International Journal of Mathematics and Computational Science
Vol. 5, No. 1, 2019, pp. 1-5
http://www.aiscience.org/journal/ijmcs
ISSN: 2381-7011 (Print); ISSN: 2381-702X (Online)

A Suggested Method from MAD-Median and Ridge Regression Estimator

Irtefaa A. Neamah*

Department of Mathematics, University of Kufa, Najaf, Iraq

Abstract

For the important of the estimation strategies and its big role in the sciences, we look forward to developing a suggested method based on the other classical methods. This is the main aim of this paper. One of the most important statistical distributions is Weibull distribution. So, we look forward to study more strategies to estimate its parameters shape and scale. In this paper, we study two known methods in estimation. There are ridge regression RR and Mad/Median estimator. But, we decided to create a new suggested method to know what is the best of them. So, the suggested method is a composition between the ridge regression and the mad/median estimator. To know the best result, it be supported this suggestion by a simulation study. Besides that, the comparison based on the most uses criteria is mean squares error MSE. The result was the suggested method gives more efficient result at most of the sample sizes.

Keywords

Estimation, Ridge Regression, Mad/Median Estimator, Weibull Distribution, Mean Square Error, Simulation Study

Received: July 21, 2016 / Accepted: September 2, 2016 / Published online: April 17, 2019

1. Introduction

Always, the applied researches are concerned with a determine the model of the issue. Especially, the problem that is related in errors of the results.

Recently, the need for the methods of estimation has been increased, because the importance of estimation in solving many of the problems of everyday life. However, the most important branches of statistical inference develop daily, in this research I address some of the pre-existing methods to study estimation such as ridge regression, Mad/Median estimator and then, and my suggested new method to get best results of them.

The ridge regression concept was firstly used by Horel and Kennard (1970). They used a nonzero value as a ridge regression parameter which called k. In other word, the mean square error of ridge regression is known as k. [1]

In 1977, Stone & Heewijk checked graphical procedure to estimate the weibull distribution parameters. [2]

The development of six techniques of estimation by composition between minimum distance estimation and maximum likelihood was by Gallagher & Moore. [3]

Many procedures are processed and suggested in literature for the estimation in Many statistical distributions. One of the most famous distributions is Weibull distribution. Because it is extensively used in many fields of life. [4]

Cacciari & etal (2002) proposed a new estimator for the two Weibull distribution parameters, which are capable of providing point estimates that are both efficient (and unbiased) and robust. [4]

In (2007), Sultan, Mahmoud & Saleh, used the setup proposed by Balakrishnan and Aggarwala (2000), for get the best linear unbiased estimates by approximation to two weibull distribution parameters. [5]

* Corresponding author
E-mail address: irtefaa.radhi@uokufa.edu.iq

The most required property for any estimator is the minimum variance and unbiasedness. So, to know the best, it must know the error value. For that, using the Mean Squared Error (MSE). The truth verification is satisfying by which method has less mean square error. More benefit to context of least square is ridge estimation. [6]

Many authors have studied the developing a suggested estimation method using ridge regression, such as, Hefnawy and Farag [7], Aslam [8].

Depend on least median squares, Kafi and others proposed a new modification for estimation by ridge regression and robust which has has a high breakdown point. [9]

Kibria and Banik suggested five estimators using ridge regression based on simulation study. [10]

Although, the ridge regression, Mad/Median estimator are well known used methods, but statisticians need to improve them with time. So, for this reason, I looked forward to suggest a new strategy to amend the estimation style.

This method consisted the improvement that is the composition of the two methods, Mad/Median and ridge regression estimation in other effective one.

This new suggestion uses Weibull probability distribution function to estimate its two parameters scale and shape parameters. To make this claim is good enough, I included a study of numerical simulation to compare the results among them and see which gives the best results. Finally, it is shown that the suggested method gives a better estimation than the others.

