computer sciences and mathematics forum

Proceeding Paper

Estimation of the Conditional Hazard Function with a Recursive Kernel from Censored Functional Ergodic Data †

Hadjer Kebir * and Boubaker Mechab

Laboratory of Statistic and Processus Stochastic, University of Djillali Liabes, L.P 89, Sidi Bel Abbes 22000, Algeria; mechaboub@yahoo.fr

* Correspondence: hadjer_kebir@yahoo.com

† Presented at the 1st International Online Conference on Mathematics and Applications, 1–15 May 2023; Available online: https://iocma2023.sciforum.net/.

Abstract: In this paper, we propose a non-parametric estimator of the conditional hazard function weighted on the recursive kernel method given an explanatory variable taking values in a semi-metric space when the scalar response is censored. Under the ergodicity condition, we establish the convergence rate of this estimator.

Keywords: conditional hazard function; censored data; functional ergodic data; recursive kernel estimate

Citation: Kebir, H.; Mechab, B. Estimation of the Conditional Hazard Function with a Recursive Kernel from Censored Functional Ergodic Data. *Comput. Sci. Math. Forum* **2023**, 7, 16. https://doi.org/10.3390/IOCMA2023-14412

Academic Editor: Antonio Di Crescenzo

Published: 28 April 2023

1. Introduction

The functional estimate has been a topic of great interest in the statistical literature. To obtain a summary of the current state of non-parametric functional data, we refer to the work in [1,2]. The hazard function, also known as the risk function, is a concept commonly used in survival analysis and reliability theory. It plays an important role in statistics and arises in a variety of fields, including econometrics, epidemiology, environmental science, and many others. The work in [3] is an important contribution to the conditional hazard rate for functional covariates in an infinite-dimensional space. Censored data are a type of data in which the values are incomplete or partially known. We consider a type of right-censored data, where the observation is known to be above a certain threshold, but the exact value is unknown. For example, if we are studying the time until a light bulb fails, we might know that the bulb lasted at least 500 h, but we do not know exactly how long it lasted beyond that. Ergodic data have represented a rising interest in this domain over the past few years. It is an essential postulate in statistical physics for analyzing the thermodynamic characteristics of gases, atoms, electrons, or plasmas. Ergodic theory enables us to circumvent intricate probabilistic computations related to the mixing condition. In our setting, we study the almost sure convergence of the kernel estimator of the conditional hazard function; we consider a recursive estimate with strictly stationary, censored, and ergodic observations. It is worth noting that the recursive estimate has a benefit in that the smoothing parameter is tied to the observation (X_i, Y_i), which enables us to continuously update our estimator as we receive new observations. To combine censored data and ergodic theory, we refer to the work in [4]. They estimated the conditional quantile using censored and ergodic data.

2. Materials and Methods

In practice, it is possible to coincide with censored variables, that is, instead of observing the lifetimes, we observe the censored lifetimes. This problem is usually modeled by considering $T_1, ..., T_n$, a sequence of lifetimes which satisfy some kind of dependency, and $C_1, ... C_n$, which is a sequence of i.i.d censored random variables with a common unknown continuous distribution function G. We observe only the n pairs (Y_i, δ_i), where $Y_i = min\{T_i, C_i\}$ and $\delta_i = \mathbb{I}_{\{T_i \le C_i\}}$, $1 \le i \le n$, where $\mathbb{I}_A$ denotes the indicator function of the set A. To ensure the identifiability of the model, we assume that C_i and (T_i, X_i) $(1 \leq i \leq n)$ are independent. Let $(X_i, T_i)_{i=1,\dots,n}$ be a sequence of strictly stationary ergodic processes with the same distribution. We also assume X_i takes values in a semi-metric space $(\mathcal{F}, d)$, whereas T_i are real-valued random variables. In addition, for insuring good mathematical properties of the functional nonparametric methods, we establish our asymptotic results on the concentration properties of small balls of the probability measure of the functional variable.

We define the function hazard h^x for $y \in \mathbb{R}$ and $F^x(y) < 1$ by

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}.$$

To this aim, we first introduce the recursive double-kernel-type pseudo-estimator $\tilde{F}^x$ of the conditional distribution function F^x defined by

$$\tilde{F}^x(t) = \frac{\displaystyle\sum_{i=1}^{n} \frac{\delta_i}{\bar{G}(Y_i)} K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(t - Y_i))}{\displaystyle\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}, \qquad \forall t \in \mathbb{R}.$$

where K is the kernel, H is a strictly increasing distribution function, a_i, b_i are sequences of positive real numbers such that $\lim_{n \to +\infty} a_n = \lim_{n \to +\infty} b_n = 0$, and $\bar{G}(.) = 1 - G(.)$.

