

A generalized conditional Kaplan-Meier estimator for uncertain or unknown censorship.

Mikael ESCOBAR-BACH
Université d'Angers

Olivier GOUDET
Université d'Angers

February 24, 2020

Abstract

Along with the analysis of time-to-event data, it is common to assume that only a partial information is given at hand. In the presence of *right-censored* data with covariates, the conditional Kaplan-Meier estimator (also referred as the Beran estimator) is known to propose a consistent estimate for the lifetimes conditional survival function. However, a necessary condition is the clear knowledge of whether each individual is censored or not. In this paper, we consider the natural problem of the censorship uncertainty, that is, when the censoring indicator is reduced to a random indicator $P \in [0, 1]$ that characterizes the degree of certainty. We propose a new estimator for the conditional survival function adapted from the conditional Kaplan-Meier estimator and establish its asymptotic normality. Based on this, we further study the supervised learning problem where the conditional survival function is to be predicted with no censorship indicators. To this aim, we investigate various approaches estimating the conditional expectation for P . Along with the theoretical results, we illustrate how the estimators work for small samples by means of a simulation study and show their practical applicability with the analysis of patient with breast cancer.

Key Words: survival analysis; covariates; right-censoring; survival function; supervised regression.

1 Introduction

As the volume of data expands, the problem of missing/censoring data has been increasingly present in many fields of statistical applications. In the classical literature of survival analysis, the study of the duration time preceding an event of interest is considered with series of random censors, that might prevent the capture of the whole survival time. This feature is known as *censoring* and arises from restrictions that depend from the nature of the study. For instance, this feature is particularly present in medicine with studies of the survival times before the recovery/decease from a specific chronic disease. Indeed, a patient lifetime or time to occurrence will not continue if it exceeds the study follow-up time, or because of the early withdrawal of the individual from the population under study. This is denoted as a *right-censored* observation time, indicating that the event of interest might only occur beyond and represents the most commonly studied censoring type. One can find other examples of right-censoring in a large scope of domains, like in economic (unemployment duration time), social (spending time before marriage or child-bearing) or in actuarial science (life-insurance or time before claim).

We consider the random vector $(Y, C, P, X) \in \mathbb{R} \times \mathbb{R} \times [0, 1] \times \mathbb{R}^p$ and assume that we only observe (T, P, X) where $T := \min(Y, C)$, P is a the partial censorship indicator and X an explanatory random covariable. Intuitively, P stands for the censorship information on the couple (Y, C) , that is greater is P and the more we suspect $Y \leq C$ while P small highly suggests that $Y > C$. Formally, we assume in the sequel that $\forall t \in \mathcal{S}_T \subset \mathbb{R}$ and $\forall x \in \mathcal{S}_X \subset \mathbb{R}^p$

$$\mathbb{E}[P|T = t, X = x] = \mathbb{P}(T \leq C|T = t, X = x) \quad (1.1)$$

where \mathcal{S}_T and \mathcal{S}_X are respectively the support for the observations T and X . Note that one can retrieve the classical censorship model when $P = \delta := \mathbb{1}_{\{Y \leq C\}}$. The conditional distribution functions of the survival and the censoring times are respectively denoted by F and G . Recall also that the right end points of the support of the distributions F and G are respectively denoted by $\tau_S(x)$ and $\tau_C(x)$. Finally, we denote F^\leftarrow the generalized inverse function for the distribution function F given by

$$F^\leftarrow(\alpha|x) := \inf\{t, F(t|x) \geq \alpha\}, \quad \alpha \in (0, 1).$$

We will work under minimal conditions on the distribution functions, though we have to impose the usual identification assumption that T and C are independent, which implies that $H(t|x) := \mathbb{P}(T \leq t|X = x)$ satisfies $1 - H(t|x) = (1 - F(t|x))(1 - G(t|x))$. Here, the right end point for the support of H_x is denoted $\tau_H(x) = \min(\tau_C(x), \tau_S(x))$. In the sequel, we will also use the notations

$$H^u(t|x) = \mathbb{P}(T \leq t, \delta = 1|X = x) = \int_{-\infty}^t (1 - G(s^-|x))dF(s|x)$$

for the sub-distribution function of the uncensored observations and $\Lambda(\cdot|x)$ for the cumulative hazard function given by

