Discrete Weibull-geometric distribution

Y O Mabel, M Novita and S Nurrohmah
Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
Corresponding author’s email: mila.novita@sci.ui.ac.id

Abstract. The failure-time distribution is used to describe the life of a device, material, or structure. One of the most commonly used distributions to analyze failure-time data is Weibull distribution. The Weibull distribution is very flexible for modeling varying types of failure rates data with constant, increasing and decreasing shapes, but cannot be used for modeling data with unimodal failure rates. Because of this, we developed the Weibull distribution using compounding method to produce a Weibull-Geometric distribution. The Weibull-Geometric distribution is useful for modeling failure rates data with a unimodal shape. In studies of failure, time to failure is frequently measured in the number of cycles to failure or shocks to failure and become a discrete random variable. Hence, this paper discusses the formation of distribution that more appropriate to modeling discrete failure data. The discrete distribution is obtained by discretizing the continuous Weibull-Geometric distribution. The discretization is carried out by maintaining one of characteristics of the continuous distribution, that is, its survival function. The result distribution, called Discrete Weibull-Geometric distribution (DWG) has unimodal failure rates, right skew, and more appropriate at modeling discrete failure data. At the end of this paper, the Discrete Weibull-Geometric distribution is used to illustrate the real dataset and show that DWG distribution is the appropriate model.

Keywords: Discrete data, discrete Weibull distribution, Weibull distribution, Weibull-geometric distribution

1. Introduction
The lifespan of devices, materials, organisms, etc., is important in reliability studies such as biological and engineering sciences. The important part is a mathematical description of the length of life with the failure distribution. The failure-time distribution is used to describe life of a device, material, or structure. There was a distribution that used to analyze failure time data, that is exponential, gamma, and Weibull distribution [1]. But the most commonly used distribution is the Weibull distribution, because it is more flexible than other two distribution. The Weibull distribution has constant, increasing and decreasing failure rates, so that the Weibull distribution can model various types of data.

In failure time data, there was another shape of data, that is, unimodal failure rates shape. The Weibull distribution is not useful for modeling data with indication of unimodal failure rates. Because of this, in 2011, Barreto-Souza et al. [2] developed Weibull distribution using compounding method to produce a Weibull-Geometric distribution, and shows that the new distribution is different with the Weibull distribution. The new distribution is useful for modeling failure rate data with a unimodal shape.

In studies of failure, time to failure is often measured in number until failure occurs than the length of life to failure. For example, the number of cycles until failure, the number of shocks until failure,
rounds fired until failure, etc. Then the observation is the number of time periods that successfully
completed before failure, and the number of lifetimes is counted as a discrete random variable.
Therefore, a continuous distribution is not appropriate for modeling this type of data. In 1975, Nakagawa
and Osaki modeled a discrete failure data using Discrete Weibull Distribution. But, the Discrete Weibull
distribution cannot model data with indicates unimodal failure rate. Hence, this paper discusses the
formation of distribution that more appropriate to modeling discrete failure data in varying failure rate
shape. In this paper, a discrete distribution is obtained by discretizing the continuous of the Weibull-
Geometric distribution. The resulting distribution, called Discrete Weibull-Geometric distribution
(DWG) has unimodal failure rates, right skew, and is more appropriate at modeling discrete failure data
than the Discrete Weibull distribution.

2. Methodology

2.1. Formation of Weibull-geometric distribution

Suppose the random variables \( Z_i \) are independent and have identical distribution (iid) with pdf:

\[
f_{Z_i}(x; \theta) = f_Z(x; \theta); \quad \theta = (\theta_1, \theta_2, ..., \theta_k); \quad k \geq 1; x > 0; \theta > 0
\]

and \( N \) is a discrete random variable with geometric distribution having pmf:

\[
p_n = (1 - p)p^{n-1}; \quad 0 < p < 1; n = 1, 2, 3, ...
\]

Let \( X = \min\{Z_i\}_{i=1}^N \), then the pdf and cdf of random variable \( X \) is obtained as

\[
f_X(x) = \frac{(1 - p)f_Z(x; \theta)}{[1 - p(1 - F_Z(x; \theta))]^2}
\]

and

\[
F_X(x) = \frac{F_Z(x; \theta)}{1 - p(1 - F_Z(x; \theta))}
\]

Equation 1 and equations 2 are pdf and cdf of compound geometric class with \( X \) as a random variable
that follows continuous distribution [3].

