The Annals of Applied Probability
2002, Vol. 12, No. 3, 944–971

RANDOM CENSORING IN SET-INDEXED SURVIVAL ANALYSIS

BY B. GAIL IVANOFF[1] AND ELY MERZBACH[2]

University of Ottawa and Bar-Ilan University

Using the theory of set-indexed martingales, we develop a general model for survival analysis with censored data which is parameterized by sets instead of time points. We define a set-indexed Nelson–Aalen estimator for the integrated hazard function with the presence of a censoring by a random set which is a stopping set. We prove that this estimator is asymptotically unbiased and consistent. A central limit theorem is given. This model can be applied to cases when censoring occurs in geometrical objects or patterns, and is a generalization of models with multidimensional failure times.

1. Introduction. The aim of this work is to study random censoring in the framework of set-indexed survival analysis. Examples in which such situations occur are many: one such example is the study of survival with liver cirrhosis [2] in which the basic time is the duration of the trial, but survival depends also on the age of the patient and on the shape and size of the liver. Other relevant medical examples concern the development of cancer tumors. In models involving time to failure, the "time" parameter(s) of interest may include not only chronological time but also mileage for warranties on cars, cumulative exposure, etc. Another type of example concerns estimation of windowed spatial processes [3, 4, 12], which consider statistical problems arising from the observation of a spatial process of geometrical objects or patterns. Typically, the process is observed through a bounded observation window which can be considered as a sort of random censoring. This kind of model has diverse areas of application, including environmental science (aerial photography of forests) (cf. [22]), medical science (cf. [10]) and metallurgy and mining.

To our knowledge, there does not exist a literature on the theory of set-indexed survival analysis. Therefore, using the tools developed in [14], we will present a first step toward a general model for set-indexed survival analysis in the framework of martingale theory.

Some results do exist for the two-parameter case; we refer the reader to two excellent books [2] and [13] and to the references cited therein for a survey of the known results. Other recent results in this direction may be found in [9], and

Received November 2000; revised December 2001.

[1]Supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

[2]Work done at the University of Ottawa.

AMS 2000 subject classifications. 62G05, 60G42, 60G55.

Key words and phrases. Censoring, stopping set, set-indexed martingale, hazard function, survival analysis, Volterra equation, estimator, central limit theorem.

in [15] for function-indexed nonparametric tests; for a survey see also [20]. As explained in [2], for many years much of the literature concentrated on independent two-sample tests. The question of dependence is examined more closely in [13], but not from the point of view of multidimensional martingales. A two-parameter martingale approach is analysed by Pons in [17], showing that some properties of the multivariate Nelson–Aalen estimator can be obtained exactly as in the univariate case. However, Pons' structure presupposes the filtration to be a product filtration and requires independence of the coordinates. Her approach will be shown to be a special case of the set-indexed framework described here.

In this paper, we deal with the following model for set-indexed survival processes: the processes are indexed by a family $\mathcal{A}$ of compact subsets of a locally compact Hausdorff topological space T equipped with a positive measure λ on the Borel subsets of T. The two-parameter case, studied in [1, 13, 17–19] is obtained when $T = \mathbf{R}_+^2$, $\mathcal{A}$ is the class of rectangles $\{[0, t] : t \in \mathbf{R}_+^2\}$ and λ is Lebesgue measure. In all of the preceding references, it is implicit in the model that all components of the multivariate survival time are nonnegative. Since we impose no such restriction, it will be seen that our more general model can also be applied to spatial data.

In Section 2, we define the main tools of survival analysis in the set-indexed framework: the point process, the survivor function and the hazard function. We compute a kind of compensator for the point process and present a Volterra equation.

In Section 3, we introduce censoring in terms of a *stopping set*. The model presented here may be seen as a generalization of [17] and [18]. In certain cases, the censoring is in fact a "filtering" model [2]. Under an independence assumption, we shall see that this gives us a multiplicative model.

The theme of Section 4 is estimation. Using martingale methods, we propose a Nelson–Aalen type estimator for the integrated hazard function and prove that it is asymptotically unbiased and consistent.

In Section 5, a central limit theorem for the estimator is proved. Statistical techniques involving the estimator and potential areas of application are discussed in Section 6.

2. Preliminaries and model specification.

The framework will be essentially the same as in [14] and we refer to this book for the details. Let T be a locally compact Hausdorff space and λ a measure on $\mathcal{B}$, the Borel sets of T, which is finite on compact sets. All processes will be indexed by a class $\mathcal{A}$ of compact connected subsets of T, and we assume that the measure λ does not charge the borders of sets from $\mathcal{A}$.

In what follows, for any class of sets $\mathcal{D}$, the class of finite unions of sets from $\mathcal{D}$ will be denoted by $\mathcal{D}(u)$. In the terminology of [14], we assume that $\mathcal{A}$ is an *indexing collection*:

DEFINITION 2.1. A nonempty class $\mathcal{A}$ of compact, connected subsets of T is called an *indexing collection* if it satisfies the following:

(i) $\varnothing \in \mathcal{A}$, and $\forall A \in \mathcal{A}$, $A^\circ \neq A$ if $A \neq \varnothing$ or T. In addition, there is an increasing sequence (B_n) of sets in $\mathcal{A}(u)$ such that $T = \bigcup_{n=1}^\infty B_n^\circ$. [Hence, for every $A \in \mathcal{A}$, there exists $n = n(A)$ such that $A \subseteq B_n$.]

(ii) $\mathcal{A}$ is closed under arbitrary intersections and if $A, B \in \mathcal{A}$ are nonempty, then $A \cap B$ is nonempty. If (A_i) is an increasing sequence in $\mathcal{A}$ and there exists n such that $A_i \subseteq B_n$ for every i, then $\overline{\bigcup_i A_i} \in \mathcal{A}$. [Such a sequence (A_i) is called bounded.]

(iii) The σ-algebra generated by $\mathcal{A}$, $\sigma(\mathcal{A}) = \mathcal{B}$, the collection of all Borel sets of T.

(iv) *Separability from above*: There exists an increasing sequence of finite subclasses $\mathcal{A}_n = \{A_1^n, \dots, A_{k_n}^n\}$ of $\mathcal{A}$ closed under intersections and satisfying $\varnothing$, $B_n \in \mathcal{A}_n(u)$ [B_n is defined in (i) above] and a sequence of functions $g_n : \mathcal{A} \to \mathcal{A}_n(u) \cup \{T\}$ such that:

- (a) g_n preserves arbitrary intersections and finite unions [i.e., $g_n(\bigcap_{A\in\mathcal{A}'} A) = \bigcap_{A\in\mathcal{A}'} g_n(A)$ for any $\mathcal{A}' \subseteq \mathcal{A}$, and if $\bigcup_{i=1}^k A_i = \bigcup_{j=1}^m A'_j$, then $\bigcup_{i=1}^k g_n(A_i) = \bigcup_{j=1}^m g_n(A'_j)$],
- (b) for each $A \in \mathcal{A}$, $A \subseteq (g_n(A))^\circ$,
- (c) $g_n(A) \subseteq g_m(A)$ if $n \geq m$,
- (d) for each $A \in \mathcal{A}$, $A = \bigcap_n g_n(A)$,
- (e) if $A, A' \in \mathcal{A}$ then for every n, $g_n(A) \cap A' \in \mathcal{A}$, and if $A' \in \mathcal{A}_n$ then $g_n(A) \cap A' \in \mathcal{A}_n$.
- (f) $g_n(\varnothing) = \varnothing \; \forall n$.

(v) Every countable intersection of sets in $\mathcal{A}(u)$ may be expressed as the closure of a countable union of sets in $\mathcal{A}$.

NOTE. The symbol $\subset$ indicates strict inclusion, and $\overline{(\cdot)}$ and $(\cdot)^\circ$ denote, respectively, the closure and the interior of a set.

For $t \in T$, define the following sets:

The "past" of t: $A_t = \cap A$, $A \in \mathcal{A}$, $t \in A$.
The "future" of t: $E_t = \cap B^c$, $B \in \mathcal{A}$, $t \notin B$.

We assume that E_t is closed. By (iii) above it follows that $s \neq t$ if and only if $A_s \neq A_t$, which in turn implies that $A_t \cap E_t = \{t\}$, since it is easy to see that $s \in E_t$ if and only if $t \in A_s$, $\forall s, t \in T$. Therefore, there is a natural partial order induced on T by the indexing collection $\mathcal{A}$: $s \leq t$ if and only if $A_s \subseteq A_t$. Thus, $A_t = \{s : s \leq t\}$ and $E_t = \{s : s \geq t\}$.

We shall define the semialgebra $\mathcal{C}$ to be the class of all subsets of T of the form

$$C = A \setminus B,\; A \in \mathcal{A},\; B \in \mathcal{A}(u).$$

$\mathcal{C}$ is closed under intersections and any set in $\mathcal{C}(u)$ may be expressed as a finite disjoint union of sets in $\mathcal{C}$. Note that if $B = \bigcup_{i=1}^{k} A_i \in \mathcal{A}(u)$, without loss of generality we can require that for each i, $A_i \not\subseteq \bigcup_{j \neq i} A_j$. Such a representation of $B \in \mathcal{A}(u)$ will be called *extremal*. If $C = A \setminus B,\ A \in \mathcal{A},\ B \in \mathcal{A}(u)$, then the representation of C is called extremal if that of B is. Unless otherwise stated, it will always be assumed that all representations of sets in $\mathcal{A}(u)$ and C are extremal. For each finite subsemilattice $\mathcal{A}_k$, let $\mathcal{C}^{\ell}(\mathcal{A}_k)$ denote the *left-neighborhoods* of $\mathcal{A}_k$; that is, sets of the form $C = A \setminus \bigcup_{A' \in \mathcal{A}_k, A \not\subseteq A'} A'$, $A \in \mathcal{A}_k$. Clearly, $\mathcal{C}^{\ell}(\mathcal{A}_k)$ partitions B_k (B_k as defined in Definition 2.1) and by separability from above, the points of T are separated by the sets $A, A \in \bigcup_k \mathcal{A}_k$, and so the sequence $(\mathcal{C}^{\ell}(\mathcal{A}_k))_k$ is a dissecting system for T.

The following assumption about the structure of $\mathcal{C}$ will be required.

