Generalized exponential Marshall-Olkin distribution

R Mohammad, D Lestari and S Devila
Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: dian.lestari@sci.ui.ac.id

Abstract. The distribution of generalized exponential was invented by Rameshwar D. Gupta and Debasis Kundu in 2007. The distribution was the result of a generalized transformation of the exponential distribution. This paper explained the generalized exponential Marshall-Olkin distribution which is the result of the expansion of the generalized exponential distribution using the Marshall-Olkin method. The generalized exponential Marshall-Olkin distribution has a more flexible form than the previous distribution, especially in its hazard function which has various forms so that it can represent survival data better. The flexibility characteristic is due to the addition of new parameters to the generalized exponential Marshall-Olkin distribution.

We explained some characteristics of the Marshall-Olkin generalized exponential distribution such as the probability density function (PDF), cumulative distribution function (CDF), survival function, hazard function, mean, and moments. Parameter estimation was conducted using the maximum likelihood method. In the application, it was shown data with generalized exponential Marshall-Olkin (GEMO) distribution. The GEMO distribution was modelled to the waiting time data until the damage to a lamp. The data was taken from Aarset data (1987). The results of modelling the waiting time data until the damage to a lamp on the distribution of GEMO and was compared to the distribution of alpha power Weibull. A comparison of models using Akaike information criteria (AIC) and Bayesian information criteria (BIC) shows that the distribution of GEMO is more suitable in modelling the lamp damage waiting time data than the distribution of alpha power Weibull.

Keywords: Generalized transformation, hazard function, Marshall-Olkin, maximum likelihood estimation

1. Introduction
In analyzing survival data, a probability distribution is needed. Several probability distributions are used to analyze or model survival data such as exponential, Weibull and gamma. The modeling of survival data is based on the shape of the hazard function [1]. The hazard function has a variety of shapes, such as the monotonous form of up, down, constant and non-monotonous form of the bathtub and upside-down bathtub [1].

The three distributions mentioned above can have an increase or unimodal probability density function (PDF) and a monotonous hazard function. Unfortunately, none of them have a non-monotonous hazard function; for example, to analyze lifetime data, exponential distribution has been used quite effectively [2], even though it is often described as a flexible distribution. However, the hazard function of exponential distribution only has a constant shape [3]. In 2007, the generalized exponential (GE) distribution was invented by Gupta et al. [4]. It was the result of the expansion of the
exponential distribution through generalized transformation and can be used to analyze survival data. The generalized exponential distribution has a shape of rising, falling, and constant hazard rates [4], so it has not been able to model survival data which has non-monotonous hazard form of the bathtub and upside-down bathtub.

Let $X$ be a random variable continuous non-negative having an exponential distribution with parameter $\lambda > 0$. The probability density function (PDF) of exponential distribution is given by [5]

$$f_E(x; \lambda) = \lambda e^{-\lambda x}, \text{for } x \geq 0 \quad (1)$$

The cumulative distribution function (CDF) of exponential distribution is given by:

$$F_E(x; \lambda) = \int_{-\infty}^{x} f(t) dt = 1 - e^{-\lambda x}, \text{for } x \geq 0 \quad (2)$$

Survival function of exponential distribution is given by:

$$S(x; \lambda) = 1 - F(x) = e^{-\lambda x}, \text{for } x \geq 0 \quad (3)$$

The failure rate or hazard function $h_E(x) = \lambda$

The generalized exponential distribution results from the generalized transformation

$$F_{GE}(x) = [F_E(x)]^\alpha \quad (4)$$

Therefore, $X$ is a two parameter generalized exponential random variable having a CDF given by [4]

$$F_{GE}(x; \alpha, \lambda) = \left( 1 - e^{-\lambda x} \right)^\alpha; \quad x > 0 \quad (5)$$

It follows PDF of the generalized exponential given by

$$f_{GE}(x; \alpha, \lambda) = \alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}; \quad x > 0 \quad (6)$$

It follows the hazard function of the generalized exponential given by

$$h_{GE}(x; \alpha, \lambda) = \frac{\alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}}{1 - (1 - e^{-\lambda x})^\alpha}; \quad x > 0 \quad (7)$$

with $x, \alpha,$ and $\lambda > 0$. $\alpha$ is the shape parameter and $\lambda$ is the scale parameter [4].