From this pseudo-estimator, we deduce a pseudo-estimator $\tilde{f}^x$ of the conditional density f^x by

$$\tilde{f}^x(t) = \frac{\displaystyle\sum_{i=1}^{n} \frac{\delta_i}{\bar{G}(Y_i)} b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(t - Y_i))}{\displaystyle\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}, \qquad \forall t \in \mathbb{R}.$$

where H' is the derivative of H.

In practice, G is unknown, and one can estimate it using the Kaplan and Meier (1958) estimate $\bar{G}_n(.)$, defined as:

$$\bar{G}_n(t) = \begin{cases} \displaystyle\prod_{i=1}^{n} \left(1 - \frac{1 - \delta_{(i)}}{n - i + 1}\right)^{\mathbb{I}_{\{Y_{(i)} \leq t\}}} & \text{if } t < Y_{(n)}, \\ 0 & \text{Otherwise.} \end{cases}$$

where $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ are the order statistics of $(Y_i)_{1 \leq i \leq n}$ and $\delta_{(i)}$ is concomitant with $Y_{(i)}$.

Thus, the feasible estimator of the conditional distribution function $F^x(t)$ is given by

$$\widehat{F}^x(t) = \frac{\displaystyle\sum_{i=1}^{n} \frac{\delta_i}{\bar{G}_n(Y_i)} K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(t - Y_i))}{\displaystyle\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}, \qquad \forall t \in \mathbb{R}.$$

We deduce an estimator for a conditional density $f^x(t)$, defined as

$$\widehat{f}^x(t) = \frac{\displaystyle\sum_{i=1}^{n} \frac{\delta_i}{\bar{G}_n(Y_i)} b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(t - Y_i))}{\displaystyle\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}, \qquad \forall t \in \mathbb{R}.$$

We estimate the conditional hazard function $\widehat{h}^x$ by

$$\widehat{h}^x(t) = \frac{\widehat{f}^x(t)}{1 - \widehat{F}^x(t)}. \quad \forall t \in \mathbb{R}.$$

Remark 1. *The Kaplan–Meier estimator is not recursive and the use of such an estimator can slightly penalize the efficiency of our estimator in terms of computational time.*

3. Results

To establish the almost sure convergence of $\widehat{h}^x$, we need to include the following assumptions:

Assumptions 1.
(H1)

$$\begin{cases} \text{(i) The function } \phi(x,h) := \mathbb{P}(X \in B(x,h)) > 0, \forall h > 0. \\ \text{(ii) For all } i = 1,\dots,n \text{ there exists a deterministic function } \phi_i(x,.) \text{ such that almost surely} \\ 0 < \mathbb{P}(X_i \in B(x,h)|\mathcal{F}_{i-1}) \le \phi_i(x,h), \forall h > 0. \\ \text{and } \phi_i(x,h) \to 0 \text{ as } h \to 0. \\ \text{(iii) For all sequences } (h_i)_{i=1,\dots,n} > 0,\ \dfrac{\sum_{i=1}^{n} \mathbb{P}(X_i \in B(x,h_i)|\mathcal{F}_{i-1})}{\sum_{i=1}^{n} \phi(x,h_i)} \to 1 \end{cases}$$

where $B(x,h) := \{x' \in \mathcal{F}/d(x',x) < h\}$.

(H2) (i) The conditional distribution function F^x *is such that* $\forall t \in \mathcal{S}$, $\exists \beta > 0$, $\inf_{t\in\mathcal{S}}(1 - F^x(t)) > \beta$, $\forall$ $(t_1,t_2) \in \mathcal{S} \times \mathcal{S}$, $\forall$ $(x_1,x_2) \in N_x \times N_x$,

$$|F^{x_1}(y_1) - F^{x_2}(y_2)| \le C_1(d(x_1,x_2)^{\beta_1} + |t_1 - t_2|^{\beta_2}),\ \beta_1 > 0,\ \beta_2 > 0.$$

(ii) The density f^x *is such that* $\forall t \in \mathcal{S}$, $\exists \alpha > 0$, $f^x(t) < \alpha$, $\forall$ $(t_1,t_2) \in \mathcal{S} \times \mathcal{S}$, $\forall$ $(x_1,x_2) \in N_x \times N_x$,

$$|f^{x_1}(t_1) - f^{x_2}(t_2)| \le C_1(d(x_1,x_2)^{\beta_1} + |t_1 - t_2|^{\beta_2}),\ \beta_1 > 0,\ \beta_2 > 0.$$

(H3) $\forall (t_1,t_2) \in \mathbb{R}^2, |H^{(j)}(t_1) - H^{(j)}(t_2)| \le C|t_1 - t_2|$ *for* $j = 0,1$.
$\int |w|^{\beta_2} H^{(1)}(w)dw < \infty$, $\int H'^2(t)dw < \infty$.