2. Weibull Distribution

Ernst Hjalmar Waldoddi Weibull was born in 1887. [11]

Two approaches of the engineering sciences could found by some researchers, which give the clear way to the Weibull distribution that was in 1930s. [11]

The statisticians interested in the Weibull distribution of more than five hundred years ago, since it was found by Waloddi Weibull* in (1951). In the worldwide fields uses, one of the popular and famous statistical distribution to computations of theory and applications directions, this was the Weibull distribution. This distribution has been used in a wide variety of areas, including reliability engineering. [11]

The probability density function (p.d.f.) of the random variable X is said to be Weibull distribution with shape and scale parameter, α and β, respectively, if it is given by:

$$f(x,\alpha,\beta)=\frac{\alpha}{\beta}x^{\alpha-1}e^{\frac{-x^{\alpha}}{\beta}} \tag{1}$$

Where,

$$x,\alpha, and\beta \geq 0$$

1. when α<1 then the failure will be less over time.
2. But if α=1 refers to the rate of failure is fixed over time.
3. On the other hand, if α>1 that means the rate of failure is increasing over time.

The cumulative distribution function (c.d.f) of the Weibull distribution is mathematically given as: [11]

$$F(x) = 1 - exp\left(-\frac{x^{\alpha}}{\beta}\right)$$

3. Mad/Median Estimator

One of the most important quantities for the statistical distributions are the median and the median absolute deviation. Consequently, the median formula is:

$$MED\ (n) = X_{\left(\frac{(n+1)}{2}\right)}$$

if n is odd,

$$MED\ (n) = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\left(\frac{n}{2}\right)+1\right)}}{2},$$

if n is even.

In the same side, the sample median absolute deviation is:

$$MAD\ (n) = MED(\ |X_i - MED\ (n)|), i = 1, \dots, n$$

Olive has considered MAD/Median estimator. This the proposed procedure introduces a new simple estimator. If it be supposed that the random variable X has a Weibull Distribution with p.d.f. in formula (1). Then log-Weibull distribution is the smallest extreme value distribution, X~SEV (θ, σ),), then the p.d.f. of X is:

$$f(x) = \frac{1}{\sigma}exp\left(\frac{x-\theta}{\sigma}\right)exp\left[-exp\left(\frac{x-\theta}{\sigma}\right)\right]$$

Where, x and θ are real, and σ > 0.

On the other hand,

$$\Theta = log\left(\beta^{\frac{1}{\alpha}}\right)$$

And,

$$\sigma = \frac{1}{\alpha}$$

Now, let

$$\hat{\sigma} = \frac{MAD\ (X)}{0.767049}$$

And,

$$\hat{\theta} = MED(X) - log(log(2))\hat{\sigma}$$

$$\hat{\alpha}_{MAD} = \frac{1}{\hat{\alpha}} \quad (2)$$

And,

$$\hat{\beta}_{MAD} = exp(\frac{\hat{\theta}}{\hat{\sigma}}) \quad (3)$$

This method was proposed by Olive. [12]

4. Ridge Regression Estimator

The regression analysis of a broader statistical methods commonly used in various fields of science. As it shows the relationship between the variables in the form of an equation to estimate the parameters inferred on the importance of this relationship, and the strength and direction.

It is worth mentioning, that one of the statistical methods, which has a lot of the famous statistical, characterizes is the classical ordinary least squares (OLS). [9]

The amendment of the classical least squares method (OLS) that lets the regression coefficients for the biased estimators. The preferred estimators are being unbiased estimators. That is mean it is closed to the true value of the parameter, also have the smaller MSE.

The standard model of the multiple linear regression can be considered as:

$$Y = X\beta + \quad (4)$$

The Y is (n × 1) vector of variable values which are independent, X is (n × p) matrix with the matrix (1 × p) is variables values of predictor P with, and (n × 1) is the vector of random variables ε. In addition, the unknown coefficients of regression (p×1) is a vector β.