$$\Lambda(t|x) = \int_{-\infty}^t \frac{dH^u(s|x)}{1 - H(s^-|x)}.$$

2 Conditional distribution function estimators

The conditional Kaplan-Meier estimator or the so-called Beran estimator of the distribution F is defined as follows. Let assume an independent and identically distributed (i.i.d.) n -sized sample drawn from the classical censoring model and denoted $\{(T_i, \delta_i, X_i)\}_{1 \leq i \leq n}$. Denote the i -th order statistic of T_1, \dots, T_n by $T_{(i)}$, and denote the corresponding censoring indicator and covariable by $\delta_{(i)}$ and $X_{(i)}$ respectively. In the absence of ties, the Beran estimator is given by

$$F_n(t|x) = 1 - \prod_{T_{(i)} \leq t} \left(1 - \frac{W_h(x - X_{(i)})}{1 - \sum_{j=1}^{i-1} W_h(x - X_{(j)})} \right)^{\delta_{(i)}},$$

where for any $i = 1, \dots, n$,

$$W_h(x - X_i) = \frac{K_h(x - X_i)}{\sum_{j=1}^n K_h(x - X_j)},$$

and $K_h(\cdot) = K(\cdot/h)/h^p$ with K a kernel function and $h = h_n$ a non-random positive sequence such that $h_n \rightarrow 0$ as $n \rightarrow \infty$.

Adapted to our context, we assume from now on that we only observe the i.i.d. n -sized sample $\{(Y_i, P_i, X_i)\}_{1 \leq i \leq n}$ drawn the triplet (T, C, P, X) . The new estimator of the distribution function F_x is similarly defined as

$$\hat{F}_n(t|x) = 1 - \prod_{T_{(i)} \leq t} \left(1 - \frac{W_h(x - X_{(i)})}{1 - \sum_{j=1}^{i-1} W_h(x - X_{(j)})} \right)^{P_{(i)}},$$

where the product over an empty set is defined to be 1. In order to study further study our new estimator, we introduce the empirical estimators for the distribution function H , H_u and Λ respectively given by

$$H_n(t|x) = \sum_{i=1}^n W_h(x - X_i) \mathbb{1}_{\{T_i \leq t\}},$$

$$H_n^u(t|x) = \sum_{i=1}^n W_h(x - X_i) \mathbb{1}_{\{T_i \leq t, \delta_i = 1\}}, \quad \hat{H}_n^u(t|x) = \sum_{i=1}^n W_h(x - X_i) \mathbb{1}_{\{T_i \leq t\}} P_i$$

and

$$\Lambda_n(t|x) = \int_{-\infty}^t \frac{dH_n^u(s|x)}{1 - H_n(s^-|x)}, \quad \hat{\Lambda}_n(t|x) = \int_{-\infty}^t \frac{d\hat{H}_n^u(s|x)}{1 - H_n(s^-|x)}.$$

We finally denote f_n as the kernel estimator for the density function f with $\forall x \in \mathbb{R}^p$

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i).$$

2.1 Asymptotic properties

In this section, we study the asymptotic properties of the estimator $\hat{F}_n(t|x)$. We also assume from now on that $x \in S_X$ defines a fixed reference position. Due to the regression context, we need some Hölder-type conditions on the distribution functions H , H^u and F and on the density function f of the covariate X . Let $\|\cdot\|$ be the Euclidean norm in \mathbb{R}^p .

Assumption (H). There exist $0 < \eta, \eta' \leq 1$ and $c > 0$ such that for any $t, s \in \mathbb{R}$ and any $x_1, x_2 \in S_X$,

- (H.1) $|f(x_1) - f(x_2)| \leq c\|x_1 - x_2\|^\eta$,
- (H.2) $|H(t|x_1) - H(t|x_2)| \leq c\|x_1 - x_2\|^\eta$,
- (H.3) $|H^u(t|x_1) - H^u(s|x_2)| \leq c(\|x_1 - x_2\|^\eta + |t - s|^{\eta'})$,
- (H.4) $|F(t|x) - F(s|x)| \leq c|t - s|^{\eta'}$.

Also, some common assumptions on the kernel function need to be imposed.

Assumption (K). Let K be a bounded density function in \mathbb{R}^p with support S_K included in the unit ball of \mathbb{R}^p with respect to the euclidean norm.

Assumption (A). The distribution functions $F(\cdot|x)$ and $G(\cdot|x)$ are continuous on $(-\infty, T]$.

