Advances and Applications in Statistics
© 2016 Pushpa Publishing House, Allahabad, India
Published Online: April 2016
http://dx.doi.org/10.17654/AS048020157
Volume 48, Number 2, 2016, Pages 157-167 **ISSN: 0972-3617**

ADDING A NEW PARAMETER IN FLEXIBLE WEIBULL DISTRIBUTION USING MARSHALL-OLKIN MODEL AND ITS APPLICATION

W. M. Afify

Department of Statistics, Mathematics and Insurance
Faculty of Commerce
Kafrelsheikh University
Egypt
e-mail: waleedafify@yahoo.com

Abstract

In this paper, we introduce new family distribution. Using the odds ratio model approach suggested by Marshall-Olkin [16], a new model called Marshall-Olkin flexible Weibull distribution is proposed. We discuss estimation of the model parameters under type I, type II and complete samples. An application to a real data set is presented for illustrative purposes.

1. Introduction

Statistical distributions are very useful in describing and predicting real world phenomena. In many practical situations, we find that several of usual probability distribution functions found in application or do not provide an adequate fit.

Received: December 7, 2015; Revised: January 18, 2016; Accepted: February 1, 2016
2010 Mathematics Subject Classification: 62-XX.
Keywords and phrases: Marshall-Olkin model, flexible Weibull distribution, likelihood function, censored data, Monte Carlo and bootstrap simulations.
Communicated by K. K. Azad

Adding parameters to a well-established distribution is an effective way to enlarge the behavior range of this distribution and to obtain more flexible family of distributions to model various types of data. This has motivated researchers seeking and developing new and more flexible distributions. Model of proportional odds ratio plays an important role in the survival and reliability analysis. It is defined as the ratio of the distribution function to survival function. Marshall and Olkin [17] introduced an interesting method of adding a new parameter to an existing distribution. The resulting distribution is known as Marshall-Olkin extended distribution. The Marshall-Olkin family of distributions is also known as the proportional odds family "proportional odds model" or family with tilt parameter.

In this paper, adding a new parameter in flexible Weibull distribution to introduce a new family of distributions is discussed. In particular, starting with a survival function $\overline{F}(x)$, the one parameter family of survival functions

$$\overline{G}(x;\, \alpha) = \frac{\alpha\overline{F}(x)}{1 - \overline{\alpha}\overline{F}(x)} = \frac{\alpha\overline{F}(x)}{F(x) - \alpha\overline{F}(x)}; \quad 0 < X < \infty, \quad \alpha > 0, \quad (1.1)$$

where $\overline{\alpha} = 1 - \alpha.$ Then $\{G(x) > 0\}$ is said to be a *proportional odds family* with underling distribution $\overline{F}(x).$ Note that $\overline{G}(x;\, 1) = \overline{F}(x).$

The probability density function of the Marshall-Olkin given by (1.1) has easily-computed densities. In particular, if $F(x)$ has a density and hazard rate $r_F(x),$ then $G(x)$ has the density $g(x)$ given by

$$g(x;\, \alpha) = \frac{\alpha f(x)}{\{1 - \overline{\alpha}\overline{F}(x)\}^2} \quad (1.2)$$

and hazard rate

$$r(x;\, \alpha) = \frac{r_F(x)}{1 - \overline{\alpha}\overline{F}(x)}, \quad -\infty < X < \infty. \quad (1.3)$$

The odds ratio plays an important role in the survival and reliability analysis. It is mainly used to compare two groups. It is defined as the ratio of the distribution function to survival function. In 1997, Marshall-Olkin proposed a new method for adding a parameter to a family of distributions. Some special cases discussed in the literature include the Marshall-Olkin extensions of the Pareto distribution (Ghitany [6]), Weibull distribution (Ghitany et al. [7] and Zhang and Xie [22]), Lomax distribution (Ghitany et al. [8]), gamma distribution (Ristić et al. [20]) and linear failure rate distribution (Ghitany and Kotz [9]). Also, Gupta et al. [11] compared this family and the original distribution with respect to some stochastic orderings and also investigated thoroughly the monotonicity of the failure rate of the resulting distribution when the baseline distribution is taken as Weibull. Gómez-Déniz [10] presented a new generalization of the geometric distribution using the Marshall-Olkin scheme. Economou and Caroni [5] showed that Marshall-Olkin extended distributions have a proportional odds property and Caroni [2] presented some Monte Carlo simulations considering hypothesis testing on the parameter α for the extended Weibull distribution. Maximum likelihood estimation in Marshall-Olkin family is given in Lam and Leung [14] and Gupta and Peng [12]. Nanda and Das [19] investigated the tilt parameter of the Marshall-Olkin extended family.

