Kaplan-Meier V- and U-statistics

Tamara Fernández and Nicolás Rivera

Gatsby Computational Neuroscience Unit
University College London
Sainsbury Wellcome Centre
25 Howland Street
London W1T 4JG
United Kingdom
e-mail: t.a.fernandez@ucl.ac.uk

Computer Laboratory
University of Cambridge
William Gates Building
15 JJ Thomson Ave
Cambridge CB3 0FD
United Kingdom
e-mail: nicolas.rivera@cl.cam.ac.uk

Abstract: In this paper, we study Kaplan-Meier V- and U-statistics respectively defined as $\theta(\widehat{F}_n) = \sum_{i,j} K(X_{[i:n]}, X_{[j:n]})W_i W_j$ and $\theta_U(\widehat{F}_n) = \sum_{i \neq j} K(X_{[i:n]}, X_{[j:n]})W_i W_j / \sum_{i \neq j} W_i W_j$, where $\widehat{F}_n$ is the Kaplan-Meier estimator, $\{W_1, \ldots, W_n\}$ are the Kaplan-Meier weights and $K : (0, \infty)^2 \to \mathbb{R}$ is a symmetric kernel. As in the canonical setting of uncensored data, we differentiate between two asymptotic behaviours for $\theta(\widehat{F}_n)$ and $\theta_U(\widehat{F}_n)$. Additionally, we derive an asymptotic canonical V-statistic representation of the Kaplan-Meier V- and U-statistics. By using this representation we study properties of the asymptotic distribution. Applications to hypothesis testing are given.

MSC 2010 subject classifications: Primary 62N02, 62G20; secondary 62G30.
Keywords and phrases: Kaplan-Meier estimator, right-censoring, V-statistics.

Received December 2018.

1. Introduction

Let F be a distribution on $(0, \infty)$ of interest. In this paper we study the estimation of parameters of the form $\theta(F) = \int_0^\infty \int_0^\infty K(x, y)dF(x)dF(y)$, where $K : (0, \infty)^2 \to \mathbb{R}$ is a measurable and symmetric function, commonly known as kernel function. If we consider i.i.d. samples $X_1, \ldots, X_n$ from F, the standard estimators for $\theta(F)$ are the canonical V- and U-statistics,

$$\theta(\tilde{F}_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(X_i, X_j), \quad \text{and} \quad \theta_U(\tilde{F}_n) = \frac{1}{n(n-1)} \sum_{i \neq j} K(X_i, X_j),$$

respectively, where $\tilde{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}}$ denotes the empirical distribution.

Under some regularity conditions $\theta(\tilde{F}_n) \to \theta(F)$ and $\theta_U(\tilde{F}_n) \to \theta(F)$ almost surely as n approaches infinity. It is of interest to study the limit distribution of the differences $\theta(\tilde{F}_n) - \theta(F)$ and $\theta_U(\tilde{F}_n) - \theta(F)$. The standard theory of V- and U-statistics distinguishes between two asymptotic behaviours for the distribution of the errors $\theta_U(\tilde{F}_n) - \theta(F)$ and $\theta(\tilde{F}_n) - \theta(F)$ when the number of samples tends to infinity. These two asymptotic regimes, widely known as *degenerate* and *non-degenerate*, are characterised by the behaviour of the variance of the projection $\phi : \mathbb{R}_+ \to \mathbb{R}$, defined as

$$\phi(x) = \mathbb{E}(K(x, X_2)) = \int_0^\infty K(x, y) dF(y). \tag{1.1}$$

Assume that $\mathbb{E}(K(X_1, X_2)^2) < \infty$. On one hand, we say that we are in the non-degenerate regime if $\mathbb{V}ar(\phi(X_1)) > 0$, where we have that

$$\sqrt{n}(\theta_U(\tilde{F}_n) - \theta(F)) \xrightarrow{\mathcal{D}} N(0, 4\mathbb{V}ar(\phi(X_1))).$$

On the other hand, we are in the degenerate regime if $\mathbb{V}ar(\phi(X_1)) = 0$, where it holds that

$$n(\theta_U(\tilde{F}) - \theta(F)) \xrightarrow{\mathcal{D}} \sum_{i=1}^\infty \lambda_i(\xi_i^2 - 1),$$

where $\lambda_1, \lambda_2, \ldots$ are constants in $\mathbb{R}$, and $\xi_1, \xi_2, \ldots$ are independent standard Gaussian random variables. Similar results hold for V-statistics under slightly stronger assumptions. Indeed, if the extra condition $\mathbb{E}(|K(X_1, X_1)|) < \infty$ holds, then the exact same limit is obtained in the non-degenerate regime, while for the degenerate regime, $n(\theta(\tilde{F}_n) - \theta(F)) \xrightarrow{\mathcal{D}} \mathbb{E}(K(X_1, X_1)) + \sum_{i=1}^\infty \lambda_i(\xi_i^2 - 1)$. We refer to the book of Koroljuk and Borovskich [19] for a comprehensive account of the theory of V- and U- statistics.

In this paper, we study the analogue of V- and U-statistics in the setting of right-censored data, that usually appears in Survival Analysis applications, in which we observe samples of the form $(X_i, \Delta_i)_{i=1}^n$ where $X_i \in (0, \infty)$ and $\Delta_i \in \{0, 1\}$. Here, $\Delta_i = 1$ indicates that X_i is an actual sample from F, while $\Delta_i = 0$ indicates that X_i corresponds to a right-censored observation. Similar to the uncensored setting, we are interested on estimating $\theta(F)$, however, as the data is right-censored, it is not possible to compute the canonical V- and U-statistics as the empirical distribution $\tilde{F}_n$ is not available in this setting. Instead, we propose to replace the empirical distribution by the Kaplan-Meier estimator $\widehat{F}_n$ which is the standard estimator for F in the setting of right-censored data.

The Kaplan-Meier V- and U-statistics are defined as

$$\theta(\widehat{F}_n) = \sum_{i=1}^n \sum_{j=1}^n W_i W_j K(X_{[i:n]}, X_{[j:n]})$$

and

$$\theta_U(\widehat{F}_n) = \frac{\sum_{i \neq j} W_i W_j K(X_{[i:n]}, X_{[j:n]})}{\sum_{i \neq j} W_i W_j},$$

respectively, where W_i, are the so-called Kaplan-Meier weights and $X_{[i:n]}$ denotes the i-th order statistic. In this paper we conveniently write $\theta(\widehat{F}_n)$ as

$$\theta(\widehat{F}_n) = \int_0^\infty \int_0^\infty K(x, y) d\widehat{F}_n(x) d\widehat{F}_n(y),$$

which is known in the literature as a Kaplan-Meier double integral.

Asymptotic properties of Kaplan-Meier integrals have been studied by several authors. For the simplest case of univariate functions, Central Limit Theorems for $\int_0^\infty K(x) d\widehat{F}_n(x)$ were obtained in full generality by Stute [28] and Akritas [2]. Stute [28] achieved the result by expressing the Kaplan-Meier estimator as the sum of i.i.d. random variables plus some asymptotically negligible terms. By using this representation, Stute was able to deal with more general functions K than preceding approaches (e.g. [16, 29]), however the terms in the i.i.d. representation are quite complicated, leading to strong assumptions. This problem is aggravated in the case of distributions with atoms. Akritas [2] improved the result of Stute [28], obtaining weaker conditions in a more general framework. This was accomplished by using the martingale arguments developed by Gill [15, 16], and the identities and inequalities developed by Ritov and Wellner [23] and Efron and Johnstone [11].

Multiple Kaplan-Meier integrals were studied by Gijbels and Veraverbeke [14] and by Bose and Sen [6]. Gijbels and Veraverbeke [14] studied a simplification of the problem which considers the class of truncated Kaplan-Meier integrals $\int_0^t \ldots \int_0^t K(x_1, \ldots, x_m) \prod_{j=1}^m \widehat{F}_n(x_j)$, where t is a fixed value, avoiding integration over the whole support of the observations. Then, by using an asymptotic i.i.d. sum representation of the Kaplan-Meier estimator $\widetilde{F}_n$ together with integration by parts, the authors derive an almost sure canonical V-statistic representation of $\theta(\widehat{F}_n) - \theta(F)$ up to an error of order $O(n^{-1}\log(n))$. While this result allows them to derive limiting distributions in the non-degenerate case (by scaling by $\sqrt{n}$), it is not possible to obtain results for the degenerate case since the error is too large to be scaled by n. Moreover, their representation is restricted to continuous distribution functions F. Bose and Sen [6] analysed Kaplan-Meier double integrals in a more general setting by using a generalisation of the i.i.d. representation of the Kaplan-Meier estimator derived by Stute [28] for uni-dimensional Kaplan-Meier integrals. By using this representation, Bose and Sen [6] were able to write the Kaplan-Meier double integral as a V-statistic plus some error terms. Nevertheless, similar to the univariate case, the error terms, that appear as consequence of using this approximation, are quite complicated and thus, dealing with them requires very strong and somewhat artificial conditions which are hard to verify in practice.

An alternative estimator of $\theta(F)$ in the right-censored setting is obtained by using the so-called *Inverse Probability of censoring weighted* (IPCW) estimator of F introduced by [24], which can be seen as a simplification of the Kaplan-Meier estimator. IPCW U-statistics coincide with Kaplan-Meier U-statistics when the survival times are continuous and the largest sample point is uncensored [24]. IPCW U-statistics were studied by Datta et al. [7], however, they only provide results for the non-degenerate regime.

There are several works that study limit distributions of V- and U-statistics in the setting of dependent data (see [25, 9, 30, 10, 8, 3, 4]). Nevertheless, most of these results are tailored for specific types of dependency and thus they are not suitable or it is not clear how to translate these results into our setting. The recent approach of Beutner and Zähle [4] provides a general framework which can be applied to right-censored data, however its application is limited to very well-behaved cases. This is mainly because such an approach is based on an integration-by-parts argument, requiring the function K to be locally of bounded variation, and thus denying the possibility of working with simple kernels like $K(x, y) = \mathbb{1}_{\{x+y>1\}}$. Also, it is required to establish convergence of $\sqrt{n}(\widehat{F}_n - F)$ to a limit process (under an appropriate metric), leading to stronger conditions than the ones considered in our approach. Moreover, such a general result is less informative about the limit distribution than ours.

In this paper, we obtain limit results for Kaplan-Meier V-statistics. Our proof is based on two steps. First, we find an asymptotic canonical V-statistic representation of the Kaplan-Meier V-statistic, and second, we use such a representation to obtain limit distributions under an appropriate normalisation. We also obtain similar results for Kaplan-Meier U-statistics. Our results not only provide convergence to the limit distribution, but we also find closed-form expressions for the asymptotic mean and variance.

Applications to goodness-of-fit are provided. In particular, we study a slight modification of the Cramer-von Mises statistic under right-censoring that can be represented as a Kaplan-Meier V-statistic. Under the null hypothesis, we find its asymptotic null distribution, and we obtain closed-form expressions for the asymptotic mean and variance under a specific censoring distribution. Our results agree with those obtained by Koziol and Green [20]. We also provide an application to hypothesis testing using the Maximum Mean Discrepancy (MMD), a popular distance between probability measures frequently used in the Machine Learning community. Under the null hypothesis and assuming tractable forms for F and G, we obtain the asymptotic limit distribution, as well as the asymptotic mean and variance of the test-statistic.

Our results hold under conditions that are quite reasonable, in the sense that they require integrability of terms that are very close to the variance of the limit distribution. Compared to the closest works to ours, the approach of Bose and Sen [6] and the IPCW approach [7], our conditions are much weaker and easy to verify. We explicitly compare such conditions in Section 2.3.

1.1. Notation

We establish some general notation that will be used throughout the paper. We denote $\mathbb{R}_+ = (0, \infty)$. Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be an arbitrary right-continuous function, we define $f(x-) = \lim_{h \to 0} f(x - |h|)$ and $(\Delta f)(x) = f(x) - f(x-)$. In this work we make use of standard asymptotic notation [17] (e.g. O_p, o_p, Θ_p, etc.) with respect to the number of sample points n. In order to avoid large parentheses, we write $X = O_p(1)Y$ instead of $X = O_p(Y)$, especially

if the expression for Y is very long. Given a sequence of stochastic processes $(W_n(x) : x \in \mathcal{X})$, depending on the number of observations n, and a function $f(x) > 0$, we say that $W_n(x) = O_p(f(x))$ uniformly for all $x \in A_n$, if and only if $\sup_{x \in A_n} \frac{|W_n(x)|}{f(x)} = O_p(1)$, where $A_n \subseteq \mathcal{X}$ is a set that may depend on n.

1.1.1. Right-censored data

Right-censored data consists of pairs $((X_i, \Delta_i))_{i=1}^n$, where $X_i = \min(T_i, C_i)$ denotes the minimum between survival times of interest $T_i \overset{i.i.d.}{\sim} F$ and censoring times $C_i \overset{i.i.d.}{\sim} G$, and $\Delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$ is an indicator of whether we actually observe the survival time T_i or not, that is, $\Delta_i = 1$ if $T_i \leq C_i$, and $\Delta_i = 0$ otherwise. We assume the survival times T_i's are independent of the censoring times C_i's which is known as *non-informative* censoring, and it is a standard assumption in applications.

We denote by $S(x) = 1 - F(x)$ and by $\Lambda(x) = \int_{(0,x]} S(t-)^{-1} dF(t)$, respectively, the survival and cumulative hazard functions associated with the survival times T_i. The common distribution function associated with the observed right-censored times $X_i = \min\{T_i, C_i\}$ is denoted by H. Note that $1 - H(x) = (1 - G(x))S(x)$ due to the non-informative censoring assumption. For simplicity, we assume that F and G are measures on $\mathbb{R}_+$, otherwise, we can apply an increasing transformation to the random variables, e.g. e^{X_i}. Notice that we do not impose any further restriction to the distribution functions, particularly, F and G are allowed to share discontinuity points.

1.1.2. Kaplan-Meier estimator

The Kaplan-Meier estimator [18] is the non-parametric maximum likelihood estimator of F in the setting of right-censored data. It is defined as $\widehat{F}_n(x) = \sum_{X_{[i:n]} \leq x} W_i$, where $W_i = \frac{\Delta_{[i:n]}}{n} \prod_{j=1}^{i-1} \left(1 + \frac{1 - \Delta_{[j:n]}}{n-j}\right)$ are the so-called Kaplan-Meier weights, $X_{[i:n]}$ is the i-th order statistic of the sample $X_1, \ldots, X_n$, and $\Delta_{[i:n]}$ is its corresponding censor indicator. To be very precise, ties within uncensored times or within censored times are ordered arbitrarily, while ties among uncensored and censored times are ordered such that censored times appear later. Observe that when all the observations are uncensored, that is, when $\Delta_i = 1$ for all i, each weight W_i becomes to $1/n$ and thus $\widehat{F}_n$ becomes the empirical distribution of $T_1, \ldots, T_n$. Finally, we denote by $\widehat{S}_n(x) = 1 - \widehat{F}_n(x)$ the corresponding estimator of $S(x)$.

1.1.3. Counting processes notation

In this work we use standard Survival Analysis/Counting Processes notation. For each $i \in \{1, \ldots, n\}$ we define the individual and pooled counting processes by $N_i(t) = \mathbb{1}_{\{X_i \leq t\}} \Delta_i$ and $N(t) = \sum_{i=1}^n N_i(t)$ respectively. Notice that the

previous processes are indexed in $t \geq 0$. Similarly, we define the individual and pooled risk functions by $Y_i(t) = \mathbb{1}_{\{X_i \geq t\}}$ and $Y(t) = \sum_{i=1}^{n} Y_i(t)$, respectively.

We assume that all our random variables are defined in a common filtrated probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, where $\mathcal{F}_t$ is generated by

$$\{\mathbb{1}_{\{X_i \leq s, \Delta=1\}}, \mathbb{1}_{\{X_i \leq s, \Delta=0\}} : 0 \leq s \leq t, i \in \{0, \ldots, n\}\},$$

and the $\mathbb{P}$-null sets of $\mathcal{F}$. It is well-known that $N_i(t)$, $N(t)$ and the Kaplan-Meier estimator $\widehat{F}_n(t)$ are adapted to $(\mathcal{F}_t)_{t \geq 0}$, and that $Y_i(t)$ and $Y(t)$ are predictable processes with respect to $(\mathcal{F}_t)_{t \geq 0}$. Yet another well-known fact is that $N_i(t)$ is increasing and its compensator is given by $\int_{(0,t]} Y_i(s) d\Lambda(s)$. We define the individual and pooled $(\mathcal{F}_t)$-martingales by $M_i(t) = N_i(t) - \int_{(0,t]} Y_i(s) d\Lambda(s)$ and $M(t) = \sum_{i=1}^{n} M_i(t)$, respectively.

For a martingale W, we denote by $\langle W \rangle$ its predictable variation process, and by $[W]$ its quadratic variation process.

Due to the simple nature of the processes that appear in this work, i.e. counting processes, checking integrability and/or square-integrability is very simple and thus we state these properties without giving an explicit proof. For more information about counting processes in survival analysis we refer to [13].

1.1.4. Interior and exterior regions

Let $I_H = \{t > 0 : H(t-) < 1\}$ be the interval in which $X_i = \min\{T_i, C_i\}$ takes values. Define $\tau = \tau_H = \sup\{t : 1 - H(t) > 0\}$ and notice that $\tau \in I_H$ if and only if H has a discontinuity at τ.