For convenient reasons, we study the asymptotic behaviour for our estimator through $\hat{\Lambda}_n$ with the approximation $\hat{F}_n \approx \exp(-\hat{\Lambda}_n)$ as given in the following lemma

Lemma 2.1 *Let $\tau_0 < \tau_1 < \tau_H(x)$. Then we have almost surely that*

$$\begin{aligned} \sup_{t \in [\tau_0, \tau_1]} \left| 1 - \hat{F}_n(t|x) - \exp(-\hat{\Lambda}_n(t|x)) \right| &\leq \|K\|_\infty \frac{\hat{H}_n^u(T|x)}{nh^p f_n(x) (1 - H_n(\tau_1|x))^2} \\ &= \mathcal{O}_{\mathbb{P}}((nh^p)^{-1}). \end{aligned}$$

In particular, it turns out that

$$\sup_{t \in [\tau_0, \tau_1]} \sqrt{nh^p} \left| 1 - \hat{F}_n(t|x) - \exp(-\hat{\Lambda}_n(t|x)) \right| = o_{\mathbb{P}}(1)$$

when $f(x) > 0$.

This result allows us to focus our study on $\hat{\Lambda}_n$ instead of \hat{F}_n . We hereby state the almost sure representation for our new estimator.

Theorem 2.1 Under the assumptions (\mathcal{H}) , (\mathcal{K}) and (\mathcal{A}) , for any $\tau_0 < \tau_1 < \tau_H(x)$, we have for $\tau_0 \leq t \leq \tau_1$ and $nh^{2\eta+p}|\log h| = \mathcal{O}(1)$

$$\hat{F}_n(t|x) - F(t|x) = \sum_{i=1}^n W_h(x - X_i) \hat{g}(t, T_i, \delta_i, P_i|x) + r_n(t|x)$$

where

$$\begin{aligned} \hat{g}(t, T_i, \delta_i, P_i|x) &= (1 - F(t|x)) \left\{ \int_{-\infty}^t \frac{\mathbb{1}_{\{T_i < s\}} - H(s|x)}{(1 - H(s|x))^2} dH^u(s|x) + \frac{\mathbb{1}_{\{T_i \leq t, \delta_i = 1\}} - H^u(t|x)}{1 - H(t|x)} \right. \\ &\quad \left. - \int_{-\infty}^t \frac{\mathbb{1}_{\{T_i \leq s, \delta_i = 1\}} - H^u(s|x)}{(1 - H(s|x))^2} dH(s|x) + \frac{(P_i - \delta_i) \mathbb{1}_{\{T_i \leq t\}}}{1 - H(T_i|x)} \right\} \end{aligned}$$

and

$$\sup_{\tau_0 \leq t \leq \tau_1} |r_n(t|x)| = \mathcal{O}_{\mathbb{P}}((nh^p)^{-3/4} |\log h|^{3/4}).$$

This allows us to obtain the main result of this subsection, which is the weak convergence of the estimator \hat{F}_n as a process in $\ell^\infty[\tau_0, \tau_1]$ for any $\tau_1 < \tau_H(x)$ and for fixed x . Here, for any set S , the space $\ell^\infty(S)$ is the space of bounded functions defined on S endowed with the uniform norm.

Theorem 2.2 Assume (\mathcal{A}) , (\mathcal{H}) and (\mathcal{K}) , and assume that $f(x) > 0$, $nh^p |\log h|^{-3} \rightarrow \infty$ and $nh^{2\eta+p-q} |\log h|^{-1} = \mathcal{O}(1)$ for some $q > 0$. Then, for any $\tau_0 < \tau_1 < \tau_H(x)$, the process

$$\left\{ (nh^p)^{1/2} (\hat{F}_n(t|x) - F(t|x)), \quad t \in [\tau_0, \tau_1] \right\}, \quad (2.1)$$

converges weakly in $\ell^\infty[\tau_0, \tau_1]$ to a continuous mean-zero Gaussian process $Z(\cdot|x)$ with covariance function

$$\hat{\Gamma}(t, s|x) = \frac{\|K\|_2^2}{f(x)} (1 - F(t|x))(1 - F(s|x)) \int_{-\infty}^{t \wedge s} \frac{\mathbb{E}[P^2|T = y, X = x]}{(1 - H(y|x))^2} dH(y|x).$$

3 Missing at random application

$$\hat{P}_n(t|x) = \sum_{i=1}^n \bar{W}_b((t - T_i, x - X_i)) \sigma_i$$

where for any $i = 1, \dots, n$, $\sigma_i = P_i \xi_i$ and

$$\bar{W}_b(x - X_i) = \frac{L_b(t - T_i, x - X_i)}{\sum_{j=1}^n L_b(t - T_j, x - X_j) \xi_j}$$

and $L_b(\cdot) = L(\cdot/b)/b^{p+1}$ with $L : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ a kernel function and $b = b_n$ a non-random positive sequence such that $b_n \rightarrow 0$ and $h = o(b)$ as $n \rightarrow \infty$.

