Estimation of the Conditional Survival Function with Interval-Censored Data

LUÍS MEIRA-MACHADO
Centre of Mathematics
University of Minho
Guimarães
PORTUGAL

Abstract: - This paper focuses on the estimation of the conditional survival function in the presence of interval-censored data. While estimation is relatively straightforward for qualitative covariates using Turnbull's estimator, the case of continuous covariates, particularly under interval censoring, poses additional methodological challenges. We describe existing kernel-based approaches, such as those using Nadaraya-Watson weights, applicable under right censoring. The main contribution of this work is the proposal of two novel methodologies tailored for estimating the conditional survival function when data are interval-censored. These methods use kernel weights and a modified EM algorithm to provide practical ways to solve this challenging problem.

Key-Words: - survival analysis, interval censoring, conditional survival function, nonparametric estimation.

Received: February 27, 2025. Revised: April 29, 2025. Accepted: May 23, 2025. Published: July 22, 2025.

1 Introduction

The estimation of the survival function is an important goal in survival analysis, as it provides the probability that an individual will not experience the event before a specific time. Traditionally, this function is estimated nonparametrically using the Kaplan-Meier estimator [1], which is widely appreciated for its simplicity and interpretability. However, the Kaplan-Meier estimator is designed specifically for right-censored data, the most common form of censoring, where the event of interest has not occurred by the end of the observation period. Right censoring is common in clinical trials, for example when patients leave early or the study ends before the event happens.

Another common form of censoring is interval censoring, where the exact time of the event is unknown but is known to occur within a specific time interval. This type of censoring is prevalent in medical studies with periodic follow-ups, dental studies where check-ups happen at regular intervals, or reliability studies where inspections are not frequent. A simple but often inaccurate way to handle interval-censored data involves imputing the event time using strategies like left endpoint, right endpoint, or midpoint imputation. While these methods offer simplicity, they can lead to biased estimates and misrepresent the underlying survival distribution.

A comprehensive overview of statistical methods for analyzing interval-censored failure time data can be found in [2], which includes both classical and modern approaches, including maximum likelihood estimation, multiple imputation, and Bayesian techniques. His work also discusses extensions to recurrent events and dependent censoring, making it a key reference for understanding the methodological landscape of interval-censored data analysis.

Fortunately, there are more appropriate methods for handling interval-censored data. One such method is given in [3], a nonparametric maximum likelihood estimator (NPMLE) specifically designed for interval-censored data. The so-called Turnbull estimator generalizes the Kaplan-Meier estimator and allocates probability masses across disjoint intervals identified from the observed data. It accurately captures the survival function without the need for imputation, providing a more reliable analysis.

The problem becomes more complicated when we want to estimate survival based on other variables. When dealing with a single qualitative covariate, the problem can be addressed straightforwardly through stratification, applying the Turnbull estimator separately to each stratum defined by the covariate levels. However, challenges arise when the covariate is continuous, as stratification is no longer viable.

In right-censored data, this issue has been addressed using the method described in [4], which extends the Kaplan-Meier approach by incorporating weights to smooth over the covariate space. This estimator enables the estimation of a conditional survival function by weighting observations based on their proximity in the covariate space, thus accommodating continuous covariates.

This work proposes an alternative to the method in [4], tailored for interval-censored data. By integrating kernel weighting and an adapted Expectation-Maximization (EM) algorithm, we extend the applicability of nonparametric survival estimation to complex censoring structures. A related EM-based kernel smoothing approach was previously introduced in [5], where the Turnbull estimator was generalized to incorporate continuous covariates. A distinct framework based on influence functions and local efficiency was also introduced in [6].

This method is useful in many real-world cases where survival times are interval-censored and continuous covariates are present. For instance, in longitudinal medical studies, patients are often assessed during routine visits, resulting in interval-censored data; yet, biomarkers such as age, blood pressure, or gene expression levels, typically continuous, play a crucial role in prognosis and risk stratification. Similarly, in epidemiological research, exposure levels to environmental pollutants or dietary components may be modeled as continuous covariates while health outcomes like disease onset remain interval-censored due to periodic screenings.

