



Original Article

Integration of prior information in Kaplan Meier estimator using Bayesian approach

Hamimes Ahmed^{a,b*} and Benamirouche Rachid^b

^a Faculty of Medicine, University of Constantine3, Algeria.

^b National Higher School of statistics and applied economics, Tipaza, Algeria.

ARTICLE INFO

Article history:

Received 30 November 2020

Revised 25 December 2020

Accepted 28 December 2020

Keywords:

Bayesian approach;
nonparametric survival;
Kaplan Meier;
Bayesian estimator.

ABSTRACT

As part of this contribution, we will illustrate the effectiveness of the Bayesian approach in estimating durations; we suggest a new definition of the Kaplan Meier Bayesian estimator based on a stochastic approximation under an informative prior. For this reason, based on the lognormal distribution, we have unconjugated a priori distributions. This method of processing makes it possible to assume that the use of the a priori data with the various suggested methods is sensitive to the choices of the parameters added.

Faculty of Natural Sciences and Life, University of El Oued

1. Introduction

The study of censored lifetimes is used in various fields of research and various possibilities for modeling these data have been suggested. The first area of use and collection of survival data was in the biomedical sciences where it is used for therapeutic or epidemiological trials. In economics, we look for the time spent unemployed, in a job or between two employers, the duration of a transport trip, the life of a business or the amount of a “revolving” type loan. In Bayesian inference, survival analysis has gained increasing attention in recent years, but remains limited due to the scarcity of specialized software (one of the causes of this scarcity is the difficulty of automating Bayesian analyzes by compared to the frequentist approach), as well as the force of habit and the difficulty of adopting a particular statistical definition. The great importance of Bayesian inference in frequentist approaches is its great clarity and its theoretically consistent methodology, which allow us to deduce the results of richer and more direct explanations than those given by the classical

approach. Unless we can assume a priori that the distribution of survival time obeys a parametric rule, thanks to many nonparametric methods, the most important of which is Kaplan-Meier, we can estimate the survival function S . Several works have been based on the development of this estimator. Khizanov and Maïboroda (2015), proposed a modification based on a mixing model with various concentrations. Rosse and Zieliński (1999) used local smoothing of the Kaplan-Meier estimator based on an approximation by the Weibull distribution function. Rosse and Zieliński (2002), introduce a Kaplan-Meier estimator based on a Weibull distribution approximation. Shafiq Mohammad et al (2007), presented a weighting of the Kaplan Meier estimator for heavily censored data under the sine function. Previous studies neglected the study of the prior parameter distribution in improving this type of estimate. In this article, we use some prior information integration methods to improve the estimation quality of the Kaplan Meier model. For this purpose, we have unconjugated a priori

* Corresponding author : Hamimes Ahmed Tel.: 077794624

E-mail address: ahmedhamimes@yahoo.com

Peer review under responsibility of University of El Oued

DOI : <http://doi.org/10.57056/ajb.v1i2.30>

distributions based on the lognormal distribution. This way of processing makes it possible to conclude that the use of the a priori information with the different ways proposed is sensitive to the choices of the parameters introduced.

2. Bayes formula and posterior distribution

In the discrete case and if we consider $(H_i)_{i \in \{1, \dots, n\}}$ the set of all the hypotheses (or else the causes or the circumstances) of the occurrence of an event E of non-zero probability. If the effect E occurs at the same time as one and only one of the H_i , i.e.

$$E = (E \cap H_1) \cup (E \cap H_2) \cup \dots \cup (E \cap H_n),$$

according to the theory of total probabilities we write

$$\begin{aligned} P(E) &= P(E \cap H_1) + P(E \cap H_2) + \dots \\ &\quad + P(E \cap H_n) \\ &= \sum_{i=1}^n P(E \cap H_i) = \sum_{i=1}^n P(E/H_i)P(H_i) \\ &= E^H(P(E/H_i)), \end{aligned}$$

the conditional probability of a cause H_i is given by

$$P(H_i/E) = \frac{P(E \cap H_i)}{P(E)} = \frac{P(E/H_i)P(H_i)}{E^H(P(E/H_i))}, \quad (1)$$

equation (1) is Bayes' theory (or rule).

