

Ivana Pobocikova – Zuzana Sedliackova *

THE LEAST SQUARE AND THE WEIGHTED LEAST SQUARE METHODS FOR ESTIMATING THE WEIBULL DISTRIBUTION PARAMETERS – A COMPARATIVE STUDY

In this paper we study the performance of the least square method and the weighted least square method for estimating the Weibull distribution parameters. In engineering practice these methods are commonly used due to their simplicity. The estimates of the parameters can be calculated easily by the closed-form formula. We consider three estimators of the cumulative distribution function and the weight factor proposed by Bergman (1986). The methods are compared in terms of the root mean square error and sample size. The comparison is based on the Monte Carlo simulation. The comparison shows that the weight factor improves the accuracy of the estimation the Weibull distribution parameters.

Keywords: Weibull distribution, parameter estimation, least square method, weighted least square method, root mean square error.

1. Introduction

The Weibull distribution is one of the widely used distributions in engineering practice. It is named after Walodi Weibull (1887–1979), who popularized its use in the theory of reliability, especially for metallurgical failure models. Moreover, the Weibull distribution is useful for description of the life time of the machine components, for description of mechanical properties of the materials as fatigue of materials and strength of materials.

We consider the two parameter Weibull distribution. The probability density function of the Weibull distribution with parameters $c > 0$ and $\delta > 0$, abbreviated $W(c, \delta)$, is given by

$$f(x) = \frac{c}{\delta^c} x^{c-1} \exp\left(-\left(\frac{x}{\delta}\right)^c\right)$$

where $x > 0$, c is the shape parameter and δ is the scale parameter.

The cumulative distribution function of the Weibull distribution is

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\delta}\right)^c\right), \quad x > 0. \tag{1}$$

The mean μ and the variance σ^2 of the Weibull distribution are

$$\mu = \delta \Gamma\left(1 + \frac{1}{c}\right),$$

$$\sigma^2 = \delta^2 \left[\Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right) \right],$$

where $\Gamma(a)$ is the gamma function defined by $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$, $a > 0$.

The failure rate function of the Weibull distribution is given by

$$H(x) = \frac{c}{\delta^c} x^{c-1}.$$

Fig. 1 shows the effect of the shape parameter c on the density function for different values c and $\delta = 1$. Fig. 2. shows the effect of the scale parameter δ on the density function for different values δ and $c = 2$.

In this paper we study the performance of the methods for estimating the Weibull distribution parameters c and δ . The estimates of the parameters c and δ can be obtained in more ways. The commonly used methods are the maximum likelihood method (MLM), the method of moments (MOM), the least square method (LSM) and the weighted least square method (WLSM). The MLM is the most popular for its efficiency and good properties, but the calculation is complicated. The estimates of the parameters can be obtained only numerically. Several authors have studied and compared performance of the methods for estimating the Weibull distribution parameters, e. g. Bergman [1], Chu and Ke [2], Faucher and Tyson [3], Lu, Chen and Wu [4], Trustrum and Jayatilaka [5], Wu, Zhou and Li [6], Zerda [7].

Here, we consider the least square method (LSM) and the weighted least square method (WLSM), each with three estimators of the cumulative distribution function $F(x)$. In engineering practice these methods are commonly used due to their simplicity. The estimates of the parameters can be calculated easily by the closed-form formula. The methods are compared using the Monte Carlo simulation. The comparison is based on the root mean square error (RMSE) and the sample size n . Based on the simulation study we recommend the methods which have better performance.

* Ivana Pobocikova, Zuzana Sedliackova

Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Zilina, Slovakia, E-mail: ivana.pobocikova@fstroj.uniza.sk

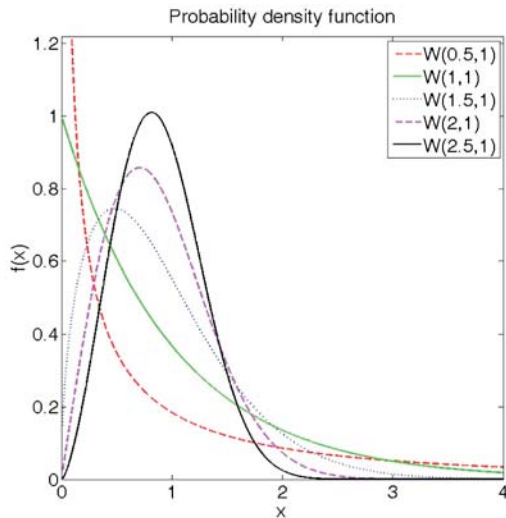


Fig. 1 Shows the effect of the shape parameter c on the density function for different values c and $\delta = 1$