The ridge estimator is $\hat{\beta}_{Rid}$ be, so for any k ≥ 0

$$\hat{\beta}_{Rid} = (X'X + K * eye)^{-1}.X'Y \quad (5)$$

Where the identity matrix is eye, and the K value computed by:

$$K = \frac{ps_e^2}{\hat{\beta}'_{Ols}\,\hat{\beta}_{Ols}}$$

The vector of estimators is $\hat{\beta}_{Ols}$ by (OLS):

$$\hat{\beta}_{Ols} = (X'X)^{-1}.X'Y$$

For more,

$$s_e^2 = \frac{\sum_i^n e_i^2}{n-p}$$

where e^2 is the mean squared error by (OLS). [1]& [13]

5. The Suggested Estimator

In this study, it has been proposed a new method to estimate the two parameters of Weibull distribution. This proposed method involves the composition of two estimation methods. They are Mad/Median and ridge regression estimator. The strategy to create the new suggested method is start with compute the Mad/Median estimator as mentioned in previous pages. After then, the last method overlap with ridge regression as follows, recall equations (2) and (3):

$$\hat{\alpha}_{MAD} = \frac{1}{\hat{\alpha}}$$

$$\hat{\beta}_{MAD} = exp(\frac{\hat{\theta}}{\hat{\sigma}})$$

Suppose these parameters estimations as an initial parameter in ridge regression:

$$\hat{\beta} = \begin{bmatrix} \hat{\alpha}_{Mad} \\ \hat{\beta}_{Mad} \end{bmatrix}$$

So, $Y = X\hat{\beta} + \varepsilon$

And, $K = \frac{ps_e^2}{\hat{\beta}'_{Mad}\,\hat{\beta}_{Mad}}$

Finally, the suggested estimator will be as:

$$\hat{\beta}_{new} = (X'*X + \sum K * eye(p) \ \ (X'*Y))^{-1}$$

Where, the number of parameter P is the size of identity matrix.

6. Computer Simulation

That the great technological advances that have happened on the computer field during the last three decades, it contributed significantly to the spread of the use of methods. Simulation in formulating and solving mathematical models, complex statistical, and of course this is reflected the use of simulation methods in the teaching and learning. [14]

The computer simulation is executing a study in order to explore the pattern the two parameters of Weibull distribution, shape α and scale β. Independently, generate a random sample of Weibull distribution with α=1.5 and β = 1 in some different sample sizes, n= (10, 25, 30, 60, 75, 100, 150, 200). The iteration for the procedure of this simulation is (r=5000) times. For each sample size n, estimate α and β by three previous methods. The criteria of the comparison among the results is mean squares error (MSE). That is:

$$MES = E\left[\left(\theta - \hat{\theta}\right)^2\right]$$

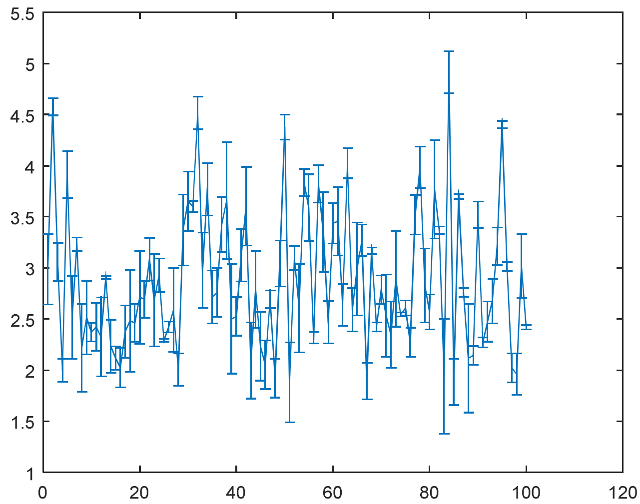

Figure 1. Error bar for ridge regression in sample size 100.

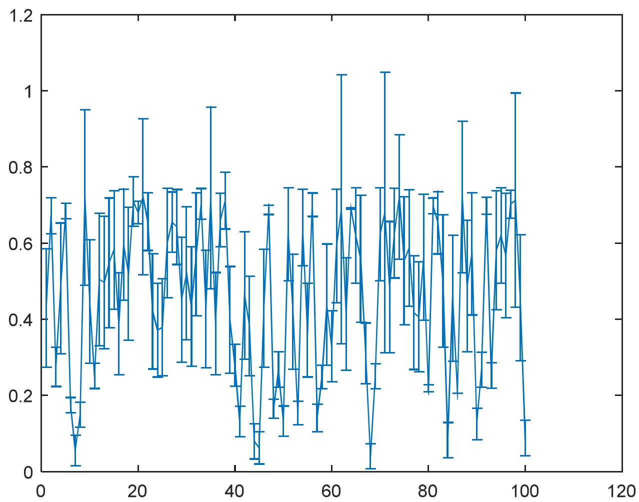

Figure 2. Error bar for new method in sample size 100.

