uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Alberto Carabarin Aguirre
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Mathematics)
GRADE / DEGREE

Department of Mathematics and Statistics
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Set-indexed Survival Analysis with Generalized Censoring

TITRE DE LA THÈSE / TITLE OF THESIS

Gail Ivanoff
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

John Braun

David McDonald

Rafal Kulik

Mohamedou Ould-Haye

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

SET-INDEXED SURVIVAL ANALYSIS WITH GENERALIZED CENSORING

Alberto Carabarin Aguirre

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Mathematics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. Program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Your file Votre référence
ISBN: 978-0-494-48392-3
Our file Notre référence
ISBN: 978-0-494-48392-3

Abstract

This thesis focuses on the problem of survival analysis of data subject to generalized censoring by an arbitrary adapted random set. We consider the problem of estimation of both the cumulative hazard and the distribution function. A problem with observability of certain events needed to construct the Nelson-Aalen estimator for the cumulative hazard is resolved by introducing a new model based on a more general type of random set. Under the framework of this censoring scheme, we reconstruct the Nelson-Aalen estimator and propose three different estimators for the distribution function: the first two come from a reverse probability equation and the last is a path-dependent estimator. Functional central limit theorems are proven for each of these estimators which, coupled with bootstrapping methods, allow us to explore some applications, like tests for independence of the components in the bivariate case, a test of hazard rate order and copula estimation. The novelty of this work is twofold: first, the techniques developed may be applied not only to distributions on Euclidean spaces, but also to distributions on more general metric spaces, and second, the generalized censoring mechanism subsumes and extends the usual models used in multivariate survival analyis.

Acknowledgements

I would like to thank my supervisor, Prof. Gail Ivanoff, for all her help, academic and otherwise. Her ideas and suggestions always kept me on the right path throughout the process of writing this thesis. Working with her has been an amazing experience that allowed me to have a close look at the way a true professional does her job. Professor Ivanoff, it has been a pleasure and a privilege.

I would also like to thank Consejo Nacional de Ciencia y Tecnología (CONACYT), México, for the financial support they have provided over the past three years, which made it possible for me to complete my research.

I am also indebted to the Faculty of Graduate and Postdoctoral Studies of the University of Ottawa for the funding they so kindly provided at the beginning of my doctoral program.

Finally, I must mention the Department of Mathematics at the University of Ottawa. I only have kind words for all the professors I came in contact with, as well as for the administrative staff. I am truly thankful for all their help.

Dedication

This thesis is dedicated to my parents and Adriana. Each of them, in their own way, have had a great hand in molding me into the person that I am today; it is only fitting that the fruit of these past five years of work be dedicated to them.

Contents

List of Figures

Chapter 1

Introduction

1.1 Origin and Motivation

The kind of problems we are set to face in this work are, in essence, those that survival analysis is usually concerned with. We are given a random sample coming from a multivariate distribution F, the observations have been censored in some way or another, and our goal is to construct different estimators of relevant functions, such as the integrated hazard or F itself.

This topic has been explored by several different authors who have put forward their own estimators, each with key advantages and some drawbacks. Most of them, however, concern themselves with only a particular censoring scheme, thus limiting the application of the estimators to very specific problems and data sets. This is one of the main motivations for the present thesis: we want to build a mathematical structure that can handle very general types of censoring mechanisms, since this would enable us to greatly expand the range of problems where our estimators would be of use.

A first step in this direction can be found in [11], where Ivanoff and Merzbach defined the censoring mechanism in a set-indexed framework by introducing the concept of a stopping set, a random set with a particular kind of measurability. The stopping sets acted as windows, since we could observe the data points only through them, and they allowed Ivanoff and Merzbach to consider more sophisticated censoring schemes than the usual multivariate generalizations of right-censoring; however, the stopping sets were still restricted to certain shapes.

The theory of stopping sets was expanded by Ivanoff and Merzbach to more general random sets called anti-clouds in [12]. That name conveys both the authors' source of inspiration and the role those sets would play in the survival analysis context: as they explain in that paper, the authors were thinking of the example of aerial photography, where pictures are taken from high in the air and clouds may interfere with the observation of whatever the subject of interest is. Anti-clouds, then, are the complement of the clouds: the regions where we can actually observe the data points. The difference between the stopping sets and the anti-clouds is, quite simply, that anti-clouds have virtually no restrictions in the shape they can take; obviously, this makes for much more realistic censoring schemes.

Once armed with the stopping set and the anti-cloud concepts, Ivanoff and Merzbach went after the estimation of the cumulative hazard function. The measurability requirement imposed on these random sets allowed them to use the theory of set-indexed martingales –previously developed by themselves in [9]– to produce a set-indexed Nelson-Aalen estimator for the cumulative hazard. In the case of censoring by stopping sets, they were able to prove consistency and asymptotic unbiasedness for their estimate, as well as asymptotic normality of its finite-dimensional distributions. The Nelson-Aalen estimator in the presence of censoring by clouds was more challenging, due to the geometric nature of the clouds; nevertheless, they showed the estimator to be consistent and asymptotically unbiased.

That story did not have an altogether happy ending. Aside from the lack of a functional central limit theorem for the estimator, there was a considerable problem in the anti-cloud case. Even though the mathematical structure and the technical details were all done correctly and presented no faults, it turns out that, when it comes to constructing the Nelson-Aalen estimator at a time t, a very important process would not be observable given the available information at t under the most common data structures. In practice, this would render the estimator unusable most of the time.

We are now at the point where this thesis was born. Our first goal was to somehow amend this problem and produce a working estimator for the cumulative hazard, while still preserving the censoring model of Ivanoff and Merzbach. It happens that the root of the observability problem lies in the measurability requirement of the anti-clouds. By asking for a slightly different measurability condition, it was possible for us to achieve our objective. We call the new kind of random set that came from this modification a $*$-anti-cloud. If the complements of these sets, the $*$-clouds, act as the censoring in the survival model, then we are able to construct a Nelson-Aalen estimate that can actually be observed. It is also possible to transition from the anti-cloud model to the $*$-anti-cloud structure, so that we can calculate the cumulative hazard estimate when working under the first model. The price we pay is a likely loss of previously uncensored data that will need to be discarded.

1.2 Previous Work on the Subject

In one-dimension, the study of the survival function $S(t) := P(Y \geq t)$ or, equivalently, the distribution function $F(t) = 1 - S(t)$, is intimately tied to that of the hazard function, $h(t) := -S'(t)/S(t)$. It is a well-known fact that the hazard completely determines the distribution of Y because of the relation $S(t) = \exp(-H(t))$, where $H(t) := \int_0^t h(s)ds$ is the cumulative (or integrated) hazard function (See, for example [1]). Therefore, the estimation of the cumulative hazard is a very desirable road to obtain, in turn, an estimate of the survival function. Under censoring, the Nelson-Aalen and the Kaplan-Meier estimators are widely used to estimate the cumulative hazard and survival functions, respectively.

Unfortunately, in higher dimensions the relationship between the survival and the hazard functions is not as straightforward. In two dimensions, for example, the survival function is determined by the hazard and both marginal distributions. Even more, in this case $F(t) \neq 1 - S(t)$. For a good exposition of the representation of a bivariate survival function S in terms of the hazard and the marginals, we refer the reader to [8] or [14]. Some authors elect to stay with the hazard-based approach made possible by the aforementioned representation. Usually, one-dimensional Kaplan-Meier estimates are used for the marginal survival functions; what will change is the way the hazard is estimated. As Kalbfleisch and Prentice explain in [14], the simplest of this type of estimators is due to Bickel, who uses the ratio of the number of failure points that were uncensored in both components at (t_1, t_2) to the number of pairs at risk at that point. The Dabrowska estimator and the Prentice-Cai estimator represent attempts at improving this approach, by considering not only the ratio of the number of double failure times to the at-risk set, but also the ratio of the number of failures on one coordinate when the other is still alive to the at-risk set.

Under univariate censoring, where both components of a bivariate random variable $Y = (Y_1, Y_2)$ are independently censored by the same single censoring variable C and we observe $(Y_{ij} \wedge C_i, I_{\{Y_{ij} \leq C_i\}})$ for $j = 1, 2$, Lin and Ying (See [16]) opted out of the hazard route and suggested the estimate of the survival function at a point (t_1, t_2) given by

$$\frac{\sum_{i=1}^{n} I_{\{Y_{i1} \wedge C_i \geq t_1, Y_{i2} \wedge C_i \geq t_2\}}}{n\hat{G}(t_1 \wedge t_2)},$$

where $\hat{G}$ is the Kaplan-Meier estimator of the survival function G of the censoring times; this was motivated by the representation of the survival function $S(t_1, t_2) = P(Y_1 \wedge C \geq t_1, Y_2 \wedge C \geq t_2)/G(t_1 \wedge t_2)$. Hougaard, however, does not recommend this estimate because it sometimes leads to negative masses at certain points, and in some cases it places mass on censoring points (See [8]).

Yet another approach was explored by Campbell and Földes in [5]. Back to the usual bivariate censoring, where (Y_{i1}, Y_{i2}) are independently censored by (C_{i1}, C_{i2}) and we are allowed to observe $(Y_{ij} \wedge C_{ij}, I_{\{Y_{ij} \leq C_{ij}\}})$ for $i = 1, \ldots, n$, $j = 1, 2$, they proposed a path-dependent estimator for the survival function based on the relation $S(t_1, t_2) =$

$S(t_1,0)P(Y_2 > t_2|Y_1 > t_1)$. In essence, they projected vertically all points to the left of t_1, regardless of their second component, to the line $Y_2 = 0$ and calculated the Kaplan-Meier estimate of $S(t_1, 0)$ using the sample $(Y_{i1} \wedge C_{i1}, I_{\{Y_{i1} \leq C_{i1}\}})$; then, they projected horizontally all remaining points under t_2 and to the right of t_1 to the line $Y_1 = t_1$, and calculated the Kaplan-Meier estimator with the sample $(Y_{i2} \wedge C_{i2}, I_{\{Y_{i2} \leq C_{i2}\}})$ for which $Y_{i1} \wedge C_{i1} > t_1$. Finally, their estimator is the product of those two Kaplan-Meier estimators. By reversing their process, projecting the points to the line $Y_1 = 0$, they arrive at another estimator. Later, Tsai and Crowley generalized the idea of projecting points onto a path and then work with one-dimensional data in [21], although they did it in the context of univariate censoring. They managed to show that their proposed estimator for the survival function has the smallest variance among all path-dependent estimators. The path they took to any point (t_1, t_2) is the one that follows a diagonal straight line from $(0, 0)$ to $(t_1 \wedge t_2, t_1 \wedge t_2)$ and then a straight line from $(t_1 \wedge t_2, t_1 \wedge t_2)$ to (t_1, t_2). A weak point of both the Campbell-Földes and the Tsai-Crowley estimators however, lies in that they may not actually be genuine survival functions: they may allow points with negative mass.

In [3] Burke focused, unlike the previously mentioned authors, on estimation of the distribution function F. Considering the usual bivariate model with right censoring in each coordinate, his idea is to put to use the relation

$$F(t_1, t_2) = \int^{t_2} \int^{t_1} \{G(s_1, s_2)\}^{-1} d\tilde{F}(s_1, s_2), \tag{1.2.1}$$

where G is the survival function of the censoring variables and $\tilde{F}$ is the subdistribution of the uncensored times:

$$\tilde{F}(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2, I_{\{Y_1 \leq C_1\}} = I_{\{Y_2 \leq C_2\}} = 1).$$

He then proceeds to estimate that subdistribution empirically, while estimating G via either of the two Campbell-Földes estimators. His estimators do satisfy the monotonicity requirements of a distribution. Recently, Lu and Burke expanded this idea in [17] to include the estimation of linear functionals of the distribution, this time in the presence of univariate censoring.

Lastly, Pons proposed a martingale-based estimator of the cumulative hazard of two right-censored survival times in [18], which she then used to produce a test for independence of the two survival times by comparing it to the product of the usual one-dimensional Nelson-Aalen estimates of both marginal cumulative hazards.

As we can already see from this sample of available estimators, the focus has been bestowed on many aspects, but the issue of actually exploring more general versions of censoring has been largely ignored. Also absent from the literature is the consideration of more abstract spaces, moving beyond the usual Euclidean spaces. Our goal is to employ Ivanoff and Merzbach's framework to cover these unexplored areas, to provide some sort of unifying language for many different kinds of censoring by showing that this general censoring structure encompasses some of the previously mentioned estimators as particular cases, and also to come up with new estimators that possess desirable qualities, such as consistency and proper monotonic behavior in the case of the distribution function F. We are also extremely interested in proving functional central limit theorems for our estimators, since this property would yield estimators for many functionals of F or of the cumulative hazard H by the continuous mapping theorem.

1.3 Overview

We start off by outlining the general setting in which we will be developing our survival model. Chapter 2 is devoted to the basic definitions of the spaces, filtrations and processes employed throughout this thesis. As we have mentioned, this framework is essentially the same that Ivanoff and Merzbach presented in [12], which in turn relied heavily on [9] and [11]. As an example, we elected to stay within the confines of the unit square for ease of exposition, but note that these ideas can be expanded to the whole plane; even more, we could consider locally compact complete separable metric spaces by emulating Definition 2.1 in [12].

Chapter 3 takes a look at the stopping set and anti-cloud structures of Ivanoff and Merzbach as a preamble to the formal definition of the $*$-anti-cloud model. What follows is a discussion of the properties of these new random sets, including a stopping theorem; most of the properties are shared with the anti-cloud sets, although we do find a difference in the stopping theorem, where we will lose adaptedness of the stopped process as a consequence of the new measurability condition for $*$-anti-clouds.

With the fundamental concepts in place, we turn our focus to the study of the survival model. In Chapter 4 we build the Nelson-Aalen estimator of the cumulative hazard under censoring by $*$-clouds, correcting the observability problem of the clouds model. We also expand on the asymptotic results of [12], as we are able to prove a functional central limit theorem for our estimator by means of Hadamard-differentiability and the functional Delta-Method. We also give the covariance structure for the estimator, and we show it to be consistent and asymptotically unbiased. Next, we consider some applications for the Nelson-Aalen estimator. Following in the footsteps of Pons in [18], we give a test of independence for the components of $Y = (Y_1, Y_2)$ under three different types of data structure, all within the framework of the $*$-anti-clouds, by comparing our estimate of the joint cumulative hazard to the product of the marginal Nelson-Aalen estimators. We use a bootstrap approach to apply our test and also come up with confidence bands for the difference of the Nelson-Aalen estimator and the product of the two marginal estimators. The bootstrap is proven to be valid by noting that the classes of functions needed to construct the estimator are all Donsker classes, in which case the Delta-Method in its bootstrap variant guarantees the convergence of the necessary processes. We also develop a test of hazard rate order given two different samples with a common censoring mechanism, as well as an estimate for the hazard of a copula. The chapter ends with an example, where we apply the previous results to FGM copulas.

Chapters 5 and 6 focus on estimation of the distribution function F. We do not use a hazard-based approach for this purpose because, as we remarked before, the relationship between the hazard and the distribution function outside of the comfort of one-dimensional Euclidean space is not straightforward, if it exists at all. For Chapter 5, we borrow the reverse probability approach of Burke in [3]. We consider two different cases: 'observable' censoring and 'ordered' censoring. The former means that each censoring set is completely observable, as it would be, for example, in aerial photography, where we could actually see the shape of the cloud that obscures a part of the data; the latter refers to a situation such that the censoring set takes its values in an increasing class of lower layers. An example of ordered censoring could appear in radar readings, where the observable regions are concentric circles or semicircles that increase in size as their radius expands. Univariate censoring also falls in this category. In both cases, we start with the relation

$$\mu(A) = \int_A (P(t \in \xi))^{-1} \tilde{\mu}(dt), \tag{1.3.1}$$

where A is a Borel set, μ is the distribution of the data points Y_i, ξ represents the uncensored region and $\tilde{\mu}$ is the subdistribution of the uncensored observations, i.e., $\tilde{\mu}(A) = P(Y \in A \cap \xi)$. Equation (1.3.1) is a generalization of (1.2.1). We estimate the subdistribution empirically in both the observable and the ordered censoring schemes, while $P(t \in \xi)$ is estimated differently for each case. In the presence of observable censoring, we can use its empirical counterpart; in the ordered case, the basic idea is to project both the censoring and the data points into a strictly increasing map going from $[0,1]$ to the class of lower layers, and then use the resulting one-dimensional data to arrive at an estimate. We can prove that both estimators of the distribution function follow functional central limit theorems, we give the asymptotic variance of each, and we use them to construct tests of independence and estimates for the copula associated to the distribution.

The estimators of Chapter 5 are easy to construct, but the specific censoring required is a drawback, as is the fact that they are not dynamic estimators. In

Chapter 6 we aim to arrive at an estimator of F that incorporates more general censoring schemes and that is dynamic (in the sense that it only uses data available at the time it is calculated, instead of needing the whole data set). Under the structure we provide in this Chapter, we can even consider more general spaces for our random variables to take their values in. We project data points into a path –a continuous and strictly increasing map– as well as the censoring to reduce the problem to one dimension, which leads us to path dependent estimators of both the distribution and the survival function, depending on the way we project those points. Survival estimation in this case is a generalization of Tsai and Crowley's approach in [21], but our focus lies in the estimation of the distribution function. Of course, there are many ways of choosing a path, so we study in detail the case of ordered censoring, where it is actually possible to choose a unique optimal path –that is, a path that generates more uncensored points than any other. In this particular scheme, we can find a dynamic estimator that discards no uncensored observations, follows a functional central limit theorem and is a measure that only puts mass on the observations and on some sup's of observations. It is also possible to use this estimator in tests of independence and copula estimation.

Finally, the Appendix gives a brief overview of the Nelson-Aalen and Kaplan-Meier estimators for right-censored univariate data.

Chapter 2

Framework and definitions

This work is motivated by the censoring model presented in [12]. We attempt to modify that mechanism in order to avoid certain 'observability' issues that arose in the construction of the Nelson-Aalen estimator for the integrated hazard function. This will lead us to a new type of random set that will be introduced in Chapter 3; in Chapter 4 we will explain more precisely how these new random sets help us eliminate such problems. Before that however, we must present the general setting in which these ideas will be developed.

2.1 The Set-up

The framework we will be using is very similar to that used in [9], [11] and [12], except for some details. Let T be a compact complete separable metric space and λ a finite measure on $\mathcal{B}$, the Borel sets of T. A simple example is $T = [0,1]^2$, and we will return to this space in several sections of this thesis. Preserving the notation in [12], if $\mathcal{D}$ is an arbitrary class of sets, then $\mathcal{D}(u)$ (respectively, $\mathcal{D}(cu)$) will denote the class of finite (respectively, countable) unions of sets from $\mathcal{D}$. For D an arbitrary subset of T, let T_D be a countable dense subset of D. We will use '$\subset$' to indicate strict inclusion; moreover, $\overline{D}$ denotes the closure and D° denotes the interior of D. All processes will be indexed by sets belonging to an indexing collection $\mathcal{A}$:

Definition 2.1.1 *A nonempty class $\mathcal{A}$ of compact, connected subsets of T is called an* indexing collection *if it satisfies the following:*

1. $\emptyset, T \in \mathcal{A}$, *and for* $A \in \mathcal{A}$, $A^\circ \neq A$ *if* $A \neq \emptyset$ *or* T.

2. $\mathcal{A}$ *is closed under arbitrary intersections and if* $A, B \in \mathcal{A}$ *are nonempty, then* $A \cap B$ *is nonempty. Denote* $\emptyset' = \cap_{A\in\mathcal{A}, A\neq\emptyset} A$. *If* (A_i) *is an increasing sequence in* $\mathcal{A}$, *then* $\overline{\cup_i A_i} \in \mathcal{A}$.

3. *The σ-algebra generated by* $\mathcal{A}$, $\sigma(\mathcal{A}) = \mathcal{B}$.

4. Separability from above:

   There exists an increasing sequence of countable subclasses $\mathcal{A}_n = \{A_1^n, A_2^n, ...\}$ *of* $\mathcal{A}$ *closed under intersections and satisfying* $\emptyset, \emptyset' \in \mathcal{A}_n(u)$. *As well, there exists a sequence of functions* $g_n : \mathcal{A} \to \mathcal{A}_n(u)$ *such that*

   (a) g_n *preserves arbitrary intersections and finite unions (i.e.* $g_n(\cap_{A\in\mathcal{A}'} A) = \cap_{A\in\mathcal{A}'} g_n(A)$ *for any* $\mathcal{A}' \subseteq \mathcal{A}$, *and if* $\cup_{i=1}^k A_i = \cup_{j=1}^m A'_j$, *then* $\cup_{i=1}^k g_n(A_i) = \cup_{j=1}^m g_n(A'_j)$*)*,

   (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* $A \in \mathcal{A}$, $A = \cap_n g_n(A)$,

   (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) $g_n(\emptyset) = \emptyset \ \forall n$.

Definition 2.1.2 *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: $E_t = \cap {B^c}_{B\in\mathcal{A},\ t\notin B}$.

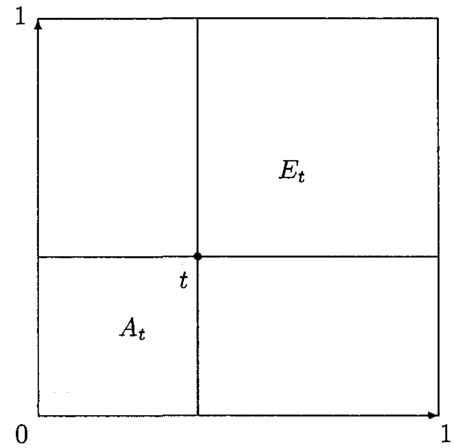

Figure 2.1: The past and the future of a point t in $[0,1]^2$.

We assume that E_t is closed.

Lemma 2.1.3 *There is a partial order induced on T by the indexing collection $\mathcal{A}$: $s \leq t$ if and only if $A_s \subseteq A_t$.*

Proof: This result was stated in [12]; we will carry out the proof. It's clear that $A_t \subseteq A_t\ \forall t \in T$, and that $A_s \subseteq A_t$, $A_t \subseteq A_u$ implies $A_s \subseteq A_u$ for $s, t, u \in T$. Now, from 3 in Definition 2.1.1 it follows that $s \neq t$ if and only if $A_s \neq A_t$. Plus, $s \in E_t$ if and only if $t \in A_s$; indeed,

$$
\begin{aligned}
t \in A_s &\Leftrightarrow t \in A\ \forall A \in \mathcal{A} : s \in A \\
&\Leftrightarrow s \in A \Rightarrow t \in A\ \forall A \in \mathcal{A} \\
&\Leftrightarrow t \in A^c \Rightarrow s \in A^c\ \forall A \in \mathcal{A} \\
&\Leftrightarrow s \in E_t.
\end{aligned}
$$

Hence $A_t \cap E_t = \{t\}$. Now, suppose $A_s \subseteq A_t$ and $A_t \subseteq A_s$. The first inclusion implies that $s \in A_t$, and so $t \in E_s$. But the second inclusion says that $t \in A_s$, and so $t \in A_s \cap E_s$, which implies that $t = s$ and hence $A_t = A_s$. $\square$

The previous lemma allows us to write $A_t = \{s : s \leq t\}$ and $E_t = \{s : s \geq t\}$. In terms of this partial order, $\emptyset'$ is the minimal element of T.

The following definitions can be found in [12], but we're going to reproduce them here for clarity of exposition. We define $\mathcal{C}$ 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 $\mathcal{C}$; in other words, $\mathcal{C}$ is a semi-algebra. Suppose that $B = \cup_{i=1}^k A_i \in \mathcal{A}(u)$. If $A_i \not\subseteq \cup_{j \neq i} A_j$ for each i, then we say that this representation is *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. Without loss of generality it will always be assumed that all representations of sets in $\mathcal{A}(u)$ and $\mathcal{C}$ are extremal.

For each $\mathcal{A}_k$, let $\mathcal{C}^\ell(\mathcal{A}_k)$ denote the *left-neighbourhoods* of $\mathcal{A}_k$: that is, all nonempty sets of the form $C = A \setminus \bigcup_{A' \in \mathcal{A}_k, A \not\subseteq A'} A', A \in \mathcal{A}_k$. It's clear that $\mathcal{C}^\ell(\mathcal{A}_k)$ partitions T, and as observed in [11], the sequence $(\mathcal{C}^\ell(\mathcal{A}_k))_k$ is a dissecting system for T. If $t \in T$, let $C_k(t)$ be the set in $\mathcal{C}^\ell(\mathcal{A}_k)$ containing t. The class of finite unions of left-neighbourhoods of $\mathcal{A}_k$ is denoted by $\mathcal{C}^\ell(\mathcal{A}_k)(u)$, and the class of countable unions will be denoted by $\mathcal{C}^\ell(\mathcal{A}_k)(cu)$.

Example 2.1.4 As an example of the type of structure we will be using, consider $T = [0,1]^2$. Then the class $\mathcal{A} = \{[0,t] : t \in T\}$ satisfies the conditions of Definition 2.1.1, and $\mathcal{C}(u)$ consists of all finite unions of disjoint rectangles of the form $(s,t]$ for $s,t \in T$.

For $T = [0,1]^2$ and $A \in T$, A will be called a *lower layer* if $[0,t] \subseteq A,\ \forall t \in A$. The class $\mathcal{A} = \{A \subseteq T : A \textit{ is a lower layer}\}$ is also an indexing collection. In both cases we have that $A_t = [0,t]$ and $E_t = [t,1]$, and both indexing collections induce the usual partial order in $[0,1]^2$.

It is important to note that we can define other indexing collections that induce different partial orders. Example 2.4 in [11] illustrates this point.

We will require some extra assumptions. The first one is about the structure of $\mathcal{C}$, while the second one is concerned with the sub-semilattices $\mathcal{A}_k$:

Assumption 2.1.5 *If $C \in \mathcal{C}$, then there exist sets $D_1, ..., D_m \in \mathcal{A}$ such that $C = A \backslash \cup_{i=1}^m D_i$ is an extremal representation, and if $A' \in \mathcal{A}$, $A' \cap C = \emptyset$, then $A' \subseteq \cup_{i=1}^m D_i$. This is called a* maximal *representation of C.*

Assumption 2.1.6 *For each $C \in \mathcal{C}^\ell(\mathcal{A}_k)$, there exists a point $t_{C^-} \in T$ such that $C \subseteq E_{t_{C^-}}$, and*

$$\lim_{k\to\infty} \sup_{C\in\mathcal{C}^\ell(\mathcal{A}_k)} \sup_{t\in C} d(t_{C^-}, t) = 0,$$

where d is the metric on T. Moreover, t_{C^-} can be chosen so that $t_{C^-} \in \partial C$.

The point t_{C^-} acts as a lower bound on the points in C: $t_{C^-} \leq t \; \forall t \in C$. For example, if $\mathcal{A} = \{[0,t] : t \in T\}$ and $C = (t, t']$, then $t_{C^-} = t$. As observed in [12], Assumption 2.1.6 implies that

$$\lim_{k\to\infty} \left(\sup_{t\in T} diam(C_k(t)) \right) = 0,$$

where '$diam(\cdot)$' denotes the diameter of a set.

Let $(\Omega, \mathcal{F}, P)$ be any complete probability space.

Definition 2.1.7 *A* filtration *indexed by $\mathcal{A}$ is a class of complete sub-σ-fields of $\mathcal{F}$, $\{\mathcal{F}_A : A \in \mathcal{A}\}$ such that*

- *If $A \subseteq B$, then $\mathcal{F}_A \subseteq \mathcal{F}_B \; \forall A, B \in \mathcal{A}$.*
- *$\mathcal{F}_{\cap A_i} = \bigcap \mathcal{F}_{A_i}$ for any decreasing sequence (A_i) in $\mathcal{A}$ (Monotone outer-continuity).*

We continue to follow the definitions in [12]. If $B \in \mathcal{A}(u)$, then $\mathcal{F}_B^0 = \bigvee_{A\in\mathcal{A},\ A\subseteq B} \mathcal{F}_A$. The σ-algebras $\{\mathcal{F}_B^0 : B \in \mathcal{A}(u)\}$ are complete and increasing, but they may not be monotone outer-continuous. To avoid this problem, we can define $\mathcal{F}_B = \bigcap_n \mathcal{F}_{g_n(B)}^0$

for $B \in \mathcal{A}(u)$. Next, for $C \in \mathcal{C}(u) \setminus \mathcal{A}$, let $\mathcal{G}_C = \bigcap_{A \in \mathcal{A},\ A \cap C \neq \emptyset} \mathcal{F}_A$ and $\mathcal{G}^*_C = \bigvee_{B \in \mathcal{A}(u),\ B \cap C = \emptyset} \mathcal{F}_B$, and for $A \in \mathcal{A}$, $A \neq \emptyset$, define $\mathcal{G}_A = \mathcal{G}^*_A = \mathcal{F}_\emptyset$. It was noted in [9] that $\mathcal{G}_C \subseteq \mathcal{G}^*_C$, and that $\{\mathcal{G}^{(*)}_C\}$ is a decreasing family of σ-fields: if $C \subseteq C'$, then $\mathcal{G}^{(*)}_{C'} \subseteq \mathcal{G}^{(*)}_C$. Finally, if $C = A \setminus B$ is a maximal representation of C, then $\mathcal{G}^*_C = \mathcal{F}_B$. This was shown in [9].

Definition 2.1.8 *The following definitions can be found in [12]:*

- *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 $E[\| X_A \|] < \infty\ \forall A \in \mathcal{A}$.*

- *A process $X : \mathcal{A} \to \mathbf{R}$ is* increasing *if for every $\omega \in \Omega$, the function $X_{\cdot}(\omega)$ can be extended to a finitely additive function on $\mathcal{C}$ satisfying $X_{\emptyset'}(\omega) = 0$ and $X_C(\omega) \geq 0$, $\forall C \in \mathcal{C}$, and such that $X_{\cdot}(\omega)$ is monotone outer-continuous (i.e. if (A_n) is a decreasing sequence of sets in $\mathcal{A}(u)$ such that $\cap_n A_n \in \mathcal{A}(u)$, then $\lim_n X_{A_n}(\omega) = X_{\cap_n A_n}(\omega)$).* (It is clear that any trajectory of an increasing process can be extended to a σ-finite measure on $\mathcal{B}$. Note that an increasing process is not necessarily adapted.)

- *An integrable process $M = \{M_A, A \in \mathcal{A}\}$ is called a* strong martingale *if it is adapted and for any $C \in \mathcal{C}$, $E[M_C|\mathcal{G}^*_C] = 0$. 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 and the difference $X - \overline{X}$ is a pseudo-strong martingale.*

Before moving on to the next Chapter dealing with adapted random sets, we should note that although we are restricting ourselves to a compact set T, all of the definitions and developments throughout this work can easily be expanded to a locally compact space using the structure of domains found in Definition 2.1 of [12].

Chapter 3

The Censoring Mechanism

In this Chapter we develop the random sets that will represent the censoring in our survival model. We start by recalling the definitions of stopping sets and anti-clouds given by Ivanoff and Merzbach, and then we define the concept of $*$-anti-cloud. The rest of the Chapter is concerned with the mathematical properties of these random sets.

3.1 Clouds and $*$-clouds

We are going to start by giving the definition of a stopping set, a notion that was first introduced in [9].

Definition 3.1.1 $\xi : \Omega \to \mathcal{A}(u)$ *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* $\{\omega : \emptyset = \xi(\omega)\} \in \mathcal{F}_\emptyset$.

As noted in [12], there is a simple but important equivalence to the above definition of a stopping set, which will make possible to extend this concept to a class of random sets taking their values in $\mathcal{B}$.

Lemma 3.1.2 $\xi : \Omega \to \mathcal{A}(u)$ *is a stopping set if and only if for any* $t \in T$, $\{\omega : t \in \xi(\omega)\} \in \mathcal{F}_{A_t}$.