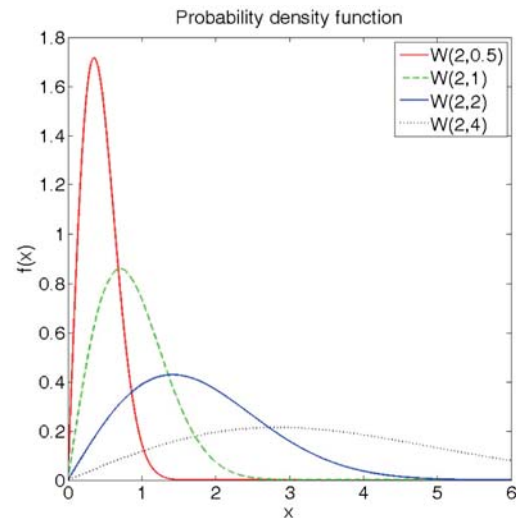


Fig. 2 Shows the effect of the scale parameter δ on the density function for different values δ and $c = 2$

2. Estimation of the parameters of the Weibull distribution

In this section we introduce the methods for estimating the Weibull distribution parameters. The estimates of the parameters c and δ denote \hat{c} and $\hat{\delta}$, respectively. Let X_1, X_2, \dots, X_n be a random sample of size n from the Weibull distribution $W(c, \delta)$ and let x_1, x_2, \dots, x_n be a realization of a random sample.

Least square method

Now, the cumulative distribution function (1) will be transformed to a linear function. From (1) by two logarithmic calculations we obtain

$$\ln[-\ln(1 - F(x))] = c \ln x - c \ln \delta \tag{2}$$

Let $Y = \ln[-\ln(1 - F(x))]$, $X = c \ln x$, $\beta_1 = c$ and $\beta_0 = -c \ln \delta$. Then the equation (2) can be written as

$$Y = \beta_1 X + \beta_0$$

Now let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the order statistics of X_1, X_2, \dots, X_n and let $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ be observed ordered observations. To estimate the values of the cumulative distribution function $F(x)$ we can use the following methods

$$\hat{F}(x_{(i)}) = \frac{i}{n+1} \quad (\text{the mean rank}) \tag{3}$$

$$\hat{F}(x_{(i)}) = \frac{i - 0.5}{n}, \tag{4}$$

$$\hat{F}(x_{(i)}) = \frac{i - 0.3}{n + 0.4} \quad (\text{the median rank}) \tag{5}$$

where i denotes the i^{th} smallest value of $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, $i = 1, 2, \dots, n$.

The estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the regression parameters β_0 and β_1 minimize the function

$$\begin{aligned} Q(\beta_0, \beta_1) &= \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 = \\ &= \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 \ln x_{(i)})^2. \end{aligned}$$

Therefore, the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the parameters β_0 and β_1 are given by

$$\hat{c} = \hat{\beta}_1 = \frac{n \sum_{i=1}^n \ln x_{(i)} \ln[-\ln(1 - \hat{F}(x_{(i)}))] - \sum_{i=1}^n \ln x_{(i)} \sum_{i=1}^n \ln[-\ln(1 - \hat{F}(x_{(i)}))]}{n \sum_{i=1}^n \ln^2 x_{(i)} - \left(\sum_{i=1}^n \ln x_{(i)}\right)^2},$$

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n \ln[-\ln(1 - \hat{F}(x_{(i)}))] - \hat{c} \frac{1}{n} \sum_{i=1}^n \ln x_{(i)}.$$

The estimate $\hat{\delta}$ of the parameter δ is given by

$$\hat{\delta} = \exp\left(-\frac{\hat{\beta}_0}{\hat{c}}\right) = \exp\left(-\frac{\sum_{i=1}^n \ln[-\ln(1 - \hat{F}(x_{(i)}))] - \hat{c} \sum_{i=1}^n \ln x_{(i)}}{\hat{c} n}\right).$$