Therefore, using compound geometric class, let \( X \) be a random variable that follows Weibull
distribution with pdf

\[
g(x) = \alpha \beta x^{\alpha-1}e^{-(\beta x)^\alpha}; \quad x > 0; \alpha > 0; \beta > 0
\]

Then, the pdf and cdf of the Weibull-Geometric (WG) distribution can be obtained, respectively, as follow:

\[
f(x) = \frac{\alpha \beta x^{\alpha-1}(1 - p)e^{-(\beta x)^\alpha}}{[1 - p e^{-(\beta x)^\alpha}]^2}; \quad x > 0
\]

and

\[
F(x) = \frac{1 - e^{-(\beta x)^\alpha}}{1 - p e^{-(\beta x)^\alpha}}; \quad x > 0
\]

for the parameters \( \alpha \) and \( \beta \) is positive, and \( 0 < p < 1 \). \( \alpha \), \( \beta \) and \( p \) are the shape, scale and mixing
parameters, respectively.
The survival and hazard functions of the WG distribution are given:

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

(3)

and

\[
h(x) = \frac{\alpha \beta^\alpha x^{\alpha - 1}}{1 - p e^{-(\beta x)^{\alpha}}}; \quad x > 0
\]

Figure 1 illustrates the possible shapes of the hazard function for various parameter values. The plots shown that the WG distribution is quite flexible for the hazard function. It shows that hazard functions can take decreasing, increasing, and unimodal shapes. Decreasing shape occurs for \(0 < \alpha < 1\) and increasing shape when \(\alpha > 1\). The hazard function can take the unimodal shape only when \(\alpha > 1\) with large parameter values of \(p\) that approach 1.

The distribution has an interpretation for the situation where the failure occurs because of an unknown number of \(N\), of initial defects of same kind (a number of components of a device, for example). Their lifetimes are represented by \(Z_i\) and the defect is only detected after the causing a failure. When a component fails, it will be repaired immediately. Then the assumption distribution models the time to the first failure \(X\).

Figure 1. Hazard function of weibull-geometric distribution.
2.2. Formation of discrete Weibull-geometric distribution

The pmf of discrete Weibull-Geometric (DWG) distribution is obtained by maintaining one of the characteristics of Weibull-Geometric (WG) distribution, i.e. survival function based on discretization methods by Kemp in Chakraborty [4], as defined by

\[
p_y = S_X(y) - S_X(y + 1)
\]

where \( Y = [X] \), \([X]\) is the largest integer less than or equal to \( X \) and \( S_X(\cdot) \) is the survival function of the random variable \( X \). Thus, the pmf of discrete Weibull-Geometric distribution based on equation 3 can be written as follows:

\[
p_y = \frac{(1 - p)e^{-(\beta y)^a}}{1 - pe^{-(\beta y)^a}} - \frac{(1 - p)e^{-(\beta(y+1))^a}}{1 - pe^{-(\beta(y+1))^a}}
\]

\[
= \frac{(1 - p)(e^{-(\beta y)^a} - e^{-(\beta(y+1))^a})}{[1 - pe^{-(\beta y)^a}][1 - pe^{-(\beta(y+1))^a}]}
\]

with reparameterization \( \lambda = e^{-\beta a} \), obtained:

\[
p_y = \frac{(1 - p)(\lambda^{y} - \lambda^{(y+1)a})}{[1 - p\lambda^{y}][1 - p\lambda^{(y+1)a}]}, \quad y = 0,1,2,\ldots
\]

(4)

The characteristics of the pmf is shown on the graph (figure 2). Figure 2 shown that the pmf of DWG distribution have decreasing and unimodal shapes. The decreasing shape occurs for \( \alpha \leq 1 \), and the unimodal shape occurs for \( \alpha > 1 \).

In general, based on pmf of DWG distribution in equation 4, the cdf and survival function of its distribution is written as follows:

\[
F(y) = \frac{1 - \lambda^{(y+1)a}}{1 - p\lambda^{(y+1)a}}, \quad n \leq y < n + 1; \quad n = 0,1,2,\ldots
\]

and

\[
S(y) = \frac{(1 - p)\lambda^{(y+1)a}}{1 - p\lambda^{(y+1)a}}, \quad n \leq y < n + 1; \quad n = 0,1,2,\ldots
\]

Figure 2. pmf of discrete weibull-geometric distribution.
Hazard function of discrete distribution is defined by Xekalaki [5] as follows:

\[ h(t) = \Pr(T = t | T > t - 1) = \frac{\Pr(T = t)}{\Pr(T > t - 1)} \]