Another prominent application appears in dental research, where outcomes such as tooth decay or implant failure are monitored over check-ups that occur at fixed intervals, and covariates like bone density or oral hygiene scores vary continuously across individuals. In industrial reliability studies, components may be inspected periodically, leading to interval-censored failure times, while stress levels, temperature, or usage patterns serve as continuous explanatory variables.

The remainder of this paper is structured as follows: Section 2 reviews Turnbull's estimator for interval-censored data with qualitative covariates. Section 3 addresses the estimation of the conditional survival function with continuous covariates, presenting two kernel-based methodologies tailored for interval-censored data. Finally, Section 4 offers concluding remarks and outlines directions for future research.

2 Estimation of Survival with Interval Censoring

2.1 Notation and Definitions

Let T denote the survival time with cumulative distribution function (CDF) $F(t)$ and survival function $S(t) = 1 - F(t)$. In the context of interval-censored data, the exact survival time T_i is not observed; instead, it is only known to lie within an interval $[L_i, R_i)$. The observed data thus consist of n independent realizations $\{(L_i, R_i)\}_{i=1}^{n}$, where L_i and R_i define the left and right endpoints of the censoring interval for individual i.

In the special case of right censoring, the interval reduces to $[L_i, \infty)$, meaning that only a lower bound is known. Conversely, in left censoring, the interval is of the form $[0, R_i)$, indicating that the event occurred before R_i. When both endpoints are finite, we have general interval censoring.

2.2 Turnbull's Estimator

The Nonparametric Maximum Likelihood Estimator (NPMLE) introduced in [3] provides an estimate of the survival function $S(t)$ in the presence of interval censoring. Unlike the Kaplan-Meier estimator, which is specifically designed for right-censored data, Turnbull's estimator accounts for the additional uncertainty introduced by censoring intervals. The estimation procedure consists of the following key steps:

Step 1: Construction of Disjoint Intervals

Identify m disjoint intervals $\{[p_j, q_j)\}_{j=1}^{m}$ over which the survival function is estimated. These intervals are obtained from the union of the observed censoring intervals across all subjects. Specifically, they are constructed from the set of unique left and right endpoints of the observed intervals $\mathcal{I}_j = [p_j, q_j)$ where p_j and q_j are observed interval limits. No other observed endpoints exist between p_j and q_j.

Step 2: Probability Mass Allocation

Assign probability masses s_j to each Turnbull interval, ensuring that the estimated survival probabilities satisfy the constraints:

$$\sum_{j=1}^{m} s_j = 1, \quad s_j \geq 0.$$

Step 3: Likelihood Maximization

The likelihood function for interval-censored data is given by:

$$L(s) = \prod_{i=1}^{n} \left(\sum_{j=1}^{m} \alpha_{ij} s_j \right), \tag{1}$$

where the indicator function α_{ij} is defined as:

$$\alpha_{ij} = \begin{cases} 1, & \text{if } [p_j, q_j) \subseteq [L_i, R_i), \\ 0, & \text{otherwise.} \end{cases}$$

This likelihood is maximized using the Expectation-Maximization (EM) algorithm, which iteratively updates the probability mass distribution until convergence.

3 Estimation of the Conditional Survival Function

In this section, we address the problem of estimating the conditional survival function, denoted as $S(t \mid x)$, which represents the probability of surviving beyond time t given a covariate value $X = x$. Accurately estimating $S(t \mid x)$ is essential for understanding the effect of covariates on survival probabilities and for making informed predictions in survival analysis.

3.1 Estimation Approaches

When the covariate X is categorical, the estimation process is relatively straightforward, as the data can be stratified by covariate levels, allowing Turnbull's estimator to be applied separately to each stratum. However, when the covariate is continuous, stratification becomes impractical, requiring alternative estimation techniques.

