

Size Biased X-Geometric with Statistical Properties and Applications

Brijesh P. Singh¹, Tanya Singh² and Utpal Dhar Das³

Department of Statistics, Institute of Science
Banaras Hindu University, Varanasi-221005

¹ Email: brijesh@bhu.ac.in

² Email (Corresponding Author): tanya6702@gmail.com

³ Email: utpal.statmath@gmail.com

ABSTRACT

In this study, a weighted distribution named Size Biased X-Geometric Distribution (SBXGD) is introduced and explored. Various statistical properties along with Lorenz curve, stochastic ordering and maximum likelihood estimation are discussed. Also bias and mean square error for the parameter is estimated through simulation. Some real data sets are used to examine the suitability of model.

Keywords: Size biased X-Geometric distribution; Lorenz curve; stochastic ordering; maximum likelihood estimation.

2000 Mathematics Subject Classification: 49K05, 49K15, 49S05.

1 Introduction

(Fisher, 1934) was the first to suggest a size biased distribution to reflect ascertainment bias, and weighted distributions were formalized in a unifying theory (Rao, 1965). When observations from a sample are recorded with unequal probability, such as in probability proportional to size (PPS) designs, this distribution emerges naturally (Gove, 2003). They also appear in a variety of biomedical applications, including the study of family history and disease, survival and intermediate events, the latency period of AIDS in patients who receive blood transfusions, and cell labeling with radioactive tracers (Zelen, 1974; Zelen and Feinleib, 1969). When the researchers record the observations according to certain stochastic model, the recorded observations will not have the original distribution unless each observation is given an equal chance of being recorded (Ahmad et al. 2020). In this scenario, standard distributions fail to deliver adequate results, necessitating the use of weighted versions of the distributions. If we take X as a random variable which follows a distribution $P(x; \theta)$, with unknown parameter θ , the corresponding weighted distribution is of the form.

$$P^\delta(x; \theta) = \frac{\delta(x)P(x; \theta)}{E[\delta(x)]}$$

www.ceser.in/ceserp
www.ceserp.com/cp-jour

where $\delta(x)$ is a non-negative weight function such that $E[\delta(x)]$ exists.

A special case of interest arises when the weight function is of the form $\delta(x) = x^\alpha$. Such distributions are known as size biased distributions of order α and used by many authors (Patil and Ord, 1976; Patil, 1981; Mahfoud and Patil, 1982).

$$f_\alpha^*(x; \theta) = \frac{x^\alpha f(x; \theta)}{E(X^\alpha)}$$

where $E(X^\alpha) = \int x^\alpha f(x; \theta) dx$ is the α^{th} raw moment of $f(x, \theta)$ and $\alpha = \frac{1}{2}$ was found to provide a good fit to data on the number of albino children in a family ascertained through affected children. When $\alpha = 1$, it is simply called size biased, and has size biased pdf is defined as

$$f^*(x; \theta) = \frac{xf(x; \theta)}{E(X)}$$

The concept of weighted distribution is a useful tool in the selection of appropriate model for observed data, specially when samples are drawn without a proper frame (Patil et al., 1986). Different authors have reviewed and studied the various weighted probability models and illustrated their applications in different fields. (Dey et al., 2015) discussed weighted exponential distribution with its properties and different methods of estimation. Recently (Singh and Das, 2020) discussed statistical properties of a weighted distribution and its application to waiting time data. (Singh et al., 2021) discussed a weighted discrete distribution for the under dispersed data. (Bodhisuwan et al., 2016) introduced the discrete weighted Lindley distribution and (Khan et al., 2018) discussed the weighted modified weibull distribution. (Hassan and Mir, 2008) discussed bayesian estimation of size biased generalized geometric series distribution and its applications. Also (Kilany, 2016) have obtained the weighted version of lomax distribution.

In this study, we have introduced different version of weighted distribution considering geometric distribution as a base distribution and explore its particular case named as Size Biased X-Geometric Distribution (SBXGD). A simulation study is carried out to evaluate the performance of parameter using different sample sizes, also obtained average bias and mean square error. We have tried to show the applicability of a SBXGD by considering four data sets. These data sets are fitted to both SBXGD and SBGD based on some well known goodness of fit statistics.

2 Proposed Distribution

We have considered a geometric distribution $P^*(x)$ as baseline distribution and a weight function $w(x)$ is considered as

$$P^*(x) = \left(\frac{\lambda}{1+\lambda}\right) \left(\frac{1}{1+\lambda}\right)^x \text{ and } w(x) = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)}$$

Then the weighted distribution is define as

$$P(x) = \frac{w(x)P^*(x)}{E[w(x)]} \quad (2.1)$$

Now we know that the weight function

$$\begin{aligned}\frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} &= \binom{x+r-1}{r-1} = \binom{x+r-1}{x} = \frac{(x+r-1)(x+r-2)\dots(r+1)r}{x!} \\ &= \frac{(-1)^x(-r)(-r-1)\dots(-r-x+2)(-r-x+1)}{x!} \\ &= (-1)^x \binom{-r}{x}\end{aligned}\quad (2.2)$$

and the expectation of the weight function

$$\begin{aligned}E[w(x)] &= \sum_{x=0}^{\infty} \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \frac{\lambda}{1+\lambda} \left(\frac{1}{1+\lambda}\right)^x \\ &= \frac{\lambda}{1+\lambda} \sum_{x=0}^{\infty} \binom{-r}{x} \left(\frac{-1}{1+\lambda}\right)^x \\ &= \frac{\lambda}{1+\lambda} \left(1 - \frac{1}{1+\lambda}\right)^{-r} \\ &= \left(\frac{\lambda}{1+\lambda}\right)^{-r+1}\end{aligned}\quad (2.3)$$

from equation (2.1) and (2.3) , we have a weighted geometric distribution as

$$P(x) = \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \left(\frac{\lambda}{1+\lambda}\right)^r \left(\frac{1}{1+\lambda}\right)^x \quad (2.4)$$

