INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

A Fuzzy Approach to Kalbfleisch - Prentice Estimator

R Jaisankar[1]*, **H Haripriya**[2]

1 Professor, Department of Statistics, Bharathiar University, Coimbatore, 641046, Tamil Nadu, India
2 Research Scholar, Department of Statistics, Bharathiar University, Coimbatore, 641046, Tamil Nadu, India

Check for updates

OPEN ACCESS

Received: 09-01-2025

Accepted: 24-02-2025

Published: 03-03-2025

Citation: Jaisankar R, Haripriya H (2025) A Fuzzy Approach to Kalbfleisch - Prentice Estimator. Indian Journal of Science and Technology 18(8): 635-639. https://doi.org/10.17485/IJST/v18i8.65

* **Corresponding author**.

r_jaisankar@rediffmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To amalgamate the fuzziness present in the survival data with the randomness by evolving new types of survival estimators. **Methods:** This study discusses the methodology, which incorporates fuzziness present in the survival data, and the estimation of Kalbfleisch – Prentice estimator for the survival function. **Findings:** The hazard and survival functions of the fuzzified version of the Kalbfleisch—Prentice estimator are obtained and compared with the classical type. There is a subtle difference before and after fuzzification, and the median survival time also shows variation. **Novelty:** So far, no procedure has been suggested to address the fuzziness that prevails in the data in the case of Kalbfleisch –Prentice estimator; this study has suggested a methodology for the same.

Keywords: Survival Data; Covariates: Fuzzification; Non-Parametric Estimate; Kalbfleisch - Prentice Estimator

1 Introduction

The significance of Survival Analysis is derived from its examination of data related to the lifespans of biological organisms. While lifetimes can be described using various probability distributions, a significant challenge in analyzing survival data is the issue of censoring, which leads to a preference for non-parametric methods among experts in the field. One of the methods developed is attributed to Kaplan Meier (1958) [1], which is the most recognized approach and can be applied in situations involving heavily censored data. It estimates the Survival function through conditional probabilities derived from both censored and observed events, relying on the assumptions of independence between censored and survival times, non-informative censoring, and homogeneity of Risk. Nelson-Aalen (1972) [2] introduced an alternative non-parametric estimator for the purpose of estimating the cumulative hazard function. The primary limitation of these estimators is their inability to integrate the covariates.

Cox (1972) [3] proposed a method for estimating the hazard function that relies on the proportional hazard assumption and establishes linear relationships with the covariates. Two non-parametric estimators have been proposed for the baseline hazard function of the Cox model. The first method is based on Breslow (1972) [4], which is effective for large datasets with minimal ties and where event times are nearly continuous. The second method, attributed to Kalbfleisch-Prentice (1973) [5], is suitable even for smaller datasets with frequent ties.

In survival data, survival times are usually continuous, however they are frequently recorded as discrete values. Continuous observations are obviously imprecise, and this also applies to survival times (Viertl (1995)[6]). When survival times are affected by certain continuous factors, they may also be regarded as fuzzy. Therefore, in addition to unpredictability, the inherent fuzziness in the data must be considered and handled to provide more trustworthy survival and hazard estimates. This study proposes a method to integrate fuzziness into the Kalbfleisch–Prentice estimator (1973)[5].

1.1 The Kalbfleisch-Prentice Estimator

The Kalbfleisch-Prentice estimator (1973)[5] is a novel estimator that estimates the survival function using discrete failure times to approach continuous function, developed to address the lacuna of the Kaplan Meier estimator on its inability to adjust for covariates. Hence, it may be considered as a counterpart of the Kaplan Meier estimator (1958)[1].

The Kalbfleisch-Prentice estimator (1973)[5] has a wide application in survival analysis and competing risk models for estimating the cumulative incidence function (CIF), for measuring the probability of a specific type of event occurring in the setting of conflicting hazards.

Let $D(t_j)$ be the set of patients who failed at the time t_j. For distinct failure time $t, j = 1, \ldots, m$,

$$t_1 < t_2 < \ldots < t_m.$$

Let p_j denote the conditional probability of surviving at the time t_j for a subject at baseline. The baseline survival function can be estimated as

$$\widehat{S}_{0,KP}(t) = \prod_{j:t_j \leq T} \hat{p}_j^{\tau_j} \tag{1}$$

where τ_j is the indicator of whether the event occurred $(\tau_i$=1) or was censored (τ_0=1)

The likelihood function LF is given by,

$$LF = \prod_{j=1}^{m} \prod_{l \in D(t_j)} \left(1 - p_j^{exp(\widehat{\beta}^T x^*)}\right) \prod_{k \in R(t_j) - D(t_j)} p_j^{exp(\widehat{\beta}^T x^*)} \tag{2}$$

where $R\left(t_j\right)$ is the Risk set, that is the set of individuals at risk just before time t.

