INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 07-09-2023

Accepted: 28-12-2023

Published: 20-01-2024

Citation: Jaisankar R, Varshini KSP (2024) On Addressing Censoring in Survival Data Using Fuzzy Theory. Indian Journal of Science and Technology 17(4): 312-316. https://doi.org/10.17485/IJST/v17i4.2288

* **Corresponding author**.

r_jaisankar@buc.edu.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

On Addressing Censoring in Survival Data Using Fuzzy Theory

R Jaisankar[1]*, K S Parvatha Varshini[2]

1 Professor, Department of Statistics, Bharathiar University, Coimbatore, Tamil Nadu, India
2 Research Scholar, Department of Statistics, Bharathiar University, Coimbatore, Tamil Nadu, India

Abstract

Objectives: The subject of survival analysis mainly concentrated on the behavior of the survival times of specific organisms or objects, which is the time at which a specified event occurs from an initial epoch. The problem of censoring adds complexity to survival analysis, and specific types of tools have been developed to deal with censored observations, some of which are nonparametric and semi-parametric approaches. When the proportion of censored observations becomes higher, it is natural that the results obtained from any type of survival analysis are less reliable. In this article, an attempt has been made to meaningfully incorporate the censored observations in survival data using fuzzy theory. **Method:** Taking Kaplan-Meier's procedure as a base, the fuzzy numbers are used for censored observations. **Findings:** In the presence of censored observations, ascertaining the median survival time is always obscure; hence, fuzzy methodology has been adopted in this article. The corresponding survival curves are drawn, and the median survival time is estimated with reference to an example. **Novelty:** Incorporating fuzzy theory into Kaplan-Meier's procedure, particularly regarding censoring, is a novel methodology discussed here. So, the presence of impreciseness in statistical data with particular reference to survival analysis has not been considered much in the literature. The present work has addressed this issue and proposes a new type of Kaplan-Meier estimator for addressing censoring.

Keywords: Censoring; Survival Curve; Product Limit Estimator; Fuzzy sets; and Fuzzy Numbers

1 Introduction

Survival analysis has been developed to deal with the problems involved in clinical setups, such as estimating the remission time of patients or getting a cure from disease when using different treatment protocols, etc. Its application in the present scenario has increased considerably among many disciplines. The data for survival analysis consists of the time taken for a well-defined event of interest to occur with information about some variables or attributes that influence the event's occurrence. A significant problem to be dealt with in the survival analysis is the presence of censored observations, which may mainly arise because a subject is lost during follow-up or the event of interest does not occur to the subject even till the end of the study. Such incomplete information makes the survival analysis tedious. It would be hard to say whether and when the censored subject has experienced the event or not. Omitting the information about the censored subjects may dilute the reliability of the results obtained from the analysis, and incorporating the incomplete information unhesitatingly may also create some lacunae with reference to the point of perfection.

Grzegorzewski and Hryniewicz[1] investigated a generalization of the exponential model that acknowledges ambiguity in lifetimes. They use fuzzy numbers to model imperfect lives. It is assumed that the censorship is ambiguous. Grzegorzewski[2], Grzegorzewski, and Hryniewicz[3] suggested that a fuzzy method for estimating a mean lifetime in the presence of ambiguous data, particularly an uncertain number of failures, estimated the mean time to failure, which is also observable as a fuzzy number, and determined the accompanying confidence interval using fuzzy numbers. Jaisankar R and Parvatha[4] proposed an approach for a Fuzzy Kaplan-Meier estimator based on fuzzified survival times and compared it to the traditional method.

The prime familiar nonparametric procedures such as the Kaplan-Meier estimator for calculating the Survival probabilities[5], Kian L Pokorny et al.[6] proposed a fuzzy product limit estimator that could be utilized even with extensive censoring. They demonstrated that it is more trustworthy than the classical estimator. Semi-parametric procedures like Cox-Regression[7] are developed to combat censored observations.

A failure must be clearly defined for the Conventional estimators, which might not always be met in an actual case. Investigators may report highly ambiguous descriptions that may arise due to various factors, including perceptions of subjective and imprecise failures, imprecise records, etc. Hence, the survival times of censored subjects are mostly fuzzy, and it would be objective to use the fuzzy theory in such circumstances. This article attempts to deal with censored observations using fuzzy concepts. The proposed methodology facilitates the use of fuzzy numbers in the conventional Kaplan-Meier survival procedure to obtain more reliable results, which is the main contribution of the present work.

