Mathematical and Computational Applications, Vol. 13, No. 2, pp. 129-136 2008.

A DATA DRIVEN PARAMETER ESTIMATION FOR THE THREE-PARAMETER WEIBULL POPULATION FROM CENSORED SAMPLES

Ahmed M. M. Sultan
Egypt Air Force, Cairo, Egypt
amsultan_52@yahoo.com

Abstract- A method is described for the calculation of the three-parameter Weibull distribution function from censored samples. The method introduces a data driven technique based on an adapted Gaussian like kernel to match the censoring scheme. The method minimizes the Cramer von Mises distance from a non-parametric density estimate and the parametric estimate at the order statistics. The maximum likelihood estimators are found and a comparison is made with the new estimator. A Monte Carlo experiment of size 1000 is conducted to test the performance of the new parameter estimation technique. The mean integrated square error is taken as a measure of the closeness of the estimated density and the true density.

Key Words- Non-parametric density, Weibull censored samples, Gaussian kernel, type II censoring, hybrid methods, Cramer von Mises statistic

1. INTRODUCTION

The method of moment, the method of maximum likelihood, and other methods have considered the estimation of the parameters of Weibull population based on a censored sample. In this paper, an approach using an adapted non-parametric density estimation is introduced as a methodology for the parameter estimation. Section 2 discusses the solution of the log likelihood equations for the censored sample. The method of solution is a modification of the classical Newton-Raphson iterative scheme. The method is based on the numerical solution of the log likelihood equation using a quasi Newton method and an active set strategy to maximize the log likelihood function subject to simple bounds on the distribution parameters. The method is surveyed and a stopping rule is stated. In section 3 the application of a non-parametric density estimator to obtain estimates of the parameters of the three-parameter Weibull distribution from a censored sample is discussed. An adapted kernel is used which is a Gaussian like kernel with a finite right tail. A Monte Carlo comparison of the maximum likelihood estimators and the minimum distance estimators is given using the integrated squared error (ISE) between the true density and the estimated true model. Samples of size 10, 20, and 30 censored at the 7th , 15th , and 20th order statistic respectively are used. The experiment is done for thousand Monte Carlo repetitions. A comparison is made between the maximum likelihood estimators and the new estimators for location parameter 10, with scale parameter 5 and 10 and for shape parameter 3, 4,5, and 6 in tables and figures. The results are shown in section 4. These results indicate an improvement of the new method over the classical maximum likelihood method.

2. MAXIMUM LIKELIHOOD PARAMETER ESTIMATION

The maximum likelihood estimation for the parameters of the Weibull distribution has been studied extensively for complete and censored samples. The studies include those by Harter and Moore (1965, 1967) where they studied the maximum likelihood estimation of the gamma and Weibull population, from censored samples. They also studied the asymptotic variances and covariances of maximum likelihood estimators from censored samples from Weibull and gamma populations. Cohen (1965) studied the maximum likelihood estimation in the Weibull distribution based on complete and censored samples. He also studied (1975) the multi-censored sampling in case of the three-parameter Weibull distribution. Some results on complete and censored sampling from the three parameter Weibull distribution were shown by Wycoff et al.(1980). Cohen et. al. (1984) introduced modified estimators for the parameters of the three-parameter Weibull with smaller biases and smaller variances.
The probability density function of the three-parameter Weibull denoted by W(γ, β, δ)with location γ, scale β, and shape δ is given by:

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

where $\gamma < x < \infty$, $\delta > 0$, $\beta > 0$. The corresponding cumulative distribution function is given by:

$$F(x;\gamma,\beta,\delta)=1-\exp\left\{-\left[\frac{x-\gamma}{\beta}\right]^{\delta}\right\}$$

Now, consider that a sample of size n has been censored at the r^{th} order statistic using a type II censoring mechanism. The resulting density for the first r order statistics will be given as:

$$f\left(x_{(1)},\ldots,x_{(r)}\right)=\frac{n!}{(n-r)!}\left(\prod_{i=1}^{r}f\left(x_{(i)}\right)\right)\left[1-F\left(x_{(r)}\right)\right]^{n-r}$$

$$=\frac{n!}{(n-r)!}\left[\prod_{i=1}^{r}\frac{\delta}{\beta}\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-1}\exp\left\{-\left[\frac{x_{(i)}-\gamma}{\beta}\right]^{\delta}\right\}\right]\left[\exp\left\{-\left[\frac{x_{(r)}-\gamma}{\beta}\right]^{\delta}\right\}\right]^{n-r}$$

