Journal of Mathematics and Statistics 9 (1): 12-17, 2013
ISSN 1549-3644

doi:10.3844/jmssp.2013.12.17 Published Online 9 (1) 2013 (http://www.thescipub.com/jmss.toc)

ESTIMATION OF WEIBULL PARAMETERS USING A RANDOMIZED NEIGHBORHOOD SEARCH FOR THE SEVERITY OF FIRE ACCIDENTS

[1]Soontorn Boonta, [2]Anchalee Sattayatham and [1]Pairote Sattayatham

[1]Institute of Science, School of Mathematics,
Suranaree University of Technology, NakhonRatchasima 30000, Thailand
[2]Faculty of Liberal Arts, Mahidol University, Bangkok 10400, Thailand

Received 2012-12-19; Revised 2013-02-16; Accepted 2013-03-15

ABSTRACT

In this study, we applied Randomized Neighborhood Search (RNS) to estimate the Weibull parameters to determine the severity of fire accidents; the data were provided by the Thai Reinsurance Public Co., Ltd. We compared this technique with other frequently-used techniques: the Maximum Likelihood Estimator (MLE), the Method of Moments (MOM), the Least Squares Method (LSM) and the weighted least squares method (WLSM) and found that RNS estimates the parameters more accurately than do MLE, MOM, LSM or WLSM.

Keywords: Weibull Distribution, Parameter Estimation, Randomized Neighborhood Search

1. INTRODUCTION

The problem of estimating parameters in actuarial science is an important issue. Choosing an appropriate estimator is very important. In practice, constructive methods for parameter estimation are needed. The Maximum Likelihood Estimator (MLE), the Method of Moments (MOM), the Least Squares Method (LSM) and the Weighted Least Squares Method (WLSM) are frequently used for parameter estimation. Here, we consider the problem of the estimation of Weibull parameters. Many authors have investigated various aspects of this problem. Seyit and Ali (2009) presented power density method for Weibull parameters estimation. El-Mezouar (2010) proposed the Coefficient of Variation (CV) estimator comparing with Cran (1988) of the estimation of Weibull parameters. Yeliz *et al.* (2011) compared the method based on quantiles, maximum spacing method, MLE, MOM, LSM and WLSM for Weibull parameters estimation.

In this study, we propose the Randomized Neighborhood Search technique (RNS) for the estimation of the Weibull parameters for the claim severity of fire accidents; the data were provided by the Thai Reinsurance Public Co., Ltd. Five estimation methods (MLE, MOM, LSM, WLSM and RNS) were used to estimate the Weibull parameters. Based on chi-squared value, RNS estimates the parameters more accurately than do MLE, MOM, LSM or WLSM.

2. MATERIALS AND METHODS

2.1. Weibull Distribution

Catastrophe insurance covers large insurance losses that happen infrequently, but have payouts for claims. Examples include large-scale fire, windstorm or flood insurance. In case of catastrophes, claim severity has heavy tails. The Weibull distribution with a shape parameter of less than one and a scale parameter greater than zero is a clear example of heavy-tailed distribution. The probability density and cumulative distribution function forth three-parameter Weibull random variable X, in which each is defined by Equation 1 and 2:

Corresponding author: Soontorn Boonta, Institute of Science, School of Mathematics, Suranaree University of Technology, NakhonRatchasima 30000, Thailand

$$f(x;\alpha,\beta,\gamma)=\frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right) \quad (1)$$

And:

$$F(x;\alpha,\beta,\gamma)=1-\exp\left(-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right) \quad (2)$$

where, $\alpha>0$, $\beta>0$ and $\gamma>0$ and are the shape, scale and location parameters respectively. In this study, we consider claim severity x with a cost greater than 20 million baht. Thus we set $\gamma = 20$ Let $y = x-\gamma$. It then follows from (1) and (2) that for each $y\geq 0$:

$$f(y;\alpha,\beta)=\frac{\alpha}{\beta}\left(\frac{y}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{y}{\beta}\right)^{\alpha}\right)$$

And Equation 3:

$$F(y;\alpha,\beta)=1-\exp\left(-\left(\frac{y}{\beta}\right)^{\alpha}\right) \quad (3)$$

2.2. Estimation of the Weibull Parameters

2.2.1. Maximum Likelihood Estimator (MLE)

Let $y_1, y_2, \ldots, y_n$ be a random sample for the Weibull distribution, then the likelihood function L is defined as Equation 4:

$$L(y_1,y_2,\ldots,y_n;\alpha,\beta)=\prod_{i=1}^{n}\frac{\alpha}{\beta}\left(\frac{y_i}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{y_i}{\beta}\right)^{\alpha}\right) \quad (4)$$