The article is organized as follows. In Section 2, we demonstrate the maximum likelihood estimates and the asymptotic variance-covariance matrix of the unknown parameters from the different sample schemes. Finally, some lifetime data sets are used to illustrate that the Marshall-Olkin flexible Weibull distribution can be used for the data under analysis by using simulation results which are performed in Section 3.

2. Censored Data

A random variable X is said to have a *flexible Weibull distribution* with parameters $\lambda, \beta > 0$ if its probability density function, cumulative function and survival function are given by

$$f(x; \lambda, \beta) = \left(\lambda + \frac{\beta}{x^2}\right) e^{\lambda x - \frac{\beta}{x}} e^{-e^{\lambda x - \frac{\beta}{x}}}, \tag{2.1}$$

$$F(x; \lambda, \beta) = 1 - e^{-e^{\lambda x - \frac{\beta}{x}}}, \tag{2.2}$$

$$\overline{F}(x; \lambda, \beta) = e^{-e^{\lambda x - \frac{\beta}{x}}}, \tag{2.3}$$

respectively.

Now using (1.1) and (2.3), the survival functions of Marshall-Olkin flexible Weibull distribution with parameters $\lambda, \beta > 0$ and $0 < \alpha < \infty$ take the form as following:

$$\overline{G}(x; \alpha) = \frac{\alpha e^{-e^{\lambda x - \frac{\beta}{x}}}}{1 - \overline{\alpha} e^{-e^{\lambda x - \frac{\beta}{x}}}}. \tag{2.4}$$

Also, the probability density function with parameters λ, β and α becomes

$$g(x; \alpha, \lambda, \beta) = \frac{\alpha\left(\lambda + \frac{\beta}{x^2}\right) e^{\lambda x - \frac{\beta}{x}} e^{-e^{\lambda x - \frac{\beta}{x}}}}{[1 - \overline{\alpha} e^{-e^{\lambda x - \frac{\beta}{x}}}]^2}. \tag{2.5}$$

In a typical life test, n specimens are placed under observation and as each failure occurs the time is noted. Finally, at some pre-determined fixed time T or after pre-determined fixed number of sample specimens fail r, the test is terminated. In both of these cases, the data collected consist of observations $x_{(1)}, x_{(2)}, \ldots, x_{(r)}$ plus the information that $(n - r)$ specimens survived beyond the time of termination, T in the former case and $x_{(r)}$ in the latter. When T is fixed and r is thus a random variable, censoring is said to be of *single censored type* I; also, when r is fixed and the time of termination T is a random variable, censoring is said to be of *single censored type* II. In both type I and type II censoring, Cohen [4] gave the likelihood function as

$$L \propto \prod_{i=1}^{r} g(x_{(i)})[\overline{G}(x_0)]^{n-r}, \tag{2.6}$$

where $g(x_{(i)})$ and $\overline{G}(x_0)$ are the density and survival functions, respectively, and in type I, the time of termination at $x_0 = T$ and in the type II at $x_0 = x_{(r)}$. If $r = n$, then equation (2.6) reduces to complete samples.

