A Discrete Analogue of the Continuous Marshall-Olkin Weibull Distribution with Application to Count Data

Festus C. Opone\(^1\)*, Elvis A. Izekor\(^1\), Innocent U. Akata\(^1\) and Francis E. U. Osagiede\(^2\)

\(^1\) Department of Statistics, University of Benin, Nigeria
e-mail: festus.opone@physci.uniben.edu

\(^2\) Department of Mathematics, University of Benin, Nigeria

Abstract

In this paper, we introduced the discrete analogue of the continuous Marshall-Olkin Weibull distribution using the discrete concentration approach. Some mathematical properties of the proposed discrete distribution such as the probability mass function, cumulative distribution function, survival function, hazard rate function, second rate of failure, probability generating function, quantile function and moments are derived. The method of maximum likelihood estimation is employed to estimate the unknown parameters of the proposed distribution. The applicability of the proposed discrete distribution was examined using an over-dispersed and under-dispersed data sets.

1. Introduction

The theory of discretization generally arises when it becomes seemingly difficult to measure the life length of a product or device on a continuous scale. Situation of such arises when the observed lifetimes are preferably recorded on a discrete scale than on a continuous analogue. For examples the number of times devices are switched on/off, the number of days a patient stays in an observation ward, and the number of weeks/months/years a cancer patient survives after treatment etc. Although, classical discrete distributions such as the Poisson, Geometric, Binomial and Negative Binomial distributions have been developed to handle such situation, there is still need to introduce more flexible discrete distributions especially those arising from the discretization of continuous distributions to handle more sophisticated real-life phenomena.
Several methods of discretizing a continuous distribution have been widely considered in literature. [3] gave a comprehensive survey on the different methods of generating discrete probability distributions as analogues of continuous distributions which include; the discretization method based on the survival function, the discretization method based on probability mass function (infinite series), the discretization method based on the cumulative distribution function, the discretization method based on the hazard function, discretization method based on reverse hazard function, the difference equation analogues of Pearsonian differential equation and the two-stage composite method etc. The method of discretization by survival function also referred to as the discrete concentration approach was proposed by [13].

Let \( X \) be a random variable associated to a continuous probability distribution with survival function \( S_X(x) \), [9] defined the probability mass function (pmf) of a discrete random variable \( Y = \lfloor X \rfloor \), where \( \lfloor X \rfloor \) indicates the smallest integer part or equal to \( X \), as:

\[
P(Y = k) = \sum_{j=0}^{1} (-1)^j S_X(k + j) = S_X(k) - S_X(k + 1). \tag{1}
\]


In this paper, using the same method of discretization defined in equation (1), we introduced the discrete analogue of the continuous Marshall-Olkin Weibull (DMOW) distribution. The rest Sections of this paper are as follows: Section 2 proposes the new discrete Marshall-Olkin Weibull (DMOW) distribution and presents some of its mathematical properties, the estimation of the unknown parameters of the proposed distribution is presented in Section 3, while Section 4 presents the application of the proposed distribution to two count data sets. Section 5 concludes the paper.
2. The Proposed Discrete Marshall-Olkin Weibull (DMOW) Distribution

[12] introduced the Marshall-Olkin family of distributions obtained by adding a scale parameter \( \alpha \), which they called the “tilt parameter” to an existing distribution. They derived the Marshall-Olkin Weibull distribution with the survival function given by

\[
S(x) = \frac{\alpha e^{-\theta x}}{1 - \alpha e^{-\theta x}} \quad x > 0, \quad \alpha, \theta, \beta > 0,
\]

(2)

where \( \alpha = (1 - \alpha) \) is called a “tilt parameter”.

By substituting equation (2) into (1), we defined the discrete analogue of the continuous Marshall-Olkin Weibull distribution with probability mass function (pmf) expressed as

\[
P(x) = \frac{\alpha \{\gamma e^{\beta x} - \gamma^{(x+1)} e^{\beta x}\}}{1 - \alpha \gamma e^{\beta x}} \quad x > 0, \quad \alpha, \beta > 0, \quad \gamma = e^{-\theta}.
\]

(3)

The probability mass function of the discrete Marshall-Olkin Weibull distribution defined in equation (3) can be expressed as an infinite mixture (with weights \( \alpha \alpha^k \)) of discrete Weibull distributions with parameter \( \gamma^{(k+1)} \), \( k = 0, 1, 2, \ldots \) and \( 0 < \alpha < 1 \).