Assumption (\mathcal{L}). L and satisfies Assumption (\mathcal{K}) in \mathbb{R}^{p+1} , there exists $\delta, m > 0$ such that $B_0(\delta) \subset S_L$ and $L(u) \geq m$ for all $u \in B_0(\delta)$, and L belongs to the linear span (the set of finite linear combinations) of functions $k \geq 0$ satisfying the following property: the subgraph of k , $\{(s, u) : k(s) \geq u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s, u) : q(s, u) \geq \varphi(u)\}$, where q is a polynomial on $\mathbb{R}^{p+1} \times \mathbb{R}$ and φ is an arbitrary real function.

Lemma 3.1 Assume that there exists $d > 0$ such that $f(x) \geq d, \forall x \in S_X \subset \mathbb{R}^p$, f is bounded, and (\mathcal{K}_2) and (\mathcal{F}) hold. Consider a sequence c tending to 0 as $n \rightarrow \infty$ such that for some $q > 1$

$$\frac{|\log c|^q}{nc^p} \longrightarrow 0.$$

Also assume that there exists an $\varepsilon > 0$ such that for n sufficiently large

$$\inf_{x \in S_X} \lambda(\{u \in B_0(1) : x - cu \in S_X\}) > \varepsilon, \quad (3.1)$$

where λ denotes the Lebesgue measure. Then, for any $0 < \eta < \min(\eta_{F_1}, \eta_{F_2})$, we have

$$\sup_{(y,x) \in \mathbb{R} \times S_X} |F_{n,j}(y|x) - F_j(y|x)| = o_{\mathbb{P}} \left(\max \left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta \right) \right), \text{ for } j = 1, 2.$$

4 Appendix

4.1 Proofs of Section 2

Proofs of Lemma 2.1. The proof here is similar to that of Lemma 3.6 in [?]. According to the mean value theorem, we have for any $t \in [\tau_0, \tau_1]$

$$\begin{aligned} \left| 1 - \hat{F}_n(t|x) - \exp(-\hat{\Lambda}_n(t|x)) \right| &\leq \left| r(t) \left[-\log(1 - \hat{F}_n(t|x)) - \hat{\Lambda}_n(t|x) \right] \right| \\ &\leq \left| -\log(1 - \hat{F}_n(t|x)) - \hat{\Lambda}_n(t|x) \right| \end{aligned} \quad (4.1)$$

where $r(t)$ lies between $1 - \hat{F}_n(t)$ and $\exp(-\hat{\Lambda}_n(t))$. For any $0 < y < 1$, we have the inequality $0 < -\log(1 - y) - y < \frac{y^2}{1 - y}$. This yields in (4.1) to

$$\begin{aligned} \left| 1 - \hat{F}_n(t|x) - \exp(-\hat{\Lambda}_n(t|x)) \right| &\leq \sum_{T_i \leq t} P_i \frac{W_h(x - X_i)^2}{(1 - H_n(T_i^-|x))(1 - H_n(T_i|x))} \\ &\leq \frac{\sum_{T_i \leq t} P_i W_h(x - X_i)^2}{(1 - H_n(\tau_1|x))^2} \\ &\leq \|K\|_\infty \frac{\sum_{T_i \leq t} P_i W_h(x - X_i)}{nh^p f_n(x)(1 - H_n(\tau_1|x))^2} \end{aligned}$$

and the lemma follows.

Proofs of the Theorem 2.1. The proofs follow the same arguments than to that of Theorem 3.1 in [?]. We thus refer to the former paper for the technical specifications and only provide here the main arguments for our estimator. According to the decomposition

$$\begin{aligned} \hat{F}_n(t|x) - F(t|x) &= \hat{F}_n(t|x) - 1 + \exp(-\hat{\Lambda}_n(t|x)) + \exp(-\Lambda_n(t|x)) - \exp(\hat{\Lambda}_n(t|x)) \\ &\quad 1 - F_n(t|x) - \exp(\Lambda_n(t|x)) + F_n(t|x) - F(t|x) \end{aligned}$$

we have combining Lemma 2.1 and Theorem 3.1 in [?] that

$$\begin{aligned} &\hat{F}_n(t|x) - F(t|x) \\ &= F_n(t|x) - F(t|x) + \exp(-\Lambda_n(t|x)) - \exp(-\hat{\Lambda}_n(t|x)) + \mathcal{O}_{\mathbb{P}}((nh^p)^{-1}) \\ &= \sum_{i=1}^n W_h(x - X_i) g(t, T_i, \delta_i|x) + \exp(-\Lambda_n(t|x)) - \exp(-\hat{\Lambda}_n(t|x)) + \mathcal{O}_{\mathbb{P}}((nh^p)^{-3/4} |\log h|^{3/4}). \end{aligned}$$

