EIMJ
المجلة الإلكترونية الشاملة متعددة التخصصات
العدد الثالث والأربعون (شهر 12) 2021

Monti-Carlo Simulation to Investigate Estimating the performance of Bayesian Parametric Methods for the Survivor Function in Censored Data

[1]Fakhereldeen E. E. Musa, [2]Alshaikh.A.Shokeralla, [3]Salem Al Zahrani, [4] Hussam M. Gubara

[1] Department of Statistics, Faculty of Mathematical Sciences and Statistics, Al-Neelain University, Khartoum, Sudan. Email: fakhryalhaj@gmail.com.

[2]Department of Mathematics, Faculty of Sciences and Art, ALBaha University, AlBaha, KSA
Email: sshokeralla@gmail.com

[3]Department of Mathematics, Faculty of Sciences and Art, ALBaha University, AlBaha, KSA
Email: salem.bb@hotmail.com

[4]Department of Mathematics, Faculty of Mathematical Sciences and Statistics, AL-Neelain University, Khartoum, Sudan
Email: hussam.m.a86@hotmail.com

Abstract:

The aim of this paper is to investigate the performance of some parametric Bayesian estimators of the survivor function with respect to bias and efficiency when data contain censoring. The study is primarily based on Mote Carlo experiments. The methods used were: the likelihood method, Bayesian with exponential prior and Bayesian with gamma prior. It is shown that the Bayesian method with gamma prior has the best performance.

Keywords: Bayesian method, parametric, Monte Carlo methods, survivor function, censoring, Likelihood.

1. Introduction:

The survivor function, denoted by $S(t)$ is the probability that an individual survives at least up to time t . If T is a random variable representing the survival time of an individual, then $S(t)$ is defined formally as:

$$S(t) = Prob(T > t)$$

Several parametric and nonparametric estimation methods are suggested in the literature for the estimation of $S(t)$. In this paper our discussion will cover Bayesian and non-Bayesian methods. These will be considered in case of censored data. Analysis of survival data when survival times follow a known distribution is also extensively investigated through Bayesian and non – Bayesian approaches. Sinha (1986) determined the Bayes estimates of the reliability function and the hazard rate of the Weibull failure time distribution by employing squared error loss function. Abdel-Wahid and Winterbottom (1987) studied the approximate Bayesian estimates for the Weibull reliability function and hazard rate from censored data by employing a new method that has the potential of reducing the number of terms in Lindley procedure. Omari and Ibrahim (2011) conducted a study on Bayesian survival estimator for Weibull distribution with censored data using squared error loss function with Jeffreys prior amongst others. Noortwijkl and Gelder (2000) applied Bayesian estimation of quantiles for the purpose of flood prevention assuming sea water levels exponentially distribution with unknown value of the mean. Linear loss, squared-error loss, and modified linex loss are three types of loss functions are considered.

Guure, and Ibrahim (2012) compared the classical maximum likelihood against the Bayesian estimators using an informative prior and a proposed data-dependent prior known as generalized non informative prior. The Bayesian estimation was considered under three loss functions. A simulation was conducted under different sample sizes and the mean squared error, absolute bias are used for comparison. Sulistianingsih et al. (2017) used Bayesian estimation method under Linex Loss function for Survival model followed an exponential distribution and they considered Gamma distribution as prior and likelihood function produces a gamma distribution as posterior distribution. Calabria and Pulcini (1996) derived Bayes estimates of the parameters and functions thereof in the left-truncated exponential distribution. Both the non-informative prior and an informative prior on the reliability level at a prefixed time value are considered. The statistical performance of the Bayes estimates is compared to those of the maximum likelihood ones through the risk function. Palacio and Leisen (2018) focus for estimating multivariate survival functions. Their model extends the work of Epifani and Lijoi (2010) to an arbitrary dimension and allows to model the dependence among survival times of different groups of observations. Theoretical results about the posterior behavior of the underlying dependent vector of completely random measures are provided. The performance of the model is tested on a simulated dataset arising from a distributional Clayton copula. Guure et al. (2012) applied Bayesian estimation, for the two-parameter Weibull distribution using extension of Jeffreys' prior information with three loss functions, and Syuan-Rong and Shuo-Jye (2011) considered Bayesian estimation and prediction for Weibull model with progressive censoring.