Define $\tau_n = \max\{X_1, \ldots, X_n\}$. We denote by $I = (0, \tau_n]$ the *interior* region in which we observe data points, and by $E = I_H \setminus I$ the *exterior* the region. Notice that both I and E depend on τ_n even if we do not explicitly write it.

In this work, the integral symbol $\int_a^b$ means integration over the interval $(a, b]$, unless we state otherwise. An exception to this rule is when we integrate over the interval I_H, in which case, instead of writing $\int_{I_H}$, we write $\int_0^\tau$ and define $\int_0^\tau = \int_{(0,\tau]}$ if $\tau \in I_H$ and $\int_0^\tau = \int_{(0,\tau)}$ if $\tau \notin I_H$.

Lastly, let $g : \mathbb{R}_+ \to \mathbb{R}$ be an arbitrary function, then $\int_{\mathbb{R}_+} g(x) d\widehat{F}_n(x) = \int_0^\tau g(x) d\widehat{F}_n(x) = \int_0^{\tau_n} g(x) d\widehat{F}_n(x)$ as $\widehat{F}_n(t) = \widehat{F}_n(\tau_n)$ for all $t > \tau_n$. The same holds when integrating with respect the martingale $M(t)$.

1.1.5. Efron and Johnstone's forward operator

We consider the forward operator $A : L^2(F) \to L^2(F)$, independently introduced by Ritov and Wellner [23] and Efron and Johnstone [11], defined by

$$(Ag)(x) = g(x) - \frac{1}{S(x)} \int_x^\infty g(s) dF(s).$$

For a bivariate function $g : \mathbb{R}_+^2 \to \mathbb{R}$ such that $g \in L^2(F \times F)$, we denote by A_i, the operator A applied only to the i-th coordinate of the function g, e.g. $(A_2 g)(x, y) = g(x, y) - \frac{1}{S(y)} \int_y^\tau g(x, s) dF(s)$. Note that A_1 and A_2 commute.

Similarly, for $g : \mathbb{R}_+ \to \mathbb{R}$ such that $g \in L^2(F)$, we define $\widehat{A} : L^2(F) \to L^2(F)$, as

$$(\widehat{A}g)(x) = \begin{cases} g(x) - \frac{1}{S(x)} \int_x^{\tau_n} g(s) dF(s), & 0 < x \leq \tau_n \\ 0, & x > \tau_n \end{cases}.$$

Observe that the difference between $\widehat{A}$ and the forward operator A is the upper limit of integration. For bivariate functions, we define $\widehat{A}_i$ as the operator $\widehat{A}$ applied only to the i-th coordinate.

Notice that for $x \leq \tau_n$,

$$((\widehat{A} - A)g)(x) = \frac{1}{S(x)} \left(\int_x^\infty g(s) dF(s) - \int_x^{\tau_n} g(s) dF(s) \right)$$

$$= \frac{1}{S(x)} \int_{\tau_n}^\infty g(s) dF(s). \tag{1.2}$$

Also, notice that if $g(x) = 1$ is a constant function, then $(Ag)(x) = 1 - \frac{1}{S(x)} \int_x^\infty 1 dF(s) = 0$, and $(\widehat{A}g)(x) = 1 - \frac{S(x) - S(\tau_n)}{S(x)} = \frac{S(\tau_n)}{S(x)}$. Finally, observe that the definitions of A and $\widehat{A}$ depend only on F and it does not consider the censoring distribution G.

2. Main results

The Kaplan-Meier V-statistic associated with $\theta(F)$ is defined by

$$\theta(\widehat{F}_n) = \int_0^{\tau_n} \int_0^{\tau_n} K(x, y) d\widehat{F}_n(x) d\widehat{F}_n(y) = \sum_{i=1}^n \sum_{j=1}^n W_i W_j K(X_{[i:n]}, X_{[j:n]}),$$

where the second equality follows from the definition of the Kaplan-Meier estimator $\widehat{F}_n$.

Bose and Sen [5] proved that

$$\theta(\widehat{F}_n) \overset{a.s.}{\to} \theta(F; \tau) = \int_0^\tau \int_0^\tau K(x, y) dF(x) dF(y),$$

as n approaches infinity. Notice that the limit $\theta(F; \tau)$ has a dependency on τ since the data is right-censored, and thus we do not observe any survival time beyond the time τ. Consider the difference $\theta(\widehat{F}_n) - \theta(F; \tau)$, and notice it can be decomposed into two error terms:

$$\theta(\widehat{F}_n) - \theta(F; \tau) = \int_0^\tau \int_0^\tau K(x, y)(d\widehat{F}_n(x) d\widehat{F}_n(y) - dF(x) dF(y))$$

$$= 2\alpha(F, \widehat{F}_n) + \beta(F, \widehat{F}_n), \tag{2.1}$$

where

$$\alpha(F, \widehat{F}_n) = \int_0^\tau \phi(y)d(\widehat{F}_n - F)(y),$$

where the projection ϕ is given by

$$\phi(y) = \int_0^\tau K(x,y)dF(x), \tag{2.2}$$

and

$$\beta(F, \widehat{F}_n) = \int_0^\tau \int_0^\tau K(x,y)d(\widehat{F}_n - F)(x)d(\widehat{F}_n - F)(y). \tag{2.3}$$

Note that in our setting, our definition of ϕ integrates up to τ, instead of the whole support of F as in Equation (1.1) used in the uncensored setting. To ease notation, we write α and β instead of $\alpha(F, \widehat{F}_n)$ and $\beta(F, \widehat{F}_n)$, respectively. Each error term α and β can be seen as a first and second order approximation of the difference $\theta(\widehat{F}_n) - \theta(F; \tau)$. That being said, we expect that the error term α is of a much larger order than β. Indeed, it holds that $\alpha = O_p(n^{-1/2})$ and $\beta = O_p(n^{-1})$. This suggests the use of two different scaling factors, splitting our main result into two cases: the *non-degenerate* and the *degenerate* case. In the first case, we will show that, under appropriate conditions, $\sqrt{n}(\theta(\widehat{F}_n) - \theta(F; \tau))$ converges in distribution to a zero-mean normal random variable when n approaches infinity. This result will follow from proving a normal limit distribution for the scaled error term $\sqrt{n}\alpha$ and an in-probability convergence to zero of the scaled error $\sqrt{n}\beta$. In the degenerate case, the error term α is trivially 0, and thus we only care about the term β. We will show that $n\beta$ converges in distribution to a linear combination of (potentially infinity) independent χ^2 random variables plus a constant. From these results, we will be able to derive analogue results to those in the canonical V-statistics setting.

To express our results and conditions, we define the kernel $K' : \mathbb{R}_+^2 \to \mathbb{R}$ as $K'(x,y) = (A_1 A_2 K)(x,y)$, which, by the definition of the operators A_1 and A_2, is equal to

$$K(x,y) - \int_x^\tau K(s,y)\frac{dF(s)}{S(x)} - \int_y^\tau K(x,t)\frac{dF(t)}{S(y)} + \int_x^\tau \int_y^\tau K(s,t)\frac{dF(s)}{S(x)}\frac{dF(t)}{S(y)}. \tag{2.4}$$

We introduce two sets of conditions, one for the non-degenerate case, and one for the degenerate case.

Condition 2.1 (non-degenerate case: scaling factor $\sqrt{n}$)**.** Assume the following conditions hold:

i) $\int_0^\tau \int_0^\tau \frac{K(x,y)^2}{(1-G(x-))}dF(y)dF(x) < \infty$,

ii) $\int_0^\tau \left(\frac{S(x-)}{1-G(x-)}\right)^{1/2} |K'(x,x)|dF(x) < \infty$.

Condition 2.2 (degenerate case: scaling factor n)**.** Assume the following conditions hold:

i) $\int_0^\tau \int_0^\tau \frac{K(x,y)^2}{(1-G(x-))(1-G(y-))} dF(x)dF(y) < \infty$,

ii) $\int_0^\tau \frac{|K'(x,x)|S(x)}{1-H(x-)} dF(x) < \infty$.

Notice that Condition 2.2 implies Condition 2.1.

2.1. Results for Kaplan-Meier V-statistics

i) The non-degenerate case: $\sqrt{n}$-scaling: Equation (2.1) states $\sqrt{n}(\theta(\widehat{F}_n) - \theta(F;\tau)) = 2\sqrt{n}\alpha + \sqrt{n}\beta$. Recall that α is defined in terms of the projection ϕ defined in Equation (2.2). Then, the main result follows under Condition 2.1 from a standard application of the Central Limit Theorem (CLT) derived by Akritas [2] for univariate Kaplan-Meier integrals and by proving $\sqrt{n}\beta \xrightarrow{\mathbb{P}} 0$. Akritas proved

$$\sqrt{n}\alpha \xrightarrow{\mathcal{D}} N(0,\sigma^2),$$

where

$$\sigma^2 = \int_0^\tau \frac{S(x)}{1-H(x-)}(A\phi)^2(x)dF(x). \qquad (2.5)$$

As noticed by Efron and Johnstone [11], σ^2 is finite if $\int_0^\tau \frac{\phi(x)^2}{1-G(x-)} dF(x) < \infty$, which is implied by Condition 2.1.

Theorem 2.3. *Under Condition 2.1, it holds*

$$\sqrt{n}\alpha \xrightarrow{\mathcal{D}} N(0,\sigma^2), \quad and \quad \sqrt{n}\beta \xrightarrow{\mathbb{P}} 0,$$

and thus

$$\sqrt{n}(\theta(\widehat{F}_n) - \theta(F)) \xrightarrow{\mathcal{D}} N(0,4\sigma^2),$$

where $\sigma < \infty$ is given in Equation (2.5).

ii) The degenerate case: n-scaling. The previous result considers that $\sigma^2 > 0$. Notice this is not satisfied if $\alpha = 0$ as it implies $\sigma^2 = 0$. In turn, $\alpha = 0$ can be deduced from either of the following conditions of the projection ϕ defined in Equation (2.2): i) $\phi(x) = 0$, F-a.s., or ii) $\phi(x) = c$, F-a.s., for some non-zero constant c and $\widehat{S}_n(\tau) = S(\tau)$ a.s. for all n large enough. In the theory of V- and U-statistics, these conditions are known as the *degeneracy properties*. In such a case, the $\sqrt{n}$-scaling does not capture the nature of the asymptotic distribution of $\theta(\widehat{F}_n) - \theta(F;\tau)$, suggesting that we need to consider a larger scaling factor.

Recall $\alpha = \int_0^\tau \phi(x)(d\widehat{F}_n - dF)(x)$. Then, it is straightforward to verify, that i) $\phi(x) = 0$, F-a.s., implies $\alpha = 0$, and that ii) if $\phi(x) = c$, F-a.s., with $c \neq 0$,

then $\alpha = c(\widehat{F}_n(\tau) - F(\tau)) = 0$ if and only if $\widehat{F}_n(\tau) = F(\tau)$ a.s. for all large n (notice this condition is trivially satisfied in the uncensored case). In those cases the information of the limit distribution is contained in the term β.

Define $J : (I_H \times \{0,1\})^2 \to \mathbb{R}$

$$J((x,r),(x',r')) = \int_0^\tau \int_0^\tau \frac{K'(s,t)}{(1-G(s-))(1-G(t-))} dm_{x,r}(s) dm_{x',r'}(t), \quad (2.6)$$

where $dm_{x,r}(s) = r\delta_x(s) - \mathbb{1}_{\{x \geq s\}} d\Lambda(s)$, and notice that $dm_{X_i, \Delta_i} = dM_i$, and recall that M_i is the i-th individual martingale defined in Section 1.1.3.

Theorem 2.4. *Under Condition 2.2, it holds that*

$$n\beta \xrightarrow{\mathcal{D}} \int_0^\tau \frac{S(x)}{1 - H(x-)} K'(x,x) dF(x) + \Psi,$$

where $\Psi = \sum_{i=1}^\infty \lambda_i(\xi_i^2 - 1)$, $\xi_i \overset{i.i.d}{\sim} N(0,1)$ *and the* λ_i*'s are the eigenvalues associated with the integral operator* $T^J : L^2(X, \Delta) \to L^2(X, \Delta)$, *defined as*

$$(T^J f)(X_1, \Delta_1) = \mathbb{E}\left(J((X_1, \Delta_1), (X_2, \Delta_2)) f(X_2, \Delta_2) | (X_1, \Delta_1)\right),$$

where $L^2(X, \Delta)$ *is the space of square-integrable functions with respect the measure induced by* (X_1, Δ_1).

Moreover, $\mathbb{E}(\Psi) = 0$ *and*

$$\mathbb{E}(\Psi^2) = 2 \int_0^\tau \int_0^\tau \frac{K'(x,y)^2 S(x) S(y)}{(1 - H(x-))(1 - H(y-))} dF(x) dF(y).$$

An immediate consequence of the previous Theorem is the asymptotic behaviour of the degenerate case for the Kaplan-Meier V-statistic.

Corollary 2.5. *Suppose one of the following degeneracy conditions hold:*

   i) $\phi(x) = 0$, *F-a.s.*,
   ii) $\phi(x) = c$, *F-a.s. for some non-zero constant* c, *and exists* $N > 0$ *such that* $\widehat{S}_n(\tau) = S(\tau)$ *a.s. for all* $n \geq N$.

Then, under Condition 2.2,

$$n(\theta(\widehat{F}_n) - \theta(F; \tau)) \xrightarrow{\mathcal{D}} \int_0^\tau \frac{S(x)}{1 - H(x-)} K'(x,x) dF(x) + \Psi,$$

where $\Psi = \sum_{i=1}^\infty \lambda_i(\xi_i^2 - 1)$ *is defined as in Theorem 2.4.*

As a part of the proof, we find an asymptotic representation of β as a canonical V- statistic, this representation is as following.

Theorem 2.6. *Under Condition 2.2 it holds that*

$$\beta = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n J((X_i, \Delta_i), (X_j, \Delta_j)) + o_p(n^{-1}).$$

All the results of this section are proved in Section 5 except for Theorem 2.6, which is proved in Section 9.

2.2. Kaplan-Meier U-statistics

The Kaplan-Meier U-statistic is defined by

$$\theta_U(\widehat{F}_n) = \frac{\sum_{i \neq j} W_i W_j K(X_{[i:n]}, X_{[j:n]})}{\sum_{i \neq j} W_i W_j} = \frac{\theta(\widehat{F}_n) - \sum_{i=1}^{n} K(X_i, X_i) W_i^2}{\sum_{i \neq j} W_i W_j},$$

where the second equality follows from adding and subtracting the diagonal term $(\sum_{i=1}^{n} K(X_i, X_i) W_i^2)/(\sum_{i \neq j} W_i W_j)$.

Without loss of generality, assume $\theta(F; \tau) = 0$. Then, the asymptotic distribution of $\theta_U(\widehat{F}_n)$ can be related to the one for $\theta(\widehat{F}_n)$ by analysing the asymptotic behaviour of $\sum_{i \neq j} W_i W_j$ and $\sum_{i=1}^{n} K(X_i, X_i) W_i^2$. For the first term, Bose and Sen [5] proved that $\sum_{i \neq j} W_i W_j \overset{a.s.}{\to} F(\tau)^2$. For the second term we enunciate the following result, which is proved in Appendix D.

Lemma 2.7. *If $\int_0^\tau |K(x,x)| \sqrt{\frac{S(x-)}{1-G(x-)}} dF(x) < \infty$, then*

$$\sqrt{n} \sum_{i=1}^{n} K(X_i, X_i) W_i^2 = o_p(1).$$

Additionally, if $\int_0^\tau \frac{|K(x,x)|}{1-G(x-)} dF(x) < \infty$, then

$$n \sum_{i=1}^{n} K(X_i, X_i) W_i^2 = \int_0^\tau \frac{K(x,x)}{1-G(x-)} dF(x) + o_p(1).$$

The previous Lemma combined with the results obtained in the previous section allow us to deduce the following results for Kaplan-Meier U-statistics.

Corollary 2.8. *Assume Condition 2.1, and additionally assume that we have $\int_0^\tau |K(x,x)| \sqrt{\frac{S(x-)}{1-G(x-)}} dF(x) < \infty$. Then, it holds that*

$$\sqrt{n} \theta_U(\widehat{F}_n) \overset{\mathcal{D}}{\to} N\left(0, 4\sigma^2/F(\tau)^4\right).$$

Corollary 2.9. *Suppose one of the following degeneracy conditions hold:*

i) $\phi(x) = 0$, F-a.s.,
ii) $\phi(x) = c$, F-a.s. for some non-zero constant c, and exists $N > 0$ such that $\widehat{S}_n(\tau) = S(\tau)$ a.s. for all $n \geq N$.

Assume Condition 2.2 and that $\int_0^\tau \frac{|K(x,x)|}{1-G(x-)} dF(x) < \infty$. Then, it holds that

$$n \theta_U(\widehat{F}_n) \overset{\mathcal{D}}{\to} \frac{1}{F(\tau)^2} \left(\int_0^\tau \left(\frac{S(x)K'(x,x)}{1-H(x-)} - \frac{K(x,x)}{1-G(x-)} \right) dF(x) + \Psi \right),$$

where Ψ is as in Theorem 2.4.

2.3. Analysis of Conditions 2.1 and 2.2, and comparison with related works

In this section we discuss our conditions, and we compare them with the work of Bose and Sen [6] and Datta et al. [7].