It thus only remain to derive the almost sure representation for $\Lambda_n - \hat{\Lambda}_n$. Indeed, by Taylor's expansion,

$$\begin{aligned} \exp(-\Lambda_n(t|x)) - \exp(-\hat{\Lambda}_n(t|x)) &= (\hat{\Lambda}_n(t|x) - \Lambda_n(t|x)) \exp(-\Lambda_n(t|x)) \\ &\quad + \frac{1}{2} (\Lambda_n(t|x) - \hat{\Lambda}_n(t|x))^2 \exp(-\Lambda_n(t|x)) (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

However, we have by definition that

$$\begin{aligned}
\hat{\Lambda}_n(t|x) - \Lambda_n(t|x) &= \int_{-\infty}^t \frac{d(\hat{H}_n^u - H_n^u)(s|x)}{1 - H_n(s^-|x)} \\
&= \sum_{T_i \leq t} \frac{(P_i - \delta_i) \mathbb{1}_{\{T_i \leq t\}}}{1 - H(T_i|x)} + \int_{-\infty}^t \frac{1}{1 - H(s|x)} - \frac{1}{1 - H_n(s|x)} d(\hat{H}_n^u - H_n^u)(s|x) \\
&\quad + \int_{-\infty}^t \frac{1}{1 - H_n(s^-|x)} - \frac{1}{1 - H_n(s|x)} d(\hat{H}_n^u - H_n^u)(s|x) \\
&=: \sum_{T_i \leq t} \frac{(P_i - \delta_i) \mathbb{1}_{\{T_i \leq t\}}}{1 - H(T_i|x)} + R_{n,1}(t) + R_{n,2}(t).
\end{aligned}$$

Clearly, for any $t \in \mathbb{R}$ the estimators $H_n(t)$ and $H_n(t^-)$ might only differs of one jump of size $(nh^p)^{-1}$, which gives us

$$\begin{aligned}
\left| \frac{1}{1 - H_n(s^-|x)} - \frac{1}{1 - H_n(s|x)} \right| &= \frac{|H_n(s|x) - H_n(s^-|x)|}{(1 - H_n(s^-|x))(1 - H_n(s|x))} \\
&\leq \frac{|H_n(s) - H_n(s^-)|}{(1 - H_n(\tau_1|x))^2} = \mathcal{O}_{\mathbb{P}}((nh^p)^{-1})
\end{aligned}$$

where the last term is uniform in $s \in \mathbb{R}$, yielding that $\sup_{t \in [\tau_0, \tau_1]} |R_{n,1}(t)| = \mathcal{O}_{\mathbb{P}}((nh^p)^{-1})$ for n large enough, since we have almost surely $H_n(\tau_1|x) \rightarrow H(\tau_1|x) < 1$ as $n \rightarrow +\infty$.

Next, we have

$$\begin{aligned}
R_{n,2}(t) &= \int_{-\infty}^t \frac{1}{1 - H(s|x)} - \frac{1}{1 - H_n(s|x)} d(\hat{H}_n^u - H^u)(s|x) \\
&\quad + \int_{-\infty}^t \frac{1}{1 - H(s|x)} - \frac{1}{1 - H_n(s|x)} d(H^u - H_n^u)(s|x) \\
&=: R_{n,3}(t) + R_{n,4}(t).
\end{aligned}$$

In order to uniformly bound the remaining terms $R_{n,3}(t)$ and $R_{n,4}(t)$, we again refer to the proof of Theorem 3.1 in [?]. Indeed H_n , H_n^u and \hat{H}_n^u share the same asymptotic behavior as stated in Lemma 3.1 from [?], that is

$$\begin{aligned}
\sup_{t \in \mathbb{R}} |H_n(t|x) - H(t|x)| &= \mathcal{O}_{\mathbb{P}}((nh^p)^{-1/2} |\log h|^{1/2}), \\
\sup_{t \in \mathbb{R}} |H_n^u(t|x) - H^u(t|x)| &= \mathcal{O}_{\mathbb{P}}((nh^p)^{-1/2} |\log h|^{1/2}), \\
\sup_{t \in \mathbb{R}} |\hat{H}_n^u(t|x) - H^u(t|x)| &= \mathcal{O}_{\mathbb{P}}((nh^p)^{-1/2} |\log h|^{1/2})
\end{aligned}$$

where the arguments for \hat{H}_n^u are the same as H_n and H_n^u . Since H and H^u are continuous, it yields from the proof of Lemma 2.1 in [?] that

$$\sup_{t \in [\tau_0, \tau_1]} |R_{n,3}(t)| = \mathcal{O}_{\mathbb{P}}((nh^p)^{-3/4} |\log h|^{3/4}) \quad \text{and} \quad \sup_{t \in [\tau_0, \tau_1]} |R_{n,4}(t)| = \mathcal{O}_{\mathbb{P}}((nh^p)^{-3/4} |\log h|^{3/4})$$

which show our result. \square

Proofs of the Theorem 2.2. Here the proof follows the same that of Theorem 3.2 in [?]. We thus mostly refer to this proof by adapting each milestones to our estimator. We also keep the same notations for convenience.