Weighted least square method

We suppose that the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the regression parameters β_0 and β_1 minimize the function

$$\begin{aligned} Q(\beta_0, \beta_1) &= \sum_{i=1}^n w_i (Y_i - \beta_0 - \beta_1 X_i)^2 = \\ &= \sum_{i=1}^n w_i (Y_i - \beta_0 - \beta_1 \ln x_{(i)})^2 \end{aligned}$$

where w_i is the weight factor, $i = 1, 2, \dots, n$. Therefore, the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the parameters β_0 and β_1 and are given by

$$\hat{c} = \hat{\beta}_1 = \frac{\sum_{i=1}^n w_i \sum_{j=1}^n w_j \ln x_{(i)} \ln[-\ln(1 - \hat{F}(x_{(i)}))] - \sum_{i=1}^n w_i \ln x_{(i)} \sum_{j=1}^n w_j \ln[-\ln(1 - \hat{F}(x_{(j)}))]}{\sum_{i=1}^n w_i \sum_{j=1}^n w_j \ln^2 x_{(i)} - \left(\sum_{i=1}^n w_i \ln x_{(i)}\right)^2},$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n w_i \ln[-\ln(1 - \hat{F}(x_{(i)}))] - \hat{c} \sum_{i=1}^n w_i \ln x_{(i)}}{\sum_{i=1}^n w_i}.$$

Then the estimate $\hat{\delta}$ of the parameter δ is given by

$$\hat{\delta} = \exp\left[\frac{\sum_{i=1}^n w_i \ln[-\ln(1 - \hat{F}(x_{(i)}))] - \hat{c} \sum_{i=1}^n w_i \ln x_{(i)}}{\hat{c} \sum_{i=1}^n w_i}\right].$$

In this paper we use the weight factor proposed by Bergman [1]

$$w_i = \left[(1 - \hat{F}(x_{(i)})) \ln(1 - \hat{F}(x_{(i)}))\right]^2, i = 1, 2, \dots, n.$$

3. Monte Carlo simulation

We generate by the Monte Carlo simulation the random samples from the Weibull distribution and compare the performance of the methods for estimating the Weibull distribution parameters mentioned above. In simulation study we consider the LSM and the WLSM each with three methods for estimating the cumulative distribution function. Thus together we compare six methods. We denote the methods with the estimators of the cumulative distribution function (3) as LSM_1, WLSM_1, with (4) as LSM_2, WLSM_2 and with (5) as LSM_3, WLSM_3.

We consider sample sizes $n = 5$ to 100, $\delta = 1$ and several values of the parameters $c = 0.5, 1.5, 2.5$ representing decreasing, increasing and concave, increasing and convex failure rate functions respectively. All possible combinations of the parameters c, δ and sample sizes n are considered. For each combination c, δ and n we generate by the Monte Carlo simulation $N = 5000$ random samples from the Weibull distribution. For each of six methods we obtain 5000 estimates $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_{5000}$ of the parameter c and 5000 estimates $\hat{\delta}_1, \hat{\delta}_2, \dots, \hat{\delta}_{5000}$ of the parameter δ . Then we compute for each method the sample means $\bar{c}, \bar{\delta}$ and the sample variances s_c^2, s_δ^2 , i. e.

$$\bar{c} = \frac{1}{5000} \sum_{i=1}^{5000} \hat{c}_i, \quad \bar{\delta} = \frac{1}{5000} \sum_{i=1}^{5000} \hat{\delta}_i,$$

$$s_c^2 = \frac{1}{4999} \sum_{i=1}^{5000} (\hat{c}_i - \bar{c})^2, \quad s_\delta^2 = \frac{1}{4999} \sum_{i=1}^{5000} (\hat{\delta}_i - \bar{\delta})^2.$$

To compare the performance of the various methods we compute the sample root mean square error (*RMSE*) defined by

$$RMSE = \sqrt{\frac{1}{5000} \sum_{i=1}^{5000} [(\hat{c}_i - \bar{c})^2 + (\hat{\delta}_i - \bar{\delta})^2]}.$$

The estimates with smaller variance and *RMSE* are preferred. The simulations and the calculation are performed in the Matlab system.

4. Comparison of the methods

In this section we summarize the performance of the methods for estimating the Weibull distribution parameters. The methods are compared in terms of the *RMSE* and sample size n .

The results of the comparison for selected sample sizes $n = 5, 10, 30, 50, 100$ are summarized in Tables 1, 2, 3. The tables show the sample means, the sample variances and the sample *RMSE*. Figures 3, 4, 5 show illustrative plots of the *RMSE* for $n = 5$ to 50 (left), for $n = 50$ to 100 (right).