The paper aimed to find the best estimator of survival function. Maximum likelihood method, Bayesian method with exponential as prior distribution and Bayesian method with gamma as prior distribution are used for estimation. The comparison between these estimation methods is done through a Monte-Carlo Simulation with different sample sizes.

2. Methods of estimation for censored data:

Suppose the survivor data, in sample of size n, is such that times t_1, t_2, ..., t_r are uncensored and time t^*_1, t^*_2, ..., t^*_{n-r} are censored

The likelihood function is:

$$L(\lambda) = \prod_{i=1}^{r} f(t_i) \prod_{i=1}^{n-r} s(t^*_i)$$

Letting

$$\delta_i = \{0 \;\; if \;\; t_i \; is\, censored \qquad 1 \; if \;\; not$$

we may write the likelihood function as

$$\prod_{i=1}^{r} \{f(t_i)\}^{\delta_i} \left\{S(t^*_i)\right\}^{1-\delta_i}$$

If

$$f(t) = \lambda e^{-\lambda t}$$

We know that

$$S(t) = e^{-\lambda t}$$

In this case the likelihood function becomes

$$L(\lambda) = \prod_{i=1}^{r} \left[\lambda e^{-\lambda t_i}\right]^{\delta_i} \left[e^{-\lambda t_i}\right]^{1-\delta_i} = \prod_{i=1}^{r} \lambda^{\delta_i} e^{-\lambda t_i} \qquad (1)$$

2.1 Likelihood Estimator:

To obtain the maximum likelihood estimator of $S(t)$ we first estimate λ. Taking logs of both sides of (5.8) we get:

$$L(\lambda) = \sum_{i}^{r} \delta_i log\lambda - \lambda \sum_{i}^{r} t_i = rlog\lambda - \lambda \sum_{i}^{r} t_i$$

since only r of the δ are nonzero.

Taking partial derivatives with respect to λ and equating to zero we get as a maximum likelihood estimator of λ

$$\hat{\lambda}_4 = \frac{r}{\sum_{i=1}^{r} t_i} \qquad (2)$$

so that

$$\hat{S}_4(t) = exp\left[- \frac{r}{\sum_{i=1}^{r} t_i} t\right] \qquad (3)$$

is the maximum likelihood estimator of the survivor function when the data is censored.

2.2 Bayesian Estimator with Exponential Prior:

If the parameter λ of the exponential distribution is assumed $Exp(\lambda')$ we see that the posterior distribution of λ is proportional to the product of the likelihood function and the prior i.e.

$$Posterior \propto \lambda^r e^{-\lambda \sum_{i=1}^{n} t_i} \lambda' e^{-\lambda\lambda'}$$

$$\propto \lambda^r e^{-\lambda(\lambda' + \sum_{i=1}^{n} t_i)}$$

This is $gamma(r + 1, \lambda' + \sum_{i=1}^{n} t_i)$ which has a mean

$$\lambda_4 = \frac{r+1}{\lambda' + \sum_{i=1}^{r} t_i} \qquad (4)$$

$$\hat{\lambda}' = \frac{m}{\sum_{i}^{n} \hat{\mu}_i}$$

substituted this we get as the Bayesian estimator of λ

$$\hat{\lambda}_5 = \frac{r+1}{\sum_{i=1}^{n} t_i + \frac{m}{\sum_{i}^{n} \hat{\mu}_i}} \qquad (5)$$

so that

$$\hat{S}_5(t) = exp \qquad (6)$$

is the estimator in this case.

2.3 Bayesian estimator with gamma prior:

If λ in $Exp(\lambda)$ is assumed to be $gamma\ (\alpha, \beta)$ we can see that the posterior distribution of λ is $gamma\ (r + \alpha, \beta + \sum_{i}^{n} t_i)$. The estimates $\hat{\alpha}$ and $\hat{\beta}$ of section (2) will serve as moment estimates $\alpha\ and\ \beta$. Hence the Bayesian estimator of λ is:

$$\hat{\lambda}_6 = \frac{r+\hat{\alpha}}{\hat{\beta} + \sum_{i=1}^{n} t_i} \qquad (7)$$

and

$$\hat{S}_6(t) = exp\left[- \frac{r+\hat{\alpha}}{\hat{\beta} + \sum_{i=1}^{n} t_i}\, t\right] \qquad (8)$$

is the estimator of the survivors function.

3. Simulation Experiments:

To examine the performance of the models presented in this study, we carry out a number of Monte Carlo experiments.

1- Generate a population of survival times, part of it censored; from an exponential distribution $Exp(\lambda)$ with know λ .