ASSUMPTION 2.2. *For any $B \in \mathcal{A}(u)$, if $C = A \setminus \bigcup_{i=1}^{k} A_i \in \mathcal{C}$ and if $A \subseteq B$, then there exist sets $D_1, \dots, D_m \in \mathcal{A},\ D_i \subseteq B,\ i = 1, \dots, m$, such that $C = A \setminus \bigcup_{i=1}^{m} D_i$ is an extremal representation, and if $A' \in \mathcal{A},\ A' \subseteq B,\ A' \cap C = \varnothing$, then $A' \subseteq \bigcup_{i=1}^{m} D_i$. This is called a* maximal *representation of C in B.*

Numerous examples of topological spaces T and indexing collections $\mathcal{A}$ satisfying the preceding assumptions may be found in [14].

EXAMPLE 2.3. Our framework generalizes the usual multiparameter setting: if $T = \mathbf{R}_+^d$, then the class $\mathcal{A} = \{[0, t] : t \in \mathbf{R}_+^d\}$ satisfies all the assumptions and in this instance, the class $\mathcal{C}(u)$ consists of all finite unions of disjoint rectangles of the form $(s, t],\ s, t \in \mathbf{R}_+^d$. More generally, we can allow $\mathcal{A}$ to consist of all the *lower layers* of $\mathbf{R}_+^d$: a set A is a lower layer if $[0, t] \subseteq A,\ \forall t \in A$. In both cases, it is easily seen that $A_t = [0, t]$ and $E_t = [t, \infty)$.

EXAMPLE 2.4. The following simple but very important generalization of the preceding example is appropriate for modelling spatial data. We may let T be any subset of $\mathbf{R}^d$ of the form $[-a, b],\ a, b \in \mathbf{R}_+^d$ and let $\mathcal{A}$ be the class of lower layers in T, as defined above. Note that the partial order induced by the sets $A_t = [0, t]$ is no longer the usual partial order on $\mathbf{R}^d$: we have that $(s_1, \dots, s_d) = s \leq t = (t_1, \dots, t_d)$ if and only if s and t lie in the same quadrant and $|s_i| \leq |t_i|$, $i = 1, \dots, d$.

EXAMPLE 2.5. A third example (the "history of the world") models space–time data. Here, $T = \overline{B(0, t_0)}$ (compact ball of radius t_0 in $\mathbf{R}^3$), $\mathcal{A} = \{A_{R(a,b,c,d),t};\ 0 \leq a < b < 2\pi, -\pi \leq c < d \leq \pi,\ t \in [0, t_0]\}$, where the set

$$A_{R(a,b,c,d),t} := \{(r\cos\theta\cos\tau, r\sin\theta\cos\tau, r\sin\tau);\ \theta \in [a, b],\ \tau \in [c, d],\ r \in [0, t]\}$$

can be interpreted as the history of the region

$$R(a,b,c,d) = \{(\cos\theta\cos\tau, \sin\theta\cos\tau, \sin\tau);\ \theta \in [a,b],\ \tau \in [c,d]\}$$

of the earth from the beginning until time t. [Here θ represents the longitude of the generic point in the region $R(a,b,c,d)$, while τ is the latitude.] Hence, $\mathcal{A}$ can be identified with the history of the world up to time t_0.

Now, let $(\Omega, \mathcal{F}, P)$ be any complete probability space. A filtration (indexed by $\mathcal{A}$) is a class of complete sub-σ-fields of $\mathcal{F}$ $\{\mathcal{F}_A : A \in \mathcal{A}\}$ which satisfies the following conditions:

1. $\forall A, B \in \mathcal{A}$, $\mathcal{F}_A \subseteq \mathcal{F}_B$, if $A \subseteq B$.
2. Monotone outer-continuity: $\mathcal{F}_{\cap A_i} = \cap \mathcal{F}_{A_i}$ for any decreasing sequence (A_i) in $\mathcal{A}$.
3. For consistency in what follows, if $T \notin \mathcal{A}$, define $\mathcal{F}_T = \mathcal{F}$.

We may associate various σ-algebras with sets in $\mathcal{A}(u)$ and $\mathcal{C}(u)$. If $B \in \mathcal{A}(u)$, then $\mathcal{F}_B^0 = \vee_{A \in \mathcal{A}, A \subseteq B} \mathcal{F}_A$. The σ-algebras $\{\mathcal{F}_B^0 : B \in \mathcal{A}(u)\}$ are complete and increasing, but not necessarily monotone outer-continuous. Thus, we define for $B \in \mathcal{A}(u)$: $\mathcal{F}_B = \bigcap_n \mathcal{F}^0_{g_n(B)}$. Next, for $C \in \mathcal{C}(u) \setminus \mathcal{A}$, let $\mathcal{G}_C^* = \vee_{B \in \mathcal{A}(u), B \cap C = \varnothing} \mathcal{F}_B$, and for $A \in \mathcal{A}$, define $\mathcal{G}_A^* = \mathcal{F}_\varnothing$. We note that $\{\mathcal{G}_C^*\}$ is a decreasing family of σ-fields: if $C \subseteq C'$, then $\mathcal{G}_{C'}^* \subseteq \mathcal{G}_C^*$.

DEFINITION 2.6. A ($\mathcal{A}$-indexed) stochastic process $X = \{X_A : A \in \mathcal{A}\}$ is a collection of random variables indexed by $\mathcal{A}$, and is said to be adapted if X_A is $\mathcal{F}_A$-measurable, for every $A \in \mathcal{A}$. X is said to be integrable if $E[| X_A |] < \infty$. A process $X : \mathcal{A} \to \mathbf{R}$ is increasing if (almost surely) X can be extended to a finitely additive process on $\mathcal{C}$, $X(\varnothing) = 0$ and $X_C \geq 0$, $\forall C \in \mathcal{C}$.

Next, we specify the model. Still assuming that $(\Omega, \mathcal{F}, P)$ is a complete probability space, let $Y : \Omega \to T$ be a T-valued random variable, and denote by $\mu = \mu_Y$ its distribution function: $\mu(B) = P\{Y \in B\}$. The *survival function* associated with Y is $S(t) = \mu(E_t)$. [Note, however, that $\mu(A_t) \neq 1 - S(t)$.] We assume that μ is absolutely continuous with respect to λ and denote by μ' the Radon–Nikodym derivative of μ with respect to λ on the Borel sets of T.

Note that the usual multivariate survival model (cf. Example 2.3) requires that all components of Y be nonnegative. In contrast, our model imposes no such restriction (cf. Examples 2.4 and 2.5).

We are now in position to define the *hazard function* h and the *integrated hazard function* H of Y:

DEFINITION 2.7. For $t \in T$, the hazard function of Y is h where

$$h(t) = \frac{\mu'(t)}{S(t)}.$$

If $S(t) = 0$, $h(t)$ is defined to be zero.

The integrated hazard function of Y is H where

$$H(t) = \int_{A_t} h(u)\lambda(du)$$

and more generally,

$$H_A = \int_A h(u)\lambda(du) \qquad \text{for any } A \in \mathcal{A}.$$

Heuristically,

$$h(t) = \lim_{n\to\infty} \left(\frac{P(Y \in g_n(A_t)|Y \in E_t)}{\lambda(g_n(A_t) \cap E_t)} \right),$$

when this limit exists.

PROPOSITION 2.8 (The Volterra equation). *For all* $t \in T$, $S(t) = 1 - \int_{E_t{}^c} S(s) \times h(s)\lambda(ds)$.

PROOF.

$$\begin{aligned} S(t) &= \mu(E_t) \\ &= 1 - \int_{E_t{}^c} \mu(E_s)\mu(ds)/\mu(E_s) \\ &= 1 - \int_{E_t{}^c} \mu(E_s)h(s)\lambda(ds) \\ &= 1 - \int_{E_t{}^c} S(s)h(s)\lambda(ds). \end{aligned}$$ □

COMMENT. Clearly, μ is determined by S; however, as is well known (cf. [2]), in general h does not completely determine S (or μ). In $\mathbf{R}_+^2$ for example, knowledge of the marginal survival functions is also required. However, it should be noted that the dependence structure of μ is reflected in the structure of h.

Let $N = \{N_A,\ A \in \mathcal{A}\} = \{I_{\{Y\in A\}},\ A \in \mathcal{A}\}$ be the single jump process associated with Y and $\{\mathcal{F}_A^Y, A \in \mathcal{A}(u)\}$ its minimal filtration: $\mathcal{F}_A^Y = \sigma\{N_B : B \in \mathcal{A}, B \subseteq A\} \cup \{\mathcal{P}_0\}$ ($\mathcal{P}_0$ is the class of P-null sets). (It is straightforward to show that $\mathcal{F}^Y$ is monotone outer-continuous.) The process N can clearly be extended to an additive process on the more general index sets $\mathcal{A}(u)$ and $\mathcal{C}$ (and in fact to $\mathcal{B}$).

In what follows, any random process indexed by $\mathcal{A}$ will be assumed to be additive. As in [14], a process $M = \{M_A,\ A \in \mathcal{A}\}$ is called a *strong martingale* if it is adapted and for any $C \in \mathcal{C}$, $E[M_C \mid \mathcal{G}_C^*] = 0$. If the process M is not adapted, it will be called a *pseudo-strong martingale*. A process $\overline{X}$ is called a **-compensator* of the process X if it is increasing ($\overline{X_C} \geq 0$ for all $C \in \mathcal{C}$) and the difference $X - \overline{X}$ is a pseudo-strong martingale. The *-compensator is not necessarily unique unless some sort of predictability condition is imposed (cf. [14]).

In the next result, we compute a compensator for the single jump process. In the two-parameter case, it was done first in [1]; see also [9] and for $T \subseteq \mathbf{R}^d$ see [14].

PROPOSITION 2.9. *The process $\overline{N}$ defined by*

$$\overline{N}_A = \int_{A \cap A_Y} \mu(E_u)^{-1} \mu(du) = \int_{A \cap A_Y} h(u)\lambda(du) = \int_A I_{\{Y \in E_u\}} h(u)\lambda(du)$$

is a *-compensator of the process N with respect to its minimal filtration, *where* $A_Y(\omega) = A_{Y(\omega)}$.