On taking the logarithms of (4), differentiated with respect to β and α and equal to zero, one gets:

$$\frac{\partial \ln L}{\partial\beta}=-\frac{n}{\beta}+\frac{1}{\beta^{\alpha+1}}\sum_{i=1}^{n}(y_i)^{\alpha}=0,$$

$$\frac{\partial \ln L}{\partial\alpha}=\frac{n}{\alpha}-n\ln\beta+\sum_{i=1}^{n}\ln y_i-\sum_{i=1}^{n}\left(\frac{y_i}{\beta}\right)^{\alpha}\ln\left(\frac{y_i}{\beta}\right)=0$$

After solving the above two equations, we obtain Equation 5 and 6:

$$\beta=\left(\frac{1}{n}\sum_{i=1}^{n}y_i{}^{\alpha}\right)^{\frac{1}{\alpha}} \quad (5)$$

$$\alpha=\left[\frac{\sum_{i=1}^{n}(y_i)^{\alpha}\ln y_i}{\sum_{i=1}^{n}(y_i)^{\alpha}}-\frac{1}{n}\sum_{i=1}^{n}\ln y_i\right]^{-1} \quad (6)$$

The value α has to be obtained from (6) by Newton-Raphson and then α is inserted into (5) to obtain β.

2.3. Methods of Moments (MOM)

We know that the kth moment μ_k for the Weibull distribution is given by:

$$\mu_k=\beta^{k}\Gamma\left(1+\frac{k}{\alpha}\right)$$

where, $\Gamma(t)$ defines the gamma function as:

$$\Gamma(t)\int_{0}^{\infty}e^{-x}x^{t-1}dx, t>0$$

In particular, the mean μ (the first moment) and the variance σ^2 are Equation 7 and 8:

$$\mu=\beta\Gamma\left(1+\frac{1}{\alpha}\right) \quad (7)$$

$$\sigma^2=\mu_2-(\mu)^2=\beta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right] \quad (8)$$

The coefficient of variation CV for the Weibull distribution can be determined as follows Equation 9:

$$CV=\frac{\sigma}{\mu}=\frac{\sqrt{\Gamma(1+\frac{2}{\alpha})-\Gamma^2(1+\frac{1}{\alpha})}}{\Gamma(1+\frac{1}{\alpha})} \quad (9)$$

The shape parameter α as appears in (9) will be determined by bisection and the scale β may be calculated from (7).

Another method of moment has been proposed by Cran (1988). Let $x_{(1)}\leq x_{(2)}\leq\ldots\leq x_{(n)}$ be an ordered random sample of the cumulative distribution function $F_{(y)}$ as in (3). Then $F_{(y)}$ can be estimated by $S_n(x)$ where:

$$S_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \dfrac{r}{n}, & x_{(r)} \le x < x_{(r+1)},\ r = 1,\ldots,n-1 \\ 1, & x_{(n)} \le x. \end{cases}$$

Then the population moment μ_k is estimated by:

$$m_k = \int_0^\infty [1 - S_n(x)]^k dx$$
$$= \sum_{r=0}^{n-1} \left(1 - \frac{r}{n}\right)^k (x_{(r+1)} - x_{(r)}),\ \ x_{(0)} = 0$$

He expresses the parameters in terms of lower order moment as follows:

$$\alpha = (\ln 2)\left(\ln(\mu_1 - \mu_2) - \ln(\mu_2 - \mu_4)\right)^{-1}$$

And:

$$\beta = \mu_1 \left(\Gamma(1 + \frac{1}{\alpha})\right)^{-1}$$

Therefore, α and β can be obtained by substituting m_1, m_2 and m_4 for μ_1, μ_2 and μ_4 respectively.

2.4. Least Squares Method (LSM)

We note from (3) that a probability F_i is assigned to each y_i. Since true value of F_i is unknown, a prescribed estimator must to be used. The following four expressions which are often used to define the probability estimator Equation 10a-10d.

$$F_i = \frac{i - 0.5}{n} \tag{10a}$$

$$F_i = \frac{i}{n+1} \tag{10b}$$

$$F_i = \frac{i - 0.3}{n + 0.4} \tag{10c}$$

$$F_i = \frac{i - 3/8}{n + 1/4} \tag{10d}$$

where, F_i is the probability for the ith ranked y_i and n is the sample size.