The plot of the hazard function of generalized exponential distribution is shown in figure 1. Based on figure 1, the hazard function of generalized exponential distribution is more flexible than exponential distribution because it has three shapes, increasing, decreasing, and constant types.

![Figure 1](image-url)
The Marshall-Olkin method is a method of adding new parameters to a distribution that will be expanded, it was invented by Albert Marshall and Ingram Olkin [6]. This method was called the expanded Marshall-Olkin distribution [6] that offers several advantages in various behaviors than the basic distribution from which they are derived. One advantage is that the hazard function of the expanded distribution depends on the parameter value added $\alpha$ and can be applied to model real situations better than the basic distribution [3].

Suppose a distribution with its survival function $S(x)$. Then survival function from the Marshall-Olkin extended distribution which is symbolized $\hat{S}(x)$ is defined by [6]

$$\hat{S}(x, \theta) = \frac{\theta \hat{F}(x)}{1 - (1 - \theta)\hat{F}(x)} , -\infty < x < \infty, \theta > 0$$ (8)

If $G(x, \theta)$, $g(x, \theta)$ and $r(x, \theta)$ are the cdf, pdf and hazard rate functions corresponding to $\hat{S}$, then [3]

$$G(x, \theta) = \frac{F(x)}{1 - \theta(1 - F(x))} ; -\infty < x < \infty; \theta > 0, \bar{\theta} = 1 - \theta$$ (9)

$$g(x, \theta) = \frac{\alpha f(x)}{[1 - \bar{\theta}F(x)]^2} ; -\infty < x < \infty; \theta > 0, \bar{\theta} = 1 - \theta$$ (10)

$$r(x, \theta) = \frac{h(x)}{1 - \bar{\theta}F(x)} ; -\infty < x < \infty; \theta > 0, \bar{\theta} = 1 - \theta$$ (11)

where $h(x)$ is the hazard rate function corresponding to $f(x)$ and $f(x), F(x)$ is the PDF and CDF corresponding to $\hat{F}(x)$.

In 2015, M. Ristic and Debasis Kundu used Marshall-Olkin method to develop Generalized Exponential (GE) distribution into Generalized Exponential Marshall Olkin (GEMO) distribution which has a more flexible shape of hazard functions such as up, down, bathtub, and upside- down bathtub [2].

2. GEMO distribution

2.1. GEMO distribution

If $X$ is a random variable continuous non negative having GEMO distribution with parameters $\alpha, \lambda > 0$ and $\theta > 0$, by substituting equation 5 and equation 6 into equation 8, equation 9, equation 10 and equation 11, we have CDF of GEMO is

$$G_{GEMO}(x; \alpha, \lambda, \theta) = \frac{(1 - e^{-\lambda x})^\alpha}{(1 - e^{-\lambda x})^\alpha(1 - \theta) + \theta} ; x > 0$$ (12)

PDF of GEMO is

$$g_{GEMO}(x; \alpha, \lambda, \theta) = \frac{\alpha \lambda \theta(1 - e^{-\lambda x})^{\alpha - 1} e^{-\lambda x}}{[(1 - e^{-\lambda x})^\alpha(1 - \theta) + \theta]^2} ; x > 0$$ (13)

Hazard rate function of GEMO is

$$r_{GEMO}(x; \alpha, \lambda, \theta) = \frac{\alpha e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha - 1}}{[\theta + (1 - \theta)(1 - e^{-\lambda x})^\alpha][1 - (1 - e^{-\lambda x})^\alpha]} ; x > 0$$ (14)
Survival function of GEMO

\[ \hat{G}_{Gemo}(x; \alpha, \lambda, \theta) = 1 - \frac{(1 - e^{-\lambda x})^\alpha}{(1 - e^{-\lambda x})^\alpha (1 - \theta) + \theta}; x > 0 \quad (15) \]