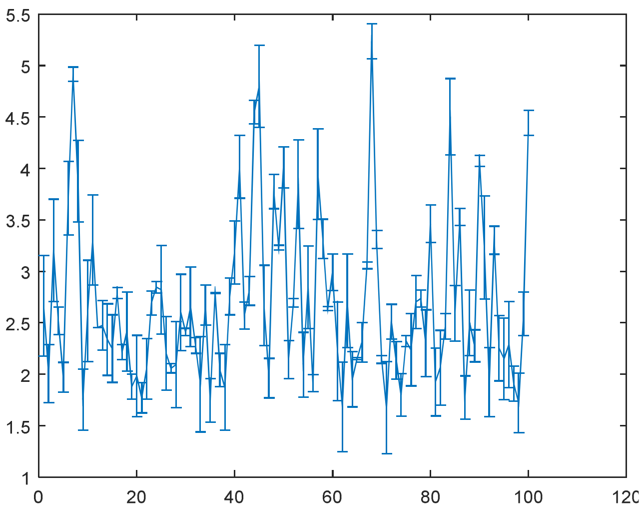

Figure 3. Error bar for mad/Median method in sample size 100.

Table 1. The results of S=simulation based on 5000 iterations in ridge regression RR, Mad/Median and suggested method with α= 1.5 & β=1.

n	Method	$\hat{\alpha}$	MSE	$\hat{\beta}$	MSE
	RR	1.8603	0.2597*10^{-4}	1.1597	0.510*10^{-4}
10	Mad/Median	1.1661	2.2299*10^{-4}	1.2430	1.1813*10^{-4}
	Suggested	1.6415	0.4521*10^{-4}	1.3413	0.0193*10^{-4}
	RR	1.8542	0.2509*10^{-4}	1.1070	0.0229*10^{-4}
25	Mad/Median	1.0137	4.7296*10^{-4}	0.8861	2.5954*10^{-4}
	Suggested	1.6413	0.7876*10^{-4}	0.7862	0.0199*10^{-4}
	RR	1.9800	0.4607*10^{-4}	1.0298	0.0018*10^{-4}
30	Mad/Median	1.9401	3.8730*10^{-4}	1.2098	8.7991*10^{-4}
	Suggested	2.4645	0.5501*10^{-4}	1.1477	0.0077*10^{-4}
	RR	2.0893	0.6945*10^{-4}	0.9189	0.0132*10^{-4}
60	Mad/Median	1.1962	1.8458*10^{-4}	1.5107	5.2170*10^{-4}
	Suggested	1.6466	0.4058*10^{-4}	1.5586	0.0046*10^{-4}
	RR	2.0067	0.5134*10^{-4}	1.0715	0.0102*10^{-4}
75	Mad/Median	1.2618	1.1348*10^{-4}	1.1790	6.4089*10^{-4}
	Suggested	1.7951	0.5687*10^{-4}	1.1706	0.0001*10^{-4}
	RR	1.8603	0.2597*10^{-4}	1.1597	0.0510*10^{-4}
100	Mad/Median	1.1138	2.9835*10^{-4}	1.1165	2.7126*10^{-4}
	Suggested	1.6245	0.5217*10^{-4}	1.1063	0.0002*10^{-4}
	RR	1.9480	0.4015*10^{-4}	1.0049	0.0000
150	Mad/Median	1.2314	1.4432*10^{-4}	1.0295	1.7358*10^{-4}
	Suggested	1.7067	0.4518*10^{-4}	1.0381	0.0001*10^{-4}
	RR	2.0454	0.5948*10^{-4}	0.9536	0.0043*10^{-4}
200	Mad/Median	1.3229	6.2763*10^{-4}	1.0775	1.2012*10^{-4}
	Suggested	1.8119	0.4784*10^{-4}	1.1126	0.0025*10^{-4}

7. Results and Conclusion

In this paper, the computation strategy depends on the numerical results from simulation. The procedure of this simulation contains three statistical methods, which are Mad/Median estimator, ridge regression, and the suggested method of this paper. These methods are applied to estimate the two parameters of weibull distribution, shape and scale α and β respectively.