By applying the logarithm to (3), we get a linear form:

$$\ln\ln\left[\frac{1}{1-F}\right] = \alpha \ln y - \alpha \ln \beta \tag{11}$$

The shape parameter α can be obtained from the slope term in (11) and the scale parameter β can be solved from the intercept term.

2.5. Weighted Least Squares Method (WLSM)

For this method, we follow the technique given by Wu *et al*. (2006). Equation (11) can be rewritten in the form Y = mS+b, where:

$$Y = \ln\ln\left[\frac{1}{1-F}\right],\ m = \alpha,\ S = \ln y \ \text{ and } \ b = -\alpha \ln \beta$$

WLSM is based on the hypothesis that a straight line fitting must minimize the weighted sum of the squares of deviations for the data Y_i from the fitting function $Y(S_i)$, so the equation:

$$l^2 = \sum_{i=1}^{n} W_i (Y_i - b - mS_i)^2$$

gives the minimum value. By solving $\frac{\partial l^2}{\partial m} = \frac{\partial l^2}{\partial b} = 0$, we compute:

$$m = \alpha = \frac{\sum_{i=1}^{n} W_i \sum_{i=1}^{n} S_i Y_i W_i - \sum_{i=1}^{n} S_i W_i \sum_{i=1}^{n} Y_i W_i}{\sum_{i=1}^{n} W_i \sum_{i=1}^{n} S_i^2 W_i - (\sum_{i=1}^{n} S_i W_i)^2},$$

$$b = \frac{\sum_{i=1}^{n} Y_i W_i - \alpha \sum_{i=1}^{n} S_i W_i}{\sum_{i=1}^{n} W_i}$$

where, W_i is the weight factor for the ith datum point. The parameter β can be calculated from:

$$\beta = \exp\left(-\frac{b}{m}\right)$$

It is clear that LSM is a special case of WLSM at $W_i = 1$.

They used the weight factor based on the theory of error propagation Equation 12a and 12b:

$$W_i = \left[(1-F_i)\ln(1-F_i)\right]^2 \quad (12a)$$

$$W_i = 3.3F_i - 27.5\left[1-(1-F_i)^{0.025}\right] \quad (12b)$$

Similar to LSM, the probability F for each datum ranked in ascending order is also approximated by F_i as shown from (10a) to (10d).

We consider a data set of fire insurance claims in Thailand from 2000 to 2004. These data were provided by the Thai Reinsurance Public Co., Ltd. They consist of the claim times and the claim severity x_i. The amount y_i as shown in **Table 1**, is represents amounts above 20 million baht, i.e., $y_i = x_i-20$. For convenience, we still call the amount y_i claim severity.

Table 2 shows the shape parameters α and scale parameters β using different estimation methods for the data found in **Table 1**.

2.6. Chi-Squared

Chi-squared is defined as:

$$x^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where, k is the total number of intervals, O_i is the observed frequency for intervali, E_i is the expected frequency for interval i and:

$$E_i = n[F(y_i) - F(y_{i-1})],\ i = 1,\ 2, \ldots, F(y_0) = 0$$

Here n is the sample size, F is the cumulative distribution function as in (3) and y_i, y_{i-1} are the endpoints of the interval.

We performed the chi-squared goodness of fit test for all methods in **Table 2**. The null hypothesis H_0: data is assumed for Weibull (α,β). We found that the chi-squared value is less than the chi-squared critical value for degree of freedom 4 at a significance level of 0.05 For example, H_0: data is the assumed Weibull (α = 0.9286, β = 30.0055). The chi-squared critical valuefor degree of freedom 4 at a significance level of 0.05 is 9.49, whereas the chi-squared value is 4.0569 (**Table 3**). Thus we can assume that the distribution of the data (**Table 1**) is Weibull at a 5% degree of significance.

Table 1. Claim times and claim severity y_1 (million baht)

2000					
6-Mar	12-Mar	12-Mar	25-Mar	13-Jul	26-Aug
15.5	6.4	44.9	107.3	37.7	1.8
3-Sep	24-Oct				
47.3	28.5				
2001					
16-Jan	28-Jan	17-Feb	22-Feb	9-Mar	19-Jun
3.6	2.3	64.6	1.4	31.5	0.7
20-Jun	5-Jul	6-Aug	24-Aug	18-Sep	23-Oct
20.1	9.3	6.7	12.4	56.5	13.2
29-Nov	1-Dec				
5.7	40.2				
2002					
27-Jan	2-Mar	10-Apr	13-Apr	2-Jun	23-Aug
112.2	0.9	45.8	35.3	13	2.1
26-Oct	29-Oct				
4.2	24.4				
2003					
9-Jan	5-Feb	8-Apr	14-Apr	7-May	23-Nov
0.4	10.8	49.9	102.7	138.9	13.1
2004					
2-Jan	2-Jan	2-Jan	7-Feb	28-Feb	5-Mar
40	84.3	9.2	43.1	70	7.2
14-Mar	22-Apr	8-Jul	1-Nov	24-Dec	
2.4	7.5	37.2	14.2	33.2	