Using the series representation in [4],

\[
(1 - Z)^{-j} = \sum_{k=0}^{\infty} \frac{\Gamma(j + k)}{\Gamma(j + 1) \Gamma(j)} Z^k, \quad |Z| < 1, \quad j > 0,
\]

(4)

equation (2) can be rewritten as

\[
S(x) = \alpha \sum_{k=0}^{\infty} \alpha^k \left( \gamma^{(k+1)} \right)^{x+1},
\]

so that,

\[
P(x) = S(x) - S(x + 1) = \alpha \sum_{k=0}^{\infty} \alpha^k \left( \gamma^{(k+1)} x^\beta - \gamma^{(k+1)(x+1)} \right).
\]

(5)
The graphical representation of the probability mass function of the discrete Marshall-Olkin Weibull (DMOW) distribution for varying parameters of the distribution is shown in Figure 1.

![Figure 1. Probability mass function (pmf) of the DMOW distribution.](image)

The plots in Figure 1 reveal that the pmf of the DMOW distribution can be decreasing, left-skewed, right-skewed unimodal and symmetric which proves flexibility in handling any nature of real data sets.

2.1 Distribution and survival functions

The cumulative distribution function of the DMOW distribution is defined by

\[
G(x) = 1 - S(x) + P(X = x) = 1 - \frac{\gamma^{\alpha (x+1)^\beta}}{1 - \alpha \gamma^{(x+1)^\beta}}, \quad x > 0, \quad \alpha, \beta > 0, \quad 0 < \gamma < 1, \tag{6}
\]

the corresponding survival function of the DMOW distribution is obtained as

\[
\overline{G}(x) = \frac{\alpha \gamma^{(x+1)^\beta}}{1 - \alpha \gamma^{(x+1)^\beta}}, \quad x > 0, \quad \alpha, \beta > 0, \quad 0 < \gamma < 1. \tag{7}
\]
2.2 Hazard rate, reversed hazard rate and second rate of failure

The hazard rate, reversed hazard rate and the second rate of failure of the discrete Marshall-Olkin Weibull distribution are respectively defined by

\[
h(x) = \frac{P(X = x)}{G(x)} = \frac{\gamma^\beta - \gamma^{(x+1)^\beta}}{\gamma^{(x+1)^\beta \left[1 - \alpha \gamma^\beta\right]}}, \quad (8)
\]

\[
r^*(x) = \frac{P(X = x)}{G(x)} = \frac{\alpha \gamma^\beta - \gamma^{(x+1)^\beta}}{\left[1 - \gamma^{(x+1)^\beta}\right] \left[1 - \alpha \gamma^\beta\right]}, \quad (9)
\]

and

\[
r^{**}(x) = \log \left\{ \frac{G(x)}{G(x+1)} \right\} = \log \left\{ \frac{\gamma^{(x+1)^\beta}}{\gamma^{(x+2)^\beta}} \left( \frac{1 - \alpha \gamma^{(x+2)^\beta}}{1 - \alpha \gamma^{(x+1)^\beta}} \right) \right\}, \quad \alpha, \beta > 0, \; 0 < \gamma < 1. \quad (10)
\]

Figures 2 and 3 respectively show the plots of the cumulative distribution function and the hazard rate function of the discrete Marshall-Olkin Weibull distribution for varying values of its parameters.

Figure 2. Cumulative distribution function (cdf) of the DMOW distribution.
Figure 3. Hazard rate function of the DMOW distribution.

Clearly, Figure 3 shows that the hazard rate of the DMOW distribution exhibits a decreasing, increasing, bathtub-shape and inverted bathtub-shape properties.

2.3 Probability generating function, moments and quantile function

Let $X$ be a discrete random variable defined in the non-negative integers $\{0, 1, 2, ...\}$. The probability generating function $G_X(S)$ of $X$ is defined by

$$G_X(S) = E(S^X) = \sum_{x=0}^{\infty} S^x P(X = x).$$

Using equation (11), the probability generating function of the DMOW distribution is obtained as

$$G_X(S) = \sum_{x=0}^{\infty} S^x P_x = \sum_{x=0}^{\infty} S^x \left\{ \gamma^\beta - \gamma^{(x+1)\beta} \right\} \frac{1 - \alpha \gamma^\beta}{\left(1 - \alpha \gamma (x+1)\beta\right)}.$$
A Discrete Analogue of the Continuous Marshall-Olkin Weibull Distribution …

\[ S = 1 + \alpha (S - 1) \sum_{x=1}^{\infty} x^{-1} \frac{\gamma^x}{1 - \alpha \gamma^x}. \]  