Taking the logarithm for the above density gives the following log-likelihood function:

$$L^{*}=\log\left[\frac{n!}{(n-r)!}\right]+r\left[\log\delta-\log\beta\right]+\left[\sum_{i=1}^{r}\log\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-1}\right]-\sum_{i=1}^{r}\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta}-(n-r)\left[\frac{x_{(r)}-\gamma}{\beta}\right]^{\delta}$$

The partial derivatives for the log-likelihood function with respect to the three unknown parameters (γ,β,δ) are:

$$\frac{\partial L^*}{\partial \gamma}=\sum_{i=1}^{r}\frac{(\delta-1)\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-2}\left(\frac{-1}{\beta}\right)}{\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-1}}-\sum_{i=1}^{r}\delta\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-1}\left(\frac{-1}{\beta}\right)-\delta(n-r)\left(\frac{x_{(r)}-\gamma}{\beta}\right)^{\delta-1}\left(\frac{-1}{\beta}\right)$$

$$=\sum_{i=1}^{r}\frac{(\delta-1)\left(\frac{-1}{\beta}\right)}{\left(\frac{x_{(i)}-\gamma}{\beta}\right)}+\left(\frac{\delta}{\beta}\right)\sum_{i=1}^{r}\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-1}+\left(\frac{\delta}{\beta}\right)(n-r)\left(\frac{x_{(r)}-\gamma}{\beta}\right)^{\delta-1}$$

$$\frac{\partial L^*}{\partial \beta}=-\frac{r}{\beta}+\sum_{i=1}^{r}\frac{(\delta-1)\left(\frac{x_{(i)}-\gamma}{\beta}\right)}{\left(\frac{x_{(i)}-\gamma}{\beta}\right)}+\delta\sum_{i=1}^{r}\left(\frac{x_{(i)}-\gamma}{\beta^2}\right)\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta-1}+\delta(n-r)\left(\frac{x_{(r)}-\gamma}{\beta}\right)^{\delta-1}\left(\frac{x_{(r)}-\gamma}{\beta^2}\right)$$

$$=-\frac{r\delta}{\beta}+\frac{\delta}{\beta}\sum_{i=1}^{r}\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta}+\frac{\delta(n-r)}{\beta}\left(\frac{x_{(r)}-\gamma}{\beta}\right)^{\delta}$$

$$\frac{\partial L^*}{\partial \delta}=\frac{r}{\delta}+\sum_{i=1}^{r}\log\left(\frac{x_{(i)}-\gamma}{\beta}\right)-\sum_{i=1}^{r}\left[\left(\frac{x_{(i)}-\gamma}{\beta}\right)^{\delta}\log\left(\frac{x_{(i)}-\gamma}{\beta}\right)\right]-(n-r)\left(\frac{x_{(r)}-\gamma}{\beta}\right)^{\delta}\log\left(\frac{x_{(r)}-\gamma}{\beta}\right)$$

Equating the partial derivatives to zero and solving the system of nonlinear equations simultaneously gives an estimator $\hat{\Theta}_l=(\hat{\gamma},\hat{\beta},\hat{\delta},)$ that maximizes the log-likelihood function and maximizes the likelihood function as well.

The system of the 3-non linear equations for the maximum likelihood in $\Theta=(\gamma,\beta,\delta)$ is solved using a numerical technique. The method is known as the hybrid method. This method is basically an iterative method based on Newton-Raphson method. Such methods need to compute 3 components of L and 9 entries of L'. Several other modifications are introduced by Powell (1970) to relieve such a problem by computing the difference approximations instead of the direct computation of L'.

In Powell's iterative scheme the derivative is not just scaled by a small factor but by introducing a negative multiple of the gradient of $L(\Theta)$ such that the direction for the correction in the different iterations will be sensible when the Jacobian becomes almost singular.

For details about cases when method can, and different factors that affect the running time of the method, see Powell (1970).

An accuracy of .01 was used for the absolute difference between two successive Θ 's while the Euclidean norm accuracy was relaxed since the mean integrated square error (MISE) criteria is to be used latter for the comparison and the interest was in the convergence of the Θ parameter mainly.