The a posteriori distribution and defined by:

$$\begin{aligned} \pi(\theta/x) &= \frac{f(x/\theta) \times \pi(\theta)}{\int_{\theta_0} f(x/\theta) \times \pi(\theta) d\theta} \\ &= \frac{f(x/\theta) \times \pi(\theta)}{m(x)} \end{aligned} \quad (2)$$

this a posteriori distribution is the combination of

- $f(x/\theta)$ the density function of x knowing the value of the random variable θ .
- $\pi(\theta)$ models the a priori density function on θ .
- $m(x)$ the marginal distribution of x .

Expression (2) represents what we know and what we do not know before about the parameter considering the observed data; it is also the update of $\pi(\theta)$ after observing our sample.

Once we have the data, the quantity $m(x)$ is a normalization constant, which guarantees that $\pi(\theta/x)$ is indeed a probability distribution. We can write:

$$\pi(\theta/x) \propto f(x/\theta) \times \pi(\theta) \quad (3)$$

Equation (3) shows that Bayesian inference satisfies the likelihood principle: a posteriori, the information from the data comes exclusively from the likelihood $f(x/\theta)$.

2.1. The Monte Carlo Method by Markov Chains (MCMC)

The Bayesian approach requires simulating sample sequences following the distributions of interest. However, these distributions are often multivariate, of non-standard type, so it is necessary to use sophisticated simulation methods such as Markov chains where a sequence of random variables $(\theta_t)_{t \geq 1}$ is a Markovian chain of first order if the conditional distribution $\pi(\theta_t/\theta_{t-1}, \theta_{t-2}, \dots, x) = \pi(\theta_t/\theta_{t-1}, x)$. The convergence of this chain is linked to a new concept called ergodicity. A sequence is said to be ergodic if it is possible to go from any realization a in the chain to any realization b . In another way the sequence of conditional distributions $\{\pi(\theta_t/\theta_0, x)\}$ converges to the marginal distribution $\pi(\theta/x)$ for any value θ_0 . In other words, the dependence of the chain on θ_0 gradually fades. Suppose that this convergence occurs before T . the statistical mean,

$$\frac{1}{T} \sum_{i=T+1}^{T+n} g(\theta_i),$$

almost surely converges to $E(g(\theta))$.

The jump rules are governed by the kernel transition $K(\theta_t, \theta_{t+1})$, which is a mechanism describes the probability of moving from one state to another based. These rules constitute an important element in the ergodicity theorem.

Definition 1 (The Monte Carlo method by Markov Chains (MCMC)).

The Markov chain Monte Carlo method is any method that produces an ergodic Markov chain whose stationary distribution is the distribution of interest.

Note that in the case of θ_t i.i.d we use the Strong Distribution of Large Numbers (LFDN) to make the approximation; on the other hand, in the case of θ_t generated by the MCMC methods, the ergodic theorem is used instead of LFDN because we have a dependency structure between the θ_t defined by the Markov chain Robert and Casella (2004).

The two most popular algorithms are the Metropolis-Hastings algorithm and the Gibbs sampling algorithm presented below.

2.1.1. Gibbs sampling

Gibbs sampling is the most used MCMC algorithm, introduced by Geman and Geman (1984) within the framework of the restoration of satellite images, the idea of this algorithm is simple, to raise the difficulties of calculations in a complex model composed of n-parameters $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ where the distribution $\pi(\theta_j/x)$ of a parameter / subset of θ is non-standard, and when the full distribution $\pi(\theta_j/\theta_{-j}x)$ will often be standard, it is possible to simulate prints

$$\theta_j \sim \pi(\theta_j/\theta_{-j}x),$$

generate what will be called a Gibbs sequence (artificial sequences) as follows:

The algorithm breaks down into the following points:

1. Initialize $\theta^0 = \theta_{(1)}^0, \theta_{(2)}^0, \dots, \theta_{(d)}^0$, which is the first vector of elements in the string.
2. Pose $t \leftarrow 0$.
3. To go from step t to step $t + 1$:

$$\left\{ \begin{array}{l} \text{Generate } \theta_1^{(t+1)} \text{ by simulating according to the law } \pi(\theta_1^{(t+1)}/\theta_2^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_d^{(t)}, x) \text{ (an iteration)} \\ \text{Generate } \theta_2^{(t+1)} \text{ by simulating according to the law } \pi(\theta_2^{(t+1)}/\theta_1^{(t)}, \theta_3^{(t)}, \dots, \theta_d^{(t)}, x) \\ \vdots \\ \text{Generate } \theta_d^{(t+1)} \text{ by simulating according to the law } \pi(\theta_d^{(t+1)}/\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{d-1}^{(t)}, x) \end{array} \right.$$