(H4) K is a function with support (0,1) such that $0 < C_1\mathbb{I}_{[0,1]} < K(t) < C_2\mathbb{I}_{[0,1]} < \infty$, *where* $\mathbb{I}_A$ *is the indicator function.*

(H5) (i) $\lim_{n\to+\infty} n^{-1} \sum_{i=1}^{n} \dfrac{a_i^{\beta_1}\phi_i(x,a_i)}{\phi(x,a_i)} = 0$, *(ii)* $\lim_{n\to+\infty} n^{-1} \sum_{i=1}^{n} \dfrac{b_i^{\beta_2}\phi_i(x,a_i)}{\phi(x,a_i)} = 0$.

(H6) $\lim_{n\to+\infty} \dfrac{\varphi_{n,j}(x)\log n}{n^2} = 0$, *where* $\varphi_{n,j}(x) = \sum_{i=1}^{n} \dfrac{b_i^{-j}\phi_i(x,a_i)}{\phi^2(x,a_i)}$ *for* $j = 0,1$.

(H7) $(C_n)_{n\ge1}$ *and* $(X_n,T_n)_{n\ge1}$ *are independent.*

(H8) G has a bounded first derivative $G^{(1)}$.

Theorem 1. *Under hypotheses (H1)–(H7), we have:*
$\sup_{t\in\mathcal{S}} |\widehat{h}^x(t) - h^x(t)| =$

$$O\left(n^{-1}\sum_{i=1}^{n} \frac{a_i^{\beta_1}\phi_i(x,a_i)}{\phi(x,a_i)}\right) + O\left(n^{-1}\sum_{i=1}^{n} \frac{b_i^{\beta_2}\phi_i(x,a_i)}{\phi(x,a_i)}\right) + O\left(\sqrt{\frac{\varphi_{n,1}(x)\log n}{n^2}}\right) \text{ a.s.} \tag{1}$$

Proof of Theorem 1. The proof of this theorem is based on the following decomposition and lemmas below:

$$\widehat{h}^x(t) - h^x(t) = \frac{1}{1-\widehat{F}^x(t)}\left[\widehat{f}^x(t) - f^x(t)\right] + \frac{h^x(t)}{1-\widehat{F}^x(t)}\left[\widehat{F}^x(t) - F^x(t)\right]. \tag{2}$$

□

Lemma 1. *Under hypotheses (H1), (H2)(i) and (H3)–(H7), we have:*
$\sup_{t\in\mathcal{S}} |\widehat{F}^x(t) - F^x(t)| =$

$$O\left(n^{-1}\sum_{i=1}^{n}\frac{a_i^{\beta_1}\phi_i(x,a_i)}{\phi(x,a_i)}\right) + O\left(n^{-1}\sum_{i=1}^{n}\frac{b_i^{\beta_2}\phi_i(x,a_i)}{\phi(x,a_i)}\right) + O\left(\sqrt{\frac{\varphi_{n,0}(x)\log n}{n^2}}\right)\ a.s. \tag{3}$$

Lemma 2. *Under hypotheses (H1), (H2)(ii) and (H3)–(H7), we have:*
$\sup_{t\in\mathcal{S}} |\widehat{f}^x(t) - f^x(t)| =$

$$O\left(n^{-1}\sum_{i=1}^{n}\frac{a_i^{\beta_1}\phi_i(x,a_i)}{\phi(x,a_i)}\right) + O\left(n^{-1}\sum_{i=1}^{n}\frac{b_i^{\beta_2}\phi_i(x,a_i)}{\phi(x,a_i)}\right) + O\left(\sqrt{\frac{\varphi_{n,1}(x)\log n}{n^2}}\right)\ a.s. \tag{4}$$

Lemma 3. *Under the hypotheses of Lemma 1, we have:*

$$\exists\, \delta > 0 \ \text{ such that } \sum_{n=1}^{\infty}\mathbb{P}\left\{\inf_{t\in\mathcal{S}}|1-\widehat{F}^x(t)| \le \delta\right\} < \infty. \tag{5}$$

4. Discussion

This contribution concerns a recursive nonparametric estimation of the conditional hazard function in the presence of a functional explanatory variable when the scalar response is right censored in the ergodic case. As asymptotic results, we have established the almost sure convergence. Concerning the assumptions, they can be divided into three categories: structural assumptions, assumptions of the explanatory variable, and technical assumptions.

Author Contributions: Both authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: The work received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Springer Science+Business Media, Inc.: Berlin/Heidelberg, Germany, 2006.
2. Ramsay, J.O.; Silverman, B.W. *Applied Functional Data Analysis*; Springer Series in Statistics; Springer: New York, NY, USA, 2002; Volume 1.

3. Ferraty, F.; Rabhi, A.; Vieu, P. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Rev. Math. Pures Appl.* **2008**, *53*, 1–18.
4. Chaouch, M.; Khardani, S. Randomly censored quantile regression estimation unsing functional stationary ergodic data. *J. Nonparametr. Stat.* **2015**, *27*, 65–87. [CrossRef]