(12)

The mean of the discrete Marshall-Olkin Weibull distribution can be obtained by taking the first derivative of equation (12) and setting \( S = 1 \), yielding

\[ \mu = G'_X (1) = \alpha \sum_{x=1}^{\infty} \frac{x^\beta}{1 - \alpha x^\beta}. \]

Further derivatives (higher) of the generating function of the DMOW distribution yield the corresponding raw moment about the origin of the distribution. Hence, the variance \( \sigma^2 \), coefficient of skewness \( S_k \), and coefficient of kurtosis \( K_s \) of the DMOW distribution are expressed as

\[ \sigma^2 = \mu'_2 - \mu^2 \]

\[ = \alpha \sum_{x=1}^{\infty} (2x - 1) \frac{\gamma^x}{1 - \alpha \gamma^x} - \left[ \alpha \sum_{x=1}^{\infty} \frac{\gamma^x}{1 - \alpha \gamma^x} \right]^2. \]

\[ S_k = \frac{\mu'_3 - 3 \mu'_2 \mu + 2 \mu^3}{(\mu'_2 - \mu^2)^{1/2}}, \]

\[ K_s = \frac{\mu'_4 - 4 \mu'_3 \mu + 6 \mu'_2 \mu^2 - 3 \mu^4}{(\mu'_2 - \mu^2)^2}. \]

Table 1 shows the nature of the Mean \( \mu \), variance \( \sigma^2 \), coefficient of skewness \( S_k \), and coefficient of kurtosis \( K_s \) of the DMOW distribution for varying values of the parameters.
From Table 1, we observed that for varying values of the parameters, the mean ($\mu$) is either greater or less than the variance ($\sigma^2$) which makes the distribution flexible for analyzing under-dispersed ($\mu > \sigma^2$) and over-dispersed ($\mu < \sigma^2$) data sets. The DMOW distribution also exhibits positively-skewed ($S_k > 0$), and negatively-skewed ($S_k < 0$) shapes which conform with the plots of the pmf of the distribution as shown in Figure 1.

The recurrence relation for generating probabilities of the DMOW distribution is given by

$$P_{x+1} = \frac{\{\gamma^{(x+1)} - \gamma^{(x+2)}\}\{1 - \alpha\gamma^\beta\}}{\{1 - \alpha\gamma^{(x+2)}\}\{\gamma^\beta - \gamma^{(x+1)}\}} P_x, \quad x = 0, 1, 2, \ldots$$

(13)

where,

$$P_0 = \frac{\{1 - \gamma\}}{\{1 - \alpha\gamma\}}.$$

The $p$th quantile of the DMOW distribution denoted by $Q_X (p)$ is given by

$$Q_X (p) = G^{-1}(x),$$
A Discrete Analogue of the Continuous Marshall-Olkin Weibull Distribution …

\[ p = \frac{1 - \gamma^{(x+1)^\beta}}{1 - \alpha \gamma^{(x+1)^\beta}}, \]

\[ (p\alpha - 1)\gamma^{(x+1)^\beta} = p - 1, \]

\[ (x + 1)^\beta \log \gamma = \log \left[ \frac{p - 1}{(p\alpha - 1)} \right]. \]

\[ Q_X(p) = \left\{ \frac{\log \left( \frac{p - 1}{p\alpha - 1} \right)}{\log \gamma} \right\}^{1/\beta} - 1. \quad (14) \]

The median of the DMOW distribution is obtained from equation (14) by substituting \( p = 1/2 \), yielding

\[ Q_X(1/2) = \left\{ \frac{\log \left( 1 - \alpha \right)}{\log \gamma} \right\}^{1/\beta} - 1. \quad (15) \]

3. Parameter Estimation

Suppose \( x = (x_1, x_2, ..., x_n) \) is a random sample of size \( n \) from the discrete Marshall-Olkin Weibull distribution with probability mass function defined in equation (3), then the log-likelihood function is given by

\[ \ell = \prod_{i=1}^{n} P_{x_i} \]

\[ = \sum_{i=1}^{n} \log \left\{ \frac{\alpha \left\{ \gamma_i^{\beta} - \gamma^{(x_i+1)^\beta} \right\}}{\left\{ 1 - \alpha \gamma^{\beta} \right\} \left\{ 1 - \alpha \gamma^{(x_i+1)^\beta} \right\}} \right\}, \]