4. Change the value of t to $t \leftarrow t + 1$, and go to 3.

Each turn of this algorithm describes a first order Markov chain due to the dependence on the previous realization. When the conditional distribution of each iteration is positive and for all θ_{-j} , the chain is ergodic which implies the convergence of chain after iteration n towards its distribution of interest $\pi(\theta/x)$ independently of the initial state $\theta^{(0)}$. So we write for all functions $g: E \rightarrow IR$ integrable, we have:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T g(\theta^t) \sim E^\pi[g(\theta)] \quad (4)$$

2.1.2. Metropolis Hastings algorithm (MH)

The Metropolis-Hastings method is the first of the MCMC methods, it was developed by

Metropolis et al (1953) in the fifties initially for particle physics, and generalized by Hastings (1970) in a more statistical framework in the year 1970. The M-H algorithm can be written as follows (Begin (2010))

Iteration 0: Give the initial value to θ_0 .

Iteration t: Update θ_t through $\theta_{t+1}, t = (1, 2, \dots)$, as follows:

1. Generate $\theta^* \sim q(\theta^*/\theta_t)$
2. Calculate the α -value according to the Metropolis Hastings formula:

$$\alpha(\theta_t, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*) q(\theta_t/\theta^*)}{\pi(\theta_t) q(\theta^*/\theta_t)} \right\}$$

3. Accept θ^* with the probability α such that :

$$\theta_{t+1} = \begin{cases} \theta^* & \text{with the probability } \alpha(\theta_t, \theta^*) \\ \theta_t & \text{with the probability } [1 - \alpha(\theta_t, \theta^*)] \end{cases}$$

The probability $\alpha(\theta_t, \theta^*)$ is called the acceptance ratio of H-M, and the acceptance rate of H-M represents the average of the acceptance probabilities over all iterations given by:

$$\bar{\alpha} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \alpha(\theta_t, \theta_t^*)$$

3. The Kaplan Meier Estimator

The Kaplan-Meier estimator is defined by:

$$\hat{S}(t) = \prod_{i: t_i < t} \left(1 - \frac{d_i}{n_i} \right) \quad (5)$$

This estimator is also called by Anglo-Saxon statisticians "Product Limit Estimations (PLE)". This estimator, which is a generalization of the notion of empirical distribution function, is based on the following idea: to survive after a time t is to be alive just before t and not to die at time t. , if t_i represents an instant during which there is the observation of at least one event, then the probability of survival at time t_i is equal to the probability of having survived before t_i multiplied by the "conditional" probability of surviving at time t_i .

The use of the term "conditional" means here that it is about the probability of surviving time t_i knowing that the individuals were survivors in t_i :

$$S(t_i) = P(X > t_i / X \geq t_i) * S(t_{i-1})$$

Let us call d_i and c_i , the numbers of individuals who, respectively, know the event and exit from observation at t_i . The number n_i of individuals subject to the risk of experiencing the event at t_i corresponds to the set of individuals who, just before this instant t_i was reached, had neither known the observed event, nor were out

4. The Bayesian conception of the Kaplan Meier estimator

In the frequentist approach, the number of deaths in the interval of time is a realization of a Binomial distribution written by:

$$d_i \sim \text{bin}(n_i, q_i)$$

or

$$q_i = 1 - \frac{d_i}{n_i} \quad (6)$$

From a Bayesian perspective we assume an a priori for q_i , and when the distribution used in the case of proportions is that of Beta, we set:

$$q_i \sim \text{beta}(\alpha, \beta) \quad (7)$$

In the Bayesian approach, equation (7) replaced by the a priori distribution of q_i , this a priori distribution has several important characteristics in our situation:

For the hyperparameters (α, β) , we find several propositions:

A vague a priori distribution, it is a proper distribution with a very large variance, according to this distribution, the a priori distribution is considered to be weak informative, and one uses this distribution for the regularization and the stabilization, it provides solutions in the use of algorithms. We pose:

$$q_i \sim \beta(0,01,0,01) \quad (8)$$

5. Integration of the a priori distribution

Informative priors represent the subjective way of thinking where the a priori is based on the information available on the parameter obtained. Among these methods we find the histogram approach, the relative likelihood and the conjugate approach remains the most standard solution in the informative framework. The conjugate a priori distributions can be

partially justified by an invariance reasoning. It is also possible to increase the robustness of the conjugate distributions by hierarchical models, before the rise of numerical computation, these families were practically the only ones which allowed computations to succeed. However, cases often arise in which an unconjugated prior is desirable, despite the increased mathematical difficulty. For example, generic databases often express epistemic uncertainty in terms of a lognormal distribution, which is not conjugated to the binomial likelihood function. In addition, the conjugate priors have relatively light tails and may influence the results too much in cases where there is little data that conflicts with the priors. The estimate provided by the data will generally lie in the tail of the prior distribution in such cases, where the prior probability is very low. In this section, we describe how to make an inference with a lognormal prior, which is a commonly encountered unconjugated prior.

The beta distribution is a bit more complicated algebraically. The mean is equal to $\alpha / (\alpha + \beta)$ and the variance is a complicated expression in terms of the parameters α, β . The expression of the variance can be more conveniently rewritten in terms of the mean as $\text{mean} (1 - \text{mean}) / (\alpha + \beta + 1)$, and it can be solved for α, β .

Development of an unconjugated (lognormal) prior - One of the things that makes the lognormal distribution attractive as a priori in PRA (Probabilistic Risk Analysis) is the ease with which it can encode uncertainty on a parameter which varies over several orders of magnitude. The uncertainty encoded by the lognormal distribution is generally not provided in terms of the distribution parameters (l and s) needed by OpenBUGS. More commonly, information is given in terms of a median or mean value and an error factor, or sometimes in terms of upper and lower bounds.

In this method if the number of deaths in the interval of time is a realization of a Binomial distribution written by:

$$d_i \sim \text{bin}(n_i, q_i)$$

so

$$q_i \sim \text{log-normal}(\mu; \sigma^2) \quad (9)$$

using the properties of the lognormal distribution, any of these sets of information can be translated into the l and s parameters needed

by OpenBUGS, as shown in the script snippets below.

Use the following lines if the median and the error factor are given:

```
mu <-log (median); tau <-pow (log (EF) /1.645, -2)
```

Use the following lines if the mean and error factor are given:

```
mu <-log (mean) - pow (log (EF) /1.645, 2) / 2; tau <-pow (log (EF) /1.645, -2)
```

Use the following lines if the median and upper limit are given:

```
mu <-log (median); tau <-pow (log (upper / median) /1.645, -2)
```

6. Application

In this section, survival function is estimated in a clinical study for two pharmaceuticals (placebo and prednisolone), this example uses survival times for 42 patients with chronic active hepatitis. These patients were randomized into two equal groups, one was treated with prednisolone, the other received a placebo (see Held, 2010). In this example, patients with prednisolone are used.

Table 1. Survival Bayesian Kaplan Meier.

Time	Total No of Deaths	Total No of censored	No at risk	Kaplan Meier
2	1	0	21	0.9545
6	1	0	20	0.9082
12	1	0	19	0.8624
54	1	0	18	0.8169
56	0	1	17	0.8164
68	1	0	16	0.7686
89	1	0	15	0.7198
96	1	0	14	0.6233
125	0	1	13	0.6228
128	0	1	12	0.6223
131	0	1	11	0.6218
140	0	1	10	0.6211
141	0	1	9	0.6204
143	1	0	8	0.5414
145	0	1	7	0.5406
146	1	0	6	0.4502
148	0	1	5	0.4492
162	0	1	4	0.4482
168	1	0	3	0.2985
173	0	1	2	0.2969
181	0	1	1	0.294

If we assume the absence of any a priori information on the estimated survival model, the choice of an uninformative a priori is

obvious. We use in this article a conjugate prior distribution such that $\alpha = \beta = 0.01$.