Fuzzy algebra and logic are now used in survival analysis to deal with uncertain failure times. Viertl[8–10] expanded the statistical estimation of the reliability characteristics by considering the fact that the real-time data are more or less fuzzy. Viertl and Shafiq[11–13] generalized the Kaplan Meier estimator for fuzzy survival time observations and for the parametric survival models.

2 The Kaplan Meier Estimator (KM)

In 1958, Edward L. Kaplan and Paul Meier[5] worked together on a landmark work on how to handle incomplete observations in lifetime data, and they observed the survival and hazard curves using conditional probabilities. Since then, it becomes familiar conventional method of handling different survival periods with censored data. The Kaplan-Meier method also posits random censoring, independent censoring times, and independent survival times with the same distribution and constant probability of survival, in addition to the standard assumption of independence. The censored are assumed to have lived until after the time at which their final known existence. The Kaplan-Meier estimator of the survivor function at times, for $t_{(k)} \leq t < t_{(k+1)}$ is given by,

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{(n_j - d_j)}{n_j}$$

where, $t_1, t_2, t_3, \ldots n d_{1,} d_{2,} d_{3,} \ldots n_1, n_2, n_{3,} \ldots$ are the corresponding number of patients remaining in the cohort.

3 Methodology

The estimator suggested by Kaplan-Meier[5] is based on conditional probability, which is determined based on the number of fatalities and the number of persons at risk. As these are discrete quantities, it is not appropriate to consider them fuzzy. At first, using Kaplan-Meier's procedure the survival probabilities are calculated. One of the main drawbacks of the Kaplan-Meier estimator is that the survival probability cannot be calculated for the censored subjects and hence it remains the same till the next death occurs. As it is a case, in order to overcome this at least to some extent, the current methodology, which may be called the fuzzy Kaplan-Meier estimator is developed by considering the censored survival times as fuzzy.

In many practical situations even though the final event of interest is clearly defined, there are certain stages which leads to the final event may be identified in the course of the study. For example, in cardiology, the heart of a person may function even with small defect in the valve which may not leads to complete heart failure. Such a functioning cannot be considered as a case of perfect heart functioning but only a diminished functioning. Similarly, an HIV infected person may be identified for pre-AIDS condition which may lead to a final event like full-blown AIDS or death. This may be described as a state of partial occurrence of the event. If the final event is death or failure such conditions may be called as a partial or non-critical failure. Likewise, a subject who is censored would not be identified for the status of either death or survival and hence his status is vague. However, one may access the possibility for that the censored person to experience the event. Such kind of evaluation is possible through some notions of possibility theory like possibility of dominance or necessity of dominance which are helpful when someone wants to measure the degree for some imprecisely defined qualities. The possibility of attaining the event can be linguistically defined as slightly possible, highly possible, nearly sure etc., and for each of which one may assign weights $s_i \in (0,1)$ arbitrarily. The theory of fuzzy would be helpful in dealing with such conditions.

If the lifetimes are not crisp but vague, one must deal not only with failures but also with non-failures in particular with the censored subject. So, in order to take into account such non-failures, let C_i be the time at which the i-th subject is being censored. Let S denote the set of all subjects which are not yet experienced the event of interest. If $s_i = 1$ for a subject, which indicates that it does not experience the event before the C_i while in the case of $s_i = 0$ then it might have experienced the event before or at the time C_i. If $s_i \varepsilon (0,1)$ then the subject unknown for its failure or survival. Hence, the set S is a fuzzy set with finite support and core. As the status of the censored is not known after C_i, one may assign for the censored subject the degree of belongingness $s_i = \mu_S(i)$ to S where $s_i \varepsilon [0,1]$, for $i = 1,2,\ldots,n$. One may also consider an alternative set T which may contain the subjects who attained the event so that their corresponding degree of belongingness is $t_i = \mu_T(i) = 1 - s_i$.