This is similar to the negative binomial distribution if $p = \frac{\lambda}{1+\lambda}$ and $q = \frac{1}{1+\lambda}$

2.1 Particular case of weighted geometric distribution

Case 1. When $r=1$ in equation (2.4) then we get

$$P^*(x) = \frac{\lambda}{1+\lambda} \left(\frac{1}{1+\lambda}\right)^x \quad ; \quad x = 0, 1, 2, 3, \dots, \lambda > 0 \quad (2.5)$$

This is geometric distribution with expectation $E(x) = \frac{1}{\lambda}$ and the size biased geometric distribution is as follows

$$P_1(x) = x\lambda \left(\frac{\lambda}{1+\lambda}\right) \left(\frac{1}{1+\lambda}\right)^x \quad ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.6)$$

and the expectation of the size biased geometric distribution is $E(x) = \frac{\lambda+2}{\lambda}$.

The corresponding cumulative distribution function (cdf) and survival function are:

$$G_1(x) = 1 - \frac{1}{(1+\lambda)^x} \quad ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.7)$$

and

$$S_1(x) = \frac{1}{(1+\lambda)^x} \quad ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.8)$$

Thus the hazard of SBGD is given by

$$h_1(x) = \frac{x\lambda^2}{1+\lambda} \quad ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.9)$$

The hazard of SBGD is depending on x , therefore as x is increasing hazard of SBGD is also increasing however hazard of geometric distribution is constant.

Case 2. When $r=2$ in equation (2.4) then we get

$$P_2(x) = (x+1) \left(\frac{\lambda}{1+\lambda} \right)^2 \left(\frac{1}{1+\lambda} \right)^x ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.10)$$

This distribution is named as X -geometric distribution and the expectation is $\frac{2}{\lambda}$ and the size biased X -geometric distribution is as follows

$$P_{2sb}(x) = \frac{\lambda^3}{2(1+\lambda)^{x+2}} x(x+1) ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.11)$$

Average of the size biased X -geometric distribution is $E(x) = \frac{\lambda+3}{\lambda}$. The corresponding cumulative distribution function and survival rate function are:

$$G_{2sb}(x) = 1 - \frac{1}{(1+\lambda)^x} \left[1 + \frac{\lambda x(2 + (1-x)\lambda)}{2(1+\lambda)^2} \right] ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.12)$$

and

$$S_{2sb}(x) = \frac{1}{(1+\lambda)^x} \left[1 + \frac{\lambda x(2 + (1-x)\lambda)}{2(1+\lambda)^2} \right] ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.13)$$

Thus the hazard function of SBXGD is given by

$$h_{2sb}(x) = \frac{\lambda^3 x(x+1)}{[2(1+\lambda)^2 + \lambda x(2 + (1-x)\lambda)]} ; \quad x = 1, 2, 3, \dots, \lambda > 0 \quad (2.14)$$

This is also increasing hazard.

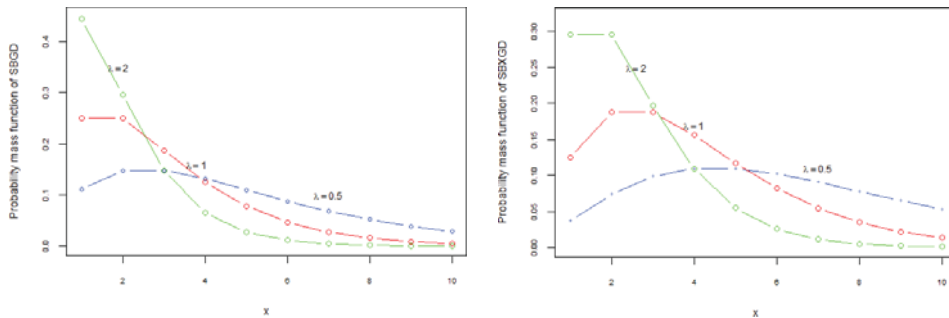


Figure 1: Probability mass function of (i) Size biased geometric distribution and (ii) Size biased X -geometric distribution respectively as a function of x for $\lambda = 0.5, \lambda = 1$ and $\lambda = 2$

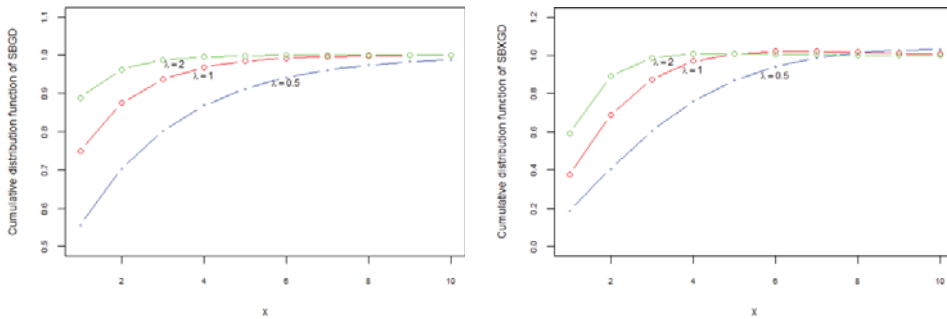


Figure 2: Cumulative distribution function of (i) Size biased geometric distribution and (ii) Size biased X-geometric distribution respectively as a function of x for $\lambda = 0.5, \lambda = 1$ and $\lambda = 2$