\[ = n \log \alpha + \sum_{i=1}^{n} \log \left( \gamma_i^{\beta} - \gamma^{(x_i+1)^\beta} \right) - \sum_{i=1}^{n} \log \left( 1 - \alpha \gamma^{\beta} \right) - \sum_{i=1}^{n} \log \left( 1 - \alpha \gamma^{(x_i+1)^\beta} \right). \quad (16) \]
The maximum likelihood estimates say \( \hat{\phi} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \), can be obtained by solving the system of non-linear equation \( \frac{\partial \ell}{\partial \phi} = 0 \). Hence, taking the first derivative of equation (16) with respect to the parameter vector \( \phi \), we have

\[
\frac{\partial \ell}{\partial \alpha} = n - \sum_{i=1}^{n} \gamma^{x_i+1} - \sum_{i=1}^{n} \gamma^{x_i+1} + \alpha \gamma^{x_i+1},
\]

\[
\frac{\partial \ell}{\partial \beta} = \log \gamma \sum_{i=1}^{n} x_i \gamma^{x_i} \log x_i - (x_i + 1) \gamma^{x_i} \log(x_i + 1) + \alpha \log \gamma \sum_{i=1}^{n} x_i \gamma^{x_i} \log x_i \gamma^{x_i} + \alpha \gamma^{x_i} \log(x_i + 1) + \alpha \gamma^{x_i} \gamma^{x_i+1}.
\]

\[
\frac{\partial \ell}{\partial \gamma} = \sum_{i=1}^{n} x_i \gamma^{x_i+1} - (x_i + 1) \gamma^{x_i+1} + \alpha \sum_{i=1}^{n} x_i \gamma^{x_i} + \alpha \gamma^{x_i+1} + \alpha \gamma^{x_i+1}.
\]

Since there is no close form expression for the MLE of the discrete Marshall-Olkin Weibull distribution, it becomes difficult to obtain analytical solutions for the unknown parameter estimates, thus a standard numeric optimization algorithm such as the Newton-Raphson Iterative Scheme is employed to optimize the log-likelihood function. The \texttt{fitdistplus} package in R statistical software is used to evaluate the maximum likelihood estimates of the DMOW distribution.

4. Applications

In this section, we attempt to illustrate the flexibility of the proposed DMOW distribution using two real data sets. The first data set records the number of strikes in UK coal mining industries in four successive week periods during 1948-1959 reported in [7]. The data set is under-dispersed with \( \mu = 0.99 \) and \( \sigma^2 = 0.74 \). The second data set is a biological data set which records the number of Hemocytometer yeast cell counts per square reported in [1]. The data set is over-dispersed with \( \mu = 0.68 \) and \( \sigma^2 = 0.81 \).

The goodness of fit of the DMOW distribution is compared with the following existing discrete distributions:
(1) Discrete Weibull (DW) distribution ([13]) with pmf given by
\[ P_x = \gamma^x \beta - \gamma^{(x+1)\beta}, \]

(2) Exponentiated Discrete Weibull (EDW) distribution ([14]) with pmf given by
\[ P_x = \left(1 - \gamma^{(x+1)\beta}\right)^\alpha - \left(1 - \gamma^\beta\right)^\alpha, \]

(3) Generalized Geometric (GG) distribution ([4]) with pmf given by
\[ P_x = \frac{\alpha \gamma^x (1 - \gamma)}{(1 - \alpha \gamma^{x+1})(1 - \alpha \gamma^x)}, \]

(4) Discrete Generalized Exponential (DGE) distribution ([15]) with pmf given by
\[ P_x = \left(1 - \gamma^{(x+1)\beta}\right)^\alpha - \left(1 - \gamma^\beta\right)^\alpha, \]

(5) Discrete Marshall-Olkin Generalized Exponential (DMOGE) distribution ([2]) with pmf given by
\[ P_x = \frac{\beta \left(1 - \gamma^x\right)^\alpha}{\beta + \beta \left(1 - \gamma^x\right)^\alpha} - \frac{\beta \left(1 - \gamma^{(x+1)\beta}\right)^\alpha}{\beta + \beta \left(1 - \gamma^{(x+1)\beta}\right)^\alpha}, \]

(6) Discrete Bur XII (DBXII) distribution ([11]) with pmf given by
\[ P_x = \gamma^\log(1+x^\alpha) - \gamma^\log[(1+(x+1)^\alpha)]. \]

The parameter estimates, log-likelihood (\(\ell\)) and the expected frequencies were obtained using the estimated probabilities considering the parameter estimates for each distribution. The bolded expected frequencies refer to those closer to the corresponding observed frequencies. The chi-square (\(\chi^2\)) goodness-of-fit statistics with the corresponding p-value is employed to evaluate the fit of each distribution to determine which distribution fit better for the two data sets under study.
### Table 2. Goodness-of-fit of each distribution for data set 1.