PROOF. Let $C = A \setminus D_n$ be a maximal representation of $C \in \mathcal{C}$ in B_n (B_n is as defined in Definition 2.1). $(\mathcal{F}^Y_{D_n})$ is an increasing sequence of σ-fields and it is clear by definition that $\mathcal{G}^*_C$ is generated by $\bigcup_n \mathcal{F}^Y_{D_n}$. Therefore, it suffices to show that for every $C \in \mathcal{C}$ and for every n,

$$E\big[N_C - \overline{N}_C \mid \mathcal{F}^Y_{D_n}\big] = 0.$$

First we observe that

$$E\big[N_C \mid \mathcal{F}^Y_{D_n}\big] = I_{\{N_{D_n}=0\}} \mu(C)\big(1 - \mu(D_n)\big)^{-1}.$$

Next, note that if $u \in C$, then $E_u \subseteq D_n^c$, since if $D_n = \bigcup_{i=1}^{k_n} A_{n,i}$ $(A_{n,i} \in \mathcal{A})$, then $u \notin A_{n,i}$, $i = 1, \dots, k_n$, and $E_u \subseteq \bigcap_i A^c_{n,i} = D_n^c$. Now, since $C \cap A_Y = \varnothing$ if $Y \in D_n$, then

$$\begin{aligned}
E\big[\overline{N}_C \mid \mathcal{F}^Y_{D_n}\big] &= E\bigg[\int_{C \cap A_Y} (\mu_Y(E_u))^{-1} \mu(du) \big| \mathcal{F}^Y_{D_n}\bigg] \\
&= I_{\{N_{D_n}=0\}} \int_{D_n^c} \bigg[\int_{C \cap A_v} (\mu(E_u))^{-1} \mu(du)\bigg] \big(1 - \mu(D_n)\big)^{-1} \mu(dv) \\
&= I_{\{N_{D_n}=0\}} \big(1 - \mu(D_n)\big)^{-1} \int_C \int_{E_u \cap D_n^c} \mu(dv) \big(\mu(E_u)\big)^{-1} \mu(du) \\
&= I_{\{N_{D_n}=0\}} \mu(C) \big(1 - \mu(D_n)\big)^{-1}. \qquad \square
\end{aligned}$$

COMMENT. Proposition 2.9 shows that we have a sort of *multiplicative* model, in analogy to the classic model on $\mathbf{R}_+$, since $\overline{N}_A = \int_A I_{\{Y \in E_u\}} h(u)\lambda(du)$. Notice that in general the *-compensator is not adapted since the event $\{Y \in E_u\}$ does not necessarily belong to $\mathcal{F}^Y_{A_u}$. However, $\overline{N}$ does satisfy a certain predictability property. We denote by $\mathcal{P}^*$ the (*-predictable) σ-field generated on the product space $(\Omega \times T, \mathcal{F} \otimes \mathcal{B})$ by the class of rectangles $\{F_C \times C : C \in \mathcal{C}, F_C \in \mathcal{G}^*_C\}$. It is shown at the end of the proof of Proposition 4.1 in the Appendix that $\{(\omega, u) : Y(\omega) \in E_u\} \in \mathcal{P}^*$.

3. Set-indexed censoring. In this section, we add censoring to the survival process. Since in most cases, the censoring is not deterministic, we define the censoring mechanism in terms of a *stopping set* ξ.

DEFINITION 3.1. $\xi : \Omega \to \mathcal{A}(u)$ is a stopping set with respect to a filtration $\mathcal{F}$ if for any $A \in \mathcal{A}$, $\{\omega : A \subseteq \xi(\omega)\} \in \mathcal{F}_A$ and $\{\omega : \varnothing = \xi(\omega)\} \in \mathcal{F}_\varnothing$.

Generally the observation of the random variable Y is "right-censored" or in our setting is "outer-censored," that is, Y is observed not on A but only on a subset of the form $A \cap \xi$. Such a model is natural from a geographical point of view; for example, if one is counting diseased trees from the air, observations can only be taken in regions not obscured by cloud cover. For the case in which $T = [-1, 1]^2$ and $\mathcal{A}$ is the class of lower layers, this is illustrated in Figure 1. Those values of Y which can be observed are indicated with $\bullet$ and those which are obscured outside of ξ are indicated with $\circ$. The outer-censored case is the most important example of incomplete observation; however, there may be other observational plans of interest where some information is available after the censoring. Such a concept is called "filtering" in [2]. In such a case, we may assume that we have an increasing sequence of stopping sets $\varnothing = \eta_0 \subseteq \xi_1 \subseteq \eta_1 \subseteq \xi_2 \subseteq \eta_2 \cdots \subseteq \eta_k \subseteq \xi_k$ and that we cannot observe Y if $Y \in \bigcup_i \eta_i \setminus \xi_i$. For the sake of clarity this more general structure will not be pursued here, but clearly can be analyzed in a similar manner.

Formally, therefore, given a T-valued random variable Y whose associated single jump process N is adapted to a filtration $\mathcal{F}$, and an $\mathcal{F}$-stopping set ξ, the *censored* jump process N^ξ is defined by the "stopped" process

$$N_A^\xi = N_{A\cap\xi} = I_{\{Y \in A\cap\xi\}}, \qquad A \in \mathcal{A}(u).$$

The fact that ξ is a stopping set ensures that the censored process N^ξ is measurable and adapted (cf. [14], Lemma 1.5.9).

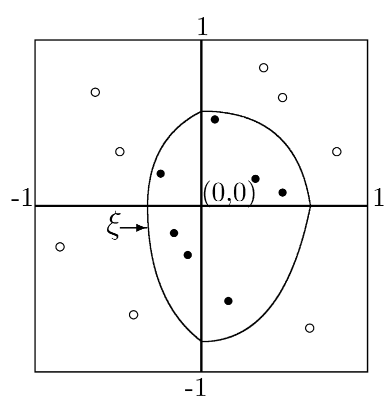

FIG. 1. *Censoring by a stopping set ξ in* $[-1, 1]^2$.

It will be necessary to assume a type of independence between the censoring mechanism and the random variable being observed.

DEFINITION 3.2. Let Y be a T-valued random variable and let $\mathcal{F}^Y$ be the minimal filtration generated by its associated jump process N. Let $\mathcal{F}$ be a filtration such that $\mathcal{F}_A^Y \subseteq \mathcal{F}_A$ $\forall A \in \mathcal{A}(u)$ and let ξ be an $\mathcal{F}$-stopping time. ξ is:

(i) weakly independent of Y if the *-compensator of N with respect to $\mathcal{F}$ is the same as the *-compensator with respect to $\mathcal{F}^Y$;

(ii) independent of Y if $\mathcal{F}^Y$ is independent of $\sigma(\{A \subseteq \xi\} : A \in \mathcal{A})$ and $\mathcal{F}_A = \mathcal{F}_A^Y \vee \sigma(\{A' \subseteq \xi\} : A' \in \mathcal{A},\ A' \subseteq A), \forall A \in \mathcal{A}$.

It is easily seen that independence implies weak independence. However, the reverse is not true: if $\mathcal{F} = \mathcal{F}^Y$, then trivially ξ and Y are weakly independent but not independent in general. *We shall always assume that ξ and Y are weakly independent.*

When it exists, we may define the hazard function of the censored process to be

$$h^\xi(t) = \lim_{n\to\infty}\left(\frac{P(Y \in g_n(A_t) \mid Y \in E_t,\ \xi \not\subseteq E_t^c)}{\lambda(g_n(A_t) \cap E_t)}\right).$$

When ξ is weakly independent of Y, $h^\xi = h$, and the following lemma shows that the model with censoring remains multiplicative.

PROPOSITION 3.3. *If ξ is weakly independent of Y, then the stopped process $\overline{N}^\xi$ is a *-compensator of N^ξ, where*

$$\begin{aligned}\overline{N}_A^\xi &= \overline{N}_{A\cap\xi}\\ &= \int_{A\cap\xi\cap A_Y} \mu(E_u)^{-1}\mu(du)\\ &= \int_{A\cap\xi\cap A_Y} h(u)\lambda(du)\\ &= \int_A I_{\{Y\in E_u\}} I_{\{u\in\xi\}} h(u)\lambda(du).\end{aligned}$$

PROOF. Using the same notation as in the proof of Proposition 2.9, it suffices to show that for every $C \in \mathcal{C}$ and for every n,

$$E[N_C^\xi - \overline{N}_C^\xi \mid \mathcal{F}_{D_n}] = 0,$$

where $C = A \setminus D_n$ is the maximal representation of C in B_n. By additivity, this is equivalent to showing that

$$E[N_{A\cap\xi} - \overline{N}_{A\cap\xi} \mid \mathcal{F}_{D_n}] = E[N_{A\cap D_n\cap\xi} - \overline{N}_{A\cap D_n\cap\xi} \mid \mathcal{F}_{D_n}].$$

This follows by the weak independence of ξ and from the proof of Theorem 3.3.1 of [14], observing that this proof remains valid for pseudo-strong martingales. $\square$

COMMENT. In two-dimensional survival analysis, censoring is defined in a more restricted manner: a failure time $Y \in \mathbf{R}_+^2$ may be partially observed, given a random censoring time $\tau = (\tau_1, \tau_2)$; that is, we observe $Y \wedge \tau = (Y_1 \wedge \tau_1, Y_2 \wedge \tau_2)$ and $I_{\{Y_i \le \tau_i\}}$, $i = 1, 2$. Expressing this in terms of sets, letting $A_t = [0, t]$, $A_Y = [0, Y]$ and $\xi = [0, \tau]$, the counting process of censored times is (cf. [17])

$$N^{\xi}(t) = N^{\xi}_{A_t} = I_{\{Y \wedge \tau \le t, Y_1 \le \tau_1, Y_2 \le \tau_2\}} = I_{\{Y \in A_t \cap \xi\}}.$$

Thus, our framework extends the usual bivariate censoring model.