Proof: Suppose that ξ is a stopping set. Then

$$\{\omega : t \in \xi(\omega)\} = \{\omega : A_t \subseteq \xi(\omega)\} \in \mathcal{F}_{A_t}.$$

On the other hand, suppose $\{\omega : t \in \xi(\omega)\} \in \mathcal{F}_{A_t}$ $\forall t \in T$. Using the fact that T is separable, we have that

$$\{\omega : A \subseteq \xi(\omega)\} = \bigcap_{t \in A} \{\omega : t \in \xi(\omega)\} = \bigcap_{t \in T_A} \{\omega : t \in \xi(\omega)\} \in \bigvee_{t \in T_A} \mathcal{F}_{A_t} \subseteq \mathcal{F}_A,$$

where T_A denotes a countable dense subset of A. □

Recall that a closed set D is a *domain* if $D = \overline{D^\circ}$. As in [12], let $\mathcal{K}$ be the class of domains D in T such that $\lambda(\partial D) = 0$, where λ is the measure defined at the beginning of Section 2.1, and let $\mathcal{L}$ be the class of open sets that are complements of sets in $\mathcal{K}$.

Definition 3.1.3 *A random set $\eta : \Omega \to \mathcal{B}$ is an* adapted random set *if for any $t \in T$, $\{\omega : t \in \eta(\omega)\} \in \mathcal{F}_{A_t}$.*

- *An adapted random set ρ taking its values in $\mathcal{L}$ is a* cloud.
- *An adapted random set ξ taking its values in $\mathcal{K}$ is an* anti-cloud.

The concept of anti-cloud was introduced by Ivanoff and Merzbach in [12] as a generalization of a stopping set. Now we are going to define a new kind of random set, analogous to the previous ones; but before doing that, we need another assumption.

Assumption 3.1.4 *Define $D_t = \overline{E_t^c}$. Then $D_t \in \mathcal{A}(u)$ $\forall$ $t \in T$. Also, if (t_n) is a sequence such that $t \leq t_n$ $\forall n$ and $t_n \to t$, then $D_t = \bigcap D_{t_n}$.*

This is crucial to some later developments, since $E_t^\circ = T \setminus D_t$ and so we will have that $E_t^\circ \in \mathcal{C}$. We may interpret E_t° as the 'strict' future of t, and we shall define the 'wide' history of t to be $\mathcal{F}_t^* := \mathcal{G}_{E_t^\circ}^* = \mathcal{F}_{D_t}$. We will also need this to deal with some right-continuity results in some of the theorems that follow, especially in the proof of Lemma 3.1.13.

Remark 3.1.5 We observe that if $C \in \mathcal{C}$ and if $C \subseteq E_t^\circ$, then $\mathcal{F}_t^* \subseteq \mathcal{G}_C^*$, since the σ-algebras $\{\mathcal{G}_C^*\}$ are decreasing. On the other hand, if $t \in C$ and $C = A \setminus B$ is a maximal representation of C, then $B \bigcap E_t^\circ = \emptyset$ and so $\mathcal{G}_C^* = \mathcal{F}_B \subseteq \mathcal{F}_{D_t} = \mathcal{F}_t^*$. Indeed, if $s \in B \bigcap E_t^\circ$, then we would have $s > t$ and $s \in B$, which would imply that $t \in B$, contradicting the assumption that $t \in C$. So $B \subseteq (E_t^\circ)^c = D_t$ and hence $\mathcal{G}_C^* \subseteq \mathcal{F}_t^*$.

Now we are in position to define the $*$-sets:

Definition 3.1.6 *$\xi : \Omega \to \mathcal{A}(u)$ is a $*$-*stopping set *if for any $t \in T$, $\{\omega : t \in \xi(\omega)\} \in \mathcal{F}_t^*$.*

- *A random set $\eta : \Omega \to \mathcal{B}$ is a $*$-*adapted random set *if for any $t \in T$, $\{\omega : t \in \eta(\omega)\} \in \mathcal{F}_t^*$.*
- *A $*$-adapted random set ρ taking its values in $\mathcal{L}$ is a $*$-*cloud.
- *A $*$-adapted random set ξ taking its values in $\mathcal{K}$ is an $*$-*anti-cloud.

As with adapted random sets and anti-clouds (cf. Ivanoff and Merzbach, [12]), we have that finite intersections and unions of $*$-adapted random sets are $*$-adapted random sets, and if η is a $*$-adapted random set and $B \in \mathcal{B}$, then $\eta \cap B$ is also a $*$-adapted random set. Also, $*$-anti-clouds are not generally closed under intersection, since the intersection of domains may not be a domain. The same happens with $A \cap \xi$ for ξ a $*$-anti-cloud and $A \in \mathcal{A}$.

Lemma 3.1.7 *Let ξ be a $*$-anti-cloud. For any $B \in \mathcal{K}$, the following are in $\mathcal{F}$:*

- *$\{\omega : B \subseteq \xi(\omega)\}$;*
- *$\{\omega : \xi(\omega) \subseteq B\}$;*
- *$\{\omega : \xi(\omega) = B\}$.*

Proof: The proof goes through exactly as in Lemma 3.4 in [12]. $\square$

Now we will focus on the measurability of $X_{A\cap\xi}(\omega) := X_{A\cap\xi(\omega)}(\omega)$ as a map from $(\Omega, \mathcal{F}, P)$ to $\mathbf{R}$, where X is an additive $\mathcal{A}$-indexed process and ξ is a $*$-anti-cloud. For that, we need to define certain sets that approximate ξ in some sense. This was done first in [12]:

Definition 3.1.8 *Let ξ be a $*$-anti-cloud.*

- $\xi_k^+ \in \mathcal{C}^\ell(\mathcal{A}_k)(u)$ *is defined by:*

$$\xi_k^+ = \bigcup_{C\in\mathcal{C}^\ell(\mathcal{A}_k),\ \overline{C}\cap\xi\neq\emptyset} C.$$

- $\xi_k^- \in \mathcal{C}^\ell(\mathcal{A}_k)(u)$ *is defined by:*

$$\xi_k^- = \bigcup_{C\in\mathcal{C}^\ell(\mathcal{A}_k),\ \overline{C}\subseteq\xi^\circ} C.$$

- $g_k(\xi) \in \mathcal{C}^\ell(\mathcal{A}_k)(u)$ *is defined by:*

$$g_k(\xi) = \bigcup_{C\in\mathcal{C}^\ell(\mathcal{A}_k),\ t_C-\in\xi} C.$$

We have that $\xi_k^- \subseteq \xi^\circ \subseteq \xi \subseteq \xi_k^+$, but ξ_k^+ and ξ_k^- are not $*$-adapted random sets. For this reason we need $g_k(\xi)$; we will prove that it is a $*$-adapted random set. The next lemma will clarify how each of these sets approximates ξ.

Lemma 3.1.9 *If ξ is a $*$-anti-cloud, then*

1. $\xi = \bigcap_k \xi_k^+$.
2. $\xi^\circ = \bigcup_k \xi_k^-$.
3. $\xi^\circ \subseteq \liminf_k g_k(\xi) \subseteq \limsup_k g_k(\xi) \subseteq \xi$.

Proof: The proof is exactly the same as the one for Lemma 3.7 in [12], since these properties are of a geometric nature and hence don't depend on the $*$-anti-cloud observability. □

The following theorem and its corollaries were proven in [12] for anti-clouds. We give the statements for $*$-anti-clouds; the proofs are identical to their anti-cloud counterparts.

Theorem 3.1.10 *Let X be an increasing process (or the difference of two increasing processes) on $\mathcal{A}$ and suppose that ξ is a $*$-anti-cloud. Then for any $A \in \mathcal{A}$, both $X_{A\cap\xi}$ and $X_{A\cap\partial\xi}$ are random variables.*

Corollary 3.1.11 *Let X be an increasing process on $\mathcal{A}$ and suppose that ξ is a $*$-anti-cloud. If $P(X_{\partial\xi} = 0) = 1$, then for $A \in \mathcal{A}$, $X_{A\cap\xi} = \lim_k X_{A\cap g_k(\xi)}$ a.s.*

Corollary 3.1.12 *Let ξ be a $*$-anti-cloud.*

1. *If $B \in \mathcal{C}^\ell(\mathcal{A}_k)(cu)$, then both $\{\omega : B = \xi_k^+(\omega)\}$ and $\{\omega : B = \xi_k^-(\omega)\}$ are in $\mathcal{F}$.*
2. *If K is a compact set, then $\{\omega : \xi(\omega) \cap K \neq \emptyset\} \in \mathcal{F}$.*

The next Lemma is analogous to Lemma 3.11 in [12] for anti-clouds, but the proof needs some modifications.

Lemma 3.1.13
1. *If ξ is a $*$-anti-cloud, then $g_k(\xi)$ is a $*$-adapted random set taking its values in $\mathcal{C}^\ell(\mathcal{A}_k)(u)$.*
2. *If ξ is a $*$-adapted random set taking its values in $\mathcal{C}^\ell(\mathcal{A}_k)(u)$ and*

$$(\emptyset, \xi] := \{(\omega, t) \in \Omega \times T : t \in \xi(\omega)\},$$

then

$$(\emptyset, \xi] = \bigcup_{C \in \mathcal{C}^\ell(\mathcal{A}_k)} F_C \times C,$$

where $F_C \in \mathcal{G}_C^*$.

Proof: 1. Adaptedness is a simple consequence of the definition of $g_k(\xi)$:

$$\{\omega : t \in g_k(\xi(\omega))\} = \{\omega : C_k(t) \subseteq g_k(\xi(\omega))\}$$

$$= \{\omega : t_{C_k(t)}- \in \xi(\omega)\} \in \mathcal{F}^*_{t_{C_k(t)}-} \subseteq \mathcal{G}^*_{C_k(t)} \subseteq \mathcal{F}^*_t,$$

using the fact that $C_k(t) \subseteq E^\circ_{t_{C_k(t)}-}$ and both parts of Remark 3.1.5.

2. Note that for $C \in \mathcal{C}^\ell(\mathcal{A}_k)$,

$$F_C = \{\omega : C \subseteq \xi(\omega)\} = \{\omega : t \in \xi(\omega)\} \in \mathcal{F}^*_t$$

for every $t \in C$. Let $(t_n) \in C$ be a monotone decreasing sequence converging to t_{C^-}. The existence of such a sequence is guaranteed because $t_{C^-} \in \partial C$. Then we have that

$$F_C \in \bigcap_n \mathcal{F}^*_{t_n} = \bigcap_n \mathcal{F}_{D_{t_n}} = \mathcal{F}_{\cap_n D_{t_n}} = \mathcal{F}_{D_{t_{C^-}}} = \mathcal{F}^*_{t_{C^-}} \subseteq \mathcal{G}^*_C,$$

where the second equality follows from monotone outer-continuity of the σ-fields $\{\mathcal{F}_{D_{t_n}}\}$ and the third comes from Assumption 3.1.4. $\square$

Now we deal with the filtered process $X^\xi_A := X_{\xi \cap A}$ when $X_{\partial\xi} = 0$ a.s.

Theorem 3.1.14 *Let ξ be a $*$-anti-cloud and $X = Y - W$, where Y and W are increasing processes such that $Y_{\partial\xi} = W_{\partial\xi} = 0$ a.s.*

1. *If X is a (pseudo)-strong martingale, then X^ξ is a pseudo-strong martingale.*

2. *X^ξ will not generally be adapted, even if X is.*

Proof:

1. We have to verify the pseudo-strong martingale property: for any $C \in \mathcal{C}$,

$$E[X^\xi_C \mid \mathcal{G}^*_C] = 0. \tag{3.1.1}$$

Suppose that $C \in \mathcal{C}^\ell(\mathcal{A}_n)(u)$ for some n. Noting that if $D \in \mathcal{C}^\ell(\mathcal{A}_k)$, $F_D = \{D \in g_k(\xi)\} = I_{\{t_{D^-} \in\ \xi\}}$, then by part 2 of Lemma 3.1.13, for $k \geq n$,

$$X_C^{g_k(\xi)} = \sum_{D \in \mathcal{C}^\ell(A_k)} I_{\{t_{D^-} \in\ \xi\}} X_{D \cap C}. \tag{3.1.2}$$

If X satisfies the pseudo-strong martingale property, then

$$\begin{aligned} E[X_C^{g_k(\xi)} \mid \mathcal{G}_C^*] &= \sum_{D \in \mathcal{C}^\ell(A_k)} E\left[I_{\{t_{D^-} \in\ \xi\}} X_{D \cap C} \mid \mathcal{G}_C^*\right] \\ &= \sum_{D \in \mathcal{C}^\ell(A_k)} E\left[E[I_{\{t_{D^-} \in\ \xi\}} X_{D \cap C} \mid \mathcal{G}_{C \cap D}^*] \mid \mathcal{G}_C^*\right] \\ &= \sum_{D \in \mathcal{C}^\ell(A_k)} E\left[I_{\{t_{D^-} \in\ \xi\}} E[X_{D \cap C} \mid \mathcal{G}_{C \cap D}^*] \mid \mathcal{G}_C^*\right] \\ &= 0, \end{aligned}$$

since $\{t_{D^-} \in\ \xi\} \in \mathcal{F}_{t_{D^-}}^* \subseteq \mathcal{G}_D^* \subseteq \mathcal{G}_{C \cap D}^*$. Now, by additivity, $X_C^{g_k(\xi)} \to_k X_C^\xi$ a.s., and since X is the difference of increasing processes, we have that

$$E[X_C^\xi \mid \mathcal{G}_C^*] = E[\lim_{k \to \infty} X_C^{g_k(\xi)} \mid \mathcal{G}_C^*] = \lim_{k \to \infty} E[X_C^{g_k(\xi)} \mid \mathcal{G}_C^*] = 0.$$

Next, consider a general set $C \in \mathcal{C}$. If $C = A \setminus B$ is its maximal representation, define $C_k = g_k(A) \setminus g_k(B)$. Using additivity of the process and outer-continuity of the g_k, we have that $X_C^\xi = \lim_{k \to \infty} X_{C_k}^\xi$. Now, since $C_k \in \mathcal{C}^\ell(\mathcal{A}_k)(u)$, the first part of the proof says that $E[X_{C_k}^\xi \mid \mathcal{G}_{C_k}^*] = 0$. Then, since we have $\mathcal{G}_C^* \subseteq \mathcal{G}_{C_k}^*$,

$$\begin{aligned} E[X_C^\xi \mid \mathcal{G}_C^*] &= E[\lim_{k \to \infty} X_{C_k}^\xi \mid \mathcal{G}_C^*] = \lim_{k \to \infty} E[X_{C_k}^\xi \mid \mathcal{G}_C^*] = \\ &= \lim_{k \to \infty} E\left[E[X_{C_k}^\xi \mid \mathcal{G}_{C_k}^*] \mid \mathcal{G}_C^*\right] = 0, \end{aligned}$$

where the second equality follows by uniform integrability, a consequence of X being the difference of increasing processes.

2. For $A \in \mathcal{A}$, $D \in \mathcal{C}^\ell(\mathcal{A}_n)$ and $A \cap D \neq \emptyset$, we have that

$$\{t_{D^-} \in\ \xi\} \in \mathcal{G}_D^* \not\subseteq \mathcal{F}_A,$$

so X^ξ may not be adapted even if X is. $\square$

Chapter 4

Estimating the Cumulative Hazard Function

The goal of this Chapter is to develop a Nelson-Aalen estimator for the cumulative (or integrated) hazard function in a general survival model, in which the data points are elements of the set T and are observable only on a $*$-anti-cloud. Since the results that we have gotten so far for the $*$-adapted random sets closely resemble those found in [12] for adapted random sets, it is to be expected that most of the developments in Section 4 of [12] will hold, with some appropriate changes.

4.1 The Survival Model and the Estimator $\hat{H}^{(n)}$

The model we will use is the same as the one in [11] and [12]. Assume that $(\Omega, \mathcal{F}, P)$ is a complete probability space. Let $Y : \Omega \to T$ be a T-valued random variable, and $\mu(B) = P\{Y \in B\}$ its distribution function. The *survival function* associated with Y is $S(t) = \mu(E_t)$. We assume that μ is absolutely continuous with respect to λ (λ as in Chapter 2) and denote by μ' the Radon-Nikodym derivative of μ with respect to λ on the Borel sets of T. Finally, we will assume throughout this section that the indexing collection is $\mathcal{A} = \{A_t : t \in T\}$.

Definition 4.1.1 • *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*

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

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 $\mathcal{F}^Y = \{\mathcal{F}^Y_A, A \in \mathcal{A}(u)\}$ its minimal filtration: $\mathcal{F}^Y_A = \sigma\{N_B : B \in \mathcal{A}, B \subseteq A\} \cup \{\mathcal{P}_0\}$, where $\mathcal{P}_0$ is the class of P-null sets. This filtration is in fact outer-continuous, as was proved in [10]. N is increasing, and it was proved in [11] that it has a $*$-compensator:

Proposition 4.1.2 *([11], 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 $A_Y(\omega) = A_{Y(\omega)}$.

Suppose we have a T-valued random variable Y whose associated single jump process N is adapted to a filtration $\mathcal{F}$ and an $\mathcal{F}$-$*$-cloud ρ with corresponding $*$-anti-cloud $\xi = \rho^c$. The filtered jump process $N^\xi_\cdot := I_{\{Y \in \xi \cap \cdot\}}$ corresponds to observing occurrences of Y only on the complement of the $*$-cloud; then we can say that N has been filtered by the $*$-cloud ρ.

Example 4.1.3 This example shows that our framework includes the usual bivariate censoring model, as presented in [18]. Let $Y = (Y_1, Y_2) \in [0,1]^2$ be a two-dimensional failure time on a probability space $(\Omega, \mathcal{F}, P)$, and suppose that $\mathcal{F}$ satisfies $\mathcal{F}^Y_t \subseteq \mathcal{F}_t$ for every $t \in [0,1]^2$. Let $t = (t_1, t_2) \in [0,1]^2$, $\mathcal{F}^{(1)}_{t_1} = \bigvee_{t_2} \mathcal{F}_{(t_1,t_2)}$, $\mathcal{F}^{(2)}_{t_2} = \bigvee_{t_1} \mathcal{F}_{(t_1,t_2)}$, and let $\tau = (\tau_1, \tau_2)$ be a two-dimensional censoring time, where τ_i is an $\mathcal{F}^{(i)}$-stopping time for $i = 1, 2$.

Suppose we can observe $Y \wedge \tau = (Y_1 \wedge \tau_1, Y_2 \wedge \tau_2)$ and $I_{\{Y_i \leq \tau_i\}}$, $i = 1, 2$. Now we can express everything in terms of sets, letting $A_t = [0, t]$, and $\xi = [0, \tau]$. ξ is a $*$-stopping set, since

$$\{t \in \xi\} = \{(t_1, t_2) \in [0, \tau_1] \times [0, \tau_2]\} = \{t_1 \leq \tau_1\} \cap \{t_2 \leq \tau_2\} \in \mathcal{F}^{(1)}_{t_1} \otimes \mathcal{F}^{(2)}_{t_2} \subseteq \mathcal{F}^*_t.$$

Finally, the counting process of censored times is

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

In what follows, if $\xi : \Omega \to \mathcal{K}$, let $\mathcal{F}^\xi$ denote the minimal ($\mathcal{A}$-indexed) filtration with respect to which ξ is a $*$-anti-cloud.

Definition 4.1.4 *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 $\mathcal{F}^Y_A \subseteq \mathcal{F}_A$ $\forall A \in \mathcal{A}(u)$ and let ξ be an $\mathcal{F}$-$*$-anti-cloud. ξ is*

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

2. *independent of Y if $\mathcal{F}^Y$ is independent of $\mathcal{F}^\xi$ and $\mathcal{F}_A = \mathcal{F}^Y_A \vee \mathcal{F}^\xi_A$, $\forall A \in \mathcal{A}$.*

Lemma 4.1.5 *Suppose that ξ is a $*$-anti-cloud, that Y and ξ are weakly independent and that the filtration $\mathcal{F}$ satisfies $\mathcal{F}_A = \mathcal{F}^Y_A \vee \mathcal{F}^\xi_A$, $\forall A \in \mathcal{A}$. If N is the single-jump process associated with Y, then $(N - \overline{N})^\xi$ is a pseudo-strong ($\mathcal{F}$-)martingale.*

Proof: This follows from Theorem 3.1.14. $\square$

Now we are able to define the Nelson-Aalen estimator of the integrated hazard function using filtered data. We will use the same definitions as in [12].

Henceforth, we assume that we have a sequence of i.i.d. T-valued random variables (Y_i) with the same distribution as Y, as well as a sequence (ξ_i) of $*$-anti-clouds independent of the $Y_i's$. Define the following processes:

$$\begin{aligned} N_A^{(n)} &= \sum_{i=1}^{n} I_{\{Y_i \in A\}}, \\ Z_n(t) &= \sum_{i=1}^{n} I_{\{Y_i \in E_t\}} I_{\{t \in \xi_i\}}, \\ \overline{N}_A^{(n)} &= \int_A Z_n(t) h(t) \lambda(dt), \\ N^{(n)}{}_A^{\xi} &= \sum_{i=1}^{n} I_{\{Y_i \in A \cap \xi_i\}}. \end{aligned} \tag{4.1.1}$$

By independence and Lemma 4.1.5, $\overline{N}^{(n)}$ is a $*$-compensator for $N^{(n)\xi}$, and the process

$$M_{\cdot}^{(n)} = N^{(n)\xi}_{\cdot} - \int_{\cdot} Z_n(t) h(t) \lambda(dt) \tag{4.1.2}$$

is a pseudo-strong martingale with respect to $\mathcal{F}$, the minimal filtration generated by the sequences (Y_i) and (ξ_i). Since

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

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

$$\hat{H}_A^{(n)} = \int_A \frac{N^{(n)\xi}(dt)}{Z_n(t)} = \sum_{\{i: Y_i \in A \cap \xi_i\}} (Z_n(Y_i))^{-1}. \tag{4.1.3}$$

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,

$$\hat{H}_A^{(n)} - \tilde{H}_A^{(n)} = \int_A \frac{J_n(t)}{Z_n(t)} M^{(n)}(dt).$$

Proposition 4.1.6 *$\hat{H}^{(n)} - \tilde{H}^{(n)}$ is a pseudo-strong martingale.*

Proof: We have that

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

so we must show that for every $C \in \mathcal{C}$ and $F \in \mathcal{G}^*_C$,

$$\int_F \int_C \frac{J_n(\omega,t)}{Z_n(\omega,t)} M^{(n)}(\omega, dt) P(d\omega) = 0. \tag{4.1.4}$$

We note that $\frac{J_n(t)}{Z_n(t)}$ is a finite linear combination of random variables of the form

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

where $a_i(t) = 1 - I_{\{Y_i \in E_t^c\}}$ or $I_{\{Y_i \in E_t^c\}}$ and $b_i(t) = I_{\{t \in \xi_i\}}$ or $1 - I_{\{t \in \xi_i\}}$. Therefore, (4.1.4) will hold if

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

for every $C \in \mathcal{C}$ and $F \in \mathcal{G}^*_C$.

From part 3 of Lemma 3.1.9, we have that

$$I_{\{t \in \xi_i^0\}} \leq \liminf_k I_{\{t \in g_k(\xi_i)\}} \leq \limsup_k I_{\{t \in g_k(\xi_i)\}} \leq I_{\{t \in \xi_i\}}. \tag{4.1.5}$$

Using part 2 of Lemma 3.1.13,

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

where $F_{k,h} \in \mathcal{G}^*_{C_{k,h}}$, so

$$\liminf_k I_{\{t \in g_k(\xi_i)\}} = \liminf_k \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k)} I_{F_{k,h} \times C_{k,h}} \tag{4.1.6}$$

and

$$\limsup_k I_{\{t \in g_k(\xi_i)\}} = \limsup_k \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k)} I_{F_{k,h} \times C_{k,h}}. \tag{4.1.7}$$

Now we are going to show that

$$\{(\omega,t) : Y_i(\omega) \in E_t^c\} = \bigcup_k \bigcup_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} G_{k,h} \times C_{k,h}, \tag{4.1.8}$$

where $G_{k,h} \in \mathcal{G}^*_{C_{k,h}}$. Consider $\{Y_i \in E_t^c\}$; by definition $E_t^c = \bigcup_{t\notin A,\ A\in\mathcal{A}} A$. By separability from above, $t \notin A$ if and only if there exists k such that $t \notin g_k(A)$, and so

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

This union is increasing in k, since the classes $\mathcal{A}_k$ are increasing. By definition of a left-neighborhood, 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

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

From this (4.1.8) follows. Now, since the sets $(C_{k,h})_h$ are disjoint, for each (ω,t) we have that

$$I_{\{Y_i\in E_t^c\}} = I_{\{\bigcup_k \bigcup_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} G_{k,h}\times C_{k,h}\}} = \lim_k \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} I_{G_{k,h}\times C_{k,h}}. \tag{4.1.9}$$

Combining (4.1.6), (4.1.7) and (4.1.9), and taking (4.1.5) into account, it follows that $\frac{J_n(t)}{Z_n(t)}$ is a finite linear combination of random variables $W_1(t), \ldots, W_m(t)$ where $0 \le W_j(t) \le 1$ for each j, $0 \le W_j(t) - V_j(t) \le I_{\{t\in\partial\xi\}}$ for a random variable $V_j(t)$, and there exist events $H_{k,h} \in \mathcal{G}^*_{C_{k,h}}$ such that

$$\begin{aligned} W_j(\omega,t) &\ge \limsup_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} I_{H_{k,h}\times C_{k,h}}(\omega,t) && (4.1.10)\\ &\ge \liminf_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} I_{H_{k,h}\times C_{k,h}}(\omega,t) && (4.1.11)\\ &\ge V_j(\omega,t). && (4.1.12)\end{aligned}$$

So all we need now is to show that

$$\int_F \int_C W_j(\omega,t) M^{(n)}(\omega,dt) P(d\omega) = 0, \tag{4.1.13}$$

where $F \in \mathcal{G}_C^*$; we may assume that $C \in \mathcal{C}^\ell(\mathcal{A}_k)(u)$ for every k without any loss of generality, since we may suitably augment $\mathcal{A}_k$. In order to do this, note that μ is absolutely continuous with respect to λ and ξ takes its values in $\mathcal{K}$ to get both

$$\begin{aligned} 0 \le \int_F \int_C (W_j - V_j) dN^{(n)\xi} dP &\le \int_F \int_C I_{\{t \in \partial\xi\}} dN^{(n)\xi} dP \\ &= \int_F \int_{\partial\xi} dN^{(n)\xi} dP \\ &\le E[N_{\partial\xi}^{(n)\xi}] = 0 \end{aligned}$$

and

$$\begin{aligned} 0 \le \int_F \int_C (W_j - V_j) d\overline{N}^{(n)} dP &\le \int_F \int_C I_{\{t \in \partial\xi\}} d\overline{N}^{(n)} dP \\ &= \int_F \int_{\partial\xi} d\overline{N}^{(n)} dP \\ &\le E[\overline{N}_{\partial\xi}^{(n)}] = 0, \end{aligned}$$

hence the equalities

$$\begin{aligned} \int_F \int_C W_j \; dN^{(n)\xi} dP &= \int_F \int_C V_j \; dN^{(n)\xi} dP, \\ \int_F \int_C W_j \; d\overline{N}^{(n)} dP &= \int_F \int_C V_j \; d\overline{N}^{(n)} dP. \end{aligned}$$

Now we can use (4.1.10)-(4.1.12) to get

$$\begin{aligned} \int_F \int_C W_j \; dN^{(n)\xi} dP &\ge \int_F \int_C \limsup_{k\to\infty} \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k)} I_{H_{k,h} \times C_{k,h}} dN^{(n)\xi} dP \\ &\ge \int_F \int_C \liminf_{k\to\infty} \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k)} I_{H_{k,h} \times C_{k,h}} dN^{(n)\xi} dP \\ &\ge \int_F \int_C V_j \; dN^{(n)\xi} dP, \end{aligned}$$

as well as

$$\begin{aligned} \int_F \int_C W_j \; d\overline{N}^{(n)} dP &\ge \int_F \int_C \limsup_{k\to\infty} \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k)} I_{H_{k,h} \times C_{k,h}} d\overline{N}^{(n)} dP \\ &\ge \int_F \int_C \liminf_{k\to\infty} \sum_{C_{k,h} \in \mathcal{C}^\ell(\mathcal{A}_k)} I_{H_{k,h} \times C_{k,h}} d\overline{N}^{(n)} dP \\ &\ge \int_F \int_C V_j \; d\overline{N}^{(n)} dP, \end{aligned}$$

which justify the calculations

$$
\begin{aligned}
\int_F \int_C W_j \, dN^{(n)\xi} dP &= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} \int_F \int_C I_{H_{k,h}\times C_{k,h}} dN^{(n)\xi} dP \\
&= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} \int_{H_{k,h}\cap F} \int_{C_{k,h}\cap C} dN^{(n)\xi} dP \\
&= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k), C_{k,h}\subseteq C} \int_{H_{k,h}\cap F} N^{(n)\xi}_{C_{k,h}} dP
\end{aligned}
$$

and

$$
\begin{aligned}
\int_F \int_C W_j \, d\overline{N}^{(n)} dP &= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} \int_F \int_C I_{H_{k,h}\times C_{k,h}} d\overline{N}^{(n)} dP \\
&= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k)} \int_{H_{k,h}\cap F} \int_{C_{k,h}\cap C} d\overline{N}^{(n)} dP \\
&= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k), C_{k,h}\subseteq C} \int_{H_{k,h}\cap F} \overline{N}^{(n)}_{C_{k,h}} dP.
\end{aligned}
$$

But then we have that

$$
\begin{aligned}
\int_F \int_C W_j \, dM^{(n)} dP &= \int_F \int_C W_j \, dN^{(n)\xi} dP - \int_F \int_C W_j \, d\overline{N}^{(n)} dP \\
&= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k), C_{k,h}\subseteq C} \int_{H_{k,h}\cap F} (N^{(n)\xi}_{C_{k,h}} - \overline{N}^{(n)}_{C_{k,h}}) dP \\
&= \lim_{k\to\infty} \sum_{C_{k,h}\in\mathcal{C}^\ell(\mathcal{A}_k), C_{k,h}\subseteq C} \int_{H_{k,h}\cap F} M^{(n)}_{C_{k,h}} dP \\
&= 0,
\end{aligned}
$$

where the last equality follows since $\{\mathcal{G}^*_C\}$ is a decreasing family and so $H_{k,h} \cap F \in \mathcal{G}^*_{C_{k,h}}$ whenever $C_{k,h} \subseteq C$. This proves (4.1.13), and the proof of the proposition is complete. □

So far, we have been able to develop a censoring mechanism based on the new random sets called $*$-anti-clouds. Our work has led us to an estimator for the integrated hazard function under that type of censoring, but so far we haven't explained the reasons for wanting to switch from the anti-cloud model presented in [12] to our model. We will address this point right now.

Example 4.1.7 Ideally, we would like to have all the information regarding the events in $\mathcal{F}_t^*$ at time t. Unfortunately, this is generally not possible in practice, and we have to settle for somewhat more limited information. Specifically, if ξ is a fully observable anti-cloud, we can only observe

$$\mathcal{H}_t = \sigma\{I_{\{Y\in A\cap\xi\}}, I_{\{s\in\xi\}} : \ A \subseteq D_t, \ s \in D_t\} \subseteq \mathcal{F}_t^*.$$

We need the event $I_{\{Y\in E_t\}}I_{\{t\in\xi\}}$ to be observable in order to construct the estimator for the integrated hazard function. But the problem is that $I_{\{Y\in E_t\}}I_{\{t\in\xi\}}$ is not $\mathcal{H}_t$-measurable, and so the estimator cannot be used. Indeed, suppose $t \in \xi$ and $D_t \cap \xi^c \neq \emptyset$. If Y is not observed in $D_t \cap \xi$, it is impossible to know whether $Y \in D_t \cap \xi^c$ or $Y \in E_t$. This situation is illustrated in Figure 4.1.