We begin by analysing Condition 2.1, used in the non-degenerate regime, which implies Theorem 2.3. Recall that Efron and Johnstone [11] showed that the variance of the limit distribution in Theorem 2.3 is finite if

$$\int_0^\tau \frac{(\phi(x))^2}{1 - G(x-)} dF(x) = \int_0^\tau \frac{1}{1 - G(x-)} \left(\int_0^\tau K(x,y) dF(y) \right)^2 dF(x) < \infty, \tag{2.7}$$

whereas our Condition 2.1.i requires

$$\int_0^\tau \int_0^\tau \frac{K(x,y)^2}{1 - G(x-)} dF(y) dF(x) < \infty,$$

which is very close to term in Equation (2.7). Indeed, there is just one *Cauchy-Schwarz* inequality gap from the condition of Efron and Johnstone [11], suggesting little room for improvement. On the other hand, Condition 2.1.ii is a standard condition to deal with the diagonal term that appears in the V-statistic representation. It is only used in Lemma 9.1 and it is usually much simpler to verify due to the multiplicative factor $\sqrt{S(x-)}$ that appears in the integral, which makes the tail much lighter.

We compare our Condition 2.1 with the conditions of Theorem 1 of Bose and Sen [6], which establishes the same limit result as our Theorem 2.3 under different conditions. Theorem 1 of [6] requires our Condition 2.2.i (which implies our Condition 2.1.i), together with three extra conditions involving the function

$$C(x) = \int_0^{x-} \frac{dG(s)}{S(s)((1 - G(s))^2)}. \tag{2.8}$$

For example, one of the extra conditions required is

$$\int_0^\tau \int_0^\tau \frac{|K(x,y)|C(x)C(y)}{(1 - G(x-))(1 - G(y-))} dF(x) dF(y) < \infty, \tag{2.9}$$

which, compared to our Condition 2.1.i, is much harder to satisfy as the function $C(x)$ grows much faster than $(1 - G(x-))^{-1}$ when x approaches infinity. Indeed, by assuming that G and F are continuous distributions, it is not hard to verify that $(\frac{\partial}{\partial x} C(x))/(\frac{\partial}{\partial x}(1 - G(x))^{-1}) = 1/S(x)$. Therefore, unless the kernel $K(x,y)$ decays very fast, Equation (2.9) is very hard to satisfy. In example 2.10 below, we show that $C(x)$ can grow exponentially faster than $(1 - G(x))^{-1}$.

We continue by comparing our Condition 2.1 with the ones of Theorem 1 of Datta et al. [7]. Theorem 1 of [7] requires $\int_0^\tau \frac{\phi(x)^2}{1-G(x-)} dF(x) < \infty$, which is the

condition of Efron and Johnstone [11] for finiteness of the variance. However, it also requires

$$\int_0^\tau \left(\frac{(\phi - A\phi)(x)}{1 - G(x-)} \right)^2 \frac{dG(x)}{1 - G(x-)} = \int_0^\tau \frac{S(x)}{1 - G(x)} (\phi - A\phi)(x)^2 dC(x) < \infty,$$

which is very hard to satisfy as it involves the function $C(x)$ defined in Equation (2.8).

Example 2.10. Let $K(x,y) = xy$, $S(x) = e^{-x}$ and $1 - G(x) = S(x)^a$ with $a > 0$. Note that $\phi(x) = x$, $(A\phi)(x) = -1$ and $K'(x,y) = (A_1 A_2 K)(x,y) = 1$. From Theorem 2.3, we have that $\sqrt{n}(\theta(\widehat{F}_n) - \theta(F)) \xrightarrow{\mathcal{D}} N(0, 4\sigma^2)$, where $\sigma = \int_0^\infty e^{-x+ax} dx$ (see Equation (2.5)) is finite if and only if $a < 1$.

In this setting, our Conditions 2.1.i and 2.1.ii are

$$\int_0^\infty \int_0^\infty x^2 y^2 e^{-x+ax} e^{-y} dx dy < \infty, \quad \text{and} \quad \int_0^\infty e^{-(3/2-a/2)x} < \infty,$$

respectively, which are satisfied for $a < 1$. Hence, our conditions are the best possible in this case, as the variance σ^2 of the limit distribution is finite if and only if $a < 1$.

Bose and Sen's approach [6] requires the finiteness of the expression in Equation (2.9). In this example, $C(x) = a \frac{e^{x+ax}-1}{1+a}$, then Equation (2.9) is equal to

$$\int_0^\tau \int_0^\tau \frac{|K(x,y)|C(x)C(y)}{(1 - G(x-))(1 - G(y-))} dF(x)dF(y) = \left(\int_0^\infty x \frac{a(e^{x+ax}-1)}{(1+a)e^{-ax}} e^{-x} dx \right)^2,$$

which is infinite for all $a > 0$. We deduce that Theorem 1 of [6] cannot be applied in this setting.

IPCW's approach [7] requires

$$\int_0^\infty x^2 e^{ax} e^{-x} dx < \infty, \quad \text{and} \quad \int_0^\infty a(x+1)^2 e^{2ax} dx < \infty.$$

While the first equation is satisfied for $a < 1$, the second equation cannot be satisfied for any $a > 0$. Hence, Theorem 1 of [7] does not hold in this setting.

In the previous example note that $C(x) = a \frac{e^{x+ax}-1}{1+a}$ grows exponentially fast with x. Therefore, to use the respective theorems of [6] and [7] we will need to use kernels that decay exponentially fast.

We continue by analysing Condition 2.2 which is used in the degenerate case. Observe that the integral of Condition 2.2.ii is equal to the first moment of the limit distribution of Theorem 2.4, thus this condition cannot be avoided. The variance of the limit distribution in Theorem 2.4, is given by

$$\int_0^\tau \frac{K'(x,y)^2 S(x)S(y)}{(1 - H(x-))(1 - H(y-))} dF(x)dF(y), \tag{2.10}$$

while our Condition 2.2.i requires

$$\int_0^\tau \frac{K(x,y)^2}{(1-G(x-))(1-G(y-))} dF(x)dF(y) < \infty,$$

which ensures the finiteness of (2.10). Recall that K', defined in Equation (2.4), is given by

$$K(x,y) - \int_x^\tau K(s,y)\frac{dF(s)}{S(x)} - \int_y^\tau K(x,t)\frac{dF(t)}{S(y)} + \int_x^\tau \int_y^\tau K(s,t)\frac{dF(s)}{S(x)}\frac{dF(t)}{S(y)}.$$

From here, if we consider continuous distributions, we observe that the expression in our condition is similar to the variance given in (2.10). If we consider an appropriate kernel K, it may happen that some terms in K' cancel each other, resulting in a kernel K' of much smaller order than K. An example of this is the kernel $K(x,y) = (x-c)(y-c)$, where $c = \int x dF(x)$ and $S(x) = e^{-x}$. Note this kernel is similar to the previous example, but we subtract c to make it degenerate. In this setting we have $K'(x,y) = 1$, hence it is easier to have finite variance than to satisfy our condition. However, in general cases we do not expect to have cancellation between the terms in K' and thus K and K' should be of similar order, making our Condition 2.2.i sufficient and necessary.

Up to the best of our knowledge, the work of Bose and Sen [6] is the only one that establishes results for the degenerate case in a general setting. Compared to their result, our conditions are better since their Theorem 2 has the same requirements as their Theorem 1, i.e. conditions involving the function $C(x)$, including Equation (2.9) which, as we saw in our previous example, is very hard to satisfy. Indeed, if we repeat Example 2.10 with the kernel $K(x,y) = (x-c)(y-c)$, the conditions of Theorem 2.4 are satisfied for $a < 1$ (in which case the asymptotic variance is well-defined), while the conditions of Bose and Sen are not satisfied.

3. Applications

We give two examples of applications that motivated us to study Kaplan-Meier V-statistics. First we analyse a slight variation of the Cramer-Von Mises statistic that allows us to treat it as a Kaplan-Meier V-statistic. In our second application, we measure goodness-of-fit via the Maximum Mean Discrepancy (MMD), a popular distance between probability measures frequently used in the Machine Learning community.

Example 3.1 (Cramér-von Mises test-statistic)**.** Consider the problem of testing the null hypothesis $H_0 : F = F_0$ against the general alternative $H_1 : F \neq F_0$. The Cramér-von Mises statistic measures the closeness between F and F_0 by computing

$$\int_{\mathbb{R}_+} (F(x) - F_0(x))^2 dF_0(x). \tag{3.1}$$

When F is a probability distribution function, it can be verified that Equation (3.1) equals to

$$\theta(F) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} K(x_1, x_2) dF(x_2) dF(x_1), \qquad (3.2)$$

where

$$K(x, y) = \int_{\mathbb{R}_+} (\mathbb{1}\{x \leq t\} - F_0(t))(\mathbb{1}\{y \leq t\} - F_0(t)) dF_0(t).$$

Under the null hypothesis $H_0 : F = F_0$, we estimate $\theta(F) = 0$ by using Equation (3.2), replacing F by the Kaplan-Meier estimator $\widehat{F}_n$. Then, our test-statistic is

$$\theta(\widehat{F}_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} W_i W_j K(X_{[i:n]}, X_{[j:n]}). \qquad (3.3)$$

Notice that the equality between Equations (3.1) and (3.2) is only valid when F is a probability distribution, unfortunately, the Kaplan-Meier estimator $\widehat{F}_n$ is not always a probability distribution, indeed, $\widehat{F}_n$ is a probability distribution if and only if the largest observation is uncensored, thus $\theta(\widehat{F}_n)$ is slightly different from the Cramér-von Mises test-statistic $\int_0^\infty (\widehat{F}_n(x) - F_0(x))^2 dF_0(x)$.

Under the null hypothesis, we observe two different asymptotic behaviours of our test-statistic $\theta(\widehat{F}_n)$, one for $F_0(\tau) < 1$ and the other for $F_0(\tau) = 1$. To see this, for $x \in I_H$, consider the projection ϕ defined in Equation (2.2), which in this case is given by

$$\begin{aligned}
\phi(x) &= \int_0^\tau K(x, y) dF_0(y) \\
&= \int_0^\infty (F_0(t \wedge \tau) - F_0(t) F_0(\tau)) \left(\mathbb{1}\{x \leq t\} - F_0(t)\right) dF_0(t),
\end{aligned}$$

and notice that if $F_0(\tau) < 1$, then $\phi(x)$ does not satisfy the degeneracy condition of Corollary 2.5. Thus, by Theorem 2.3, it holds that $\sqrt{n}(\theta(\widehat{F}_n) - \theta(F_0, \tau))$ is asymptotically normally distributed. On the other hand, if $F_0(\tau) = 1$, then $\phi(x)$ satisfies the degeneracy condition of Corollary 2.5, indeed, we have that $\phi(x) = 0$ for all $x \in I_H$. Hence, under Condition 2.2, Corollary 2.5 applies, concluding that $n\theta(\widehat{F}_n)$ is asymptotically distributed as the weighted sum of i.i.d. χ_1^2 random variables plus some constant term.

For comparison purposes, we consider the alternative formulation of the Cramer-von Mises statistic by Koziol and Green [20]. They consider the random integral $\Phi_n = \int_0^\infty (\bar{F}_n(t) - F_0(t))^2 dF_0(t)$, where $\bar{F}_n$ is exactly as the Kaplan-Meier estimator, but they force $\bar{F}_n(\tau_n) = 1$ even if the largest observation is censored. For simplicity of the analysis, Koziol and Green [20] assumed that the censoring distribution satisfies $1 - G(t) = S_0(t)^\gamma$ for $\gamma < 2$ and that F_0 is a continuous distribution. Then, based on Gaussian processes arguments, they proved
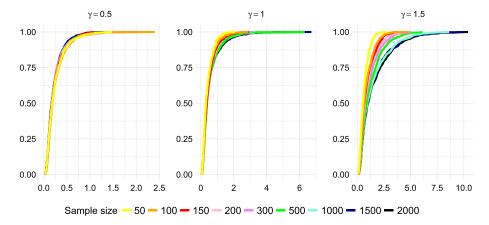
FIG 1. *Empirical distribution functions of $n\theta(\widehat{F}_n)$ for different sample sizes and values of γ. Each empirical distribution was computed using 1000 independent realisations.*

that $n\Phi_n \xrightarrow{\mathcal{D}} \Phi$ where Φ denotes (a potentially infinite) linear combination of $\chi_1^2 - 1$ independent random variables, and that

$$\mathbb{E}(\Phi) = \frac{1}{3(2-\gamma)}, \quad \text{and} \quad \mathbb{V}ar(\Phi) = \frac{2}{9(5-\gamma)(2-\gamma)}.$$

Using our techniques, we consider $\theta(\widehat{F}_n)$ as in Equation (3.3). In this case, we get $K(x,y) = S_0(\max\{x,y\}) + \frac{F_0(x)^2 + F_0(y)^2}{2} - \frac{2}{3}$ and $K'(x,y) = \frac{S_0(\max\{x,y\})^3}{3S_0(x)S_0(y)}$ (recall the definition of K' in Equation (2.4)). By choosing $1 - G(t) = S_0(t)^\gamma$, it holds $F_0(\tau) = 1$ which satisfies the degeneracy condition, as $\phi(x) = \int_0^\tau K(x,s)dF_0(s) = 0$. Then, if $\gamma < 1$, the conditions of Corollary 2.5 are satisfied and thus

$$n\theta(\widehat{F}_n) \xrightarrow{\mathcal{D}} \int_0^\infty \frac{S_0(x)}{3S_0(x)^\gamma}dF_0(x) + \Psi,$$

where Ψ is as in Theorem 2.4. Recall that $\mathbb{E}(\Psi) = 0$, then the asymptotic mean is given by $\frac{1}{3(2-\gamma)}$ and the asymptotic variance is given by

$$2\sum_{k=1}^\infty \lambda_k^2 = 2\int_0^\tau \int_0^\tau \frac{K'(x,y)^2}{S_0(x)^\gamma S_0(y)^\gamma}dF_0(y)dF_0(x) = \frac{2}{9(5-\gamma)(2-\gamma)}.$$

Our result suggests that our estimator and the one considered by Koziol and Green [20] have similar behaviours, even when rescaled by n. In Figure 1, we show simulations of the empirical distribution of $n\theta(\widehat{F}_n)$ for different sample sizes n, and $\gamma \in \{0.5, 1, 1.5\}$. For $\gamma = 0.5$ we can observe a clear convergence of the distribution functions as predicted by our results. The plot for $\gamma = 1.5$ shows a shift of the distribution functions as the sample size increases, suggesting divergence. The simulations for $\gamma = 1$ are, unfortunately, not very revealing.

Example 3.2 (Maximum mean discrepancy). Let $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ be a reproducing kernel Hilbert space of real-valued functions with reproducing kernel denoted by K. Denote by $\mathcal{P}$ the set of all probability distribution functions on $\mathbb{R}_+$, we define the map $\mu_. : \mathcal{P} \to \mathcal{H}$ by $\mu_F(\cdot) = \mathbb{E}_{X \sim F}(K(\cdot, X))$ for any distribution function $F \in \mathcal{P}$. A reproducing kernel is called characteristic if the map μ is injective [26]. It is worth mentioning that most of the standard positive-definite kernels (e.g. Gaussian and Ornstein-Uhlenbeck) are characteristic. In such a case, the map μ allows us to establish a proper distance between probability measures in terms of the norm of the space $\mathcal{H}$. That is, given two probability distributions F and F_0, we define their distance by

$$\|\mu_F - \mu_{F_0}\|_{\mathcal{H}} = \sqrt{\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} K(x,y)(dF(x) - dF_0(x))(dF(y) - dF_0(y))}. \quad (3.4)$$

Also, under the conditions stated above, such distance coincides with the Maximum mean discrepancy with respect to the unit ball of $\mathcal{H}$, which is defined as follows

$$MMD_{\mathcal{H}}(F, F_0) = \sup_{f \in \mathcal{H}, \|f\|_{\mathcal{H}} \leq 1} \mathbb{E}_F(f(X)) - \mathbb{E}_{F_0}(f(X)). \quad (3.5)$$

In the uncensored setting, the Maximum mean discrepancy has been used in a variety of testing problems. Indeed, in the simplest case, we can assess if our data points are generated from a distribution F_0 by comparing it with the empirical distribution $\tilde{F}_n$. By using the equivalency between Equation (3.4) and (3.5), we deduce that $MMD(\tilde{F}_n, F_0)^2$ is a V-statistic. This fact allows us to easily derive the relevant asymptotic results to construct a statistical test.

In the setting of right-censored data we study $MMD(\widehat{F}_n, F_0)^2$ using the Kaplan-Meier estimator $\widehat{F}_n$. By using Equations (3.4) and (3.5), our test-statistic can be written as

$$MMD(\widehat{F}_n, F_0)^2 = \int_0^\tau \int_0^\tau K(x,y)(d\widehat{F}_n(x) - dF_0(x))(d\widehat{F}_n(y) - dF_0(y)).$$

Notice that $MMD(\widehat{F}_n, F_0)^2$ coincides with β defined in Equation (2.3). Hence, under the null hypothesis $H_0 : F = F_0$, and Condition 2.2, Theorem 2.4 states

$$nMMD(\widehat{F}_n, F_0)^2 = n\beta(\widehat{F}_n, F_0) \xrightarrow{\mathcal{D}} \int_0^\tau \frac{K'(x,x)}{1 - G(x)} dF_0(x) + \Psi,$$

where Ψ is as in Theorem 2.4. Notice that Theorem 2.4 does not require the degeneracy condition of Corollary 2.5.