The GEMO hazard function depends on the parameters of \(\alpha, \lambda,\) and \(\theta.\) \(\lambda\) is for scale parameters; therefore, the GEMO hazard function curve does not depend on \(\lambda.\) The figure 2, figure 3, figure 4 and figure 5 show the curve of the hazard function of GEMO with different parameter values. The GEMO hazard function has more flexible than the previous distribution because it has four primary curve shapes formed, increasing, decreasing, bathtub, and upside-down bathtub shapes [2] by substituting variations of \(\theta\) parameter. The bathtub shape of GEMO hazard function curve is formed if \(0 < \alpha < 1\) and \(\theta > (1 + \alpha) / (2\alpha),\) as shown in figure 2, if \(\alpha > 1\) and \(\theta > (1 + \alpha) / (2\alpha),\) then the GEMO hazard function form increasing shape as in figure 3. The upside-down bathtub shape as in figure 4 is formed if \(\alpha > 1\) and \(0 < \theta < (1 + \alpha) / (2\alpha),\) and for the decreasing shape, if \(0 < \alpha < 1\) and \(0 < \theta < (1 + \alpha) / (2\alpha),\) the shape is formed as in figure 5.

2.2. Moment

The \(n\)th moments of GEMO will be explained in this section. If \(X\) is a random variable of the Generalized Exponential Marshall Olkin distribution with parameters \(\alpha, \lambda,\) and \(\theta > 0,\) accordingly, the \(n\)th moment of GEMO is

\[ E[X^n] = \int_0^\infty x^n f(x) \, dx \quad (16) \]

then

\[ E[X^n] = \sum_{j=0}^{\infty} \frac{1}{\theta} \left( 1 - \frac{1}{\theta} \right)^j \frac{\alpha(j+1)}{\lambda^n} \Gamma(n+1) \sum_{i=0}^{\infty} (-1)^i c(\alpha(j+1)-1,i) \frac{1}{(1+i)^{n+i+1}} \quad (17) \]

2.3. Mean

In this section we will discuss the Mean (\(\mu\)) of GEMO. If \(X\) is a random variable of generalized exponential Marshall-Olkin distribution with parameters \(\alpha, \lambda,\) and \(\theta > 0\) then the mean of GEMO for the first moment is

\[ \mu = \sum_{j=0}^{\infty} \frac{1}{\theta} \left( 1 - \frac{1}{\theta} \right)^j \frac{\alpha(j+1)}{\lambda} \sum_{i=0}^{\infty} (-1)^i c(\alpha(j+1)-1,i) \frac{1}{(1+i)^2} \quad (18) \]
2.4. Variance

In this section we will discuss the Variance of GEMO. Let $X$ be a random variable of the Generalized Exponential Marshall Olkin distribution with parameters $\alpha$, $\lambda$, and $\theta > 0$. Then the Variance of GEMO is

$$\sigma^2 = E[X^2] - \mu^2$$

(19)

$$\sigma^2 = \sum_{j=0}^{\infty} \left(1 - \frac{1}{\theta}\right)^j \frac{\alpha(j+1)}{\lambda^2} \sum_{i=0}^{\infty} (-1)^i \alpha(j + 1 - 1, i) \frac{1}{(1 + i)^2} - \left(\sum_{j=0}^{\infty} \left(1 - \frac{1}{\theta}\right)^j \frac{\alpha(j+1)}{\lambda} \sum_{i=0}^{\infty} (-1)^i \alpha(j + 1 - 1, i) \frac{1}{(1 + i)^2}\right)^2$$

(20)

3. Parameter estimation

Parameter estimation using maximum likelihood method will be explained here. If $X_1, X_2, \ldots, X_n$ are random samples of GEMO distribution, then the likelihood function is:

$$L(x_1, x_2, \ldots, x_n; \alpha, \lambda, \theta) = (\alpha \lambda \theta)^n \prod_{i=1}^{n} (1 - e^{-\lambda x_i})^{\alpha - 1} e^{-\lambda x_i} \prod_{i=1}^{n} \frac{1}{[1 - e^{-\lambda x_i}]^\alpha (1 - \theta) + \theta]^2}$$