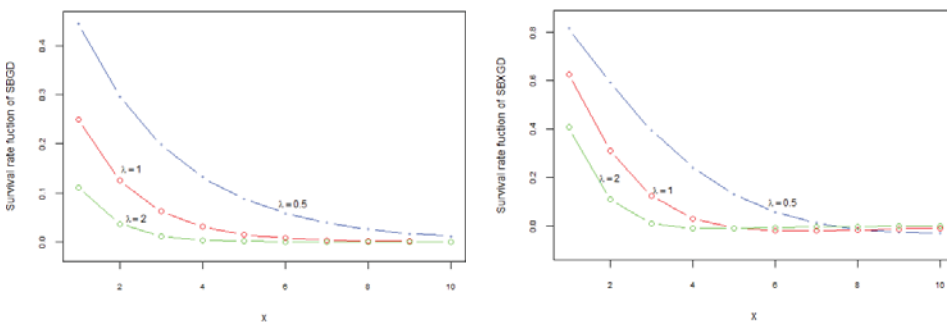


Figure 3: Survival rate function of (i) Size biased geometric distribution and (ii) Size biased X-geometric distribution respectively as a function of x for $\lambda = 0.5, \lambda = 1$ and $\lambda = 2$

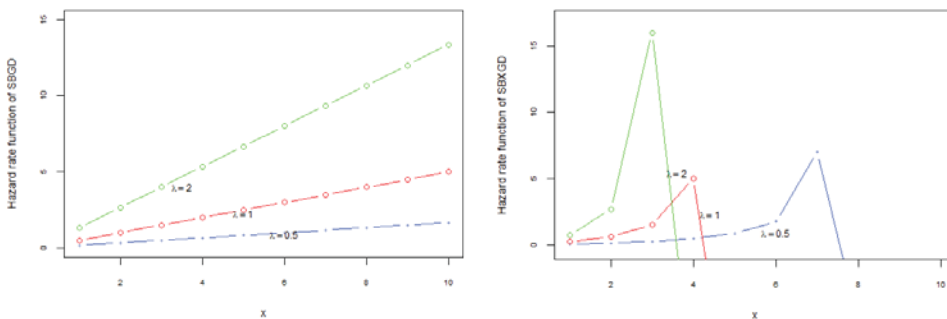


Figure 4: Hazard rate function of (i) Size biased geometric distribution and (ii) Size biased X-geometric distribution respectively as a function of x for $\lambda = 0.5, \lambda = 1$ and $\lambda = 2$

3 Mathematical Properties

3.1 Mode

Let X be a discrete, non-negative, integer-valued random variable such that $P_{2sb}(X = x) = p_x$, $x = 1, 2, 3, \dots$. Now if $X = x$ is the mode, then the following condition (Broca, 2006) must be satisfied:

$$p_{x-1} \leq p_x \geq p_{x+1} \quad (3.1)$$

For application purposes, the above condition may be broken down into two inequalities: $\frac{p_x}{p_{x-1}} \geq 1$ and $\frac{p_{x+1}}{p_x} \leq 1$. Satisfying both of these inequalities will lead to the identification of the mode(s) almost instantly.

Then mode for the size biased X -geometric distribution:

Case 1.

$$\frac{p(x)}{p(x-1)} = \frac{\frac{\lambda^3 x(x+1)}{2(1+\lambda)^{x+2}}}{\frac{\lambda^3 x(x-1)}{2(1+\lambda)^{x+1}}} \geq 1 \quad (3.2)$$

$$\implies \frac{x+1}{x-1} \frac{1}{1+\lambda} \geq 1 \quad (3.3)$$

$$\implies x \leq \frac{2+\lambda}{\lambda} \quad (3.4)$$

Case 2. If x has been replaced with $x+1$ and the direction of the inequality has reversed; this obviates the necessity of computations (16) and (17), immediately leading to:

$$x+1 \geq \frac{2+\lambda}{\lambda} \implies x \geq \frac{2}{\lambda} \quad (3.5)$$

Since x is a non-negative integer which also must satisfy $\frac{2}{\lambda} \leq x \leq \frac{2+\lambda}{\lambda}$, so clearly $x = [\lambda]$ where $[\lambda]$ denotes the greatest integer that does not exceed λ . If the parameter λ is itself a positive integer then, by virtue of $\frac{2}{\lambda} \leq x \leq \frac{2+\lambda}{\lambda}$, the size biased X -geometric distribution is bimodal with two model located at $x = \frac{2}{\lambda}$ and $x = \frac{2+\lambda}{\lambda}$. Similarly, by using the above method to find mode, we have obtained that size biased geometric distribution is also bimodal with two model located at $x = \frac{1}{\lambda}$ and $x = \frac{1+\lambda}{\lambda}$.