To correct this, we can define $\xi^*(\omega) = \{t : D_t \cap \xi^c(\omega) = \emptyset\}$. ξ^* is a $*$-anti-cloud, since

$$\{t \in \xi^*\} = \{D_t \cap \xi^c = \emptyset\} = \{D_t \subseteq \xi\} = \bigcap_{s\in D_t} \{s \in \xi\} = \bigcap_{s\in T_{D_t}} \{s \in \xi\} \in \mathcal{H}_t \subseteq \mathcal{F}_t^*,$$

recalling that T_{D_t} stands for a countable dense subset of D_t. Figure 4.2 pictures the $*$-anti-cloud ξ^* corresponding to the anti-cloud ξ in Figure 4.1.

Now consider $I_{\{Y\in E_t\}}I_{\{t\in\xi^*\}}$. We have that if $t \in \xi^*$, then $D_t \subseteq \xi$, and so

$$\begin{aligned} \{Y \in E_t\} \cap \{t \in \xi^*\} &= \{Y \in D_t\}^c \cap \{t \in \xi^*\} \\ &= \{Y \in D_t \cap \xi\}^c \cap \{t \in \xi^*\} \in \mathcal{H}_t. \end{aligned}$$

This means that $I_{\{Y\in E_t\}}I_{\{t\in\xi^*\}}$ is $\mathcal{H}_t$-measurable and hence observable, and now we are able to calculate the estimator.

Of course, there are instances where it is not possible to observe $I_{\{t\in\xi\}}$ whenever the observation Y happens before time t; in other words, when the anti-cloud is not

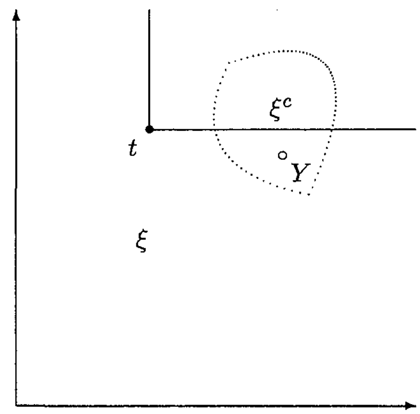

Figure 4.1: The observability problem: we don't know whether $Y \in E_t$.

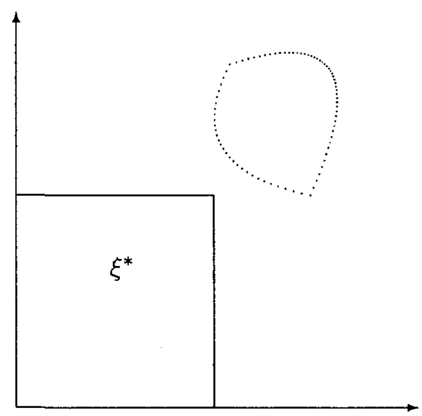

Figure 4.2: The resulting $*$-anti-cloud ξ^*.

fully observable and $Y \in A_t$. In this case, the information available up to time t consists of

$$\mathcal{H}_t = \sigma\{I_{\{Y \in A \cap \xi\}}, I_{\{s \in \xi \cap E_Y^c\}} : A \subseteq D_t,\ s \in D_t\} \subseteq \mathcal{F}_t^*;$$

nevertheless, we can still observe the event $I_{\{Y \in E_t\}} I_{\{t \in \xi^*\}}$: indeed, using the fact that $\{Y \in E_t\} = \bigcap_{s \in T_{D_t}^\circ} \{s \in E_Y^c\}$, we have that

$$\begin{aligned}\{Y \in E_t\} \cap \{t \in \xi^*\} &= \bigcap_{s \in T_{D_t}^\circ} (\{s \in E_Y^c\} \cap \{s \in \xi\}) \\ &= \bigcap_{s \in T_{D_t}^\circ} \{s \in E_Y^c \cap \xi\} \in \mathcal{H}_t.\end{aligned}$$

Example 4.1.8 We now consider a generalization of Example 4.1.3. Again, let $Y = (Y_1, Y_2) \in [0,1]^2$ be a two-dimensional failure time on a probability space $(\Omega, \mathcal{F}, P)$, $t = (t_1, t_2) \in [0,1]^2$ with $\mathcal{F}_t^Y \subseteq \mathcal{F}_t$ for every $t \in [0,1]^2$ and $\mathcal{F}_{t_i}^{(i)}$ as in Example 4.1.3 for $i = 1, 2$; but instead of having a single two-dimensional censoring time, we will consider a finite sequence of $\mathcal{F}^{(i)}$-stopping times $0 = \eta_{i1} \leq \nu_{i1} \leq \eta_{i2} \leq \nu_{i2} \leq \ldots \leq \eta_{in_i} \leq \nu_{in_i}$ for $i = 1, 2$. Let $s = (s_1, s_2)$ and $u = (u_1, u_2)$ and suppose we can observe

$$\mathcal{H}_t = \{I_{\{Y_i \leq s_i, Y_i \in \bigcup_{j=1}^{n_i} [\eta_{ij}, \nu_{ij}]\}}, I_{\{Y_i \in (\nu_{ij}, \eta_{i,j+1}), \eta_{i,j+1} \leq u_i\}} : s, u \in A_t;\ i = 1, 2\}.$$

An illustration of this kind of situation can be seen in Figure 4.3. This situation could arise in a laboratory, where test animals are under continuous observation during the day, but not at night.

Let $\xi^* = \{\bigcup_{j=1}^{n_1} [\eta_{1j}, \nu_{1j}] \times \bigcup_{j=1}^{n_2} [\eta_{2j}, \nu_{2j}]\}$. We have that

$$\begin{aligned}\{t \in \xi^*\} &= \{(t_1, t_2) \in \bigcup_{j=1}^{n_1} [\eta_{1j}, \nu_{1j}] \times \bigcup_{j=1}^{n_2} [\eta_{2j}, \nu_{2j}]\} \\ &= \{t_1 \in \bigcup_{j=1}^{n_1} [\eta_{1j}, \nu_{1j}]\} \cap \{t_2 \in \bigcup_{j=1}^{n_2} [\eta_{2j}, \nu_{2j}]\} \\ &= (\bigcup_{j=1}^{n_1} \{t_1 \in [\eta_{1j}, \nu_{1j}]\}) \cap (\bigcup_{j=1}^{n_2} \{t_2 \in [\eta_{2j}, \nu_{2j}]\}) \\ &\in \mathcal{F}_{t_1}^{(1)} \otimes \mathcal{F}_{t_2}^{(2)} \subseteq \mathcal{F}_t^*,\end{aligned}$$

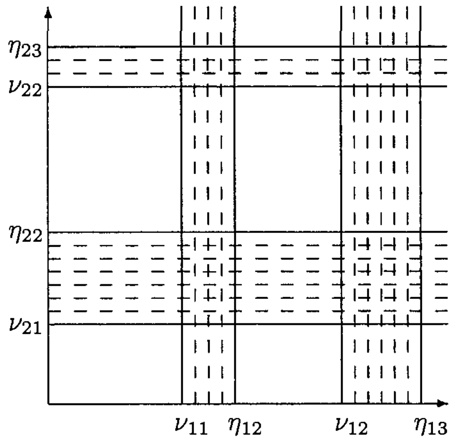

Figure 4.3: We can observe the failure times in the unshaded area.

so ξ^* is a $*$-anti-cloud. Then $I_{\{Y \in E_t\}} I_{\{t \in \xi^*\}}$ can be observed and we get our estimator for the integrated hazard function. $\square$

4.2 Asymptotic Behavior of $\hat{H}^{(n)}$

In this Section, we will assume that $T = [0,1]^2$. Our objective is to prove a functional central limit theorem for the T-indexed process $U_t^{(n)} := \sqrt{n}(\hat{H}_{A_t}^{(n)} - H_{A_t})$. In order to do this, we are going to employ a general form of the Delta-Method, which in turn is based on the concept of Hadamard differentiability. The following definition is taken from [22].

Definition 4.2.1 *Let* $\mathbf{D}$ *and* $\mathbf{E}$ *be normed spaces. A map* $\phi : \mathbf{D}_\phi \subseteq \mathbf{D} \to \mathbf{E}$ *is Hadamard-differentiable at* $\theta \in \mathbf{D}_\phi$ *if there is a continuous linear map* $\phi'_\theta : \mathbf{D} \to \mathbf{E}$ *such that*

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h) \tag{4.2.1}$$

as $n \to \infty$ *for all converging sequences* $t_n \to 0$ *and* $h_n \to h$ *such that* $\theta + t_n h_n \in \mathbf{D}_\phi$ *for every* n.

If we require in our previous definition that every $h_n \to h$ has $h \in \mathbf{D}_o$ for some set $\mathbf{D}_o \subseteq \mathbf{D}$, then the map is called *Hadamard-differentiable tangentially* to $\mathbf{D}_o$. The derivative need then be defined on $\mathbf{D}_o$ only.

For the rest of the thesis we will make extensive use of the following spaces. If X is an arbitrary set, then the Banach space $l^\infty(X)$ is the set of all functions $f : X \to \mathbf{R}$ that are bounded uniformly and equipped with the norm $||f|| = \sup_x |f(x)|$. Furthermore, $D[0,1]^2$ is the Banach space of all functions $f : [0,1]^2 \to \mathbf{R}$ continuous from the upper right quadrant and with limits from the other quadrants, equipped with the uniform norm. Finally, $BV_M[0,1]^2$ denotes the space of all functions in $D[0,1]^2$ with total variation bounded by M. Products of these spaces will always be equipped with a product norm.

The next Lemma will be very helpful in the proof of the functional central limit theorem. It is very similar to Lemma 3.9.17 in [22], the difference being that we are working on two-dimensional spaces and that we are considering functions in a larger space. The proof is also similar to its counterpart in [22].

Lemma 4.2.2 *For each fixed M, the maps $\phi : l^\infty[0,1]^2 \times BV_M[0,1]^2 \to \mathbf{R}$ and $\psi : l^\infty[0,1]^2 \times BV_M[0,1]^2 \to D[0,1]^2$,*

$$\phi(A,B) = \int_{(0,1]} AdB, \qquad \psi(A,B)(t) = \int_{(0,t]} AdB \tag{4.2.2}$$

are Hadamard-differentiable tangentially to $C[0,1]^2 \times D[0,1]^2$ at each (A,B) in $l^\infty[0,1]^2 \times BV_M[0,1]^2$ such that $\int |dA| < \infty$, and the derivatives are given by

$$\phi'_{A,B}(\alpha,\beta) = \int Ad\beta + \int \alpha dB, \qquad \psi'_{A,B}(\alpha,\beta)(t) = \int_{(0,t]} Ad\beta + \int_{(0,t]} \alpha dB,$$

where $\int Ad\beta$ is defined via the two-dimensional integration by parts formula found in Theorem 8.8. of [7].

Proof: Let $\alpha_t \to \alpha$ and $\beta_t \to \beta$, where $(\beta_t), \beta \in D[0,1]^2$, $\alpha \in C[0,1]^2$ and $(\alpha_t) \in l^\infty[0,1]^2$. Define $A_t = A + t\alpha_t$ and $B_t = B + t\beta_t$, and we consider only perturbations such that (A_t, B_t) is contained in the given domain. Using (4.2.1), we have that

$$\begin{aligned}\frac{\int A_t dB_t - \int AdB}{t} - \int Ad\beta_t - \int \alpha_t dB &= \int \alpha_t d(B_t - B) \\ &= \int \alpha d(B_t - B) \\ &+ \int (\alpha_t - \alpha) d(B_t - B).\end{aligned}$$

Since A is of bounded variation, the derivative map is continuous with respect to the uniform norm, so we only need to show that the last expression converges to zero. For the last term, we have that

$$\left|\int (\alpha_t - \alpha) d(B_t - B)\right| \le 2M \left\|\alpha_t - \alpha\right\|_\infty \to 0,$$

since $\alpha_t \to \alpha$. As for the first term, we are assuming that $\alpha \in C[0,1]^2$, so there exists $k \in \mathbf{N}$ such that α varies less that ϵ on each block $\pi_{ij} = (\frac{i-1}{2^k}, \frac{i}{2^k}] \times (\frac{j-1}{2^k}, \frac{j}{2^k}]$ for $i, j = 1, \ldots 2^k$. Let $\overline{\alpha}$ be the discretization that is constant and equal to $\alpha(\frac{i}{2^k}, \frac{j}{2^k})$ on the block π_{ij}. Then we have

$$\begin{aligned}\left|\int \alpha d(B_t - B)\right| &\le 2M \left\|\alpha - \overline{\alpha}\right\|_\infty + \sum_{i=1}^{2^k}\sum_{j=1}^{2^k} \left|\alpha(\frac{i}{2^k}, \frac{j}{2^k})\right|\left|(B_t - B)(\pi_{ij})\right| \\ &+ |\alpha(1)||(B_t - B)(1)|,\end{aligned}$$

where $(B_t - B)(\pi_{ij})$ denotes the increment of $B_t - B$ over the block π_{ij}; that is,

$$\begin{aligned}(B_t - B)(\pi_{ij}) &= (B_t - B)(\frac{i}{2^k}, \frac{j}{2^k}) - (B_t - B)(\frac{i}{2^k}, \frac{j-1}{2^k}) \\ &- (B_t - B)(\frac{i-1}{2^k}, \frac{j}{2^k}) + (B_t - B)(\frac{i-1}{2^k}, \frac{j-1}{2^k}).\end{aligned}$$

The first term can be made arbitrarily small by the choice of ϵ, while the second term is bounded by $(4 \cdot 2^{2k} + 1) \left\|B_t - B\right\|_\infty \left\|\alpha\right\|_\infty$ and hence converges to zero as $t \to 0$ for every fixed partition.

The proof for ψ is the same as that for ϕ, since the same bounds still work uniformly for all $t \in [0,1]^2$. $\square$

Now that we have the definition of Hadamard differentiability and its refinement, we can take a look at the Delta-Method. Its proof can be found in [22].

Theorem 4.2.3 **(Delta – Method)**

Let $\mathbf{D}$ *and* $\mathbf{E}$ *be normed spaces. Let* $\phi : \mathbf{D}_\phi \subseteq \mathbf{D} \to \mathbf{E}$ *be Hadamard-differentiable at* θ *tangentially to* $\mathbf{D}_o$*. Let* $X_n : \Omega_n \to \mathbf{D}_\phi$ *be maps with* $r_n(X_n - \theta) \Rightarrow X$ *for some sequence of constants* $r_n \to \infty$*, where* X *is separable and takes its values in* $\mathbf{D}_o$*. Then* $r_n(\phi(X_n) - \phi(\theta)) \Rightarrow \phi'_\theta(X)$*.*

Note that the Delta-Method only requires that ϕ be differentiable at the single point θ, a fact that will be crucial for our purposes.

Before moving on to the central limit theorem, we have to make some assumptions that will allow us to apply the Delta-Method to our processes.

Assumption 4.2.4 *For all* $s, t \in T$*,* $P(s \in \xi) - P(s, t \in \xi) \leq K\,|s-t|$*, where* K *is a constant and* $|\cdot|$ *denotes the Euclidean norm.*

Assumption 4.2.5 $P(t \in \xi)$ *is continuous in* $t \in T$ *and there exists* $\epsilon > 0$ *such that* $P(t \in \xi) > \epsilon$ *for every* $t \in T$*.*

Assumption 4.2.6 *The survival function* S *satisfies a Lipschitz condition of order 1.*

Theorem 4.2.7 *Under Assumptions 4.2.4, 4.2.5 and 4.2.6,* $\frac{1}{\sqrt{n}}(Z_n(\cdot) - E(Z_n(\cdot))) \Rightarrow \Delta$ *and* $\frac{1}{\sqrt{n}}(N^{(n)}{}^{\xi}_{A.} - E(N^{(n)}{}^{\xi}_{A.})) \Rightarrow \Gamma$*, where* Δ *and* Γ *are tight Gaussian processes on* $C[0,1]^2$ *and* $D[0,1]^2$ *respectively. Moreover,* $U^{(n)} \Rightarrow G$ *in* $l^\infty[0,\tau]^2$ *for every* τ *such* *that $S(\tau) > 0$, where G is a mean-zero Gaussian process such that for $t \in T$,*

$$G_t = \int_{[0,t]} \frac{d\Gamma(u)}{S(u)P(u \in \xi)} - \int_{[0,t]} \frac{\Delta(u)}{(S(u)P(u \in \xi))^2} dP(Y \in A_u, Y \in \xi).$$

The first term in G_t is defined by integration by parts.

Again, this Theorem is very similar to Example 3.9.19 in [22]. One of the differences, obviously, is that we are working on two dimensions instead of one. Another resides in our censoring mechanism and the information we are provided with; a consequence of this is that Z_n, the survivor function process, will not be in $D[0,1]^2$, the space that plays the role of $D[0,1]$ in our model.

Proof: Let $W_t = I_{\{Y \in E_t\}}$ and $V_t = I_{\{t \in \xi\}}$. Note that Assumption 4.2.4 implies that $E\,|V_t - V_s| \leq C\,|t-s|$, where C is some constant, since

$$E\,|V_t - V_s| = E|I_{\{t\in\xi\}} - I_{\{s\in\xi\}}| = P(t \in \xi, s \in \xi^c) + P(s \in \xi, t \in \xi^c).$$

Then we have that

$$\begin{aligned}
&E|I_{\{Y\in E_t\}}I_{\{t\in\xi\}} - I_{\{Y\in E_s\}}I_{\{s\in\xi\}}|) \\
&= E\left[E(|W_tV_t - W_tV_s|)|W_t, W_s\right] \\
&= E\left[|V_t - V_s|\,|W_t = W_s = 1\right] P(W_t = W_s = 1) \\
&+ E\left[V_t|W_t = 1, W_s = 0\right] P(W_t = 1, W_s = 0) \\
&+ E\left[V_s|W_t = 0, W_s = 1\right] P(W_t = 0, W_s = 1) \\
&\leq C\,|t-s|\,S(s \vee t) + P(t \in \xi)(S(t) - S(s \vee t)) \\
&+ P(s \in \xi)(S(s) - S(s \vee t)) \\
&\leq K^*\,|t-s|,
\end{aligned}$$

where the first inequality follows from the independence of the processes V and W and the observation at the beginning of this proof; the second is a consequence of Assumption 4.2.6. But by a nearly identical argument to that in Example 2.11.14 in

[22] -we use two-dimensional blocks instead of closed intervals- this implies that the sequence of processes defined by

$$\frac{1}{\sqrt{n}}[\sum_{i=1}^{n} I_{\{Y_i \in E_t\}} I_{\{t \in \xi_i\}} - nS(t)P(t \in \xi)] = \frac{1}{\sqrt{n}}[Z_n(t) - E(Z_n(t))]$$

converges in distribution to a tight Gaussian process Δ on $l^\infty[0,1]^2$. In fact, the limiting process is continuous, but we will show this in the Lemma following this proof. Moreover, since the ξ_i's are i.i.d., we also have that the sequence of processes defined by

$$\frac{1}{\sqrt{n}}[\sum_{i=1}^{n} I_{\{Y_i \in A_t \cap \xi_i\}} - nP(Y \in A_t, Y \in \xi)]$$

converges in distribution to another tight Gaussian process Γ on $D[0,1]^2$. This follows simply from the CLT for empirical processes, since we are working with a subdistribution. This means that

$$\frac{1}{\sqrt{n}}[(Z_n(\cdot), N^{(n)\xi}_{A.}) - (E(Z_n(\cdot)), E(N^{(n)\xi}_{A.}))] \Rightarrow (\Delta, \Gamma)$$

on $l^\infty[0,1]^2 \times D[0,1]^2$, where (Δ, Γ) is a Gaussian process on $C[0,1]^2 \times D[0,1]^2$.

Recalling (4.1.3), we note that the estimator depends on the pair $(\frac{Z_n}{n}, \frac{N^{(n)\xi}}{n})$ through the maps

$$(A, B) \longrightarrow (\frac{1}{A}, B) \longrightarrow \int \frac{1}{A} dB. \tag{4.2.3}$$

It is a consequence of Lemma 4.2.2 that the map 4.2.3 is Hadamard-differentiable tangentially to $C[0,1]^2 \times D[0,1]^2$ on a domain of the type $\{(A, B) : \int |dB| \leq M, A \geq \epsilon\}$ for given M and $\epsilon > 0$, at every point (A, B) such that $1/A$ is of bounded variation. Whenever t is restricted to an interval $[0,\tau]^2$ such that $S(\tau) > 0$, the pair $(\frac{Z_n}{n}, \frac{N^{(n)\xi}}{n})$ is contained in this domain with probability tending to 1 for $M \geq 1$ and ϵ sufficiently small. The derivative map is given by

$$(\alpha, \beta) \longmapsto \int (1/A) d\beta - \int (\alpha/A^2) dB.$$

Let $r_n = \sqrt{n}$, $X_n = (\frac{Z_n}{n}, \frac{N^{(n)\xi}}{n})$, $\theta = (S(t)P(t \in \xi), P(Y \in A_t, Y \in \xi))$ and $X = (\Delta, \Gamma)$. Now we can apply the Delta-Method to conclude that

$$U^{(n)}_{\cdot} = \sqrt{n}(\hat{H}^{(n)}{}_{A.} - H_{A.}) \Rightarrow G_{\cdot},$$

where

$$G_t = \int_{[0,t]} \frac{d\Gamma(u)}{S(u)P(u \in \xi)} - \int_{[0,t]} \frac{\Delta(u)}{(S(u)P(u \in \xi))^2} dP(Y \in A_u, Y \in \xi)$$

is again Gaussian. As in Lemma 4.2.2, the first term in the limiting process has to be defined by integration by parts, since Γ may not be of bounded variation.

The covariance structure for G will be given later in this section. □

Following the notation in [11], let $D_n(t) := n^{-1/2}[Z_n(t) - E(Z_n(t))]$ and $C_n(t) := n^{-1/2}[N^{(n)}{}^{\xi}_{A_t} - E(N^{(n)}{}^{\xi}_{A_t})]$.

Lemma 4.2.8 *Under the same conditions as Theorem 4.2.7, the sequence of processes* $\frac{1}{\sqrt{n}}[Z_n(t) - E(Z_n(t))]$ *converges in distribution in* $l^\infty[0,1]^2$ *to a tight Gaussian process* Δ *taking its values on* $C[0,1]^2$.

Proof: The convergence in distribution of $D_n(\cdot)$ to a tight Gaussian process has already been shown in the proof of Theorem 4.2.7. It remains to show that the limiting process is continuous.

From Assumptions 4.2.4, 4.2.5 and 4.2.6, we have that

$$\begin{aligned}
E|D_n(s) - D_n(t)|^2 &= S(s)P(s \in \xi) + S(t)P(t \in \xi) - 2S(s \vee t)P(s,t \in \xi) \\
&\quad -[S(s)P(s \in \xi) - S(t)P(t \in \xi)]^2 \\
&= [S(s)P(s \in \xi) - S(s \vee t)P(s,t \in \xi)] + [S(t)P(t \in \xi) \\
&\quad -S(s \vee t)P(s,t \in \xi)] - [S(s)P(s \in \xi) - S(t)P(t \in \xi)]^2 \\
&= [S(s) - S(s \vee t)]P(s \in \xi) + [S(t) - S(s \vee t)]P(t \in \xi) \\
&\quad +S(s \vee t)[P(s \in \xi) - P(s,t \in \xi)] \\
&\quad +S(s \vee t)[P(t \in \xi) - P(s,t \in \xi)] \\
&\quad -[S(s)P(s \in \xi) - S(t)P(t \in \xi)]^2 \\
&\leq K_1\,|s-t| \\
&\leq \frac{K_1}{(\ln|s-t|)^2},
\end{aligned}$$

whenever $|s-t| < 1$. Then Condition (1.15) of Theorem 1.4 in [2] is satisfied for $\alpha = 1$ and $0 < \eta < 1$, which implies that $\Delta \in C[0,1]^2$ (See discussion in p. 20 of [2] for an extension of that Theorem to multidimensional Euclidean spaces). □

The next proposition and its corollary are identical to their counterparts in [12] (Proposition 4.6 and Corollary 4.7). Again, since the only change is that we are working with $*$-anti-clouds instead of anti-clouds and we don't make use of the new observability requirements, the proofs given in [12] are still valid and hence will not be included.

Proposition 4.2.9 *For $C, D \in \mathcal{C}$,*

$$\begin{aligned}\mathrm{Cov}\left(M_C^{(n)}, M_D^{(n)}\right) &= n\int_{C\cap D} S(t)P(t\in\xi)h(t)\lambda(dt) \\ &+ n\int\int_{l(C,D)} \mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)\lambda(ds)\lambda(dt),\end{aligned} \tag{4.2.4}$$

where

$$\begin{aligned} l(C,D) &= \{(c,d)\in C\times D : c\in A_d^c\cap E_d^c\} \\ &= \{(c,d)\in C\times D : d\in A_c^c\cap E_c^c\}.\end{aligned}$$

Corollary 4.2.10 *Let g be a continuous function on T. Then $\int_\cdot g(t)M^{(n)}(dt)$ is a pseudo-strong martingale and for $C, D\in\mathcal{C}$,*

$$\begin{aligned}&\mathrm{Cov}\left(\int_C g(s)M^{(n)}(ds), \int_D g(t)M^{(n)}(dt)\right) \\ &= n\left[\int_{C\cap D} g^2(t)S(t)P(t\in\xi)h(t)\lambda(dt)\right. \\ &+ \left.\int\int_{l(C,D)} g(s)g(t)\mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)\lambda(ds)\lambda(dt)\right].\end{aligned} \tag{4.2.5}$$

Now that we have Proposition 4.2.9 and Corollary 4.2.10, we are in position to give the covariance structure for the limiting process in Theorem 4.2.7.

Lemma 4.2.11 *The covariance structure for the process G in Theorem 4.2.7 is given, for $C, D \in \mathcal{C}$, by*

$$\begin{aligned}\mathrm{Cov}(G_C, G_D) &= \int_{C\cap D} (S(t)P(t\in\xi))^{-1} h(t)\lambda(dt) \qquad (4.2.6)\\ &+ \int\int_{l(C,D)} \frac{\mu(E_s\cap E_t)P(s,t\in\xi)}{S(s)S(t)P(s\in\xi)P(t\in\xi)} h(s)h(t)\lambda(ds)\lambda(dt).\end{aligned}$$

Proof: For any $A \in \mathcal{A}$,

$$\begin{aligned}\sqrt{n}(\hat{H}_A^{(n)} - H_A) &= \sqrt{n}\int_A \frac{1}{Z_n(t)} M^{(n)}(dt)\\ &= \int_A \frac{n}{Z_n(t)} \cdot \frac{M^{(n)}(dt)}{\sqrt{n}}\\ &= \int_A \left(\frac{n}{Z_n(t)} - \frac{1}{S(t)P(t\in\xi)}\right)\frac{M^{(n)}(dt)}{\sqrt{n}} \qquad (4.2.7)\\ &+ \int_A \frac{1}{S(t)P(t\in\xi)} \cdot \frac{M^{(n)}(dt)}{\sqrt{n}}. \qquad (4.2.8)\end{aligned}$$

We can show that (4.2.7) converges in probability to 0 with exactly the same argument used in the proof of Theorem 5.1 in [11], where it was shown that the term (25) converges to 0 in probability. In fact, in [11] it was assumed the existence of an appropriate function space where the sequence $(D_n(\cdot))$ defined by $D_n(t) := n^{-1/2}(Z_n(t) - E[Z_n(t)])$ converges. In our case though, we don't need to make this assumption since we have already shown the convergence of this sequence under Assumptions 4.2.4 and 4.2.6. Then we have that the only term that contributes to the covariance is (4.2.8), and so by applying Corollary 4.2.10 we get that the covariance structure is given by (4.2.6). □

Remark 4.2.12 In the proof of Lemma 4.2.11 we wrote

$$\begin{aligned}\sqrt{n}(\hat{H}_A^{(n)} - H_A) &= \int_A \left(\frac{n}{Z_n(t)} - \frac{1}{S(t)P(t\in\xi)}\right)\frac{M^{(n)}(dt)}{\sqrt{n}}\\ &+ \int_A \frac{1}{S(t)P(t\in\xi)} \cdot \frac{M^{(n)}(dt)}{\sqrt{n}},\end{aligned}$$

with the first term in the right-hand side converging to 0 in probability. This equality actually gives us the central limit theorem for the finite-dimensional distributions of

$\sqrt{n}(\hat{H}^{(n)} - H)$, since the second term in the right-hand side is the normalized sum of n i.i.d. processes, as noted in [11]. This comes up short of giving us the functional CLT though, because we would still have to prove tightness -and that calculation is not very pleasant at all. It is for this reason that we believe the use of the Delta-Method is a more elegant approach to this particular problem.

To complete the discussion about the Nelson-Aalen estimator, we will establish its asymptotic unbiasedness and its consistency as an estimator of the integrated hazard over a more general class of sets. For the rest of this Section, let $\mathcal{A}$ denote the lower layers in $[0,1]^2$ (See Example 2.1.4). We will need some additional definitions, as well as the following assumption.

Assumption 4.2.13 *For every $A \in \mathcal{A}$, $P(A \subseteq \xi) > 0$. Additionally,*

$$\lim_{\epsilon \to 0} \sup_{t \in T} P(t \in (\partial \xi)^{\epsilon}) = 0. \tag{4.2.9}$$

As it was done in Section 4 of [12], recall Definition 3.1.8 to define

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

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

and note that

$$Z_n^{k-}(t) \leq Z_n(t) \leq Z_n^{k+}(t). \tag{4.2.10}$$

Define D_n^{k+}, D_n^{k-} in an analogous way to D_n, only replacing Z_n by Z_n^{k+}, Z_n^{k-} respectively. These two sequences converge weakly to tight Gaussian processes in $l^{\infty}[0,1]^2$ taking their values on $C[0,1]^2$ by the same arguments as those used in the proof of Lemma 4.2.8.

The proofs for the following uniform law of large numbers, as well as for the Theorems stating the asymptotic unbiasedness and consistency of the Nelson-Aalen estimator, are essentially the same as those for Lemma 4.9 and Theorems 4.8 and 4.10 in [12], and so they will be omitted.

Lemma 4.2.14 *For $A \in \mathcal{A}$,*

$$\sup_{t \in A} \left| \frac{Z_n(t)}{n} - S(t)P(t \in \xi) \right| \to_P 0$$

as $n \to \infty$.

Theorem 4.2.15 *Let $A \in \mathcal{A}$. If $P(A \subseteq \xi) > 0$, then $\hat{H}_A^{(n)}$ is an asymptotically unbiased estimator of H_A.*

Theorem 4.2.16 *Let $A \in \mathcal{A}$. Under Assumptions 4.2.4, 4.2.5, 4.2.6 and 4.2.13, $\hat{H}_A^{(n)}$ is a consistent estimator of H_A.*

4.3 Applications

For all our applications, we will assume that $T = [0, 1]^2$.

Test of Independence

We now concern ourselves with the construction of a test of independence of the components (Y_1, Y_2) of the random vector $Y \in T$. We will follow closely the development of the ideas presented in [18].

As noted in [11] and [18], when Y_1 and Y_2 are independent, the cumulative hazard is the product of the marginal hazards. Therefore, we have to be able to estimate the marginal hazards in order to obtain a test of independence. Assumptions 4.2.4, 4.2.5 and 4.2.6 will hold throughout this section. We will consider three different types of data structure that can arise in a given experiment.

- Model 1: Assume we have a sequence of i.i.d T-valued random variables $(Y_i) = (Y_{1,i}, Y_{2,i})$ with the same distribution as Y, and a sequence (ξ_i^*) of $*$-anti-clouds independent of the $Y_i's$. Let $\mathcal{F}$ be the filtration $\mathcal{F} = \mathcal{F}^Y \vee \mathcal{F}^\xi$, and define the filtrations $\mathcal{F}^{(1)}$, $\mathcal{F}^{(2)}$ by $\mathcal{F}_{t_1}^{(1)} = \bigvee_{t_2} \mathcal{F}_{(t_1,t_2)}$ and $\mathcal{F}_{t_2}^{(2)} = \bigvee_{t_1} \mathcal{F}_{(t_1,t_2)}$ for any $(t_1, t_2) \in T$.