(21)

while the logarithm in the likelihood function is given by

$$\log L(x_1, x_2, \ldots, x_n; \alpha, \lambda, \theta) = n \log ((\alpha \lambda \theta) + \sum_{i=1}^{n} x_i + (-1) \sum_{i=1}^{n} \log[1 - e^{-\lambda x_i}])$$

$$- \left(\sum_{j=0}^{\infty} \left(1 - \frac{1}{\theta}\right)^j \frac{\alpha(j+1)}{\lambda} \sum_{i=0}^{\infty} (-1)^i \alpha(j + 1 - 1, i) \frac{1}{(1 + i)^2}\right)^2$$

(22)

The first derivative will be done in log $L(\alpha, \lambda, \theta)$ of the parameter to be estimated. The first derivative is then equated with 0.

$$\frac{\partial \log L(\alpha, \lambda, \theta)}{\partial \lambda} = 0$$

(23)
The solution of equation 23, equation 24, and equation 25 can be obtained by numerical methods.

4. Application

In this chapter, data will be modeled using the generalized exponential Marshall-Olkin distribution. The data presents the waiting time for damage to a lamp and taken from Aarset [7] which is shown in table 1.

Furthermore, it will be shown that the GEMO distribution can model the waiting time data until damage to a lamp. In addition, the waiting time data until damage to a lamp is also modeled using the alpha power Weibull (APW) distribution. Descriptive statistics of data are also shown in table 2.

In figure 6, data are shown in the form of histograms, and parameter estimation is performed using the maximum likelihood method presented in table 3.

From table 3, we obtained $0 < \alpha = 0.256902 < 1$ and $\theta = 14.8209 > (1 + \alpha) / 2\alpha = 10.4$. It shows the shape of the GEMO hazard function in the form of a bathtub as shown in figure 8.

| Table 1. Data waiting time for damage to a lamp. |
|-----------------------------------------------|
| 0.1  | 0.2  | 1    | 1    | 1    | 1    | 1    | 2    | 3    | 6    |
| 7    | 11   | 12   | 18   | 18   | 18   | 18   | 18   | 21   | 32   |
| 36   | 40   | 45   | 46   | 47   | 50   | 55   | 60   | 63   | 63   |
| 67   | 67   | 67   | 67   | 72   | 75   | 79   | 82   | 82   | 83   |
| 84   | 84   | 84   | 85   | 85   | 85   | 85   | 85   | 86   | 86   |
Table 2. Descriptive statistics

| Statistic                | Value          |
|--------------------------|----------------|
| N Valid                  | 50             |
| Mean                     | 45.686         |
| Median                   | 48.5           |
| Mode                     | 1.10a          |
| Std. Deviation           | 32.8352        |
| Variance                 | 1078.153       |
| Skewness                 | -0.142         |
| Std. Error of Skewness   | -0.377         |
| Kurtosis                 | -1.627         |
| Std. Error of Kurtosis   | 0.662          |
| Minimum                  | 0.1            |
| Maximum                  | 86             |
| Sum                      | 2284.3         |

Figure 6. The histogram of Aarset lamp damage data [7].

Table 3. Maximum likelihood estimation for Aarset lamp damage data [7].

| No | Distribution | Parameter 1 | Parameter 2 | Parameter 3 | Parameter 4 |
|----|--------------|-------------|-------------|-------------|-------------|
| 1  | GEMO         | 0.256902    | -           | 0.03596     | 14.8209     |
| 2  | APW          | 4.505796    | 0.836265    | 0.058405    | -           |

4.1. Goodness of fit distribution models
The goodness of fit distribution models is shown in Table 4. Based on the Kolmogorov-Smirnov compatibility test with a significance level of $\alpha = 0.05$ with a value of $W_{1-\alpha} = 0.188$, it is shown that all values of statistics $T$ are smaller than 0.188, so all models are convenient in defining the data.

4.2. Fitted models distribution
In this section, it will be seen which of the two distributions is more appropriate for modeling the data. Akaike information criteria (AIC) are determined by $-2 \log L + 2q$ while Bayesian information criteria
(BIC) are defined by $-2\log L + q \log n$, with $n$ as the number of data sizes and $q$ as the number of parameters. The best model will be determined based on the smallest value of AIC, and BIC. Next, AIC and BIC values from the GEMO and APW distributions are presented in table 5, showing that the GEMO distribution is more suitable for modeling the data.

