

Performance of parametric Bayesian Methods for estimating the survivor function in uncensored data using Monte-Carlo simulation

Mohammed Elamin Hassan¹, Fakhreldeen Elhaj Esmial Musa²

¹Assistant Professor, Applied Statistics, Sudan Academy of Banking and Financial Sciences, Khartoum, Sudan

²Assistant Professor, Statistics, -Al-Neelain University, Faculty of Mathematical Sciences and Statistics, Department of Statistics, Khartoum, Sudan

¹aminhassan1976@gmail.com, ²fakhryalhaj@gmail.com

Received: 28/8/2021

Revised: 24/10/2021

Accepted: 17/11/2021

DOI: <https://doi.org/10.31559/GJEB2021.11.3.7>



This file is licensed under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)

Performance of parametric Bayesian Methods for estimating the survivor function in uncensored data using Monte-Carlo simulation

Mohammed Elamin Hassan

Assistant Professor, Applied Statistics, Sudan Academy of Banking and Financial Sciences, Khartoum, Sudan
aminhassan1976@gmail.com

Fakhereldeen Elhaj Esmial Musa

Assistant Professor, Statistics, -Al-Neelain University, Faculty of Mathematical Sciences and Statistics, Department of Statistics, Khartoum, Sudan
fakhryalhaj@gmail.com

Received: 28/8/2021 Revised: 24/10/2021 Accepted: 17/11/2021 DOI: <https://doi.org/10.31559/GJEB2021.11.3.7>

Abstract: The paper aimed to investigate the performance of some parametric survivor function estimators based on Bayesian methodology with respect to bias and efficiency. A simulation was conducted based on Monte Carlo experiments with different sample sizes different (10, 30, 50, 75, 100). The bias and variance of mean square Error $V(MSE)$ were selected as the basis of comparison. The methods of estimation used in this study are Maximum Likelihood, Bayesian with exponential as prior distribution and Bayesian with gamma as prior distribution. A Monte Carlo Simulation study showed that the Bayesian method with gamma as prior distribution was the best performance than the other methods. The study recommended that.

Keywords: Bayesian method, parametric; Monte Carlo simulation; survivor function; censoring; Maximum Likelihood method.

1. Introduction

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

$$S(t) = \text{Prob}(T > t)$$

Several parametric and nonparametric estimation methods are suggested in the literature for the estimation of $S(t)$. The paper focuses on the Bayesian and non-Bayesian methods. These will be considered in case of uncensored data. Maximum likelihood estimator (MLE) is quiet efficient and very popular both in literature and practice. Bayesian approach has been employed for estimating parameters. Some research has been done to compare MLE and that of the Bayesian approach in estimating the survival function and the parameters of the Weibull distribution. 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. Noortwijk 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.

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. 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. Palacio and Leisen (2018) focused on estimating multivariate survival functions. Their model extends the work of Epifani and Lijoi (2010) to an arbitrary dimension and allows modeling the dependence among survival times of different groups of observations. 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.

Kutal and Qian (2018) utilized the maximum likelihood method to estimate a non-mixture cure model for right-censored data. A simulation study is based on Fréchet susceptible distribution to evaluate the performance of the method. Thamrin and et-al (2018) used Weibull distribution to model and analyze data on the survival time. They were examined the performance of the Bayesian estimator using conjugate prior information for estimating the parameters of Weibull distribution with censored survival data for dengue fever. Abbas et al. (2020) compared the maximum likelihood and Bayesian estimators to estimate the parameters of Gumbel type-II distribution based on the type-II censored data using the Bayesian framework. Tekindal, Yonar, and Kader (2021) evaluated the performance of parametric method for Left-Censored Data Using Covid-19. They produced uncensored data according to different parameters of each distribution. The most appropriate distributions used for left-censored data in Parametric Inverse Hazard Models were found as Generalized Inverse Weibull as well as Log-Logistic, Log-Normal, Inverse Normal and Gamma distributions.

1.1. Objective of the study

The main objective of this paper is to investigate the effect of some factors on the bias and efficiency, as measured by the mean square error, of certain parametric methods for the estimation of the survivor function. To complete the picture the performance of Bayesian methods is compared to that of some classical methods.

1.2. Importance of the study

By throwing more light on the behavior of the methods under various conditions, it hoped that the outcome of the study will help guide users of the methods in selecting the method that is expected to give relatively the best performance.

The rest of the paper is arranged as follows: Section 2 contains the derivative of the parameters under maximum likelihood estimator and Bayesian estimators approach. This is followed by simulation experiments in Section 3. Results of simulation are in Section 4 and finally the conclusion is in section 5.