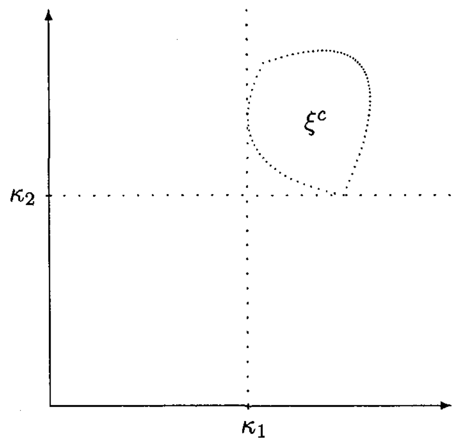

Figure 4.4: The stopping times κ_1 and κ_2

For each i, let

$$\begin{aligned} \kappa_{1,i} &:= \sup\{t_1 : (t_1, 0) \in \xi_i^*\} \\ \kappa_{2,i} &:= \sup\{t_2 : (0, t_2) \in \xi_i^*\}, \end{aligned}$$

see Figure 4.4; note that $\kappa_{1,i}$ is an $\mathcal{F}^{(1)}$-stopping time: for any $t_1 \in [0, 1]$,

$$\{\kappa_{1,i} \le t_1\} = \{t_1 < \kappa_{1,i}\}^c = \{(t_1, 0) \in \xi_i^*\}^c \in \mathcal{F}^*_{(t_1,0)} = \mathcal{F}^{(1)}_{t_1}.$$

Similarly, we can show that $\kappa_{2,i}$ is an $\mathcal{F}^{(2)}$-stopping time; then $\xi_{j,i} := [0, \kappa_{j,i}]$ are $*$-anti-clouds with respect to $\mathcal{F}^{(j)}$ for $j = 1, 2$. Taken separately, the $Y_{j,i}$ are censored on the intervals $\xi^c_{j,i} := (\kappa_{j,i}, 1]$ for $j = 1, 2$, $i = 1, 2, \ldots$.

• Model 2: Recall Example 4.1.8, illustrated in Figure 4.3. We have a sequence of i.i.d. T-valued random variables $(Y_i) = (Y_{1,i}, Y_{2,i})$ with the same distribution as Y on a probability space $(\Omega, \mathcal{F}, P)$. The filtration $\mathcal{F}$ is such that $\mathcal{F}^Y_t \subseteq \mathcal{F}_t$ for every $t = (t_1, t_2) \in [0, 1]^2$, and we let $\mathcal{F}^{(1)}_{t_1} = \bigvee_{t_2} \mathcal{F}_{(t_1,t_2)}$, $\mathcal{F}^{(2)}_{t_2} = \bigvee_{t_1} \mathcal{F}_{(t_1,t_2)}$. We also have a sequence of $\mathcal{F}^{(j)}$-stopping times $0 = \eta^{(i)}_{j1} \le \nu^{(i)}_{j1} \le \eta^{(i)}_{j2} \le \nu^{(i)}_{j2} \le \ldots \le \eta^{(i)}_{jn_j} \le \nu^{(i)}_{jn_j}$, for $i = 1, 2, \ldots$ and $j = 1, 2$. If we let $s = (s_1, s_2)$, $u = (u_1, u_2)$, we can observe

$$\mathcal{H}_t = \{I_{\{Y_{j,i} \le s_j, Y_{j,i} \in \bigcup_{k=1}^{n_j} [\eta^{(i)}_{jk}, \nu^{(i)}_{jk}]\}}, I_{\{Y_{j,i} \in (\nu^{(i)}_{jk}, \rho^{(i)}_{j,k+1}), \rho_{j,k+1} \le u_j\}} : s, u \in A_t; j = 1, 2\},$$

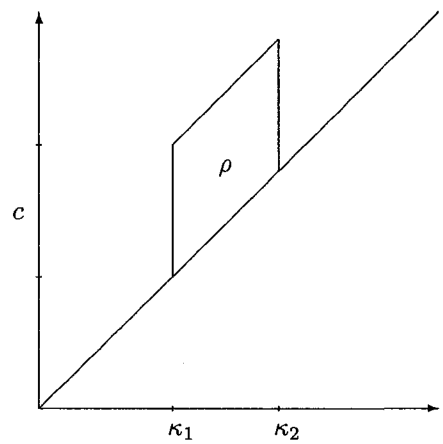

Figure 4.5: Model 3.

where $i = 1, 2, \ldots$ We showed in Example 4.1.8 that if we define the set $\xi_i^* = \{\bigcup_{k=1}^{n_1}[\eta_{1k}^{(i)}, \nu_{1k}^{(i)}] \times \bigcup_{k=1}^{n_2}[\eta_{2k}^{(i)}, \nu_{2k}^{(i)}]\}$, then ξ_i^* is a $*$-anti-cloud with respect to $\mathcal{F}_{(t_1,t_2)}$ for $i = 1, 2, \ldots$.

• Model 3: Suppose now that we have a sequence of i.i.d. T-valued random variables $(Y_i) = (Y_{1,i}, Y_{2,i})$ with the same distribution as Y, as well as a sequence of i.i.d. $[0, 1]$-valued random variables $\kappa_{1,i} < \kappa_{2,i}$ for $i = 1, 2, \ldots$. Define the filtration $\mathcal{F}_{t_1}^{\kappa_1,\kappa_2} = \sigma\{I_{\{\kappa_{1,i} \leq s_1\}}, I_{\{\kappa_{2,i} \leq s_1\}} : s_1 \leq t_1\}$, and let $\mathcal{F}_{(t_1,t_2)} := \mathcal{F}_{(t_1,t_2)}^Y \vee \mathcal{F}_{t_1}^{\kappa_1,\kappa_2}$, as well as $\mathcal{F}_{t_1}^{(1)} = \bigvee_{t_2} \mathcal{F}_{(t_1,t_2)}$, $\mathcal{F}_{t_2}^{(2)} = \bigvee_{t_1} \mathcal{F}_{(t_1,t_2)}$. Let $\rho_i := \{(t_1, t_2) : \kappa_{1,i} < t_1 < \kappa_{2,i},\ t_1 < t_2 < t_1 + c\}$ represent the censored region, where c denotes a constant. Then it's easy to see that ρ_i is a sequence of $*$-clouds, since the $\kappa_{j,i}$ are stopping times. This kind of data structure can arise when, for example, Y_1 is the time of start of pregnancy and Y_2 is the time of onset of tuberculosis. At κ_1, it is found that the test used for diagnosing tuberculosis is dangerous for a pregnant woman, so once the pregnancy starts it becomes impossible to diagnose that disease for the next nine months. At time κ_2 though, a new test becomes available and the tuberculosis times are no longer censored. See Figure 4.5.

In this case, we would censor the pairs $(Y_{1,i}, Y_{2,i})$ on the extended region $\xi_i^{*c} =$

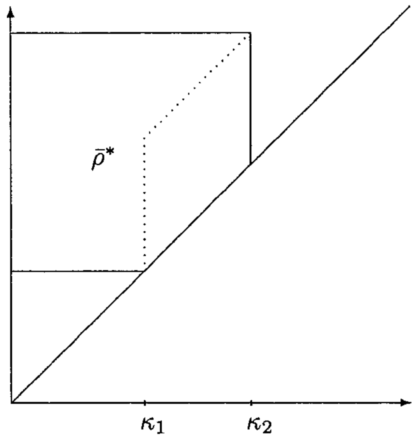

Figure 4.6: The region ρ^*.

$\rho_i^* := \{(t_1, t_2) : \kappa_{1,i} < t_1 < \kappa_{2,i},\ t_1 < t_2 < \kappa_{2,i} + c\} \bigcup \{(t_1, t_2) : t_1 < \kappa_{1,i},\ \kappa_{1,i} < t_2 < \kappa_{2,i} + c\}$ in order to be able to observe the indicators $I_{\{Y \in E_t\}} I_{\{t \in \xi_i^*\}}$ needed to obtain the Nelson-Aalen estimator, as illustrated in Figure 4.6. Note that ρ_i^* is also a sequence of $*$-clouds, since all the endpoints of the intervals in the definition of each ρ_i^* are stopping times with respect to both $\mathcal{F}^1_{t_1}$ and $\mathcal{F}^2_{t_2}$; hence

$$\{(t_1, t_2) \in \rho_i^*\} \in \mathcal{F}^1_{t_1} \otimes \mathcal{F}^2_{t_2} \subseteq \mathcal{F}^1_{t_1} \vee \mathcal{F}^2_{t_2} = \mathcal{F}^*_{(t_1,t_2)}.$$

Taken separately, Y_1 is not censored at all, since it is always possible to determine whether the individual is pregnant or not, but Y_2 is censored on the interval $(\kappa_{1,i}, \kappa_{2,i} + c)$.

In each of these three models, we end up with the same type of one and two-dimensional structure: the pair (Y_1, Y_2) is filtered on an $\mathcal{F}$-$*$-cloud, and each Y_j is filtered by an $\mathcal{F}^j$-$*$-cloud ξ_j^c, a union of random open intervals with endpoints that are $\mathcal{F}^j$-stopping times for $j = 1, 2$. We would like to note that in classical one-dimensional problems, the intervals are usually taken to be left-open and right-closed in order to ensure predictability; but since we are assuming that $P(Y_{\partial\xi} = 0) = 1$, the endpoints of the intervals can actually be ignored. In fact, that assumption can be weakened in models 1 and 2 by using the predictable sets $\bigcup_{k=1}^{n_j} (\eta_{jk}^{(i)}, \nu_{jk}^{(i)}]$ in one dimension and the sets $\bigcup_{k=1}^{n_1}(\eta_{1k}^{(i)}, \nu_{1k}^{(i)}] \times \bigcup_{k=1}^{n_2}(\eta_{2k}^{(i)}, \nu_{2k}^{(i)}]$ in two dimensions, which also have some predictability properties.

Now we are in position to define the analogous one-dimensional processes

$$\begin{aligned}
N_1^{(n)\xi_1}(t_1) &= \sum_{i=1}^{n} I_{\{Y_{1,i}\le t_1\}} I_{\{Y_{1,i}\in\xi_{1,i}\}} \\
N_2^{(n)\xi_2}(t_2) &= \sum_{i=1}^{n} I_{\{Y_{2,i}\le t_2\}} I_{\{Y_{2,i}\in\xi_{2,i}\}} \\
Z_{n,1}(t_1) &= \sum_{i=1}^{n} I_{\{Y_{1,i}\ge t_1\}} I_{\{t_1\in\xi_{1,i}\}} \\
Z_{n,2}(t_2) &= \sum_{i=1}^{n} I_{\{Y_{2,i}\ge t_2\}} I_{\{t_2\in\xi_{2,i}\}},
\end{aligned}$$

as well as the processes $C_{n,j}(t_j) = n^{-1/2}[N_j^{(n)\xi_j}(t_j) - E(N_j^{(n)\xi_j}(t_j))]$ and $D_{n,j}(t_j) = n^{-1/2}[Z_{n,j}(t_j) - E(Z_{n,j}(t_j))]$ for $j = 1, 2$. We will also make use of the one-dimensional analogues of the process $M^{(n)}$:

$$M_j^{(n)}(\cdot) = N_j^{(n)\xi_j}(\cdot) - \int_{\cdot} Z_{n,j}(u_j)h(u_j)\lambda(du_j), \qquad j = 1, 2.$$

The cumulative marginal hazards are estimated by

$$\hat{H}_j^{(n)}(t_j) = \int_{[0,t_j]} \frac{N_j^{(n)\xi_j}(du_j)}{Z_{n,j}(u_j)}, \qquad j = 1, 2;$$

see also Appendix.

Lemma 4.3.1 *$n^{1/2}(\hat{H}^{(n)} - H, \hat{H}_1^{(n)} - H_1, \hat{H}_2^{(n)} - H_2) \Rightarrow (G, G_1, G_2)$ in the space $l^\infty[0,\tau]^2 \times l^\infty[0,\tau_1] \times l^\infty[0,\tau_2]$ for every $\tau = (\tau_1, \tau_2)$ such that $S(\tau) > 0$, where the vector (G, G_1, G_2) is jointly Gaussian with continuous sample paths.*

Proof: We start very similarly to the proof of Lemma 4.1 in [18]. We already saw in the proof of Theorem 4.2.7 that (D_n, C_n) converges to (Δ, Γ) on $l^\infty[0,1]^2 \times D[0,1]^2$. We also know that $(D_{n,j}, C_{n,j})$ converges to a joint Gaussian process $(\Delta_j, \Gamma_j), j = 1, 2$, on $l^\infty[0,1] \times D[0,1]$, since the same arguments for the convergence of (D_n, C_n) still apply. Therefore, the joint sequence $(D_n, D_{n,1}, D_{n,2}, C_n, C_{n,1}, C_{n,2})$ is tight in the product of topologies of the space $l^\infty[0,1]^2 \times (l^\infty[0,1])^2 \times D[0,1]^2 \times (D[0,1])^2$. This, coupled with the fact that its finite-dimensional distributions converge to those of a Gaussian process with continuous sample paths, give the convergence of the sequence.

Next, note that both the estimators for the cumulative marginal hazards depend on the pairs $(\frac{Z_{n,j}}{n}, \frac{N_j^{(n)\xi_j}}{n})$, $j = 1, 2$, through the map (4.2.3), where the integral is defined on $\mathbf{R}$ instead of $\mathbf{R}^2$. Using the same arguments as in Theorem 4.2.7,we can show that $n^{1/2}(\hat{H}_j^{(n)} - H_j) \Rightarrow G_j$ in $D[0, \tau_j]$ for every τ_j such that $S_j(\tau_j) > 0$, where G_j are mean-zero Gaussian processes for $j = 1, 2$. As a consequence of this fact and the joint convergence of $(D_n, D_{n,1}, D_{n,2}, C_n, C_{n,1}, C_{n,2})$, we get the desired result. $\square$

As was the case in [18], in order to make our test of independence between $(Y_{1,i})$ and $(Y_{2,i})$, $i = 1, 2, \ldots$, we will take the difference between the two-dimensional Nelson-Aalen estimator and the estimator under the hypothesis of independence. Define $V_n := n^{1/2}(\hat{H}^{(n)} - \hat{H}_1^{(n)}\hat{H}_2^{(n)})$ on $[0, \tau]$, where $\tau = (\tau_1, \tau_2)$. We could calculate the variance of V_n, but as Pons remarked in [19], its practical estimation is not easy and therefore not very useful. That is why we prefer to use a bootstrap test based on V_n, and our efforts will now turn towards that goal.

We shall use the Delta-Method for the bootstrap, and for that we need to make sure that the classes of functions needed are Donsker classes. Let $X_1, \ldots, X_n$ be a sample of random elements in a measurable space $(\mathcal{X}, \mathcal{A})$ with distribution P, let $\mathcal{G}$ be a collection of measurable functions $g : \mathcal{X} \to \mathbf{R}$ and let P_n be the empirical measure of the X_i, which induces a map from $\mathcal{G}$ to $\mathbf{R}$ defined by $P_n g := n^{-1} \sum_{i=1}^n g(X_i)$. The following definition is taken from [22]:

Definition 4.3.2 *A Donsker class is a class $\mathcal{G}$ for which $G_n := \sqrt{n}(P_n - P) \Rightarrow G$ in $l^\infty(\mathcal{G})$, where the limit G is a tight Borel measurable element in $l^\infty(\mathcal{G})$.*

Ivanoff and Merzbach observed in [12] that an anti-cloud is measurable as a random closed set as a consequence of part 2 of their Corollary 3.10; we still have an identical result in Corollary 3.1.12 for $*$-anti-clouds, so these random sets are also measurable in the same sense. Now assume that we can see the pair (Y_i, ξ_i). Then we can take (Y_i, ξ_i) to be the random elements X_i in the previous definition. Let P be the joint distribution of the pair (Y_i, ξ_i), let P_n denote its empirical measure, $\tilde{P}_n$ the bootstrap empirical distribution and $\tilde{G}_n$ the bootstrap empirical process defined by $\tilde{G}_n := \sqrt{n}(\tilde{P}_n - P_n)$. For a more complete description of these elements, as well as for the exact statement of the Delta-Method for bootstrap in probability -which we will use a few lines ahead- refer to Chapters 3.6 and 3.9 of [22], respectively.

Recalling that $T = [0,1]^2$, let $\mathcal{G}^1 = \{g_t^1 : t \in T\}$ and $\mathcal{G}^2 = \{g_t^2 : t \in T\}$ be defined, respectively, by $g_t^1(Y_i, \xi_i) := I_{\{Y_i \in E_t\}} I_{\{t \in \xi_i\}}$ and $g_t^2(Y_i, \xi_i) := I_{\{Y_i \in A_t\}} I_{\{Y_i \in \xi_i\}}$. Then both $\mathcal{G}^1$ and $\mathcal{G}^2$ are Donsker classes, since both were shown to be convergent in distribution to tight Gaussian limits on $l^\infty[0,1]^2$ in the proof of Theorem 4.2.7. We can apply Theorem 2.10.6 in [22] to find that the pair $(\mathcal{G}^1, \mathcal{G}^2)$ is also a Donsker class, since each one is uniformly bounded. In fact, we could have concluded so without using that result, since we showed in the proof of Theorem 4.2.7 that $\frac{1}{\sqrt{n}}[(Z_n(\cdot), N^{(n)}{}^{\xi}_{A.}) - (E(Z_n(\cdot)), E(N^{(n)}{}^{\xi}_{A.}))] \Rightarrow (\Delta, \Gamma)$ on $l^\infty[0,1]^2 \times D[0,1]^2$, where (Δ, Γ) is a Gaussian process on $C[0,1]^2 \times D[0,1]^2$, which in our current notation means that $\tilde{G}_n$ converges in distribution when it is indexed by $(\mathcal{G}^1, \mathcal{G}^2)$. It is also clear that $\mathcal{G}^1$ and $\mathcal{G}^2$ have finite envelope functions -that is, for each $\mathcal{G}^j$, $j = 1, 2$, there exists a function that bounds $g(x)$ for every x and every $g \in \mathcal{G}^j$, $j = 1, 2$-, because both classes are comprised exclusively of indicator functions. Hence by Theorem 3.6.1 in [22], the conditions (3.9.9) of [22] are met. Furthermore, we have shown that the map (4.2.3) is Hadamard-differentiable tangentially to $C[0,1]^2 \times D[0,1]^2$ on a certain domain. Then we only need to apply Theorem 3.9.11 in [22], the Delta-Method for bootstrap in probability, to get the conditional convergence given (Y_i, ξ_i) of the bootstrapped Nelson-Aalen estimator. Similar arguments lead to the same conclusion for the one-dimensional bootstrapped Nelson-Aalen estimators.

The above is summarized in the next Lemma, which is an analogue of Lemma 4.3.1 for the bootstrapped estimators.

Lemma 4.3.3 *Assume we can observe the pairs* (Y_i, ξ_i) *for* $i = 1, 2, \ldots,$ *and let* $(\tilde{Y}_i, \tilde{\xi}_i)$ *denote the bootstrap sample. Then if* $\tilde{H}^{(n)}, \tilde{H}_1^{(n)}, \tilde{H}_2^{(n)}$ *represent the bootstrapped versions of the joint and marginal Nelson-Aalen estimators, we have that* $n^{1/2}(\tilde{H}^{(n)} - \hat{H}^{(n)}, \tilde{H}_1^{(n)} - \hat{H}_1^{(n)}, \tilde{H}_2^{(n)} - \hat{H}_2^{(n)})$ *converges conditionally given* (Y_i, ξ_i) *to the Gaussian process* (G, G_1, G_2) *of Lemma 4.3.1.*

Now that we have seen that the bootstrapping is working correctly under the assumption of complete observability of the pairs (Y_i, ξ_i), we note that the bootstrapped version of the statistics only involve the product of indicator functions $q_i = I_{\{Y_i \in A_t\}} I_{\{Y_i \in \xi_i\}}$ and $r_i = I_{\{Y_i \in E_t\}} I_{\{t \in \xi_i\}}$. However, we can drop the assumption of complete observability, since those indicators can still be constructed because the events involved are observable under the usual data structure. Therefore, we can in fact bootstrap directly from the observed values q_i, r_i and come up with the same estimators.

We can deduce from Lemma 4.3.1 that

$$(n^{1/2}[\hat{H}^{(n)}(t_1, t_2) - \hat{H}_1^{(n)}(t_1)\hat{H}_2^{(n)}(t_2) - (H(t_1, t_2) - H_1(t_1)H_2(t_2))])$$

converges in distribution to a mean-zero Gaussian process; indeed, following the proof of Proposition 4.2 in [18], we can rewrite $V_n = n^{1/2}(\hat{H}^{(n)} - \hat{H}_1^{(n)}\hat{H}_2^{(n)})$ as

$$\begin{aligned} V_n &= n^{1/2}(\hat{H}^{(n)} - H) - n^{1/2}(\hat{H}_1^{(n)} - H_1)\hat{H}_2^{(n)} - n^{1/2}(\hat{H}_2^{(n)} - H_2)H_1 \\ &+ n^{1/2}(H - H_1 H_2), \end{aligned}$$

hence

$$\begin{aligned} n^{1/2}\left(\hat{H}^{(n)} - \hat{H}_1^{(n)}\hat{H}_2^{(n)} - (H - H_1 H_2)\right) &= n^{1/2}(\hat{H}^{(n)} - H) \\ &- n^{1/2}(\hat{H}_1^{(n)} - H_1)(\hat{H}_2^{(n)} - H_2) \\ &- n^{1/2}(\hat{H}_1^{(n)} - H_1)H_2 \\ &- n^{1/2}(\hat{H}_2^{(n)} - H_2)H_1, \end{aligned}$$

which converges weakly to $G - H_2G_1 - H_1G_2$. If we can completely observe (Y_i, ξ_i), we can construct $\tilde{H}^{(n)}, \tilde{H}_1^{(n)}, \tilde{H}_2^{(n)}$ and Lemma 4.3.3 would guarantee us that

$$\sup\left|n^{1/2}[\tilde{H}^{(n)}(t_1,t_2) - \tilde{H}_1^{(n)}(t_1)\tilde{H}_2^{(n)}(t_2) - (\hat{H}^{(n)}(t_1,t_2) - \hat{H}_1^{(n)}(t_1)\hat{H}_2^{(n)}(t_2))]\right| \quad (4.3.1)$$

converges in distribution to the sup of the absolute value of that same Gaussian process. Then we can bootstrap from (q_i, r_i) jointly and find c_α so that $(1-\alpha)100\%$ of the absolute values of (4.3.1) fall below c_α, which would give us $(1-\alpha)100\%$ uniform confidence bands for the difference $H - H_1H_2$.

Finally, for our test of independence $H_\circ : H = H_1H_2$ vs. $H_\circ : H \neq H_1H_2$, note that under the null hypothesis, $P\left(\sup|V_n(t_1,t_2)| < c_\alpha\right) \approx 1-\alpha$. Therefore, we reject $H_\circ$ if $\sup n^{1/2}\left|\hat{H}^{(n)}(t_1,t_2) - \hat{H}_1^{(n)}(t_1)\hat{H}_2^{(n)}(t_2)\right| > c_\alpha$.

Test of Hazard Rate Order

Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be independent random samples from the distributions F and G respectively, and suppose that there is a common censoring mechanism in the form of a sequence of independent $*$-clouds. We are interested in testing whether the distributions are equal on a particular set or, more generally, on a suitable class of sets, against the alternative that there is a difference in the hazard rates.

We start with the case where we look at the hazards on a fixed set A. The null hypothesis is $H_\circ : F = G$, and we can test it either against a single-sided alternative such as $H_A^F < H_A^G$, or the double-sided alternative $H_A^F \neq H_A^G$. Tests using a one-sided alternative leads us to tests of hazard rate order such as $H_\circ : F = G$ vs. $H_1 : h^F < h^G$ on A, since this alternative would imply $H_A^F < H_A^G$. It is natural to consider the difference between the Nelson-Aalen estimators $\hat{H}^F$ and $\hat{H}^G$ of the integrated hazards, so we would like to use the test statistic

$$W_A^{(N)} = \sqrt{\frac{nm}{N}}\left(\hat{H}_A^{(n)F} - \hat{H}_A^{(m)G}\right),$$

where $N = n + m$.

We still have that both $U_A^{(n)F} = \sqrt{n}(\hat{H}_A^{(n)F} - H_A^F)$ and $U_A^{(m)G} = \sqrt{m}(\hat{H}_A^{(m)G} - H_A^G)$ converge to independent mean-zero Gaussian limits U_A^F and U_A^G respectively. This can be shown by applying the map ϕ in Theorem 4.2.2 to the pair $(\frac{Z_n}{n}, \frac{N_A^{(n)\xi}}{n})$ and repeating the calculations found in the proof of Theorem 4.2.7. We can write

$$W_A^{(N)} = \sqrt{\frac{m}{N}} U_A^{(n)F} - \sqrt{\frac{n}{N}} U_A^{(m)G} + \sqrt{\frac{nm}{N}} \left(H_A^F - H_A^G\right).$$

Under the null hypothesis of equality of the cumulative hazards on the set A, and assuming $\frac{n}{n+m} \to \lambda$ as $n, m \to \infty$, we get that

$$W_A^{(n)} \Rightarrow \sqrt{1-\lambda} U_A^F - \sqrt{\lambda} U_A^G.$$

The limit variable has the same distribution as U_A^F under the null hypothesis. Then we have a test of asymptotic level α for a two-sided alternative if we reject the null hypothesis whenever $\left|W_A^{(N)}\right| > w^{(N)}$, where we choose $w^{(N)}$ such that $w^{(N)} \to w^F = \inf\left\{t : P(\left|U_A^F\right| > t) \le \alpha\right\}$. A test with a one-sided alternative would be handled similarly; we just have to remove the absolute values.

We can also consider testing for the hazard rate over a whole class, as long as it is Donsker in order to preserve the Gaussian limits. An example would be to take the class of rectangles $\mathcal{F} = \{A_z = [0, z] : z \in [0, 1]^2\}$. All the previous remarks about $W^{(N)}$ are still valid, but now we have to take a look at $\sup_z \left|W_{A_z}^{(N)}\right|$ and $\sup_z W_{A_z}^{(N)}$ as the test statistics, as well as choosing $w^{(N)} \to w^F = \inf\left\{t : P(\sup_z \left|U_{A_z}^F\right| > t) \le \alpha\right\}$ and $w^{(N)} \to w^F = \inf\left\{t : P(\sup_z U_{A_z}^F > t) \le \alpha\right\}$ for two-sided and one-sided alternatives, respectively.

Assuming we can fully observe the pairs (X_i, ξ_i), (Y_j, ξ_j') for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, we determine the appropriate value for $w^{(N)}$ by bootstrapping from the pooled sample $S^{(N)} = (X_1, \xi_1), \ldots, (X_n, \xi_n), (Y_1, \xi_1'), \ldots, (Y_m, \xi_m')$, where we choose any of these observations with probability $1/N$, as was done in Section 3.7.2 in [22]. Define $J = \lambda F + (1-\lambda)G$, and note that under the null hypothesis, $H^J = H^F$. We assign the first n elements from the resampling to F, and the rest to G. Let $\ddot{H}^{(n,N)F}$ and $\ddot{H}^{(m,N)G}$ be the Nelson-Aalen estimators for the pooled sample assigned to F and G respectively, and $\ddot{H}^{(N)}$ the estimator for the complete pooled sample. Set

$$\ddot{W}^{(N)} = \sqrt{\frac{nm}{N}} \left(\ddot{H}^{(n,N)F} - \ddot{H}^{(m,N)G} \right).$$

But note that by Hadamard-differentiability and Theorem 3.7.6 in [22], we know that $\sqrt{n}\left(\ddot{H}^{(n,N)F} - \ddot{H}^{(N)}\right) \Rightarrow U_1^J$ and $\sqrt{m}\left(\ddot{H}^{(m,N)G} - \ddot{H}^{(N)}\right) \Rightarrow U_2^J$, where U_1^J and U_2^J are independent mean-zero Gaussian processes with covariance as defined in Lemma 4.2.11, with S equal to the survival function of J and $h = J'/(\text{survival of } J)$. Then if we let $\lambda_N = n/N$,

$$\ddot{W}^{(N)} = \sqrt{n(1-\lambda_N)} \left(\ddot{H}^{(n,N)F} - \ddot{H}^{(N)} \right) - \sqrt{m\lambda_N} \left(\ddot{H}^{(m,N)G} - \ddot{H}^{(N)} \right),$$

which converges in distribution to $\sqrt{1-\lambda}U_1^J - \sqrt{\lambda}U_2^J$, also a Gaussian process equal in distribution to U_1^J. Then, as remarked in Section 3.7.2 of [22], we can use

$$\ddot{w}_N = \inf\left\{ t : P(\ddot{W}^{(N)} > t) \leq \alpha \right\}$$

as critical values for a one-sided test, and

$$\ddot{w}_N = \inf\left\{ t : P(|\ddot{W}^{(N)}| > t) \leq \alpha \right\}$$

for the two-sided test.

Once again, as we did following the statement of Lemma 4.3.3, we can see that the bootstrapped version of our test statistic only uses the functions $q_i = I_{\{X_i \in A_t\}} I_{\{X_i \in \xi_i\}}$, $r_i = I_{\{X_i \in E_t\}} I_{\{t \in \xi_i\}}$, $q'_j = I_{\{Y_j \in A_t\}} I_{\{Y_j \in \xi'_j\}}$ and $r'_j = I_{\{Y_j \in E_t\}} I_{\{t \in \xi'_j\}}$; then we can safely drop the assumption of complete observability, given that even without it we can still construct the appropriate estimators. Now we can say that the test can be performed by bootstrapping directly from the recorded values q_i, r_i, q'_j, r'_j and following the procedure described above.

The Hazard of a Copula

Suppose we have a sequence (X_i, Y_i) of i.i.d. bivariate random vectors with continuous distribution J and continuous, strictly increasing marginals F and G. Let $\mathcal{F}^\circ_{(s,t)} = \sigma\{(X_i, Y_i) : X_i \le s, Y_i \le t\}$, and let $\mathcal{F}$ be a filtration such that $\mathcal{F}^\circ_{(s,t)} \subseteq \mathcal{F}_{(s,t)}$ for every $(s,t) \in [0,1]^2$. Moreover, define $\mathcal{F}^{(1)}_s = \bigvee_t \mathcal{F}_{(s,t)}$ and $\mathcal{F}^{(2)}_t = \bigvee_s \mathcal{F}_{(s,t)}$. As usual, each pair (X_i, Y_i) will be censored by an $\mathcal{F}$-$*$-cloud ξ^c_i. Let $C(p,q) = J(F^{-1}(p), G^{-1}(q))$, the copula function for J, and let $\overline{C}$ be the survival function for the copula: $\overline{C}(p,q) = 1 - p - q + C(p,q)$. A quick calculation for the hazard of the copula, denoted $h^c(p,q)$, shows that

$$\begin{aligned} h^c(p,q) &= \frac{\frac{\partial^2}{\partial p \partial q} C(p,q)}{\overline{C}(p,q)} \\ &= \frac{\frac{\partial^2}{\partial p \partial q} J(F^{-1}(p), G^{-1}(q))}{1 - p - q + C(p,q)} \\ &= \frac{j(F^{-1}(p), G^{-1}(q))}{(1 - p - q + C(p,q)) f(F^{-1}(p)) g(G^{-1}(q))}, \end{aligned}$$

where f, g and j are the densities of F, G and J, respectively. Then, integrating the hazard and using the change of variables $s = F^{-1}(u)$, $t = G^{-1}(v)$, we get

$$\begin{aligned} H^c(p,q) &= \int_0^q \int_0^p h^c(u,v)\,du\,dv \\ &= \int_0^q \int_0^p \frac{j(F^{-1}(u), G^{-1}(v))}{(1 - u - v + C(u,v)) f(F^{-1}(u)) g(G^{-1}(v))} du\,dv \\ &= \int_0^{G^{-1}(q)} \int_0^{F^{-1}(p)} \frac{j(s,t)}{1 - F(s) - G(t) + J(s,t)} ds\,dt \\ &= H(F^{-1}(p), G^{-1}(q)). \end{aligned} \tag{4.3.2}$$

In other words, calculating the cumulative hazard of the copula at (p,q) is the same as calculating the cumulative hazard of the original distribution at the point $(F^{-1}(p), G^{-1}(q))$. We would like to estimate the cumulative hazard of the copula, and the equation above seems to suggest that we could do so by looking at the Nelson-Aalen estimator for the distribution J and replacing $(F^{-1}(p), G^{-1}(q))$ with the quantiles of the respective Kaplan-Meier estimates of the marginals F and G (See Section IV.3.1 in [1]).