<table>
<thead>
<tr>
<th>Number of Strikes</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DMOW</td>
<td>DMOGE</td>
</tr>
<tr>
<td>0</td>
<td>46</td>
<td>46.14</td>
</tr>
<tr>
<td>1</td>
<td>76</td>
<td>75.17</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>26.40</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6.93</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Parameter Estimates
- $\alpha = 0.1466$
- $\beta = 2.7120$
- $\gamma = 0.9420$
- $\alpha = 4.6120$
- $\beta = 1.1353$
- $\gamma = 0.2177$
- $\alpha = 2.9170$
- $\beta = 1.1996$
- $\gamma = 0.3406$
- $\alpha = 4.7947$
- $\beta = 0.81$
- $\gamma = 0.84$
- $\alpha = 12.0575$
- $\gamma = 0.5941$
- $\alpha = 12.0575$
- $\gamma = 0.5941$
- $\alpha = 4.6533$
- $\gamma = 0.5941$

| $\chi^2 (df)$   | 0.8997 (1)          | 1.4140 (1)         | 1.5032 (1) | 3.2523 (2) | 1.3553 (2) | 3.0078 (2) | 5.534 (2)  |
| $p$-value       | 0.3430             | 0.2340             | 0.2200     | 0.1970    | 0.2440    | 0.2220    | 0.0620     |

### Table 3. Goodness-of-fit of each distribution for data set 2.

<table>
<thead>
<tr>
<th>Number of Hemo. yeast</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DMOW</td>
<td>DMOGE</td>
</tr>
<tr>
<td>0</td>
<td>213</td>
<td>212.89</td>
</tr>
<tr>
<td>1</td>
<td>128</td>
<td>127.27</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>41.23</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>13.49</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3.97</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Parameter Estimates
- $\alpha = 0.1945$
- $\beta = 1.8972$
- $\gamma = 0.8188$
- $\alpha = 1.8447$
- $\beta = 0.9650$
- $\gamma = 0.2957$
- $\alpha = 0.9435$
- $\beta = 1.3059$
- $\gamma = 0.4845$
- $\alpha = 1.2719$
- $\beta = 0.4651$
- $\gamma = 0.2938$
- $\alpha = 1.8096$
- $\beta = 0.2938$
- $\gamma = 0.2685$
- $\alpha = 2.3695$
- $\beta = 0.2938$
- $\gamma = 0.3324$
- $\alpha = 2.4174$

| $\ell$          | -445.8829          | -446.4779          | -446.4810 | -446.4518 | -446.4783 | -446.6230 | -451.3099 |
| $\chi^2 (df)$   | 2.1868 (2)          | 2.9101 (2)         | 3.3081 (2) | 3.2187 (2) | 2.9347 (2) | 3.3820 (2) | 7.5618 (2) |
| $p$-value       | 0.3351             | 0.2334             | 0.1913    | 0.2000    | 0.2305    | 0.1843    | 0.0228    |

http://www.earthlinepublishers.com
Tables 2 and 3 present the parameter estimates, log-likelihood, observed and expected frequencies, and $\chi^2$ statistics with the corresponding $p$-value of each distribution for the two data sets. The better distribution corresponds to the one having the maximized log-likelihood and $p$-value and the least $\chi^2$ statistics value. Thus, the results obtained from these Tables suggest that the DMOW distribution performs reasonably better than the compared discrete distributions for the two data sets.

5. Conclusion

A discrete analogue of the continuous Marshall-Olkin Weibull distribution has been introduced using the discrete concentration approach. Some mathematical properties of the proposed DMOW distribution were derived and it observed that the probability mass function of the distribution accommodates a decreasing, left-skewed, right-skewed unimodal and symmetric shapes while the hazard rate function exhibits a decreasing, an increasing, bathtub and inverted bathtub shapes properties. The method of maximum likelihood estimation was employed to estimate the unknown parameters of the proposed distribution and finally, two count data sets (over-dispersed and under-dispersed) was used to examine the applicability of the DMOW distribution.

References


This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted, use, distribution and reproduction in any medium, or format for any purpose, even commercially provided the work is properly cited.