The algorithm did not converge in a few cases (number in bold in tables 1-3 in the first column) which were excluded from the Monte Carlo results. This happened because the method was searching for a zero of the system of nonlinear equations $L(\Theta)$ = 0 by minimizing the quadratic form $L^T(\Theta)L(\Theta)$ or the sum of squares of the maximum likelihood equations. In this case, the minimum would not give a zero of the system. The same initial guess is chosen for all the different Monte Carlo samples of size 1000.

The results from the previous for sample sizes 10, 20, and 30 censored at the 7th, 15th, and 20th respectively are shown. The parameters used for the Monte Carlo experimentation are 3, 4, 5, and 6 for the shape parameter, 5 and 10 for the scale parameter, and 10 for the location parameter.

3. MINIMUM DISTANCE ESTIMATION

In this section of the paper, we find estimators of the parameters γ, β, δ. These estimators are the minimum distance estimators that minimize a goodness of fit statistic. This goodness of fit statistic is taken as the Cramer von Mises statistic W^2 which measures the integral of the squared difference between the density and the sample empirical distribution function. This W^2 is defined as:

$$W^2 = n\int_{-\infty}^{\infty}\left[F_o(x) - \hat{F}(x)\right]^2 dF_o(x)$$

where $\hat{F}(x)$ is the sample empirical distribution function and $F_0(x)$ is a completely specified distribution function. The corresponding computational form is:

$$W^2 = \sum_{i=1}^{n}\left[F_0(x_{(i)}) - \frac{i-0.5}{n}\right]^2 + \frac{1}{12n}$$

This computational form uses the step function $\frac{i-0.5}{n}$ as an estimator for $\hat{F}(x)$. The basic notion in this section of the paper is to implement the concept of nonparametric density estimation to replace the step function representing the sample empirical distribution function. Of course to do that it is needed to define a kernel and the parameter to be used with that kernel. In our case an adapted Gaussian kernel together with a heuristic or empirical choice for the window width are introduced. First, the definition of the new adapted Gaussian kernel was driven by how to benefit from the fact that the sample is right censored sample. Also, the definition takes care of that the sample is an ordered sample. This adapted kernel takes the form:

$$K(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\phi(\tau)} e^{-\frac{1}{2}x^2} & -\infty < x < \tau \\ 0 & x \geq \tau \end{cases}$$

where τ determines a threshold from the right that gives a zero weight to the x-values beyond that τ and $\phi(\tau)$ is the C.D.F of the standard normal distribution. This τ value will be used to compensate the ordering of the sample in which case it will consider the information that $X_{(i)} \leq X_{(i+1)}$ for all $1 \leq i \leq r$ with r as the right censoring limit. This can simply be shown from the way the kernel or the bump is placed over each observation. This kernel is placed over each observation such that a zero weight (mass) is given for observation $X_{(i)}$ at and beyond $X_{(i+1)}$ for all observations other than $X_{(r)}$.

While for $X_{(r)}$, the threshold is arbitrary chosen to be at multiples of $X_{(r)}$ (taken at 5 multiples of $X_{(r)}$ in our case).

Thus, the kernel at order statistic $X_{(i)}$ will be :

$$K_{(i)}(x) = \begin{cases} \dfrac{1}{h\sqrt{2\pi}\phi\left(X_{(i+1)}\right)} e^{-\frac{1}{2}\left(\frac{x-X_{(i)}}{h}\right)^2} & -\infty < x < X_{(i+1)} \\ 0 & x \geq X_{(i+1)} \end{cases}$$

for $1 \leq i \leq r$-1, and

$$K_{(i)}(x) = \begin{cases} \dfrac{1}{h\sqrt{2\pi}\phi\left(5X_{(r)}\right)} e^{-\frac{1}{2}\left(\frac{x-X_{(Ri)}}{h}\right)^2} & -\infty < x < 5X_{(r)} \\ 0 & x \geq 5X_{(r)} \end{cases}$$

for $i = r$.