Now, one must determine a fuzzy counterpart of the observed failures, say, q. Grzegorzewski (2001) suggested several methods of failure counting techniques. One may consider only just significant failures or to treat all types of failures (partial failures) in the same way. This may yield two methods which may be called optimistic (or liberal) and pessimistic (conservative) with reference to the points of view, respectively. According to the optimistic viewpoint, the number of failures counted as,

$$Q_{opt} = \sum_{i=1}^{n} I(t_i = 1) = n - \sum_{i=1}^{n} I(s_i > 0)$$

where I is the indicator function and the number of failures obtained from a pessimistic viewpoint is

$$Q_{pes} = n - \sum_{i=1}^{n} I(s_i = 1) = \sum_{i=1}^{n} I(t_i > 0)$$

In general, if only failures with some degree of belongingness, say, φ is to be considered, then we have

$$Q_{\varphi} = n - \sum_{i=1}^{n} I(t_i > \varphi) = \sum_{i=1}^{n} I(s_i \geq 1 - \varphi)$$

where $\varphi \in (0,1)$.

The failure counting methods described above are, in some ways, crisp and reductive. Actually, they discard the entire information on specific degrees of belongingness and use only a portion of it. However, it is occasionally beneficial to consider all available information. Then the following failure counting approach can be used.

$$Q^c = \sum_{i=1}^{n} t_i = |T| = n - \sum_{i=1}^{n} s_i = n - |S|$$

where $|T|$ and $|S|$ denotes the cardinalities of the fuzzy sets T and S.

4 Results and Discussions

The following self-coined example has been taken for the purpose of illustration of the proposed methodology. The survival times (t_i), the number of deaths (d_i) with 13 subjects are given in the following Table 1. Using the classical Kaplan Meier procedure, the cumulative survival probabilities $S(t)$ are calculated. The curves corresponding to each of these probabilities are drawn and presented Figure 1.

Figure 2 shows the fuzzy optimistic and pessimistic approaches of survival curves for the Kaplan-Meier estimator.

It is to be noted that as the failure counting of optimistic procedure is same as that of the classical Kaplan-Meier procedures their corresponding curves coincides but pessimistic curve differs it is seen that the median survival obtained by both the classical and optimistic method is 7 which is lesser than that of pessimistic point of view. As comparing survival curves is one of the prime objectives of survival analysis drawing this kind of curve may provide some additional understanding about the pattern of survival times.

Table 1. Classical Kaplan Meier Estimator

Time(t_i)	Status	Number of Patients at Risk(d_i)	Cumulative Survival Probability$S(t)$
0	1	13	0.9231
2	1	12	0.8462
3	1	11	0.7693
4	1	10	0.6924
6	1	9	0.6155
7	1	8	0.5386
9	0		
12	1	6	0.4489
16	1	5	0.3591
17	1	4	0.2693
23	0		
27	1	2	0.1347
31	1	1	0