The conditional survival function $S(t \mid x)$ can be estimated using different approaches. Parametric methods, such as Exponential, Weibull or Log-Normal distributions, assume a specific functional form for the survival distribution, which can lead to efficient estimates if the model is correctly specified but may introduce bias if the assumptions are violated. Semi-parametric methods, like the Cox proportional hazards model, provide greater flexibility by allowing an unspecified baseline hazard while still imposing a structured relationship between covariates and survival time. In contrast, non-parametric methods make minimal assumptions about the survival distribution, offering greater adaptability, particularly when the underlying relationship between covariates and survival time is complex or unknown.

In this work, we focus on a non-parametric approach, specifically employing kernel-weighting methods for estimation.

3.2 Kernel-Weighted Estimation with Continuous Covariates

3.2.1 Nadaraya-Watson Weights

For a continuous covariate X, the weight assigned to the i-th observation when estimating at a specific point x_0 using the Nadaraya-Watson estimator introduced in [7], [8] is:

$$W(x_0, X_i, h) = \frac{K_h(X_i - x_0)}{\sum_{j=1}^{n} K_h(X_j - x_0)}, \qquad (2)$$

where $K_h(\cdot)$ is a kernel function with bandwidth h, which controls the degree of smoothing.

The use of kernel weights for smoothing over continuous covariates follows the framework of local polynomial regression, as described in [9], offering improved bias properties particularly near boundary regions.

The choice of bandwidth h is crucial, as it determines the trade-off between bias and variance in the resulting estimate. A smaller h leads to lower bias but higher variance, while a larger h smooths more but may obscure important structure in the data.

For practical purposes, a data-driven bandwidth selector can be used. One widely adopted method is the direct plug-in kernel (DPK) selector, which is implemented in the `dpik` function from the `KernSmooth` package in `R`. This approach, as introduced in [10], provides a reliable, automatic choice of bandwidth that performs well in a wide range of settings.

3.2.2 Handling Right Censoring with IPCW

In survival analysis, right censoring occurs when the event time T is only known to be greater than an observed censoring time C. That is, instead of observing T_i, we observe

$$\widetilde{T}_i = \min(T_i, C_i), \quad \Delta_i = I(T_i \le C_i),$$

where $\widetilde{T}_i$ represents the observed time, and Δ_i is an event indicator taking the value 1 if the event is observed ($T_i \le C_i$) and 0 if the observation is censored ($T_i > C_i$).

To handle right censoring, we use Inverse Probability of Censoring Weighting (IPCW) [11]. Assuming the support of T is contained within that of C given X, we obtain:

$$E[I(T \le t) \mid X] = E\left[\frac{I(\widetilde{T} \le t)\Delta}{1 - G_X(\widetilde{T}^-)} \mid X\right], \quad (3)$$

where $G_X(t)$ is the conditional censoring CDF:

$$G_X(t) = P(C \le t \mid X).$$

This function can be estimated nonparametrically as described in [4]:

$$\widehat{G}_X(t) = 1 - \prod_{\widetilde{T}_i \le t} \left(1 - \frac{W(X_i, X, h)(1 - \Delta_i)}{\sum_{j:\widetilde{T}_j \ge \widetilde{T}_i} W(X_j, X, h)}\right). \quad (4)$$

3.2.3 Simple Imputation Approaches for Interval Censoring

A simple approach to estimate the conditional survival function under interval censoring consists of imputing the event time to a representative point within the censoring interval. The most common choices for imputation are the lower bound (L_i), the upper bound (R_i), or the midpoint of the interval, defined as:

$$M_i = \frac{L_i + R_i}{2}.$$

Among these options, midpoint imputation is the most commonly used, as it provides a natural compromise between assuming the event occurs at the earliest or latest possible time. Once the imputed event times are determined, standard right-censored survival analysis methods can be applied. In the non-parametric setting, the method described in [4] can be employed to estimate the conditional survival function:

$$\widehat{S}(t \mid X = x) = \prod_{\widetilde{T}_i \leq t} \left(1 - \frac{W(X_i, X, h)\Delta_i}{\sum_{j:\widetilde{T}_j \geq \widetilde{T}_i} W(X_j, X, h)}\right), \quad (5)$$

where $\widetilde{T}_i$ represents the imputed event time and Δ_i is the event indicator (taking the value 1 for observed events and 0 for censored observations). The weights $W(X_i, X, h)$ are determined using a kernel function to account for the continuous nature of the covariate.