It is evident that the *RMSE* of the least square method is in many cases much larger than the *RMSE* of the weighted least square method for the case studies in this paper. The weight factor improves the accuracy of the estimation the Weibull distribution parameters. When n gets larger the *RMSE* of all methods tends to be smaller.

For the sample size $n \geq 10$ and for $c = 1.5, 2.5$ the comparison shows that the *RMSE* of the LSM_1 is in many cases much larger than the other methods. The LSM_2 and the LSM_3 are comparable methods in many cases in terms of the *RMSE*. The *RMSE* of the LSM_3 is slightly larger than the *RMSE* of the LSM_2. The *RMSE* of the WLSM_2 is larger than the WLSM_1 and the WLSM_3. The *RMSE* of the WLSM_3 is slightly larger than the *RMSE* of the WLSM_1, both methods are comparable for $n \geq 40$. In general, the WLSM_1 provides the best estimates of the Weibull distribution parameters than the other methods in terms of the *RMSE*.

For the small sample size $5 \leq n < 10$ and for $c = 1.5, 2.5$ the comparison shows that in general the *RMSE* of the WLSM_1 outperforms the other methods. The *RMSE* of the LSM_1 is only slightly larger than the *RMSE* of the WLSM_1. The LSM_1 provides good results in these cases.

For the sample size $n \geq 10$ and for $c = 0.5$ the comparison shows that in general the *RMSE* of the WLSM_2 outperforms the other methods. The *RMSE* of the WLSM_3 is slightly larger than the *RMSE* of the WLSM_2. The *RMSE* of the LSM_2 and the

Simulation results of the parameter estimation for real parameters $c = 0.5, \delta = 1$ Table 1

Sample size	Method	\bar{c}	$\bar{\delta}$	s_c^2	s_δ^2	RMSE
n = 5	LSM_1	0.4508	1.6745	0.0745	2.6689	1.7889
	LSM_2	0.6007	1.3947	0.1311	1.8255	1.4568
	LSM_3	0.5271	1.4977	0.1014	2.1039	1.5663
	WLSM_1	0.4319	1.6060	0.0651	2.4445	1.6974
	WLSM_2	0.5485	1.3710	0.1071	1.8767	1.4572
	WLSM_3	0.4940	1.4555	0.0855	2.0588	1.5334
n = 10	LSM_1	0.4363	1.3946	0.0205	0.9475	1.0619
	LSM_2	0.5311	1.2439	0.0298	0.7336	0.9075
	LSM_3	0.4860	1.3034	0.0252	0.8129	0.9645
	WLSM_1	0.4345	1.3256	0.0182	0.8595	0.9939
	WLSM_2	0.5036	1.2197	0.0272	0.7586	0.9132
	WLSM_3	0.4734	1.2595	0.0226	0.7937	0.9403
n = 30	LSM_1	0.4557	1.1618	0.0076	0.2252	0.5108
	LSM_2	0.5048	1.0972	0.0090	0.1957	0.4627
	LSM_3	0.4822	1.1246	0.0084	0.2078	0.4816
	WLSM_1	0.4713	1.1046	0.0066	0.1984	0.4655
	WLSM_2	0.4966	1.0721	0.0079	0.1898	0.4503
	WLSM_3	0.4864	1.0847	0.0073	0.1930	0.4557
n = 50	LSM_1	0.4652	1.1087	0.0047	0.1198	0.3708
	LSM_2	0.5011	1.0637	0.0053	0.1083	0.3430
	LSM_3	0.4848	1.0831	0.0051	0.1131	0.3540
	WLSM_1	0.4823	1.0640	0.0040	0.1108	0.3452
	WLSM_2	0.4976	1.0450	0.0045	0.1080	0.3383
	WLSM_3	0.4915	1.0524	0.0043	0.1091	0.3408
n = 100	LSM_1	0.4760	1.0563	0.0025	0.0533	0.2440
	LSM_2	0.4991	1.0283	0.0027	0.0501	0.2315
	LSM_3	0.4888	1.0405	0.0026	0.0515	0.2363
	WLSM_1	0.4898	1.0258	0.0022	0.0525	0.2355
	WLSM_2	0.4974	1.0166	0.0023	0.0519	0.2333
	WLSM_3	0.4944	1.0202	0.0023	0.0521	0.2341