For the sake of simplicity, let us consider K as the Ornstein-Uhlenbeck kernel given by $K(x,y) = e^{-|x-y|}$, and let $S_0(x) = e^{-x}$ and $1 - G(x) = S_0(x)^\gamma$ (notice that for this choice of parameters $\tau = \infty$). A tedious computation shows that

$$K'(x,y) = \begin{cases} \frac{1}{2}(1 - (x-y))e^{-(x-y)} & x > y \\ \frac{1}{2}(1 - (y-x))e^{-(y-x)} & x \leq y \end{cases}.$$
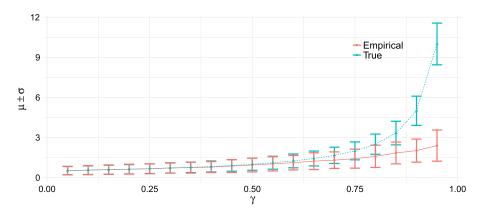
FIG 2. *Empirical vs asymptotic mean of $nMMD(\widehat{F}_n, F_0)$ for different values of $\gamma < 1$. For each value of the mean μ and $\widehat{\mu}$, we plot the intervals $[\mu - \sigma, \mu + \sigma]$ and $[\widehat{\mu} - \widehat{\sigma}, \widehat{\mu} + \widehat{\sigma}]$, where $\widehat{\sigma}^2$ and σ^2 denote the empirical and asymptotic variance respectively. We use a fixed sample size $n = 3000$.*

Then, under the null hypothesis and Condition 2.2, which is satisfied for $\gamma < 1$, it holds

$$nMMD(\widehat{F}_n, F_0)^2 \xrightarrow{\mathcal{D}} \frac{1}{2(1 - \gamma)} + \Psi.$$

Since $\mathbb{E}(\Psi) = 0$, the asymptotic mean is given by $\frac{1}{2(1-\gamma)}$ and the asymptotic variance corresponds to

$$2\sum_{k=1}^{\infty}\lambda_k^2 = 2\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} K'(x,y)^2 e^{-(1-\gamma)(x+y)}dxdy = \frac{5 - 4\gamma + \gamma^2}{2(\gamma - 3)^3(\gamma - 1)}.$$

In Figure 2, we compare the empirical mean and variance with the mean and variance of the limit distribution. We repeat this experiment 1000 times for different values of γ and a fixed sample size of 3000 data points. We observe that as γ approaches 1 the empirical estimation starts to get far away from the mean and variance predicted by our result, suggesting a slow convergence rate.

4. Conclusions and final remarks

In this work we studied the limit distribution of Kaplan-Meier V- and U-statistics under two different regimes: the degenerate and the non-degenerate. Our results hold under very simple conditions and, in practice, we just need to check the finiteness of two simple integrals. Compared to previous approaches our results are much simpler to state and the conditions required to apply them are much easy to satisfy and verify. Additionally, our result gives more information about the limit distribution, e.g. we give closed-form expressions for the

asymptotic mean and variance, as well as an asymptotic canonical V-statistic representation of the Kaplan-Meier V-statistic.

We give a few comments about our results. First, in the canonical case (uncensored data), U-statistics are preferred over V-statistics due to several reasons. Arguably, the most important reason is that U-statistics are unbiased while V-statistics are, in general, biased. The bias of V-statistics implies that limit theorems need to deal with the behaviour of the biased part of the estimator, resulting in stronger conditions in the statement of the results. In the right-censored setting, it does not seem to be a major difference between U- and V-statistics, and indeed, V-statistics are easier to work with as they can be represented by an integral with respect to the Kaplan-Meier estimator. Furthermore, due to the complex structure of the Kaplan-Meier weights, the Kaplan-Meier U-statistics are usually biased as opposed to its canonical counterpart, losing their main advantage over V-statistics.

Second, we think that our proof can be implemented in the settings of random kernels K_n that depend on the data points $(X_i, \Delta_i)_{i=1}^n$ as long certain regularity conditions hold, namely, i) K_n' is predictable in the sense of Definition 6.6, ii) it exists a deterministic kernel $\bar{K}$ such that $\sup_{x,y \leq \tau_n} |K_n(x,y)/\bar{K}(x,y)| = O_p(1)$, iii) K_n converges in probability to some deterministic kernel K, iv) $\bar{K}$ and K satisfy Conditions 2.1 or 2.2, depending on the case of interest.

Third, our analysis can be extended to kernels of dimension greater than two by using the same underlying ideas exposed in this work. Nevertheless, the statements and proofs of the results become much more complicated due to long computations that come from the fact that the core of our proof strategy relies on decomposing the integration region into Interior and Exterior regions, thus, as the number of integrals grows so do the possible combinations of these type of regions. We do not include these type of results as they do not add much value to the current work, especially because U- and V-statistics of dimension two are the most common in applications.

Finally, after the publication of the first preprint of this paper a few works have followed the path of using MMD distances to hypothesis testing in the setting of right-censored data. Particularly, Matabuena and Padilla [21] implemented our MMD example for the two-sample problem, and extended it to Energy Distances, which are a generalisation of the MMD. Their analysis is a direct application of Theorems 2.3 and 2.4, and Corollary 2.5. In a similar direction, in [12] the authors studied MMD distances in the context of hazard functions, obtaining as test-statistic a double integral with respect to the Nelson-Aalen estimator. Due to the relationship between the Nelson-Aalen estimator and the counting process martingale $M(t)$, the asymptotic analysis of their test-statistic is carried out by using the techniques of this paper. Somewhat related is the work of Rindt et al. [22], where an MMD independent test for right-censored data is presented. Their test-statistic is a Kaplan-Meier quadruple integral, however, under their null hypothesis, such a test-statistic becomes the product two Kaplan-Meier double integrals, thus its asymptotic analysis follows from our results.

5. Proofs I: Road map

In order to keep our proofs as tidy as possible and to emphasise the key steps without the distraction of messy computations, we give a list of intermediate steps that are needed to carry out the proof of our main results.

Recall that Equation (2.1) states $\theta(\widehat{F}_n) - \theta(F; \tau) = 2\alpha + \beta$, where $\alpha = \int_0^\tau \phi(x)d(\widehat{F}_n - F)(x)$ and $\phi(x) = \int_0^\tau K(x, y)dF(y)$. We analyse α and β individually.

5.1. Treatment of α

We distinguish between two cases, when $\phi(x)$, defines in Equation (2.2), satisfies the degeneracy condition stated in the Corollary 2.5 and when it does not. For the first case, observe that $\alpha = 0$ holds trivially. For the second case, an application of the Central Limit Theorem (CLT) of Akritas [2] gives us the asymptotic behaviour of α.

Theorem 5.1 (Akritas [2]). *Let $\phi : (0, \infty) \to \mathbb{R}$ be such that $\int_0^\tau \frac{\phi(x)^2}{1-G(x-)}dF(x) < \infty$, then*

$$\sqrt{n} \int_0^\tau \phi(x)d(\widehat{F}_n - F)(x) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\sigma^2 = \int_0^\tau \frac{S(s)}{1-H(s-)}(A\phi)^2(s)dF(s) < \infty$.

Then, by applying the previous CLT, we obtain the following Corollary.

Corollary 5.2. *Consider $\phi(x) = \int_0^\tau K(x, y)dF(y)$. Then, under Condition 2.1, $\int_0^\tau \frac{\phi(x)^2}{1-G(x-)}dF(x) < \infty$, and thus*

$$\sqrt{n}2\alpha \xrightarrow{\mathcal{D}} N(0, 4\sigma^2),$$

where $\sigma^2 = \int_0^\tau \frac{S(s)}{1-H(s-)}(A\phi)^2(s)dF(s) < \infty$.

5.2. Treatment of β

Let R_1, R_2 be two subsets of $\mathbb{R}_+$. Denote by $\beta_{R_1 \times R_2}$, the integral

$$\beta_{R_1 \times R_2} = \int_{R_1} \int_{R_2} K(x, y)d(\widehat{F}_n - F)(x)d(\widehat{F}_n - F)(y).$$

Observe that $\beta = \beta_{I_H \times I_H} = \beta_{I_H^2}$, and that β can be decomposed into $\beta = \beta_{I^2} + \beta_{(I^2)^c}$, where $(I^2)^c = (E \times E) \cup (I \times E) \cup (E \times I)$. We recall that I, E and I_H are defined in Section 1.1.4. To avoid extra parentheses, we write I^{2c} instead of $(I^2)^c$.

In Section 7, we study the asymptotic properties of $\beta_{I^{2c}}$, obtaining as main result the following Lemma.

Lemma 5.3. *Under Condition 2.1, it holds that $\sqrt{n}\beta_{I^{2c}} = o_p(1)$, and under Condition 2.2, it holds that $n\beta_{I^{2c}} = o_p(1)$.*

The handling of the term β_{I^2} is far more complicated since it contains all the important information about the limit distribution. In Section 8, we transform β_{I^2} into a more tractable object by performing a *change of measure*, where instead of integrating with respect to $d(\widehat{F}_n - F)$, we integrate with respect to the measure $dM = dN - Yd\Lambda$. This is done by using Duhamel's equations (Proposition 6.3). The main result of Section 8 is the following.

Lemma 5.4. *Under Condition 2.1, it holds that*

$$\sqrt{n}\beta_{I^2} = \sqrt{n}\int_{I^2}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}K'(x,y)dM(x)dM(y) + o_p(1), \qquad (5.1)$$

and under Condition 2.2, it holds that

$$n\beta_{I^2} = n\int_{I^2}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}K'(x,y)dM(x)dM(y) + o_p(1), \qquad (5.2)$$

where the kernel K' is defined in Equation (2.4).

5.3. Proof of Theorem 2.3

In order to prove Theorem 2.3, we require the following intermediate result, which is formally proved in Section 9.

Lemma 5.5. *Under Condition 2.1 it holds that*

$$\sqrt{n}\int_{I^2}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}K'(x,y)dM(x)dM(y) = o_p(1).$$

Proof of Theorem 2.3. Under Condition 2.1, Lemmas 5.3, 5.4 and 5.5 prove $\sqrt{n}\beta \xrightarrow{\mathbb{P}} 0$, which together with Corollary 5.2 concludes the result. $\square$

5.4. Proof of Theorem 2.4

We proceed to give proof to Theorem 2.4 and Corollary 2.5. Observe that under Condition 2.2, Lemma 5.3, and Equation (5.2) of Lemma 5.4 yield

$$n\beta = \frac{1}{n}\int_{I^2}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{(Y(x)/n)(Y(y)/n)}K'(x,y)dM(x)dM(y) + o_p(1). \qquad (5.3)$$

Theorem 2.6, states that $\widehat{S}_n(x-)$ and $Y(x)/n$ in Equation (5.3) can be substituted by their respective limits, $S(x-)$ and $1 - H(x-)$, obtaining that

$$n\beta = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{I_H^2}\frac{K'(x,y)}{(1-G(x-))(1-G(y-))}dM_i(x)dM_j(y) + o_p(1). \qquad (5.4)$$

The proof of Theorem 2.4 follows by noticing that the leading term in Equation (5.4) is a degenerate V-statistic.

Proof of Theorem 2.4. Equation (5.4) states

$$\beta = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} J((X_i, \Delta_i), (X_j, \Delta_j)) + o_p(n^{-1}),$$

where J is the kernel defined in Equation (2.6), thus we deduce that β is a canonical V-statistic up an error of order $o_p(n^{-1})$. From Condition 2.2, we can deduce that $\mathbb{E}(J((X_1, \Delta_1), (X_2, \Delta_2))^2) < \infty$, and that J satisfies the following degeneracy condition

$$\mathbb{E}(J((X_i, \Delta_i), (x, r))) = 0,$$

for all $(x, r) \in I_H \times \{0, 1\}$ (see Appendix C). Therefore, by applying standard results for degenerate V-statistics, e.g. [19, Theorem 4.3.2.], it holds

$$n\beta \xrightarrow{\mathcal{D}} \mathbb{E}(J((X_1, \Delta_1), (X_1, \Delta_1))) + \Psi,$$

where $\Psi = \sum_{i=1}^{\infty} \lambda_i(\xi_i^2 - 1)$, $\xi_1, \xi_2, \dots$ are i.i.d. standard normal random variables, and the λ_i's are the eigenvalues of the integral operator $T_J : L^2(X, \Delta) \to L^2(X, \Delta)$ associated with J.

Finally, note that $\mathbb{E}(\Psi) = 0$, and that

$$\mathbb{E}(\Psi^2) = 2 \sum_{i=1}^{\infty} \lambda_i^2 = 2\mathbb{E}(J((X_1, \Delta_1), (X_2, \Delta_2))^2)$$

$$= 2 \int_0^\tau \int_0^\tau \frac{S(x)S(y)K'(x, y)^2}{(1 - H(x-))(1 - H(y-))} dF(x) dF(y),$$

where the last equality is formally verified in Appendix C. $\qquad\square$

The rest of the paper is devoted to prove Lemmas 5.3, 5.4 and 5.5, and Theorem 2.6.

6. Proofs II: Preliminary results

The following results are going to be used several times in this paper.

6.1. Some results for counting processes

Proposition 6.1. *The following results holds a.s.*

i) $\lim_{n \to \infty} \sup_{t \le \tau} |\widehat{S}_n(t) - S(t)| = 0$,
ii) $\lim_{n \to \infty} \sup_{t \le \tau} |Y(t)/n - H(t-)| = 0$, *and*
iii) *for every fixed $t^\star > 0$ such that $1 - H(t^\star-) > 0$,*

$$\sup_{t \le t^\star} \left| n\widehat{S}_n(t-)/Y(t) - 1/(1 - G(t-)) \right| \to 0.$$

Item i. is due to Stute and Wang [27], item ii. is the Glivenko-Cantelli theorem, and item iii. follows from the two previous items.

Proposition 6.2. *Let $\beta \in (0,1)$, then the following results hold true:*

i) $\mathbb{P}(\widehat{S}_n(t) \leq \beta^{-1} S(t), \quad \forall t \leq \tau_n) \geq 1 - \beta,$
ii) $\mathbb{P}\left(Y(t)/n \geq \beta\{1 - H(t-)\}, \quad \forall t \leq \tau_n\right) \geq 1 - e(1/\beta)e^{-1/\beta},$
iii) $\mathbb{P}\left(Y(t)/n \leq \beta^{-1}\{1 - H(t-)\}, \quad \forall t \leq \tau_n\right) \geq 1 - \beta,$ *and*
iv) $\mathbb{P}\left(n(1 - H(\tau_n)) \leq \beta^{-1}\right) \geq 1 - e^{-\beta^{-1}}.$

i.e. i) $\sup_{t \leq \tau_n} \frac{\widehat{S}_n(t)}{S(t)} = O_p(1)$, *ii) and iii)* $\sup_{t \leq \tau_n} Y(t)/(n(1 - H(t-)) = \Theta_p(1)$, *and iv)* $n(1 - H(\tau_n)) = O_p(1)$.

Items i. and ii. are due to Gill [16]. Item iii. is from [15, Theorem 3.2.1], and Item iv. is due to Yang [29].

Yet another useful result is the so-called Duhamel's Equation.

Proposition 6.3 (Prop. 3.2.1 of [15])**.** *For all $x > 0$ such that $S(x) > 0$,*

$$d\widehat{F}_n(x) = dF(x) - \left(\int_{(0,x)} \frac{\widehat{S}_n(s-)}{S(s)Y(s)} dM(s)\right) dF(x) + \frac{\widehat{S}_n(x-)}{Y(x)} dM(x). \quad (6.1)$$

Equation (6.1) allows us to deduce two important results. Firstly,

$$\int_0^{\tau_n} \phi(x)(d\widehat{F}_n(x) - dF(x))$$
$$= \int_0^{\tau_n} \phi(x)\frac{\widehat{S}_n(x-)}{Y(x)} dM(x) - \int_0^{\tau_n} \int_{(0,x)} \phi(x)\frac{\widehat{S}_n(s-)}{S(s)Y(s)} dM(s)dF(x)$$
$$= \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{Y(x)} \left(\phi(x) - \int_x^{\tau_n} \phi(s)\frac{dF(s)}{S(x)}\right) dM(x)$$
$$= \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{Y(x)}(\widehat{A}\phi)(x)dM(x), \quad (6.2)$$

where $\widehat{A}$ is the operator defined in Section 1.1. Secondly, for a 2-dimensional kernel K, we obtain

$$\int_0^{\tau_n} \int_0^{\tau_n} K(x,y)(d\widehat{F}_n(x) - dF(x))(d\widehat{F}_n(y) - dF(y))$$
$$= \int_0^{\tau_n} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}(\widehat{A}_2\widehat{A}_1K)(x,y)dM(x)dM(y). \quad (6.3)$$

6.2. Some convergence theorems

We state, without proof, the following elementary result that is useful to prove that a sequence of (random) integrals converge to zero in probability.