In figure 7, cdf of GEMO is more suitable for describing the empirical distribution function of Aarset lamp damage data (1987) and figure 8 shows that the hazard function of GEMO based on Aarset lamp damage data (1987) has a bathtub shape, where the hazard graph is relatively high initially, then decreases, before increasing again over time [7].

5. Results and discussion

The Marshall-Olkin method is applied to GE which results in a new distribution called GEMO. It has the advantage of modeling data that has a non-monotonous and monotonous hazard shape compared to GE. APW distribution can actually model Aarset lamp damage data (1987) but AIC and BIC show that GEMO is better at modeling data than APW [7].

A hazard function of GEMO is generated from Aarset lamp damage data (1987) to form of a bathtub [7]. It is divided into 3 phases. The first is called burn-in phase, where the hazard function decreases because of initial damage to the lamp, for example, through a design error in the lamp. The second phase is called the useful life phase when the hazard function is constant, the lamp works properly and there is no damage at all. The third phase is called the wear-out phase when the hazard function increases, probably due to aging lamp components.

Table 4. Summary of fitted distributions

| Distribution | T   |
|--------------|-----|
| GEMO         | 0.151037 |
| APW          | 0.17499  |

Table 5. AIC and BIC Values

| Model  | AIC   | BIC  |
|--------|-------|------|
| GEMO   | 474.174 | 479.91 |
| APW    | 526.146 | 531.882 |

Figure 7. Plot of the empirical and fitted cumulative distribution functions based on Aarset lamp damage data (1987)

Figure 8. Plot of hazard function GEMO based on Aarset lamp damage data (1987) in bathtub shape
The hazard function of GEMO based on Aarset data (1987) [7], which has a bathtub shape, is very useful for analysis survival related to decision making and cost analysis. Placement of the time points when a process starts to stabilize or when the hazard function starts to increase is important in its application, for example, such as determining the lamp’s warranty time.

6. Conclusion
In this paper the Marshall-Olkin method is applied to expand the generalized exponential distribution called the Marshall-Olkin generalized exponential distribution. Some properties of our proposed GEMO ($\alpha$, $\lambda$, $\theta$) distribution were derived. The addition of variation $\theta$, formed four different major shapes of the GEMO hazard function’s curve: increasing, decreasing, bathtub, and upside down bathtub shapes. Therefore, GEMO is more flexible and solved the disadvantages of the other distributions in providing the bathtub or upside-down bathtub forms of the hazard function. Hazard function of GEMO will have an increasing shape if $\alpha > 1$ and $\theta > (1 + \alpha) / (2\alpha)$, a decreasing shape if $0 < \alpha < 1$ and $0 < \theta < (1 + \alpha) / (2\alpha)$, a bathtub shape if $0 < \alpha < 1$ and $\theta > (1 + \alpha) / (2\alpha)$, and an upside-down bathtub shape if $\alpha > 1$ and $0 < \theta < (1 + \alpha) / (2\alpha)$. The maximum likelihood method was utilized to obtain the estimation of parameters. The results of simulation on the lamp damage waiting time data show that the generalized exponential Marshall-Olkin distribution is more suitable for modeling the lamp damage waiting time data compared to the alpha power Weibull distribution.

References
[1] Lawless J F 2003 Statistical Models and Methods for Lifetime Data 2nd edition (New Jersey: John Wiley and sons)
[2] Ristić M M and Kundu D 2015 Metron 73 317-33
[3] K K Jose 2011 Int. Statistical Inst.: Proc. 58th World Statistical Congress Dublin pp 3918-23
[4] Gupta R D and Kundu D 2007 J. Stat. Plan. Infer. 137 3537-47
[5] Gupta A, Zeng W B and Wu Y 2010 Probability and Statistical Models Foundations for Problems in Reliability and Financial Mathematics. (Swiss: Birkhäuser publisher) pp 23-43
[6] Bdair O M 2012 J. Appl. Stat. 19 13-29
[7] Aarset M V 1987 IEEE T. Reliab. 36 106-08