In general, likelihood function in different sample schemes will be obtained as follows:

$$L \propto \prod_{i=1}^{r} \frac{\alpha f(x_{(i)})}{\{1 - \overline{\alpha}\overline{F}(x_{(i)})\}^2} [\overline{G}(x_0)]^{n-r}.$$

Taking logarithm likelihood function with survival function $\overline{G}(x_0)$ and probability density function $g(x_{(i)})$ based on equation (2.6) is given by

$$\log L = C + r \log \alpha + \sum_{i=1}^{r} \log f(x_{(i)}; \lambda, \beta) - 2\sum_{i=1}^{r} \log\{1 - \overline{\alpha}\overline{F}(x_{(i)}; \lambda, \beta)\}$$

$$+ (n - r)\log[\overline{G}(x_0; \alpha, \lambda, \beta)].$$

Taking the partial derivative of the log-likelihood function with respect to $\Theta = (\alpha, \lambda, \beta)$, where Θ is an unknown parameter which corresponds to new distribution where (λ, β) corresponds to the parameter of the baseline distribution. The log-likelihood function for Θ is

$$\frac{\partial \log L}{\partial \alpha} = \frac{r}{\alpha} - 2\sum_{i=1}^{r} \frac{\overline{F}(x_{(i)}; \lambda, \beta)}{1 - \overline{\alpha}\overline{F}(x_{(i)}; \lambda, \beta)} + (n - r)\frac{\frac{\partial}{\partial \alpha}\overline{G}(x_0; \Theta)}{\overline{G}(x_0; \Theta)}. \tag{2.7}$$

After the previous equation is equal to zero $\hat{\alpha}$ is

$$\hat{\alpha} = \frac{r}{2\sum_{i=1}^{r} \frac{\overline{F}(x_{(i)}; \lambda, \beta)}{1 - \overline{\alpha}\overline{F}(x_{(i)}; \lambda, \beta)} - (n - r)\frac{\frac{\partial}{\partial \alpha}\overline{G}(x_0; \Theta)}{\overline{G}(x_0; \Theta)}}, \tag{2.8}$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{r} \frac{\frac{\partial}{\partial \lambda} f(x_{(i)}; \lambda, \beta)}{f(x_{(i)}; \lambda, \beta)} + \sum_{i=1}^{r} \frac{2\bar{\alpha}}{1 - \bar{\alpha}\bar{F}(x_{(i)}; \lambda, \beta)} \frac{\partial}{\partial \lambda} \bar{F}(x_{(i)}; \lambda, \beta)$$

$$+ (n - r) \frac{\frac{\partial}{\partial \lambda} \bar{G}(x_0; \Theta)}{\bar{G}(x_0; \Theta)}, \tag{2.9}$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{r} \frac{\frac{\partial}{\partial \beta} f(x_{(i)}; \lambda, \beta)}{f(x_{(i)}; \lambda, \beta)} + \sum_{i=1}^{r} \frac{2\bar{\alpha}}{1 - \bar{\alpha}\bar{F}(x_{(i)}; \lambda, \beta)} \frac{\partial}{\partial \lambda} \bar{F}(x_{(i)}; \lambda, \beta)$$

$$+ (n - r) \frac{\frac{\partial}{\partial \beta} \bar{G}(x_0; \Theta)}{\bar{G}(x_0; \Theta)}. \tag{2.10}$$

Setting above equations to zero and solving them simultaneously yields the maximum likelihood estimate $\hat{\Theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\beta})$ of $\Theta = (\alpha, \lambda, \beta)$. These equations cannot be solved analytically and statistical software can be used to solve them numerically.

The asymptotic variance-covariance matrix of the estimators of the parameters is obtained by inverting the Fisher information matrix in which elements are negative of expected values of the second partial derivatives of the logarithm of the likelihood function.

Therefore, the approximate sample information matrix will be

$$I(\Theta) = \left[\frac{\partial^2 \ln L}{\partial \Theta^2} \right]_{\hat{\Theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\beta})} \tag{2.11}$$

(see Cohen [3]). For large n $(n \geq 50)$, matrix (2.11) is a reasonable approximation to the inverse of the Fisher information matrix. Not that closed form expressions of the expected values of these second order partial derivatives are not readily available. These terms can be evaluated by using numerical methods. Furthermore, define $V = \lim_{n \to \infty} nI_1^{-1}(\hat{\alpha}, \hat{\lambda}, \hat{\beta})$. The joint asymptotic distribution of the maximum likelihood estimators of α and θ is multivariate normal (see Lawless [15]).