Lemma 6.4. *Let $(\mathcal{X}, \mathcal{B}, \mu)$ be a σ-finite measure space. Let $(R_n(x) : x \in \mathcal{X})$ be a sequence of stochastic process indexed on $\mathcal{X}$. We assume that $R_n(\cdot)$ is measurable with respect to $\mathcal{B}$ (for any fixed realisation of R_n). Suppose that*

i) For each $x \in \mathcal{X}$, $R_n(x) \to 0$ almost surely as n tends to infinity, and
ii) it exists a deterministic non-negative function $R : \mathcal{X} \to [0, \infty)$ such that

$$\sup_{x \in \mathcal{X}} \frac{|R_n(x)|}{R(x)} = O_p(1),$$

and that $\int_{x \in \mathcal{X}} R(x)\mu(dx) < \infty$.

Define the sequence of random integrals $I_n = \int_{\mathcal{X}} R_n(x)\mu(dx)$, then $I_n = o_p(1)$.

6.3. Some martingale results

For a given martingale W, we denote by $\langle W \rangle$ and $[W]$, respectively, the predictable and quadratic variation processes associated with W. It is particularly useful to remember that for counting process martingales M_i and M we have that

$$d\langle M_i \rangle_t = (1 - \Delta\Lambda(y))Y_i(t)d\Lambda(t) \text{ and } d\langle M \rangle_t = (1 - \Delta\Lambda(y))Y(t)d\Lambda(t),$$

and note that $1 - \Delta\Lambda(y) = S(y)/S(y-)$.

In our proofs we will constantly use the Lenglart-Rebolledo inequality [13, Theorem 3.4.1], in particular, we will use the fact that if W is a submartingale depending on the number of observations n, with compensator R, then the Lenglart-Rebolledo inequality implies that $\sup_{t \leq T} W(t) \overset{\mathbb{P}}{\to} 0$ for any stopping time T such that $R(T) \overset{\mathbb{P}}{\to} 0$. Here limits are taken as n approaches infinity. Throughout the proofs, we may not explicitly write the dependence on n when writing a stochastic process, e.g. the martingale $M(t)$ depends on all data points.

In this work we will often encounter (sub)martingales with extra parameters, and we will integrate with respect to them. A particular case is stated in the following lemma, whose proof is very simple (and thus omitted).

Lemma 6.5. *Let $(\mathcal{X}, \mathcal{B}, \mu)$ be a σ-finite measure space. Consider the stochastic process $(M_y(t) : t \geq 0, y \in \mathcal{X})$, and assume that*

i) For every fixed $y \in \mathcal{X}$, $M_y(t)$ is a square-integrable $(\mathcal{F}_t)$-martingale, and
ii) for every $t \geq 0$, $\mathbb{E}\left(\int_{\mathcal{X}} M_y(t)^2 \mu(dy)\right) < \infty$.

Then, for fixed $B \in \mathcal{B}$, the stochastic process $W(t) = \int_B M_y(t)^2 d\mu(y)$ is an $(\mathcal{F}_t)$-submartingale, and its compensator, R, is given by $R(t) = \int_B \langle M_y \rangle(t) d\mu(y)$.

Another interesting type of stochastic processes that appear in our proofs are double integrals with respect to martingales. Define the process

$$W(t) = \int_{C_t} h(x, y)dM(x)dM(y),$$

where $C_t = \{(x,y) : 0 < x < y \leq t\}$, and $C_0 = \emptyset$. The natural questions are whether $W(t)$ defines a proper martingale with respect to $(\mathcal{F}_t)_{t\geq0}$ and, if that is the case, what is its predictable variation process (if it exists). We answer these questions below.

Definition 6.6. Define the predictable σ- algebra $\mathcal{P}$ as the σ-algebra generated by the sets of the form

$$(a_1, b_1] \times (a_2, b_2] \times X, \text{ where } 0 \leq a_1 \leq b_1 < a_2 \leq b_2, X \in \mathcal{F}_{a_2},$$

and $\{0\} \times \{0\} \times X$ with $X \in \mathcal{F}_0$.

Let $C = \{(x,y) : 0 < x < y < \infty\}$. A process $(h(x,y) : (x,y) \in C)$ is called elementary predictable if it can be written as a finite sum of indicator functions of sets belonging to the predictable σ-algebra $\mathcal{P}$. On the other hand, if a process h is $\mathcal{P}$-measurable then it is the almost sure limit of elementary predictable functions.

Straightforwardly from Definition 6.6 we get the following proposition.

Proposition 6.7. *If $h_1(t)$ and $h_2(t)$ are predictable w.r.t. $(\mathcal{F}_t)_{t\geq0}$, then $h(x,y) = h_1(x)h_2(y)$ is $\mathcal{P}$-measurable. Also, all deterministic functions are $\mathcal{P}$-measurable.*

Theorem 6.8. *Let h be a $\mathcal{P}$-measurable process, and suppose that for all $t \geq 0$ it holds that*

$$\mathbb{E}\left(\int_{C_t} |h||dM(x)dM(y)|\right) < \infty. \tag{6.4}$$

Then,

$$W(t) = \int_{C_t} h(x,y)dM(x)dM(y)$$

is a martingale on $\mathbb{R}_+$ with respect to the filtration $(\mathcal{F}_t)_{t\geq0}$.
Moreover, if

$$\mathbb{E}\left(\int_{(0,t]} \left(\int_{(0,y)} |h(x,y)||dM(x)|\right)^2 d\langle M\rangle(y)\right) < \infty, \tag{6.5}$$

then $W(t)$ is a square-integrable $(\mathcal{F}_t)$-martingale with predictable variation process $\langle W\rangle$ given by

$$\langle W\rangle(t) = \int_{(0,t]} \left(\int_{(0,y)} h(x,y)dM(x)\right)^2 \frac{Y(y)S(y)}{S(y-)^2}dF(y).$$

The proof of Theorem 6.8 is given in Appendix B.

6.4. Forward operators

Lemma 6.9. *Under Condition 2.1, it holds that*

$$\int_0^\tau \int_0^\tau \frac{(A_1 K)^2(x,y)S(x)}{1 - H(x-)} dF(x)dF(y) < \infty, \tag{6.6}$$

and that

$$\int_0^\tau \int_0^\tau \frac{(A_1 A_2 K)^2(x,y)S(x)S(y)}{1 - H(x-)S(y-)} dF(x)dF(y) < \infty, \tag{6.7}$$

Under Condition 2.2, it holds that

$$\int_0^\tau \int_0^\tau \frac{(A_1 K)^2(x,y)S(x)}{(1 - H(x-))(1 - G(y-))} dF(x)dF(y) < \infty, \tag{6.8}$$

and that

$$\int_0^\tau \int_0^\tau \frac{(A_1 A_2 K)^2(x,y)S(x)S(y)}{(1 - H(x-))(1 - H(y-))} dF(x)dF(y) < \infty. \tag{6.9}$$

The proof of Lemma 6.9 is given in Appendix A.

7. Proofs III: Exterior region

7.1. Proof of Lemma 5.3

In this section we prove Lemma 5.3. Recall that $I^{2c} = (E \times E) \cup (E \times I) \cup (I \times E)$, and, by the symmetry of the kernel K, $\beta_{I^{2c}} = \beta_{E \times E} + 2\beta_{E \times I}$. Then, the result holds by Lemma 7.1 which states $\beta_{E \times E} = o_p(n^{-1/2})$ under Condition 2.1, and $\beta_{E \times E} = o_p(n^{-1})$ under Condition 2.2, and by Lemmas 7.2 and 7.3, which state that $\beta_{E \times I} = o_p(n^{-1/2})$ under Condition 2.1, and $\beta_{E \times I} = o_p(n^{-1})$ under Condition 2.1, respectively.

In the rest of Section 7, we enunciate and prove Lemmas 7.1, 7.2 and 7.3.

Lemma 7.1. *Under Condition 2.1, it holds that $\beta_{E \times E} = o_p(n^{-1/2})$. Moreover, under Condition 2.2, it holds that $\beta_{E \times E} = o_p(n^{-1})$.*

Proof. First, we prove $\beta_{E \times E} = o_p(n^{-1/2})$ under Condition 2.1. Observe that $d\widehat{F}_n(t) = 0$ for all $t > \tau_n$, thus $d(\widehat{F} - F)(x)d(\widehat{F} - F)(y) = dF(x)dF(y)$ in the region $E \times E$. Then, by the Cauchy-Schwartz's inequality, it holds

$$\beta_{E \times E} = \int_{\tau_n}^\tau \int_{\tau_n}^\tau K(x,y)dF(x)dF(y) \leq S(\tau_n)\sqrt{\int_{\tau_n}^\tau \int_{\tau_n}^\tau K(x,y)^2 dF(x)dF(y)}.$$

Multiplying by $\sqrt{(1 - G(\tau_n))/(1 - G(\tau_n))} = 1$, we get

$$\sqrt{n}\beta_{E \times E} \le \sqrt{n \frac{(1 - H(\tau_n))S(\tau_n)}{1 - G(\tau_n)} \int_{\tau_n}^{\tau} \int_{\tau_n}^{\tau} K(x,y)^2 dF(x) dF(y)}$$

$$\le \sqrt{n(1 - H(\tau_n))S(\tau_n) \int_{\tau_n}^{\tau} \int_{\tau_n}^{\tau} \frac{K(x,y)^2}{1 - G(x-)} dF(x) dF(y)} = o_p(1),$$

where the last equality follows from the facts that $n(1 - H(\tau_n)) = O_p(1)$ by Proposition 6.2.iv, and that the double integral tends to 0 since $\tau_n \to \tau$ when n tends to infinity, and by Condition 2.1.

Following the same argument, under Condition 2.2, we get

$$n\beta_{E \times E} \le n(1 - H(\tau_n))\sqrt{\int_{\tau_n}^{\tau} \int_{\tau_n}^{\tau} \frac{K(x,y)^2 dF(x) dF(y)}{(1 - G(x-))(1 - G(y-))}} = o_p(1),$$

since $n(1 - H(\tau_n)) = O_p(1)$, and since the double integral tends to 0 by Condition 2.2, together with the fact that $\tau_n \to \tau$.   □

Lemma 7.2. *Under Condition 2.2, it holds that $\beta_{E \times I} = o_p(n^{-1})$.*

Proof. We start by noticing that if τ is a point of discontinuity of H then $\tau_n = \tau$ almost surely for a sufficiently large n. Consequently, the set $E \times I$ is empty and thus the statement above holds trivially. Therefore, we assume that τ is a continuity point of H.

Replacing Equation (6.2) in $\beta_{E \times I}$ yields

$$|n\beta_{E \times I}| = \left| n \int_{\tau_n}^{\tau} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{Y(x)} \left(\widehat{A}_1 K\right)(x,y) dM(x) dF(y) \right|$$

$$= \left| n \int_{\tau_n}^{\tau} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{Y(x)} \left(\left(\widehat{A}_1 - A_1 + A_1\right) K\right)(x,y) dM(x) dF(y) \right|.$$

Recall that Equation (1.2) states that for $x \le \tau_n$,

$$\left(\left(\widehat{A}_1 - A_1\right) K\right)(x,y) = \frac{1}{S(x)} \int_{\tau_n}^{\tau} K(s,y) dF(s),$$

then

$$n \int_{\tau_n}^{\tau} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{Y(x)} \left(\left(\widehat{A}_1 - A_1\right) K\right)(x,y) dM(x) dF(y)$$

$$= n \int_{\tau_n}^{\tau} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{S(x)Y(x)} \int_{\tau_n}^{\tau} K(s,y) dF(s) dM(x) dF(y)$$

$$= L(\tau_n) \left(n \int_{\tau_n}^{\tau} \int_{\tau_n}^{\tau} K(s,y) dF(s) dF(y) \right) = L(\tau_n)\beta_{E \times E} = o_p(1), \qquad (7.1)$$

where we define $L(\tau_n) = \int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{S(x)Y(x)}dM(x)$. To verify the last equality, we use Lemma 7.1, and the fact that $L(\tau_n) = O_p(1)$, which follows from Lemma 2.4 of Gill [16] that states that $L(\tau_n) = 1 - \widehat{S}_n(\tau_n)/S(\tau_n)$, and then by Proposition 6.2.i, we get $L(\tau_n) = O_p(1)$.

Therefore, from Equation (7.1), we deduce

$$|n\beta_{E\times I}| = \left|n\int_{\tau_n}^{\tau}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)}{Y(x)}(A_1K)(x,y)dM(x)dF(y)\right| + o_p(1) \qquad (7.2)$$

$$= \left|n\int_{\tau_n}^{\tau}M_y^\star(\tau_n)dF(y)\right| + o_p(1),$$

where

$$M_y^\star(t) = \int_0^t\frac{\widehat{S}_n(x-)}{Y(x)}(A_1K)(x,y)dM(x).$$

We proceed to show that $n\int_{\tau_n}^{\tau}M_y^\star(\tau_n)dF(y) = o_p(1)$ which implies that $n\beta_{E\times I} = o_p(1)$.

Notice that for any fixed $y \in \mathbb{R}_+$, $M_y^\star(t)$ is a square-integrable $(\mathcal{F}_t)$-martingale. By applying the Cauchy-Schwartz's inequality, we obtain

$$\left|n\int_{\tau_n}^{\tau}M_y^\star(\tau_n)dF(y)\right| \le n^{1/2}S(\tau_n)^{1/2}\left(n\int_{\tau_n}^{\tau}M_y^\star(\tau_n)^2dF(y)\right)^{1/2}$$

$$\le n^{1/2}\frac{(1-H(\tau_n))^{1/2}}{(1-G(\tau_n))^{1/2}}\left(n\int_{\tau_n}^{\tau}M_y^\star(\tau_n)^2dF(y)\right)^{1/2}$$

$$\le n^{1/2}(1-H(\tau_n))^{1/2}\left(n\int_{\tau_n}^{\tau}\frac{M_y^\star(\tau_n)^2}{1-G(y-)}dF(y)\right)^{1/2}$$

$$= O_p(1)\left(n\int_{\tau_n}^{\tau}\frac{M_y^\star(\tau_n)^2}{1-G(y-)}dF(y)\right)^{1/2},$$

where the last equality follows from Proposition 6.2.iv. We proceed to prove that

$$n\int_{\tau_n}^{\tau}\frac{M_y^\star(\tau_n)^2}{1-G(y-)}dF(y) = o_p(1).$$

Notice that the previous equation considers random integration limits. Our first step will be to prove that τ_n can be replaced by a deterministic value, say T_n, without affecting the result we wish to prove.

Let $C > 0$ be a large constant, define $T_n = \inf\{t > 0 : H(t) \ge 1 - C/n\}$ and the event $B_n = \{1 - C/n \le H(\tau_n)\}$. By Proposition 6.2.iv, it holds $\mathbb{P}(B_n) \ge 1 - e^{-C}$ and, by the definition of T_n, we have that $\{\tau_n \ge T_n\} \subseteq B_n$. Since $\lim_{C\to\infty}\mathbb{P}(B_n) = 1$, it is enough to prove that

$$n\int_{T_n}^{\tau}\frac{M_y^\star(\tau_n)^2}{1-G(y-)}dF(y) = o_p(1).$$

Observe that by Lemma 6.5, the process

$$t \to n \int_{T_n}^{\tau} \frac{M_y^{\star}(t)^2}{1 - G(y-)} dF(y)$$

is an $(\mathcal{F}_t)$-submartingale with compensator, evaluated at $t = \tau_n$, given by

$$
\begin{aligned}
n \int_{T_n}^{\tau} \frac{\langle M_y^{\star} \rangle(\tau_n)}{1 - G(y-)} dF(y) &= n \int_{T_n}^{\tau} \int_0^{\tau_n} (A_1 K)^2(x, y) \frac{\widehat{S}_n(x-)^2 S(x)}{Y(x)S(x-)^2} \frac{dF(x)dF(y)}{(1 - G(y-))} \\
&= O_p(1) \int_{T_n}^{\tau} \int_0^{\tau_n} (A_1 K)^2(x, y) \frac{S(x)}{(1 - H(x-))} \frac{dF(x)dF(y)}{(1 - G(y-))} \\
&= o_p(1),
\end{aligned}
$$

where the second equality is due to Propositions 6.2.i and 6.2.ii, and the third equality holds by noticing that $T_n \to \tau$ and that

$$\int_0^{\tau} \int_0^{\tau} (A_1 K)^2(x, y) \frac{S(x)}{(1 - H(x-))} \frac{dF(x)dF(y)}{(1 - G(y-))} < \infty,$$

by Equation (6.8) of Lemma 6.9 under Condition 2.2. We conclude then that

$$n \int_{T_n}^{\tau} \frac{\langle M_y^{\star} \rangle(\tau_n)}{1 - G(y-)} dF(y) = o_p(1),$$

which, by the Lenglart-Rebolledo inequality, implies

$$n \int_{T_n}^{\tau} \frac{M_y^{\star}(\tau_n)^2}{1 - G(y-)} dF(y) = o_p(1).$$

Since the previous result is valid in the event B_n, which can be chosen with arbitrarily large probability, we conclude

$$n \beta_{E \times I} = O_p(1) \left(n \int_{\tau_n}^{\tau} \frac{M_y^{\star}(\tau_n)^2}{1 - G(y-)} dF(y) \right)^{1/2} + o_p(1) = o_p(1),$$

finishing our proof. □

Lemma 7.3. *Under Assumption 2.1 it holds that $\beta_{E \times I} = o_p(n^{-1/2})$.*

Proof. Following the same steps of the proof of Lemma 7.2, it holds

$$
\begin{aligned}
n^{1/2} \beta_{E \times I} &= n^{1/2} \int_{\tau_n}^{\tau} \int_0^{\tau_n} \frac{\widehat{S}_n(x)}{Y(x)} (A_1 K)(x, y) dM(x) dF(y) + o_p(1) \\
&= n^{1/2} \int_{\tau_n}^{\tau} M_y^{\star}(\tau_n) dF(y) + o_p(1).
\end{aligned}
$$

where $M_y^\star(t) = \int_0^t \frac{\widehat{S}_n(x-)}{Y(x)}(A_1 K)(x,y)dM(x)$ is an $(\mathcal{F}_t)$-martingale for every fixed $y \in \mathbb{R}_+$. Let T_n be the same deterministic sequence used in the proof of Lemma 7.2. Then, it suffices to show that

$$n^{1/2}\int_{T_n}^{\tau} M_y^\star(\tau_n)dF(y) = o_p(1).$$

By the Cauchy-Schwartz's inequality

$$|n^{1/2}\beta_{E\times I}| \leq n^{1/2}\int_{T_n}^{\tau} M_y^\star(\tau_n)dF(y) + o_p(1)$$

$$\leq \left(n\int_{T_n}^{\tau} M_y^\star(\tau_n)^2 dF(y)\right)^{1/2} + o_p(1). \qquad (7.3)$$

Moreover, by Lemma 6.5, $n\int_{T_n}^{\tau} M_y^\star(t)^2 dF(y)$ is an $(\mathcal{F}_t)$-submartingale with compensator, evaluated at $t = \tau_n$, given by

$$n\int_{T_n}^{\tau} \langle M_y^\star \rangle(\tau_n)dF(y) = \int_{T_n}^{\tau}\int_0^{\tau_n} \frac{\widehat{S}(x-)^2}{Y(x)/n}(A_1 K)^2(x,y)\frac{S(x)}{S(x-)^2}dF(x)dF(y)$$

$$= O_p(1)\int_{T_n}^{\tau}\int_0^{\tau_n}(A_1 K)^2(x,y)\frac{S(x)}{1-H(x-)}dF(x)dF(y),$$

where the second equality holds by Propositions 6.2.i and 6.2.ii. We prove that the compensator in the previous equation converges to zero by noticing that $T_n \to \tau$, and that

$$\int_0^{\tau}\int_0^{\tau}(A_1 K)^2(x,y)\frac{S(x)}{1-H(x-)}dF(x)dF(y) < \infty,$$

which holds due to Equation (6.6) of Lemma 6.9 under Condition 2.1.