3.1 Experiment (I): Likelihood estimation:

1. Take a sample of size n and let the number of uncensored times in the sample be r and the censored $(n - r)$. Denoted this sample by i.
2. Calculate $\hat{\lambda}_4$ using (2) and denoted by $\hat{\lambda}_{4i}$
3. Calculate $\hat{S}_4(t)$ using (3) and denoted it by $\hat{S}_{4i}(t)$
4. Calculate the residuals

$$r_{ij} = \hat{s}_{4i}\left(t_j\right) - s_i\left(t_j\right) \quad , \quad j = 1, \dots, n$$

where $\hat{s}_{4i}\left(t_j\right)$ is the value of the estimate of the survivor function (using (2)) at the j the survival time in sample i and $s_i\left(t_j\right)$ is the value of the true survival function (for the censored data) at survival time j in sample i.

5. Calculate the mean square error for sample i

$$MSE_{4i} = \frac{\sum_j^n r_{ij}^{\ 2}}{n}$$

6. Repeats steps 1-5 $for\ i = 1, \dots, N$ (very large).
7. Calculate the average mean error for $\hat{s}_4(t)$

$$MSE_4 = \frac{\sum_j^N MSE_{4i}}{N}$$

and the variance

$$V_4 = \frac{\sum_{i}^{N}(MSE_{4i} - MSE_4)^2}{N}$$

3.2 Experiment (II): Bayesian Estimation with exponential prior:

1. Select m samples each of size n and calculates their means $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_m$.
2. Calculate

$$\bar{\hat{\mu}} = \frac{\sum_{i}^{n} \hat{\mu}_i}{m}$$

3. Take a sample of size n and obtain

$$\hat{\lambda}_{5i} = \frac{r+1}{\sum_{j=1}^{n} t_j + \frac{m}{\sum_{j}^{n} \hat{\mu}_j}} \quad , \qquad j = 1, \ldots, m$$

where the i in $\hat{\lambda}_{5i}$ refers to the order of the sample

4. Calculate $\hat{S}_5(t)$ using (5) and denoted by $\hat{S}_{5i}(t)$
5. Calculate the residual

$$r_{ij} = \hat{S}_{5i}(t_j) - S_i(t_j) \quad , \quad j = 1, \ldots, n$$

where $\hat{s}_{5i}(t_j)$ is the value of the estimate of the survivor function (using (6)) at the j the survival time in sample i and $S_i(t_j)$ is the value of the true survival function (for the censored data) at survival time j in sample i

6. Calculate the mean square error for sample i

$$MSE_{6i} = \frac{\sum_{j}^{n} r_{ij}^{\ 2}}{n}$$

7. Repeats steps $1 - 5\ for\ i = 1, \ldots, N$ (very large).
8. Calculate the average mean error for $\hat{S}_1(t)$

$$MSE_5 = \frac{\sum_j^N MSE_{5i}}{N}$$

and the variance

$$V_5 = \frac{\sum_i^N (MSE_{5i} - MSE_5)^2}{N}$$

3.3 Experiment (III): Bayesian Estimation with gamma prior:

1. Take m samples each of size n and calculate their means $\frac{1}{\hat{\mu}_1}, \frac{1}{\hat{\mu}_2}, \ldots, \frac{1}{\hat{\mu}_m}$.
2. Calculate

$$\bar{\hat{\mu}} = \frac{\sum_i^n \frac{1}{\hat{\mu}_i}}{m}$$

and

$$\hat{\sigma}^2 = \frac{\sum_i^n (\frac{1}{\hat{\mu}_i} - \bar{\hat{\mu}})^2}{m}$$

3. Obtain the moment estimators $\hat{\alpha}\ and\ \hat{\beta}$ of $\alpha\ and\ \beta$ respectively in section (2).
4. Take a sample of size n and obtain

$$\hat{\lambda}_{6i} = \frac{n + \hat{\alpha}}{\sum_{j=1}^{n} t_j + \hat{\beta}} \quad , \qquad j = 1, \ldots, n$$

where the i in $\hat{\lambda}_{3i}$ refers to the order of the sample

5. Calculate $\hat{S}_{6i}(t)$ using (7) and denoted by $\hat{S}_{6i}(t)$
6. Calculate the residuals

$$r_{ij} = \hat{S}_{6i}\left(t_j\right) - S_i\left(t_j\right) \quad , \quad j = 1, \dots, n$$

where $\hat{S}_{3i}\left(t_j\right)$ is the value of the estimate of the survivor function (using (8)) at the j the survival time in sample i and $S_i\left(t_j\right)$ is the value of the true survival function (for the censored data) at survival time j in sample i