Table 2. Shape α and scale β parametersusing various estimation methods

Method	Type	W_i	F_i	α	β
1	MLE	-	-	0.8633	28.8668
2	MOM (CV)	-	-	0.9286	30.0055
3	MOM (Cran)	-	-	0.9552	30.4239
4	LSM_1	-	10a	0.8580	28.6168
5	LSM_2	-	10b	0.7984	29.1888
6	LSM_3	-	10c	0.8310	28.8602
7	LSM_4	-	10d	0.8405	28.7721
8	WLSM_1	12a	10a	0.7647	29.9050
9	WLSM_2	12a	10b	0.7455	30.1924
10	WLSM_3	12a	10c	0.7571	30.0176
11	WLSM_4	12a	10d	0.7600	29.9750
12	WLSM_5	12b	10a	0.7967	29.2036
13	WLSM_6	12b	10b	0.7710	29.5150
14	WLSM_7	12b	10c	0.7868	29.3166
15	WLSM_8	12b	10d	0.7907	29.2713

Table 3. Chi-Squared, α = 0.9286 and β = 30.0055

Row^i	y_i	$F(y_i) - F(y_{i-1})$	E_i	O_i	$(O_i-E_i)^2/E_i$
1	6	0.20094	9.4441	11	0.2563
2	12	0.14658	6.8892	7	0.0018
3	18	0.11571	5.4384	6	0.0580
4	30	0.16883	7.9350	3	3.0692
5	42	0.11295	5.3088	7	0.5388
6	66	0.12996	6.1082	7	0.1302
7	∞	0.12503	5.8763	6	0.0026
	Totals	1	47	47	4.0569

3. RESULTS

3.1. Randomized Neighborhood Search (RNS)

Randomized neighborhood search is a numerical optimization method whose objective functions may be discontinuous and non-differentiable. This optimization is also known as a direct-search or derivative-free method. Randomized neighborhood search operates by iterative random moving from the initial solution to a better solution. The RNS algorithm is as follows:

Step 1 : Start from the initial parameters α and β. Compute the chi-squared value.

Step 2 : Randomly change the value α to α' and β to β'. We can do this by choosing a uniform variate μ from the interval [0,1] and let:

$$\alpha' = \alpha + 2(0.5 - u)(0.1998),$$
$$\beta' = \beta + 2(0.5 - u)(4.995)$$

Step 3 : Compute chi-squared value with α' and β'.

Step 4 : Compare the chi-squared values which were obtained from steps 1 and 3.

If the chi-squared value of step 3 is greater than or equal to that of step 1, then repeat step 2.

If not, we set $\alpha = \alpha'$, $\beta = \beta'$ and then go on to step 2.

Step 5 : Repeat until a termination criterion is met (adequate fitness reached).

From **Table 1**, we compute the mean (μ) and variance (σ^2):

$$\mu = 31.055319$$
$$\sigma^2 = 1,120.3337743$$

When we replace μ and σ^2 in (8) and then approximate α by bisection, we get α = 0.9286. The approximate value of β = 30.0055 can be obtained from (7). These two parameters α and β will be used as the initial parameters for the RNS algorithm. We iterated RNS 10,000 times and obtained the results shown in **Table 4**.

Table 5 shows the shape parameters α, scale parameters β and chi-squared value using different estimation methods.