4. Estimation. The goal is to estimate the integrated (set-indexed) hazard function H [recall that $H_A = \int_A h(t)\lambda(dt)$] using censored data. The pseudo-strong martingale structure of the multiplicative model allows us to develop an estimator in analogy to the classical Nelson–Aalen estimator on $\mathbf{R}_+$. To be precise, suppose that a sequence of i.i.d. T-valued random variables (Y_i) with the same distribution as Y is given, as well as a sequence (ξ_i) of stopping sets. We shall assume that for every i, j, ξ_i is an $\mathcal{F}$-stopping time weakly independent of Y_j. We do not assume at this point that the ξ_i's are independent (for example, it is possible that $\xi_i = \xi$, $\forall i$). Define the following processes:

$$Z_n(t) = \sum_{i=1}^{n} I_{\{Y_i \in E_t\}} I_{\{t \in \xi_i\}},$$

$$\overline{N}^{(n)}_A = \int_A Z_n(t)h(t)\lambda(dt).$$

Note that Z_n is indexed by T, while $\overline{N}^{(n)}$ is indexed by $\mathcal{A}$ and is additive. We see that Z_n can be viewed as the *survivor function* process. Furthermore, by independence and Proposition 3.3, $\overline{N}^{(n)}$ is a *-compensator for $N^{(n)\xi}$, where $N^{(n)\xi}_A = \sum_{i=1}^{n} I_{\{Y_i \in A \cap \xi_i\}}$, and the process

$$M^{(n)}_{\cdot} = N^{(n)\xi}_{\cdot} - \int_{\cdot} Z_n(t)h(t)\lambda(dt)$$

is a pseudo-strong martingale with respect to $\mathcal{F}$. Therefore, exactly as in the classical case, we have

$$N^{(n)\xi}(dt) = Z_n(t)h(t)\lambda(dt) + M^{(n)}(dt).$$

Regarding $M^{(n)}$ as noise, we are led to a set-indexed version of the Nelson–Aalen estimator for H_A:

$$\hat{H}^{(n)}_A = \int_A \frac{N^{(n)\xi}(dt)}{Z_n(t)} = \sum_{\{i : Y_i \in A \cap \xi_i\}} \big(Z_n(Y_i)\big)^{-1}.$$

Define

$$J_n(t) = I_{\{Z_n(t) > 0\}}$$

and

$$\tilde{H}_A^{(n)} = \int_A J_n(t)h(t)\lambda(dt).$$

Trivially, $\tilde{H}_A^{(n)} = H_A$ if $P(J_n(t) = 1\ \forall t \in A) = 1$. We observe that since E_t is closed, $N^{(n)^\xi}(dt) > 0$ only if $J_n(t) = 1$. Thus,

$$\hat{H}^{(n)} - \tilde{H}^{(n)} = \int_\cdot \frac{J_n(t)}{Z_n(t)} M^{(n)}(dt). \tag{1}$$

Clearly, the asymptotic behavior of the estimator $\hat{H}^{(n)}$ depends on the properties of the pseudo-strong martingale $M^{(n)}$. First, we note that the fact that the censoring mechanism is a stopping set yields the following important proposition, whose proof is given in the Appendix.

PROPOSITION 4.1. *$\hat{H}^{(n)} - \tilde{H}^{(n)}$ is a pseudo-strong martingale.*

Next, we will need to determine the covariance structure of $\int_\cdot g(t)M^{(n)}(dt)$, where g is a continuous increasing function on T. [We call a function $g : T \to \mathbf{R}$ increasing (resp., decreasing) if it is increasing (decreasing) with respect to the partial order "less than or equal to", induced by $\mathcal{A}$ on T.] This will follow from the covariance structure of $M^{(n)}$. This is nontrivial, as the increments of a pseudo-strong martingale are not necessarily uncorrelated.

COMMENT. For the remainder of the paper, we shall assume that $\xi_1, \xi_2, \dots$ are i.i.d. and independent of the Y_i's.

PROPOSITION 4.2. *For $C, D \in \mathcal{C}$,*

$$\begin{aligned} &\operatorname{Cov}(M_C^{(n)}, M_D^{(n)}) \\ &\quad = n\int_{C\cap D} S(t)P(t \in \xi)h(t)\lambda(dt) \\ &\qquad + n\iint_{\mathcal{I}(C,D)} \mu(E_s \cap E_t)P(s,t \in \xi)h(s)h(t)\lambda(ds)\lambda(dt), \end{aligned} \tag{2}$$

where $\mathcal{I}(C, D) = \{(c, d) \in C \times D : c \in A_d^c \cap E_d^c\} = \{(c, d) \in C \times D : d \in A_c^c \cap E_c^c\}$.

PROOF. We use a decomposition similar to that in [17]. For $A \in \mathcal{A}$ and $t \in T$, define

$$C_A^{(n)} := \frac{N^{(n)^\xi}_A - E[N_A^{(n)^\xi}]}{\sqrt{n}}, \tag{3}$$

$$D_n(t) := \frac{Z_n(t) - E[Z_n(t)]}{\sqrt{n}}. \tag{4}$$

We have

$$n^{-1/2}M^{(n)}(A) = C_A^{(n)} - \int_A D_n(t)h(t)\lambda(dt). \tag{5}$$

We observe that

$$E\Big[N_A^{(n)\xi}\Big] = n\int_A S(t)P(t \in \xi)h(t)\lambda(dt),$$
$$E[Z_n(t)] = nS(t)P(t \in \xi).$$

For $C, D \in \mathcal{C}$,

$$\begin{aligned}
&\operatorname{Cov}\big(C_C^{(n)}, C_D^{(n)}\big) \\
&\quad = n^{-1}E\Bigg[\Bigg(\sum_1^n\bigg(I_{\{Y_i \in C\cap\xi_i\}} - \int_C S(s)P(s \in \xi)h(s)\lambda(ds)\bigg)\Bigg) \\
&\qquad\qquad \times\Bigg(\sum_1^n\bigg(I_{\{Y_j \in D\cap\xi_j\}} - \int_D S(t)P(t \in \xi)h(t)\lambda(dt)\bigg)\Bigg)\Bigg] \\
&\quad = P(Y \in C\cap D\cap\xi) \\
&\qquad + \int_D\int_C S(s)S(t)P(s \in \xi)P(t \in \xi)h(s)h(t)\lambda(ds)\lambda(dt) \\
&\qquad - P(Y \in C\cap\xi)\int_D S(t)P(t \in \xi)h(t)\lambda(dt) \\
&\qquad - P(Y \in D\cap\xi)\int_C S(s)P(s \in \xi)h(s)\lambda(ds).
\end{aligned}$$

The last equality follows by independence of the Y_i's and ξ_i's. Noting that

$$\mu(dt) = S(t)h(t)\lambda(dt),$$

it is easy to see that for any $C \in \mathcal{C}$,

$$P(Y \in C\cap\xi) = \int_C P(t \in \xi)S(t)h(t)\lambda(dt) \tag{6}$$

and so

$$\begin{aligned}
\operatorname{Cov}\big(C_C^{(n)}, C_D^{(n)}\big) &= \int_{C\cap D} P(t \in \xi)S(t)h(t)\lambda(dt) \\
&\quad - \int_D\int_C S(s)S(t)P(s \in \xi)P(t \in \xi)h(s)h(t)\lambda(ds)\lambda(dt).
\end{aligned} \tag{7}$$

For $s, t \in T$, again using the independence of the Y_i's and ξ_i's,

$$
\begin{aligned}
&E[D_n(s)D_n(t)] \\
&\quad = n^{-1} E\Bigg[\Bigg(\sum_1^n \big(I_{\{Y_i \in E_s\}} I_{\{s \in \xi_i\}} - S(s)P(s \in \xi)\big)\Bigg) \\
&\qquad\qquad \times \Bigg(\sum_1^n \big(I_{\{Y_j \in E_t\}} I_{\{t \in \xi_j\}} - S(t)P(t \in \xi)\big)\Bigg)\Bigg] \qquad (8)\\
&\quad = \mu(E_s \cap E_t)P(s, t \in \xi) - S(s)S(t)P(s \in \xi)P(t \in \xi),
\end{aligned}
$$

which immediately implies that

$$
\begin{aligned}
&\operatorname{Cov}\Big(\int_C D_n(s)h(s)\lambda(ds), \int_D D_n(t)h(t)\lambda(dt)\Big) \\
&\quad = \int_D \int_C \mu(E_s \cap E_t)P(s, t \in \xi)h(s)h(t)\lambda(ds)\lambda(dt) \qquad (9)\\
&\qquad - \int_D \int_C S(s)S(t)P(s \in \xi)P(t \in \xi)h(s)h(t)\lambda(ds)\lambda(dt).
\end{aligned}
$$

Next,

$$
\begin{aligned}
&\operatorname{Cov}\Big(C_C^{(n)}, \int_D D_n(t)h(t)\lambda(dt)\Big) \\
&\quad = n^{-1} E\Bigg[\Bigg(\sum_1^n \Big(I_{\{Y_i \in C \cap \xi_i\}} - \int_C S(s)P(s \in \xi)h(s)\lambda(ds)\Big)\Bigg) \\
&\qquad\qquad \times \Bigg(\sum_1^n \int_D \Big(I_{\{Y_j \in E_t\}} I_{\{t \in \xi_j\}} h(t)\lambda(dt) \\
&\qquad\qquad\qquad - \int_D S(t)P(t \in \xi)h(t)\lambda(dt)\Big)\Bigg)\Bigg] \\
&\quad = E\Big[\int_D I_{\{Y \in C \cap E_t \cap \xi\}} I_{\{t \in \xi\}} h(t)\lambda(dt)\Big] \qquad (10)\\
&\qquad - E\Big[\int_D I_{\{Y \in C \cap \xi\}} S(t)P(t \in \xi)h(t)\lambda(dt)\Big] \qquad (11)\\
&\qquad - E\Big[\int_D \int_C I_{\{Y \in E_t\}} I_{\{t \in \xi\}} P(s \in \xi)S(s)h(s)h(t)\lambda(ds)\lambda(dt)\Big] \qquad (12)\\
&\qquad + \Big[\int_D \int_C P(s \in \xi)P(t \in \xi)S(s)S(t)h(s)h(t)\lambda(ds)\lambda(dt)\Big]. \qquad (13)
\end{aligned}
$$

In (10), note that $\{Y \in E_t \cap \xi\} \subseteq \{t \in \xi\}$, and so by (6),

$$(10) = \int_D \int_{C\cap E_t} P(s \in \xi) S(s) h(s) \lambda(ds) h(t) \lambda(dt)$$

(14)

$$= \int_D \int_{C\cap E_t} P(s, t \in \xi) \mu(E_s \cap E_t) h(s) h(t) \lambda(ds) \lambda(dt),$$

where the last equality follows since, for $s \in E_t$, $P(s \in \xi) = P(s, t \in \xi)$ and $S(s) = \mu(E_s) = \mu(E_s \cap E_t)$.