Take the estimated $\beta^T = \widehat{\beta}^T$ from the partial likelihood and maximize the likelihood function LF with regard to p_j, which is the solution of the equation

$$\sum_{l \in D(t_j)} \frac{exp(\widehat{\beta}^T x_l)}{\left(1 - p_j^{exp(\widehat{\beta}^T x_l^*)}\right)} = \sum_{k \in R(t_j)} exp(\widehat{\beta}^T x_k) \tag{3}$$

where x_j is the set of covariates for a subject who died at t_j. Assuming no ties, the solution $\hat{p}_j$ is

$$\hat{p}_j = \left[1 - \frac{exp(\widehat{\beta}^T x_j)}{\sum_{k \in R(t_j)} exp(\widehat{\beta}^T x_k)}\right]^{exp(-\widehat{\beta}^T x_j)} \tag{4}$$

Accordingly, the KP estimated survival function for a subject with covariates $X = x*$ is:

$$\widehat{S}_{KP}(t|X = x*) = \left[\widehat{S}_{0,KP}(T)\right]^{exp\left(\widehat{\beta}^T x^*\right)} = \left[\prod_{j:t_j \leq T} \hat{p}^{\tau_j}\right]^{exp\left(\widehat{\beta}^T x^*\right)} \tag{5}$$

2 Methodology

Let $t_1, t_2, \ldots, t_m$ represents the survival times of m subjects taken in a study. As stated earlier, the survival times are continuous in nature but recorded as discrete quantities, and hence they are more or less fuzzy.

This kind of fuzziness is to be addressed together with the inherent randomness present in the data for any type of analysis. In this paper, a methodology has been proposed to fulfill this objective for the Kalbfleisch-Prentice non-parametric survival estimator. The procedure for the same has been given below:

In the first step, the survival times are converted into Triangular fuzzy numbers with a specified fuzzification factor. Triangular fuzzy numbers are more commonly used fuzzy numbers which are mainly defined on three points: The minimum possible value (lower), the most likely value (middle), and the maximum possible value (upper). Let them be denoted by the triplet $(\Delta^*_{iL}, \Delta^*_{iM}, \Delta^*_{iU})$ respectively. Let $\Delta^*_i, (i = 1, 2, \ldots, m.)$ represents the fuzzified survival times. Let X_j, (j = 1,2,…,n) represent the covariates which are expected to influence the lifetime. If there are some covariates that are of continuous in nature, a similar procedure of fuzzification which converts them into triangular fuzzy numbers is to be adopted.

Let, $\vartheta_{\Delta^*_i}(t_i)$, ($t_i$ represents the survival times) be the membership functions of the fuzzified survival times that assigns a degree of membership to each of the time points, whose values range from 0 to 1, where 0 indicates no membership and 1 indicates full membership. The fuzzified survival time Δ^*_i can be represented together with the membership function as a fuzzy set $\Delta^*_i = (t, \vartheta_{\Delta^*_i}(t))$.

After converting the survival times as well as the continuous covariates as fuzzy numbers, the $\alpha - cuts$ of each of them for a specified level of αwere found. A crisp interval which contains all the points with a membership degree greater than or equal to α, where the value of α ranges from 0 to 1 is the $\alpha - cuts$ of that particular fuzzy number. Let the $\alpha - cuts$ for the survival times are represented by $\Delta^*_{i\alpha}$, given by,

$$\Delta^*_{i\alpha} = \{t_i \,|\, \vartheta_{\Delta^*_i}(t)\} \geq \alpha \tag{6}$$

Each $\alpha - cut$ is characterised by upper and lower level values. A similar procedure has been adopted for the continuous covariates. These values of the $\alpha - cuts$ are used for defuzzificiation of the fuzzy number. The defuzzificiation may be done by any method, but the centroid method is preferred. The centroid (Centre of Gravity Method (COG) (1965)) of a fuzzy set Δ^*_i, denoted as T_i, is calculated as given by,

$$COG(T_i) = \frac{\sum_{i=p}^{p} t_i.\vartheta_{\Delta^*_i}(t)}{\sum_{i=p}^{p} \vartheta_{\Delta^*_i}(t)} \tag{7}$$

So also to the case of fuzzified covariates of the model was done. Using the defuzzified values, the Kalbfleisch – Prentice estimator was calculated with the help of "Python Software" and the results so obtained may be compared with the classical Kalbfleisch – Prentice estimator.