Overall, the approach relies on the Theorem 19.28 in [?]. Hence, we introduce some notations borrowed from the theory of weak convergence of empirical processes. Firstly, for any class \mathcal{F} of bounded and measurable functions over a metric space (\mathcal{T}, d) and any probability measure Q and $\epsilon > 0$, define the covering number $N(\mathcal{F}, L_2(Q), \epsilon)$ as the minimal number of $L_2(Q)$ -balls of radius ϵ needed to cover \mathcal{F} . We say that the class \mathcal{F} is VC if one can find $A > 0$ and $\nu > 0$ such that for any probability measure Q and $\epsilon > 0$,

$$N(\mathcal{F}, L_2(Q), \epsilon \|F\|_{Q,2}) \leq \left(\frac{A}{\epsilon}\right)^{\nu},$$

where $0 < \|F\|_{Q,2}^2 = \int F^2 dQ < \infty$ and F is an envelope function of the class \mathcal{F} . Additionally, we also define the uniform entropy integral as

$$J(\delta, \mathcal{F}, L_2) = \int_0^\delta \sqrt{\log \sup_{\mathcal{Q}} N(\mathcal{F}, L_2(Q), t \|F\|_{Q,2})} dt,$$

where \mathcal{Q} is the set of all probability measures Q . Next, let P denote the law of the vector (T, δ, X) and define the expectation under P as $Pf = \int f dP$ for any real-valued measurable function f . By integration by parts we can rewrite the function \hat{g} as

$$\begin{aligned} \hat{g}(t, T, \delta, P|x) &= (1 - F(t|x)) \left\{ \frac{\mathbb{1}_{\{\delta=1, T \leq t\}}}{1 - H(T|x)} - \int_{-\infty}^{T \wedge t} \frac{dH^u(s|x)}{(1 - H(s|x))^2} + \frac{(P - \delta)\mathbb{1}_{\{T \leq t\}}}{1 - H(T|x)} \right\} \\ &= g(t, T, \delta|x) + (1 - F(t|x)) \frac{(P - \delta)\mathbb{1}_{\{T \leq t\}}}{1 - H(T|x)} \end{aligned} \quad (4.2)$$

and define the sequence of classes $\hat{\mathcal{F}}_n$ with functions taking values in $E = \mathbb{R} \times \{0, 1\} \times [0, 1] \times \mathcal{S}_X$ as

$$\begin{aligned} \hat{\mathcal{F}}_n &= \left\{ (u, v, v', w) \rightarrow \hat{f}_{n,t}(u, v, v', w), t \in [\tau_0, \tau_1] \right\} \\ &= \left\{ (u, v, v', w) \rightarrow \sqrt{h^p} K_h(x - w) \hat{g}(t, u, v, v'|x), t \in [\tau_0, \tau_1] \right\}, \end{aligned}$$

embedded with the envelope function $E_n(u, v, v', w) = \sqrt{h^p} K_h(x - w)M$, $M > 0$ being an appropriate constant since $\hat{g}(\cdot|x)$ is uniformly bounded.

The weak convergence of the stochastic process (2.1) follows from the four conditions (6.6), (6.7), (6.8) and (6.9) as described in [?] page 26. However, (6.6) and (6.9) has already been proven and it only remains to show (6.7) and (6.8), that is

$$\sup_{|t-s| \leq \delta_n} P(\hat{f}_{n,t} - \hat{f}_{n,s})^2 \longrightarrow 0 \text{ for every } \delta_n \searrow 0, \quad (4.3)$$

$$J(\delta_n, \hat{\mathcal{F}}_n, L_2) \longrightarrow 0 \text{ for every } \delta_n \searrow 0. \quad (4.4)$$