3.2 Moments and Related Measures

Moment generating function of SBXGD and SBGD are respectively

$$M_x(t) = \frac{\lambda^3 e^t}{(1 + \lambda - e^t)^3} \quad (3.6)$$

and

$$M'_x(t) = \lambda^2 e^t (1 + \lambda - e^t)^{-2} \quad (3.7)$$

The first four moments about the origin for SBXGD are

$$\begin{aligned}\mu'_1 &= \frac{\lambda + 3}{\lambda}; \quad \mu'_2 = \frac{\lambda^2 + 9\lambda + 12}{\lambda^2}; \quad \mu'_3 = \frac{\lambda^3 + 21\lambda^2 + 72\lambda + 60}{\lambda^3}; \\ \mu'_4 &= \frac{\lambda^4 + 45\lambda^3 + 300\lambda^2 + 600\lambda + 360}{\lambda^4}\end{aligned}$$

Using the above expression, we get central moments of SBXGD which is given by

$$\mu_1 = \frac{\lambda + 3}{\lambda}; \quad \mu_2 = \frac{3(\lambda + 1)}{\lambda^2}; \quad \mu_3 = \frac{3(\lambda + 1)(\lambda + 2)}{\lambda^3}; \quad \mu_4 = \frac{3(\lambda + 1)(\lambda^2 + 15\lambda + 15)}{\lambda^4}$$

using central moments of SBXGD, we get the coefficient of variation (γ), coefficient of skewness ($\sqrt{\beta_1}$) and the coefficient of kurtosis (β_2) as

$$\gamma = \frac{\sqrt{3(\lambda + 1)}}{(\lambda + 3)}; \quad \sqrt{\beta_1} = \frac{\lambda + 2}{\sqrt{3(\lambda + 1)}}; \quad \beta_2 = \frac{\lambda^2 + 15\lambda + 15}{3(\lambda + 1)}$$

Similarly, we obtain first four moments about the origin for SBGD as

$$\begin{aligned}\mu'_1 &= \frac{\lambda + 2}{\lambda}; \quad \mu'_2 = \frac{\lambda^2 + 6\lambda + 6}{\lambda^2}; \quad \mu'_3 = \frac{\lambda^3 + 14\lambda^2 + 36\lambda + 24}{\lambda^3}; \\ \mu'_4 &= \frac{\lambda^4 + 30\lambda^3 + 150\lambda^2 + 240\lambda + 120}{\lambda^4}\end{aligned}$$

Thus central moments of SBGD are

$$\mu_1 = \frac{\lambda + 2}{\lambda}; \quad \mu_2 = \frac{2(\lambda + 1)}{\lambda^2}; \quad \mu_3 = \frac{2(\lambda + 1)(\lambda + 2)}{\lambda^3}; \quad \mu_4 = \frac{2(\lambda^3 + 13\lambda^2 + 24\lambda + 12)}{\lambda^4}$$

The coefficient of variation (γ), skewness ($\sqrt{\beta_1}$) and the kurtosis (β_2) for SBGD are:

$$\gamma = \frac{\sqrt{2(\lambda + 1)}}{(\lambda + 2)}; \quad \sqrt{\beta_1} = \frac{\lambda + 2}{\sqrt{2(\lambda + 1)}}; \quad \beta_2 = \frac{\lambda^3 + 13\lambda^2 + 24\lambda + 12}{(\lambda + 1)^2}$$

3.3 Mean Deviation

The mean deviation about the mean and the median is defined as

$$\delta_1(X) = \sum_{x=1}^{\infty} |x - \mu| P_{2sb}(x)$$

and

$$\delta_2(X) = \sum_{x=1}^{\infty} |x - M| P_{2sb}(x)$$

respectively, where $\mu = E(x)$ and $M = \text{Median}(X)$. These measures can be calculated using the relationship that

$$\begin{aligned}
 E|X - m| &= \sum_{x=1}^{\infty} (m - x)P_{2sb}(x) + \sum_{x=m}^{\infty} (x - m)P_{2sb}(x) \\
 &= 2 \sum_{x=1}^{\infty} (m - x)P_{2sb}(x) \\
 &= 2 \left(mG_{2sb}(m) - \sum_{x=1}^{\infty} xP_{2sb}(x) \right)
 \end{aligned} \tag{3.8}$$

The expression of mean deviation about the mean and the median for SBXGD as

$$\delta_1(X) = 2 \left[\mu - \frac{3 + \lambda}{\lambda} + \frac{1}{2(1 + \lambda)^\mu} \left(\frac{3\lambda^2\mu^2 + (2\lambda - 1)\lambda\mu + (\lambda + 1)(\lambda + 2)}{2(1 + \lambda)^2} + \frac{3 + \lambda}{\lambda} \right) \right] \tag{3.9}$$

and

$$\delta_2(X) = 2 \left[M - \frac{3 + \lambda}{\lambda} + \frac{1}{2(1 + \lambda)^M} \left(\frac{3\lambda^2M^2 + (2\lambda - 1)\lambda M + (\lambda + 1)(\lambda + 2)}{2(1 + \lambda)^2} + \frac{3 + \lambda}{\lambda} \right) \right] \tag{3.10}$$

Similarly, we find the expression of mean deviation about the mean and the median for SBGD which is written as respectively

$$\delta'_1(X) = 2 \left[(\mu - 1) + \frac{(\mu - 1)}{(1 + \lambda)^\mu} + \frac{1}{(1 + \lambda)^{\mu+1}} \left(\mu^2\lambda - \frac{2((1 + \lambda)^{\mu+1} - 1)}{\lambda} \right) \right] \tag{3.11}$$

and

$$\delta'_2(X) = 2 \left[(M - 1) + \frac{(M - 1)}{(1 + \lambda)^M} + \frac{1}{(1 + \lambda)^{M+1}} \left(M^2\lambda - \frac{2((1 + \lambda)^{M+1} - 1)}{\lambda} \right) \right] \tag{3.12}$$