Simulation results of the parameter estimation for real parameters $c = 1.5, \delta = 1$ Table 2

Sample size	Method	\bar{c}	$\bar{\delta}$	s_c^2	s_δ^2	RMSE
n = 5	LSM_1	1.3407	1.0925	0.7156	0.1192	0.9320
	LSM_2	1.7878	1.0277	1.2710	0.1060	1.2084
	LSM_3	1.5683	1.0527	0.9784	0.1107	1.0471
	WLSM_1	1.2914	1.0764	0.6761	0.1170	0.9178
	WLSM_2	1.6453	1.0164	1.1606	0.1095	1.1363
	WLSM_3	1.4792	1.0394	0.9075	0.1117	1.0105
n = 10	LSM_1	1.2949	1.0598	0.1758	0.0564	0.5270
	LSM_2	1.5759	1.0202	0.2545	0.0519	0.5590
	LSM_3	1.4421	1.0363	0.2154	0.0536	0.5232
	WLSM_1	1.2884	1.0413	0.1562	0.0554	0.5079
	WLSM_2	1.4920	1.0108	0.2348	0.0542	0.5378
	WLSM_3	1.4033	1.0226	0.1948	0.0545	0.5091

n = 30	LSM_1	1.3653	1.0326	0.0676	0.0185	0.3245
	LSM_2	1.5126	1.0135	0.0801	0.0176	0.3131
	LSM_3	1.4447	1.0217	0.0743	0.0180	0.3095
	WLSM_1	1.4098	1.0162	0.0569	0.0181	0.2888
	WLSM_2	1.4853	1.0059	0.0677	0.0181	0.2932
	WLSM_3	1.4549	1.0099	0.0630	0.0181	0.2885
n = 50	LSM_1	1.3945	1.0218	0.0433	0.0112	0.2570
	LSM_2	1.5022	1.0080	0.0485	0.0107	0.2435
	LSM_3	1.4532	1.0140	0.0462	0.0109	0.2438
	WLSM_1	1.4442	1.0083	0.0369	0.0110	0.2261
	WLSM_2	1.4897	1.0022	0.0412	0.0110	0.2286
	WLSM_3	1.4716	1.0046	0.0394	0.0110	0.2264
n = 100	LSM_1	1.4276	1.0155	0.0240	0.0054	0.1868
	LSM_2	1.4968	1.0065	0.0255	0.0053	0.1755
	LSM_3	1.4657	1.0104	0.0248	0.0053	0.1774
	WLSM_1	1.4711	1.0052	0.0200	0.0054	0.1619
	WLSM_2	1.4938	1.0022	0.0211	0.0054	0.1628
	WLSM_3	1.4847	1.0034	0.0206	0.0054	0.1620

Simulation results of the parameter estimation for real parameters $c = 2.5, \delta = 1$ Table 1

Sample size	Method	\bar{c}	$\bar{\delta}$	s_c^2	s_δ^2	RMSE
n = 5	LSM_1	2.2223	1.0423	1.5070	0.0387	1.2745
	LSM_2	2.9630	1.0048	2.6661	0.0361	1.7076
	LSM_3	2.5993	1.0194	2.0564	0.0370	1.4503
	WLSM_1	2.1367	1.0329	1.3909	0.0383	1.2498
	WLSM_2	2.7187	0.9975	2.4000	0.0377	1.5764
	WLSM_3	2.4459	1.0112	1.8685	0.0377	1.3816
n = 10	LSM_1	2.1727	1.0325	0.5197	0.0197	0.8047
	LSM_2	2.6437	1.0094	0.7517	0.0188	0.8894
	LSM_3	2.4194	1.0188	0.6367	0.0191	0.8140
	WLSM_1	2.1559	1.0212	0.4450	0.0196	0.7637
	WLSM_2	2.4961	1.0028	0.6570	0.0196	0.8225
	WLSM_3	2.3478	1.0100	0.5497	0.0195	0.7696
n = 30	LSM_1	2.2736	1.0151	0.1894	0.0064	0.4973
	LSM_2	2.5190	1.0038	0.2245	0.0062	0.4807
	LSM_3	2.4060	1.0086	0.2082	0.0063	0.4727
	WLSM_1	2.3494	1.0056	0.1656	0.0064	0.4412
	WLSM_2	2.4747	0.9994	0.1977	0.0064	0.4525
	WLSM_3	2.4243	1.0018	0.1839	0.0064	0.4427
n = 50	LSM_1	2.3228	1.0145	0.1193	0.0040	0.3935
	LSM_2	2.5024	1.0062	0.1332	0.0039	0.3703
	LSM_3	2.4206	1.0098	0.1269	0.0039	0.3704
	WLSM_1	2.4116	1.0065	0.0997	0.0039	0.3339
	WLSM_2	2.4876	1.0029	0.1112	0.0040	0.3396
	WLSM_3	2.4573	1.0043	0.1065	0.0040	0.3350
n = 100	LSM_1	2.3805	1.0082	0.0646	0.0019	0.2844
	LSM_2	2.4960	1.0028	0.0686	0.0019	0.2656
	LSM_3	2.4441	1.0052	0.0669	0.0019	0.2682
	WLSM_1	2.4580	1.0021	0.0539	0.0019	0.2399
	WLSM_2	2.4959	1.0003	0.0568	0.0019	0.2424
	WLSM_3	2.4808	1.0010	0.0556	0.0019	0.2407