7. Calculate the mean square error for sample i

$$MSE_{6i} = \frac{\sum_{j}^{n} {r_{ij}}^2}{n}$$

8. Repeats steps 1-7 $for\ i = 1, \dots, N$ (very large).
9. Calculate the average mean error for $\hat{S}_6(t)$

$$MSE_6 = \frac{\sum_{j}^{N} MSE_{6i}}{N}$$

and the variance

$$V_6 = \frac{\sum_{i}^{N} \left(MSE_{6i} - MSE_6\right)^2}{N}$$

4. Result Monte - Carlo simulation and Discussion:

This section compares, through a simulation experiment the performance of likelihood estimators and Bayesian estimators in censored data. Maximum likelihood estimators are considered for under the exponential distribution for both types of data. On the other hand the Bayesian estimators are investigated for both Gamma prior and Exponential prior.

Discussion is focused on performance of the methods with respect to bias and efficiency under various sample size.

For all cases an exponentially distributed population of size 10000 is simulated contains 10% censer. Sample sizes 10,30,50,75 and 100 are selected, with 1000 repetitions for each. For stability purposes 10 iteration are performed.

Table (1), Table (2) and Table (3) show estimators of the population parameter as well as the mean square and bias of the estimates of the survival function using likelihood and Bayesian methods with exponential prior and gamma prior respectively when the population contain censoring. Based on practical experience data is generated that contains 10% survival times.

4.1 Likelihood Estimation:

Table (1) gives the estimate of exponential parameter (the true of which is 55), the mean of the estimate, as well as the bias for various sample size. The table also gives the bias in the estimates of the survivor function and their mean squares and variance.

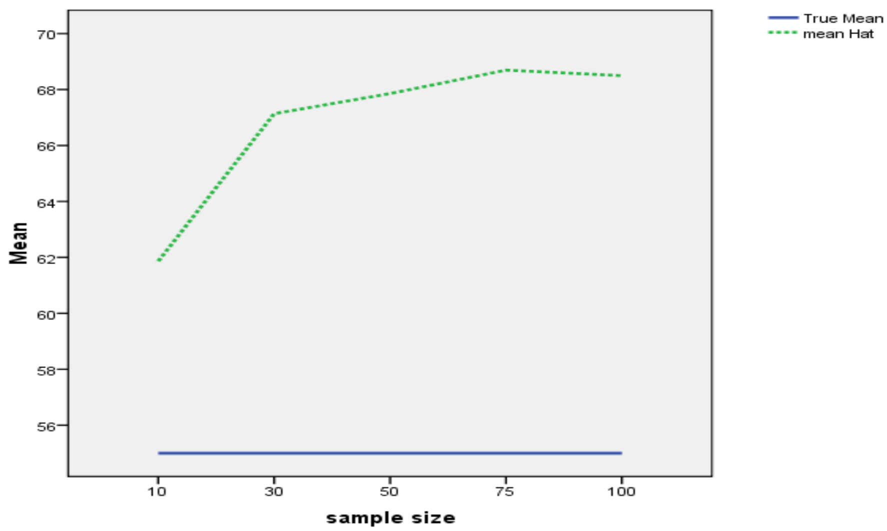
Table (1): Bias in the estimates of the exponential parameter and survivor function together with the MSE of the survivor function using maximum likelihood method:

N	$\hat{\lambda}$	$Mean\,\frac{1}{\hat{\lambda}}$	Bias	Bias_st	MSE	VAR
10	0.016167	61.86401	-6.86401	0.073044	0.01211	0.00027
30	0.01491	67.13832	-12.1383	0.096797	0.006739	6.05E-05
50	0.014748	67.85796	-12.858	0.064518	0.0055	3.1E-05
75	0.014567	68.69542	-13.6954	0.05717	0.005053	1.83E-05
100	0.014616	68.49616	-13.4962	0.049119	0.004643	1.35E-05