Second, the optimal value of the window width h (in the MISE sense) depends on the choice of the kernel K, the underlying unknown density f(x) and the sample size i.e.

$$h_{opt} = f_1(K).f_2(f(x)).f_3(n)$$

with explicit expression for h_{opt} given as:

$$h_{opt} = m_2^{-2/5}\left\{\int K^2(t)dt\right\}^{1/5}\left\{\int f''^2(x)dx\right\}^{-1/5} n^{-1/5}$$

where m_2 denotes the kernel second moment. A reasonable approximation for this optimal value for basically a normal sample was suggested to be $h = kn^{\frac{-1}{5}}$ where k is a real constant. Although this approximation simplifies the optimal expression for the window width and works fine with the normal distribution it is not as good for other distributions. An alternative for computing the window width that is more efficient computationally and gives a good improvement in this application is to choose an empirical h which equals c$sr^{-\frac{1}{5}}$ where s represents the data driven parameter from the censored sample and r represents the censored sample size. This s is equal to $\beta_g\sqrt{\Gamma\left(\dfrac{\delta_g+2}{\delta_g+1}\right)-\Gamma^2\left(\dfrac{\delta_g+1}{\delta_g}\right)}$.The (β_g, δ_g) are initial guess for both the scale and shape parameters of the Weibull density. These are chosen as scaled sample standard deviation of the censored sample with scale 4.0 and 3.0 for both values respectively. Suggested h together with the adapted kernel showed an improvement in MISE besides being simple, without a need for extensive computations.

The following figure (Fig. 1.) shows an example for the use of this new non-parametric density with the introduced kernel and the chosen window width.

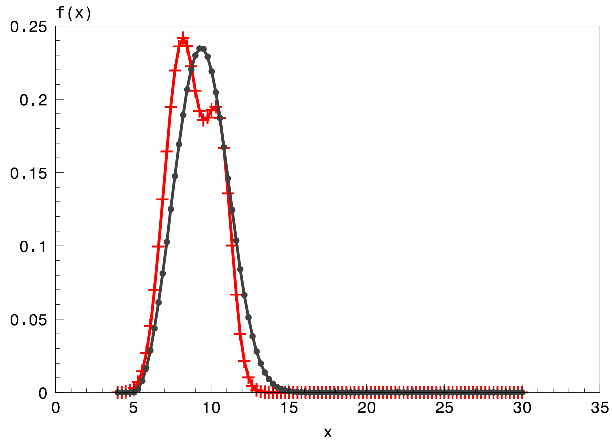

Fig 1. Example for use of adapted kernel
True density Estimated density

The sample used in the example is of size 20 censored at the 15th ordered observation and is from a Weibull distribution with location parameter 5, scale parameter 5, and shape parameter 3. The 15 ordered statistics are 6.957001, 7.009424, 7.609188, 7.930433, 8.007868, 8.371517, 8.414957, 8.495595, 9.018893, 9.484709, 10.187010, 10.485670, 10.487470, 10.692950, and 10.794190. The data driven window width in this case is h=0.7654

4. METHODOLGY

Both results from MLE computations and the new technique are shown in the following tables (Table 1, Table 2, and Table 3).

Table 1. Results from M.C size 1000 for type II Right Censored sample of size 7 out of 10

Weibull (loc., sca., sha.)	$MISE_{CvM}$	$MISE_{MLE}$
W(10,5,3) **3**	0.04152560 *(0.0719672)*	0.21541840 *(0.12253522)*
W(10,5,4) **8**	0.04926788 *(0.05453460)*	0.26278644 *(0.11724960)*
W(10,5,5) **6**	0.05344808 *(0.06940708)*	0.28162700 *(0.13792380)*
W(10,5,6) **5**	0.05973878 *(0.10077130)*	0.27194753 *(0.16992954)*
w(10,10,3) **2**	0.02534832 *(0.02905686)*	0.03724256 *(0.05392057)*
W(10,10,4) **1**	0.01101317 *(0.02811573)*	0.05318439 *(0.07417205)*
W(10,10,5) **2**	0.00731237 *(0.02531788)*	0.09116063 *(0.09960468)*
W(10,10,6) **5**	0.00676097 *(0.02890469)*	0.12782822 *(0.11734741)*

Table 2. Results from M.C size 1000 for type II Right Censored sample of size 15 out of 20