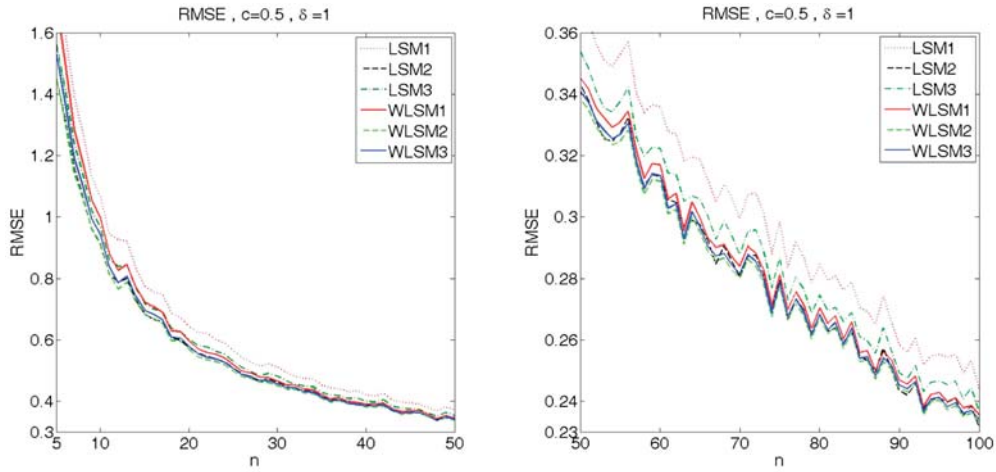


Fig. 3 Root mean square error for real parameters $c = 0.5, \delta = 1$

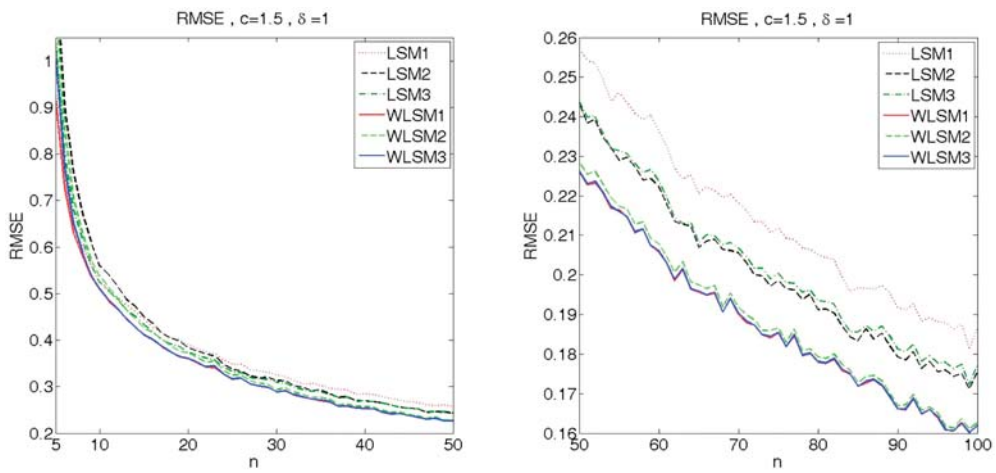


Fig. 4 Root mean square error for real parameters $c = 1.5, \delta = 1$

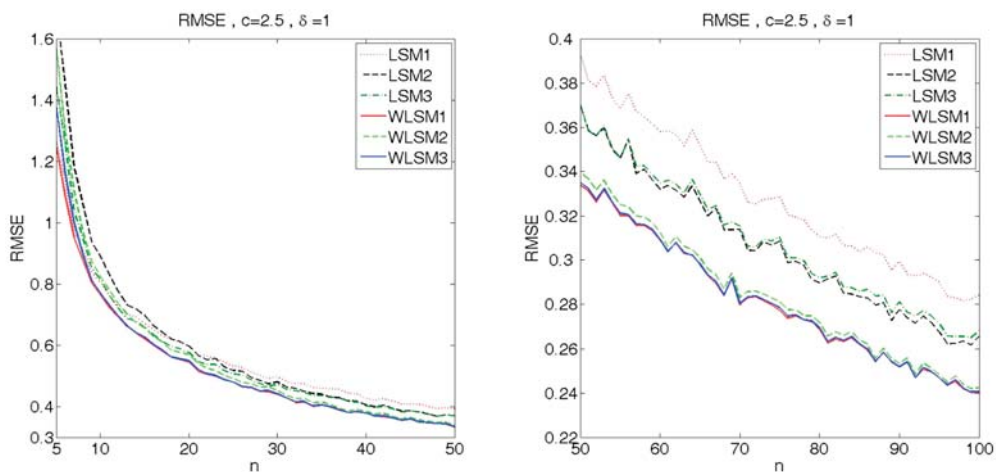


Fig. 5 Root mean square error for real parameters $c = 2.5, \delta = 1$

WLSM_1 follow. The *RMSE* of the LSM_1 is in many cases much larger than the other methods.

For the sample size $5 \leq n < 10$ and for $c = 0.5$ the *RMSE* of the LSM_2 is only slightly larger than the *RMSE* of the WLSM_2. The *RMSE* of the WLSM_3 and the LSM_3 follow. The *RMSE* of the LSM_1 is much larger than the *RMSE* of the other methods.

5. Concluding remarks

In this paper we compared the performance of the methods for estimating the Weibull distribution parameters in terms of the *RMSE* and sample size n . The comparison was based on the Monte Carlo simulation. The comparison shows that the weight factor improves the accuracy of the estimation the Weibull distribution

parameters. The WLSM_1 performs the best in terms of the *RMSE* than the other methods for majority cases studied in this paper and for all sample sizes, the WLSM_3 follows as the second good choice. Except the case when $c = 0.5$, the WLSM_2 performs to be the best for all the sample sizes than the other methods. The good choice is in this case for middle and large sample sizes the WLSM_3.

