, $q \in (0.89 \mp 0.078)$ and $\alpha \in (1.47 \mp 0.255)$.

4 Conclusions

A new discrete distribution, called the geometric odds ratio (GOR) distribution, motivated by the odds ratio of the geometric distribution and the fact that it provides greater flexibility in order to analyze various discrete data is introduced. Moreover, the GOR distribution is a special member of a general class of discrete odds ratio (DOR) distributions, which was introduced for the first time in the present paper. The GOR distribution is appropriate for modeling both over- and under-dispersed data and can be a unimodal or a bimodal discrete distribution. Moreover, the hazard rate function of a GOR distribution can be increasing, decreasing and bathtub-shaped. That is, the GOR distributions can be used as improved models for analyzing failure data in discrete case.

Appendix

Here the rival models, indicated in Table 4, are briefly introduced.
1) The exponentiated discrete Weibull (EDW) distribution of Nekoukhou and Bidram (2015), for $y \in \mathbb{N}_0$, has the following PMF

$$\begin{aligned} f(y; p, \alpha, \gamma) &= \{1 - p^{(y+1)^\alpha}\}^\gamma - \{1 - p^{y^\alpha}\}^\gamma \\ &= \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\gamma}{j} \{p^{jy^\alpha} - p^{j(y+1)^\alpha}\}, \end{aligned}$$

where $0 < p < 1$, $\alpha > 0$, $\gamma > 0$ and $\binom{\gamma}{j} = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1-j)j!}$. For integer $\gamma > 0$, the above sum stops at γ.
2) Discrete Weibull distribution of Nakagawa and Osaki (1975), with PMF

$$f(y; p, \alpha) = (1 - p^{(y+1)^\alpha}) - (1 - p^{y^\alpha}),$$

is a special case of the EDW distribution, when $\gamma = 1$.
3) Discrete generalized exponential distribution, $DGE(\gamma, p)$, of Nekoukhou et al. (2013) with PMF

$$f(y; p, \gamma) = \{1 - p^{(y+1)}\}^\gamma - \{1 - p^y\}^\gamma,$$

can be considered as another special case of the EDW distribution, by choosing $\alpha = 1$.
4) A generalization of the geometric distribution has been introduced by Gómez-Déniz (2010). The generalized geometric (GG) distribution of Gómez-Déniz (2010), for $y \in \mathbb{N}_0$, has the following PMF

$$f(y; \alpha, \theta) = \frac{\alpha\theta^y(1-\theta)}{\{1 - (1-\alpha)\theta^{y+1}\}\{1 - (1-\alpha)\theta^y\}}$$

in which $\alpha > 0$ and $0 < \theta < 1$ are the model parameters.

Acknowledgement

The authors would like to thank anonymous referees and also the Editor-in-Chief.

References

Consul, P.C. and Jain, G.C. (1973). A generalization of the Poisson distribution. *Technometrics*, **15(4)**, 791-799.

Ferguson, T.S. (1996). *A Course in Large Sample Theory.* Chapman and Hall, London.

Gómez-Déniz E. (2010). Another generalization of the geometric distribution. *Test*, **19**, 399-415.

Gupta, R.C. (2011). Bivariate odds ratio and association measures. *Statistical Papers*, **52**, 125-138.

Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, **41(2)**, 173-188.

McDonald J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica,* 52, 647-664.

Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24(5), 300-301.

Nekoukhou, V., Alamatsaz, M.H. and Bidram, H. (2013). Discrete generalized exponential distribution of a second type. *Statistics*, **47**(4), 876-887.

Nekoukhou, V. and Bidram, H. (2015). The exponentiated discrete Weibull distribution. *SORT*, **39**(1), 127-146.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders.* vol 3–44. Springer, New York.

Sankaran, P.G. and Jayakumar, K. (2008). On proportional odds models. *Statistical Papers*, **49**, 779-789.

Unnikrishnan Nair N. and Sankaran, P.G. (2015). Odds function and odds rate for discrete lfetime distributions. *Communications in Statistics—Theory and Methods*, **44**, 4185-4202.

Warde, W.D., and Katti. S.K. (1971). Infinite divisibility of discrete distributions. *The Annals of Mathematical Statistics*, **42** (3), 1088-1090.