Now, suppose for a moment that F and G are known. Define the pseudo-observations $(X_i^*, Y_i^*) := (F(X_i), G(Y_i))$, so that the distribution for (X_i^*, Y_i^*) is the copula C. The observable region then becomes $\xi^* = \{(F(X_i), G(Y_i)) : (X_i, Y_i) \in \xi\}$, and the new filtration is $\mathcal{F}^*(p,q) = \mathcal{F}(F^{-1}(p), G^{-1}(q))$. Using Equation (4.1.3), we have

$$\hat{H}^{(n)c}(p,q) = \sum_{\{i:(X_i^*,Y_i^*)\le(p,q),(X_i^*,Y_i^*)\in\xi_i^*\}} (Z_n^*(X_i^*,Y_i^*))^{-1}. \tag{4.3.3}$$

But note that

$$\begin{aligned} Z_n^*(u,v) &= \sum_{i=1}^n I_{\{(X_i^*,Y_i^*)\in E_{u,v}\}} I_{\{(u,v)\in\xi_i^*\}} \\ &= \sum_{i=1}^n I_{\{(X_i,Y_i)\in E_{(F^{-1}(u),G^{-1}(v))}\}} I_{\{(F^{-1}(u),G^{-1}(v))\in\xi_i\}} \\ &= Z_n(F^{-1}(u), G^{-1}(v)), \end{aligned}$$

so

$$\begin{aligned} (4.3.3) &= \sum_{\{i:(X_i,Y_i)\le(F^{-1}(p),G^{-1}(q)),(X_i,Y_i)\in\xi_i\}} (Z_n(X_i,Y_i))^{-1} \\ &= \hat{H}^{(n)}(F^{-1}(p), G^{-1}(q)), \end{aligned}$$

and hence we have an empirical analogue of (4.3.2).

As we can see from the previous equation, the next step is to estimate the quantiles F^{-1} and G^{-1}. We will do so by taking the appropriate quantiles of the Kaplan-Meier estimators $\hat{F}$, $\hat{G}$ of F and G respectively (See Appendix); then our estimator will be $\hat{H}^{(n)}(\hat{F}^{-1}, \hat{G}^{-1})$, where $\hat{F}^{-1}(p) = \inf\{s : \hat{F}(s) \ge p\}$ and $\hat{G}^{-1}(q)$ is defined in a similar manner. It is interesting to note that $A_{(\hat{F}^{-1}(p),\hat{G}^{-1}(q))}$ is still a $*$-anti-cloud: indeed, since $\hat{F}(s)$ and $\hat{G}(t)$ are adapted to $\mathcal{F}^1(s)$ and $\mathcal{F}^2(t)$ respectively,

$$\begin{aligned} (s,t) \in A_{(\hat{F}^{-1}(p),\hat{G}^{-1}(q))} &= \left\{s \le \hat{F}^{-1}(p)\right\} \cap \left\{t \le \hat{G}^{-1}(q)\right\} \\ &= \left\{\hat{F}(u) < p,\ \forall\, u < s\right\} \cap \left\{\hat{G}(v) < q,\ \forall\, v < t\right\} \\ &\in \mathcal{F}_s^1 \vee \mathcal{F}_t^2 \subseteq \mathcal{F}_{(s,t)}^*. \end{aligned}$$

An immediate consequence of this is that our estimator still maintains a pseudo-strong martingale structure. Moreover, under Assumptions 4.2.4, 4.2.5 and 4.2.6, the composition map

$$\begin{aligned}(Z_n, N^{(n)\xi}, Z_{n,1}, N_1^{(n)\xi_1}, Z_{n,2}, N_2^{(n)\xi_2}) &\xrightarrow{\phi_1} (Z_n, N^{(n)\xi}, \hat{H}_1^{(n)}, \hat{H}_2^{(n)})\\ &\xrightarrow{\phi_2} (Z_n, N^{(n)\xi}, \hat{F}, \hat{G})\\ &\xrightarrow{\phi_3} (\hat{H}^{(n)}, \hat{F}, \hat{G})\\ &\xrightarrow{\phi_4} (\hat{H}^{(n)}, \hat{F}^{-1}, \hat{G}^{-1})\\ &\xrightarrow{\phi_5} \hat{H}^{(n)} \circ (\hat{F}^{-1}, \hat{G}^{-1}) \qquad (4.3.4)\end{aligned}$$

is Hadamard-differentiable tangentially to $C[0,1]^2 \times BV_M[0,1]^2 \times (C[0,1] \times BV_M[0,1])^2$ on a domain analogous to map (4.2.3): the maps ϕ_1 and ϕ_3 are Hadamard-differentiable as a consequence of Theorem 4.2.7, ϕ_2's Hadamard differentiability comes from Lemma 3.9.30 in [22], and ϕ_4 and ϕ_5 are Hadamard-differentiable by Example 3.9.24 and Lemma 3.9.25, respectively, in [22] again. This means that the process $\sqrt{n}(\hat{H}^{(n)c}_{\cdot} - H^c_{\cdot})$ converges in $l^\infty[0,1]^2$ to a Gaussian limit for every (τ_1, τ_2) such that $S(F^{-1}(\tau_1), G^{-1}(\tau_2)) = \overline{C}(\tau_1, \tau_2) > 0$.

An example: The FGM copula

We can apply the previous results to Farlie-Gumbel-Morgenstern (FGM) copulas. Suppose C_{J_1} and C_{J_2} are two FGM copulas with parameters θ_{J_1} and θ_{J_2}, respectively, and we are interested in testing $H_\circ : C_{J_1} = C_{J_2}$ vs. $H_1 : C_{J_1} < C_{J_2}$. For FGM copulas, this is equivalent to testing $H_\circ' : \theta_{J_1} = \theta_{J_2}$ vs. $H_1' : \theta_{J_1} < \theta_{J_2}$. This test can be performed by 'translating' these conditions to those of a test of hazard rate order: we look for a specific set A under which $\theta_{J_1} < \theta_{J_2}$ is equivalent to $h^{C_{J_1}} < h^{C_{J_2}}$, and then we can apply the results of the previous sections.

The FGM copula is defined by $C_\theta(u,v) = uv + \theta uv(1-u)(1-v)$. Straightforward calculations yield $\overline{C}_\theta(u,v) = (1-u)(1-v) + \theta uv(1-u)(1-v)$, as well as

$\frac{\partial^2 C}{\partial u \partial v} = 1 + \theta(1-2u)(1-2v)$. Then $h^{C_{J_1}} < h^{C_{J_2}}$ becomes

$$\frac{1+\theta_{J_1}(1-2u)(1-2v)}{(1-u)(1-v)+\theta_{J_1}uv(1-u)(1-v)} < \frac{1+\theta_{J_2}(1-2u)(1-2v)}{(1-u)(1-v)+\theta_{J_2}uv(1-u)(1-v)}.$$

Multiplying both sides by $(1-u)(1-v)$ gives us

$$\frac{1+\theta_{J_1}(1-2u)(1-2v)}{1+\theta_{J_1}uv} < \frac{1+\theta_{J_2}(1-2u)(1-2v)}{1+\theta_{J_2}uv},$$

which in turn is equivalent to

$$(\theta_{J_2}-\theta_{J_1})uv < (\theta_{J_2}-\theta_{J_1})(1-2u)(1-2v)$$

after cross-multiplying and rearranging terms. Under the alternative hypothesis $\theta_{J_1} < \theta_{J_2}$, the previous inequality becomes

$$uv < (1-2u)(1-2v),$$

and finally

$$1-2u-2v+3uv > 0.$$

This means that, under $H_1' : \theta_{J_1} < \theta_{J_2}$, we have $h^{C_{J_2}}(u,v) - h^{C_{J_1}}(u,v) > 0$ for every (u,v) lying in the set $A = \{(u,v) : 1-2u-2v+3uv > 0\}$. Let $N = n+m$. Since $\int_A (h^{C_{J_2}} - h^{C_{J_1}})d\lambda = H_A^{C_{J_2}} - H_A^{C_{J_1}}$, according to our previous results the appropriate test statistic is $W_A^{(N)} = \sqrt{\frac{nm}{N}}(\hat{H}_A^{(m)C_{J_2}} - \hat{H}_A^{(n)C_{J_1}})$, which would get large on the set A if $\theta_{J_2} > \theta_{J_1}$. The Hadamard-differentiability of $\hat{H}_A^{(n)C_J}$ follows from an application of the composition map (4.3.4) using $N_A^{(n)\xi}$ in this case.

We need to consider two different cases for calculating the critical values for our test. If, from the standpoint of the experiment being performed, it is reasonable to think the marginals from the two distributions are similar, then we will assume they are equal and so we proceed with the bootstrapping as indicated in the Test of Hazard Rate Order subsection. Under $H_\circ$, we know that $W_A^{(N)}$ converges to a mean-zero Gaussian limit $U_A^{C_{J_1}}$, and our test of level α would reject $H_\circ$ whenever $W_A^{(N)} = \sqrt{\frac{nm}{N}}(\hat{H}_A^{(m)C_{J_2}} - \hat{H}_A^{(n)C_{J_1}}) > \ddot{w}_N$, where $\ddot{w}_N = \inf\left\{t : P(\ddot{W}^{(N)} > t) \le \alpha\right\}$. On the other hand, it may well be that we can't possibly make the assumption that the marginals are equal in a given experiment. Then we can no longer pool the samples and treat them as one big sample as we did before.

If this is the case, then we will start by bootstrapping each sample separately. Let $\ddot{H}_A^{(n)C_{J_1}}$ and $\ddot{H}_A^{(m)C_{J_2}}$ denote the bootstrapped Nelson-Aalen estimators. Now let

$$\ddot{W}_A^{(n,m)} = \sqrt{\frac{nm}{N}} \left(\ddot{H}_A^{(m)C_{J_2}} - \ddot{H}_A^{(n)C_{J_1}} - (\hat{H}_A^{(m)C_{J_2}} - \hat{H}_A^{(n)C_{J_1}}) \right),$$

which can be rewritten as

$$\sqrt{\frac{n}{N}} \left(\sqrt{m}(\ddot{H}_A^{(m)C_{J_2}} - \hat{H}_A^{(m)C_{J_2}}) \right) - \sqrt{\frac{m}{N}} \left(\sqrt{n}(\ddot{H}_A^{(n)C_{J_1}} - \hat{H}_A^{(n)C_{J_1}}) \right).$$

Assuming that $\frac{n}{n+m} \longrightarrow \lambda$ as $n, m \longrightarrow \infty$, the last expression converges conditionally given the original samples to $\sqrt{\lambda} U_A^{C_{J_2}} - \sqrt{1-\lambda} U_A^{C_{J_1}}$ by Lemma 4.3.3, where $U_A^{C_{J_1}}$ and $U_A^{C_{J_2}}$ are the independent mean-zero Gaussian limits of $\sqrt{n}(\hat{H}_A^{(n)C_{J_1}} - H_A^{C_{J_1}})$ and $\sqrt{m}(\hat{H}_A^{(m)C_{J_2}} - H_A^{C_{J_2}})$ respectively. Under the null hypothesis, this limit has the same distribution as $U_A^{C_{J_1}}$. Then we have a test of asymptotic level α if we reject the null hypothesis whenever $W_A^{(N)} > \ddot{w}_{n,m}$, where $\ddot{w}_{n,m} = \inf \left\{ t : P(\ddot{W}_A^{(n,m)} > t) \leq \alpha \right\}$.

Chapter 5

Estimating the Distribution Function

As in Chapter 4, let Y be a T-valued random variable with distribution $\mu(B) = P\{Y \in B\}$. Consider a sequence of i.i.d. T-valued random variables (Y_i) with the same distribution as Y, as well as an i.i.d. sequence of anti-clouds (ξ_i) independent of the (Y_i). We will turn away from the martingale-based approach of the past Chapters; this time we take inspiration in the main idea behind [3]. The motivation for exploring this new angle comes from the fact that, in order to avoid observability problems in the construction of the martingale estimators, we had to omit many fully observed data points when we transitioned from anti-clouds to $*$-anti-clouds. The advantage of the approach in this Chapter is that it allows us to consider such points for our estimators.

5.1 Observable Censoring

The objective of this Section is to construct an estimator for the distribution function μ under the assumption that the clouds are completely observed. This condition would be met, for example, when we deal with aerial photography -as it was mentioned in Section 6 of [11]-; there we can actually see the shape of the obscured region.

Another example would be the medical scenario of Chapter 4, where we had times of start of pregnancy and times of onset of tuberculosis. In that case, we knew exactly what the censoring region would be, as it was dictated by the length of a pregnancy.

Let $\tilde{\mu}(A)$ denote the subdistribution function of the uncensored observations, defined by

$$\tilde{\mu}(A) = P(Y \in A \cap \xi) = \int_A P(t \in \xi)\mu(dt).$$

From this equality we deduce that $\tilde{\mu}(dt) = P(t \in \xi)\mu(dt)$, and so we can write

$$\mu(A) = \int_A \frac{1}{P(t \in \xi)}\tilde{\mu}(dt). \tag{5.1.1}$$

This generalizes the equation used by Burke as the basis for his estimators of the distribution function in [3]. We will use the same approach, although we estimate the necessary quantities in a different way.

Equation (5.1.1) suggests that in order to estimate μ, we need to come up with estimators for $P(t \in \xi)$ and $\tilde{\mu}(A)$. Let $\hat{\tilde{\mu}}^{(n)}(A) = \frac{1}{n}\sum_{i=1}^n I_{\{Y_i \in A \cap \xi_i\}}$ and $\hat{P}_n(t \in \xi) = \frac{1}{n}\sum_{i=1}^n I_{\{t \in \xi_i\}}$ be those estimators, and notice that $\hat{\tilde{\mu}}^{(n)}(A) = \frac{1}{n}N_A^{(n)\xi}$. It will be convenient in what follows to use the notation $K_n(t) = n\hat{P}_n(t \in \xi)$. With these estimators at hand, we are in position to define our estimator for the distribution function as follows:

$$\hat{\mu}^{(n)}(A) = \int_A \frac{1}{\hat{P}_n(t \in \xi)}\hat{\tilde{\mu}}^{(n)}(dt) = \frac{1}{n}\sum_{\{i: Y_i \in A \cap \xi_i\}} (\hat{P}_n(Y_i \in \xi))^{-1}. \tag{5.1.2}$$

In order to consider the asymptotic behavior of $\hat{\mu}^{(n)}$, we will need the following Lemma, and for the remainder of this Section we assume $T = [0,1]^2$.

Lemma 5.1.1 *Under Assumptions 4.2.4 and 4.2.5, the sequence of processes* $\frac{1}{\sqrt{n}}[K_n(t) - E(K_n(t))]$ *converges in distribution in* $l^\infty[0,1]^2$ *to a tight Gaussian process* Λ *taking its values on* $C[0,1]^2$.

Proof: This Lemma is the analogue of Lemma 4.2.8 for the process K_n, so the proofs are essentially the same, except that in this case the survival function is not involved and so the argument becomes simpler. We showed in Theorem 4.2.7 that Assumption 4.2.4 implies $E|I_{\{t\in\xi\}} - I_{\{s\in\xi\}}| \leq C\,|t-s|$, where C is some constant, and we saw that this particular condition implies the convergence in distribution of the process $\frac{1}{\sqrt{n}}[K_n(t) - E(K_n(t))]$ to a tight Gaussian process Λ on $l^\infty[0,1]^2$. It remains to show that the limiting process is continuous.

Let

$$B_n(t) := \frac{1}{\sqrt{n}}[K_n(t) - E(K_n(t))]. \tag{5.1.3}$$

We have, for a constant C_1, that

$$\begin{aligned}
E|B_n(s) - B_n(t)|^2 &= P(s\in\xi) + P(t\in\xi) - 2P(s,t\in\xi) \\
&\quad -[P(s\in\xi) - P(t\in\xi)]^2 \\
&= [P(s\in\xi) - P(s,t\in\xi)] + [P(t\in\xi) - P(s,t\in\xi)] \\
&\quad -[P(s\in\xi) - P(t\in\xi)]^2 \\
&\leq [P(s\in\xi) - P(s,t\in\xi)] + [P(t\in\xi) - P(s,t\in\xi)] \\
&\leq C_1\,|s-t| \\
&\leq \frac{C_1}{(\ln|s-t|)^2},
\end{aligned}$$

whenever $|s-t| < 1$. Then Condition (1.15) of Theorem 1.4 in [2] is satisfied for $\alpha = 1$ and $0 < \eta < 1$, which implies that $\Lambda \in C[0,1]^2$. (As we mentioned in the proof of Lemma 4.2.8, see the discussion on p. 20 of [2] for an extension of Theorem 1.4 to multidimensional Euclidean spaces) □

The estimator $\hat{\mu}^{(n)}$ is monotone nondecreasing in the usual partial order, as we can clearly deduce from the right-hand term in (5.1.2). We can show, in a very similar way to the proof of Theorem 4.10 in [12], that it is a consistent estimator of the distribution function under certain conditions.

Recalling Definition 3.1.8, define

$$K_n^{k+}(t) = \sum_{i=1}^{n} I_{\{t\in\xi_{i,k}^+\}},$$

$$K_n^{k-}(t) = \sum_{i=1}^{n} I_{\{t \in \xi_{i,k}^{-}\}},$$

and notice that

$$K_n^{k-}(t) \leq K_n(t) \leq K_n^{k+}(t). \tag{5.1.4}$$

Define B_n^{k+}, B_n^{k-} in an analogous way to (5.1.3), only replacing K_n by K_n^{k+}, K_n^{k-} respectively. These two sequences converge weakly to tight Gaussian processes in $l^\infty[0,1]^2$ taking their values on $C[0,1]^2$ by the same arguments as those used in the proof of Lemma 5.1.1.

This is enough to prove a uniform law of large numbers for K_n, which will be essential to prove the consistency of $\hat{\mu}^{(n)}$. Let $\mathcal{A}$ denote the class of lower layers of $[0,1]^2$.

Lemma 5.1.2 *For $A \in \mathcal{A}$,*

$$\sup_{t \in A} \left| \frac{K_n(t)}{n} - P(t \in \xi) \right| \to_P 0$$

as $n \to \infty$.

Proof: The proof mirrors that of Lemma 4.9 in [12], but again, as the survival function is nowhere to be seen, the argument is simplified a bit. We just need to replace $Z_n, Z_n^{k+}, Z_n^{k-}, D_n^{k+}, D_n^{k-}$ in [12] by $K_n, K_n^{k+}, K_n^{k-}, B_n^{k+}, B_n^{k-}$, respectively, omit $S(t)$ in every instance, and the proof goes through intact. □

Theorem 5.1.3 *Let $A \in \mathcal{A}$. Under Assumptions 4.2.4, 4.2.5, 4.2.6 and 4.2.13, $\hat{\mu}^{(n)}(A)$ is a consistent estimator of $\mu(A)$.*

Proof: Note that if $A = A_t$ for some $t \in T$, consistency is a simple consequence of Theorem 5.1.4. This proof extends the result for more general sets.

Let $A \in \mathcal{A}$. We can write

$$\hat{\mu}^{(n)}(A) = \int_A \frac{1}{\hat{P}_n(t \in \xi)} \hat{\hat{\mu}}^{(n)}(dt) = \int_A \frac{1}{K_n(t)} N^{(n)\xi}(dt),$$

as well as

$$\mu(A) = \int_A \mu(dt) = \int_A S(t)h(t)\lambda(dt).$$

Then, with Z_n, $M^{(n)}$ as in (4.1.1), (4.1.2) respectively,

$$\begin{aligned}
\hat{\mu}^{(n)}(A) - \mu(A) &= \int_A \frac{1}{K_n(t)} N^{(n)\xi}(dt) - \int_A S(t)h(t)\lambda(dt) \\
&= \frac{1}{n}\int_A \left(\frac{1}{K_n(t)/n} - \frac{1}{P(t\in\xi)}\right) N^{(n)\xi}(dt) \\
&+ \frac{1}{n}\int_A \frac{1}{P(t\in\xi)} N^{(n)\xi}(dt) \\
&- \int_A S(t)h(t)\lambda(dt) \\
&= \frac{1}{n}\int_A \left(\frac{1}{K_n(t)/n} - \frac{1}{P(t\in\xi)}\right) N^{(n)\xi}(dt) \qquad (5.1.5)\\
&+ \int_A \frac{1}{nP(t\in\xi)} M^{(n)}(dt) \qquad (5.1.6)\\
&+ \int_A \left(\frac{Z_n(t)}{n} - S(t)P(t\in\xi)\right) \frac{1}{P(t\in\xi)} h(t)\lambda(dt). \qquad (5.1.7)
\end{aligned}$$

But note that

$$\begin{aligned}
(5.1.5) &= \frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i\in A\cap\xi\}}\left(\frac{n}{K_n(Y_i)} - \frac{1}{P(Y_i\in\xi)}\right) \\
&\leq \sup_{t\in A} \left|\frac{n}{K_n(t)} - \frac{1}{P(t\in\xi)}\right|,
\end{aligned}$$

and

$$(5.1.7) \leq \sup_{t\in A}\left|\frac{Z_n(t)}{n} - S(t)P(t\in\xi)\right| \int_A \frac{1}{P(t\in\xi)} h(t)\lambda(dt),$$

so (5.1.5) and (5.1.7) converge to 0 in probability as $n\to\infty$ by Lemma 5.1.2 and Lemma 4.2.14, respectively.

Finally, consider (5.1.6). Let $g_n(t) = (nP(t\in\xi))^{-1}$. Then g_n is continuous by Assumption 4.2.5, and by Corollary 4.2.10,

$$\begin{aligned}
E\left[\left(\int_A g_n(t)M^{(n)}(dt)\right)^2\right] &= \frac{1}{n}\int_A \frac{S(t)h(t)}{P(t\in\xi)}\lambda(dt) \\
&+ \frac{1}{n}\int\int_{l(A,A)} \frac{\mu(E_s\cap E_t)P(s,t\in\xi)h(s)h(t)}{P(s\in\xi)P(t\in\xi)}\lambda(ds)\lambda(dt) \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Hence (5.1.6) converges in probability to 0 as $n \to \infty$ and the proof is complete. $\square$

As was the case with the Nelson-Aalen estimator, we can prove a functional central limit theorem for the process $\sigma_.^{(n)} = \sqrt{n}(\hat{\mu}^{(n)}(A_.) - \mu(A_.))$ under similar conditions to those of Theorem 4.2.7, where $A_t = [0,t]$ for $t \in T$.

Theorem 5.1.4 *Under Assumptions 4.2.4 and 4.2.5, $\sigma^{(n)} \Rightarrow Q$ in $l^\infty[0,1]^2$, where Q is a mean-zero Gaussian process such that for $t \in T$,*

$$Q_t = \int_{[0,t]} \frac{d\Gamma(u)}{P(u \in \xi)} - \int_{[0,t]} \frac{\Lambda(u)}{P(u \in \xi)} \mu(du),$$

and where Γ and Λ are the processes defined in Theorem 4.2.7 and Lemma 5.1.1 respectively. The first integral is defined via integration by parts.

Proof: From the proof of Theorem 4.2.7 we know that the process $\frac{1}{\sqrt{n}}[N_{A.}^{(n)\xi} - E(N_{A.}^{(n)\xi})] \Rightarrow \Gamma$, where Γ is a Gaussian process on $D[0,1]^2$. Moreover, Lemma 5.1.1 gives us the convergence in distribution of $\frac{1}{\sqrt{n}}[K_n(t) - E(K_n(t))]$ to a tight continuous Gaussian process Λ on $l^\infty[0,1]^2$. Then we have that

$$\frac{1}{\sqrt{n}}[(K_n(\cdot), N^{(n)\xi}_{A.}) - (E(K_n(\cdot)), E(N^{(n)\xi}_{A.}))] \Rightarrow (\Lambda, \Gamma)$$

on $l^\infty[0,1]^2 \times D[0,1]^2$, where (Λ, Γ) is a Gaussian process on $C[0,1]^2 \times D[0,1]^2$. Note that $\hat{\mu}^{(n)}$ depends on the pair $\left(\frac{K_n}{n}, \frac{N^{(n)\xi}}{n}\right)$ through the composition map (4.2.3), and we also know by Lemma 4.2.2 that this map is Hadamard-differentiable tangentially to $C[0,1]^2 \times D[0,1]^2$ on a domain of the type $\{(A,B) : \int |dB| \le M, A \ge \epsilon\}$ for given M and $\epsilon > 0$, at every point (A,B) such that $1/A$ is of bounded variation. The pair $(\frac{K_n}{n}, \frac{N^{(n)\xi}}{n})$ is contained in this domain with probability tending to 1 for $M \ge 1$ and ϵ sufficiently small. Finally, apply the Delta-Method to conclude that $\sigma_.^{(n)} = \sqrt{n}(\hat{\mu}^{(n)}(A_.) - \mu(A_.)) \Rightarrow Q_.$, where

$$Q_t = \int_{[0,t]} \frac{d\Gamma(u)}{P(u \in \xi)} - \int_{[0,t]} \frac{\Lambda(u)}{(P(u \in \xi))^2} dP(Y \in A_u \cap \xi)$$

$$
\begin{aligned}
&= \int_{[0,t]} \frac{d\Gamma(u)}{P(u \in \xi)} - \int_{[0,t]} \frac{\Lambda(u)}{(P(u \in \xi))^2} \tilde{\mu}(du) \\
&= \int_{[0,t]} \frac{d\Gamma(u)}{P(u \in \xi)} - \int_{[0,t]} \frac{\Lambda(u)}{P(u \in \xi)} \mu(du)
\end{aligned}
$$

is again Gaussian. As in Lemma 4.2.2, the first term in the limiting process has to be defined by integration by parts, since Γ may not be of bounded variation. □

Remark 5.1.5 Given the way we have set up our estimator for the distribution function, it is entirely possible that for small n, $\hat{\mu}^{(n)}(A_t)$ gives a value greater than 1. If we are interested in demanding that our estimator behave like a distribution function even for such sample sizes, we could opt for a scaled version of $\hat{\mu}^{(n)}$. Define

$$
\check{\mu}^{(n)}(A_t) := \frac{\hat{\mu}^{(n)}(A_t)}{\hat{\mu}^{(n)}(A_{(1,1)})},
$$

and note that Theorem 5.1.3 implies that $\hat{\mu}^{(n)}(A_{(1,1)}) \to_P \mu(A_{(1,1)}) = 1$. Then the new estimator $\check{\mu}^{(n)}$ preserves all the properties of $\hat{\mu}^{(n)}$, with the added bonus of being properly scaled.

The computation of the asymptotic variance of our estimate, as well as the covariance structure for the limiting Gaussian process Q, will be harder than that of the covariance of the process G (see Lemma 4.2.11). Before attempting to perform such calculations, we will rewrite the process $\sigma_t^{(n)} = \sqrt{n}(\hat{\mu}^{(n)}(A_t) - \mu(A_t))$ in a way that better suits our purposes. We have that, for $t \in T$,

$$
\begin{aligned}
\sigma_t^{(n)} &= \sqrt{n}(\hat{\mu}^{(n)}(A_t) - \mu(A_t)) \\
&= \sqrt{n}\left(\int_{A_t} \frac{1}{\hat{P}_n(s \in \xi)} \hat{\tilde{\mu}}_n(ds) - \int_{A_t} \mu(ds)\right) \\
&= \int_{A_t} \frac{1}{\hat{P}_n(s \in \xi)} \left(\sqrt{n}[\hat{\tilde{\mu}}_n(ds) - \tilde{\mu}(ds)]\right) \qquad (5.1.8) \\
&+ \sqrt{n}\left(\int_{A_t} \frac{1}{\hat{P}_n(s \in \xi)} \tilde{\mu}(ds) - \int_{A_t} \mu(ds)\right); \qquad (5.1.9)
\end{aligned}
$$

we will deal with (5.1.8) and (5.1.9) separately. First,

$$\begin{aligned}
(5.1.8) &= \int_{A_t} \left(\frac{1}{\hat{P}_n(s \in \xi)} - \frac{1}{P(s \in \xi)} \right) \left(\sqrt{n}[\hat{\tilde{\mu}}_n(ds) - \tilde{\mu}(ds)] \right) \\
&+ \int_{A_t} \frac{1}{P(s \in \xi)} \left(\sqrt{n}[\hat{\tilde{\mu}}_n(ds) - \tilde{\mu}(ds)] \right) \\
&= \int_{A_t} \frac{\sqrt{n}\left(P(s \in \xi) - \hat{P}_n(s \in \xi) \right)}{\hat{P}_n(s \in \xi) P(s \in \xi)} \left(\hat{\tilde{\mu}}_n(ds) - \tilde{\mu}(ds) \right) \qquad (5.1.10) \\
&+ \int_{A_t} \frac{1}{P(s \in \xi)} \left(\sqrt{n}[\hat{\tilde{\mu}}_n(ds) - \tilde{\mu}(ds)] \right), \qquad (5.1.11)
\end{aligned}$$

while

$$\begin{aligned}
(5.1.9) &= \sqrt{n} \left(\int_{A_t} \frac{1}{\hat{P}_n(s \in \xi)} \tilde{\mu}(ds) - \int_{A_t} \frac{1}{P(s \in \xi)} \tilde{\mu}(ds) \right) \\
&= \int_{A_t} \frac{\sqrt{n}\left(P(s \in \xi) - \hat{P}_n(s \in \xi) \right)}{\hat{P}_n(s \in \xi) P(s \in \xi)} \tilde{\mu}(ds) \\
&= \int_{A_t} \frac{\sqrt{n}\left(P(s \in \xi) - \hat{P}_n(s \in \xi) \right)}{P^2(s \in \xi)} \tilde{\mu}(ds) \qquad (5.1.12) \\
&+ \int_{A_t} \frac{\sqrt{n}\left(P(s \in \xi) - \hat{P}_n(s \in \xi) \right)}{P^2(s \in \xi)} \left(\frac{P(s \in \xi)}{\hat{P}_n(s \in \xi)} - 1 \right) \tilde{\mu}(ds). \qquad (5.1.13)
\end{aligned}$$

But notice that (5.1.10) $\to_P 0$ by the exact same reasoning as equation (29) in [11], while

$$(5.1.13) \leq \sup_{s \in A_t} \sqrt{n} |P(s \in \xi) - \hat{P}_n(s \in \xi)| \sup_{s \in A_t} |\frac{P(s \in \xi)}{\hat{P}_n(s \in \xi)} - 1| \sup_{s \in A_t} \left(\frac{1}{P^2(s \in \xi)} \right) \tilde{\mu}(A_t),$$

which converges to 0 in probability since

$$\sup_{s \in A_t} |\frac{P(s \in \xi)}{\hat{P}_n(s \in \xi)} - 1| \leq \sup_{s \in A_t} |P(s \in \xi) - \hat{P}_n(s \in \xi)| \sup_{s \in A_t} \frac{1}{\hat{P}_n(s \in \xi)} \to_P 0$$

by Lemma 5.1.2. These two facts, along with the relations $\hat{\tilde{\mu}}^{(n)}(A) = \frac{1}{n} N_A^{(n)\xi}$, $K_n(t) = n\hat{P}_n(t \in \xi)$ and equation (5.1.1), yield

$$\sigma_t^{(n)} = (5.1.11) + (5.1.12) + o_p(1)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}}\left(\int_{A_t}\frac{1}{P(s\in\xi)}N^{(n)\xi}(ds)-\int_{A_t}\frac{K_n(s)}{P(s\in\xi)}\mu(ds)\right)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\left(\frac{I_{\{Y_i\in A_t\cap\xi_i\}}}{P(Y_i\in\xi|Y_i)}-\int_{A_t}\frac{I_{\{s\in\xi_i\}}}{P(s\in\xi)}\mu(ds)\right)\right]+o_p(1) \qquad (5.1.14)\\
&= \tilde{\sigma}_t^{(n)}+o_p(1). \qquad (5.1.15)
\end{aligned}
$$

Therefore, since the process $\sigma_t^{(n)}$ is asymptotically equivalent to a sum of i.i.d. random variables, the usual CLT gives us that the asymptotic variance of the estimate $\hat{\mu}^{(n)}(A_t)$ is equal to the variance of $\tilde{\sigma}_t^{(n)}$.