While simple imputation methods are computationally convenient and easy to implement, they have significant limitations. The assumption that the event occurs exactly at L_i, R_i, or M_i can introduce bias, particularly when the censoring intervals are wide. Midpoint imputation is a common shortcut, but it assumes that the event could have happened at any time within the interval, which might not be true. Additionally, these methods do not properly account for the inherent uncertainty introduced by interval censoring. As a result, they may underestimate or overestimate survival probabilities, particularly when censoring patterns are highly variable across individuals.

3.2.4 Handling Interval Censoring with an Adapted EM Algorithm

For interval censoring, the Expectation-Maximization (EM) algorithm is adapted to incorporate kernel weights, ensuring that observations with similar covariate values contribute more significantly to the estimation process. This adaptation allows for a more flexible estimation of the conditional survival function in the presence of interval-censored data.

The algorithm begins with an initialization step, where an initial estimate of the probability masses s_j is required. A common approach is to assign uniform probability masses across the intervals or to use an empirical estimate based on midpoint imputation followed by kernel-smoothed estimation.

The kernel weights are incorporated into the Expectation-Maximization (EM) algorithm in the following way:

1. E-Step: Compute the expected number of observed events in each interval $[p_j, q_j)$ based on the current estimate of the probability masses. This is given by:

$$d'_j = \sum_{i=1}^{n} \left(W(X_i, X, h) \frac{\alpha_{ij} s_j}{\sum_{k=1}^{m} \alpha_{ik} s_k} \right). \quad (6)$$

Here, $W(X_i, X, h)$ represents a kernel weight function, which assigns greater influence to observations with covariate values close to X. The term α_{ij} is an indicator function that takes the value 1 if the interval $[p_j, q_j)$ is compatible with the censoring interval $[L_i, R_i)$ and 0 otherwise. The denominator ensures proper normalization by summing over all possible intervals.

2. M-Step: Update the probability masses s_j through normalization:

$$s_j^{\text{new}} = \frac{d'_j}{\sum_{i=1}^{n} W(X_i, X, h)}. \quad (7)$$

Since the EM algorithm is iterative, a stopping criterion is necessary to determine convergence. The most common approaches involve assessing the absolute difference in the estimated probability masses between successive iterations and terminating the algorithm when the maximum absolute change falls below a small threshold $\epsilon > 0$:

$$\max_j \left| s_j^{\text{new}} - s_j^{\text{old}} \right| < \epsilon.$$

Another alternative is to monitor the relative change in the probability masses, ensuring that the estimates stabilize within a predefined tolerance:

$$\max_j \frac{|s_j^{\text{new}} - s_j^{\text{old}}|}{s_j^{\text{old}} + \epsilon} < \tau.$$

In certain cases, evaluating the improvement in the log-likelihood function provides an effective criterion for stopping the iterations. The algorithm halts when the relative difference in log-likelihood values between consecutive iterations becomes negligibly small:

$$\frac{|L^{(t+1)} - L^{(t)}|}{|L^{(t)}|} < \delta.$$

We need to confirm that the updated probabilities still add up to 1. This is confirmed by computing:

$$\begin{aligned}
\sum_{j=1}^{m} s_j^{\text{new}} &= \sum_{j=1}^{m} \frac{d'_j}{\sum_{i=1}^{n} w_i} \\
&= \frac{1}{\sum_{i=1}^{n} w_i} \sum_{j=1}^{m} d'_j \\
&= \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \left(\sum_{j=1}^{m} \frac{\alpha_{ij} s_j}{\sum_{k=1}^{m} \alpha_{ik} s_k} \right) \\
&= \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i = 1.
\end{aligned}$$