The advantages of these recommended methods are: simple derivation, easy calculation of the estimates of the parameters by the closed-form formula. And so from this point of view these methods are very useful for engineering practice.

Acknowledgement

This research was supported by the Slovak Grant Agency VEGA through the projects No. 1/1245/12 and No. 1/0797/12.

References

- [1] BERGMAN, B.: Estimation of Weibull Parameters Using a Weight Function. *J. of Materials Science Letters* 5, pp. 611–614, 1986
- [2] CHU, Y. K., KE, CH., J.: Computation Approaches for Parameter Estimation of Weibull Distribution. *Mathematical and Computational Applications*, vol. 17, No. 1, pp. 39–47, 2012
- [3] FAUCHER, B., TYSON, W. R.: On the Determination of Weibull Parameters. *J. of Materials Science Letters* 7, pp. 1199–1203, 1988
- [4] LU, H. L., CHEN, CH. H., WU, J. W.: A Note on Weighted Least-squares Estimation of the Shape Parameter of the the Weibull Distribution. *Quality and Reliability Engineering International* 20, pp. 579–586, 2004
- [5] TRUSTRUM, K., JAYATILAKA, A. S.: On Estimating the Weibull Modulus for a Brittle Material. *J. of Material Science* 14, pp. 1080–1084, 1979
- [6] WU, D., ZHOU, J., LI, Y.: Methods for Estimating Weibull Parameters for Brittle Materials. *J. of Material Science* 41, pp. 5630–5638, 2006
- [7] ZERDA, I.: An Experimental Comparison of Popular Estimation Methods for the Weibull, Gamma, and Gompertz Distributions. *Schedae Informaticae* 20, pp. 67–82, 2011.