Source: Authors Prepared, 2021

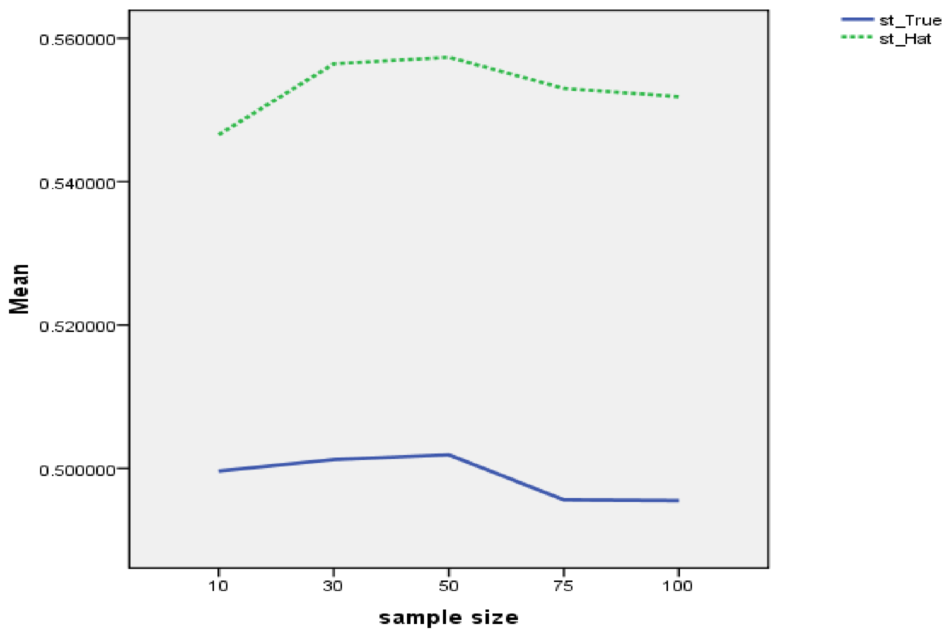
Unlike the case in uncensored data we see from the Table (1) that the bias in the estimates of the parameters increases slightly in general for larger sample size. However bias and mean square of the survivor function appears to be constant with increase in sample size.

Figure (1) and Figure (2) confirms this result for the exponential and the survivor function.



Source: Authors Prepared, 2021

Fig (1)



Source: Authors Prepared, 2021

Fig (2)

4.2 Bayesian Estimation with Exponential Distribution Prior:

Table (2) gives the estimate of exponential parameter (the true of which is 55) , the mean of the estimate , as well as the bias for various sample size . The table also gives the bias in the estimates of the survivor function and their mean squares and variance.

Table (2): Bias in the estimates of the exponential parameter and survivor function together with the MSE of the survivor function using Bayesian estimation method:

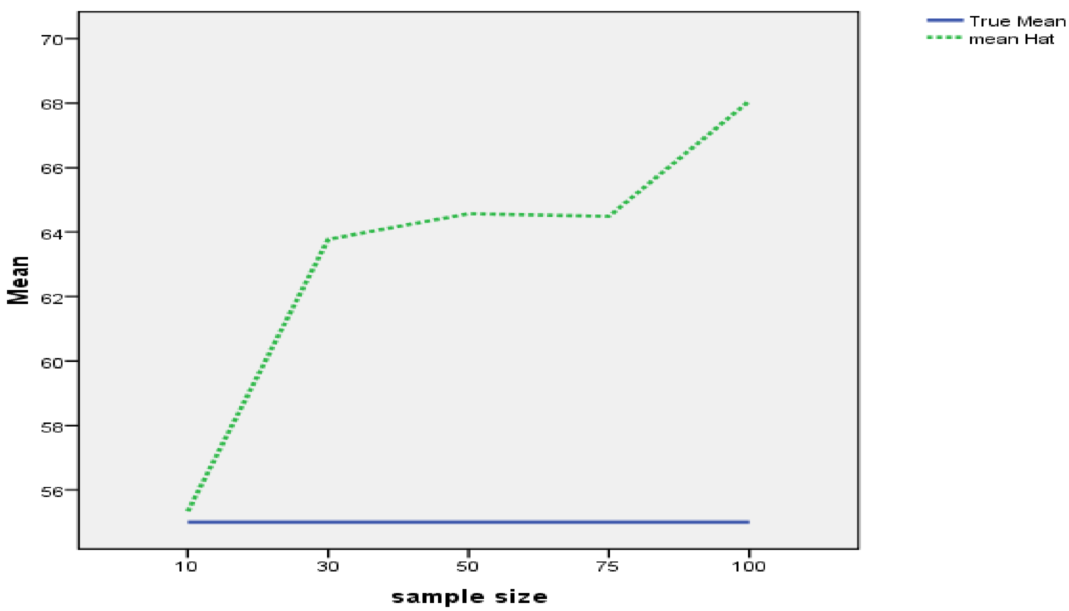
N	$\hat{\lambda}$	$Mean\,\frac{1}{\hat{\lambda}}$	Bias	Bias_st	MSE	VAR
10	0.018104	55.33814	-0.33814	0.111241	0.009136	0.00016
30	0.015712	63.77874	-8.77874	0.054745	0.004969	3.47E-05
50	0.015501	64.57379	-9.57379	0.042111	0.004047	2.3E-05
75	0.015564	64.49244	-9.49244	0.034773	0.003312	1.17E-05
100	0.01471	68.07073	-13.0707	0.062029	0.00441	1.28E-05