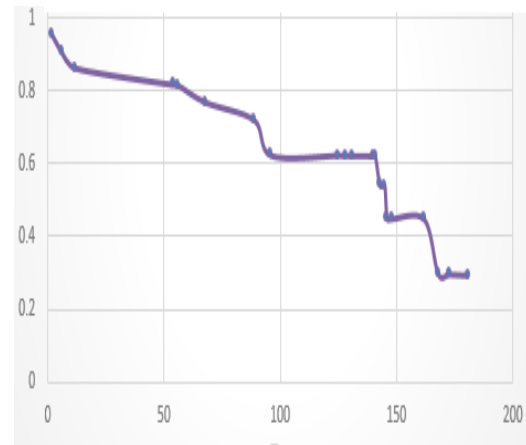


Fig 1. The survival curve estimated according to the Bayesian Kaplan-Meier method with an a priori of beta.

From Figure (1), it can be seen that at the start of the curve, 100% of the individuals in the sample are included in the treatment study of. After more than 146 days after using the treatment in the sample 50% of the patients had died. But, the treatment failure for the rest of the individuals in the sample lasts for a long time, for some it exceeds 180 days. If we assume the presence of information on the mean and the error factor as follows (prior.mean = 0.05, prior.EF = 5) then the survival function is given by the following form:

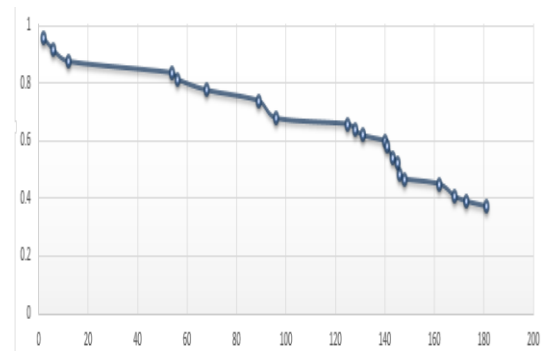


Fig 2. The survival curve estimated according to the Bayesian Kaplan-Meier method with an informative a priori.

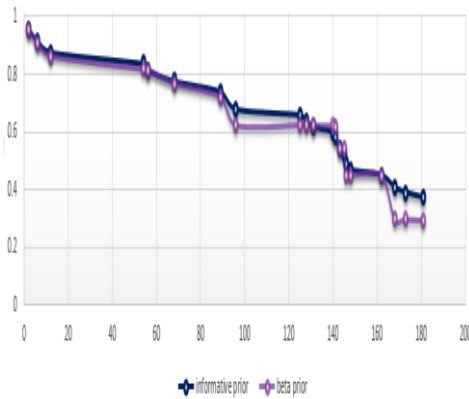


Fig3. The survival curves estimated according to the Bayesian Kaplan-Meier method with an a priori of beta and with an informative a priori.

In the comparison between the two informative and non-informative Kaplan Meier curves, we find a small difference which does not change the interpretation in our example in the tails of the survival curves and this because the probability of survival at time t_i is equal the probability of having survived before t_i multiplied by the "conditional" probability of surviving over time t_i .

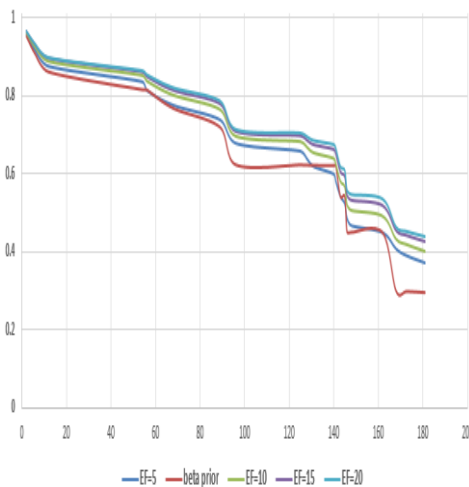


Fig 4. the survival curves estimated according to the Bayesian Kaplan-Meier method with an a priori of beta and with an informative a priori according to several error factors.

We notice from fig (4), when the error factor increases the probabilities of survival also increase, so the more the uncertainty on the a priori information increases, the more the instantaneous risk decreases and the survival increases. The error factor chosen in Figure (3) and after this study of the effects of the error factors is a plausible choice.

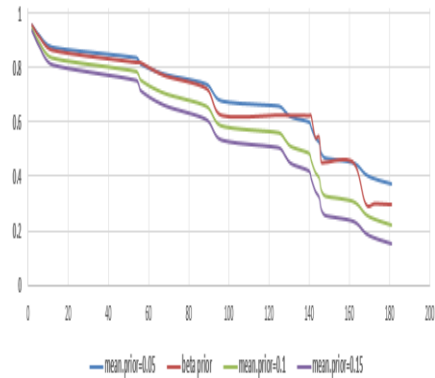


Fig 5. The survival curves estimated according to the Bayesian Kaplan-Meier method with an a priori of beta and with an informative a priori according to several a priori means.