In order to prove (4.3), without lost of generality, we have for any $s < t \in \mathbb{R}$

$$(\hat{f}_{n,t} - \hat{f}_{n,s})(u, v, v', w) = (f_{n,t} - f_{n,s})(u, v, w) + \mathbb{1}_{\{s < u \leq t\}} \frac{v' - v}{1 - H(u|x)}$$

that give us by convexity

$$((\hat{f}_{n,t} - \hat{f}_{n,s})(u, v, v', w))^2 \leq 2((f_{n,t} - f_{n,s})(u, v, w))^2 + 8 \frac{\mathbb{1}_{\{s < u \leq t\}}}{(1 - H(\tau_1|x))^2}.$$

Hence, we are able to obtain the following inequality

$$P(\hat{f}_{n,t} - \hat{f}_{n,s})^2 \leq 2P(f_{n,t} - f_{n,s})^2 + 8 \frac{\mathbb{P}(s < T \leq t)}{(1 - H(\tau_1|x))^2} = 2P(f_{n,t} - f_{n,s})^2 + 8 \frac{H(t) - H(s)}{(1 - H(\tau_1|x))^2}.$$

Since we already have that $\sup_{|t-s| \leq \delta_n} P(f_{n,t} - f_{n,s})^2 = o(1)$ in [?], (4.3) follows by the uniform continuity of the distribution function H . To prove (4.4), we use the results for the class of function \mathcal{F}_n introduced in [?] and apply them to $\hat{\mathcal{F}}_n$. Between $\hat{\mathcal{F}}_n$ and \mathcal{F}_n , we have the relation

$$\begin{aligned} \hat{\mathcal{F}}_n &= \left\{ (u, v, v', w) \rightarrow \sqrt{h^p} K_h(x - w) \left[g(t, u, v|x) + (1 - F(t|x)) \frac{(v - v') \mathbb{1}_{\{u \leq t\}}}{1 - H(u|x)} \right], t \in [\tau_0, \tau_1] \right\} \\ &\subset \mathcal{F}_n + \left\{ (u, v, v', w) \rightarrow \sqrt{h^p} K_h(x - w) (1 - F(t|x)) \frac{(v - v') \mathbb{1}_{\{u \leq t\}}}{1 - H(u|x)}, t \in [\tau_0, \tau_1] \right\} \\ &\subset \mathcal{F}_n + \left\{ (u, v, v', w) \rightarrow \sqrt{h^p} K_h(x - w) (1 - F(t|x)) \frac{v \mathbb{1}_{\{u \leq t\}}}{1 - H(u|x)}, t \in [\tau_0, \tau_1] \right\} \\ &\quad + \left\{ (u, v, v', w) \rightarrow -\sqrt{h^p} K_h(x - w) (1 - F(t|x)) \frac{v' \mathbb{1}_{\{u \leq t\}}}{1 - H(u|x)}, t \in [\tau_0, \tau_1] \right\} \\ &=: \mathcal{F}_n + \mathcal{G}_{n,1} + \mathcal{G}_{n,2}. \end{aligned}$$

According to Lemma 2.6.18 (i), (vi) and (viii) in [?], we obtain that $\mathcal{G}_{n,1}$ and $\mathcal{G}_{n,2}$ are VC with the envelope function $E'_n(u, v, v', w) = \sqrt{h^p} K_h(x - w) (1 - H(\tau_1|x))^{-1}$. Finally,

since the covering number for \mathcal{F}_n is already provided in [?] and that $\hat{\mathcal{F}}_n$ is included in the class of functions $\mathcal{F}_n + \mathcal{G}_{n,1} + \mathcal{G}_{n,2}$ with envelope function $E_n + 2E'_n$, we have using Lemma 16 in [?] for any $t > 0$

$$\begin{aligned} \sup_Q N(\hat{\mathcal{F}}_n, L_2(Q), t\|E_n + 2E'_n\|_{Q,2}) &\leq \sup_Q N(\mathcal{F}_n + \mathcal{G}_{n,1} + \mathcal{G}_{n,2}, L_2(Q), t\|E_n + 2E'_n\|_{Q,2}) \\ &\leq L \left(\frac{1}{t}\right)^V \end{aligned}$$

for some L and V . Thus, (4.4) is established since for any sequence $\delta_n \searrow 0$ and n large enough since we have

$$J(\delta_n, \hat{\mathcal{F}}_n, L_2) \leq \int_0^{\delta_n} \sqrt{\log(2^V L) - V \log(t)} dt = o(1).$$

and the weak convergence for our process is established. In order to derive the covariance structure of the limiting process, it is sufficient to see that in (4.2), \hat{g} might be written as

$$\hat{g}(t, T, \delta, P|x) = (1 - F(t|x)) \left\{ \frac{P \mathbb{1}_{\{T \leq t\}}}{1 - H(T|x)} - \int_{-\infty}^{T \wedge t} \frac{dH^u(y|x)}{(1 - H(y|x))^2} \right\}.$$