In terms of computational considerations, efficiency can be improved through strategies such as Expectation Conditional Maximization (ECM) or Parameter-Expanded EM (PX-EM), which accelerate convergence while preserving the stability of the algorithm. The use of kernel smoothing techniques helps in mitigating numerical instability, particularly in cases where the data are sparse or the censoring intervals vary widely. Furthermore, parallel computation can be employed to enhance performance when dealing with large datasets, as the E-step involves summing over all observations.

By iteratively refining the probability mass estimates, the adapted EM algorithm provides a robust approach for estimating the conditional survival function under interval censoring. The incorporation of kernel weighting accounts for the continuous nature of the covariates, leading to a more accurate and flexible survival analysis framework.

3.2.5 Kaplan-Meier-Type Estimator for Conditional Survival under Interval Censoring

The Turnbull estimator can be seen as a generalization of the Kaplan-Meier (KM) estimator to interval-censored data, where the number of events occurring within each interval $[p_j, q_j)$ is determined using an iterative algorithm. Instead of assigning events to exact time points as in the standard Kaplan-Meier setting, the nonparametric maximum likelihood estimator (NPMLE) for interval-censored data assigns probability mass to discrete regions, ensuring that the survival function is piecewise constant outside these regions.

Given its structural similarity to the Kaplan-Meier estimator, an alternative method to estimate the conditional survival function under interval censoring consists of adapting the Kaplan-Meier formula by incorporating kernel weights to account for covariate effects. This approach relies on estimating the number of failures $d'_j(x)$ and the number of individuals at risk $n'_j(x)$ at each interval $[p_j, q_j)$, weighted by a function of the covariate X.

Estimation of Risk Set and Event Counts

To compute the survival probability at any given time t conditioned on a covariate value x, it is necessary to determine the quantities:

- $d'_j(x)$, the estimated number of failures occurring in interval $[p_j, q_j)$, conditioned on $X = x$.
- $n'_j(x)$, the estimated number of individuals at risk in the same interval.

These quantities can be obtained using kernel smoothing. The estimated number of failures within the interval $[p_j, q_j)$ is given by:

$$d'_j(x) = \sum_{i=1}^{n} W(X_i, x, h) d'_{ij},$$

where d'_{ij} represents the contribution of individual i to the number of failures in $[p_j, q_j)$, and $W(X_i, x, h)$ is a kernel weight function that assigns greater influence to observations with covariate values closer to x.

Since the exact failure times are not observed, the event count must be estimated using an expectation-based approach. Following Turnbull's framework, the expected number of failures in the interval $[p_j, q_j)$ for individual i is computed as:

$$d'_{ij} = \frac{\alpha_{ij} s_j}{\sum_{k=1}^{m} \alpha_{ik} s_k}, \tag{8}$$

where α_{ij} is an indicator function that equals 1 if the interval $[p_j, q_j)$ is compatible with the censoring interval $[L_i, R_i)$, and 0 otherwise; s_j represents the estimated probability mass assigned to the interval $[p_j, q_j)$. The denominator ensures that the estimated failures are properly distributed across the possible intervals for each individual.

Similarly, the number of individuals at risk in $[p_j, q_j)$, conditioned on $X = x$, is given by:

$$n'_j(x) = \sum_{i=1}^{n} W(X_i, x, h) n'_{ij},$$

where n'_{ij} represents the number of individuals at risk in the interval $[p_j, q_j)$. Since individuals may have uncertain failure times due to interval censoring, their contribution to the risk set is also estimated in a weighted manner. A reasonable estimator is:

$$n'_{ij} = \sum_{k=1}^{n} \alpha_{kj} w_k,$$

where α_{kj} is the same indicator function used in the estimation of d'_{ij}, w_k represents the weight assigned to individual k, accounting for their presence in the risk set at interval $[p_j, q_j)$.