Table 4. Parameters α, β and chi-squared value by RNS

Times	α	β	Chi-Squared
1	0.9286000000	30.0055000000	4.0569000000
2	0.8381502696	33.0173413017	3.2266108642
3	0.8381502696	33.0173413017	3.2266108642
4	0.8381502696	33.0173413017	3.2266108642
5	0.8381502696	33.0173413017	3.2266108642
6	0.7076414583	28.7762666642	2.6481293827
7	0.7076414583	28.7762666642	2.6481293827
8	0.7076414583	28.7762666642	2.6481293827
9	0.7076414583	28.7762666642	2.6481293827
10	0.7076414583	28.7762666642	2.6481293827
20	0.7095244694	29.5717808657	2.6140305005
30	0.7148708496	26.9714010325	2.5856511398
40	0.7148708496	26.9714010325	2.5856511398
50	0.7148708496	26.9714010325	2.5856511398
60	0.7148708496	26.9714010325	2.5856511398
70	0.7148708496	26.9714010325	2.5856511398
80	0.7131632905	30.1643026571	2.5559654901
90	0.7131632905	30.1643026571	2.5559654901
100	0.7131632905	30.1643026571	2.5559654901
200	0.7160097628	28.1949938030	2.4788283052
300	0.7160097628	28.1949938030	2.4788283052
400	0.7160097628	28.1949938030	2.4788283052
500	0.7160097628	28.1949938030	2.4788283052
600	0.7160097628	28.1949938030	2.4788283052
700	0.7160097628	28.1949938030	2.4788283052
800	0.7157118026	28.4146840369	2.4783086176
900	0.7157118026	28.4146840369	2.4783086176
1,000	0.7158868924	28.4191078089	2.4745204778
2,000	0.7157825238	28.7701321511	2.4707423531
3,000	0.7158410970	28.7428309644	2.4697671190
4,000	0.7158324217	28.7679580993	2.4697088719
5,000	0.7158182813	28.8071827518	2.4696879774
6,000	0.7158147162	28.8246778976	2.4696432384
7,000	0.7158161544	28.8206039617	2.4696395832
8,000	0.7158168650	28.8183923908	2.4696393961
9,000	0.7158170080	28.8179670891	2.4696392513
10,000	0.7158169062	28.8182970081	2.4696391693

Table 5. Chi-squared value for various estimation methods

Method	Type	α	β	Chi-Squared
1	MLE	0.8633	28.8668	5.9412
2	MOM (CV)	0.9286	30.0055	4.0569
3	MOM (Cran)	0.9552	30.4239	4.4097
4	LSM_1	0.8580	28.6168	5.9758
5	LSM_2	0.7984	29.1888	3.4099
6	LSM_3	0.8310	28.8602	6.0731
7	LSM_4	0.8405	28.7721	6.0239
8	WLSM_1	0.7647	29.9050	3.7284
9	WLSM_2	0.7455	30.1924	3.4214
10	WLSM_3	0.7571	30.0176	3.8360
11	WLSM_4	0.7600	29.9750	3.7936
12	WLSM_5	0.7967	29.2036	3.4216
13	WLSM_6	0.7710	29.5150	3.6609
14	WLSM_7	0.7868	29.3166	3.4988
15	WLSM_8	0.7907	29.2713	3.4662
16	RNS	0.7158	28.8183	2.4696

4. DISCUSSION

We should apply the RNS to other distributions for parameter estimation. The RNS should be applied to a mixture models; it is using the MLE via the Expectations-Maximization (EM) algorithm (Sattayatham and Talangtam (2012) for detail). In the other, we should consider the data of truncated and/or censored data sets in further research.

5. CONCLUSION

In this study, we have used RNS to estimate the Weibull parameters for the claim severity of fire accidents that cost more than 20 million baht. **Table 5** shows RNS has the smallest chi-squared value (i.e., chi-squared value = 2.4696). Therefore RNS givesa more accurate estimation of parameters than do MLE, MOM, LSM or WLSM.

6. ACKNOWLEDGMENT

This research was partially supported by the Centre of Excellence in Mathematics, the Commission of Higher Education (CHE), Sriayudthaya Road, Bangkok 10140, Thailand.

7. REFERENCES

Cran, G.W., 1988. Moment estimators for the 3-parameter Weibull distribution. IEEE Trans. Reliab., 37: 360-363. DOI: 10.1109/24.9839

El-Mezouar, Z.C., 2010. Estimation the shape location and scale parameters of the Weibull distribution. RTA, 4: 36-40.

Sattayatham, P. and T. Talangtam, 2012. Fitting of finite mixture distributions to motor insurance claims. J. Math. Stat., 8: 49-56. DOI: 10.3844/jmssp.2012.49.56

Seyit, A.A. and D. Ali, 2009. A new method to estimate Weibull parameters for wind energy application. Energy Conver. Mange., 50: 1761-1766. DOI: 10.1016/j.enconman.2009.03.020

Wu, D, J. Zhou and Y. Li, 2006. Methods for estimating weibull parameters for brittle materials. J. Mater. Sci., 41: 5630-5638. DOI: 10.1007/s10853-006-0344-9

Yeliz, M.K., K. Mehmet and O.H. Fatih, 2011. Comparison of six different parameter estimation methods in wind power applications. Sci. Res. Essays, 6: 6594-6604. DOI: 10.5897/SRE11.549