Source: Authors Prepared, 2021

It is obvious from Table (2) that the bias in the estimates of the parameters increases with increase in sample size (Fig (3)).

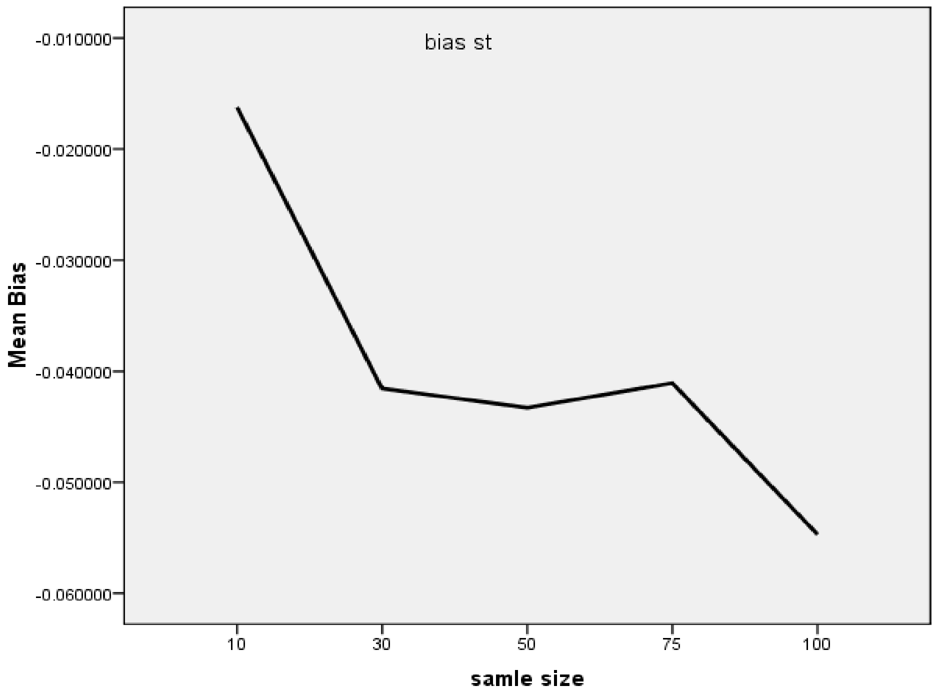
As for the survivor function, we see that its bias decrease with increase in sample size with the exception of sample size 100.

The same applies to the mean square (MSE) and variance (VAR) meaning that the efficiency of the estimate increases with sample size.



Source: Authors Prepared, 2021

Fig (3)



Source: Authors Prepared, 2021

Fig (4)

4.3 Bayesian Estimation with Gamma Distribution Prior:

Table (3) gives the estimate of exponential parameter (the true of which is 55), the mean of the estimate, as well as the bias for various sample size. The Table (3) also gives the bias in the estimates of the survivor function and their mean squares and variance.