The n'_{ij} can be computer using an alternative approach:

$$n'_j(x) = \sum_{k=j}^{m} \sum_{i=1}^{n} W(X_i, x, h) \frac{\alpha_{ik} s_k}{\sum_{r=1}^{m} \alpha_{ir} s_r}. \quad (9)$$

This later expression for the n'_{ij} accounts for the presence of an individual in the risk set of the interval $[p_j, q_j)$. The kernel weight $W(X_i, x, h)$ ensures that individuals whose covariate values X_i are close to x have a stronger influence on the estimation.

By computing d_{ij} and n'_{ij} for all intervals, it is possible to estimate the conditional survival function using the adapted Kaplan-Meier-type estimator. These estimates ensure that survival probabilities are correctly adjusted for the effects of interval censoring.

Using these quantities, the conditional survival function can be estimated as:

$$\hat{S}(t \mid X = x) = \prod_{j:p_j \le t} \left(1 - \frac{d'_j(x)}{n'_j(x)}\right). \quad (10)$$

This formula is analogous to the Kaplan-Meier estimator, adapted to the interval-censoring setting. The key distinction is that instead of assigning events to precise time points, events are distributed across interval regions, and the contribution of each observation is weighted according to the covariate structure.

3.2.6 Comparison with the EM-Based Approach

This Kaplan-Meier-type estimator differs from the kernel-weighted EM algorithm previously described in that it directly estimates survival probabilities using a recursive product-limit formula, rather than iteratively refining event probabilities. Some advantages of this approach include:
- A simpler computational framework, since it does not require the iterative maximization of likelihood functions.
- A direct interpretation similar to the Kaplan-Meier estimator, making it easier to understand and implement.
- Computational efficiency in cases where the kernel-weighted EM approach is slow to converge.

However, it also presents some challenges:
- Sensitivity to bandwidth selection: The accuracy of the estimator depends heavily on the choice of the kernel bandwidth h.
- Potential bias when censoring intervals are wide, as the method does not explicitly model the uncertainty within the censoring regions.
- Loss of efficiency compared to EM methods, particularly when the true failure time distribution inside each interval is highly skewed.

3.2.7 Implementation Considerations

The practical implementation of the two estimators requires careful selection of the kernel function and bandwidth parameter. A commonly used choice is the Gaussian kernel:

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right), \quad K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \quad (11)$$

where h is the bandwidth controlling the degree of smoothing. Alternative kernel functions, such as the Epanechnikov or triangular kernel, may also be considered depending on the data structure.

One aspect to consider in the Kaplan-Meier-type estimator is the handling of ties, particularly when multiple individuals share the same censoring intervals. If ties are frequent, the estimation of $d'_j(x)$ and $n'_j(x)$ should be adjusted to avoid overestimation of risk.

This Kaplan-Meier-type estimator provides an intuitive and computationally efficient alternative for estimating the conditional survival function under interval censoring. While it lacks some of the theoretical rigor of likelihood-based approaches such as the EM algorithm, it remains a useful option in practical settings where computational simplicity and interpretability are important.

4 Conclusion

This paper addressed the problem of estimating the conditional survival function in the presence of interval-censored data, a setting that presents unique challenges when covariates are continuous. While Turnbull's estimator remains appropriate for qualitative covariates, it does not extend naturally to the continuous case. To overcome this limitation, we introduce two nonparametric methodologies that integrate kernel-weighting techniques with adaptations of the EM algorithm and the Kaplan-Meier estimator. These approaches offer flexible and practical alternatives for analyzing complex censoring structures without relying on strong distributional assumptions.

One limitation inherent to kernel-based nonparametric methods is the so-called curse of dimensionality, which leads to a rapid degradation of performance as the number of covariates increases. In the context of right-censored data, this issue can be addressed by adopting parametric or semiparametric approaches, such as the Cox proportional hazards model, where the conditional survival function can be estimated via the estimator in [12]. However, such models rely on assumptions (e.g., proportional hazards [13]) that may not hold in practice.