2. Methodology and Methods

The paper is based heavily on Monte Carlo experiments with the analytical approach resorted to in certain areas of the research. The exponential distribution and the Weibull distribution are the most widely used to model survival data. The paper focuses on the survival uncensored data which follows an exponential distribution. Let t_1, t_2, \dots, t_n be a sample of n survival times from exponential distribution with parameter λ , i.e. $Exp(\lambda)$. Assume that the actual survival times are there without censored time. The study considered three methods of estimation of, namely:

2.1. Maximum Likelihood Estimation Method

The likelihood function is given by:

$$L(\lambda) = \prod_{i=1}^n f(t_i) = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

$$\text{Log}L(\lambda) = n \text{Log} \lambda - \lambda \sum_{i=1}^n t_i \quad (1)$$

By taking the logarithm and partial derivatives with respect to λ , the maximum likelihood estimator of λ is:

$$\hat{\lambda}_1 = \frac{n}{\sum_{i=1}^n t_i} \quad (2)$$

The distribution function is:

$$F(t) = 1 - e^{-\lambda t} \quad (3)$$

The survivor function is

$$S(t) = 1 - F(t) = e^{-\lambda t} \quad (4)$$

Substituting the estimator $\hat{\lambda}$ for λ we get:

$$\hat{S}_1(t) = \exp \left[- \frac{n}{\sum_{i=1}^n t_i} t \right] \quad (5)$$

2.2. Bayesian Estimator with Exponential Prior

Suppose that the parameter λ of exponential distribution is distributed as $Exp(\lambda')$. In this case the posterior distribution of λ is followed by Gamma distribution with parameters $(n + 1, \sum_{i=1}^n t_i + \lambda')$. The Bayesian estimator of λ can be taken as the mean of this posterior distribution. The mean is given by:

$$\lambda_2 = \frac{n+1}{\sum_{i=1}^n t_i + \lambda'} \quad (6)$$

The parameter λ' of prior distribution can be estimated through m samples of size n which are taken from the distribution $Exp(\lambda)$. The mean of each sample is $\hat{\mu} = \frac{\sum_{i=1}^n t_i}{n}$ and the maximum likelihood estimator of the mean of the exponential distribution $\frac{1}{\lambda}$. If $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_m$ are the mean of m samples, then their mean is:

$$\underline{\hat{\mu}} = \frac{\sum_{i=1}^m \hat{\mu}_i}{m} \quad (7)$$

The estimator of the mean of the prior distribution of λ is $\frac{1}{\lambda'}$. Hence the estimator of λ' can be given as:

$$\hat{\lambda}' = \frac{m}{\sum_{i=1}^m \hat{\mu}_i} \quad (8)$$

To get $\hat{\lambda}_2$, this can be done by substitute $\hat{\lambda}'$ in (6) as follows:

$$\hat{\lambda}_2 = \frac{n+1}{\sum_{i=1}^n t_i + \frac{m}{\sum_{i=1}^m \hat{\mu}_i}} \quad (9)$$

As the Bayesian estimator of λ when the population distribution is $Exp(\lambda)$ and the prior distribution is $Exp(\lambda')$. Substituting this for λ in the expression of $S(t)$ for the exponential distribution, we get:

$$\hat{S}_2(t) = \exp \quad (10)$$

As a Bayesian estimator of survivor function.

2.3. Bayesian Estimation with Gamma Prior

If the prior distribution of λ is assumed to be Gamma (α, β) , the posterior distribution of λ can be followed by gamma distribution with parameters $(n + \alpha, \sum_{i=1}^n t_i + \beta)$. The mean and variance of $gamma(\alpha, \beta)$ are respectively $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$. To estimate α and β , $\underline{\hat{\mu}}$ and $\hat{\sigma}^2$ can be computed as:

$$\underline{\hat{\mu}} = \frac{\sum_{i=1}^m \frac{1}{\hat{\mu}_i}}{m} \quad (11)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^m \left(\frac{1}{\hat{\mu}_i} - \underline{\hat{\mu}} \right)^2}{m} \quad (12)$$

Where $\underline{\hat{\mu}}$ and $\hat{\sigma}^2$ are the sample mean and variance for the prior distribution and $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_m$ are the means of m samples of size n are taken before as in section (2.2).