3. Data Analysis

In this section, we provide an application of Marshall-Olkin flexible Weibull distribution and their submodels. We provide a data analysis in order to assess the goodness-of-fit of a model which works in practice. The data have been obtained from the time between failures of secondary reactor pumps, Suprawhardana et al. [21]. The data are as follows.

Table 1. Time between failures (thousands of hours) of secondary reactor pumps

2.160	0.746	0.402	0.954	0.791	6.560	4.992	0.347
0.150	0.358	0.101	1.359	3.465	1.060	0.614	1.921
4.082	0.199	0.605	0.273	0.070	0.062	5.320	

The analysis of least square estimates for the unknown parameters in these distributions namely: Marshall-Olkin flexible Weibull distribution (M-OFWD), flexible Weibull distribution (FWD), Marshall-Olkin flexible exponential distribution (M-OFED) and exponential distribution (ED) by using the methods of least squares is defined. The least square estimators of the unknown parameters for Marshall-Olkin flexible Weibull distribution (M-OFWD), coefficient of determination (R^2) and the corresponding mean square error (MSE) are given in Table 2.

Table 2 shows that the Marshall-Olkin flexible Weibull distribution has the smallest MSE among these models which indicates that the Marshall-Olkin flexible Weibull distribution provides the best fit for the given data among all these models. Another check is to compare the respective coefficients of determination for these regression lines. We have supporting evidence that the coefficient of determination of M-OFWD is 0.982, which is higher than the coefficient of determination of M-OFED, FWD and ED. Hence, the data point from the M-OFWD has better relationship and hence this distribution is good model for lifetime data.

Table 2. Estimated parameters of different models

Distribution	Least square estimators				MSE	R^2
	α	β	λ	θ		
M-OFWD	0.971	3.369	0.451	---	0.000889	0.982
FWD	---	3.393	0.413	---	0.000956	0.938
M-OFED	1.098	---	---	0.0094	0.000937	0.974
ED	---	---	---	0.718	0.006603	0.925

Now, we apply the formal goodness-of-fit test in order to verify which distribution fits better to these data. We consider the Cramer-von Mises (W^*) and Anderson-Darling (A^*) statistics. In general, in Table 3, the smaller the values of the statistics W^* and A^*, the better of fit to the data.

Table 3. Goodness-of-fit test for different models

Distribution	Statistics	
	W^*	A^*
M-OFWD	0.087	0.042
FWD	0.091	0.331
M-OFED	0.089	0.047
ED	0.097	1.001

Also, estimate the parameters by using the bootstrap sampling method in Table 4 as follows:

Table 4. Point estimation based on bootstrap simulations

Complete				Type I censored				Type II censored			
n	α	β	λ	r	α	β	λ	T	α	β	λ
	0.827	3.002	0.561	2	0.809	2.750	0.503	1.958	0.794	2.750	0.378
5	0.827	3.003	0.562	3	0.809	2.780	0.503	1.967	0.794	2.780	0.409
	0.829	3.005	0.572	4	0.81	2.835	0.503	1.971	0.794	2.835	0.421
	0.830	3.006	0.574	5	0.811	2.894	0.503	1.998	0.795	2.894	0.443
10	0.831	3.007	0.578	6	0.811	2.901	0.503	2.005	0.795	2.901	0.443
	0.833	3.007	0.579	7	0.811	2.908	0.505	2.009	0.797	2.908	0.445
	0.835	3.009	0.579	10	0.812	2.911	0.505	2.013	0.797	2.911	0.447
30	0.838	3.011	0.58	15	0.813	2.914	0.506	2.018	0.800	2.914	0.449
	0.838	3.013	0.582	20	0.814	2.940	0.509	2.031	0.800	2.940	0.451
	0.840	3.029	0.584	30	0.816	2.946	0.514	2.041	0.800	2.946	0.451
100	0.841	3.031	0.592	40	0.817	2.947	0.515	2.058	0.801	2.947	0.452
	0.845	3.033	0.593	50	0.82	2.953	0.516	2.071	0.801	2.953	0.456

Again, Table 5 shows asymptotic variance-covariance matrix of the maximum likelihood estimators and were calculated as described in Section 3 and are given as α, β and λ.