Remark 5.1.6 Care must be taken to distinguish between the expressions $P(Y\in\xi)$ and $P(Y\in\xi|Y)$; in fact, we have

$$P(Y\in\xi)=\int_T P(Y\in\xi|Y=s)\mu(ds)=\int_T P(s\in\xi)\mu(ds).$$

Theorem 5.1.7 *Let $t\in T$. The asymptotic variance of the estimator $\hat{\mu}^{(n)}(A_t)$ is given by*

$$\int_{A_t}\frac{1}{P(s\in\xi)}\mu(ds)-\int_{A_t}\int_{A_t}\frac{P(s,u\in\xi)}{P(s\in\xi)P(u\in\xi)}\mu(ds)\mu(du).$$

Proof: The groundwork for the proof was laid in the discussion that preceded Remark 5.1.6. Since the asymptotic variance of $\hat{\mu}^{(n)}(A_t)$ is equal to the variance of $\tilde{\sigma}_t^{(n)}$, and recalling that both the Y_i's and the ξ_i's are i.i.d., we have

$$
\begin{aligned}
\text{Avar}\left(\hat{\mu}^{(n)}(A_t)\right) &= \text{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{I_{\{Y_i\in A_t\cap\xi_i\}}}{P(Y_i\in\xi|Y_i)}-\int_{A_t}\frac{I_{\{s\in\xi_i\}}}{P(s\in\xi)}\mu(ds)\right)\right]\\
&= \text{Var}\left[\frac{I_{\{Y\in A_t\}}I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)}-\int_{A_t}\frac{I_{\{s\in\xi\}}}{P(s\in\xi)}\mu(ds)\right]. \qquad (5.1.16)
\end{aligned}
$$

The expected value of the expression inside the brackets in (5.1.16) is equal to zero; indeed, conditioning on Y gives us

$$
\begin{aligned}
&E\left[\frac{I_{\{Y\in A_t\}}I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)}-\int_{A_t}\frac{I_{\{s\in\xi\}}}{P(s\in\xi)}\mu(ds)\right]\\
&= E\left[E\left(\frac{I_{\{Y\in A_t\}}I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)}-\int_{A_t}\frac{I_{\{s\in\xi\}}}{P(s\in\xi)}\mu(ds)|\,Y\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= E\left[\frac{I_{\{Y\in A_t\}}}{P(Y\in\xi|Y)}P(Y\in\xi|Y) - \int_{A_t}\frac{P(s\in\xi)}{P(s\in\xi)}\mu(ds)\right] \\
&= P(Y\in A_t) - \mu(A_t) \\
&= 0.
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
(5.1.16) &= E\left[\left(\frac{I_{\{Y\in A_t\}}I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)} - \int_{A_t}\frac{I_{\{s\in\xi\}}}{P(s\in\xi)}\mu(ds)\right)^2\right] \\
&= E\left[E\left\{\left(\frac{I_{\{Y\in A_t\}}I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)} - \int_{A_t}\frac{I_{\{s\in\xi\}}}{P(s\in\xi)}\mu(ds)\right)^2 \mid Y\right\}\right] \\
&= E\left[\frac{I_{\{Y\in A_t\}}}{P^2(Y\in\xi|Y)}P(Y\in\xi|Y)\right] \\
&- 2E\left[\frac{I_{\{Y\in A_t\}}}{P(Y\in\xi|Y)}E\left(I_{\{Y\in\xi\}}\int_{A_t}\frac{I_{\{s\in\xi\}}}{P(s\in\xi)}\mu(ds)|\, Y\right)\right] \\
&+ E\left[E\left(\int_{A_t}\int_{A_t}\frac{I_{\{s\in\xi\}}I_{\{u\in\xi\}}}{P(s\in\xi)P(u\in\xi)}\mu(ds)\mu(du)|\, Y\right)\right] \\
&= \int_{A_t}\frac{1}{P(s\in\xi)}\mu(ds) - 2E\left[\frac{I_{\{Y\in A_t\}}}{P(Y\in\xi|Y)}\int_{A_t}\frac{P(Y,s\in\xi|Y)}{P(s\in\xi)}\mu(ds)\right] \\
&+ \int_{A_t}\int_{A_t}\frac{P(s,u\in\xi)}{P(s\in\xi)P(u\in\xi)}\mu(ds)\mu(du) \\
&= \int_{A_t}\frac{1}{P(s\in\xi)}\mu(ds) - 2\int_{A_t}\frac{1}{P(u\in\xi)}\int_{A_t}\frac{P(s,u\in\xi)}{P(s\in\xi)}\mu(ds)\mu(du) \\
&+ \int_{A_t}\int_{A_t}\frac{P(s,u\in\xi)}{P(s\in\xi)P(u\in\xi)}\mu(ds)\mu(du) \\
&= \int_{A_t}\frac{1}{P(s\in\xi)}\mu(ds) - \int_{A_t}\int_{A_t}\frac{P(s,u\in\xi)}{P(s\in\xi)P(u\in\xi)}\mu(ds)\mu(du).
\end{aligned}
$$

This completes the proof. □

Theorem 5.1.8 *Let $s, t \in T$. The covariance structure for the limiting process Q in Theorem 5.1.4 is given by*

$$
\begin{aligned}
&\mathrm{Cov}(Q_t, Q_s) \\
&= \int_{A_t\cap A_s}\frac{1}{P(u\in\xi)}\mu(du) - \int_{A_t}\int_{A_s}\frac{P(u,v\in\xi)}{P(u\in\xi)P(v\in\xi)}\mu(dv)\mu(du). \quad (5.1.17)
\end{aligned}
$$

Proof: This proof is, as expected, very similar to the proof of Theorem 5.1.7. Again, note that Equation (5.1.15) gives us

$$\begin{aligned}\operatorname{Cov}(Q_t, Q_s) &= \operatorname{Cov}(\tilde{\sigma}_t^{(n)}, \tilde{\sigma}_s^{(n)}) \\ &= E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{I_{\{Y_i\in A_t\cap\xi_i\}}}{P(Y_i\in\xi|Y_i)} - \int_{A_t}\frac{I_{\{u\in\xi_i\}}}{P(u\in\xi)}\mu(du)\right)\right. \\ &\quad \left.\times\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{I_{\{Y_j\in A_s\cap\xi_j\}}}{P(Y_j\in\xi|Y_j)} - \int_{A_s}\frac{I_{\{v\in\xi_j\}}}{P(v\in\xi)}\mu(dv)\right)\right]. \quad (5.1.18)\end{aligned}$$

Since the Y_i's and the ξ_i's are i.i.d., we get that

$$\begin{aligned}(5.1.18) \quad = \quad & E\left[\frac{I_{\{Y\in A_s\cap A_t\cap\xi\}}}{P^2(Y\in\xi|Y)} - \frac{I_{\{Y\in A_t\cap\xi\}}}{P(Y\in\xi|Y)}\int_{A_s}\frac{I_{\{v\in\xi\}}}{P(v\in\xi)}\mu(dv)\right. \\ & \left. - \frac{I_{\{Y\in A_s\cap\xi\}}}{P(Y\in\xi|Y)}\int_{A_t}\frac{I_{\{u\in\xi\}}}{P(u\in\xi)}\mu(du) + \int_{A_t}\int_{A_s}\frac{I_{\{u,v\in\xi\}}}{P(u\in\xi)P(v\in\xi)}\mu(dv)\mu(du)\right].\end{aligned}$$

From here on, we proceed exactly as in the proof of Theorem 5.1.7: we condition on Y and reduce terms in the same way until we arrive at (5.1.17). □

Remark 5.1.9 It is possible to use Equation (5.1.14) and the usual multivariate CLT to prove the convergence of the finite-dimensional distributions of $\sigma_t^{(n)}$; however, as was pointed out in Remark 4.2.12, the matter of tightness would still need to be addressed if we wanted a functional CLT. The Delta-Method, fortunately, allows us to bypass the rather messy calculations needed to prove tightness.

With the asymptotic properties of our estimator firmly established, we can now focus on some applications, starting with an alternative to the test of independence presented in the previous Chapter.

Test of Independence and Confidence Bands

In Chapter 4 we constructed a test of independence for the components (Y_1, Y_2) of a random vector Y; we based it on the process $V_n = n^{1/2}(\hat{H}^{(n)} - \hat{H}_1^{(n)}\hat{H}_2^{(n)})$ to take advantage of the fact that when the components are independent, the joint cumulative hazard is the product of the marginal cumulative hazards. The same is true for the joint distribution μ of the random vector and the marginal distributions μ_1, μ_2 of its components: if Y_1 and Y_2 are independent, then $\mu = \mu_1\mu_2$, and so we would like to use our estimator $\hat{\mu}^{(n)}$ to develop another independence test.

For ease of notation, let $\hat{\mu}^{(n)}(A_t) = \hat{\mu}^{(n)}(t_1, t_2)$ for any $t = (t_1, t_2) \in T$. Define the process $U^{(n)} = n^{1/2}(\hat{\mu}^{(n)} - \hat{\mu}_1^{(n)}\hat{\mu}_2^{(n)})$, where $\hat{\mu}_1^{(n)}(t_1) = \hat{\mu}^{(n)}(t_1, 1)$ and $\hat{\mu}_2^{(n)}(t_2) = \hat{\mu}^{(n)}(1, t_2)$ are the estimators of the marginals μ_1 and μ_2 respectively. Note that sometimes the data structure makes it possible to use an estimate that employs more data points for the marginals; for example, the pregnancy/tuberculosis case allows us to estimate the marginal distribution of the pregnancy times via the empirical distribution, since those observations aren't censored at all. However, $\hat{\mu}_1^{(n)}$ and $\hat{\mu}_2^{(n)}$ work on all cases. We are not overly concerned with the calculation of the variance of $U^{(n)}$, as was the case with the variance of V_n; rather, we want to proceed with a bootstrap test.

For $(t_1, t_2) \in T$, let $Q^1_{t_1} := Q_{(t_1,1)}$ and $Q^2_{t_2} := Q_{(1,t_2)}$, and so we have by Theorem 5.1.4 that $n^{1/2}(\hat{\mu}_1^{(n)} - \mu_1) \Rightarrow Q^1$ and $n^{1/2}(\hat{\mu}_2^{(n)} - \mu_2) \Rightarrow Q^2$. Then, since

$$\begin{aligned}
U^{(n)} &= n^{1/2}(\hat{\mu}^{(n)} - \hat{\mu}_1^{(n)}\hat{\mu}_2^{(n)}) \\
&= n^{1/2}(\hat{\mu}^{(n)} - \mu) - n^{1/2}(\hat{\mu}_1^{(n)} - \mu_1)\hat{\mu}_2^{(n)} \\
&- n^{1/2}(\hat{\mu}_2^{(n)} - \mu_2)\mu_1 + n^{1/2}(\mu - \mu_1\mu_2),
\end{aligned}$$

we have that

$$\begin{aligned}
n^{1/2}\left(\hat{\mu}^{(n)} - \hat{\mu}_1^{(n)}\hat{\mu}_2^{(n)} - (\mu - \mu_1\mu_2)\right) &= n^{1/2}(\hat{\mu}^{(n)} - \mu) \\
&- n^{1/2}(\hat{\mu}_1^{(n)} - \mu_1)(\hat{\mu}_2^{(n)} - \mu_2) \\
&- n^{1/2}(\hat{\mu}_1^{(n)} - \mu_1)\mu_2 \\
&- n^{1/2}(\hat{\mu}_2^{(n)} - \mu_2)\mu_1,
\end{aligned}$$

which converges weakly to $Q - \mu_2 Q^1 - \mu_1 Q^2$, a mean-zero Gaussian process.

Now, as we did in the previous test of independence, suppose we can completely observe the pairs (Y_i, ξ_i). Recall the class of functions $\mathcal{G}^2 = \{g_t^2 : t \in T\}$, where

$g_t^2(Y_i,\xi_i) := I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}$, and define $\mathcal{G}^3 = \{g_t^3 : t\in T\}$, where $g_t^3(Y_i,\xi_i) := I_{\{t\in\xi_i\}}$. The proof of Theorem 5.1.4 showed that, as a pair, $(\mathcal{G}^2,\mathcal{G}^3)$ is a Donsker class. Moreover, both classes have finite envelope functions since they only include indicator functions. We also know that the composition map (4.2.3) is Hadamard-differentiable tangentially to $C[0,1]^2\times BV_M[0,1]^2$ on a certain domain, so we can apply Theorem 3.9.11 in [22] to get the conditional convergence given (Y_i,ξ_i) of the bootstrapped estimator to the same Gaussian process $Q-\mu_2Q^1-\mu_1Q^2$.

At this point we observe again that our estimators depend only on the indicators $p_i = I_{\{t\in\xi_i\}}$ and $q_i = I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}$, so we can in fact bootstrap directly from the observed values p_i and q_i and obtain the same estimators, which allows us to drop the assumption of complete observability. Therefore, if $\ddot{\mu}^{(n)},\ddot{\mu}_1^{(n)},\ddot{\mu}_2^{(n)}$ denote the bootstrapped versions of the joint and marginal estimators of the distribution, respectively, we have that

$$\sup_{t\leq(1,1)}\left|n^{1/2}[\ddot{\mu}^{(n)}(t_1,t_2)-\ddot{\mu}_1^{(n)}(t_1)\ddot{\mu}_2^{(n)}(t_2)-(\hat{\mu}^{(n)}(t_1,t_2)-\hat{\mu}_1^{(n)}(t_1)\hat{\mu}_2^{(n)}(t_2))]\right| \quad (5.1.19)$$

converges conditionally in outer probability to the sup of the absolute value of the Gaussian process $Q-\mu_2Q^1-\mu_1Q^2$ on $[0,1]^2$. Then we can bootstrap from (p_i,q_i) jointly and find c_α so that $(1-\alpha)100\%$ of the absolute values of (5.1.19) fall below c_α, which would give us $(1-\alpha)100\%$ uniform confidence bands for the difference $\mu-\mu_1\mu_2$.

For a test of independence $H_\circ : \mu = \mu_1\mu_2$ vs. $H_1 : \mu\neq\mu_1\mu_2$, note that under the null hypothesis, $P(\sup|U_n(t_1,t_2)|<c_\alpha)\approx 1-\alpha$. Therefore, we reject $H_\circ$ if $\sup n^{1/2}\left|\hat{\mu}^{(n)}(t_1,t_2)-\hat{\mu}_1^{(n)}(t_1)\hat{\mu}_2^{(n)}(t_2)\right| > c_\alpha$. Note that the test is consistent: under H_1, the test statistic $\sup_{t\leq(1,1)}U^{(n)}$ converges to ∞ in probability.

Copula Estimation

Consider a sequence (X_i,Y_i) of i.i.d. bivariate random vectors with continuous distribution F and marginals F_1, F_2, and suppose that each pair (X_i,Y_i) is censored by a cloud ξ_i^c. Let $C(p,q)=F(F_1^{-1}(p),F_2^{-1}(q))$ denote the copula associated with F. In Chapter 4 we discussed an estimator for the cumulative hazard of the copula, which was based on the bivariate Nelson-Aalen estimator and the Kaplan-Meier estimates of the marginals. We can also use our estimator $\hat{\mu}^{(n)}$ to construct an estimate of the copula itself.

Define $\hat{C}^{(n)}(p,q) = \hat{\mu}^{(n)}(\hat{\mu}_1^{(n)-}(p), \hat{\mu}_2^{(n)-}(q))$, where $\hat{\mu}_1^{(n)-}$ and $\hat{\mu}_2^{(n)-}$ denote the quantiles of the marginal estimators $\hat{\mu}_1^{(n)}$ and $\hat{\mu}_2^{(n)}$ respectively. $\hat{C}^{(n)}$ is the desired estimator for the copula, and note that it depends only on the pair $(K_n, N^{(n)^\xi})$. Moreover, the composition map

$$\begin{aligned}
(K_n, N^{(n)^\xi}) &\xrightarrow{\phi_1} \hat{\mu}^{(n)} \\
&\xrightarrow{\phi_2} (\hat{\mu}^{(n)}, \hat{\mu}_1^{(n)}, \hat{\mu}_2^{(n)}) \\
&\xrightarrow{\phi_3} (\hat{\mu}^{(n)}, \hat{\mu}_1^{(n)-}, \hat{\mu}_2^{(n)-}) \\
&\xrightarrow{\phi_4} \hat{\mu}^{(n)}(\hat{\mu}_1^{(n)-}, \hat{\mu}_2^{(n)-})
\end{aligned}$$

is Hadamard-differentiable tangentially to $C[0,1]^2 \times BV_M[0,1]^2$ on a domain analogous to map (4.2.3): the map ϕ_1 was shown to be Hadamard-differentiable in the proof of Theorem 5.1.4, ϕ_2 is also Hadamard-differentiable because of the definition of $\hat{\mu}_1^{(n)}$ and $\hat{\mu}_2^{(n)}$, and ϕ_3 and ϕ_4 are Hadamard-differentiable by Example 3.9.24 and Lemma 3.9.25, respectively, in [22]. Then, the Delta-Method guarantees that the process $W_n := \sqrt{n}(\hat{C}^{(n)}(\cdot) - C(\cdot))$ converges in $l^\infty[0,1]^2$ to a mean-zero Gaussian limit W.

All the remarks concerning bootstrapping for the statistic $U^{(n)}$ in the test of independence are still valid in this case: like $\hat{\mu}^{(n)}$, $\hat{C}^{(n)}$ depends only on $(K_n, N^{(n)^\xi})$ through a Hadamard-differentiable composition, so Theorem 3.9.11 in [22] guarantees the conditional convergence given (Y_i, ξ_i) of the bootstrapped estimator to the same Gaussian limit W.

5.2 Ordered Censoring

Now we turn our attention to a different kind of scenario. Our approach will still be based in the reverse probability principles that have been applied throughout Section

5.1, yet we will be introducing a twist that in some ways foreshadows the ideas of Chapter 6.

Let $Y : \Omega \to [0,1]^2$ be a random variable with distribution $\mu(B) = P\{Y \in B\}$, and let $\mathcal{S}$ denote an increasing class of lower layers. Furthermore, assume that $\mu(\partial S) = 0$ for all $S \in \mathcal{S}$. Suppose we have a sequence of i.i.d. random variables (Y_i) with the same distribution as Y, as well as a sequence of anti-clouds $\xi_i : \Omega \to \mathcal{S}$, where each Y_i is censored by ξ_i^c. By suitably augmenting $\mathcal{S}$ if necessary, we can assume that there exists a strictly increasing map $\alpha : [0,1] \to \mathcal{S}$, continuous in the Hausdorff metric, such that $\alpha(0) = (0,0)$ and $\alpha(1) = [0,1]^2$ (For subsets C and D of a metric space, the Hausdorff distance is defined as

$$h(C,D) = \sup_{x \in C} d(x,D) \;\vee\; \sup_{x \in D} d(x,C)).$$

This kind of setup can be found, for example, in radar readings, where the observed region is a quarter circle that increases as the radius expands. Another example is the univariate censoring setup of Lu and Burke in [17] (See Figure 5.1). In contrast with the situation presented in Section 5.1, here we don't need to be able to see the shape of the unobscured regions, since we know the class of shapes beforehand; however, now we fully observe ξ_i only if $Y_i \notin \xi_i$. It is our intention to estimate the distribution function μ using (5.1.2), but instead of taking $\hat{P}_n(t \in \xi) = \frac{1}{n}\sum_{i=1}^n I_{\{t \in \xi_i\}}$ as an estimate of $P(t \in \xi)$, we want to transform our two-dimensional observations into a one-dimensional set of data, and then use conventional one-dimensional methods to produce the desired estimator. This will be done via the map α, which will parametrize both the censoring and the jump points.

Let $\xi_i^\alpha = \alpha^{-1}(\xi_i)$ and consider the map $\tilde{\alpha} : [0,1]^2 \to [0,1]$ defined by $\tilde{\alpha}(t) := \inf\{u : t \in \alpha(u)\}$. Note that $\{\tilde{\alpha}(t) \le \xi_i^\alpha\} \Leftrightarrow \{t \in \alpha(\xi_i^\alpha)\} \Leftrightarrow \{t \in \xi_i\}$; in particular, $\{\tilde{\alpha}(Y_i) \le \xi_i^\alpha\} \Leftrightarrow \{Y_i \in \xi_i\}$, so in fact the 'artificial' jump points $\tilde{\alpha}(Y_i)$ are uncensored by ξ_i^α whenever the 'real' jump points Y_i are observable, and viceversa. Now, let's switch the roles of $\tilde{\alpha}(Y_i)$ and ξ_i^α and consider the former as the censoring variable and the latter as the variable of interest. Denote the survival function of ξ^α by S_{ξ^α}. Then

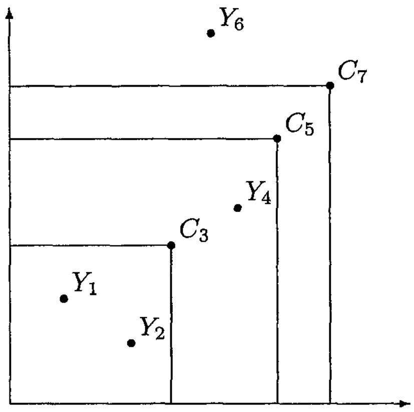

Figure 5.1: Univariate censoring, an example of ordered censoring.

we have $P(t \in \xi) = P(\tilde{\alpha}(t) \leq \xi^\alpha) = S_{\xi^\alpha}(\tilde{\alpha}(t))$, so an estimate of this survival function will give us the estimate of $P(t \in \xi)$ that we desire; since the ξ_i^α's are unidimensional random variables, we elect to use the Kaplan-Meier estimate (See Appendix), given by

$$\hat{S}_{\xi^\alpha}^{(n)}(p) = \prod_{i:\xi_i^\alpha \wedge \tilde{\alpha}(Y_i) \leq p} \left(1 - \frac{I_{\{\xi_i^\alpha \leq \tilde{\alpha}(Y_i)\}}}{\sum_{j=1}^n I_{\{\xi_j^\alpha \wedge \tilde{\alpha}(Y_j) \geq \xi_i^\alpha\}}}\right).$$

Let

$$\hat{P}_\alpha^{(n)}(t \in \xi) := \hat{S}_{\xi^\alpha}^{(n)}(\tilde{\alpha}(t)) \tag{5.2.1}$$

denote the estimate of $P(t \in \xi)$. We can plug (5.2.1) into Equation (5.1.2) to obtain the estimator

$$\hat{\mu}_\alpha^{(n)}(A) = \int_A \frac{1}{\hat{S}_{\xi^\alpha}^{(n)}(\tilde{\alpha}(t))} \hat{\hat{\mu}}_n(dt) = \frac{1}{n} \sum_{\{i: Y_i \in A \cap \xi_i\}} (\hat{S}_{\xi^\alpha}^{(n)}(\tilde{\alpha}(Y_i)))^{-1}. \tag{5.2.2}$$

In [17], Lu and Burke consider a class of estimators for linear functionals of a two-dimensional distribution under univariate censoring; it is based on the same reverse probability relation, first presented in [3], that we have exploited in this Chapter. One of their resulting estimates for the distribution function can be viewed as a particular case of ours, where the anti-clouds are rectangles of the form $[0, p] \times [0, p]$ for $p > 0$.

The next step is to consider the asymptotic behavior of $\hat{\mu}_\alpha^{(n)}$.

Theorem 5.2.1 *Let $\sigma_\alpha^{(n)}$ be the process defined by $\sigma_\alpha^{(n)}(\cdot) := \sqrt{n}(\hat{\mu}_\alpha^{(n)}(A.) - \mu(A.))$. Under Assumption 4.2.5, $\sigma_\alpha^{(n)} \Rightarrow \Delta$ in $l^\infty[0,\tau]^2$ for τ such that $P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq \tau) > 0$, where Δ is a mean-zero Gaussian process.*

Proof: The proof, as expected, is very similar to that of Theorem 5.1.4. Define the one-dimensional processes

$$\begin{aligned} N^{(n)\xi^\alpha}(p) &= \sum_{i=1}^n I_{\{\xi_i^\alpha \leq p\}} I_{\{\xi_i^\alpha \leq \tilde{\alpha}(Y_i)\}} \\ Z_n^{\xi^\alpha}(p) &= \sum_{i=1}^n I_{\{\xi_i^\alpha \geq p\}} I_{\{p \leq \tilde{\alpha}(Y_i)\}}. \end{aligned}$$

Consider the composition map

$$\begin{aligned} \left(\frac{Z_n^{\xi^\alpha}}{n}, \frac{N^{(n)\xi^\alpha}}{n}, \frac{N^{(n)\xi}}{n}\right) &\longrightarrow \left(\hat{H}_{\xi^\alpha}^{(n)}, \frac{N^{(n)\xi}}{n}\right) \\ &\longrightarrow \left(\hat{S}_{\xi^\alpha}^{(n)}, \frac{N^{(n)\xi}}{n}\right) \\ &\longrightarrow \left(\frac{1}{\hat{S}_{\xi^\alpha}^{(n)}}, \frac{N^{(n)\xi}}{n}\right) \\ &\longrightarrow \int \frac{d\hat{\hat{\mu}}_n}{\hat{S}_{\xi^\alpha}^{(n)}} = \hat{\mu}_\alpha^{(n)}, \end{aligned} \tag{5.2.3}$$

where $\hat{H}_{\xi^\alpha}^{(n)}$ denotes the Nelson-Aalen estimate (See Appendix) for the cumulative hazard function of the ξ_i^α's. The first map of the composition is Hadamard differentiable on a domain of the type $\left\{(A,B,C) : \int |dB| \leq M_1, A \geq \epsilon_1\right\}$ for given M_1 and $\epsilon_1 > 0$, at every point (A,B,C) such that $1/A$ is of bounded variation; the second map is Hadamard differentiable as long as $\int |d\hat{H}_{\xi^\alpha}^{(n)}| \leq M_2$ for some constant M_2, while the last two maps taken together are actually map (4.2.3), which we already know is Hadamard differentiable on a domain of the type $\left\{(A,B) : \int |dB| \leq M_3, A \geq \epsilon_3\right\}$ for given M_3 and $\epsilon_3 > 0$, at every point (A,B) such that $1/A$ is of bounded variation. By taking $p \in [0,\tau]$ for τ such that $P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq \tau) > 0$, $\left(\frac{Z_n^{\xi^\alpha}}{n}, \frac{N^{(n)\xi^\alpha}}{n}, \frac{N^{(n)\xi}}{n}\right)$ is contained in the first domain with probability tending to 1 for $M_1 \geq 1$ and ϵ_1

sufficiently small. On the other hand, it's also enough to take $p \in [0, \tau]$ for τ such that $P(\xi^\alpha \geq \tau) > 0$ to ensure that $\left(\hat{S}^{(n)}_{\xi^\alpha}, \frac{N^{(n)\xi}}{n}\right)$ is contained in the last domain with probability tending to 1 for $M_3 \geq 1$ and ϵ_3 sufficiently small: indeed, since $Z^{\xi^\alpha}_n$ is decreasing, we have that $\inf_{p \leq \tau} Z^{\xi^\alpha}_n(p) = Z^{\xi^\alpha}_n(\tau) \xrightarrow{P} \infty$, so condition 4.1.10 in [1] is satisfied, which in turn implies that $\sup_{p \leq \tau} |\hat{S}^{(n)}_{\xi^\alpha}(p) - S_{\xi^\alpha}(p)| \xrightarrow{P} 0$. From this point, apply the Delta-Method to get the conclusion. □

Let

$$L_t(p) := E(I_{\{Y \in A_t\}} I_{\{p < \tilde{\alpha}(Y)\}})$$

and

$$\hat{L}_t(p) := \frac{1}{n} \sum_{i=1}^{n} \frac{I_{\{Y_i \in A_t\}} I_{\{Y_i \in \xi_i\}} I_{\{p < \tilde{\alpha}(Y_i)\}}}{\hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(Y_i))}.$$

The following Lemma will allow us to give the asymptotic variance of our estimator.

Lemma 5.2.2 *Under the Conditions of Theorem 5.2.1, $L_t(p)$ is continuous in p for fixed t. Moreover, with probability one,*

$$\sup_{0 \leq p \leq \tau} |\hat{L}_t(p) - L_t(p)| = o(1) \tag{5.2.4}$$

for $t \in T$ such that $\tilde{\alpha}(t) \leq \tau$, with τ as in Theorem 5.2.1.

Proof: By definition, $L_t(p_n) \to L_t(p)$ for any decreasing sequence such that $p_n \downarrow p$. Let (p_n) be an increasing sequence such that $p_n \uparrow p$. We can write

$$L_t(p) = E(I_{\{Y \in A_t\}} I_{\{p < \tilde{\alpha}(Y)\}}) = P(Y \in A_t \setminus \alpha(p)) = \mu(A_t) - \mu(A_t \cap \alpha(p)).$$

Note that, because of the continuity of the map α, we have $\mu(A_t \cap \alpha(p_n)) \to \mu(A_t \cap \bigcup_n \alpha(p_n))$ as $n \to \infty$; moreover, $\alpha(p) \setminus \bigcup_n \alpha(p_n) \subseteq \partial(\alpha(p))$, so

$$\begin{aligned} L_t(p_n) - L_t(p) &= \mu(A_t \cap \alpha(p)) - \mu(A_t \cap \alpha(p_n)) \\ &\to \mu(A_t \cap \alpha(p)) - \mu(A_t \cap \cup_n \alpha(p_n)) \\ &\leq \mu(A_t \cap \partial(\alpha(p))) = 0 \end{aligned}$$

as $n \to \infty$. Hence $L_t(p)$ is continuous in p.