Therefore, the hazard rate of DWG distribution is obtained as follows:

\[ h(y) = \frac{1 - \lambda^{(y+1)\alpha}}{1 - p\lambda^{(y+1)\alpha}}; \quad y = 0, 1, 2, \ldots \]

The characteristics of hazard rate function is shown on the graph for varying parameter values (figure 3). It shows that the hazard rate function shapes of the DWG distribution can take decreasing, increasing, and unimodal shapes. The decreasing shape will occur for \( \alpha \leq 1 \) and the increasing shape occur for \( \alpha > 1 \). The hazard rate of DWG distribution can take a unimodal shape that occurs only for \( \alpha > 1 \) with the value of parameters \( \lambda \) and \( p \) both approaching 1.

The \( r \)-th moment of DWG distribution is given as follows:

\[ E(Y^r) = \sum_{y=0}^{\infty} y^r \left( \frac{(1 - p)(\lambda^y - \lambda^{y+1})}{[1 - p\lambda^y][1 - p\lambda^{(y+1)}]} \right) \]

where the first moment is mean (\( \mu \)). Because the complexity of the equation is quite complicated, the characteristics of DWG distribution that are explained by mean, variance, skewness coefficient, and kurtosis coefficient are presented in Table 1 for the given values of \( \alpha, p \) and \( \lambda \).

Table 1 shows that there is a mean value greater than the variance value, and a mean value smaller than the variance value, so it can be said that the DWG distribution can model data that has underdispersion and overdispersion properties. In addition, from table 1 it can be seen that for various parameter values, the value of skewness coefficient is always greater than zero, which indicates that the DWG distribution is right skew.

**Figure 3.** Hazard function of discrete weibull-geometric distribution.
2.2.1. Parameter estimation of discrete Weibull-geometric distribution. Using the maximum likelihood for estimating the value of each parameter, that is $\alpha$, $p$ and $\lambda$, with the log likelihood function:

$$
\ln \ell = n \log(1 - p) + \sum_{i=1}^{n} \log\left(\lambda_i^{\alpha_i} - \lambda^{(y_i+1)\alpha_i}\right) - \sum_{i=1}^{n} \log\left(1 - p\lambda_i^{\alpha_i}\right)
$$

$$
- \sum_{i=1}^{n} \log\left(1 - p\lambda^{(y_i+1)\alpha_i}\right)
$$

(5)

then estimate the value of $\alpha$, $p$ and $\lambda$, obtained by solving equation 3 as follows:

$$
\frac{\partial \ell}{\partial \alpha} = \log(\lambda) \sum_{i=1}^{n} \lambda_i^{\alpha_i} \log(y_i) - (y_i + 1)^\alpha \lambda^{(y_i+1)\alpha} \log(y_i + 1)
$$

$$
+ p \log(\lambda) \sum_{i=1}^{n} \lambda_i^{\alpha_i} \log(y_i) - (y_i + 1)^\alpha \lambda^{(y_i+1)\alpha} \log(y_i + 1)
$$

$$
+ p \log(\lambda) \sum_{i=1}^{n} \frac{(y_i + 1)^\alpha \lambda^{(y_i+1)\alpha} \log(y_i + 1)}{1 - p\lambda^{(y_i+1)\alpha}} = 0
$$

(6)

$$
\frac{\partial \ell}{\partial p} = \frac{-n}{1 - p} + \sum_{i=1}^{n} \lambda_i^{\alpha_i} + \sum_{i=1}^{n} \frac{(y_i + 1)^\alpha \lambda^{(y_i+1)\alpha}}{1 - p\lambda^{(y_i+1)\alpha}} = 0
$$

(7)

$$
\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \frac{\lambda_i^{\alpha_i - 1} - (y_i + 1)^\alpha \lambda^{(y_i+1)\alpha - 1}}{\lambda_i^{\alpha_i} - \lambda^{(y_i+1)\alpha}} + p \sum_{i=1}^{n} \frac{\lambda_i^{\alpha_i - 1}}{1 - p\lambda^{(y_i+1)\alpha}} + p \sum_{i=1}^{n} \frac{(y_i + 1)^\alpha \lambda^{(y_i+1)\alpha - 1}}{1 - p\lambda^{(y_i+1)\alpha}} = 0
$$

(8)

Equation 6 through equation 8 can be solved by numerical methods.