3.4 Lorenz Curve

According to (Haitovsky, 2001), the Lorenz curve (also known as the relative concentration curve) is a popular tool for displaying and analysing the size distribution of income. When the units are ordered in increasing order of their income, it connects the cumulative income fraction earned by various proportions of the population to the proportion receiving this income. The Lorenz curve for a positive random variable X is defined as the graph of the ratio

$$L(G_{2sb}(x)) = \frac{E(X/X \leq x)G_{2sb}(x)}{E(X)} \tag{3.13}$$

Against $G_{2sb}(x)$ with the properties $L(p) \leq p$, $L(0) = 0$ and $L(1) = 1$. If X represents annual income, $L(p)$ is the proportion of total income that accrues to individuals having the 100% lowest incomes. If all individual earn the same income then $L(p) = p$ for all p . The area between the line $L_2(p) = p$ and the Lorenz curve may be regarded as a measure of inequality of income, or more generally, of the variability of X as suggested by (Ghitany et al., 2008).

From the size biased X -geometric distribution in (2.11)

$$\begin{aligned}
 E(X/X \leq x)G_{2sb}(x) &= \frac{3+\lambda}{\lambda} - \frac{3+\lambda}{\lambda} \left[\frac{2(1+\lambda)^2 + x\lambda + x(1-x)\lambda^2}{2(1+\lambda)^{x+2}} \right] \\
 &\quad - \left[\frac{x(3x-1)\lambda(1+\lambda) + x^2(x+1)\lambda^2}{2(1+\lambda)^{x+2}} \right] \\
 &\quad + \left[\frac{x(3x-1)(1+\lambda) + x^2(x+1)\lambda^2}{2(1+\lambda)^{x+2}} \right] \left[\frac{2(1+\lambda)^2 + x\lambda + x(1-x)\lambda^2}{2(1+\lambda)^{x+2}} \right] \quad (3.14)
 \end{aligned}$$

Thus the Lorenz curve for SBXGD as

$$L(p) = 1 - \frac{\lambda p}{4(\lambda+3)(1+\lambda)^{2(x+2)}} \left[2(\lambda+3)(1+\lambda)^{x+2} - (x^3 + 4x^2 - x)\lambda^3 + (3x^2 - x)\lambda^2 \right] \quad (3.15)$$

where, $x = G_{2sb}^{-1}(p)$.

3.5 Stochastic Orderings

Stochastic ordering of discrete random variables is a main tool which specifies that one random variable is bigger than other. A stochastic ordering is any way to compare probability distributions, i.e., any binary relation in a set of probability distributions. The most prevalent explanation is that one distribution assigns higher probability to larger values in some sense than the other. We shall remember some basic definitions. A random variable X is said to be smaller than a random variable X^* in the

- (i) Stochastic order ($X \leq_{St} X^*$) if $G_X(x) \geq G_{X^*}(x)$ for all x ;
- (ii) Hazard rate order ($X \leq_{hr} X^*$) if $h_X(x) \geq h_{X^*}(x)$ for all x ;
- (iii) Mean residual life order ($X \leq_{mrl} X^*$) if $m_X(x) \leq m_{X^*}(x)$ for all x ;
- (iv) Likelihood ratio order ($X \leq_{lr} X^*$) if $\frac{f_X(x)}{f_{X^*}(x)}$ decreases in x .

The following implication (Shaked and Shanthikumar, 1988) is given by:

$$\begin{aligned}
 X \leq_{lr} X^* &\implies X \leq_{hr} X^* \implies X \leq_{mrl} X^* \\
 &\Downarrow \\
 &X \leq_{St} X^* \quad (3.16)
 \end{aligned}$$

The size biased X -geometric distribution are ordered with respect to the strongest likelihood ratio as shown in the following theorem.

Theorem1: Let $X \sim SBXGD(\lambda_1)$ and $X^* \sim SBXGD(\lambda_2)$. If $\lambda_1 \geq \lambda_2$ then $X \leq_{lr} X^*$ and hence $X \leq_{hr} X^*$, $X \leq_{mrl} X^*$ and $X \leq_{St} X^*$.

Proof: First we know that

$$\frac{P_X(x)}{P_{X^*}(x)} = \left(\frac{\lambda_1}{\lambda_2} \right)^3 \left(\frac{1+\lambda_2}{1+\lambda_1} \right)^2 \left(\frac{1+\lambda_2}{1+\lambda_1} \right)^x$$

since for $\lambda_1 > \lambda_2$

$$\frac{d}{dx} \frac{P_X(x)}{P_{X^*}(x)} = \left(\frac{\lambda_1}{\lambda_2}\right)^3 \left(\frac{1+\lambda_2}{1+\lambda_1}\right)^2 \left(\frac{1+\lambda_2}{1+\lambda_1}\right)^x \ln\left(\frac{1+\lambda_2}{1+\lambda_1}\right)$$

$$\frac{d}{dx} \frac{P_X(x)}{P_{X^*}(x)} = \frac{P_X(x)}{P_{X^*}(x)} \ln\left(\frac{1+\lambda_2}{1+\lambda_1}\right) < 0$$

$\frac{P_X(x)}{P_{X^*}(x)}$ is decreasing in x that is $X \leq_{lr} X^*$. The remaining statements follow from the implications in equation (30).