If you note that, the results of the simulation study showed in the figures 1, 2 and 3. These figures showed the bar chart for the error values of the three estimation methods. It can be recognizing that the error values of the Mad/Median estimator are divergent. That is mean these values are something far from the original values. While the error bar in each other methods are something convergent from the original values of the parameters.

It has been shown that the new method has given the best results especially when the increasing sample size, and it is clear that the performance of the suggested method was better than of the other methods. Therefore, it has lower mean squares error MES in almost of all sample sizes. While the results of the last two methods differed from sample to another. Finally, this suggested method can be recommended to estimate weibull distribution, parameters shape and scales rather than the other methods.

References

[1] E. Horel and R. Kennard, "Ridge Regression: Applications to Nonorthogonal Problems," *Technometrics, American Statistical Association, JSOR, USA.,* 1970.

[2] G. Stone and H. Heeswijk R. G., "Parameter Estimation for The Weibull Distribution," *IEEE Transactions on Reliability. Electr. Insul,* pp. Vol EI- 12 No. 4, August 1977.

[3] M. Gallagher and A. H. Moore, "Robust Minimum Distance Estimation Using the 3-parameter Weibull distribution," *IEEE Transactions on Reliability,* pp. Vol. 39, No. 5, 1990.

[4] M. Cacciari, G. Mazzanti, G. Montanari and J. Jacquelin, "A Robust Technique for the estimation two-Parameter Weibull Function for Complete Data Sets," *Manscript in Italy and France,* 2002.

[5] K. Sultan, M. Mahmoud and H. Saleh, "Estimation of Parameters of the Weibull Distribution Based on Progressively Censored Data," *International Mathematical Forum,* pp. 2, No. 41, 2031-2048, 2007.

[6] Z. ZANGIN, *On Median and Ridge Estimation of SURE Models,* Jönköping: doctoral dissertation at Jönköping International Business School, 2012.

[7] E. A. Hefnawy and F. A, "A combined nonlinear programming model and Kibria method for choosing ridge parameter regression," *Communications in Statistics – Simulation and Computation,,* Vols. 43 (6),. doi, 2013.

[8] M. Aslam, "Performance of Kibria's method for the heteroscedastic ridge regression model, Some Monte Carlo evidence’. Communications in Statistics – Simulation and Computation, Vol. 43 (4), 673-686. doi, 2014.

[9] D. P. Kafi, A. Robiah and A. R. Bello, "Using Ridge Least Median Squares to Estimate the Parameter by Solving Multicollinearity and Outliers Problems," *Modern Applied Science,* Vols. Vol. 9, No. 2, no. ISSN 1913-1844 E-ISSN 1913-1852, pp. 191-198, 2015.

[10] B. Kibria and S. Banik, "Some Ridge Regression Estimators and Their Performances," *Journal of Modern Applied Statistical Methods,* Vols. Vol. 15, No. 1, 206-238., no. ISSN 1538 − 9472, May, 2016.

[11] H. Renne, The Weibull Distribution, A Handbook, Giessen, Germany: Taylor & Francis Group, LLC, 2009.

[12] D. J. Olive, Applied Robust Statistics, Southern Illinois University: Department of Mathematics, 2008.

[13] Rashwan and M. E.-D. a. N. I., "Solving Multicollinearity Problem Using Ridge Regression Models," *Int. J. Contemp. Math. Sciences, Vol. 6, No. 12, 585.,* 2011.

[14] H. AHMED, K. ALANI ISAM and B. A. A. A., "Using Simulation In Teaching Simple Linear Regression," *Journal of Anbar University Pure Sciences,* vol. 2 No. 3, no. ISSN: 1991-8941, 2008.