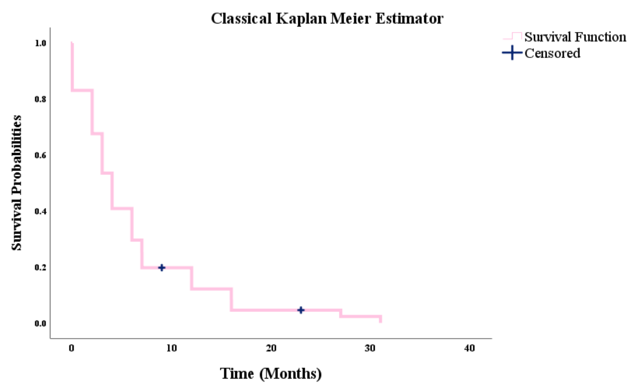

Fig 1. Classical Kaplan Meier Estimator

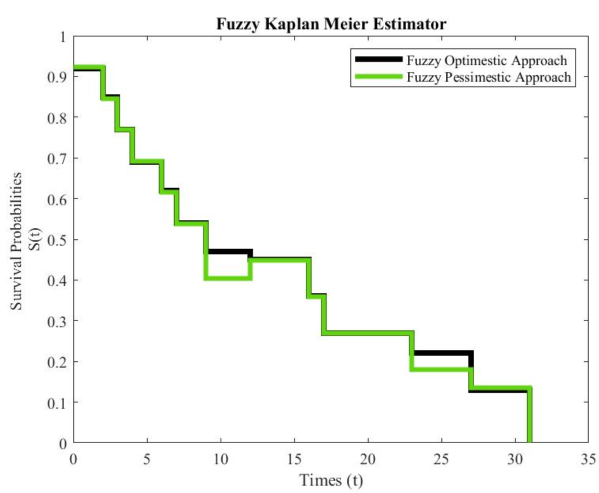

Fig 2. Fuzzy Kaplan Meier Estimator

5 Conclusion

In situations where it is requires to compare the efficacy of two or more treatments or drugs in terms of time required to get cure from a disease, survival analysis helps to make a decision for selecting the right and appropriate treatment. However, if there is a suspicion of impreciseness or inaccuracy or vagueness present in the data, the theory of fuzzy help us to tackle the problems under consideration. In survival analysis especially in the presence of censored observations ascertaining the median survival time is always in obscure and hence fuzzy methodology has been adopted in this article. The median obtained so is lesser than that of the classical method. The behavior of the survival curve can be studied with more accuracy when incorporating the fuzzy ideology which may aid to make better decisions on health protocols. It is possible to refine the methodology adopted to the cases of truncation also.

References

1) Grzegorzewski P, Hryniewicz O. Lifetime Tests for Vague Data. In: Computing with Words in Information/Intelligent Systems 2;vol. 34 of Studies in Fuzziness and Soft Computing. Heidelberg, Germany. Physica. 1999;p. 176–193. Available from: https://doi.org/10.1007/978-3-7908-1872-7_8.
2) Grzegorzewski P, Hryniewicz O. Computing with words and life data. *International Journal of Applied Mathematics and Computer Science*. 2002;12(3):337–345. Available from: https://eudml.org/doc/207591.
3) Grzegorzewski P. Estimation of the mean lifetime from vague data. In: Proceedings of the 2nd International Conference in Fuzzy Logic and Technology. 2001;p. 348–351. Available from: https://www.researchgate.net/publication/221399466_Estimation_of_the_mean_lifetime_from_vague_data.
4) Jaisankar R, Varshini KSP, Siva M. A Fuzzy Approach to Kaplan-Meier Estimator. *Mathematical Statistician and Engineering Applications*. 2022;71(3s2):1107–1114. Available from: https://www.philstat.org/special_issue/index.php/MSEA/article/view/340/336.
5) Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*. 1958;53(282):457–481. Available from: https://www.jstor.org/stable/2281868.
6) Musavi S, Pokorny KL, Poorolajal J, Mahjub H. Fuzzy survival analysis of AIDS patients under ten years old in Hamadan-Iran. *Journal of Intelligent & Fuzzy Systems*. 2015;28(3):1385–1392. Available from: https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs1422.
7) Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972;34(2):187–220. Available from: https://www.jstor.org/stable/2985181.
8) Viertl R. Statistical Methods for Non-Precise Data. 1st ed. CRC Press. 1995. Available from: https://www.routledge.com/Statistical-Methods-for-Non-PreciseData/Viertl/p/book/9780849382420.
9) Viertl R. Univariate statistical analysis with fuzzy data. *Computational Statistics & Data Analysis*. 2006;51(1):133–147. Available from: https://doi.org/10.1016/j.csda.2006.04.002.
10) Viertl R. On reliability estimation based on fuzzy lifetime data. *Journal of Statistical Planning and Inference*. 2009;139(5):1750–1755. Available from: https://doi.org/10.1016/j.jspi.2008.05.048.
11) Shafiq M, Viertl R. Generalized Kaplan Meier Estimator for Fuzzy Survival Times. *Śląski Przegląd Statystyczny*. 2015;13(19):7–14. Available from: https://doi.org/10.15611/sps.2015.13.01.
12) Shafiq M, Viertl R. Empirical reliability functions based on fuzzy life time data. *Journal of Intelligent & Fuzzy Systems*. 2015;28(2):707–711. Available from: https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs1351.
13) Shafiq M, Viertl R. On the estimation of parameters, survival functions, and hazard rates based on fuzzy life time data. *Communications in Statistics - Theory and Methods*. 2017;46(10):5035–5055. Available from: https://doi.org/10.1080/03610926.2015.1099667.