Another application of (6) gives

$$(11) = \int_D \int_C P(s \in \xi) P(t \in \xi) S(s) S(t) h(s) h(t) \lambda(ds) \lambda(dt). \tag{15}$$

Likewise,

$$(12) = \int_D \int_C P(s \in \xi) P(t \in \xi) S(s) S(t) h(s) h(t) \lambda(ds) \lambda(dt). \tag{16}$$

Finally we have that

$$\begin{aligned}
&\mathrm{Cov}\biggl(C_C^{(n)}, \int_D D_n(t) h(t) \lambda(dt)\biggr) \\
&\quad = \int_D \int_{C\cap E_t} P(s, t \in \xi) \mu(E_s \cap E_t) h(s) \lambda(ds) h(t) \lambda(dt) \qquad (17)\\
&\qquad - \int_D \int_C P(s \in \xi) P(t \in \xi) S(s) S(t) h(s) h(t) \lambda(ds) \lambda(dt).
\end{aligned}$$

Recalling the decomposition (5) and using (7), (9) and (17), we obtain

$$\begin{aligned}
&n^{-1}\,\mathrm{Cov}\bigl(M_C^{(n)}, M_D^{(n)}\bigr) \\
&\quad = \mathrm{Cov}(C_C^{(n)}, C_D^{(n)}) + \mathrm{Cov}\biggl(\int_C D_n(s) h(s) \lambda(ds), \int_D D_n(t) h(t) \lambda(dt)\biggr) \\
&\qquad - \mathrm{Cov}\biggl(C_C^{(n)}, \int_D D_n(t) h(t) \lambda(dt)\biggr) - \mathrm{Cov}\biggl(C_D^{(n)}, \int_C D_n(s) h(s) \lambda(ds)\biggr) \\
&\quad = \int_{C\cap D} P(t \in \xi) S(t) h(t) \lambda(dt) \\
&\qquad + \int_D \int_C \mu(E_s \cap E_t) P(s, t \in \xi) h(s) h(t) \lambda(ds) \lambda(dt) \\
&\qquad - \int_D \int_{C\cap E_t} \mu(E_s \cap E_t) P(s, t \in \xi) h(s) h(t) \lambda(ds) \lambda(dt) \\
&\qquad - \int_C \int_{D\cap E_s} \mu(E_s \cap E_t) P(s, t \in \xi) h(s) h(t) \lambda(dt) \lambda(ds)
\end{aligned}$$

$$= \int_{C\cap D} P(t\in\xi)S(t)h(t)\lambda(dt)$$
$$+ \iint_{\mathcal{I}(C,D)} \mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)\lambda(ds)\lambda(dt).$$

The last equality follows from the definition of $\mathcal{I}(C,D)$, noting that $\int_C\int_{D\cap E_s} = \int_D\int_{C\cap A_t}$. This completes the proof. $\square$

We must now impose some additional structure on T and the classes $\mathcal{A}_m$.

ASSUMPTION 4.3. *T is a complete separable metric space with metric d. For each $C\in\mathcal{C}^\ell(\mathcal{A}_m)$, there exist points $t_C, t_C-\in T$ such that $C\subseteq A_{t_C}\cap E_{t_C-}$ and*

$$\lim_{m\to\infty}\sup_{C\in\mathcal{C}^\ell(\mathcal{A}_m)}\sup_{t\in C}\max\bigl(d(t_C-,t),d(t_C,t)\bigr)\to 0.$$

The points t_C, t_C- act as upper and lower bounds on the points in C: $t_C-\le t\le t_C$, $\forall t\in C$. Clearly, this assumption is satisfied by the lower layers of $\mathbf{R}^d$. The following is a straightforward corollary of Proposition 4.2.

COROLLARY 4.4. *Let g be a continuous increasing (or decreasing) function on T. Then if Assumption 4.3 obtains, $\int_\cdot g(t)M^{(n)}(dt)$ is a pseudo-strong martingale and for $C,D\in\mathcal{C}$,*

$$\operatorname{Cov}\Bigl(\int_C g(s)M^{(n)}(ds),\int_D g(t)M^{(n)}(dt)\Bigr)$$
$$= n\Bigl[\int_{C\cap D} g^2(t)S(t)P(t\in\xi)h(t)\lambda(dt) \tag{18}$$
$$+ \iint_{\mathcal{I}(C,D)} g(s)g(t)\mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)\lambda(ds)\lambda(dt)\Bigr].$$

PROOF. First we observe that $\mathcal{C}^\ell(\mathcal{A}_m)$ refines $\mathcal{C}^\ell(\mathcal{A}_n)$ when $m>n$. Therefore, if $D\in\mathcal{C}^\ell(\mathcal{A}_m)$; $C\in\mathcal{C}^\ell(\mathcal{A}_n)$ and $D\subseteq C$, we may assume without loss of generality that $t_C-\le t_D-\le t_D\le t_C$.

We shall assume that g is increasing; the proof for g decreasing is analogous. We may approximate g from below by the sequence of simple functions (g_m), where

$$g_m(t)=\sum_{C\in C^\ell(\mathcal{A}_m)} g(t_C-)I_C(t).$$

It is clear by Assumption 4.3 that $g_m(t)\uparrow g(t)$ $\forall t$. Since $M^{(n)}$is a difference of positive measures, by monotone convergence (applied twice),

$$E\Bigl[\int_C g(t)M^{(n)}(dt)\Bigm|\mathcal{G}_C^*\Bigr]=\lim_{m\to\infty}E\Bigl[\int_C g_m(t)M^{(n)}(dt)\Bigm|\mathcal{G}_C^*\Bigr]$$

and

$$\begin{aligned}\operatorname{Cov}&\left(\int_C g(t)M^{(n)}(dt), \int_D g(t)M^{(n)}(dt)\right)\\ &= \lim_{m\to\infty} \operatorname{Cov}\left(\int_C g_m(t)M^{(n)}(dt), \int_D g_m(t)M^{(n)}(dt)\right).\end{aligned}$$

Recalling that the σ-algebras $\mathcal{G}_C^*$ are decreasing, it is straightforward that $\int_\cdot g(t)M^{(n)}(dt)$ is a pseudo-strong martingale; for $D \in \mathcal{C}$,

$$\begin{aligned}E&\left[\int_D g_m(t)M^{(n)}(dt)\,\Big|\,\mathcal{G}_D^*\right]\\ &= \sum_{C\in C^\ell(\mathcal{A}_m)} g(t_C-)E\big[M_{C\cap D}^{(n)} \mid \mathcal{G}_D^*\big]\\ &= \sum_{C\in C^\ell(\mathcal{A}_m)} g(t_C-)E\big[M_{C\cap D}^{(n)} \mid \mathcal{G}_{C\cap D}^* \mid \mathcal{G}_D^*\big]\\ &= 0.\end{aligned}$$

Applying Proposition 4.2,

$$\begin{aligned}n^{-1}&\operatorname{Cov}\left(\int_C g_m(t)M^{(n)}(dt), \int_D g_m(t)M^{(n)}(dt)\right)\\ &= \sum_{C_i\in\mathcal{C}^\ell(\mathcal{A}_m)}\sum_{C_j\in\mathcal{C}^\ell(\mathcal{A}_m)} g(t_{C_i}-)g(t_{C_j}-)\operatorname{Cov}\big(M_{C_i\cap C}^{(n)}, M_{C_j\cap D}^{(n)}\big)\\ &= \sum_i \int_{C_i\cap C\cap D} g^2(t_{C_i}-)S(t)P(t\in\xi)h(t)\lambda(dt)\\ &\quad+\sum_i\sum_j \iint_{\mathcal{I}(C_i\cap C, C_j\cap D)} g(t_{C_i}-)g(t_{C_j}-)\mu(E_s\cap E_t)P(s,t\in\xi)\\ &\qquad\times h(s)h(t)\lambda(ds)\lambda(dt) \qquad (19)\\ &= \int_{C\cap D} g_m^2(t)S(t)P(t\in\xi)h(t)\lambda(dt)\\ &\quad+\iint_{\mathcal{I}(C,D)} g_m(s)g_m(t)\mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)\lambda(ds)\lambda(dt).\end{aligned}$$

Letting $m\to\infty$ in (19) completes the proof. □

Bias of $\hat{H}^{(n)}$. By the pseudo-strong martingale property,

$$E\big[\hat{H}_A^{(n)}\big] = E\big[\tilde{H}_A^{(n)}\big] = \int_A P(Z_n(t)>0)h(t)\lambda(dt).$$

It follows that the bias of $\hat{H}_A^{(n)}$ is

$$E[\hat{H}_A^{(n)}] - H_A = \int_A P(Z_n(t) = 0) h(t) \lambda(dt).$$

If $\xi_1, \xi_2, \dots$ are i.i.d. and $P(A \subseteq \xi) > 0$, then for $t \in A$, $P(Z_n(t) = 0) \to 0$ as $n \to \infty$, and $\hat{H}_A^{(n)}$ is asymptotically unbiased. This will not be the case if each observation is censored by the same set.

Consistency of $\hat{H}^{(n)}$. *In what follows, we must assume that* $\inf_{t \in A} S(t) > 0$ *and* $P(A \subseteq \xi) > 0$. We will need the following lemma, which generalizes Lemma 1 of [16]. A T-indexed random process G is increasing (resp., decreasing) if a.s. each of its sample paths is increasing (decreasing).

LEMMA 4.5. *If Assumption* 4.3 *obtains and* $(G_n(t) : t \in T)_n$ *is a sequence of increasing* (*decreasing*) *processes converging in probability pointwise to a continuous function* g, *then for each* $A \in \mathcal{A}$,

$$\sup_{t \in A} | G_n(t) - g(t) | \to_P 0.$$

PROOF. Without loss of generality, we may assume in what follows that $A \subseteq B_m$, $\forall m$ (cf. Definition 2.1). Since B_1 is compact, g is uniformly continuous on A. Therefore, given $\varepsilon > 0$, there exists m_ε such that for all $m \geq m_\varepsilon$,

$$\sup_{C \in \mathcal{C}^\ell(\mathcal{A}_m)} |g(t_C) - g(t_C-)| < \varepsilon.$$

It is easy to see that

$$\begin{aligned}\sup_{t \in A} |G_n(t) - g(t)| \leq & \sup_{C \in \mathcal{C}^\ell(\mathcal{A}_m), C \subseteq B_1} |G_n(t_C) - G_n(t_C-)| \\ & + \sup_{C \in \mathcal{C}^\ell(\mathcal{A}_m), C \subseteq B_1} |G_n(t_C) - g(t_C)| \\ & + \sup_{C \in \mathcal{C}^\ell(\mathcal{A}_m), C \subseteq B_1} |g(t_C) - g(t_C-)|.\end{aligned}$$

As $n \to \infty$, the right-hand side above converges in probability to $2 \sup_{C \in \mathcal{C}^\ell(\mathcal{A}_m), C \subseteq B_1} |g(t_C) - g(t_C-)|$, which can be made arbitrarily small for m sufficiently large. Therefore, $\sup_{t \in A} |G_n(t) - g(t)| \to_P 0$. $\square$

We return to the consistency of $\hat{H}_A^{(n)}$.

THEOREM 4.6. *Let* $A \in \mathcal{A}$. *Assume that* $\inf_{t \in A} S(t) = \inf_{t \in A} \mu(E_t) > 0$, *that* S *is continuous and that the stopping sets* $\xi_1, \xi_2, \dots$ *are i.i.d. and independent of the* Y_i*'s. If* $P(A \subseteq \xi) > 0$ *and* $P(t \in \xi)$ *is continuous in* t, *then* $\hat{H}_A^{(n)}$ *is a consistent estimator of* H_A.