By equating each sample moment to its corresponding population moment, then:

$$\frac{\alpha}{\beta} = \frac{\sum_{i=1}^m \hat{\mu}_i}{m}$$

$$\frac{\alpha}{\beta^2} = \frac{\sum_{i=1}^m (\hat{\mu}_i - \underline{\hat{\mu}})^2}{m} \quad (13)$$

From equation (13), the estimates α and β can be given as:

$$\hat{\alpha} = \frac{\mu^2}{\sigma^2} \quad (14)$$

$$\hat{\beta} = \frac{\mu}{\sigma^2} \quad (15)$$

The mean of the posterior distribution is:

$$\mu = \frac{n + \alpha}{\sum_{i=1}^n t_i + \beta} \quad (16)$$

So that:

$$\hat{\lambda}_3 = \frac{n + \hat{\alpha}}{\sum_{i=1}^r t_i + \hat{\beta}} \quad (17)$$

Where $\hat{\lambda}_3$ is the Bayesian estimator of λ and then the estimator of the survivor functions in this case is:

$$\hat{S}_3(t) = \exp \left[- \frac{n + \hat{\alpha}}{\sum_{i=1}^r t_i + \hat{\beta}} t \right] \quad (18)$$

2.4. Simulation Experiments

To investigate the performance of the estimation methods with respect to bias and efficiency, the paper carried out a number of Monte Carlo experiments. This is done by generate a population of survival times (with no censored times) from an exponential distribution with known parameter (λ).

2.4.1. Experiment (I): Maximum Likelihood Estimation

1. Select a sample of size n and denoted by i .
2. Calculate $\hat{\lambda}_1$ using (2) and denoted by $\hat{\lambda}_{1i}$
3. Calculate $\hat{S}_1(t)$ using (5) and denoted it by $\hat{S}_{1i}(t)$
4. Calculate the residuals

$$r_{ij} = \hat{S}_{1i}(t_j) - S_i(t_j) \quad j = 1, \dots, n \quad (19)$$

Where $\hat{S}_{1i}(t_j)$ is the value of an estimate of survivor function at j^{th} survival time in sample i and $S_i(t_j)$ is the value of the true survival function at survival time j in sample i .

5. Calculate the mean square error for sample i

$$MSE_{1i} = \frac{\sum_{j=1}^n r_{ij}^2}{n} \quad (20)$$

6. Repeats steps 1 – 5 for $i = 1, \dots, N$ (N very large).
7. Calculate the average of mean square error and variance for $\hat{S}_1(t)$:

$$MSE_1 = \frac{\sum_{i=1}^N MSE_{1i}}{N} \quad (21)$$

$$V_1 = \frac{\sum_{i=1}^N (MSE_{1i} - MSE_1)^2}{N} \quad (22)$$

2.4.2. Experiment (II): Bayesian Estimation with Exponential Prior

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

1. Calculate $\hat{\mu}$ using (7).
2. Calculate $\hat{\lambda}_2$ using (9) and denoted by $\hat{\lambda}_{2i}$.
3. Calculate $\hat{S}_2(t)$ using (10) and denoted by $\hat{S}_{2i}(t)$
4. Calculate the residual

$$r_{ij} = \hat{S}_{2i}(t_j) - S_i(t_j) \quad j = 1, \dots, n \quad (23)$$

Where $\hat{S}_{2i}(t_j)$ is the value of an estimate of survivor function at j^{th} survival time in sample i and $S_i(t_j)$ is the value of the true survival function at survival time j in sample i .

5. Calculate the mean square error for sample i

$$MSE_{2i} = \frac{\sum_{j=1}^n r_{ij}^2}{n} \quad (24)$$

6. Repeats steps 1 – 6 for $i = 1, \dots, N$ (N very large).

8. Calculate the average of mean square error and variance for $\hat{S}_2(t)$:

$$MSE_2 = \frac{\sum_{i=1}^N MSE_{2i}}{N} \quad (25)$$

$$V_2 = \frac{\sum_{i=1}^N (MSE_{2i} - MSE_2)^2}{N} \quad (26)$$

2.4.3. Experiment (III): Bayesian Estimation with Gamma Prior:

1. Select m samples of size n and calculate their means $\frac{1}{\hat{\mu}_1}, \frac{1}{\hat{\mu}_2}, \dots, \frac{1}{\hat{\mu}_m}$.