Table 5. Asymptotic variance-covariance matrix

Sampling	Asymptotic variance-covariance matrix					
	$V(\alpha)$	$V(\beta)$	$V(\lambda)$	$Cov(\alpha, \beta)$	$Cov(\alpha, \lambda)$	$Cov(\beta, \lambda)$
Complete	0.006	0.016	0.018	–0.513	–0.726	0.138
Type I censored	0.137	0.748	0.074	–0.614	–0.742	0.179
Type II censored	0.183	0.417	0.079	–0.912	–0.777	0.243

Acknowledgement

The author thanks the anonymous referees for their valuable suggestions which led to the improvement of the manuscript.

References

[1] A. N. Balakrishnan and H. N. Nagaraja, First Course in Order Statistics, Wiley-Interscience, New York, 1992.

[2] C. Caroni, Testing for the Marshall-Olkin extended form of the Weibull distribution, Statist. Papers 51 (2010), 325-336.

[3] A. C. Cohen, Progressively censored samples in life testing, Technometrics 5 (1963), 327-329.

[4] A. C. Cohen, Maximum likelihood estimation in the Weibull distribution based on complete and censored samples, Technometrics 7 (1965), 579-588.

[5] P. Economou and C. Caroni, Parametric proportional odds frailty models, Comm. Statist. Simul. Comput. 36 (2007), 1295-1307.

[6] M. E. Ghitany, Marshall-Olkin extended Pareto distribution and its application, Int. J. Appl. Math. 18 (2005), 17-32.

[7] M. E. Ghitany, E. K. Al-Hussaini and R. A. Al-Jarallah, Marshall-Olkin extended Weibull distribution and its application to censored data, J. Appl. Stat. 32 (2005), 1025-1034.

[8] M. E. Ghitany, F. A. Al-Awadhi and L. A. Alkhalfan, Marshall-Olkin extended Lomax distribution and its application to censored data, Comm. Statist. Theory Methods 36 (2007), 1855-1866.

[9] M. E. Ghitany and S. Kotz, Reliability properties of extended linear failure-rate distributions, Probab. Eng. Inform. Sci. 21 (2007), 441-450.

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

[11] R. C. Gupta, S. Lvin and C. Peng, Estimating turning points of the failure rate of the extended Weibull distribution, Comput. Statist. Data Anal. 54 (2010), 924-934.

[12] R. D. Gupta and C. Peng, Estimating reliability in proportional odds ratio models, Comput. Statist. Data Anal. 53 (2009), 1495-1510.

[13] F. J. Kenney and S. Keeping, Mathematics of Statistics, Van Nostrand, Princeton, NJ, 1962.

[14] K. F. Lam and T. L. Leung, Marginal likelihood estimation for proportional odds models with right censored data, Lifetime Data Anal. 7 (2001), 39-54.

[15] J. F. Lawless, Statistical Models and Methods for Lifetime Data, John Wiley, New York, 1982.

[16] A. W. Marshall and I. Olkin, A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika 84 (1997), 641-652.

[17] A. W. Marshall and I. Olkin, Life Distributions, Structure of Nonparametric, Semiparametric and Parametric Families, Springer, New York, 2007.

[18] A. J. Moors, A quantile alternative for kurtosis, J. Royal Statist. Soc., Series D 37 (1998), 25-32.

[19] A. K. Nanda and S. Das, Stochastic orders of the Marshall-Olkin extended distribution, Statist. Probab. Lett. 82 (2012), 295-302.

[20] M. M. Ristić, K. K. Jose and J. Ancy, A Marshall-Olkin gamma distribution and magnification process, STARS: Stress and Anxiety Research Society 11 (2007), 107-117.

[21] M. Salman Suprawhardana, Prayoto and Sangadji, Total time on test plot analysis for mechanical components of the RSG-GAS reactor, Atom Indones 25(2) (1999).

[22] T. Zhang and M. Xie, Failure data analysis with extended Weibull distribution, Comm. Statist. Simul. Comput. 36 (2007), 579-592.