PROOF. First note that $\lim_n J_n(t) = 1$ a.s. $\forall t \in A$. Therefore,

$$\hat{H}_A^{(n)} - H_A = (\hat{H}_A^{(n)} - \tilde{H}_A^{(n)}) + (\tilde{H}_A^{(n)} - H_A),$$

and the second term on the right-hand side converges to 0 in probability as $n \to \infty$. It remains to prove that $(\hat{H}_A^{(n)} - \tilde{H}_A^{(n)}) \to_P 0$,

$$\begin{aligned}
\hat{H}_A^{(n)} - \tilde{H}_A^{(n)} &= \int_A \frac{J_n(t)}{Z_n(t)} M^{(n)}(dt) \\
&= \int_A \frac{J_n(t)}{Z_n(t)} N^{(n)\xi}(dt) - \int_A J_n(t) h(t) \lambda(dt) \\
&= \frac{1}{n} \int_A \left(\frac{J_n(t)}{Z_n(t)/n} - \frac{J_n(t)}{S(t)P(t \in \xi)} \right) N^{(n)\xi}(dt) \qquad (20) \\
&\quad + \int_A \frac{J_n(t)}{nS(t)P(t \in \xi)} M^{(n)}(dt) \qquad (21) \\
&\quad + \int_A J_n(t) \left(\frac{Z_n(t)/n}{S(t)P(t \in \xi)} - 1 \right) h(t) \lambda(dt). \qquad (22)
\end{aligned}$$

We observe that

$$\begin{aligned}
(20) &= \frac{1}{n} \sum_{i=1}^n I_{\{Y_i \in A \cap \xi\}} J_n(t) \left(\left(\frac{Z_n(Y_i)}{n} \right)^{-1} - \bigl(S(Y_i) P(Y_i \in \xi)\bigr)^{-1} \right) \\
&\le \sup_{t \in A} \bigl| \bigl(Z_n(t)/n\bigr)^{-1} - \bigl(S(t)P(t \in \xi)\bigr)^{-1} \bigr|
\end{aligned}$$

and

$$(22) \le \sup_{t \in A} \left| \frac{Z_n(t)}{n} - S(t)P(t \in \xi) \right| \int_A \frac{1}{S(t)P(t \in \xi)} h(t) \lambda(dt).$$

By the strong law of large numbers, $Z_n(t)/n \to S(t)P(t \in \xi)$ almost surely, and by applying Lemma 4.5 it follows that both (20) and (22) converge in probability to 0 as $n \to \infty$.

It remains to consider (21). Let $g_n(t) = (nS(t)P(t \in \xi))^{-1}$ and note that g_n is increasing and bounded on A,

$$(21) = \int_A g_n(t) M^{(n)}(dt) + \int_A (J_n(t) - 1) g_n(t) M^{(n)}(dt). \qquad (23)$$

The second integral in (23) converges to 0 a.s. since $J_n(t) \to 1$ a.s. as $n \to \infty$. By Corollary 4.4,

$$\begin{aligned}
&E\left[\left(\int_A g_n(t) M^{(n)}(dt) \right)^2 \right] \\
&\quad = \frac{1}{n} \int_A \frac{h(t)}{S(t)P(t \in \xi)} \lambda(dt)
\end{aligned}$$

$$+\frac{1}{n}\int\int_{\mathcal{I}(A,A)}\frac{\mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)}{S(s)S(t)P(s\in\xi)P(t\in\xi)}\lambda(ds)\lambda(dt)$$
$$\to 0.$$

Therefore, (21) converges in probability to 0 as $n\to\infty$, and so $\hat{H}_A^{(n)}$ is consistent. $\square$

COMMENT. In [2], there is a thorough discussion of how the Nelson–Aalen estimator for survival data on $\mathbf{R}_+$ may be interpreted as a nonparametric maximum likelihood estimator, rather than a method of moments estimator based on a martingale estimating equation. However, the same authors also point out that the question of a NPMLE estimator becomes much more complicated with multiparameter data. Whether or not the estimator proposed here can be regarded as a NPMLE under a suitable likelihood structure is an important open question.

5. A central limit theorem. In this section we will prove a central limit theorem for the finite-dimensional distributions (fdd) of $n^{1/2}(\hat{H}^{(n)}-\tilde{H}^{(n)})$. As these processes are indexed by $\mathcal{A}$, it is unclear whether a functional CLT exists on some appropriate function space defined on $\mathcal{A}$. A priori, this would require that the limiting Gaussian process have a regular version with all sample paths in the function space. Clearly, the class $\mathcal{A}$ cannot be too large: if $\mathcal{A}$ consists of the lower layers, then the sample paths of the limiting process will be very badly behaved, and we cannot hope for a functional CLT. We will comment further on this point at the end of the section.

Throughout this section we shall assume that Assumption 4.3 holds, that S and $P(\cdot\in\xi)$ are strictly positive continuous functions on T, and that for every $A\in\mathcal{A}$, $\inf_{t\in A}S(t)>0$ and $P(A\subseteq\xi)>0$. We continue to assume that the space (T,d) is a complete separable metric space. Additionally, we will need to suppose that we can define an appropriate function space $\mathcal{D}(T)$ on T closed under linear combinations, products and quotients (when defined) which contains all continuous functions and functions of the form $I_{\{t\in B\}}$ for $B\in\mathcal{A}(u)$, and which is equipped with a metric $d_{\mathcal{D}}$ making $\sup_{t\in A}$ continuous on compact sets $A\subseteq T$. Finally, we must also assume that uniform convergence on every $A\in\mathcal{A}$ implies convergence in $d_{\mathcal{D}}$. This is the case if $T=\mathbf{R}^d$ or $\mathbf{R}^d_+$.

As the normalized sum of i.i.d. $\mathcal{D}(T)$-valued random variables, the sequence $(D_n(\cdot))$ [as defined in (4)] converges in finite-dimensional distribution to a Gaussian process on T whose covariance structure is as defined in (8). We must assume that in fact $(D_n(\cdot))$ converges in $\mathcal{D}(T)$ and that the limiting Gaussian process has a continuous version; again, this is the case if $T=\mathbf{R}^d$ or $\mathbf{R}^d_+$.

THEOREM 5.1. *Under the assumptions above*,

$$n^{1/2}(\hat{H}^{(n)}-\tilde{H}^{(n)})\to_{fdd} G,$$

where G is a mean-zero Gaussian process on $\mathcal{A}$ with covariance structure defined as follows: *for $C, D \in \mathcal{C}$*,

$$
\begin{aligned}
&\operatorname{Cov}(G_C, G_D)\\
&\quad = \int_{C\cap D} \big(S(t)P(t\in\xi)\big)^{-1} h(t)\lambda(dt) \qquad (24)\\
&\qquad + \iint_{\mathcal{I}(C,D)} \frac{\mu(E_s\cap E_t)P(s,t\in\xi)}{S(s)S(t)P(s\in\xi)P(t\in\xi)} h(s)h(t)\lambda(ds)\lambda(dt).
\end{aligned}
$$

PROOF. The general method of proof is similar to that of [17]. Recalling (1), for any $A \in \mathcal{A}$,

$$
\begin{aligned}
n^{1/2}\big(\hat{H}_A^{(n)} - \tilde{H}_A^{(n)}\big) &= \int_A \frac{J_n(t)}{Z_n(t)/n} \frac{M^{(n)}(dt)}{\sqrt{n}}\\
&= \int_A J_n(t)\bigg(\frac{n}{Z_n(t)} - \frac{1}{S(t)P(t\in\xi)}\bigg)\frac{M^{(n)}(dt)}{\sqrt{n}} \qquad (25)\\
&\quad + \int_A \frac{J_n(t)-1}{S(t)P(t\in\xi)} \frac{M^{(n)}(dt)}{\sqrt{n}} \qquad (26)\\
&\quad + \int_A \frac{1}{S(t)P(t\in\xi)} \frac{M^{(n)}(dt)}{\sqrt{n}}. \qquad (27)
\end{aligned}
$$

We shall show that for any $A \in \mathcal{A}$, both (25) and (26) converge in probability to 0 as $n \to \infty$. The CLT then follows since

$$
\int_{\cdot} \frac{1}{S(t)P(t\in\xi)} \frac{M^{(n)}(dt)}{\sqrt{n}}
$$

is the normalized sum of n i.i.d. processes; the covariance structure is given by (18).

It remains to show that (25) and (26) converge in probability to 0 as $n \to \infty$. Since $J_n(t)$ converges a.s. to 1, this is immediate for (26). Next, recalling (5),

$$
\begin{aligned}
(25) &= \int_A J_n(t)\bigg(\frac{Z_n(t) - nS(t)P(t\in\xi)}{Z_n(t)S(t)P(t\in\xi)}\bigg)\big(D_n(t)h(t)\lambda(dt) - C^{(n)}(dt)\big)\\
&= \frac{1}{\sqrt{n}}\int_A \frac{J_n(t)D_n^2(t)}{(Z_n(t)/n)S(t)P(t\in\xi)} h(t)\lambda(dt) \qquad (28)\\
&\quad - \int_A \frac{J_n(t)D_n(t)}{(Z_n(t)/n)S(t)P(t\in\xi)} \frac{C^{(n)}(dt)}{\sqrt{n}}. \qquad (29)
\end{aligned}
$$

Since $h(t)\lambda(dt) = (S(t))^{-1}\mu(dt)$,

$$
\begin{aligned}
|(28)| &= \frac{1}{\sqrt{n}} \int_A \frac{J_n(t) D_n^2(t)}{(Z_n(t)/n)(S(t))^2 P(t \in \xi)} \mu(dt) \\
&\le \frac{1}{\sqrt{n}} \sup_{t \in A} \big(D_n^2(t)\big) \sup_{t \in A} \left(\frac{n}{Z_n(t)}\right) \sup_{t \in A} \left(\frac{1}{(S(t))^2 P(t \in \xi)}\right) \mu(A). \qquad (30)
\end{aligned}
$$

Since $P(t \in \xi) \ge P(A \subseteq \xi)$, $\forall t \in A$, $\sup_{t\in A}(((S(t))^2 P(t \in \xi))^{-1})$ is bounded. By assumption, D_n converges in $\mathcal{D}(T)$ to a continuous Gaussian process; by the continuous mapping theorem, $\sup_{t\in A}(D_n^2(t))$ converges in distribution to the square of the (a.s. finite) sup of the limiting process. We recall that $n/Z_n(t)$ converges uniformly in probability to $(S(t)P(t \in \xi))^{-1}$ on each $A \in \mathcal{A}$; this implies convergence in $\mathcal{D}(T)$, and so $\sup_{t\in A} n/Z_n(t)$ converges to $\sup_{t\in A}(S(t)P(t \in \xi))^{-1}$ which is finite. By joint tightness, (30) converges in probability to 0.