4 Estimation

Maximum Likelihood Estimation: Given a random sample $X_1, X_2, \dots, X_n$ from the size biased X -geometric distribution, the maximum likelihood estimate of λ is given by

$$\begin{aligned} \ln L &= 3n \ln \lambda + \sum_i \ln(x_i(x_i + 1)) - n \ln 2 - \left(\sum_i (x_i + 2)\right) \ln(1 + \lambda) \\ \frac{d}{d\lambda} \ln L &= \frac{3n}{\lambda} - \frac{\sum_i (x_i + 2)}{1 + \lambda} \\ \hat{\lambda} &= \frac{3}{\bar{x} - 1} \end{aligned} \quad (4.1)$$

Thus both, method of moment and maximum likelihood estimators of the parameter λ are the same. The Fisher information of size biased X -geometric distribution is

$$I_{2sb}(\lambda) = -E \left[\frac{\delta^2}{\delta \lambda^2} \ln P_{2sb}(x; \lambda) \right]$$

where

$$\frac{\delta^2}{\delta \lambda^2} \ln P_{2sb}(x; \lambda) = -\frac{3n}{\lambda^2} + \frac{\sum_i (x_i + 2)}{(1 + \lambda)^2}$$

then

$$I_{2sb}(\lambda) = E \left[\frac{3n}{\lambda^2} - \frac{\sum_i (x_i + 2)}{(1 + \lambda)^2} \right] = \frac{3n}{\lambda^2(1 + \lambda)} \quad (4.2)$$

In similar manner, we have obtain the expression of maximum likelihood and Fisher information for SBGD which is written as respectively, the ML estimate of SBGD is $\hat{\lambda} = \frac{2}{\bar{x}-1}$ and fisher information is $I_1(\lambda) = \frac{2n}{\lambda^2(1+\lambda)}$.

5 Simulation Study

Here, the simulation study for checking the performance of the maximum likelihood estimate given by equation (4.1) with respect to sample size n was carried out in the following steps:

1. Since the direct computation of equation $X = G_{2sb}^{-1}(U)$, where $U \sim U(0, 1)$, cannot be possible, so the inverse transformation method for generating random data from size biased X -geometric distribution fails. Therefore the rejection method is used for generating random

data for SBXGD with parameter λ as provided by (Ross, 2006).

2. Compute Maximum Likelihood Estimates ($\hat{\lambda}_i$) for $i = 1, 2, \dots, 20,000$.

3. Compute Biases and Mean Squared Error by given formula:

$$bias(n) = \frac{1}{20,000} \sum_{i=1}^{20,000} (\hat{\lambda}_i - \lambda) \quad \text{and} \quad MSE(n) = \frac{1}{20,000} \sum_{i=1}^{20,000} (\hat{\lambda}_i - \lambda)^2 \quad (5.1)$$

The steps are repeated for $n = 20, 40, 60, \dots, 100$ with different values of λ i.e. $\lambda = 0.05, 0.1, 0.15, 0.2$.

Table 1: Average bias of the estimator $\hat{\lambda}$

n	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.15$	$\lambda = 0.2$
20	0.000773	0.001935	0.003002	0.004258
40	0.000414	0.000788	0.001522	0.001938
60	0.000272	0.000654	0.000833	0.001417
80	0.000234	0.000426	0.000665	0.001034
100	0.000156	0.000395	0.000611	0.000878

Table 2: Average MSE of the estimator $\hat{\lambda}$

n	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.15$	$\lambda = 0.2$
20	0.000047	0.000199	0.000474	0.000879
40	0.000023	0.000094	0.000229	0.000417
60	0.000015	0.000064	0.000145	0.000279
80	0.000011	0.000047	0.000108	0.000205
100	0.000009	0.000038	0.000091	0.000164

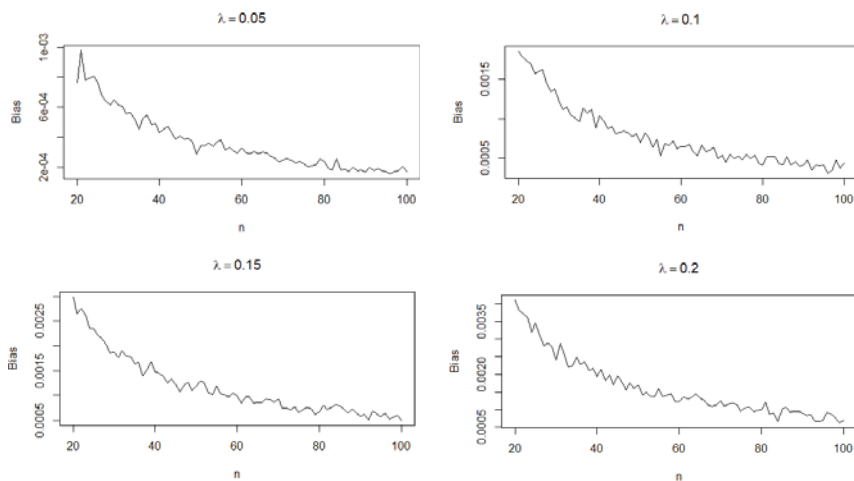


Figure 5: Bias(n) versus $n = 20, 40, 60, 80, 100$ for $\lambda = 0.05, 0.1, 0.15, 0.2$.