2. Calculate $\hat{\mu}$ and $\hat{\sigma}^2$ using equations (11) and (12) respectively

3. Obtain the moment estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β respectively.

4. Calculate $\hat{\lambda}_3$ using (17) and denoted by $\hat{\lambda}_{3i}$

5. Calculate $\hat{S}_3(t)$ using (18) and denoted by $\hat{S}_{3i}(t)$

6. Calculate the residuals

$$\begin{aligned} r_{ij} &= \hat{S}_{3i}(t_j) - S_i(t_j) \quad , \quad j = 1, \dots, n \\ r_{ij} &= \hat{S}_{3i}(t_j) - S_i(t_j) \quad j = 1, \dots, n \end{aligned} \quad (27)$$

Where $\hat{S}_{3i}(t_j)$ is the value of an estimate of survivor function at j^{th} survival time in sample i and $S_i(t_j)$ is the value of the true survival function at survival time j in sample i .

7. Calculate the mean square error for sample i

$$MSE_{3i} = \frac{\sum_{j=1}^n r_{ij}^2}{n} \quad (28)$$

8. Repeats steps 1 – 7 for $i = 1, \dots, N$ (N very large).

9. Calculate the average of mean square error and variance for $\hat{S}_3(t)$:

$$MSE_3 = \frac{\sum_{i=1}^N MSE_{3i}}{N} \quad (29)$$

$$V_3 = \frac{\sum_{i=1}^N (MSE_{3i} - MSE_3)^2}{N} \quad (30)$$

3. Analysis and Results

In this section, the study compared the performance of maximum likelihood and Bayesian estimation methods in uncensored data through a simulation experiment. Maximum likelihood estimators are considered for under the exponential distribution. In addition to, the Bayesian estimators are investigated for both Gamma prior and Exponential prior. Discussion is focused on the performance of estimation methods with respect to bias and efficiency under various sample sizes.

For all cases an exponentially distributed population of size 10000 is generated. Different Sample sizes 10,30,50,75 and 100 are selected, with 1000 repetitions for each sample and 10 iterations are performed for stability purposes. The simulation results are shown in table 1.

Table (1): The results of estimation methods for all sample sizes

Methods of estimation	n	$\hat{\lambda}$	$mean \frac{1}{\lambda}$	bias	bias_st	MSE	VAR
Likelihood Method (LES)	10	0.019961	50.1042	4.8958	0.0641	0.0077	0.00013
	30	0.01884	53.0956	1.9044	0.0409	0.0025	1.29E-05
	50	0.018507	54.0406	0.9594	0.0264	0.0014	4.49E-06
	75	0.018418	54.2983	0.7017	0.0220	0.00098	1.93E-06
	100	0.018386	54.3948	0.6052	0.0313	0.00075	1.16E-06
Bayesian Estimation with Exponential Distribution Prior (BEP)	10	0.02274	44.05512	10.94488	0.069003	0.009705	0.000208
	30	0.019813	50.50469	4.49531	0.039986	0.00278	1.74E-05
	50	0.019041	52.52497	2.47503	0.035028	0.001627	5.62E-06
	75	0.018848	53.11571	1.88429	0.018331	0.001079	2.42E-06
	100	0.018464	54.21828	0.78172	0.023356	0.000756	1.14E-06
Bayesian Estimation with Gamma Distribution Prior (BGP)	10	0.020114	49.75265	5.24735	0.054916	0.007644	0.000156
	30	0.018456	54.26077	0.73923	0.037896	0.002647	1.44E-05
	50	0.018522	54.07895	0.92105	0.037523	0.001487	5.4E-06
	75	0.018263	54.81687	0.18313	0.02267	0.001052	2.3E-06
	100	0.018715	55.48431	-0.48431	0.019347	0.000765	1.25E-06

Source: Authors prepared, 2021

Tables 1 presents the estimates of the exponential parameters, mean square error, variance and bias of survival function using likelihood and Bayesian methods with exponential prior and gamma prior under different sample sizes when the population contains no censors. The results show that the values of bias_st, MSE and VAR decrease when samples' sizes increase for all estimation methods. This indicates that the bias of the estimates of the parameters as well as the survivor function decrease with sample size which increases and the efficiency of the estimates regarding to MSE & VAR that also increases with increasing sample sizes.

Table (2): The bias and efficiency of the estimates of the exponential parameter and survivor function using LES, BEP and BGP methods

Sample Size	Bias			MSE		
	LES	BEP	BGP	LES	BEP	BGP
10	0.0641	0.0690	0.0549	0.0077	0.0097	0.0076
30	0.0409	0.0399	0.0379	0.0025	0.0028	0.0026
50	0.0264	0.0357	0.0375	0.0014	0.0016	0.0015
75	0.0220	0.0183	0.0226	0.0009	0.0011	0.0010
100	0.0313	0.0233	0.0193	0.0007	0.0007	0.0007