Weibull (loc., sca., sha.)	$MISE_{CvM}$	$MISE_{MLE}$
W(10,5,3) **6**	0.00946977 *(0.01492824)*	0.08318688 *(0.07245930)*
W(10,5,4) **4**	0.034545593 *(0.01713189)*	0.05111313 *(0.0691607)*
W(10,5,5) **2**	0.04582739 *(0.03089365)*	0.07171360 *(0.13317342)*
W(10,5,6) **1**	0.05719306 *(0.04872664)*	0.1386413 *(0.19690372)*
w(10,10,3)	0.01679135 *(0.00664824)*	0.01799473 *(0.01731703)*
W(10,10,4)	0.00524905 *(0.00338150)*	0.01699780 *(0.02617111)*
W(10,10,5)	0.0165833 *(0.00234119)*	0.01724535 *(0.02328303)*
W(10,10,6)	0.00148641 *(0.00273267)*	0.02007358 *(0.03930784)*

Table 3. Results from M.C size 1000 for type II Right Censored sample of size 20 out of 30

Weibull (loc., sca., sha.)	$MISE_{CvM}$	$MISE_{MLE}$
W(10,5,3) **3**	0.01474420 *(0.02094048)*	0.09620322 *(0.08285221)*
W(10,5,4) **7**	0.01820016 *(0.03771030)*	0.04377227 *(0.05551330)*
W(10,5,5) **2**	0.01927658 *(0.04633982)*	0.03095871 *(0.06553324)*
W(10,5,6) **19**	0.01740239 *(0.04270043)*	0.04298951 *(0.12305474)*
w(10,10,3)	0.00815912 *(0.01044164)*	0.01606623 *(0.01381363)*
W(10,10,4)	0.00566419 *(0.00830887)*	0.01821797 *(0.02310432)*
W(10,10,5)	0.00350789 *(0.00604578)*	0.0173937 *(0.01944109)*
W(10,10,6)	0.00198457 *(0.00404555)*	0.01908424 *(0.03421304)*

The tables show the resulting MISE together with its standard deviation between brackets for samples of size 10, 20, and 30 censored at the 7th , 15th , and 20th order statistic for different parameter values for both the new proposed estimator concurrently with the modified nonlinear method for solving the maximum likelihood equations. In addition, the tables show that the new technique has a significant improvement over the MLE method for shape parameters 3, 4, 5, and 6. A quick look at the results from table 2 , for example, without overgeneralizing conclusions depicts a better MISE and smaller standard deviation for the proposed method.

Thus, the new technique shows a significant improvement over the MLE method. The improvement in MISE ranges from close but yet smaller value of MISE to almost 13.5 times smaller in case of location 10, scale 10, and shape 6. The variations in h together with the corresponding variations in MISE indicate that the method is an adaptive one in the sense that the choice of the parameter h that is data dependent varies with the variation of the distribution parameters and the sample size.
The final conclusion is that the previously described method is recommended for use as an alternative to the MLE method for estimating the parameters of the Weibull distribution based on right censored samples for up to sample sizes 30.

REFERENCES

1. Cohen, A. C. and Whitten, B. J., Estimation in the three parameter Weibull distribution *Technometrics*, **17**, 347-351, 1980.
2. Cohen, A. C. and Ding, Y., Modified moment estimation for the three parameter Weibull distribution *J. Qual. Tech.*, **16**, 159-167, 1984.
3. Cohen, A. C., Maximum likelihood estimation in the Weibull distribution based on complete and censored samples. *Technometrics*, **7**, 579-588, 1965.
4. Cohen, A. C., Multicensored sampling in the three -parameter Weibull distribution. *Technometrics*, **17**, 347-351, 1975.
5. Harter, H. L. and Moore, A. H., Maximum likelihood estimation of the parameters of gamma and Weibull population from complete and censored samples. *Technometrics*, **7**, 639-643, 1965.
6. Harter, H. L. and Moore, A. H., Asymptotic variances and covariances of maximum likelihood estimators from censored samples of the parameters of Weibull and gamma distribution. *Ann. Math. Stat.*, **38**, 557-570, 1967.
7. Ortega, J. M.; Rheinboldt, W. C., 1970. *Iterative Solution of Nonlinear equations in Several Variables*. Academic Press. New York and London.
8. Powell, M. J. D., 1970. *"A hybrid method for nonlinear equations," Numerical methods for Non linear Algebraic equations*, P. Rabinwitz editor.
8. Wycoff, J. , Bain, L., and Englehardt, M., Some complete censored sampling results for the three parameter Weibull distribution *J. Statist. Comp. Simul.*, **II**, 139-151, 1980.