The last step in the proof is to show that (29) converges in probability to 0. By the same sort of argument as in the preceding paragraph and since $J_n(t) \to 1$ a.s. for each $t \in T$, $(J_n(\cdot)D_n(\cdot)/(Z_n(\cdot)/n))$ is tight and converges weakly in $\mathcal{D}(t)$ to a continuous limit. Let $\mathcal{M}(T)$ denote the space of locally finite measures on T equipped with the topology of vague convergence. It is well known that $\mathcal{M}(T)$ is a complete separable metric space since T is. By the law of large numbers, $N^{(n)\xi}_{\cdot}/n \to_{\mathrm{fdd}} \int_{\cdot} S(t)P(t \in \xi)h(t)\lambda(dt)$ and so this sequence converges weakly in $\mathcal{M}(T)$. We now have that the sequence

$$
\left(\frac{J_n(\cdot)D_n(\cdot)}{Z_n(\cdot)/n}, \frac{N^{(n)\xi}(\cdot)}{n}\right)_n
$$

is (jointly) tight in $\mathcal{D}(T) \times \mathcal{M}(T)$.

Fix $A \in \mathcal{A}$. Let $x \in \mathcal{D}(T)$, $y \in \mathcal{M}(T)$ and define

$$
q(x, y) := \int_A \frac{x(t)}{S(t)P(t \in \xi)} \big(y(dt) - S(t)P(t \in \xi)h(t)\lambda(dt)\big). \qquad (31)
$$

Since

$$
\frac{C^{(n)}}{\sqrt{n}} = \frac{N^{(n)\xi}}{n} - \int_{\cdot} S(t)P(t \in \xi)h(t)\lambda(dt),
$$

if q is continuous at points $(x, y) \in \mathcal{D}(T) \times \mathcal{M}(T)$ when x is continuous, then (29) converges in probability to 0 as required.

The final step is the proof that q is continuous. Suppose that $x_n \to x$ in $\mathcal{D}(T)$ and $y_n \to y$ in $\mathcal{M}(T)$. Letting $\rho(\cdot) = \int_{\cdot} S(t)P(t \in \xi)h(t)\lambda(dt)$,

$$
\begin{aligned}
|q(x_n, y_n) - q(x, y)| &\le \left| \int_A \frac{x_n(t)}{S(t)P(t \in \xi)} \big(y_n(dt) - y(dt)\big) \right| \qquad (32) \\
&\quad + \left| \int_A \frac{x_n(t) - x(t)}{S(t)P(t \in \xi)} \big(y(dt) - \rho(dt)\big) \right|. \qquad (33)
\end{aligned}
$$

Since $\sup_{t\in A}$ is continuous on $\mathcal{D}(T)$,

$$(33) \le \frac{\sup_{t\in A}|x_n(t)-x(t)|}{\inf_{t\in A}(S(t)P(t\in\xi))}\big(y(A)+\rho(A)\big)\to 0.$$

Assume without loss of generality that $A \in \mathcal{A}_m\ \forall m$ and fix $\varepsilon > 0$. When x is continuous, by Assumption 4.3 and uniform continuity on A, there exists a partition $\{C_1,\dots,C_k\}$ ($C_i \in \mathcal{C}\ \forall i$) of A such that

$$\max_{1\le i\le k}\sup_{s,t\in C_i}\left|\frac{x(s)}{S(s)P(s\in\xi)}-\frac{x(t)}{S(t)P(t\in\xi)}\right|<\frac{\varepsilon}{2}.$$

Again applying continuity of $\sup_{t\in A}$ on $\mathcal{D}(T)$, it follows that for every n sufficiently large,

$$\max_{1\le i\le k}\sup_{s,t\in C_i}\left|\frac{x_n(s)}{S(s)P(s\in\xi)}-\frac{x(t)}{S(t)P(t\in\xi)}\right|<\varepsilon.$$

Choose $t_1,\dots,t_k$ such that $t_i \in C_i,\ i=1,\dots,k$. Then assuming n is sufficiently large,

$$\begin{aligned}
(32)\ &\le \sum_{i=1}^{k}\left|\int_{C_i}\left(\frac{x_n(t)}{S(t)P(t\in\xi)}-\frac{x(t_i)}{S(t_i)P(t_i\in\xi)}\right)\big(y_n(dt)-y(dt)\big)\right| \\
&\quad + \sum_{i=1}^{k}\left|\int_{C_i}\frac{x(t_i)}{S(t_i)P(t_i\in\xi)}\big(y_n(dt)-y(dt)\big)\right| \\
&\le \varepsilon\big(y_n(A)+y(A)\big)+\sum_{i=1}^{k}\frac{x(t_i)}{S(t_i)P(t_i\in\xi)}|y_n(C_i)-y(C_i)| \\
&\to 2\varepsilon y(A) \qquad \text{as } n\to\infty.
\end{aligned}$$

Since ε is arbitrary, this completes the proof. $\square$

COMMENT. The question of whether or not a functional CLT exists was raised at the beginning of this section. An appropriate function space $\mathcal{D}(\mathcal{A})$ is introduced and studied in [14], and tightness conditions are given. In particular, $\mathcal{D}(\mathcal{A})$ is a generalization of the usual Skorokhod function space $D[0,\infty)$ with the J_2 topology. In the case in which $T=[a,b]^d$, the limiting Gaussian process will not have regular sample paths over the lower layers, and so in this case we will have only convergence of finite dimensional distributions. However, if we restrict our attention to the class of rectangles $[0,t]$, then (cf. [17]) we will have functional convergence in the preceding theorem. It should be noted that we may continue to assume that the process is censored by a lower layer. We conjecture that functional convergence would hold also hold over a Vapnik–Červonenkis class of sets.

6. Applications. This article has focussed on developing the theoretical basis for survival analysis with a very general censoring mechanism. As pointed out in the Introduction, our general model can be readily applied to spatial data as well as biological data, providing a wealth of new applications for survival analysis techniques. In this section, we will indicate various specific directions for future research.

We shall begin by commenting on various potential uses of the Nelson–Aalen estimator. Clearly, as in the one-dimensional case, it can be used graphically for model checking. More generally, the Nelson–Aalen estimator can be used to test a hypothesis about the hazard function of a single population or to compare the hazards of several independent populations.

In the case in which $T = [a, b]^d$ (cf. Example 2.4), as previously noted the integrated hazard does not determine the survival function S; the lower dimensional integrated hazard functions and marginal survival functions are needed as well. Therefore, as described in [2], since we can construct the one-dimensional marginal survival functions from the one-dimensional marginal integrated hazards, we can use Nelson–Aalen estimators iteratively to construct a plug-in (Volterra) estimator of the survival function.

As well, if $T = [a, b]^d$, since the hazard function contains information about the dependence structure of μ, a test of independence of $(Y_1, \dots, Y_d)$ may be based on the Nelson–Aalen estimator (cf. [17]). This follows by noting that when the components of Y are independent, the cumulative hazard is simply the product of the marginal hazards.

All of these statistical applications will be developed in detail in a future publication.

We now describe a few of the real life situations in which we foresee applications of this theory.

1. *Forestry statistics* [22]. Image-analysis methods (remote sensing, digital aerial photographs, spectometer imaging) are used for the construction of tree position maps based on blurred and noisy images of forests. In statistical modelling, two essential random processes in forest development are particularly important: mortality and regeneration. A forest stand is the result of former land use, complex ecological processes and practical forestry. Then, the corresponding spatial point pattern is an important source of information in the plant population. Natural patterns of trees often show clumping or clustering in their starting phase, caused by environmental heterogeneity, seed dispersion and competition with other species. Typically, during the evolution of a forest, there exists a tendency towards regularity. Due to several external causes, the shape and the size of the pictures are generally random and therefore can be considered as stopping sets. So the description of this model is constituted of multivariate censored data. Here the model is spatial spreading, but it may be also spatio-temporal: $T = \mathbf{R}^2 \times \mathbf{R}_+$. Adding coordinates, we may suppose that the origin belongs to each photopraph.

For the exploratory analysis of such spatial patterns, the image is observed within a bounded but variable window, which becomes bigger and bigger. The set of all the windows may be assumed to be the indexing collection $\mathcal{A}$, and the randomness of the shape of the picture is denoted by the stopping set ξ: on the complement of ξ, the image is obscured. The censored jump process,

$$N^{\xi} = \{N_A^{\xi},\ A \in \mathcal{A}\} = \left(\sum_{i=1}^{n} I_{\{Y_i \in A\cap\xi_i\}},\ A \in \mathcal{A}\right),$$

is defined by counting the number of data points $Y_1, \dots, Y_n$ in the regions $\{A \cap \xi_i,\ \ A \in \mathcal{A}\}$, $i = 1, \dots, n$. In this model, it is reasonable to assume that the censoring mechanism is independent of the variables $Y_1, Y_2, \dots, Y_n$, and Assumption 4.3 trivially holds. Therefore the results of this paper can be applied. Following Theorems 4.6 and 5.1, we get a consistent estimator of the integrated hazard function H_A which satisfies a central limit theorem.

2. *Medical and biological sciences* [2, 10]. Using the same methods, the censored survival model can be applied in a similar way to other examples in the medical and biological sciences. One such example would be monitoring the spatial and temporal spread of an epidemic when data is not available in certain regions and time periods. Set-indexed survival analysis could also aid in the diagnosis of tumors when imaging or biopsy techniques are suspected to be faulty, in which case censoring could occur.

3. *Geology and archeology.* In geological and archeological problems, the data points generally lie in a three-dimensional space and here too, random censoring may occur through constraints in excavation. Thus, set-indexed techniques could aid in the analysis of ore samples or in the search for ancient coins, for example.

4. *Material science and technology.* Applications arise in the study of the locations of point defects on a surface of a silicon wafer when one can observe only a certain portion of the wafer. A similar problem well adapted to our framework is the statistical study of silver particles on a polished steel plate [22].

APPENDIX

PROOF OF PROPOSITION 4.1. We recall that $\mathcal{C}^{\ell}(A_k)$ partitions B_k (B_k as in Definition 2.1) and that the sequence $(\mathcal{C}^{\ell}(A_k))_k$ forms a dissecting system for T.

Since

$$\hat{H}^{(n)} - \tilde{H}^{(n)} = \int_{\cdot} \frac{J_n(t)}{Z_n(t)} M^{(n)}(dt),$$

it must be shown that for every $C \in \mathcal{C}$ and $F \in \mathcal{G}_C^*$,

$$\int_F \int_C \frac{J_n(\omega, t)}{Z_n(\omega, t)} M^{(n)}(\omega, dt) P(d\omega) = 0. \tag{34}$$

$M^{(n)}(\omega, dt)P(d\omega)$ is a difference of positive measures on $\Omega \times T$, each of which is bounded above by n. For notational convenience in what follows, we shall be suppressing dependence on n, which is always assumed to be fixed.