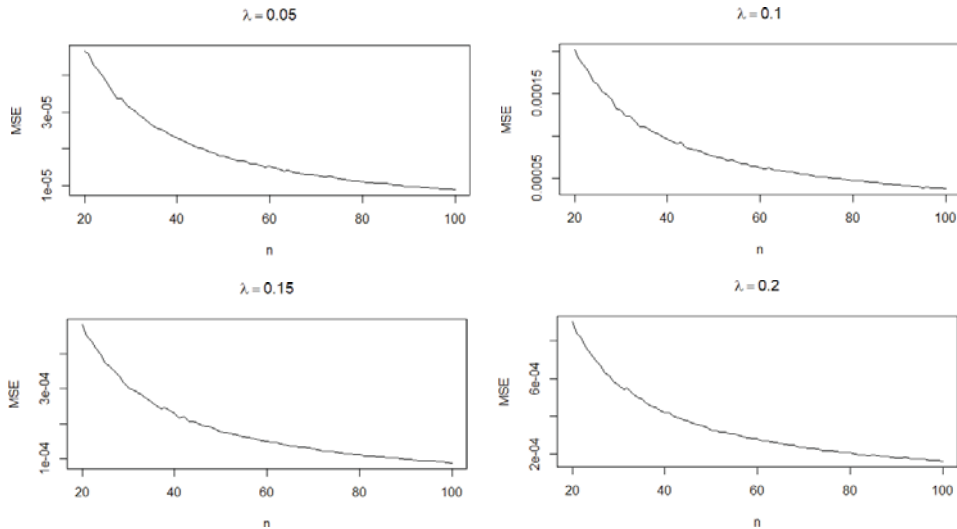


Figure 6: $MSE(n)$ versus $n = 20, 40, 60, 80, 100$ for $\lambda = 0.05, 0.1, 0.15, 0.2$.

Remark- Table 1 and Table 2 shows the result of simulation study in terms of bias and mean squared error for size biased X -geometric distribution with parameter λ .

- (i) Table 1 shows that as sample size n increases bias decreases and as parameter λ increases bias is also increases for size biased X -geometric distribution.
- (ii) Table 2 shows that as sample size n increases mean squared error decreases and as parameter λ increases mean squared error are also increases.

6 Real Data Application

In this section, we use four real data sets, given in Table 3, first data set represents an epidemic of cholera in a village in India where x is number of cholera cases in a house and n denotes number of houses with x cholera cases taken from (Dahiya and Gross, 1973). Second data set represents data on any type of abortion in Uttar Pradesh studied by (Singh and Shukla, 2017). Third data set represents data on European corn-borer provided by (McGuire et al., 1975). Fourth data set represents data on thunderstorm events at Cape Kennedy, for July provided by (Falls et al., 1971). We have fitted size biased X -geometric distribution for below four datasets and compared the result with size biased geometric distribution. Table 4, 5 and 6 provides the goodness of fit for all four data sets which clearly shows that SBXGD has smaller chi-square value, $-2 \log$ likelihood ($-2LL$), Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC) values and larger p -value than SBGD which revealed that the SBXGD provides the best fit for below datasets in comparison to SBGD.

Table 3: Distribution of various data sets

Data Set	x	1	2	3	4	5
1	n	32	16	6	1	-
2	n	33	14	4	1	-
3	n	83	36	14	2	1
4	n	80	47	26	9	2

Table 4: Fitted estimate of 1st data set

Model	Estimate	χ^2 -value	p -value	-2LL	AIC	BIC	AICc
SBXGD	5.323	0.129	0.719	109.88	111.88	113.89	111.96
SBGD	3.548	0.260	0.610	110.41	112.41	114.41	112.48

Table 5: Fitted estimate of 2nd data set

Model	Estimate	χ^2 -value	p -value	-2LL	AIC	BIC	AICc
SBXGD	6.240	0.026	0.871	95.660	97.66	99.610	97.74
SBGD	4.160	0.090	0.765	95.88	97.880	99.830	97.96

Table 6: Fitted estimate of 3rd data set

Model	Estimate	χ^2 -value	p -value	-2LL	AIC	BIC	AICc
SBXGD	5.514	0.698	0.705	269.45	271.45	274.36	271.48
SBGD	3.676	0.868	0.648	269.75	271.75	274.66	271.78

Table 7: Fitted estimate of 4th data set

Model	Estimate	χ^2 -value	p -value	-2LL	AIC	BIC	AICc
SBXGD	3.672	1.152	0.562	401.42	403.42	406.52	403.45
SBGD	2.448	1.370	0.504	402.75	404.75	407.85	404.77

7 Conclusions

In this paper, a different version of weighted distribution is introduced and studying one of its particular case i.e. a new size biased distribution named size biased X -geometric distribution (SBXGD) is proposed. Exact expression for probability mass function, cumulative distribution function, survival function and hazard function has been obtained. Moment generating function and its related measures such as moments about origin and mean, coefficient of variation, skewness and kurtosis has been obtained. Other properties like mode, mean deviation, Lorenz curve and stochastic ordering are derived. The estimation of parameter has been approached by maximum likelihood method (MLE) and we have done simulation study to understand the performance of bias and mean squared error (MSE) of the MLE of the parameter. Applying proposed distribution on real data sets reveals that size biased X -geometric distribution

(SBXGD) explains the data better than size biased geometric distribution (SBGD). Therefore the proposed distribution may be considered as potential alternative for the zero truncated data.