2.1 Numerical Illustration

For illustration purposes, the survival data of 40 patients in an outpatient Treatment Clinic was taken to check whether the time to relapse among drug users is related to patient age and drug of choice**.(** Source: Data provided courtesy of Dr. Chad. L. Cross in the textbook - Biostatistics – A Foundation for Analysis in the Health Sciences – Wayne W. Daniel, Chad L. Cross (Page no: 769). URL: https://faculty.ksu.edu.sa/sites/default/files/145_stat_-_textbook.pdf)

The data represents the self-reported time (in weeks) of the occurrence of relapse (or, the time at which the patient was lost to follow-up). Patient status is mentioned as: 0-Censored and 1-Relapse. The Drug of Choice is denoted by 1 for Opiate and 2 for other drugs. The patient's age was also taken as an additional covariate.

The methodology proposed for Fuzzy Kalbfleisch – Prentice estimator is adopted for the above data with alpha cut value = 0.2.

3 Results and Discussion

The curves of both survival and hazard functions for the Classical Kalbfleisch–Prentice estimator as well as the Fuzzified Kalbfleisch–Prentice estimator are illustrated below for comparison.

It can be observed that (From Figure 1 and Figure 2) there is a subtle difference that exist between both the survival and hazard curves before and after fuzzification. The median survival time observed using the Classical Kalbfleisch - Prentice estimator was found to be 19.00 which is greater than that of the one observed through the Fuzzified Kalbfleisch - Prentice estimator (16.29).

To test whether there is any significant difference exists between the survival curves obtained from the Classical Kalbfleisch - Prentice estimator and its fuzzified version, the log - rank test procedure was adopted with the null hypothesis that the curves are similar. The test statistic for the same is given by,

Test statistic is, $\chi^2 = \sum \frac{(O_i - E_i)^2}{V_i}$

where O_i is the observed number of events in Group i,

E_i is the expected number of events in Group i,
V_i is the variance of the difference between the observed and expected events.

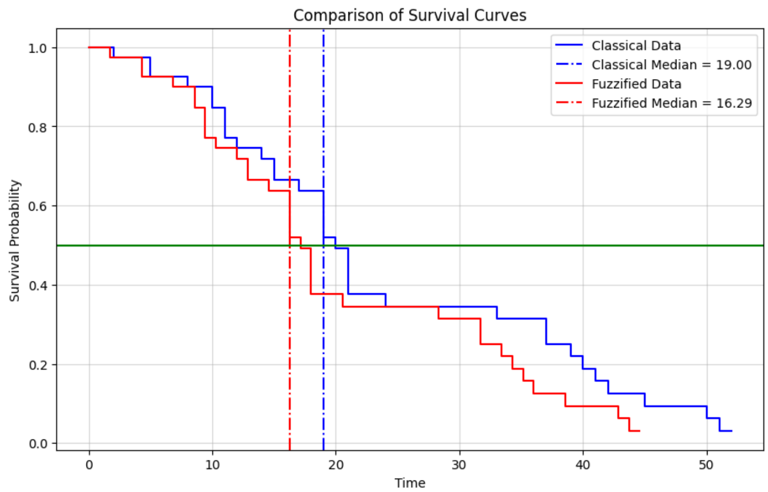

Fig 1. Comparison of Classical and Fuzzified Survival curves along with the representation of Median Survival time in both Classical and Fuzzified cases

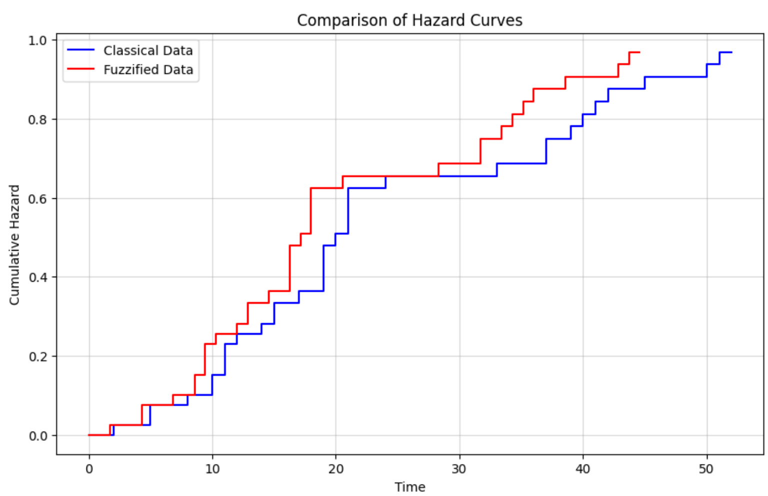

Fig 2. Comparison of Classical and Fuzzified Hazard curves

The test statistic is calculated using the defuzzified values of survival times, and the details of the results are given below:

Table 1. Log rank test results

Log Rank test Results	
Chi-square test Statistic	2.159
p – Value	0.141

On comparing the curves through the Log - rank test (Figure 3), it is observed that there is no significant difference captured in these curves. However, it depends on the selection of the fuzzification factor applied for fuzzification. However, the Kalbfleisch - Prentice estimator is only used as a baseline survival and its impact would be explicitly revealed only in its application to the Cox survival model.

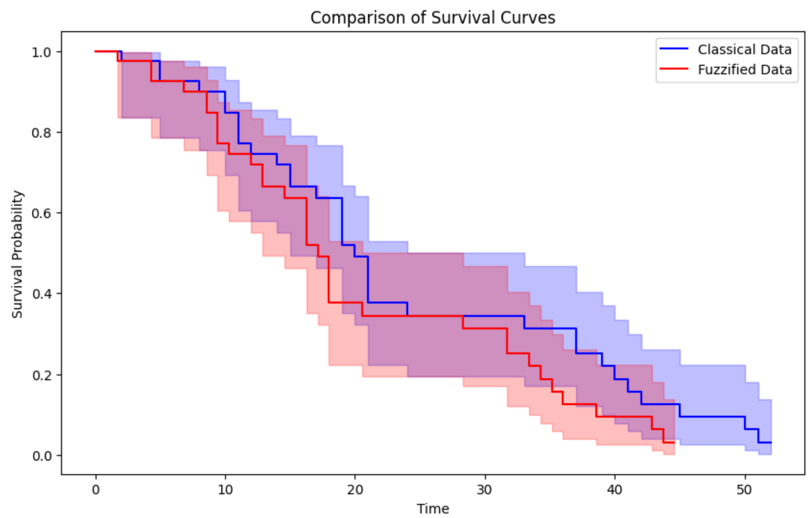

Fig 3. Comparison of Survival curves for the Log rank test

When a statistical analysis is conducted using the data observed from a random sample, it is crucial to take into account both the randomness and the fuzziness of the data, especially when the data are continuous. This is even truer in the case of the survival time data. Based on this truth, a fuzzified version of the Kalbfleisch-Prentice survival estimator was found and the corresponding values of the survival/hazard function are also estimated. The assimilation of fuzziness may make the model more realistic and hence produce accurate results since any observation of continuous nature tends to exhibit ambiguity. Ignoring this would lack the reliability of the results obtained from any type of analysis.

4 Conclusion

This study introduces a fuzzification methodology and examines its implications concerning the Kalbfleisch-Prentice estimator. It is observed that the estimates of survival probabilities and the survival curves exhibit differences when fuzzification is considered. The median survival time observed using the Classical Kalbfleisch - Prentice estimator is 19.00 which is greater than the Fuzzified Kalbfleisch - Prentice estimator (16.29). Hence, the results obtained can be regarded as more reliable since it considers both fuzziness and also randomness. It is to be noted that, the nature of survival data, the event, and the researcher's viewpoint should also be taken into consideration when choosing the fuzzification factor in this case.

References

1) Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*. 1958;53(282):457–481. Available from: https://doi.org/10.2307/2281868.
2) Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics*. 1972;14(4):945–966. Available from: https://www.tandfonline.com/doi/abs/10.1080/00401706.1972.10488991.
3) 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.
4) Breslow NE. Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*. 1975;43(1):45–57. Available from: https://www.jstor.org/stable/1402659.
5) Kalbfleisch JD, Prentice RL. Marginal likelihoods based on Cox's regression and life model. *Biometrika*. 1973;60(2):267–278. Available from: https://www.jstor.org/stable/2334538.
6) Viertl R. Statistical methods for non-precise data. CRC Press. 2009. Available from: https://web.archive.org/web/20190226175655id_/http://pdfs.semanticscholar.org/70f7/fd9a33200123536b76d399babe6d33f70f6c.pdf.