The previous result implies that $n\int_{T_n}^{\tau}\langle M_y^\star\rangle(\tau_n)dF(y) = o_p(1)$. By the Lenglart-Rebolledo inequality, we deduce $n\int_{T_n}^{\tau} M_y^\star(\tau_n)^2 dF(y) = o_p(1)$, and by substituting this result in Equation (7.3) we get $n^{1/2}\beta_{E\times I} = o_p(1)$. $\qquad\square$

8. Proofs IV: Interior region

8.1. Proof of Lemma 5.4

Proof. We start by proving Equation (5.2) under Condition 2.2. Observe that, by Equation (6.3), it holds that

$$\beta_{I^2} = \int_0^{\tau_n}\int_0^{\tau_n} \frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}(\widehat{A}_2\widehat{A}_1 K)(x,y)dM(x)dM(y).$$

We proceed to prove that $\widehat{A}_1$ and $\widehat{A}_2$ can be replaced in the previous equation by the operators A_1 and A_2, respectively. After that, Equation (5.2) follows immediately by recalling that $K'(x,y) = (A_1 A_2 K)(x,y)$.

We begin with the following equality,

$$\widehat{A}_2\widehat{A}_1 - A_2A_1 = (\widehat{A}_2 - A_2)(\widehat{A}_1 - A_1) + A_2(\widehat{A}_1 - A_1) + A_1(\widehat{A}_2 - A_2),$$

then, by the symmetry of K, we just need to prove that

$$n\int_0^{\tau_n}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}(A_1(\widehat{A}_2 - A_2)K)(x,y)dM(x)dM(y) = o_p(1), \quad (8.1)$$

and

$$n\int_0^{\tau_n}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}((\widehat{A}_2 - A_2)(\widehat{A}_1 - A_1)K)(x,y)dM(x)dM(y) = o_p(1). \tag{8.2}$$

We begin by proving Equation (8.1). From Equation (1.2) we get

$$(A_1(\widehat{A}_2 - A_2)K)(x,y) = \frac{1}{S(y)}\int_{\tau_n}^{\tau}(A_1K)(x,s)dF(s). \tag{8.3}$$

Let $L(\tau_n) = \int_0^{\tau_n}\frac{\widehat{S}(x-)}{S(x)}\frac{dM(x)}{Y(x)}$, then substituting Equation (8.3) in Equation (8.1) yields

$$n\int_0^{\tau_n}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}(A_1(\widehat{A}_2 - A_2)K)(x,y)dM(x)dM(y)$$

$$= n\int_0^{\tau_n}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}\left(\frac{1}{S(y)}\int_{\tau_n}^{\tau}(A_1K)(x,s)dF(s)\right)dM(x)dM(y)$$

$$= nL(\tau_n)\int_0^{\tau_n}\frac{\widehat{S}_n(x-)}{Y(x)}\int_{\tau_n}^{\tau}(A_1K)(x,s)dF(s)dM(x)$$

$$= O_p(1)n\left|\int_{\tau_n}^{\tau}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)}{Y(x)}(A_1K)(x,s)dM(x)dF(s)\right| = n|\beta_{E\times I}| + o_p(1),$$

where the third equality holds as $L(\tau_n) = O_p(1)$ (which is proved in Lemma 7.2), and the last equality is exactly Equation (7.2). Hence, by Lemma 7.2, we deduce that Equation (8.1) holds true.

For Equation (8.2), a similar computation yields

$$n\int_0^{\tau_n}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)}((\widehat{A}_2 - A_2)(\widehat{A}_1 - A_1)K)(x,y)dM(x)dM(y)$$

$$= nL(\tau_n)^2\int_{\tau_n}^{\infty}\int_{\tau_n}^{\infty}K(s,t)dF(s)dF(t)$$

$$= O_p(1)n\left|\int_{\tau_n}^{\infty}\int_{\tau_n}^{\infty}K(s,t)dF(s)dF(t)\right| = O_p(1)n|\beta_{E\times E}| = o_p(1),$$

where the last equality holds by Lemma 7.1.

To finish the proof of Lemma 5.4 we need to check Equation (5.1) holds under Condition 2.1, which follows from repeating the same steps but replacing the scaling factor n by $\sqrt{n}$. $\qquad\square$

9. Proofs V: Double stochastic integral

In this section we prove Lemma 5.5 and Theorem 2.6. To begin with, from Lemmas 5.3 and 5.4, we deduce that

$$\beta = \int_0^{\tau_n} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)} K'(x,y)dM(x)dM(y) + o_p(n^{-c}) \qquad (9.1)$$

holds for $c = 1/2$ under Condition 2.1, and for $c = 1$ under Condition 2.2. The form of β suggests that we need to study the double stochastic integral process given by

$$Q(t) = \int_0^t \int_0^t \frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)} K'(x,y)dM(x)dM(y).$$

The strategy to study $Q(t)$ is to consider its decomposition into a diagonal and an off-diagonal term, and to analyse them individually. To this end, we define the sets $D(t) = \{(x,y) : x = y, 0 < x \le t\}$ and $C(t) = \{(x,y) : 0 < x < y \le t\}$, and define the processes

$$Q_D(t) = \int_{D(t)} \frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)} K'(x,y)dM(x)dM(y),$$

and

$$Q_C(t) = \int_{C(t)} \frac{\widehat{S}_n(x-)\widehat{S}_n(y-)}{Y(x)Y(y)} K'(x,y)dM(x)dM(y).$$

Notice that $Q(t) = Q_D(t) + 2Q_C(t)$ follows by the symmetry of $K' = (A_1 A_2 K)$.

The proofs of Lemma 5.5 and Theorem 2.6 are an immediate consequence of the following results concerning the process $Q(t)$.

Lemma 9.1. *Under Condition 2.1 it holds that* $\sqrt{n}Q_D(\tau_n) \xrightarrow{\mathrm{P}} 0.$

Lemma 9.2. *Under Condition 2.2 it holds that*

$$nQ_D(\tau_n) = \frac{1}{n}\int_0^{\tau_n} \frac{K'(x,x)}{(1-G(x-))^2} d[M](x) + o_p(1).$$

Lemma 9.3. *Under Condition 2.1 it holds that* $\sqrt{n}Q_C(\tau_n) \xrightarrow{\mathrm{P}} 0.$

Lemma 9.4. *Under Condition 2.2 it holds that*

$$nQ_C(\tau_n) = \frac{1}{n}\int_0^{\tau_n} \int_{(0,y)} \frac{K'(x,y)}{(1-G(x-))(1-G(y-))} dM(x)dM(y) + o_p(1).$$

Proof of Lemma 5.5. Starting from Equation (9.1), we get $\sqrt{n}\beta = \sqrt{n}Q(\tau_n) + o_p(1) = \sqrt{n}Q_D(\tau_n) + 2\sqrt{n}Q_C(\tau_n) + o_p(1)$. Then, the result follows from Lemmas 9.1 and 9.3. $\square$

Proof of Theorem 2.6. From Equation (9.1) it holds that $n\beta = nQ(\tau_n) + o_p(1) = nQ_D(\tau_n) + 2nQ_C(\tau_n) + o_p(1)$. Then a direct application of Lemmas 9.2 and 9.4 yields

$$
\begin{aligned}
n\beta &= nQ_D(\tau_n) + 2nQ_C(\tau_n) + o_p(1) \\
&= \frac{1}{n}\int_0^{\tau_n}\int_0^{\tau_n}\frac{K'(x,y)}{(1-G(x-))(1-G(y-))}dM(x)dM(y) + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\int_0^{\tau_n}\int_0^{\tau_n}\frac{K'(x,y)}{(1-G(x-))(1-G(y-))}dM_i(x)dM_j(y) + o_p(1).
\end{aligned}
$$

$\square$

It just remains to prove Lemmas 9.1, 9.2, 9.3, and 9.4.

9.1. Integral over diagonal $D(t)$: Proof of Lemmas 9.1 and 9.2

Observe that $Q_D(t)$ satisfies

$$
Q_D(t) = \int_0^t \frac{\widehat{S}_n(x-)^2}{Y(x)^2}K'(x,x)d[M](x).
$$

The latter can be checked by noticing that the measure $dM(x)dM(y)$ of a small square whose main diagonal goes from (a,a) to (b,b) is $(M(b)-M(a))^2$. When b approaches a from above, we have that $(M(b)-M(a))^2 \to (\Delta M(a))^2$ (the limit is well-defined for M). Since M is the difference of two increasing processes we have that the number of discontinuities is at most countable, then $\int_0^t f(x)d[M](x) = \sum_{x\leq t} f(x)(\Delta M(x))^2$.

To analyse the process $Q_D(t)$, recall that $[M]$ is a submartingale with compensator given by $\langle M\rangle$. Thus, for any predictable process $H \geq 0$, $\int_0^t H(x)d[M](x)$ is a submartingale with compensator given by $\int_0^t H(x)d\langle M\rangle(x)$. Finally, by the Lenglart-Rebolledo inequality, if we have $\int_0^{\tau_n} H(x)d\langle M\rangle(x) \xrightarrow{\mathbb{P}} 0$, then we get that $\int_0^{\tau_n} H(x)d[M](x) \xrightarrow{\mathbb{P}} 0$.

Proof of Lemma 9.1. Define the $(\mathcal{F}_t)$-submartingale

$$
W(t) = \int_0^t \frac{\widehat{S}_n(x-)^2}{Y(x)^2}|K'(x,x)|d[M](x),
$$

and observe that $|\sqrt{n}Q_D(t)| \leq \sqrt{n}W(t)$. Thus, it is enough to prove that $\sqrt{n}W(t) = o_p(1)$. Abusing notation, denote by $\langle W\rangle(t)$ the compensator of $W(t)$. Then we will prove that $\sqrt{n}\langle W\rangle(\tau_n) = o_p(1)$, and thus, by the Lenglart-Rebolledo inequality, we will get $\sqrt{n}W(\tau_n) = o_p(1)$.

Observe that

$$
\sqrt{n}\langle W\rangle(\tau_n) = \sqrt{n}\int_0^{\tau_n}\frac{\widehat{S}_n(x-)^2}{Y(x)^2}|K'(x,x)|d\langle M\rangle(x)
$$

$$= \sqrt{n} \int_0^{\tau_n} \frac{\widehat{S}_n(x-)^2}{Y(x)^2} |K'(x,x)| \frac{Y(x)S(x)}{S(x-)^2} dF(x)$$

$$= \int_0^{\tau_n} \frac{\widehat{S}_n(x-)^2}{S(x-)^2} \frac{S(x)}{Y(x)/\sqrt{n}} |K'(x,x)| dF(x)$$

$$= O_p(1) \int_0^{\tau} \frac{\mathbb{1}_{\{x \le \tau_n\}}}{\sqrt{Y(x)}} \sqrt{\frac{S(x)}{1-G(x-)}} |K'(x,x)| dF(x),$$

where the fourth equality follows from Propositions 6.2.i and 6.2.ii. Finally, we claim that $\int_0^{\tau} \frac{1}{\sqrt{Y(x)}} \sqrt{\frac{S(x)}{1-G(x-)}} |K'(x,x)| dF(x) = o(1)$. This is verified by applying Dominated Convergence Theorem. Indeed, notice that $\frac{\mathbb{1}_{\{x \le \tau_n\}}}{\sqrt{Y(x)}} \to 0$ for each fixed $x \in I_H$, thus the integrand tends to zero. Moreover, by using that $Y(x) \ge 1$ for $x \le \tau_n$, the integrand is bounded by an integrable function due to Condition 2.1. $\qquad\square$

Proof of Lemma 9.2. Observe that it is enough to prove that

$$\frac{1}{n} \int_0^{\tau_n} \left| \frac{\widehat{S}_n(x-)^2}{Y(x)^2/n^2} - \frac{1}{(1-G(x-))^2} \right| |K'(x,x)| d[M](x) = o_p(1).$$

Write $U_n(x) = \left| \frac{\widehat{S}_n(x-)^2}{Y(x)^2/n^2} - \frac{1}{(1-G(x-))^2} \right|$, which is predictable w.r.t. $(\mathcal{F}_x)_{x \ge 0}$. Also, define the process $W(t) = \frac{1}{n} \int_0^t U_n(x)|K'(x,x)| d[M](x)$, and observe that it corresponds to an $(\mathcal{F}_t)$-submartingale. We prove that $W(\tau_n) = o_p(1)$ by using the Lenglart-Rebolledo inequality. For such, we have to prove that its compensator, which by abusing notation we denote by $\langle W \rangle$, satisfies $\langle W \rangle(\tau_n) = o_p(1)$. A simple computation shows

$$\langle W \rangle(\tau_n) = \frac{1}{n} \int_0^{\tau_n} U_n(x)|K'(x,x)| d\langle M \rangle(x)$$

$$= \frac{1}{n} \int_0^{\tau_n} U_n(x)|K'(x,x)| \frac{Y(x)S(x)}{S(x-)^2} dF(x)$$

$$= O_p(1) \int_0^{\tau} \mathbb{1}_{\{x \le \tau_n\}} U_n(x)|K'(x,x)| \frac{(1-G(x-))S(x)}{S(x-)} dF(x)$$

$$= o_p(1),$$

where the third equality follows from Proposition 6.2.iii, and the last equality follows from applying Lemma 6.4, whose conditions we proceed to verify: for the first condition, Proposition 6.1 yields $U_n(x) = \left| \frac{\widehat{S}_n(x-)^2}{Y(x)^2/n^2} - \frac{1}{(1-G(x-))^2} \right| \to 0$ for every fixed $x \in I_H$. For the second condition, set $R(x) = \frac{|K'(x,x)|S(x)}{1-H(x-)}$, which is integrable by Condition 2.2, then, Propositions 6.2.i and 6.2.ii give $U_n(x) = O_p(1) \frac{1}{(1-G(x-))^2}$, uniformly for all $x \le \tau_n$. Then

$$\mathbb{1}_{\{x \le \tau_n\}} U_n(x)|K'(x,x)| \frac{(1-G(x-))S(x)}{S(x-)}$$

$$= O_p(1)|K'(x,x)|\frac{S(x)}{(1-G(x-))S(x-)} = O_p(1)R(x),$$

uniformly for all $x \le \tau$. $\qquad\square$

9.2. Integral over off-diagonal $C(t)$: Proof of Lemmas 9.3 and 9.4

Proof of Lemma 9.3. From Theorem 6.8, $Q_C(t)$ is a square-integrable $(\mathcal{F}_t)$-martingale with mean 0. Then, by the Lenglart-Rebolledo inequality, it is enough to prove its predictable variation process, denoted by $\langle Q_C \rangle(t)$, satisfies the equality $n\langle Q_C \rangle(\tau_n) = o_p(1)$.