Source: Authors Prepared, 2021

Table 2 presents the bias and MSE of parameters estimates for the survivor function using Likelihood Method (LES), Bayesian Estimation with Exponential Distribution Prior (BEP), and Bayesian Estimation with Gamma Distribution Prior (BGP) methods. The results show that the values of Bias decreases when the sample sizes increase. Also the same results regarding to efficiency of estimates, i.e. the values of Mean Square Error (MSE) decrease when sample sizes increase. Generally, the results showed that for all estimation methods when the sample sizes increase the bias of three survivors function estimates decrease and the efficiency (MSE) increases especially for Bayesian Estimation with Exponential Distribution Prior method (smallest values for Bias and MSE). This indicates that the best performance with respect to both bias and efficiency is the BGP method (Bayesian Estimation with Gamma Distribution Prior).

4. Conclusion

The general conclusions were drawn from the results of the paper when the data is uncensored. The bias of the estimates of the parameters as well as the survivor function decrease when the sample size increases and the efficiency of the estimates also increases with increasing sample sizes. These results are applicable for all three methods used, namely; the likelihood method, Bayesian method with exponential prior and Bayesian method with gamma prior.

The results of the three methods are compared for bias estimates of parameters; the Bayesian estimates with gamma prior provide the least bias estimates followed by the likelihood method. The Bayesian estimates with exponential prior gave the weak performance (large bias). This may be the parameter value of the exponential distribution (λ) ranging between 0 and 1. Regarding the survivor function, the study found that the Bayesian estimators are superior compared to the likelihood estimators especially in the case of the Bayesian estimator with gamma prior which provided least bias estimates. For the efficiency of the estimates, the study found that the Bayesian estimator with exponential prior gave the worst performance while the likelihood estimators and Bayesian estimators with gamma prior gave almost the same level of efficiency.

Recommendations and suggestions:

As for parametric estimation, we recommend the use of a Bayesian estimator with gamma prior. In this paper, only exponential and gamma priors are examined. More research is needed to investigate the performance of the other priors. In particular the Weibull and lognormal distribution need to be considered.

References:

1. Calabria, R. & Pulcini, G. (1996). Point estimation under asymmetric loss functions for left-truncated exponential samples. *Communications in Statistics - Theory and Methods*, 25(3): 585-600. <https://doi.org/10.1080/03610929608831715>
2. Durga H. Kutal, D.H. & Qian, L. (2018). A Non-Mixture Cure Model for Right-Censored Data with Fréchet Distribution. *Stats* 1(1): 176–188. <https://doi.org/10.3390/stats1010013>
3. Epifani, I. & Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica*, 20: 1455–1484.
4. Guure, C. B. & 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](https://doi.org/10.1155/2012/329489)
5. Guure, C. B., Ibrahim, N. A. & 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*. <https://doi.org/10.1155/2012/589640>
6. Kamran Abbas, & Ket-al. (2020). Bayesian Estimation of Gumbel Type-II Distribution under Type-II Censoring with Medical Applications. *Computational and Mathematical Methods in Medicine*. 2020: 1-11. <https://doi.org/10.1155/2020/1876073>
7. Noortwijk, J.M. & 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.
8. Omari, M. A. Al & Ibrahim, N.A. (2011). Bayesian survival estimation for Weibull distribution with censored data. *Journal of Applied Sciences*, 11(2): 393–396. <https://doi.org/10.3923/jas.2011.393.396>
9. Palacio, A. R. & Leisen, F. (2018). Bayesian nonparametric estimation of survival functions with multiple-samples information. *Electronic Journal of Statistics*, 12(1):1330 –1357. <https://doi.org/10.1214/18-ejs1420>
10. Rizki, S.W, Mara, M.N., & 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. <https://doi.org/10.1088/1742-6596/855/1/012036>
11. Syuan-Rong, H. & Shuo-Jye, W. (2011). Bayesian estimation and prediction for Weibull model with progressive censoring. *Journal of Statistical Computation and Simulation*, 82(11): 1607- 1620. <https://doi.org/10.1080/00949655.2011.588602>
12. Tekindal, M.A., Yonar, H., & Kader, S. (2021). Evaluation of Parametric Method Performance for Left-Censored Data and Recommendation of Using for Covid-19 Data Analysis. *EJMO*, 5(2): 132–143.
13. Thamrin, S.A., Zoraida, M.N., Jaya, A.K. & Ansariadi. (2018). Performance of Bayesian Using Conjugate Prior Estimator for Weibull Right Censored Survival Data. *Asian Journal of Scientific Research*, 11(3): 376- 382. <https://doi.org/10.3923/ajsr.2018.376.382>