Nonparametric conditional hazard estimation with ergodic data: a recursive kernel approach

H. Kebir[1]

[1]Laboratory of statistics and stochastic processes Djillali Liabes University, Sidi Bel Abbes, Algeria

hadjer_kebir@yahoo.com

We propose a non-parametric estimator of the conditional hazard function weighted on the recursive kernel method given a explantory variable taking values in a semi-metric space and a scalar response. The functional estimate has attracted a lot of attention in the statistical literature. For an overview of the present state on nonparametric functional data (FDA), we refer to the works [1] and [2]. The hazard function, also known as the risk function, is a concept commonly used in survival analysis and reliability theory. One of the important work about the conditional hazard rate in infinite dimensional space for functional covariates is [3]. Under ergodicity condition, we establish the almost complete convergence rate of our estimator.

Let $(X_i, Y_i)_{i=1,\dots,n}$ be a sequence of strictly stationary ergodic processes. We also assume X_i take values on a semi-metric space $(\mathcal{F}, d)$ whereas Y_i are real-valued random variables. We suppose that the conditional cdf F^x of Y_i given $X_i = x$ has a continuous density f^x with respect to Lebesgue's measure on $\mathbb{R}$. We define the hazard function h^x, for $y \in \mathbb{R}$ and $F^x(y) < 1$, by

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

The recursive double kernels type estimator $\widehat{F}^x$ of F^x is defined by

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i)) H(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}, \qquad y \in \mathbb{R},$$

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

We define the recursive double kernels estimator $\widehat{f}^x$ of f^x by

$$\widehat{f}^x(y) = \frac{\sum_{i=1}^{n} b_i^{-1} K(a_i^{-1} d(x, X_i)) H'(b_i^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(a_i^{-1} d(x, X_i))}, \qquad y \in \mathbb{R},$$

where H' is the derivative of H.

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

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

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

(H1) $\phi(x,h) := \mathbb{P}(X \in B(x,h)) > 0, \forall h > 0$. where $B(x,h) := \{x' \in \mathcal{F}/d(x',x) < h\}$.

(H2) (i) $|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C_1(d(x_1,x_2)^{\beta_1} + |y_1 - y_2|^{\beta_2}), \ \beta_1 > 0, \ \beta_2 > 0.$
(ii) $|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_1(d(x_1,x_2)^{\beta_1} + |y_1 - y_2|^{\beta_2}), \ \beta_1 > 0, \ \beta_2 > 0.$

(H3) $\forall (y_1,y_2) \in \mathbb{R}^2, |H^{(j)}(y_1) - H^{(j)}(y_2)| \leq C|y_1 - y_2|$, for $j = 0, 1$.
$\int |t|^{\beta_2} H^{(1)}(t)dt < \infty, \int H'^2(t)dt < \infty.$

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

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

Theorem 1. *Under hypotheses (H1)-(H4), we have:*
$\sup_{y\in\mathcal{S}} |\widehat{h}^x(y) - h^x(y)| =$

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

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

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

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

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

1. Ferraty F., Vieu P. Nonparametric functional data analysis. —Springer Science+Business Media, Inc, 2006, xx+260 pp.
2. Ramsay J. O., Silverman B. W. Applied functional data analysis. —Springer series in statistics, 2002, 201 pp.
3. Ferraty F., Rabhi A., Vieu P. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. Revue de Mathématiques Pures et Appliquées, 2008, 53, 1–18.