From Theorem 6.8 we have that $n\langle Q \rangle(\tau_n)$ is equal to

$$n \int_0^{\tau_n} \frac{\widehat{S}_n(y-)^2 S(y)}{Y(y)S(y-)^2} \left(\int_{(0,y)} \frac{\widehat{S}_n(x-)}{Y(x)}K'(x,y)dM(x) \right)^2 dF(y)$$

$$= O_p(1) \int_0^{\tau} \frac{S(y)}{1-H(y-)} \left(\int_0^{\tau_n} \frac{\widehat{S}_n(x-)}{Y(x)}K'(x,y)\mathbb{1}_{\{x<y\}}dM(x) \right)^2 dF(y)$$

$$= O_p(1) \int_0^{\tau} \frac{S(y)}{1-H(y-)} M_y^{\star}(\tau_n)^2 dF(y), \qquad (9.2)$$

where the first equality is due to Propositions 6.2.i and 6.2.ii, and in the second equality we define $M_y^{\star}(t) = \int_0^t \frac{\widehat{S}_n(x-)}{Y(x)}K'(x,y)\mathbb{1}_{\{x<y\}}dM(x)$, which is a square-integrable $(\mathcal{F}_t)$-martingale for any fixed $y \in I_H$.

Define the process $Z(t) = \int_0^{\tau} \frac{S(y)}{1-H(y-)}M_y^{\star}(t)^2 dF(y)$, and notice that the integral in Equation (9.2) corresponds to $Z(\tau_n)$. By Lemma 6.5, $Z(t)$ is an $(\mathcal{F}_t)$-submartingale, hence we shall use the Lenglart-Rebolledo inequality to prove $Z(\tau_n) = o_p(1)$ by proving that the compensator of $Z(t)$, which by abusing notation we denote by $\langle Z \rangle(t)$, satisfies $\langle Z \rangle(\tau_n) = o_p(1)$. From Lemma 6.5, and Propositions 6.2.i and 6.2.ii, we have that

$$\langle Z \rangle(\tau_n) = \int_0^{\tau} \int_0^{\tau_n} \frac{S(y)}{1-H(y-)} \frac{\widehat{S}_n(x-)^2}{Y(x)}K'(x,y)^2\mathbb{1}_{\{x<y\}}\frac{S(x)}{S(x-)^2}dF(x)dF(y)$$

$$= O_p(1) \int_0^{\tau} \int_0^{\tau} \frac{\mathbb{1}_{\{x\le\tau_n\}}S(y)S(x)}{Y(x)(1-H(y-))}K'(x,y)^2\mathbb{1}_{\{x<y\}}dF(x)dF(y).$$

We claim that the previous quantity tends to 0 as n approaches infinity by an application of the Dominated Convergence Theorem. Indeed, notice that $\frac{\mathbb{1}_{\{x\le\tau_n\}}}{Y(x)} \to 0$ for any fixed $x \in I_H$, and that

$$\frac{\mathbb{1}_{\{x\le\tau_n\}}}{Y(x)} \frac{S(y)S(x)}{(1-H(y-))}K'(x,y)^2\mathbb{1}_{\{x<y\}} \le \frac{S(y)S(x)}{(1-H(y-))}K'(x,y)^2,$$

which is integrable by Equation (6.7) of Lemma 6.9 (Recall that $K' = A_1A_2K$). $\qquad\square$

Proof of Lemma 9.4. The result follows from proving that

$$\frac{1}{n} \int_{C(\tau_n)} \left(\frac{\widehat{S}_n(x-)\widehat{S}_n(y-)K'(x,y)}{(Y(x)/n)(Y(y)/n)} - \frac{K'(x,y)}{(1-G(x-))(1-G(y-))} \right) dM(x)dM(y)$$

is $o_p(1)$, which is equivalent to prove that

$$\frac{1}{n} \int_0^{\tau_n} \int_{(0,y)} \left(\frac{\widehat{S}_n(x-)}{Y(x)/n} - \frac{1}{1-G(x-)} \right) \frac{\widehat{S}_n(y-)K'(x,y)}{Y(y)/n} dM(x)dM(y) = o_p(1), \tag{9.3}$$

and that

$$\frac{1}{n} \int_0^{\tau_n} \int_{(0,y)} \left(\frac{\widehat{S}_n(y-)}{Y(y)/n} - \frac{1}{1-G(y-)} \right) \frac{K'(x,y)}{(1-G(x-))} dM(x)dM(y) = o_p(1). \tag{9.4}$$

We only prove Equation (9.3), as Equation (9.4) follows by similar arguments.

Define $U_n(x) = \left(\frac{\widehat{S}_n(x-)}{Y(x)/n} - \frac{1}{1-G(x-)} \right)$ which is predictable w.r.t. $(\mathcal{F}_x)_{x \geq 0}$, and define the process $W(t)$ as

$$W(t) = \int_0^t \int_{(0,y)} U_n(x) \frac{\widehat{S}_n(y-)K'(x,y)}{Y(y)} dM(x)dM(y),$$

which, by Theorem 6.8, is a square-integrable $(\mathcal{F}_t)$-martingale. We just need to prove that $W(\tau_n) = o_p(1)$. By the Lenglart-Rebolledo inequality, it is enough to check that the predictable variation process of $W(t)$, $\langle W \rangle(t)$, satisfies $\langle W \rangle(\tau_n) = o_p(1)$. From Theorem 6.8, we have

$$\langle W \rangle(\tau_n) = \int_0^{\tau_n} \left(\int_{(0,y)} U_n(x)K'(x,y)dM(x) \right)^2 \frac{\widehat{S}_n(y-)^2 S(y)}{Y(y)S(y-)^2} dF(y)$$

$$= O_p(1) \frac{1}{n} \int_0^{\tau} \left(\int_0^{\tau_n} U_n(x)K'(x,y)\mathbb{1}_{\{x<y\}} dM(x) \right)^2 \frac{S(y)}{1-H(y-)} dF(y),$$

$$= O_p(1) \frac{1}{n} \int_0^{\tau} M_y^{\star}(\tau_n)^2 \frac{S(y)}{1-H(y-)} dF(y), \tag{9.5}$$

where the second equality is due to Propositions 6.2.i and 6.2.ii, and in the third equality we define $M_y^{\star}(t) = \int_0^t U_n(x)K'(x,y)\mathbb{1}_{\{x<y\}} dM(x)$.

We proceed to check that Equation (9.5) is $o_p(1)$. Observe that for any fixed $y \in I_H$, $M_y^{\star}(t) = \int_0^t U_n(x)K'(x,y)\mathbb{1}_{\{x<y\}} dM(x)$ is a square-integrable $(\mathcal{F}_t)$-martingale, thus, by Lemma 6.5, the process $Z(t) = \frac{1}{n} \int_0^{\tau} M_y^{\star}(t)^2 \frac{S(y)dF(y)}{1-H(y-)}$ is an $(\mathcal{F}_t)$-submartingale. Note that $\langle W \rangle(\tau_n) = Z(\tau_n)$. We check that $Z(\tau_n) = o_p(1)$ by verifying that the compensator of Z, which by abusing notation we denote by $\langle Z \rangle(t)$, satisfies $\langle Z \rangle(\tau_n) = o_p(1)$. From Lemma 6.5, $\langle Z \rangle(\tau_n)$ is given by

$$\frac{1}{n}\int_0^\tau\int_0^{\tau_n} U_n(x)^2 K'(x,y)^2 \frac{Y(x)S(x)}{S(x-)^2}dF(x)\frac{S(y)}{1-H(y-)}dF(y)$$

$$= O_p(1)\int_0^\tau\int_0^\tau \mathbb{1}_{\{y\le\tau_n\}}U_n(x)^2 K'(x,y)^2\frac{(1-H(x-))S(x)S(y)}{(1-H(y-))S(x-)^2}dF(x)dF(y),$$

$$(9.6)$$

where the equality follows from Proposition 6.2.iii. We shall verify the conditions of Lemma 6.4 to prove that Equation (9.6) is $o_p(1)$. Set $R(x,y) = \frac{K'(x,y)^2 S(x)S(y)}{(1-H(x-)(1-H(y-))}$ and $R_n(x,y)$ as the integrand in Equation (9.6). To verify the first condition of Lemma 6.4, note that $R_n(x)\to 0$ for each $x\in I_H$, since $U_n(x)\to 0$ by Proposition 6.1. To verify the second condition, Propositions 6.2.i and 6.2.ii yield $U_n(x) = O_p(1)(1-G(x-))^{-1}$ uniformly for all $x\le\tau_n$, thus the integrand satisfies

$$R_n(x,y) = \mathbb{1}_{\{y\le\tau_n\}}U_n(x)^2 K'(x,y)^2\frac{(1-H(x-))S(x)S(y)}{(1-H(y-))S(x-)^2}$$

$$= O_p(1)\frac{K'(x,y)^2 S(x)S(y)}{(1-H(x-))(1-H(y-))} = O_p(1)R(x,y),$$

uniformly in x. Finally, the function $R(x,y) = \frac{K'(x,y)^2 S(x)S(y)}{(1-H(y-))(1-H(x-))}$ is integrable due to Equation (6.9) of Lemma 6.9 (recall that $K' = A_1 A_2 K$). □

Acknowledgement

References

[1] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. *Survival and event history analysis: a process point of view.* Springer Science & Business Media, 2008. MR2449233

[2] Michael G. Akritas. The central limit theorem under censoring. *Bernoulli*, 6(6):1109–1120, 2000. ISSN 1350-7265. URL https://doi.org/10.2307/3318473. MR1809737

[3] Eric Beutner and Henryk Zähle. Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli*, 18(3):803–822, 2012. ISSN 1350-7265. URL https://doi.org/10.3150/11-BEJ358. MR2948902

[4] Eric Beutner and Henryk Zähle. Continuous mapping approach to the asymptotics of U- and V-statistics. *Bernoulli*, 20(2):846–877, 2014. ISSN 1350-7265. URL https://doi.org/10.3150/13-BEJ508. MR3178520

[5] Arup Bose and Arusharka Sen. The strong law of large numbers for Kaplan-Meier U-statistics. *J. Theoret. Probab.*, 12(1):181–200, 1999. ISSN 0894-9840. URL https://doi.org/10.1023/A:1021752828590.

[6] Arup Bose and Arusharka Sen. Asymptotic distribution of the Kaplan-Meier U-statistics. *J. Multivariate Anal.*, 83(1):84–123, 2002. ISSN 0047-259X. URL https://doi.org/10.1006/jmva.2001.2039. MR1934976

[7] Somnath Datta, Dipankar Bandyopadhyay, and Glen A Satten. Inverse probability of censoring weighted u-statistics for right-censored data with an application to testing hypotheses. *Scandinavian Journal of Statistics*, 37(4):680–700, 2010. MR2779643

[8] Herold Dehling and Martin Wendler. Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *J. Multivariate Anal.*, 101(1):126–137, 2010. ISSN 0047-259X. URL https://doi.org/10.1016/j.jmva.2009.06.002. MR2557623

[9] Manfred Denker and Gerhard Keller. Rigorous statistical procedures for data from dynamical systems. *J. Statist. Phys.*, 44(1-2):67–93, 1986. ISSN 0022-4715. URL https://doi.org/10.1007/BF01010905.

[10] Isha Dewan and B. L. S. Prakasa Rao. Central limit theorem for U-statistics of associated random variables. *Statist. Probab. Lett.*, 57(1):9–15, 2002. ISSN 0167-7152. URL https://doi.org/10.1016/S0167-7152(01)00194-8. MR1911807

[11] Bradley Efron and Iain M. Johnstone. Fisher's information in terms of the hazard rate. *Ann. Statist.*, 18(1):38–62, 1990. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176347492.

[12] Tamara Fernandez and Nicolas Rivera. A reproducing kernel hilbert space log-rank test for the two-sample problem. *arXiv preprint arXiv:1904.05187*, 2019.

[13] Thomas R. Fleming and David P. Harrington. *Counting processes and survival analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1991. ISBN 0-471-52218-X. MR1100924

[14] Irène Gijbels and Noël Veraverbeke. Almost sure asymptotic representation for a class of functionals of the Kaplan-Meier estimator. *Ann. Statist.*, 19(3):1457–1470, 1991. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176348256.

[15] R. D. Gill. *Censoring and stochastic integrals*, volume 124 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1980. ISBN 90-6196-197-1. MR0596815

[16] Richard Gill. Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.*, 11(1):49–58, 1983. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176346055.

[17] Svante Janson. Probability asymptotics: notes on notation. *arXiv preprint arXiv:1108.3924*, 2011.

[18] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, 53:457–481, 1958. ISSN 0162-1459. URL http://links.jstor.org/sici?sici=0162-1459(195806)53:

282<457:NEFIO>2.0.CO;2-Z&origin=MSN. MR0093867

[19] V. S. Koroljuk and Yu. V. Borovskich. *Theory of U-statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994. ISBN 0-7923-2608-3. URL https://doi.org/10.1007/978-94-017-3515-5. MR1472486

[20] James A. Koziol and Sylvan B. Green. A Cramér-von Mises statistic for randomly censored data. *Biometrika*, 63(3):465–474, 1976. ISSN 0006-3444. URL https://doi.org/10.1093/biomet/63.3.465. MR0448695

[21] Marcos Matabuena and Oscar Hernan Madrid Padilla. Energy distance and kernel mean embeddings for two-sample survival testing. *arXiv preprint arXiv:1912.04160*, 2019.

[22] David Rindt, Dino Sejdinovic, and David Steinsaltz. Nonparametric independence testing for right-censored data using optimal transport. *arXiv preprint arXiv:1906.03866*, 2019.

[23] Ya'acov Ritov and Jon A. Wellner. Censoring, martingales, and the Cox model. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, volume 80 of *Contemp. Math.*, pages 191–219. Amer. Math. Soc., Providence, RI, 1988. URL https://doi.org/10.1090/conm/080/999013. MR0999013

[24] Glen A. Satten and Somnath Datta. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.*, 55(3):207–210, 2001. ISSN 0003-1305. URL https://doi.org/10.1198/000313001317098185.

[25] Pranab Kumar Sen. Limiting behavior of regular functionals of empirical distributions for stationary mixing processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 25:71–82, 1972/73. URL https://doi.org/10.1007/BF00533337. MR0329003

[26] Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.*, 11:1517–1561, 2010. ISSN 1532-4435. MR2645460

[27] W. Stute and J.-L. Wang. The strong law under random censorship. *Ann. Statist.*, 21(3):1591–1607, 1993. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176349273.

[28] Winfried Stute. The central limit theorem under random censorship. *Ann. Statist.*, 23(2):422–439, 1995. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176324528.

[29] Song Yang. A central limit theorem for functionals of the Kaplan-Meier estimator. *Statist. Probab. Lett.*, 21(5):337–345, 1994. ISSN 0167-7152. URL https://doi.org/10.1016/0167-7152(94)00026-3. MR1325208

[30] Ken-ichi Yoshihara. Limiting behavior of *U*-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(3):237–252, 1976. URL https://doi.org/10.1007/BF00532676. MR0418179

Appendix A: Proof of Lemma 6.9

In order to prove Lemma 6.9, we introduce the operator $R : L^2(F) \to L^2(F)$,

$$(Rg)(x) = \frac{1}{S(x)} \int_x^\infty g(t) dF(t).$$

Note that the operator A can be written as $A = \mathrm{Id} - R$, where Id is the identity operator. Additionally, for bivariate functions $K : \mathbb{R}_+^2 \to \mathbb{R}$, we define $R_1 K$ and $R_2 K$ as the operator R applied on the first and second coordinate of K, respectively. Note that R_1 and R_2 commute

Let $X \sim F$ and $g \in L^2(F)$, then we claim the operator R satisfies that

$$\mathbb{E}\left(\frac{S(X)}{S(X-)}(Rg)(X)^2 \right) \leq 4\mathbb{E}(g(X)^2). \tag{A.1}$$

The previous equation follows from Equation (4.3) of Efron and Johnstone [11], which states that

$$\mathbb{V}ar(g(X)) = \mathbb{E}\left(\frac{S(X)}{S(X-)}(Ag)(X)^2 \right). \tag{A.2}$$

Then, by using that $(Rg)^2(X) \leq 2(g(X)^2 + (Ag)(X)^2)$, we get

$$\mathbb{E}\left(\frac{S(X)}{S(X-)}(Rg)(X)^2 \right) \leq 2\mathbb{E}\left(\frac{S(X)}{S(X-)}g(X)^2 \right) + 2\mathbb{E}\left(\frac{S(X)}{S(X-)}(Ag)(X)^2 \right)$$

$$\leq 2\mathbb{E}\left(g(X)^2 \right) + 2\mathbb{V}ar(g(X)^2) \leq 4\mathbb{E}(g(X)^2),$$

where in the last step we used Equation (A.2).

Let $\Gamma(x,y) = \frac{|K(x,y)|}{\sqrt{1-G(x-)}}$ and $\Sigma(x,y) = \frac{|K(x,y)|}{\sqrt{(1-G(x-))(1-G(y-))}}$, and notice that Conditions 2.1 and 2.2 imply $\Gamma \in L^2(F \times F)$ and $\Sigma \in L^2(F \times F)$, respectively.