Table 1. Mean, variance, skewness coefficient, and kurtosis coefficient for varying parameter values.

| $\alpha$ | $\lambda$ | $p$ | Mean | Variance | Skewness | Kurtosis |
|----------|-----------|-----|------|----------|----------|----------|
| 0.5      | 0.5       | 0.1 | 1.28 | 5.09     | 2.07     | 6.68     |
| 0.5      | 0.5       | 0.86| 3.58 | 2.74     | 10.5     |
| 0.9      | 0.22      | 0.98| 6.08 | 45.16    |
| 0.5      | 0.1       | 0.092| 0.248| 8.554    | 101.38   |
| 0.5      | 0.5       | 0.86| 3.58 | 2.74     | 10.5     |
| 0.9      | 1.138     | 5.605| 2.238| 7.089    |
| 0.5      | 0.5       | 0.86| 3.58 | 2.74     | 10.5     |
| 1.5      | 0.425     | 0.465| 1.717| 6.214    |
| 2.5      | 0.343     | 0.246| 0.906| 2.409    |
| 5        | 0.333     | 0.222| 0.707| 1.5      |
3. Results and discussion

In this section, the application of distribution in modeling failure data will be illustrated. The data used is failure time data of an air conditioning system on an airplane sourced from Linhart et al. [6]. The data is presented in table 2.

The mean value of the data is \( \mu = 59.6 \), which is smaller than the variance, \( \sigma^2 = 5167.42 \); this indicates over dispersion. The histogram from the data is shown in figure 4, where it is seen that the data has right skew.

This is also explained by the value of skewness coefficient, which is greater than zero (skewness = 1.784). Therefore, the DWG and DW distribution which has characteristics in accordance with the data will be used for fitting the data. By using MLE method, the parameter value of DWG distribution is estimated at \( \alpha = 1.1765, \beta = 0.8147, \lambda = 0.9971 \) and the parameter value of DW distribution is estimated at \( \alpha = 0.78821, \beta = 0.955822 \).

Figure 5 shown that the failure rate shape of DWG distribution for the failure time of air conditioning system in an airplane dataset is unimodal and right skew.

![Figure 4. Histogram of failure time data of air conditioning system an airplane](image)

**Table 2.** Failure time data of air conditioning system an airplane.

|     | 1   | 3   | 5   | 7   | 11  | 11  | 11  | 12  | 14  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 14  | 14  | 16  | 16  | 20  | 21  | 23  | 42  | 47  | 52  |
| 71  | 71  | 87  | 90  | 95  | 120 | 120 | 225 | 246 |     |

![Figure 5. Failure rate data of DWG model](image)
Table 3. The estimated parameter and goodness-of-fit measures values.

| Model | Estimated parameter | K-S   | AIC    | BIC    | AICc   |
|-------|----------------------|-------|--------|--------|--------|
| DW    | $p = 0.9558$         | -152.62 | 311.24 | 315.45 | 312.17 |
|       | $\alpha = 0.7882$    |        |        |        |        |
| DWG   | $p = 0.8147$         | -151.55 | 309.09 | 313.29 | 310.02 |
|       | $\lambda = 0.9971$   |        |        |        |        |
|       | $\alpha = 1.1765$    |        |        |        |        |

Then test the compatibility using the value of AIC, BIC and AICc for DW and DWG models listed in table 3. In general, the data that presents smaller values for each testing criterion is a more fitting model. Table 3 shows that the value of each testing criterion for the DWG model is smaller than the value of the DW model. Then, it can be concluded that the DWG model provides a better fit than the DW model to model the failure time of air conditioning system in an airplane dataset.

4. Conclusion
In this paper, the Discrete Weibull-Geometric distribution which can be used in modeling discrete failure data has been discussed. Discrete Weibull-Geometric distribution has several characteristics, i.e. increasing, decreasing and unimodal failure rates and right skew. The application result shows that the Discrete Weibull-Geometric distribution is more appropriate for modeling discrete data than the Discrete Weibull distribution.

References
[1] Nakagawa T and Osaki S 1975 IEEE Transactions on Reliability 24 300-1
[2] Barreto-Souza W, de Morais A L and Cordeiro G M 2011 J. Stat. Comput. Sim. 81 645-57
[3] Alkarni S H 2013 The Open Statistics and Probability Journal 5 1-5
[4] Chakraborty S 2015 Journal of Statistical Distributions and Applications 2:6 1-30
[5] Xekalaki E 1983 Commun. Stat. 12 2503-9
[6] Linhart H and Zucchini W 1986 Model Selection (New York: John Wiley & Sons)