We notice in fig (5), when the proposed a priori mean increases the probabilities of survival also decreased, so the more the uncertainty on the a priori information increases, the more the instantaneous risk increases and the survival decreases. The proposed a priori mean chosen on an informative basis in Figure (4) and after this study of the effects of different a priori means is a rational choice. In this way we can conclude that the use of the a priori information with the different proposed ways is sensitive in terms of treatment, and capable of changing the survival curve differently.

Conclusion

This article has allowed us to understand how we use certain a priori information integration methods to improve the estimation quality of the Kaplan Meier model. For this purpose we have unconjugated a priori distributions based on the lognormal distribution. This way of processing makes it possible to conclude that the use of the a priori information with the different ways proposed is sensitive to the choices of the parameters introduced.

Appendices (OpenBUGS code)

```

■ model
■ {
■ for (i in 1:m1) {
■ d1[i]~dbin(q1[i],n1[i])
■ q1[i]~dlnorm(mu, tau) # Lognormal
■ prior distribution for p
■ }
■ for (i in 1:m1){
■ ce1[i]~dbin(0.01,0.01)
■ }

```

```

▪ for (i in 1:m1){
▪   qc1[i]~dbeta(0.01,0.01)
▪ }
▪ for (i in 1:m1){
▪   p1[i]<-1-q1[i]
▪ }
▪ n1[1]<- 22
▪ for(i in 2:m1){
▪   n1[i]<-n1[i-1]-d1[i-1]-ce1[i-1]
▪ }
▪ for (i in 2:m1){
▪   s1[i]<-s1[i-1]*p1[i]
▪ }
▪ s1[1]<-p1[1]
▪
▪   tau <- 1/pow(log(prior.EF)/1.645, 2) #
▪   Calculate tau from lognormal error
▪   factor
▪   # Calculate mu from lognormal prior
▪   mean and error factor
▪   mu<-log(prior.mean)
▪   pow(log(prior.EF)/1.645, 2)/2
▪ }
▪ list(m1=21,d1=c(1,1,1,1,0,1,1,2,0,0,0,0,
▪ 0,1,0,1,0,0,1,0,0),
▪ ce1=c(0,0,0,0,1,0,0,0,1,1,1,
▪ 1,1,0,1,0,1,1,0,1,1), prior.mean = 0.05,
▪ prior.EF = 5)

```

References

1. Begin J.F. *Analyse MCMC de certains modèles de diffusion avec application à la marche Européenne du carbone. Rapport Technique*, Université de Montréal, Canada, 2010.
2. Chib S., Greenberg E. *Understanding the Metropolis-Hastings algorithm*. The American Statistician. 1995, 49, 327–335.
3. Hastings W.K. *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika. 1970, 57, 97-109.
4. Held U. *Représentation graphique et comparaison de courbes de survie*. Forum Med Suisse. 2010, 10(33), 548-550.
5. Khizanov V.G., Maïboroda R. *A modified Kaplan-Meier estimator for a model of mixtures with varying concentrations*. Theor. Probability and Math. Statist. 92 (2016), 109-116.
6. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. *Equation of state calculations by fast computing machines*. The Journal of Chemical Physics. 1953, 21, 1087-1092.
7. Robert C.P. *Le choix Bayésien : principes et pratiques*. Springer. 2006.
8. Rossa A. and Zielinski R. *Locally Weibull-Smoothed Kaplan–Meier Estimator*, Institute of Mathematics Polish Academy of Sciences, Preprint 599. 1999.
9. Rossa A., & Zieliński R. *A simple improvement of the Kaplan-Meier estimator*. Communications in Statistics-Theory and Methods. 2002, 31(1), 147-158.
10. Shafiq M., Shah S., & Alamgir M. *Modified Weighted Kaplan Meier Estimator*. Pakistan Journal of Statistics and Operation Research. 2007, 3(1), 39-44.

Recommended Citation

HamimesA. & Benairouche R. Integration of prior information in Kaplan Meier estimator using Bayesian approach. *Algerian Journal of Biosciences*. 2020, 01;02:076-082.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).