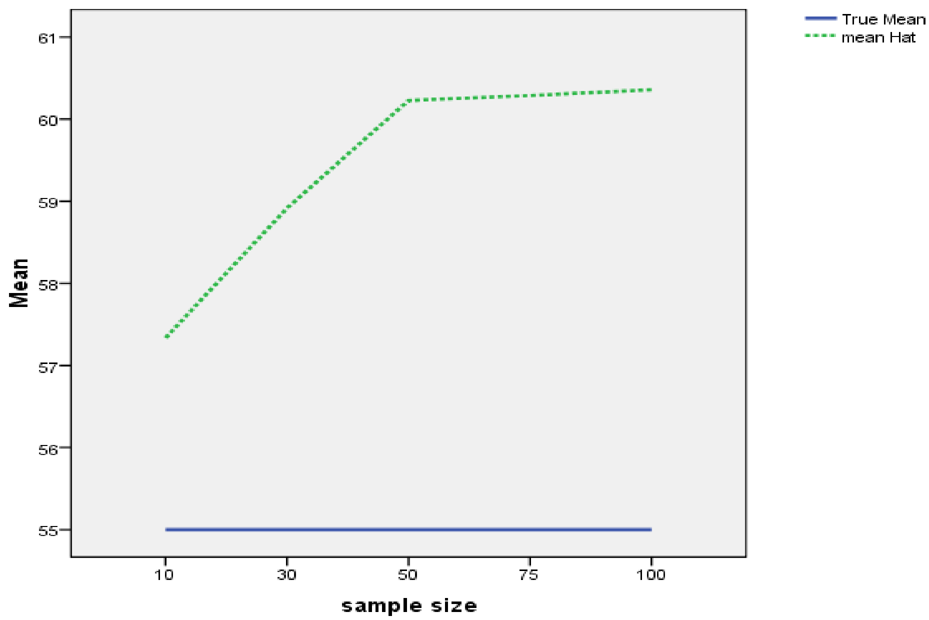
Table (3): Bias in the estimates of the exponential parameter and survivor function together with the MSE of the survivor function using Bayesian estimation method:

N	$\hat{\lambda}$	$Mean\,\frac{1}{\hat{\lambda}}$	Bias	Bias_st	MSE	VAR
10	0.017494	57.33554	-2.33554	0.035228	0.003371	2.72E-05
30	0.016992	58.91652	-3.91652	0.032418	0.001188	2.95E-06
50	0.016618	60.22775	-5.22775	0.023866	0.001102	1.75E-06

75	0.0166	60.28875	-5.28875	0.02479	0.000957	9.67E-07
100	0.016581	60.35823	-5.35823	0.02006	0.00093	8.24E-07

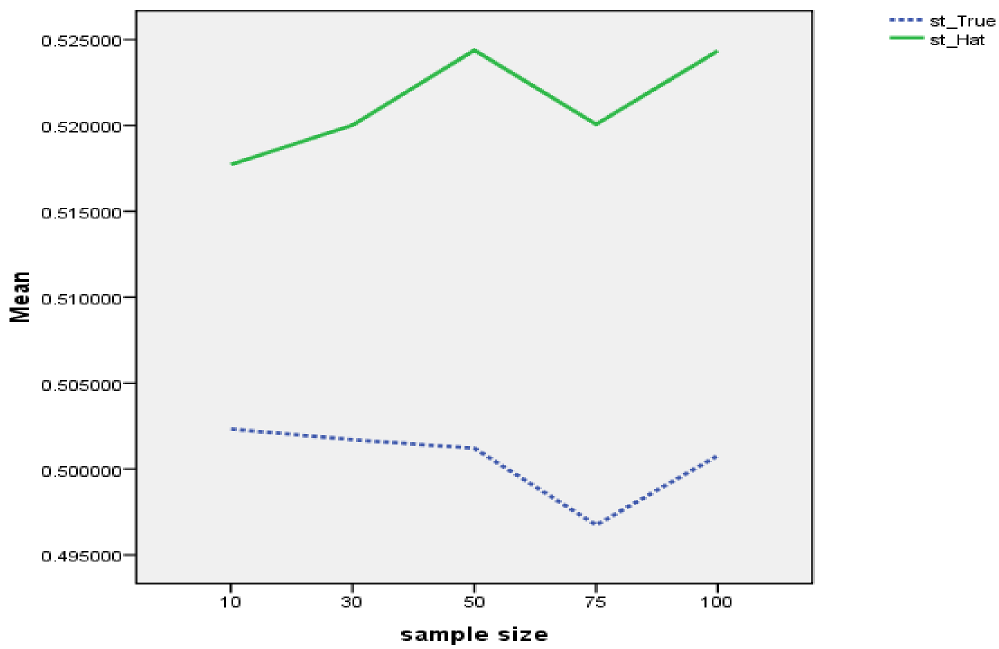
Source: Authors Prepared, 2021

Table (3) shows that the bias in the estimates of the parameters increases in sample size Fig (5) where that of survivor function decreases. The same applies to the mean square (MSE) and variance (VAR) meaning that the efficiency of the estimate increases with sample size.