Assume Condition 2.1 holds, then a simple computation shows

$$\int_0^\tau \int_0^\tau \frac{(R_1 K)(x,y)^2 S(x)}{1-H(x-)} dF(x) dF(y)$$

$$= \int_0^\tau \int_0^\tau \frac{S(x)}{1-H(x-)} \left(\frac{1}{S(x)} \int_x^\tau K(s,y) dF(s) \right)^2 dF(x) dF(y)$$

$$\leq \int_0^\tau \int_0^\tau \frac{S(x)}{S(x-)} \left(\frac{1}{S(x)} \int_x^\tau \frac{|K(s,y)|}{\sqrt{1-G(s-)}} dF(s) \right)^2 dF(x) dF(y)$$

$$= \int_0^\tau \int_0^\tau \frac{S(x)}{S(x-)}(R_1 \Gamma)(x,y)^2 dF(x) dF(y)$$

$$\leq 4 \int_0^\tau \int_0^\tau \Gamma(x,y)^2 dF(x) dF(y) < \infty, \tag{A.3}$$

where the last equation follows from Equation (A.1), and Condition 2.1. An-

other similar computation shows

$$\int_0^\tau \int_0^\tau \frac{(R_2 R_1 K)(x,y)^2 S(x) S(y)}{(1 - H(x-)) S(y-)} dF(y) dF(x)$$

$$\leq 4 \int_0^\tau \int_0^\tau \frac{(R_1 K)(x,y)^2 S(x)}{(1 - H(x-))} dF(y) dF(x)$$

$$\leq 16 \int_0^\tau \int_0^\tau \Gamma(x,y)^2 dF(y) dF(x) < \infty \tag{A.4}$$

where the first inequality follows from Equation (A.1) applied on R_2 (i.e. applied on y). The second inequality is exactly Equation (A.3).

Similar computations, show that under Condition 2.2, we have

$$\int_0^\tau \int_0^\tau \frac{(R_1 K)(x,y)^2 S(x)}{(1 - H(x-))(1 - G(y-))} dF(x) \leq 4 \int_0^\tau \int_0^\tau \Sigma(x,y)^2 dF(x) dF(y) < \infty, \tag{A.5}$$

and that

$$\int_0^\tau \frac{(R_2 R_1 K)(x,y)^2 S(x) S(y)}{(1 - H(x-))(1 - H(y-))} dF(x) dF(y)$$

$$\leq 4 \int_0^\tau \int_0^\tau \frac{S(y)}{1 - H(y-)} (R_2 \Gamma(x,y))^2 dF(x) dF(y)$$

$$\leq 16 \int_0^\tau \int_0^\tau \Sigma(x,y)^2 dF(x) dF(y) < \infty. \tag{A.6}$$

From here, under Condition 2.1, Equation (6.6) is a straightforward consequence of Equation (A.3), since

$$\int_0^\tau \int_0^\tau \frac{(A_1 K)^2(x,y) S(x)}{1 - H(x-)} dF(x) dF(y)$$

$$\leq 2 \int_0^\tau \int_0^\tau \frac{(K(x,y)^2 + (R_1 K)(x,y)^2) S(x)}{1 - H(x-)} dF(x) dF(y).$$

Also, Equation (6.7) follows directly from Equations (A.3) and (A.4) since

$$\int_0^\tau \int_0^\tau \frac{(A_1 A_2 K)^2(x,y) S(x) S(y)}{(1 - H(x-)) S(y-)} dF(x) dF(y)$$

$$\leq 4 \int_0^\tau \int_0^\tau \frac{\Phi(x,y) S(x) S(y)}{(1 - H(x-)) S(y-)} dF(x) dF(y),$$

where

$$\Phi(x,y) = K(x,y)^2 + (R_1 K)(x,y)^2 + (R_2 K)(x,y)^2 + (R_1 R_2 K)(x,y)^2.$$

Condition 2.2, together with Equations (A.5) and (A.6), yields Equations (6.8) and (6.9) by following the exact same procedure.

Appendix B: Proof of Theorem 6.8

We proceed to prove Theorem 6.8. Let h be $\mathcal{P}$-measurable. As $M(x)$ is the difference between two right-continuous increasing processes, we have

$$\int_{C_t} h(x,y)dM(x)dM(y) = \int_0^t \int_{(0,y)} h(x,y)dM(x)dM(y).$$

We proceed to prove that the process $\phi(y) = \int_{(0,y)} h(x,y)dM(x)$ is predictable with respect to the sigma-algebra $(\mathcal{F}_y)_{y \geq 0}$. For this, it is enough to verify the claim for elementary functions of $\mathcal{P}$, and then we extend the result to general functions in $\mathcal{P}$.

If $h(x,y) = X\mathbb{1}_{\{(x,y) \in (a_1,b_1] \times (a_2,b_2]\}}$ with $X \in \mathcal{F}_{a_2}$ and $0 \leq a_1 \leq b_1 \leq a_2 \leq b_2$, then

$$\int_{(0,y)} h(x,y)dM(x) = X\mathbb{1}_{(a_2,b_2]}(y) \int_{(0,y)} \mathbb{1}_{(a_1,b_1]}(x)dM(x),$$

which is predictable with respect to $(\mathcal{F}_y)_{y \geq 0}$ since both processes, $X\mathbb{1}_{(a_2,b_2]}(y)$ and $\int_{(0,y)} \mathbb{1}_{(a_1,b_1]}(x)dM(x)$, are adapted to $(\mathcal{F}_y)_{y \geq 0}$ and left-continuous. For the first process note that it is important that $X \in \mathcal{F}_{a_2}$ to ensure it is adapted, and for the second one it is key that we are integrating on $(0,y)$ instead of $(0,y]$ to ensure it is left-continuous. Therefore, the process

$$Z(t) = \int_0^t \int_{(0,y)} h(x,y)dM(x)dM(y)$$

is the integral of a predictable process, and thus Z_t is an $(\mathcal{F}_t)$-martingale. By using Equation (6.4) together with Lebesgue Dominated Convergence theorem, we extend the result to general functions h of the predictable sigma algebra $\mathcal{P}$. From Equation (6.5), we get that $Z(t)$ is a square-integrable process, and its predictable variation process is given by

$$\langle Z \rangle(t) = \int_0^t \left(\int_{(0,y)} h(x,y)dM(x) \right)^2 \frac{Y(y)S(y)}{S(y-)^2} dF(y).$$

Appendix C: Properties of J

In this section we show

   i. $\mathbb{E}(J((X_1,\Delta_1),(x,r))) = 0$ for any $(x,r) \in I_H \times \{0,1\}$.
   ii. $\mathbb{E}(J((X_1,\Delta_1),(X_1,\Delta_1))) = \int_0^\tau \frac{S(x)}{1-H(x-)} K'(x,x)dF(x)$
   iii. $\mathbb{E}(J((X_1,\Delta_1),(X_2,\Delta_2))^2) = \int_0^\tau \int_0^\tau \frac{S(x)S(y)K'(x,y)^2}{(1-H(x-))(1-H(y-))} dF(x)dF(y)$

We start with item i. To ease notation, define $\tilde{K}(x,y) = K'(x,y)/((1 - G(x-)(1-G(y-)))$ and let $dm(s) = r\delta_x(ds) - \mathbb{1}_{\{x \geq s\}}d\Lambda(s)$, then

$$J((X_1,\Delta_1),(x,r)) = \int_0^{X_1} \left(\int_0^x \tilde{K}(s,t)dm(s) \right) dM_1(t).$$

As the term inside the parenthesis is a deterministic function of t, the previous integral is just a stochastic integral with respect to the zero mean martingale M_1, then by the Optional Stopping Theorem, its expected value is 0.

We continue with item ii. Observe that

$$\mathbb{E}(J((X_1, \Delta_1), (X_1, \Delta_1)))$$
$$= \mathbb{E}\left(\int_0^{\tau_n} \int_0^{\tau_n} \frac{K'(x, y)}{(1 - G(x-))(1 - G(y-))} dM_1(x) dM_1(y)\right)$$
$$= \mathbb{E}\left(\int_0^{\tau_n} \frac{K'(x, x)}{(1 - G(x-))^2} (dM_1(x))^2\right),$$

where the last equality follows from the fact that the integral in the off-diagonal, that is, for $x \neq y$, defines a zero-mean martingale (see Definition 6.6 and Theorem 6.8). Then,

$$\mathbb{E}\left(\int_0^{\tau_n} \frac{K'(x, x)}{(1 - G(x-))^2} (dM_1(x))^2\right) = \mathbb{E}\left(\int_0^\tau \frac{K'(x, x)}{(1 - G(x-))^2} d[M_1](x)\right)$$
$$= \mathbb{E}\left(\int_0^\tau \frac{K'(x, x)}{(1 - G(x-))^2} d\langle M_1\rangle(x)\right)$$
$$= \mathbb{E}\left(\int_0^\tau \frac{K'(x, x)}{(1 - G(x-))^2} \frac{S(x) Y_1(x)}{S(x-)^2} dF(x)\right)$$
$$= \int_0^\tau \frac{S(x)}{1 - H(x-)} K'(x, x) dF(x),$$

where the first and second equalities are due to the properties of $[M_1]$ (see beginning of Section 9.1). The third equality follows from the definition of $\langle M_1\rangle$, and the last equality follows by interchanging the integral and expectation.

We finalise with item iii. Let $\tilde{K}(x, y) = K'(x, y)/((1 - G(x-))(1 - G(y-)))$, by conditional expectation, it holds

$$\mathbb{E}(J((X_1, \Delta_1), (X_2, \Delta_2))^2)$$
$$= \mathbb{E}(\mathbb{E}(J((X_1, \Delta_1), (X_2, \Delta_2))^2 | (X_2, \Delta_2)))$$
$$= \mathbb{E}\left(\mathbb{E}\left(\left(\int_0^{X_1} \int_0^{X_2} \tilde{K}(x, y) dM_2(y) dM_1(x)\right)^2 \bigg| (X_2, \Delta_2)\right)\right),$$
$$= \mathbb{E}\left(\mathbb{E}\left(\int_0^{X_1} \left(\int_0^{X_2} \tilde{K}(x, y) dM_2(y)\right)^2 d\langle M_1\rangle(x) \bigg| (X_2, \Delta_2)\right)\right), \quad \text{(C.1)}$$

where the last equality follows from the fact that, conditioned on (X_2, Δ_2), $\int_0^t \int_0^{X_2} \tilde{K}(x, y) dM_2(y) dM_1(x)$ is an $(\mathcal{F}_t)$-martingale.

Replacing with the value of $\langle M_1(x)\rangle$ in Equation (C.1), we have

$$\mathbb{E}(J((X_1, \Delta_1), (X_2, \Delta_2))^2)$$

$$= \mathbb{E}\left(\int_0^{X_1} \left(\int_0^{X_2} \tilde{K}(x,y)dM_2(y) \right)^2 \frac{S(x)Y_1(x)}{S(x-)^2}dF(x) \right)$$

$$= \mathbb{E}\left(\int_0^{\tau} \left(\int_0^{X_2} \tilde{K}(x,y)dM_2(y) \right)^2 \frac{S(x)Y_1(x)}{S(x-)^2}dF(x) \right)$$

$$= \int_0^{\tau} \mathbb{E}\left(\left(\int_0^{X_2} \tilde{K}(x,y)dM_2(y) \right)^2 \right) \frac{S(x)(1-H(x-))}{S(x-)^2}dF(x)$$

$$= \int_0^{\tau} \mathbb{E}\left(\int_0^{\tau} \tilde{K}(x,y)^2 \frac{S(y)Y_2(y)}{S(y-)^2}dF(y) \right)(1-H(x-))\frac{S(x)}{S(x-)^2}dF(x)$$

$$= \int_0^{\tau} \int_0^{\tau} \frac{K'(x,y)^2 S(x)S(y)}{(1-H(x-))(1-H(y-))}dF(x)dF(y),$$

where the third equality follows from the fact that (X_1, Δ_1) and (X_2, Δ_2) are independent and by Fubini's Theorem. The fourth inequality follows by noticing that, for fixed x, $\int_0^t \bar{K}(x,y)dM_2(y)$ is a square-integrable $(\mathcal{F}_t)$-martingale with predictable variation process given by $\int_0^t \bar{K}(x,y)^2 \frac{S(y-)}{S(y-)^2}Y_2(y)dF(y)$.

Appendix D: Proof of Lemma 2.7

Equation (3.41) of Aalen et al. [1] states that

$$\Delta\widehat{F}_n(x) = \widehat{S}_n(x-)\frac{\Delta N(x)}{Y(x)},$$

hence a Kaplan-Meier weight W_i for an uncensored observation X_i is equal to $\Delta\widehat{F}_n(X_i)$ divided by all the uncensored observations that fall exactly in X_i, i.e., the weight W_i associated with X_i equals $\frac{\Delta\widehat{F}_n(X_i)}{\Delta N(X_i)} = \frac{\widehat{S}_n(X_i-)}{Y(X_i)}$. Then

$$\sum_{i=1}^n K(X_i, X_i)W_i^2 = \sum_{i=1}^n K(X_i, X_i)\left(\frac{\widehat{S}_n(X_i-)}{Y(X_i)} \right)^2 \Delta_i$$

$$= \int_0^{\tau_n} K(x,x)\left(\frac{\widehat{S}_n(x-)}{Y(x)} \right)^2 dN(x). \tag{D.1}$$

We will first prove that $\sqrt{n}\sum_{i=1}^n K(X_i, X_i)W_i^2 = o_p(1)$ under Condition 2.1. Note that the process $\int_0^t |K(x,x)|\left(\frac{\widehat{S}_n(x-)}{Y(x)} \right)^2 dN(x)$ is an $(\mathcal{F}_t)$-submartingale, with compensator given by $Z(t) = \int_0^t |K(x,x)|\left(\frac{\widehat{S}_n(x-)}{Y(x)} \right)^2 \frac{Y(x)}{S(x-)}dF(x)$. By the Lenglart-Rebolledo inequality, it is enough to prove that $\sqrt{n}Z(\tau_n) = o_p(1)$. An application of Propositions 6.2.i and 6.2.ii shows that

$$\sqrt{n}Z(\tau_n) = \int_0^{\tau_n} |K(x,x)|\left(\frac{\widehat{S}_n(x-)}{Y(x)} \right)^2 \frac{Y(x)}{S(x-)}dF(x)$$

$$= O_p(1) \int_0^\tau \frac{\mathbb{1}_{\{x \le \tau_n\}}}{\sqrt{Y(x)}} |K(x,x)| \sqrt{\frac{S(x-)}{(1-G(x-)}} dF(x).$$

Finally, $\int_0^\tau \frac{\mathbb{1}_{\{x \le \tau_n\}}}{\sqrt{Y(x)}} |K(x,x)| \sqrt{\frac{S(x-)}{1-G(x-)}} dF(x) = o_p(1)$ by the Dominated Convergence Theorem since $Y(x) \to \infty$ for each $x \in I_H$, $Y(x) \ge 1$ for $x \le \tau_n$, and $\int_0^\tau |K(x,x)| \sqrt{\frac{S(x-)}{(1-G(x-)}} dF(x) < \infty$ by hypothesis.

We now prove that $n \sum_{i=1}^n K(X_i, X_i) W_i^2 = \int_0^\tau \frac{K(x,x)}{1-G(x-)} dF(x) + o_p(1)$. We start by proving the intermediate step

$$\frac{1}{n} \int_0^{\tau_n} |K(x,x)| \left| \left(\frac{\widehat{S}_n(x-)}{(Y(x)/n)} \right)^2 - \frac{1}{(1-G(x-))^2} \right| dN(x) = o_p(1). \qquad \text{(D.2)}$$

Denote by $U_n(x) = \left| \left(\frac{\widehat{S}_n(x-)}{(Y(x)/n)} \right)^2 - \frac{1}{(1-G(x-))^2} \right|$ which is predictable with respect to $(\mathcal{F}_x)_{x \ge 0}$. Then, observe that $\frac{1}{n} \int_0^t |K(x,x)| U_n(x) dN(x)$ is an $(\mathcal{F}_t)$-submartingale with compensator, evaluated at τ_n, given by

$$\frac{1}{n} \int_0^{\tau_n} |K(x,x)| U_n(x) \frac{Y(x)}{S(x-)} dF(x)$$

$$= O_p(1) \int_0^\tau \mathbb{1}_{\{x \le \tau_n\}} |K(x,x)| U_n(x)(1-G(x-)) dF(x), \qquad \text{(D.3)}$$

where the equality holds by Proposition 6.2.iii. We claim that Equation (D.3) equals $o_p(1)$ by Lemma 6.4, whose conditions we proceed to verify. Set $R(x) = \frac{|K(x,x)|}{1-G(x-)}$, which is integrable, and set $R_n(x)$ as the integrand in Equation (D.3). For the first condition of Lemma 6.4, note that $U_n(x) \to 0$ for all x, due to Proposition 6.1, hence $R_n(x) \to 0$. For the second condition, Propositions 6.2.i and 6.2.ii yield $U_n(x) = O_p(1)(1-G(x-))^{-2}$ uniformly in $x \le \tau_n$, and thus $R_n(x) = O_p(1)R(x)$.

Finally, the Lenglart-Rebolledo Inequality, we get that Equation (D.2) holds true. Combining (D.1) and (D.2) yields

$$n \sum_{i=1}^n K(X_i, X_i) W_i^2 = \frac{1}{n} \sum_{i=1}^n \frac{K(X_i, X_i)}{(1-G(X_i-))^2} \Delta_i + o_p(1),$$

hence, by the Law of Large numbers we obtain that $n \sum_{i=1}^n K(X_i, X_i) W_i^2 = \int_0^\tau \frac{K(x,x)}{1-G(x-)} dF(x) + o_p(1)$.