Nonparametric methods, on the other hand, offer greater flexibility and robustness, especially when the relationship between covariates and survival times is complex or unknown. Among nonparametric alternatives, single-index models provide a promising direction to mitigate the curse of dimensionality by reducing the multivariate covariate information to a single linear index. Despite their advantages, current developments in single-index models [14] are limited to the setting of right-censored data.

In this work, we chose to emphasize fully nonparametric approaches adapted to interval-censored data, highlighting their interpretability, minimal assumptions, and practical feasibility. Future work may explore the extension of single-index methodologies and other dimension-reduction techniques to interval censoring, combining the flexibility of nonparametric estimation with improved scalability in higher dimensions.

One limitation of this work is the lack of empirical validation using real-world data or simulation experiments. Although the proposed methodologies are developed with practical applications in mind, their comparative performance in terms of estimation accuracy and computational behavior has not yet been assessed. The primary objective of this paper is to introduce and describe the estimation procedures in detail. Future work will include simulation studies under a variety of censoring patterns and covariate settings, as well as applications to real datasets in biomedical and reliability contexts to illustrate the practical usefulness of the proposed approaches.

Acknowledgment:
This work is funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the UID/00013: Centro de Matemática da Universidade do Minho (CMAT/UM) Program Contract, and the project reference 2023.14897.PEX (DOI: 10.54499/2023.14897.PEX).

Declaration of Generative AI and AI-assisted Technologies in the Writing Process
The author wrote, reviewed and edited the content as needed and verifies that none utilised artificial intelligence (AI) tools were used. The author takes full responsibility for the content of the publication.

References:

[1] E.L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 457–481.

[2] J. Sun, *The Statistical Analysis of Interval-Censored Failure Time Data*, Springer, New York, 2006.

[3] B.W. Turnbull, "The empirical distribution function with arbitrarily grouped, censored, and truncated data," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 38, No. 3, 1976, pp. 290–295.

[4] R. Beran, "Nonparametric Regression with Randomly Censored Survival Data," *Technical Report*, University of California, Berkeley, 1981.

[5] M. H. Dehghan and T. Duchesne, "A generalization of Turnbull's estimator for nonparametric estimation of the conditional survival function with interval-censored data," *Lifetime Data Analysis*, vol. 17, pp. 234–255, 2011.

[6] M. van der Laan and A. Hubbard, "Estimation with interval-censored data and covariates," *Lifetime Data Analysis*, vol. 3, no. 1, pp. 77–91, 1997.

[7] E.A. Nadaraya, "On nonparametric estimates of density functions and regression curves," *Theory of Applied Probability*, Vol. 10, 1965, pp. 186–190.

[8] G.S. Watson, "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, Vol. 26, No. 4, 1964, pp. 359–372.

[9] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, CRC Press, 1996.

[10] M.P. Wand and M.C. Jones, *Kernel Smoothing*, Chapman & Hall, London, 1994.

[11] J.M. Robins, A. Rotnitzky, and L.P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, Vol. 89, No. 427, 1992, pp. 846–866.

[12] N.E. Breslow, “Discussion of the paper by D.R. Cox,” *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 34, No. 2, 1972, pp. 216–217.

[13] D.R. Cox, “Regression models and life tables (with discussion),” *Journal of the Royal Statistical Society: Series B*, Vol. 34, 1972, pp. 187–220.

[14] E. Strzalkowska-Kominiak and R. Cao, “Beran-based approach for single-index models under censoring,” *Computational Statistics*, Vol. 29, No. 5, 2014, pp. 1243–1261.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The author contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work is funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the UID/00013: Centro de Matemática da Universidade do Minho (CMAT/UM) Program Contract, and the project reference 2023.14897.PEX (DOI: 10.54499/2023.14897.PEX).

Conflict of Interest

The author has no conflict of interest to declare that is relevant to the content of this article.