Source: Authors Prepared, 2021

Fig (6)



Source: Authors Prepared, 2021

Fig (7)

4.5 All method of estimation:

Table (4) below shows bias and MSE of estimates of the survivor function using LES, BEP and BGP method. It is obvious from Table (4) that the bias in the estimates of the parameters increases with increase in sample size. As for the survivor function, we see that its bias decrease with increase in sample size. The mean square (MSE) meaning that the efficiency of the estimate increases with sample size.

Table (4): Bias in the estimates of the exponential parameter and survivor function together with the MSE of the survivor function using LES, BEP and BGP method:

Sample Size	Bias			MSE		
	LES	BEP	BGP	LES	BEP	BGP
10	0.073044	0.111241	0.035228	0.01211	0.009136	0.003371
30	0.096797	0.054745	0.032418	0.006739	0.004969	0.001188
50	0.064518	0.042111	0.023866	0.0055	0.004047	0.001102

75	0.05717	0.034773	0.02479	0.005053	0.003312	0.000957
100	0.049119	0.062029	0.02006	0.004643	0.00441	0.00093

Source: Authors Prepared, 2021

Conclusions:

We conclude in this study by observing that unlike the case in uncensored data, the bias of the estimates of the parameters by all three methods increases with the sample size when survival data is censored.

the bias of the estimates of the survivor function decreases with sample size while its efficiency increases. This so is for all three methods. When the three methods are compared with respect to bias of parameter, we see that the worst performance is that of the likelihood method which gives relatively very large bias.

The Bayesian estimator with gamma prior provided less bias than the Bayesian estimator with exponential bias. As for the survivor function we see that the Bayesian method with gamma prior gave the best performance with respect to both bias and efficiency followed by Bayesian estimator with exponential prior.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflicts of interest

The authors declare no conflicts of interest regarding the publication of this paper.

References:

1. Abdel-Wahid, A., A. and Winterbottom, A. (1987). Approximate Bayesian estimates for the Weibull reliability function and hazard rate from censored data. *Journal of Statistical Planning* and Inference, vol. 16, no. 3, pp. 277–283.
2. Palacio, A., R. and Leisen, F. (2018). Bayesian nonparametric estimation of survival functions with multiple-samples information. Electronic Journal of Statistics, vol.12:1330 –1357.
3. Calabria, R. and Pulcini, G. (1996). Point estimation under asymmetric loss functions for left-truncated exponential samples. Communications in Statistics - Theory and Methods, 25(3), 585-600.
4. Guure, C. B. and Ibrahim, N. A. (2012). Bayesian Analysis of the Survival Function and Failure Rate of Weibull Distribution with Censored Data. Mathematical Problems in Engineering, Volume 2012, Article ID 329489, doi:10.1155/2012/329489
5. Epifani, I. and Lijoi. A. (2010). Nonparametric priors for vectors of survival functions. Statistical Sinica, 20, 1455–1484.
6. Guure, C. B. and Ibrahim, N. A. and Omari, A. M. (2012). Bayesian estimation of two-parameter Weibull distribution using extension of Jeffreys' prior information with three loss functions," Mathematical Problems in Engineering. ISSN: 1935-7524.
7. Omari, M. A. Al and Ibrahim, N.A. (2011). Bayesian survival estimation for Weibull distribution with censored data. Journal of Applied Sciences, vol. 11, pp. 393–396.
8. Noortwijk, J.M. and Gelder, P. H. A. J. M. (2000). Bayesian Estimation of Quantiles for the Purpose of Flood Prevention. Conference: 26th International Conference on Coastal Engineering. DOI:10.1061/9780784404119.268,1 – 14.
9. Rizki, S.W, Mara, M.N., and Sulistianingsih, E. (2017). Survival Bayesian Estimation of Exponential-Gamma under Linex Loss Function. International Conference on Mathematics: Education, Theory and Application. IOP Conf. Series: Journal of Physics: Conf. Series 855 - 012036.
10. Sinha, S. K. (1986). Bayes estimation of the reliability function and hazard rate of a Weibull failure time distribution. Trinbago's De Statistical, vol. 1, pp. 47–56.

11. Syuan-Rong, H. and Shuo-Jye, W. (2011). Bayesian estimation and prediction for Weibull model with progressive censoring. Journal of Statistical Computation and Simulation, pp.1–14.