with the following equalities

$$\begin{aligned} \mathbb{E} \left[\frac{P \mathbb{1}_{\{T \leq t\}}}{1 - H(T|x)} \int_{-\infty}^{T \wedge s} \frac{dH^u(y|x)}{(1 - H(y|x))^2} \middle| X = x \right] &= \mathbb{E} \left[\frac{\mathbb{1}_{\{\delta=1, T \leq t\}}}{1 - H(T|x)} \int_{-\infty}^{T \wedge s} \frac{dH^u(y|x)}{(1 - H(y|x))^2} \middle| X = x \right] \\ &= \int_{-\infty}^t \frac{1}{1 - H(z|x)} \int_{-\infty}^{z \wedge s} \frac{dH^u(y|x)}{(1 - H(y|x))^2} dH^u(z|x), \end{aligned}$$

$$\mathbb{E} \left[\frac{P^2 \mathbb{1}_{\{T \leq t, T \leq s\}}}{(1 - H(T|x))^2} \middle| X = x \right] = \int_{-\infty}^{t \wedge s} \frac{\mathbb{E}[P^2 | T = y, X = x]}{(1 - H(y|x))^2} dH(y|x)$$

and

$$\begin{aligned} &\mathbb{E} \left[\int_{-\infty}^{T \wedge t} \frac{dH^u(y|x)}{(1 - H(y|x))^2} \int_{-\infty}^{T \wedge s} \frac{dH^u(y|x)}{(1 - H(y|x))^2} \middle| X = x \right] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z \wedge t} \frac{dH^u(y|x)}{(1 - H(y|x))^2} \int_{-\infty}^{z \wedge s} \frac{dH^u(y|x)}{(1 - H(y|x))^2} dH(z|x) \\ &= \int_{-\infty}^t \frac{1}{1 - H(z|x)} \int_{-\infty}^{z \wedge s} \frac{dH^u(y|x)}{(1 - H(y|x))^2} dH^u(z|x) + \int_{-\infty}^s \frac{1}{1 - H(z|x)} \int_{-\infty}^{z \wedge t} \frac{dH^u(y|x)}{(1 - H(y|x))^2} dH^u(z|x). \end{aligned}$$

Finally, we establish the continuity of the process thanks to a sufficient condition due to [?]. Indeed, let $(s, t) \in \mathbb{R}^2$ and denote

$$\bar{F}(\cdot|x) = 1 - F(\cdot|x) \quad \text{and} \quad m(y|x) = \frac{\mathbb{E}[P^2|T = y, X = x]}{(1 - H(y|x))^2}.$$

Then,

$$\begin{aligned} & \mathbb{E}[(Z(s|x) - Z(t|x))^2] \frac{f(x)}{\|K\|_2^2} \\ &= \bar{F}(s|x)^2 \int_{-\infty}^s m(y|x) dH(y|x) - 2\bar{F}(s|x)\bar{F}(t|x) \int_{-\infty}^{s \wedge t} m(y|x) dH(y|x) + \bar{F}(t|x)^2 \int_{-\infty}^t m(y|x) dH(y|x) \\ &= \bar{F}(s|x) \left[\bar{F}(s|x) \int_{s \wedge t}^s m(y|x) dH(y|x) + (F(t|x) - F(s|x)) \int_{-\infty}^{s \wedge t} m(y|x) dH(y|x) \right] \\ &\quad + \bar{F}(t|x) \left[\bar{F}(t|x) \int_{s \wedge t}^t m(y|x) dH(y|x) + (F(s|x) - F(t|x)) \int_{-\infty}^{s \wedge t} m(y|x) dH(y|x) \right] \\ &= \bar{F}(s \vee t|x)^2 \int_{s \wedge t}^{s \vee t} m(y|x) dH(y|x) + \int_{-\infty}^{s \wedge t} m(y|x) dH(y|x) (F(s|x) - F(t|x))^2 \\ &\leq \frac{c|s - t|^{\eta'}}{(1 - H(\tau_1|x))^2}. \end{aligned}$$

This yields that $\sqrt{\mathbb{E}[(Z(s|x) - Z(t|x))^2]} \leq \xi(t - s)$, with

$$\xi(t - s) = \sqrt{\frac{c}{f(x)} \frac{\|K\|_2}{1 - H(\tau_1|x)} |s - t|^{\eta'/2}}$$

being monotone and

$$\int_0^1 \frac{\xi(u)}{u|\log(u)|^{1/2}} du < +\infty. \quad \square$$

4.2 Proofs of Section 3