The proof for the second assertion is based on the same idea Lu and Burke employed to prove Theorem 3 of [17]; ours is made easier by the continuity of $L_t(p)$. Fix t such that $\tilde{\alpha}(t) \le \tau$, and let

$$\tilde{L}_t(p) := \int_{A_t} \frac{I_{\{p<\tilde{\alpha}(u)\}}}{S_{\xi^\alpha}(\tilde{\alpha}(u))} \hat{\tilde{\mu}}_n(du),$$

with $\hat{\tilde{\mu}}_n$ defined as in the beginning of Section 5.1. Note that we can write

$$\hat{L}_t(p) = \frac{1}{n}\sum_{i=1}^{n} \frac{I_{\{Y_i\in A_t\}} I_{\{Y_i\in\xi_i\}} I_{\{p<\tilde{\alpha}(Y_i)\}}}{\hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(Y_i))} = \int_{A_t} \frac{I_{\{p<\tilde{\alpha}(u)\}}}{\hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(u))} \hat{\tilde{\mu}}_n(du).$$

Letting

$$\hat{q}(p,u) = \frac{I_{\{p<\tilde{\alpha}(u)\}}}{\hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(u))} \qquad \text{and} \qquad q(p,u) = \frac{I_{\{p<\tilde{\alpha}(u)\}}}{S_{\xi^\alpha}(\tilde{\alpha}(u))},$$

we have that

$$\begin{aligned} \sup_{p,\tilde{\alpha}(u)\le\tau} |\hat{q}(p,u) - q(p,u)| &= \sup_{p,\tilde{\alpha}(u)\le\tau} \left| \frac{I_{\{p<\tilde{\alpha}(u)\}}\left(S_{\xi^\alpha}(\tilde{\alpha}(u)) - \hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(u))\right)}{\hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(u)) S_{\xi^\alpha}(\tilde{\alpha}(u))} \right| \\ &\le \sup_{\tilde{\alpha}(u)\le\tau} \left| \frac{S_{\xi^\alpha}(\tilde{\alpha}(u)) - \hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(u))}{\hat{S}^{(n)}_{\xi^\alpha}(\tilde{\alpha}(u)) S_{\xi^\alpha}(\tilde{\alpha}(u))} \right| = o(1) \end{aligned}$$

with probability 1, since S_{ξ^α} is continuous, $S_{\xi^\alpha}(\tau) > 0$ and $\sup_{p\le\tau} |\hat{S}^{(n)}_{\xi^\alpha}(p-) - S_{\xi^\alpha}(p)| \longrightarrow 0$ almost surely (see [4], [17]). Therefore, $|\hat{L}_t(p) - \tilde{L}_t(p)| = o(1)$ with probability 1. On the other hand, the Strong Law of Large Numbers gives us $\tilde{L}_t(p) \longrightarrow L_t(p)$ with probability 1, and so $\hat{L}_t(p) \longrightarrow L_t(p)$ with probability 1.

To prove uniform convergence, we start by noting that $L_t(p)$ is monotone non-increasing in p. For each positive integer l, we choose a partition $0 = p_{2^l,l} < p_{2^l-1,l}, \ldots, p_{2,l} < p_{1,l} = \tau$ such that $p_{j,l}$ is the smallest p that satisfies $L_t(p) = L_t(0)\frac{j}{2^l}$.

For a given $\epsilon > 0$, fix l such that $L_t(p_{j+1,l}) - L_t(p_{j,l}) = L_t(0)\frac{1}{2^l} < \epsilon$. Moreover, fix n large enough so that $|\hat{L}_t(p_{j,l}) - L_t(p_{j,l})| < \epsilon$ for $j = 1, 2, \ldots 2^l$; this is made possible because of pointwise convergence shown above. For any $p \le \tau$, we have that $p_{j+1,l} \le p \le p_{j,l}$ for some $j = 1, 2, \ldots 2^l$. Then $\hat{L}_t(p_{j,l}) \le \hat{L}_t(p) \le \hat{L}_t(p_{j+1,l})$ because $\hat{L}_t$ is monotone non-increasing in p.

Using these facts, we can say that

$$\hat{L}_t(p) - L_t(p) \leq \hat{L}_t(p_{j+1,l}) - L_t(p) < L_t(p_{j+1,l}) - L_t(p) + \epsilon \leq L_t(p_{j+1,l}) - L_t(p_{j,l}) + \epsilon,$$

as well as

$$\hat{L}_t(p) - L_t(p) \geq \hat{L}_t(p_{j,l}) - L_t(p) > L_t(p_{j,l}) - L_t(p) - \epsilon \geq L_t(p_{j,l}) - L_t(p_{j+1,l}) - \epsilon,$$

whence

$$-(L_t(p_{j+1,l}) - L_t(p_{j,l}) + \epsilon) < \hat{L}_t(p) - L_t(p) < L_t(p_{j+1,l}) - L_t(p_{j,l}) + \epsilon;$$

therefore,

$$|\hat{L}_t(p) - L_t(p)| < L_t(p_{j+1,l}) - L_t(p_{j,l}) + \epsilon < 2\epsilon.$$

This is true for every $p \leq \tau$. Since ϵ was chosen arbitrarily, we can conclude that $\sup_{0 \leq p \leq \tau} |\hat{L}_t(p) - L_t(p)| = o(1)$. □

Theorem 5.2.3 *Let $\tilde{\alpha}(t) \leq \tau$. Let H_{ξ^α} denote the cumulative hazard function of the ξ_i^α's and assume it is continuous. The asymptotic variance of the estimator $\hat{\mu}_\alpha^{(n)}(A_t)$ is given by*

$$\operatorname{Var}\left\{\frac{I_{\{Y \in A_t\}} I_{\{Y \in \xi\}}}{P(Y \in \xi | Y)}\right\} - \int_0^{\tilde{\alpha}(t)} \frac{L_t^2(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} dH_{\xi^\alpha}(s).$$

Proof: We will closely follow the martingale techniques used in the proofs of Theorems 1 and 2 in [17]. For $i = 1, 2 \ldots$ and $p \in [0, 1]$, let

$$\begin{aligned} N_i^{\xi^\alpha}(p) &= I_{\{\xi_i^\alpha \leq p\}} I_{\{\xi_i^\alpha \leq \tilde{\alpha}(Y_i)\}}, \\ M_i^{\xi^\alpha}(p) &= N_i^{\xi^\alpha}(p) - \int_0^p I_{\{\xi_i^\alpha \geq s, \tilde{\alpha}(Y_i) \geq s\}} dH_{\xi^\alpha}(s), \end{aligned} \tag{5.2.5}$$

as well as $J_n(p) = I_{\{p < \max_i(\xi_i^\alpha \wedge \tilde{\alpha}(Y_i))\}}$. Define

$$\theta^{(n)}(t) := \frac{1}{n} \sum_{i=1}^n \frac{I_{\{Y_i \in A_t\}} I_{\{Y_i \in \xi_i\}}}{P(Y_i \in \xi_i | Y_i)},$$

and note that for any $t \in T$ we can write our estimator (5.2.2) as

$$\hat{\mu}_{\alpha}^{(n)}(t) := \hat{\mu}_{\alpha}^{(n)}(A_t) = \frac{1}{n}\sum_{i=1}^{n}\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{\hat{P}_{\alpha}^{(n)}(Y_i\in\xi_i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{\hat{S}_{\xi^{\alpha}}^{(n)}(\tilde{\alpha}(Y_i))}.$$

Denoting $\mu(t) := \mu(A_t)$, we break down the asymptotic variance of interest in the following way:

$$\text{Avar}\{\sqrt{n}(\hat{\mu}_{\alpha}^{(n)}(t)-\mu(t))\} = \text{Avar}\{\sqrt{n}(\hat{\mu}_{\alpha}^{(n)}(t)-\theta^{(n)}(t))+\sqrt{n}(\theta^{(n)}(t)-\mu(t))\}. \quad (5.2.6)$$

Recalling that $P(t\in\xi) = S_{\xi^{\alpha}}(\tilde{\alpha}(t))$,

$$\begin{aligned}
&\sqrt{n}(\hat{\mu}_{\alpha}^{(n)}(t)-\theta^{(n)}(t))\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{\hat{S}_{\xi^{\alpha}}^{(n)}(\tilde{\alpha}(Y_i))} - \frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{S_{\xi^{\alpha}}(\tilde{\alpha}(Y_i))}\right\}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{\hat{S}_{\xi^{\alpha}}^{(n)}(\tilde{\alpha}(Y_i))}\left[\frac{S_{\xi^{\alpha}}(\tilde{\alpha}(Y_i))-\hat{S}_{\xi^{\alpha}}^{(n)}(\tilde{\alpha}(Y_i))}{S_{\xi^{\alpha}}(\tilde{\alpha}(Y_i))}\right]\right\}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{\hat{S}_{\xi^{\alpha}}^{(n)}(\tilde{\alpha}(Y_i))}\cdot\frac{1}{n}\int_{p<\tilde{\alpha}(Y_i)}\frac{\hat{S}_{\xi^{\alpha}}^{(n)}(p-)}{S_{\xi^{\alpha}}(p)}\cdot\frac{J_n(p)\sum_{j=1}^{n}dM_j^{\xi^{\alpha}}(p)}{Z_n^{\xi^{\alpha}}(p)/n}\right\}\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^{\tilde{\alpha}(t)}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}I_{\{p<\tilde{\alpha}(Y_i)\}}}{\hat{S}_{\xi^{\alpha}}^{(n)}(\tilde{\alpha}(Y_i))}\right\}\frac{\hat{S}_{\xi^{\alpha}}^{(n)}(p-)}{S_{\xi^{\alpha}}(p)}\cdot\frac{n}{Z_n^{\xi^{\alpha}}(p)}dM_j^{\xi^{\alpha}}(p)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^{\tilde{\alpha}(t)}\hat{L}_t(p)\cdot\frac{\hat{S}_{\xi^{\alpha}}^{(n)}(p-)}{S_{\xi^{\alpha}}(p)}\cdot\frac{n}{Z_n^{\xi^{\alpha}}(p)}dM_j^{\xi^{\alpha}}(p)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^{\tilde{\alpha}(t)}\frac{L_t(p)}{P(\xi^{\alpha}\wedge\tilde{\alpha}(Y)\geq p)}dM_j^{\xi^{\alpha}}(p)+o_p(1);
\end{aligned}$$

the third equality comes from [17] (see also [20], p.301), while the last is a consequence of Lemma 5.2.2, the large sample properties of the one-dimensional Kaplan-Meier estimate and the fact that $n/Z_n^{\xi^{\alpha}}(p)\to 1/P(\xi^{\alpha}\wedge\tilde{\alpha}(Y)\geq p)$ a.s.

Now, for any $p\leq\tau$, $t\in T$, let

$$\beta_{t1}^{(n)} = \sqrt{n}(\theta^{(n)}(t)-\mu(t)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{P(Y_i\in\xi_i|Y_i)}-\mu(t)\right\}$$

and

$$\beta_{t2}^{(n)} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^{\tilde{\alpha}(t)}\frac{L_t(s)}{P(\xi^{\alpha}\wedge\tilde{\alpha}(Y)\geq s)}dM_j^{\xi^{\alpha}}(s).$$

Note that $\beta_{t2}^{(n)} = \sqrt{n}(\hat{\mu}_\alpha^{(n)}(t) - \theta^{(n)}(t)) - o_p(1)$, so we have

$$(5.2.6) = \mathrm{Avar}\{\beta_{t1}^{(n)} + \beta_{t2}^{(n)}\} = \mathrm{Avar}\{\beta_{t1}^{(n)}\} + \mathrm{Avar}\{\beta_{t2}^{(n)}\} + 2\lim_{n\to\infty} E(\beta_{t1}^{(n)}\beta_{t2}^{(n)}), \quad (5.2.7)$$

since all terms can be expressed as the sum of i.i.d. random variables. The first term on the right-hand side of (5.2.7) is straightforward:

$$\begin{aligned}
\mathrm{Avar}\{\beta_{t1}^{(n)}\} &= \mathrm{Avar}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{P(Y_i\in\xi_i|Y_i)} - \mu(t)\right)\right\} \\
&= \frac{1}{n}\mathrm{Avar}\left\{\sum_{i=1}^{n}\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{P(Y_i\in\xi_i|Y_i)}\right\} \\
&= \mathrm{Var}\left\{\frac{I_{\{Y\in A_t\}}I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)}\right\}. \qquad (5.2.8)
\end{aligned}$$

For the second, note that $\beta_{t2}^{(n)}$ is a local martingale, so

$$\mathrm{Avar}\{\beta_{t2}^{(n)}\} = \lim_{n\to\infty} E(\beta_{t2}^{(n)2}) = \lim_{n\to\infty} E(\left\langle\beta_{t2}^{(n)}\right\rangle).$$

From (5.2.5) we see that

$$\left\langle M_i^{\xi^\alpha}(p)\right\rangle = \int_0^p I_{\{\xi_i^\alpha\wedge\tilde{\alpha}(Y_i)\geq s\}}dH_{\xi^\alpha}(s)$$

by applying formula II.4.2 in [1], and so

$$\begin{aligned}
\left\langle\beta_{t2}^{(n)}\right\rangle &= \frac{1}{n}\sum_{j=1}^{n}\int_0^{\tilde{\alpha}(t)}\frac{L_t^2(s)I_{\{\xi_i^\alpha\wedge\tilde{\alpha}(Y_i)\geq s\}}}{P^2(\xi^\alpha\wedge\tilde{\alpha}(Y)\geq s)}dH_{\xi^\alpha}(s) \\
&\longrightarrow \int_0^{\tilde{\alpha}(t)}\frac{L_t^2(s)}{P(\xi^\alpha\wedge\tilde{\alpha}(Y)\geq s)}dH_{\xi^\alpha}(s) \quad \text{a.s.} \\
&= \mathrm{Avar}\{\beta_{t2}^{(n)}\} \qquad (5.2.9)
\end{aligned}$$

by uniform integrability.

For the last term on the right-hand side of (5.2.7) we have

$$\begin{aligned}
&E(\beta_{t1}^{(n)}\beta_{t2}^{(n)}) \\
&= \frac{1}{n}E\left\{\sum_{i=1}^{n}\int_0^{\tilde{\alpha}(t)}\frac{L_t(s)}{P(\xi^\alpha\wedge\tilde{\alpha}(Y)\geq s)}\left[\frac{I_{\{Y_i\in A_t\}}I_{\{Y_i\in\xi_i\}}}{P(Y_i\in\xi_i|Y_i)} - \mu(t)\right]dM_i^{\xi^\alpha}(s)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= E\left\{\int_0^{\tilde{\alpha}(t)} \frac{L_t(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} \left[\frac{I_{\{Y\in A_t\}} I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)} - \mu(t)\right] dM^{\xi^\alpha}(s)\right\} \\
&= E\left\{\int_0^{\tilde{\alpha}(t)} \frac{L_t(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} \cdot \frac{I_{\{Y\in A_t\}} I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)} dM^{\xi^\alpha}(s)\right\} \\
&= E\left\{\int_0^{\tilde{\alpha}(t)} \frac{L_t(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} \cdot \frac{I_{\{Y\in A_t\}} I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)} \left\{dN^{\xi^\alpha}(s) - I_{\{\xi^\alpha\wedge\tilde{\alpha}(Y)\geq s\}} dH_{\xi^\alpha}(s)\right\}\right\} \\
&= E\left\{-\int_0^{\tilde{\alpha}(t)} \frac{I_{\{Y\in A_t\}} I_{\{Y\in\xi\}} I_{\{\xi^\alpha\wedge\tilde{\alpha}(Y)\geq s\}}}{P(Y\in\xi|Y)} \cdot \frac{L_t(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} dH_{\xi^\alpha}(s)\right\} \\
&= E\left\{-\int_0^{\tilde{\alpha}(t)} \frac{I_{\{Y\in A_t\}} I_{\{s\leq\tilde{\alpha}(Y)\}} I_{\{\tilde{\alpha}(Y)\leq\xi^\alpha\}}}{P(Y\in\xi|Y)} \cdot \frac{L_t(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} dH_{\xi^\alpha}(s)\right\} \\
&= -\int_0^{\tilde{\alpha}(t)} E\left\{\frac{I_{\{Y\in A_t\}} I_{\{s\leq\tilde{\alpha}(Y)\}} P(Y\in\xi|Y)}{P(Y\in\xi|Y)}\right\} \frac{L_t(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} dH_{\xi^\alpha}(s) \\
&= -\int_0^{\tilde{\alpha}(t)} \frac{L_t^2(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} dH_{\xi^\alpha}(s); \qquad (5.2.10)
\end{aligned}
$$

the fifth equality comes from the fact that $\{\xi^\alpha \leq \tilde{\alpha}(Y)\} \Leftrightarrow \{Y \notin \xi\}$ and so $dN^{\xi^\alpha}(s)$ cancels when multiplied by $I_{\{Y\in\xi\}}$, while the next-to-last is a consequence of Fubini's Theorem and conditioning on Y.

Finally, substitute (5.2.8), (5.2.9) and (5.2.10) in (5.2.7) to get

$$
\mathrm{Avar}\{\sqrt{n}(\hat{\mu}_\alpha^{(n)}(t) - \mu(t))\} = \mathrm{Var}\left\{\frac{I_{\{Y\in A_t\}} I_{\{Y\in\xi\}}}{P(Y\in\xi|Y)}\right\} - \int_0^{\tilde{\alpha}(t)} \frac{L_t^2(s)}{P(\xi^\alpha \wedge \tilde{\alpha}(Y) \geq s)} dH_{\xi^\alpha}(s).
$$

□

In [17], Lu and Burke were able to find a consistent estimator of $\mathrm{Avar}(\sqrt{n}\hat{\mu}_\alpha^{(n)}(t))$ at a fixed point t, and they used it to arrive at confidence bounds for the bivariate distribution function $\mu(t)$. We take a different approach for our examples and applications, since we will be making use of the functional Theorem 5.2.1. Then we will apply the bootstrapping methods presented in previous sections to find critical values for our test statistic.

Alternative Test of Independence and Copula Estimation

We can set up a test of independence by substituting the estimate $\hat{\mu}^{(n)}$ with $\hat{\mu}_\alpha^{(n)}$ in the test of Section 5.1. Then the process used to define the test statistic becomes $U_\alpha^{(n)} = n^{1/2}(\hat{\mu}_\alpha^{(n)} - \hat{\mu}_{\alpha,1}^{(n)}\hat{\mu}_{\alpha,2}^{(n)})$. Care must be taken here, however, since we cannot use $\hat{\mu}_{\alpha,1}^{(n)}(t_1) = \hat{\mu}_\alpha^{(n)}(t_1, 1)$ and $\hat{\mu}_{\alpha,2}^{(n)}(t_2) = \hat{\mu}_\alpha^{(n)}(1, t_2)$ as estimates of the marginals μ_1 and μ_2 respectively. The reason for this is that the estimate $\hat{\mu}_\alpha^{(n)}$ is only valid on $[0,\tau]^2$, whereas $\hat{\mu}^{(n)}$ didn't have that restriction. We could get a different estimate for the marginals, for example, when the data structure is as in [17]: a sequence of i.i.d. random vectors $(U_1, V_1), \ldots, (U_n, V_n)$ subjected to univariate right-censoring by the random variables $C_1, \ldots, C_n$, rendering the observations $(U_i \wedge C_i, V_i \wedge C_i, I_{\{U_i \leq C_i\}}, I_{\{V_i \leq C_i\}})$ for $i = 1, \ldots, n$. There, it is possible to use the one-dimensional Kaplan-Meier estimates for both marginals.

If the new marginal estimates $\hat{\mu}_{\alpha,1}^{(n)}, \hat{\mu}_{\alpha,2}^{(n)}$ converge weakly to mean-zero Gaussian processes Δ^1 and Δ^2 –which is the case if our data structure allows us to use the one-dimensional Kaplan-Meier estimators–, we can use the same arguments used in the previous test, with Theorem 5.2.1 replacing Theorem 5.1.4, to conclude that $n^{1/2}\left(\hat{\mu}_\alpha^{(n)} - \hat{\mu}_{\alpha,1}^{(n)}\hat{\mu}_{\alpha,2}^{(n)} - (\mu - \mu_1\mu_2)\right)$ converges weakly to the mean-zero Gaussian process $\Delta - \mu_2\Delta^1 - \mu_1\Delta^2$, with Δ as in Theorem 5.2.1.

As for bootstrapping, there are more factors to consider, although the principles remain the same. With the T-indexed processes

$$\begin{aligned} N^{(n)\xi^\alpha}(\tilde{\alpha}(t)) &= \sum_{i=1}^n I_{\{\xi_i^\alpha \leq \tilde{\alpha}(t)\}} I_{\{\xi_i^\alpha \leq \tilde{\alpha}(Y_i)\}} = \sum_{i=1}^n I_{\{t \notin \xi_i\}} I_{\{Y_i \notin \xi_i\}} \\ Z_n^{\xi^\alpha}(\tilde{\alpha}(t)) &= \sum_{i=1}^n I_{\{\xi_i^\alpha \geq \tilde{\alpha}(t)\}} I_{\{\tilde{\alpha}(t) \leq \tilde{\alpha}(Y_i)\}} = \sum_{i=1}^n I_{\{t \in \xi_i\}} I_{\{\tilde{\alpha}(t) \leq \tilde{\alpha}(Y_i)\}} \end{aligned}$$

in mind, define both $\mathcal{G}^4 = \{g_t^4 : t \in T\}$, $\mathcal{G}^5 = \{g_t^5 : t \in T\}$, where $g_t^4(Y_i, \xi_i) := I_{\{t \in \xi_i\}} I_{\{\tilde{\alpha}(t) \leq \tilde{\alpha}(Y_i)\}}$ and $g_t^5(Y_i, \xi_i) := I_{\{t \notin \xi_i\}} I_{\{Y_i \notin \xi_i\}}$, and recall the class of functions $\mathcal{G}^2 = \{g_t^2 : t \in T\}$, where $g_t^2(Y_i, \xi_i) := I_{\{Y_i \in A_t\}} I_{\{Y_i \in \xi_i\}}$. Taken together, $(\mathcal{G}^4, \mathcal{G}^5, \mathcal{G}^2)$ is a Donsker class; indeed, the classes $\mathcal{G}^4$ and $\mathcal{G}^5$ are Donsker since they each converge to a Gaussian limit (See Example 3.9.19 in [22]), while it was shown in the proof of Theorem 4.2.7 that $\mathcal{G}^2$ is also a Donsker class. These classes have finite envelope functions because they include indicator functions only. Again, if we have a data structure akin to Lu and Burke's that allows us to obtain the one-dimensional Kaplan-Meier estimates of the marginals, we also have to consider the classes comprised by functions of the type $I_{\{t_1 \leq C_i\}} I_{\{t_1 \leq U_i)\}}$, $I_{\{t_2 \leq C_i\}} I_{\{t_2 \leq V_i)\}}$, $I_{\{U_i \leq C_i\}} I_{\{U_i \leq t_1\}}$ and $I_{\{V_i \leq C_i\}} I_{\{V_i \leq t_2\}}$, where $t = (t_1, t_2)$. However, all theses classes taken together are still jointly Donsker since they converge to Gaussian limits (again, see Example 3.9.19 in [22]) and come from i.i.d. random vectors.

Assuming complete observability of the pairs (Y_i, ξ_i) for a moment, we note that the composition map (5.2.3) is Hadamard-differentiable on the domain described in the proof of Theorem 5.2.1, so we again apply Theorem 3.9.11 in [22] to get the conditional convergence given (Y_i, ξ_i) of the bootstrapped estimator to the Gaussian process $\Delta - \mu_2 \Delta^1 - \mu_1 \Delta^2$. Of course, as in previous instances, we can remove the assumption of complete observability because our estimate $\hat{\mu}_\alpha^{(n)}$ depends only on the indicators $q_i = I_{\{Y_i \in A_t\}} I_{\{Y_i \in \xi_i\}}$, $r_i = I_{\{t \in \xi_i\}} I_{\{\tilde{\alpha}(t) \leq \tilde{\alpha}(Y_i)\}}$ and $s_i = I_{\{t \notin \xi_i\}} I_{\{Y_i \notin \xi_i\}}$, so we can in fact bootstrap directly from the observed values q_i, r_i and s_i. The bootstrapped versions of the marginals are constructed similarly by using Efron's procedure ([6]; this is also discussed briefly in [1], p.322-323). Therefore, if $\ddot{\mu}_\alpha^{(n)}, \ddot{\mu}_{\alpha,1}^{(n)}, \ddot{\mu}_{\alpha,2}^{(n)}$ denote the bootstrapped version of the joint and marginal estimators of the distribution, respectively, we have that

$$\sup_{\tilde{\alpha}(t) \leq \tau} \left| n^{1/2} [\ddot{\mu}_\alpha^{(n)}(t_1, t_2) - \ddot{\mu}_{\alpha,1}^{(n)}(t_1) \ddot{\mu}_{\alpha,2}^{(n)}(t_2) - (\hat{\mu}_\alpha^{(n)}(t_1, t_2) - \hat{\mu}_{\alpha,1}^{(n)}(t_1) \hat{\mu}_{\alpha,2}^{(n)}(t_2))] \right| \tag{5.2.11}$$

converges conditionally in outer probability to the sup of the absolute value of the Gaussian process $\Delta - \mu_2 \Delta^1 - \mu_1 \Delta^2$. Then we can bootstrap from (q_i, r_i, s_i) jointly and find c_α so that $(1-\alpha)100\%$ of the absolute values of (5.2.11) fall below c_α, which would give us $(1-\alpha)100\%$ uniform confidence bands for the difference $\mu - \mu_1 \mu_2$.

For a test of independence $H_\circ : \mu = \mu_1 \mu_2$ vs. $H_\circ : \mu \neq \mu_1 \mu_2$, note that under the null hypothesis, $P\left(\sup \left| U_\alpha^{(n)}(t_1, t_2) \right| < c_\alpha \right) \approx 1 - \alpha$. Therefore, we reject $H_\circ$ if $\sup n^{1/2} \left| \hat{\mu}_\alpha^{(n)}(t_1, t_2) - \hat{\mu}_{\alpha,1}^{(n)}(t_1) \hat{\mu}_{\alpha,2}^{(n)}(t_2) \right| > c_\alpha$.

Similarly, we can give an alternative to the estimate given in the previous Section for the copula $C(p,q) = F(F_1^{-1}(p), F_2^{-1}(q))$, for (p,q) such that $\tilde{\alpha}(F_1^{-1}(p), F_2^{-1}(q)) \leq \tau$, simply by replacing $\hat{\mu}^{(n)}$, $\hat{\mu}_1^{(n)}$ and $\hat{\mu}_2^{(n)}$ with our new estimates –again, as long as we can construct Kaplan-Meier type estimates for the marginals. For example, in Lu and Burke's scheme, $\xi_i = [0, C_i]^2$, and the resulting estimate $\hat{C}_\alpha^{(n)}(p,q) = \hat{\mu}_\alpha^{(n)}(\hat{\mu}_{\alpha,1}^{(n)-}(p), \hat{\mu}_{\alpha,2}^{(n)-}(q))$ depends only on $\left(\frac{Z_n^{\xi^\alpha}}{n}, \frac{N^{(n)\xi^\alpha}}{n}, \frac{N^{(n)\xi}}{n}, \frac{N_1^{(n)C}}{n}, \frac{N_2^{(n)C}}{n}, \frac{Z_{n,1}}{n}, \frac{Z_{n,2}}{n}\right)$, with $N_j^{(n)C}, Z_{n,j}$ as defined in (4.3.1) for $j = 1, 2$. The composition map

$$\begin{aligned}
\left(\frac{Z_n^{\xi^\alpha}}{n}, \frac{N^{(n)\xi^\alpha}}{n}, \frac{N^{(n)\xi}}{n}, \frac{N_1^{(n)C}}{n}, \frac{N_2^{(n)C}}{n}, \frac{Z_{n,1}}{n}, \frac{Z_{n,2}}{n}\right) &\xrightarrow{\phi_5} (\hat{\mu}_\alpha^{(n)}, \hat{\mu}_{\alpha,1}^{(n)}, \hat{\mu}_{\alpha,2}^{(n)}) \\
&\xrightarrow{\phi_3} (\hat{\mu}_\alpha^{(n)}, \hat{\mu}_{\alpha,1}^{(n)-}, \hat{\mu}_{\alpha,2}^{(n)-}) \\
&\xrightarrow{\phi_4} \hat{\mu}_\alpha^{(n)}(\hat{\mu}_{\alpha,1}^{(n)-}, \hat{\mu}_{\alpha,2}^{(n)-})
\end{aligned}$$

is Hadamard-differentiable on a domain analogous to map (5.2.3): the map ϕ_5 is Hadamard-differentiable because of the proof of Theorem 5.2.1 and Example 3.9.19 in [22], while ϕ_3 and ϕ_4 are Hadamard-differentiable by Example 3.9.24 and Lemma 3.9.25, respectively, in [22]. Then, the Delta-Method states that the process $W_\alpha^{(n)} := \sqrt{n}(\hat{C}_\alpha^{(n)}(\cdot) - C(\cdot))$ converges in $l^\infty[0,\tau]^2$ to a mean-zero Gaussian limit $\tilde{W}$. The bootstrap works as in the case of the test of independence, since $\hat{C}_\alpha^{(n)}$ depends only on $\left(\frac{Z_n^{\xi^\alpha}}{n}, \frac{N^{(n)\xi^\alpha}}{n}, \frac{N^{(n)\xi}}{n}, \frac{N_1^{(n)C}}{n}, \frac{N_2^{(n)C}}{n}, \frac{Z_{n,1}}{n}, \frac{Z_{n,2}}{n}\right)$ through a Hadamard-differentiable composition, so Theorem 3.9.11 in [22] guarantees the conditional convergence given (Y_i, ξ_i) of the bootstrapped estimator to the same Gaussian limit $\tilde{W}$.

Chapter 6

Path-dependent Estimators of the Distribution Function

The preceeding Chapter yielded two estimators of the distribution function of a random sample under some types of censorings. We are now interested in further exploring this topic; however, we want to move in a different direction that will address some limitations inherent to the structures considered in Chapter 5.

A disadvantage of the reverse probability estimators (5.1.2) and (5.2.2) lies in the specific censoring mechanism they demand, which restricts their applicability. Furthermore, even though their construction is simple, they require the entire data set; in other words, the information up to time t is insufficient for them to produce an estimate at that point.

It is our intention to construct an estimator under more general data structures, both in terms of the censoring mechanisms being considered, as well as the very space in which the data points lie. We would also like our estimator to be dynamic, meaning that we would only need the information up to the point where the distribution function is being estimated. Our approach will be based on the transformation of the original data set into one-dimensional data via a map that we call a flow, or path; this is similar to the techniques we employed in Section 5.2, as we noted at the beginning of that Section.

6.1 Path-dependent Estimators of $1 - F$ and S

For this Chapter, we will consider a complete separable metric space (T, d). Recall that a complete distributive lattice is a partially ordered set B such that every subset of B has a sup and an inf, and for $u, s, t \in B$,

$$u \vee (s \wedge t) = (u \vee s) \wedge (u \vee t) \text{ and } u \wedge (s \vee t) = (u \wedge s) \vee (u \wedge t).$$

We will be working under Assumptions 2.1-2.4 found in [13], which are the following:

Assumption 6.1.1 *([13], Assumption 2.1) T is endowed with a partial order '$\leq$' for which T is a complete distributive lattice*

In particular, T contains a minimum element denoted by $\mathbf{0}$, as well as a maximum denoted by $\mathbf{1}$.

Recall from Definition 2.1.2 that for any $t \in T$, $A_t = \{s \in T : s \leq t\}$ denotes the strict past of such point, and $E_t = \{s \in T : s \geq t\}$ represents its future.

Assumption 6.1.2 *([13], Assumption 2.2) For every $t \in T$, both A_t and E_t are d-closed.*

Assumption 6.1.3 *([13], Assumption 2.3) The partial order on T is sufficiently rich that there exists an increasing sequence of finite sublattices T_n of T each containing $\mathbf{0}$ and $\mathbf{1}$, such that for any $t \in T$, $t \neq \mathbf{0}, \mathbf{1}$, each open neighborhood of t contains points $t', t'' \in \cup T_n$ distinct from t such that $t' < t < t''$. Each open neighborhood of $\mathbf{0}$ (respectively, $\mathbf{1}$) contains a point $t' \in \cup T_n$ distinct from $\mathbf{0}$ (respectively, $\mathbf{1}$). The elements of T_n can be thought of as the dyadics of order n.*

Assumption 6.1.4 *([13], Assumption 2.4)*

1. *For $u, s, t \in T$, $d(s \wedge u, t \wedge u) \leq d(s, t)$ and $d(s \vee u, t \vee u) \leq d(s, t)$.*

2. The sublattices T_n of Assumption 6.1.3 can be chosen so that

$$\lim_n \sup_{t \in T} d(t, t_n^+) = 0 \quad and \quad \lim_n \sup_{t \in T} d(t, t_n^-) = 0,$$

where $t_n^+ := \bigwedge_{t' \in T_n, t \leq t'} t'$ and $t_n^- := \bigvee_{t' \in T_n, t' \leq t} t'$.