Next, we observe that $\frac{J(t)}{Z(t)}$ is a finite linear combination of random variables of the form

$$\prod_1^n a_i(t)b_i(t),$$

where $a_i(t)$ is either $I_{\{Y_i \in E_t\}}$ or $I_{\{Y_i \in E_t^c\}}$ and $b_i(t)$ is either $I_{\{t \in \xi_i\}}$ or $I_{\{t \in \xi_i^c\}}$.

Therefore, (34) is true if for every $C \in \mathcal{C}$ and $F \in \mathcal{G}_C^*$,

$$\int_F \int_C \prod_1^n a_i(t)b_i(t)M(\omega, dt)P(d\omega) = 0. \tag{35}$$

For B_m as defined in Definition 2.1, we have that

$$I_{\{t \in \xi_i\}} = \lim_{m \to \infty} I_{\{t \in \xi_i \cap B_m\}} \quad \text{and} \quad I_{\{Y_i \in E_t^c\}} = \lim_{m \to \infty} I_{\{Y_i \in E_t^c \cap B_m\}}.$$

Hence, by dominated convergence, (35) follows if for every m,

$$\int_F \int_C \prod_1^n a_i^m(t)b_i^m(t)M(\omega, dt)P(d\omega) = 0, \tag{36}$$

where $a_i^m(t) = 1 - I_{\{Y_i \in E_t^c \cap B_m\}}$ or $I_{\{Y_i \in E_t^c \cap B_m\}}$, and $b_i^m(t) = I_{\{t \in \xi_i \cap B_m\}}$ or $1 - I_{\{t \in \xi_i \cap B_m\}}$, as appropriate.

For the remainder of the proof we may assume that m is fixed, and so once again for notational convenience, we may suppress dependence on m. Also, by suitably augmenting the finite subsemilattices $\mathcal{A}_k$, we may assume without loss of generality that C is a (finite) union of sets from $\mathcal{C}^\ell(\mathcal{A}_k)$ for every k.

Let ξ be a stopping set and $B = B_m$. By Lemma 1.5.6 of [14], $(g_k(\xi \cap B))_k$ is (for each ω) a decreasing sequence of stopping sets, each taking on finitely many values, and $\bigcap_k g_k(\xi \cap B) = \xi \cap B$. Thus,

$$I_{\{t \in \xi_i \cap B\}} = \lim_k I_{\{t \in g_k(\xi_i \cap B)\}}.$$

Now, as in Lemma 2.1.5 of [14],

$$\{(\omega, t) : t \in g_k(\xi \cap B)\} = \bigcup_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k), C_{k,h} \subseteq B} F_{k,h} \times C_{k,h},$$

where $F_{k,h} \in \mathcal{G}_{C_{k,h}}^*$. Therefore, for each (ω, t),

$$I_{\{t \in \xi_i \cap B\}} = \lim_k \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k), C_{k,h} \subseteq B} I_{F_{k,h} \times C_{k,h}}. \tag{37}$$

We shall show that

$$\{(\omega,t): Y_i(\omega)\in E_t^c\cap B\} = \bigcup_k \bigcup_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} G_{k,h}\times C_{k,h}, \tag{38}$$

where $G_{k,h}\in\mathcal{G}^*_{C_{k,h}}$, and that this is an increasing union in k. Since the sets $(C_{k,h})_h$ are disjoint, for each (ω,t) it follows that

$$I_{\{Y_i\in E_t^c\cap B\}} = \lim_k \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} I_{G_{k,h}\times C_{k,h}}. \tag{39}$$

Using (37) and (39) it is straightforward to see that $\prod_1^n a_i(t)b_i(t)$ is a finite linear combination of random variables of the form

$$\lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k),C_{k,h}\subseteq B} I_{H_{k,h}\times C_{k,h}},$$

where $H_{k,h}\in\mathcal{G}^*_{C_{k,h}}$. By dominated convergence,

$$\begin{aligned}
&\int_F\int_C \prod_1^n a_i(t)b_i(t)M(\omega,dt)P(d\omega)\\
&\quad=\lim_{k\to\infty}\sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k),C_{k,h}\subseteq B}\int_F\int_C I_{H_{k,h}\times C_{k,h}}M(\omega,dt)P(d\omega)\\
&\quad=\lim_{k\to\infty}\sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k),C_{k,h}\subseteq B}\int_{H_{k,h}\cap F}\int_{C_{k,h}\cap C} M(\omega,dt)P(d\omega)\\
&\quad=\lim_{k\to\infty}\sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k),C_{k,h}\subseteq C}\int_{H_{k,h}\cap F} M_{C_{k,h}}\,dP\\
&\quad=0.
\end{aligned}$$

The last equality follows since $\{\mathcal{G}^*_C\}$ is a decreasing family and so $H_{k,h}\cap F\in\mathcal{G}^*_{C_{k,h}}$ when $C_{k,h}\subseteq C$. This proves (36), and the proof of the lemma will be complete once we have verified (38).

Consider $\{Y_i\in E_t^c\cap B\}$: by definition $E_t^c\cap B=\cup_{t\notin A}(A\cap B)$. By separability from above, $t\notin A$ if and only if there exists k such that $t\notin g_k(A)$, and so

$$E_t^c\cap B=\bigcup_k\left(\bigcup_{A\in\mathcal{A}_k,t\notin A} A\cap B\right).$$

Clearly, the above union is increasing in k, since the classes $\mathcal{A}_k$ are increasing. Recalling that the class of left-neighborhoods $\mathcal{C}^\ell(\mathcal{A}_k)$ partitions B_k, it is easy to see by the definition of a left-neighborhood that if $t \in C = D \setminus \bigcup_{A' \in \mathcal{A}_k, D \not\subseteq A'} A' \in \mathcal{C}^{\ell}(\mathcal{A}_k)$, then

$$\left\{ Y_i \in \bigcup_{A \in \mathcal{A}_k, t \notin A} A \cap B \right\} = \left\{ Y_i \in \bigcup_{A' \in \mathcal{A}_k, D \not\subseteq A'} A' \cap B \right\} \in \mathcal{G}_C^*.$$

It is now straightforward that (38) follows, completing the proof. □

Acknowledgments. The authors express their appreciation to an anonymous Associate Editor for a number of thought-provoking and insightful comments. The second author thanks the first author for her hospitality.

REFERENCES

[1] AL HUSSAINI, A. and ELLIOTT, R. J. (1985). Filtrations for the two-parameter jump process. *J. Multivariate Anal.* **16** 118–139.

[2] ANDERSEN, P. K., BORGAN, O., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

[3] BADDELEY, A. J. (1998). Spatial sampling and censoring. In *Stochastic Geometry*: *Likelihood and Computation* (O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, eds.) 37–78. Chapman and Hall, London.

[4] BADDELEY, A. J. and GILL, R. D. (1997). Kaplan–Meier estimators of distance distributions for spatial point processes. *Ann. Statist.* **25** 263–292.

[5] CAMPBELL, G. (1982). Asymptotic properties of several nonparametric multivariate distribution function estimators under random censoring. In *Survival Analysis* (J. Crowley and R. A. Johnson, eds.) 243–256. IMS, Hayward, CA.

[6] DABROWSKA, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489.

[7] DABROWSKA, D. M. (1989). Kaplan–Meier estimate on the plane: Weak convergence, LIL, and the bootstrap. *J. Multivariate Anal.* **29** 308–325.

[8] DAVIDSEN, M. and JACOBSEN, M. (1991). Weak convergence of two-sided stochastic integrals, with applications to models for left-truncated survival data. In *Statistical Inference in Stochastic Processes* (N. U. Prabhu and I. V. Basawa, eds.) 167–182. Dekker, New York.

[9] DE GIOSA, M. and MININNI, R. (1999). An application of the intensity-based inference for planar point processes to bivariate survival analysis. Preprint, Univ. Bari, Italy.

[10] DURRETT, R. (1995). Spatial epidemic models, In *Epidemic Models*: *Their Structure and Relation to Data* (D. Mollison, ed.) 187–201. Cambridge Univ. Press.

[11] GEYER, C. and MOLLER, J. (1994). Simulation and likelihood inference for spatial point processes. *Scand. J. Statist.* **21** 359–373.

[12] HANSEN, M. B., BADDELEY, A. J. and GILL, R. D. (1999). First contact distributions for spatial patterns: Regularity and estimation. *Adv. Appl. Probab.* **31** 15–33.

[13] HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

[14] IVANOFF, G. and MERZBACH, E. (2000). *Set-indexed Martingales*. CRC Press, Boca Raton, FL.

[15] LIN, C. Y. and KOSOROK, M. R. (1999). A general class of function-indexed nonparametric tests for survival analysis. *Ann. Statist.* **27** 1722–1744.

[16] MCLEISH, D. L. (1978). An extended martingale invariance principle. *Ann. Probab.* **6** 144–150.

[17] PONS, O. (1986). A test of independence between two censored survival times. *Scand. J. Statist.* **13** 173–185.

[18] PONS, O. (1989). Nonparametric model and Cox model for bivariate survival data. Rap. Tech. de biometrie 89-02 INRA, Lab. de biometrie, Jouy-en-Josas, France.

[19] PONS, O. and DE TURKHEIM, E. (1991). Tests of independence for bivariate censored data based on the empirical joint hazard function. *Scand. J. Statist.* **18** 21–37.

[20] PRENTICE, R. L. (2000). Nonparametric estimation of the bivariate survivor function: Research synthesis and proposals for new estimators. Preprint, Fred Hutchinson Cancer Center, Seattle, WA.

[21] RATHBUN, S. L. and CRESSIE, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Adv. Appl. Probab.* **26** 122–154.

[22] STOYAN, D. and PENTTINEN, A. (2000). Recent applications of point process methods in forestry statistics. *Statist. Sci.* **15** 61–78.

[23] TSAI, W. Y. and CROWLEY, J. (1998). A note on nonparametric estimators of the bivariate survival function under univariate censoring. *Biometrika* **85** 573–580.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF OTTAWA
P.O. BOX 450, STN. A
OTTAWA, ONTARIO K1N 6N5
CANADA
E-MAIL: givanoff@science.uottawa.ca

DEPARTMENT OF MATHEMATICS
BAR-ILAN UNIVERSITY
52900 RAMAT GAN
ISRAEL
E-MAIL: merzbach@macs.biu.ac.il