Acknowledgment

No funding was received to conduct this study.

References

- Ahmad, Z., Hamedani, G. and Elgarhy, M. 2020. The Weighted Exponentiated Family of Distributions: Properties, Applications and Characterizations. *Journal of The Iranian Statistical Society*, **19**(1), 209-228.
- Tobin, J. 1958, 'Liquidity preference as behavior toward risk', *Review of Economic Studies* **25**(2), 65-86.
- Bakouch, H. S., Jazi, M. A. and Nadarajah, S. 2014. A new discrete distribution. *Statistics*, **48**(1), 200-240.
- Bodhisuwan, W. and Sangpoom, S. 2016. The discrete weighted Lindley distribution. *In 2016 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA), IEEE*, 99-103.
- Broca, D. S. 2006. The mode of a probability mass function. *International Journal of Mathematical Education in Science and Technology*, **37**(3), 376-378.
- Dagum, C. 1985. Lorenz Curve, encyclopedia of statistical sciences. *S. Kotz, NL Johnson and CB Read*, 5, 156-161.
- Dahiya, R. C. and Gross, A. J. 1973. Estimating the zero class from a truncated Poisson sample. *Journal of the American Statistical Association*, **68**(343), 731-733.
- Dennis, B. and Patil, G. P. 1984. The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences*, **68**(2), 187-212.
- Dey, S., Ali, S. and Park, C. 2015. Weighted exponential distribution: properties and different methods of estimation. *Journal of Statistical Computation and Simulation*, **85**(18), 3641-3661.
- Falls, L. W., Williford, W. O. and Carter, M. C. 1971. Probability distributions for thunderstorm activity at Cape Kennedy, Florida. *Journal of Applied Meteorology and Climatology*, **10**(1), 97-104.
- Fisher, R. A. 1934. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of eugenics*, **6**(1), 13-25.

- Gail, M. H. and Gastwirth, J. L. 1978. *A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve. Journal of the American Statistical Association*, **73**(364), 787-793.
- Ghitany, M. E., Atieh, B. and Nadarajah, S. 2008. Lindley distribution and its application. *Mathematics and computers in simulation*, **78**(4), 493-506.
- Gove, J. H. 2003. Estimation and applications of size-biased distributions in forestry. *Modeling Forest systems*, A. Amaro, D. Reed and P. Soares CAB International Wallingford UK, 201-212.
- Haitovsky, Y. 2001. Statistical analysis, special problems of: grouped observations.
- Hassan, A. and Mir, K. A. 2008. On Bayesian estimation of size-biased generalized geometric series distribution and its applications. *Journal of Statistical Research*, **42**(1), 117-126.
- Khan, M. N., Saeed, A. and Alzaatreh, A. 2018. Weighted Modified Weibull distribution. *Journal of Testing and Evaluation*, **47**(5), 3751-3764.
- Kilany, N. M. 2016. Weighted lomax distribution. *SpringerPlus*, **5**(1), 1-18.
- Mahfoud, M. and Patil, G. P. 1982. On weighted distributions. *Statistics and probability: essays in honor of CR Rao*, 479-492.
- McGuire, J. U., Brindley, T. A. and Bancroft, T. A. 1957. The distribution of European corn borer larvae *Pyrausta nubilalis* (HBN.), in field corn. *Biometrics*, **13**(1), 65-78.
- Patil, G. P. and Ord, J. K. 1976. On size-biased sampling and related form-invariant weighted distributions. *Sankhy: The Indian Journal of Statistics, Series B*, 48-61.
- Patil, G. P. Rao, C. R., and Ratnaparkhi, M. V. 1986. On discrete weighted distributions and their use in model choice for observed data. *Communications in Statistics-Theory and Methods*, **15**(3), 907-918.
- Patil, G. P. 1981. Studies in statistical ecology involving weighted distributions. *Proceedings of the Indian Statistical Institute Jubilee International Conference on Statistics: Applications and New Directions*, 478503.
- Rao, C.R. 1965. On discrete distributions arising out of methods of ascertainment. In: Patil, G.P. (ed.) *Classical and Contagious Discrete Distributions. Statistical Publishing Society, Calcutta*, 320-332.
- Ross, S. M. 2006. *Simulation. Elsevier Academic press.*
- Shaked, M. and Shanthikumar, J. G. 1988. Stochastic convexity and its applications *Advances in Applied Probability*, 427-446.
- Singh, Brijesh P. and Shukla U. 2017. An Extension of Poisson Distribution and its Applications in Human Reproduction. *Journal of Data Science*, **15**(1), 155-166.

- Singh, Brijesh P. and Das, U. D. 2020. Some Statistical Properties of a Weighted Distribution and Its Application to Waiting Time Data. *Asian Journal of Probability and Statistics*, **9**(1), 1-12.
- Singh, Brijesh P., Singh., G., Das, U. D. and Maurya, D. K. 2021. An Under-Dispersed Discrete Distribution and Its Application, *Journal of Statistics Applications & Probability Letters*, **8**(3), 205–213.
- Zelen, M. 1974. Problems in cell kinetics and the early detection of disease. *Reliability and biometry*, **56**(3), 701-726.
- Zelen, M. and Feinleib, M. 1969. On the theory of screening for chronic diseases. *Biometrika*, **56**(3), 601-614.