Let Y be a T-valued random variable with continuous distribution function F. Furthermore, let $Y_1, \ldots, Y_n$ be a random sample from that distribution, and for the time being we will assume that there is no censoring involved. As was done in [13], define a flow γ on T as a continuous strictly increasing map from $[0, 1]$ to T, and for any point $t \in T$, we will say that $t \in \gamma \Leftrightarrow \gamma(u) = t$ for some $u \in [0, 1]$. Throughout this Chapter, we use the terms 'flow' and 'path' interchangeably. It is pointed out in [13] that Assumptions 6.1.3 and 6.1.4 imply that, given any finite totally ordered subset $t_0 < t_1 < \ldots < t_n$ of T, there exists a flow γ connecting $t_0, t_1, \ldots, t_n$; in other words, $t_i \in \gamma$ for $i = 0, 1, \ldots, n$.

Definition 6.1.5 *Given any flow γ connecting **0** to **1**, let Y^γ and $\bar{Y}^\gamma$ be defined by*

$$Y^\gamma := \inf_{t \in \gamma} \{Y \leq t\},$$

$$\bar{Y}^\gamma := \sup_{t \in \gamma} \{t \leq Y\}.$$

We can think of Y^γ and $\bar{Y}^\gamma$ as two different projections of the random variable Y along the flow γ: Y^γ is the smallest point on γ such that $Y \leq Y^\gamma$, while $\bar{Y}^\gamma$ represents the greatest point on γ such that $Y \geq \bar{Y}^\gamma$. These projections are illustrated in Figures 6.1 and 6.2, respectively. We will focus first on Y^γ, which will be the basis for our estimation of the distribution F.

Let $\tau_i^\gamma := \gamma^{-1}(Y_i^\gamma)$ for $i = 1, \ldots, n$, and suppose $\gamma(u) = t$ for $u \in [0, 1]$. An easy calculation of the distribution function F^γ of these one-dimensional random variables

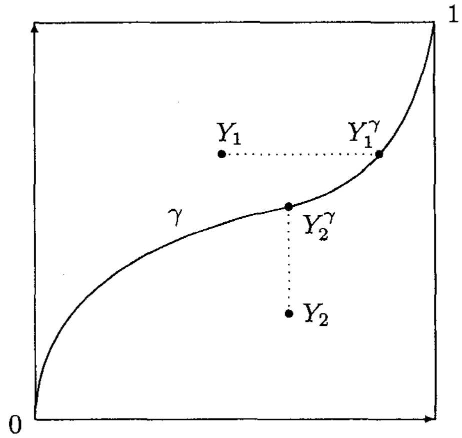

Figure 6.1: Y_i, Y_i^γ.

yields

$$\begin{aligned} P(\tau_i^\gamma \leq u) &= P(\gamma(\tau_i^\gamma) \leq \gamma(u)) \\ &= P(Y_i^\gamma \in A_{\gamma(u)}) \\ &= P(Y_i \in A_{\gamma(u)}) \\ &= F(\gamma(u)), \end{aligned}$$

hence $F^\gamma(u) = F(\gamma(u)) = F(t)$. The first equality comes from the fact that γ is strictly increasing, while the second and the third are consequences of the definitions of τ_i^γ and Y_i^γ, respectively. This means that if we somehow manage to estimate $F^\gamma(u)$, we will automatically have an estimate of $F(t)$, the original distribution function evaluated at the point $\gamma(u) = t$. The interesting part is that $\tau_1^\gamma, \ldots, \tau_n^\gamma$ are one-dimensional, so their distribution can be estimated using the classical one-dimensional methods, as outlined in the Appendix.

Define the processes

$$Z_n^\gamma(u) = \sum_{i=1}^n I_{\{\tau_i^\gamma \geq u\}} = \sum_{i=1}^n I_{\{Y_i^\gamma \geq \gamma(u)\}} = \sum_{i=1}^n I_{\{Y_i \notin A^\circ_{\gamma(u)}\}},$$

and

$$N_n^\gamma(u) = \sum_{i=1}^n I_{\{\tau_i^\gamma \leq u\}} = \sum_{i=1}^n I_{\{Y_i^\gamma \leq \gamma(u)\}} = \sum_{i=1}^n I_{\{Y_i \in A_{\gamma(u)}\}}.$$

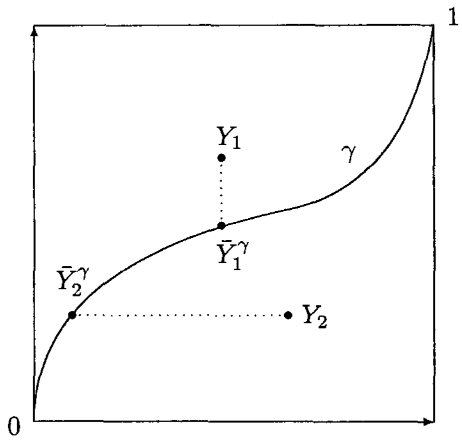

Figure 6.2: Y_i, $\bar{Y}_i^\gamma$.

Then it is well-known (see, for example, Section II.1 of [1]) that $N_n^\gamma(u) - \Lambda_n^\gamma(u)$ is a one-dimensional martingale with respect to $\mathcal{F}_\circ^\gamma(u)$, the minimal filtration generated by $\tau_1^\gamma, \ldots, \tau_n^\gamma$, where

$$\Lambda_n^\gamma(u) = \int_0^u Z_n^\gamma(u) \cdot \frac{dF^\gamma(v)}{1 - F^\gamma(v)}, \tag{6.1.1}$$

and that the Nelson-Aalen estimator of the integrated hazard $H^\gamma(u) = -\ln(1 - F^\gamma(u))$ $= -\ln(1 - F(t))$ is

$$\hat{H}^\gamma(u) = \int_0^u \frac{dN_n^\gamma(v)}{Z_n^\gamma(v)} = \sum_{i=1}^n \frac{I_{\{\tau_i^\gamma \le u\}}}{Z_n^\gamma(\tau_i^\gamma)}.$$

Similarly, the Kaplan-Meier estimate for the one-dimensional survival function $1 - F^\gamma(u)$ is given by

$$\prod_{i=1}^n \left(1 - \frac{I_{\{\tau_i^\gamma \le u\}}}{Z_n^\gamma(\tau_i^\gamma)}\right), \tag{6.1.2}$$

which also serves as an estimate of $1 - F(t)$, since $1 - F^\gamma(u) = 1 - F(\gamma(u)) = 1 - F(t)$. This estimate is, as expected, equal to the empirical distribution evaluated on $(A_{\gamma(u)})^c = (A_t)^c$, a fact that becomes clear when we rewrite (6.1.2) in the original space T as

$$(6.1.2) = \prod_{i=1}^n \left(1 - \frac{I_{\{Y_i \in A_t\}}}{\sum_{j=1}^n I_{\{Y_j \notin A^\circ_{Y_i^\gamma}\}}}\right). \tag{6.1.3}$$

It is important to note that the estimate does not depend on γ: if we select another flow, the order of the terms in the right-hand side of (6.1.3) might change, but the product will stay the same.

If we use the projection $\bar{Y}^{\gamma}$, we can arrive at an estimator for the survival function $S(t) = P(Y > t)$ in a similar way. Define $\bar{\tau}_i^{\gamma} := \gamma^{-1}(\bar{Y}_i^{\gamma})$ for $i = 1, \ldots, n$, and as before, suppose that $\gamma(u) = t$ for $u \in [0,1]$. The survival function for $\bar{\tau}^{\gamma}$ is given by

$$\begin{aligned} P(\bar{\tau}^{\gamma} > u) &= P(\gamma(\bar{\tau}^{\gamma}) > \gamma(u)) \\ &= P(\bar{Y}_i^{\gamma} \in E_{\gamma(u)}) \\ &= P(Y_i \in E_{\gamma(u)}) \\ &= S(\gamma(u)) \\ &= S(t). \end{aligned}$$

Now, since $\bar{\tau}^{\gamma}$ is unidimensional, we can write $P(\bar{\tau}^{\gamma} > u) = 1 - P(\bar{\tau}^{\gamma} \leq u) = S(t)$, and rearranging terms we get $\bar{F}^{\gamma}(u) = 1 - S(t)$. This is in contrast to the results we previously got by using the projection Y^{γ}, namely that $F^{\gamma}(u) = F(t)$; hence an estimate for $\bar{F}^{\gamma}(u)$ results in an estimate of $1 - S(t)$. Again, the advantage is that we can use one-dimensional methods for the estimation of this function.

Before we continue though, we have to note that because of the very nature of the projections $\bar{Y}_i^{\gamma}$, we will now require a slightly different data structure. When we were working with Y^{γ}, the only thing we had to know about the original observations Y_i was whether or not they belonged to the strict past of the point t, and therefore we could use the minimal filtration $\mathcal{F}(t) = \sigma\{I_{\{Y_i \leq z\}}, z \in A_t\}$. By contrast, in the case of $\bar{Y}^{\gamma}$ we have to be able to know whether the observations Y_i lie somewhere in the future of t. Therefore, the filtration $\mathcal{F}^*(t) = \sigma\{I_{\{Y_i \leq z\}}, z \in D_t\}$ for $i = 1, \ldots, n$ fits our requirements.

Now we can proceed as before by defining the usual processes

$$\bar{Z}_n^{\gamma}(u) = \sum_{i=1}^{n} I_{\{\bar{\tau}_i^{\gamma} \geq u\}} = \sum_{i=1}^{n} I_{\{\bar{Y}_i^{\gamma} \geq \gamma(u)\}} = \sum_{i=1}^{n} I_{\{Y_i \in E_{\gamma(u)}\}},$$

$$\bar{N}_n^\gamma(u) = \sum_{i=1}^n I_{\{\bar{\tau}_i^\gamma \le u\}} = \sum_{i=1}^n I_{\{\bar{Y}_i^\gamma \le \gamma(u)\}} = \sum_{i=1}^n I_{\{Y_i \notin E^\circ_{\gamma(u)}\}},$$

and by noting that $\bar{N}_n^\gamma(u) - \bar{\Lambda}_n^\gamma(u)$ is a one-dimensional martingale with respect to $\bar{\mathcal{F}}_\circ^\gamma(u)$, the minimal filtration generated by $\bar{\tau}_1^\gamma, \ldots, \bar{\tau}_n^\gamma$, where

$$\bar{\Lambda}_n^\gamma(u) = \int_0^u \bar{Z}_n^\gamma(u) \cdot \frac{d\bar{F}^\gamma(v)}{1-\bar{F}^\gamma(v)}. \tag{6.1.4}$$

The Nelson-Aalen estimator of $-\ln(1-\bar{F}^\gamma(u))$ is given by

$$\int_0^u \frac{d\bar{N}_n^\gamma(v)}{\bar{Z}_n^\gamma(v)} = \sum_{i=1}^n \frac{I_{\{\bar{\tau}_i^\gamma \le u\}}}{\bar{Z}_n^\gamma(\bar{\tau}_i^\gamma)}, \tag{6.1.5}$$

but since we know that $\bar{F}^\gamma(u) = 1 - S(t)$, (6.1.5) actually provides an estimate for $-\ln(S(t))$. Finally, the Kaplan-Meier estimate for $1-\bar{F}^\gamma(u) = S(t)$ is

$$\prod_{i=1}^n \left(1 - \frac{I_{\{\bar{\tau}_i^\gamma \le u\}}}{\bar{Z}_n^\gamma(\bar{\tau}_i^\gamma)}\right), \tag{6.1.6}$$

which again agrees with intuition: (6.1.6) is equal to

$$\prod_{i=1}^n \left(1 - \frac{I_{\{Y_i \notin E^\circ_{(s,t)}\}}}{\sum_{j=1}^n I_{\{Y_j \in E_{\bar{Y}_i^\gamma}\}}}\right),$$

so this is simply the empirical distribution evaluated on $E_{\gamma(u)} = E_t$. As was the case with (6.1.3), the above estimate is independent of the choice of flow.

Remark 6.1.6 We mentioned that $N_n^\gamma(u) - \Lambda_n^\gamma(u)$ is a one-dimensional martingale with respect to $\mathcal{F}_\circ^\gamma(u)$; in other words, that $\Lambda_n^\gamma(\cdot)$ is the compensator for the process $N_n^\gamma(\cdot)$ given the information in $\mathcal{F}_\circ^\gamma(\cdot)$. Fortunately, the compensator doesn't change if we substitute $\mathcal{F}_\circ^\gamma(\cdot)$ by the larger filtration $\mathcal{F}^\gamma(u) := \mathcal{F}(\gamma(u)) = \sigma\{I_{\{Y_i \le z\}}, z \le \gamma(u)\}$. Indeed, we have that

$$\begin{aligned} & E(N_n^\gamma(u+h) - N_n^\gamma(u) | \mathcal{F}^\gamma(u)) \\ = \;& E\Big(\sum_{i=1}^n I_{\{u < \tau_i^\gamma \le u+h\}} | \mathcal{F}^\gamma(u)\Big) \\ = \;& \sum_{i=1}^n \frac{F(\gamma(u+h)) - F(\gamma(u))}{1 - F(\gamma(u))} \cdot I_{\{\tau_i^\gamma > u\}}. \end{aligned}$$

We can use an identical argument to show that the compensator (6.1.4) also doesn't change when we replace $\bar{\mathcal{F}}^{\gamma}_{\circ}(u)$ by $\bar{\mathcal{F}}^{\gamma}(u) := \mathcal{F}^{*}(\gamma(u))$. This is going to be very useful as we introduce censoring, since we will be able to build anti-clouds into $\mathcal{F}^{\gamma}$ and $*$-anti-clouds into $\bar{\mathcal{F}}^{\gamma}$.

The Censored Model

The next step is to incorporate censoring into the model. Let $\mathcal{F}$ be a filtration such that $\mathcal{F}^{Y} \subseteq \mathcal{F}$, and consider a flow γ on T. Let $\xi_1, \ldots, \xi_n$ be a sequence of $\mathcal{F}$-anti-clouds weakly independent of $Y_1, \ldots, Y_n$, and suppose that for each i, Y_i is observable only inside ξ_i. Because of the observability issues noted in Chapter 4 –which provided the motivation for the $*$-sets–, we will not be able to use all the observations inside the anti-clouds to construct our estimators; nevertheless, depending on whether we pursue an estimate of $1 - F$ or the survival function S, the "usable" observations will vary. Obviously, the data structure will also play a very important role in this matter, as will the choice of γ.

For the moment though, we will put aside our concerns regarding observability; we feel it is best to focus first on how to project both the data points Y_i and the anti-clouds ξ_i on the flow in order to construct the full one-dimensional model. With the overall picture taken care of, we can then elaborate on the specific regions where "usable" observations lie as we treat different types of data structure.

First, we will discuss the estimation of $1-F$. Suppose $\gamma(u) = t$ for some $u \in [0, 1]$. Using Definition 6.1.5, we can obtain $Y_1^{\gamma}, \ldots, Y_n^{\gamma}$ from our original observations, and letting $\tau_i^{\gamma} := \gamma^{-1}(Y_i^{\gamma})$ for $i = 1, \ldots, n$, we get $\tau_1^{\gamma}, \ldots, \tau_n^{\gamma}$. We project the censoring on the flow by defining $\xi_i^{\gamma} := \{u \in [0, 1] : \gamma(u) \in \xi_i\}$ for $i = 1, \ldots, n$. The sets ξ_i^{γ} are $\mathcal{F}^{\gamma}$-anti-clouds, since

$$\{u \in \xi_i^{\gamma}\} = \{\gamma(u) \in \xi_i\} \in \mathcal{F}(\gamma(u)) = \mathcal{F}^{\gamma}(u).$$

Define

$$Z_n^{\gamma}(u) = \sum_{i=1}^{n} I_{\{\tau_i^{\gamma} \geq u,\ u \in \xi_i^{\gamma}\}}$$

and

$$N_n^\gamma(u) = \sum_{i=1}^n I_{\{\tau_i^\gamma \leq u,\ \tau_i^\gamma \in \xi_i^\gamma\}}.$$

Then, recalling that $F^\gamma(u) = F(\gamma(u))$, the one-dimensional Nelson-Aalen estimator

$$\int_0^u \frac{dN_n^\gamma(v)}{Z_n^\gamma(v)} = \sum_{i=1}^n \frac{I_{\{\tau_i^\gamma \leq u,\ \tau_i^\gamma \in \xi_i^\gamma\}}}{Z_n^\gamma(\tau_i^\gamma)}$$

gives us an estimate of $-\ln(1-F(\gamma(u))) = -\ln(1-F(t))$. Similarly, the Kaplan-Meier estimate for the one-dimensional survival function $1-F^\gamma(u)$,

$$\prod_{i=1}^n \left(1 - \frac{I_{\{\tau_i^\gamma \leq u,\ \tau_i^\gamma \in \xi_i^\gamma\}}}{Z_n^\gamma(\tau_i^\gamma)}\right), \tag{6.1.7}$$

provides us with an estimate of $1-F(\gamma(u)) = 1-F(t)$.

As is to be expected, the construction of the estimate for the survival function S follows along the same lines. In this case, suppose the ξ_i's are $*$-anti-clouds with respect to $\mathcal{F}$; we already know from Chapter 4 that we can't use the anti-clouds for estimating S. We project the observations $Y_1, \ldots, Y_n$ on the flow γ via Definition 6.1.5 to arrive at $\bar{Y}_1^\gamma, \ldots, \bar{Y}_n^\gamma$, and then use the equation $\bar{\tau}_i^\gamma := \gamma^{-1}(\bar{Y}_i^\gamma)$ for $i = 1, \ldots, n$ to get $\bar{\tau}_1^\gamma, \ldots, \bar{\tau}_n^\gamma$. The projection for the censoring, on the other hand, remains $\xi_i^\gamma = \{u \in [0,1] : \gamma(u) \in \xi_i\}$ for $i = 1, \ldots, n$. The projection of these $*$-anti-clouds are anti-clouds with respect to $\bar{\mathcal{F}}^\gamma$, since

$$\{u \in \xi_i^\gamma\} = \{\gamma(u) \in \xi_i\} \in \mathcal{F}^*(\gamma(u)) = \bar{\mathcal{F}}^\gamma(u).$$

The processes $\bar{Z}_n^\gamma$ and $\bar{N}_n^\gamma$ become

$$\bar{Z}_n^\gamma(u) = \sum_{i=1}^n I_{\{\bar{\tau}_i^\gamma \geq u,\ u \in \xi_i^\gamma\}},$$

$$\bar{N}_n^\gamma(u) = \sum_{i=1}^n I_{\{\bar{\tau}_i^\gamma \leq u,\ \bar{\tau}_i^\gamma \in \xi_i^\gamma\}},$$

and keeping in mind that $\bar{F}^\gamma(u) = 1 - S(\gamma(u)) = 1 - S(t)$,

$$\int_0^u \frac{d\bar{N}_n^\gamma(v)}{\bar{Z}_n^\gamma(v)} = \sum_{i=1}^n \frac{I_{\{\bar{\tau}_i^\gamma \leq u,\ \bar{\tau}_i^\gamma \in \xi_i^\gamma\}}}{\bar{Z}_n^\gamma(\bar{\tau}_i^\gamma)}$$

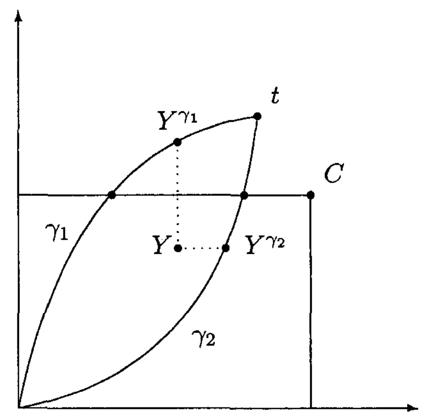

Figure 6.3: Y is recorded as a failure in the flow γ_2; it is censored in γ_1.

and

$$\prod_{i=1}^{n}\left(1-\frac{I_{\{\bar{\tau}_i^\gamma\le u,\ \bar{\tau}_i^\gamma\in\xi_i^\gamma\}}}{\bar{Z}_n^\gamma(\bar{\tau}_i^\gamma)}\right)$$

are the estimates for $-\ln(1-\bar{F}^\gamma(u)) = -\ln(S(t))$ and $1-\bar{F}^\gamma(u) = S(t)$, respectively.

As we mentioned before, the choice of estimating $1-F$ or S will have an impact on the number of observations that can be used to construct our estimates. In addition, the choice of γ can also affect the number of uncensored observations; an example of this can be found in Figure 6.3. At this point, we will consider four different types of data structure and determine how this choice affects the information available to us. In each scenario, assume γ denotes a given flow.

• Scenario 1: Suppose we are given a sequence of i.i.d. T-valued random variables (Y_i), as well as a sequence (ξ_i) of $\mathcal{F}$-anti-clouds independent of the Y_i's, where $\mathcal{F}=\mathcal{F}^Y\vee\mathcal{F}^\xi$; we also assume that for each i, Y_i is only observable inside ξ_i. For an illustration, see Figure 6.4.

If we want to estimate $1-F$, we are interested in determining whether or not any observation has actually happened before $t\in T$, and so we do not need to discard as much information as when we estimated the cumulative hazard in Chapter 4; there, we had to be able to assert, for any observable t, if Y was strictly in the future of such

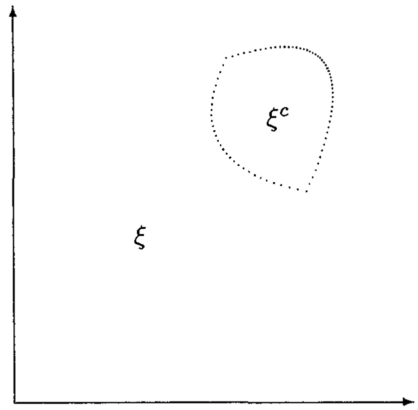

Figure 6.4: Scenario 1.

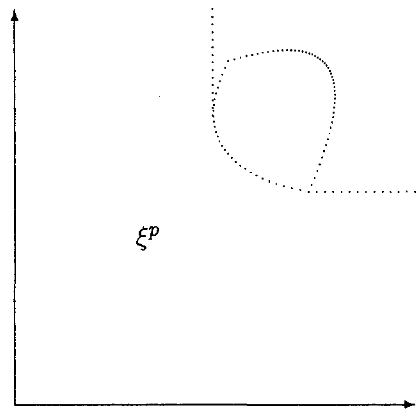

Figure 6.5: The anti-cloud ξ^p.

point. This relaxed standard leads us to define the sets $\xi_i^p := \{t \in T : A_t \cap \xi_i^c = \emptyset\}$ for $i = 1, \ldots, n$, which are still $\mathcal{F}$-anti-clouds, since $\{t \in \xi_i^p\} = \{A_t \cap \xi_i^c = \emptyset\} \in \mathcal{F}(t)$ for any $t \in T$. We can take a look at one such set in Figure 6.5. As we did before, let $\mathcal{F}^\gamma(u) := \mathcal{F}(\gamma(u))$. The projections of these sets are given by $\xi_i^\gamma := \{u \in [0,1] : \gamma(u) \in \xi_i^p\}$ for $i = 1, \ldots, n$, which are anti-clouds with respect to $\mathcal{F}^\gamma$.

In case we want to construct an estimate for the survival function S, we are forced to discard a lot of information, since the question now becomes the position of the observations with respect to the strict future of t or, equivalently, the ex-

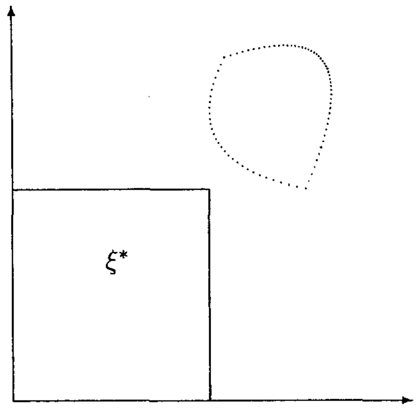

Figure 6.6: The $*$-anti-cloud ξ^*.

tended past of t. Therefore, we have to define the sets $\xi_i^* := \{t \in T : D_t \cap \xi_i^c = \emptyset\}$ for $i = 1, \ldots, n$, depicted in Figure 6.6, which we showed in Chapter 4 are $*$-anti-clouds; in other words, $\{t \in \xi_i^*\} \in \mathcal{F}^*(t)$, where $\mathcal{F}^*(t) = \sigma\{I_{\{z \leq Y_i\}}, t \leq z\}$. Let $\bar{\mathcal{F}}^\gamma(u) := \mathcal{F}^*(\gamma(u))$. The projections of these $*$-anti-clouds will be the $\bar{\mathcal{F}}^\gamma$-anti-clouds $\bar{\xi}_i^\gamma := \{u \in [0,1] : \gamma(u) \in \xi_i^*\}$ for $i = 1, \ldots, n$.

• Scenario 2: Recall the data structure found in Example 4.1.8. Let $(Y_i) = (Y_{1,i}, Y_{2,i})$ be a sequence of i.i.d. $[0,1]^2$-valued random variables. Assume there exists a filtration $\mathcal{F}$ such that $\mathcal{F}_t^Y \subseteq \mathcal{F}_t$ for every $t = (t_1, t_2) \in [0,1]^2$, and define $\mathcal{F}_{t_1}^{(1)} = \bigvee_{t_2} \mathcal{F}_{(t_1,t_2)}$, $\mathcal{F}_{t_2}^{(2)} = \bigvee_{t_1} \mathcal{F}_{(t_1,t_2)}$. We are given a finite sequence of $\mathcal{F}^{(j)}$-stopping times $0 = \eta_{j1}^{(i)} \leq \nu_{j1}^{(i)} \leq \eta_{j2}^{(i)} \leq \nu_{j2}^{(i)} \leq \ldots \leq \eta_{jn_j}^{(i)} \leq \nu_{jn_j}^{(i)}$ for $i = 1, 2, \ldots$ and $j = 1, 2$. We can observe

$$\mathcal{H}_t = \{I_{\{Y_{j,i} \leq s_j, Y_{j,i} \in \bigcup_{k=1}^{n_j}[\eta_{jk}^{(i)}, \nu_{jk}^{(i)}]\}}, I_{\{Y_{j,i} \in (\nu_{jk}^{(i)}, \rho_{j,k+1}^{(i)}), \rho_{j,k+1} \leq u_j\}} : s, u \in A_t; j = 1, 2\},$$

where $s = (s_1, s_2)$ and $u = (u_1, u_2)$. It is easy to see that $\xi_i = \{\bigcup_{k=1}^{n_1}[\eta_{1k}^{(i)}, \nu_{1k}^{(i)}] \times \bigcup_{k=1}^{n_2}[\eta_{2k}^{(i)}, \nu_{2k}^{(i)}]\}$ is an anti-cloud with respect to $\mathcal{F}$ for $i = 1, 2, \ldots$. A graphic example of this situation is found in Figure 6.7.

In this particular case we can use all the information contained in the anti-clouds to construct both estimates, because as soon as we emerge from the censored regions, we are able to determine the location of the observations with respect to that point in

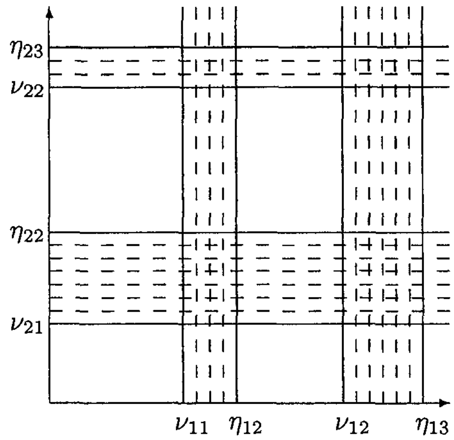

Figure 6.7: Scenario 2. The unshaded areas represent the anti-cloud ξ.

time; in other words, the data structure allows us to know whether any given observation lies in A_t, D_t or E_t for any $t \in \xi_i$. Hence, the projections of these anti-clouds on the flow γ are defined by $\xi_i^\gamma := \{u \in [0,1] : \gamma(u) \in \xi_i\}$ for $i = 1, \ldots, n$ in both the estimates of $1 - F$ and S. Again, if we define $\mathcal{F}^\gamma(u) := \mathcal{F}(\gamma(u)) = \mathcal{F}(t_1, t_2)$, the sets ξ_i^γ are $\mathcal{F}^\gamma$-anti-clouds.

• Scenario 3: Recall the medical example of Chapter 4, where we had times of start of pregnancy and times of onset of tuberculosis (See Figure 6.8). Consider a sequence of i.i.d. $[0,1]^2$-valued random variables $(Y_i) = (Y_{1,i}, Y_{2,i})$ and a sequence i.i.d. $[0,1]$-valued random variables $\kappa_{1,i} < \kappa_{2,i}$ for $i = 1, 2, \ldots$. The relevant filtrations are $\mathcal{F}_{t_1}^{\kappa_1,\kappa_2} = \sigma\{I_{\{\kappa_{1,i} \le s_1\}}, I_{\{\kappa_{2,i} \le s_1\}} : s_1 \le t_1\}$, $\mathcal{F}_{(t_1,t_2)} := \mathcal{F}^Y_{(t_1,t_2)} \vee \mathcal{F}_{t_1}^{\kappa_1,\kappa_2}$, $\mathcal{F}^{(1)}_{t_1} = \bigvee_{t_2} \mathcal{F}_{(t_1,t_2)}$ and $\mathcal{F}^{(2)}_{t_2} = \bigvee_{t_1} \mathcal{F}_{(t_1,t_2)}$. We noted in Chapter 4 that if we let $\rho_i := \{(t_1,t_2) : \kappa_{1,i} < t_1 < \kappa_{2,i},\ t_1 < t_2 < t_1 + c\}$ represent the censored regions, where c denotes a constant, then (ρ_i) is a sequence of clouds with respect to $\mathcal{F}$. It was also mentioned that taken separately, the $Y_{1,i}$'s are not censored at all, but the $Y_{2,i}$'s are censored on the intervals $(\kappa_{1,i}, \kappa_{2,i} + c)$.

For an estimate of S, we have to discard the information lying in the extended regions $\rho_i^* := \{(t_1,t_2) : \kappa_{1,i} < t_1 < \kappa_{2,i},\ t_1 < t_2 < \kappa_{2,i} + c\} \bigcup \{(t_1,t_2) : t_1 <$

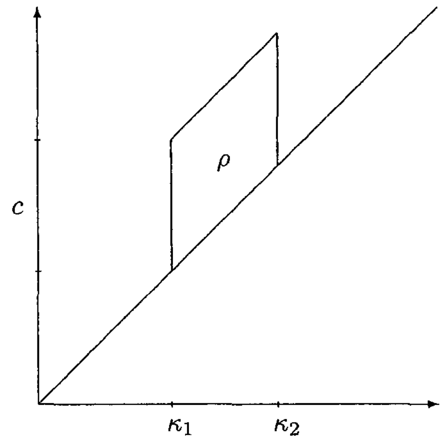

Figure 6.8: Scenario 3.

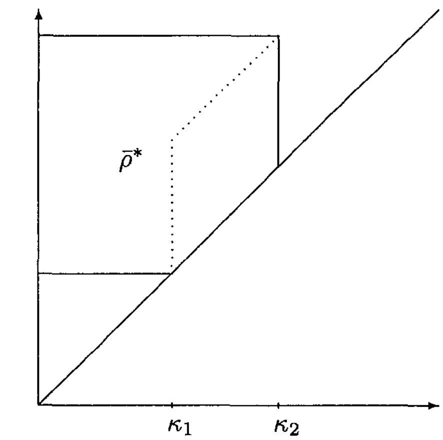

Figure 6.9: The extended region ρ^*.

$\kappa_{1,i},\ \kappa_{1,i} < t_2 < \kappa_{2,i} + c\}$, which we showed in Chapter 4 to be $*$-clouds with respect to $\mathcal{F}$. For an illustration, refer to Figure 6.9. Then the projections of the $*$-anti-clouds $(\rho_i^*)^c$ on the flow are defined by $\bar{\xi}_i^\gamma := \{u \in [0,1] : \gamma(u) \in (\rho_i^*)^c\}$ for $i = 1, 2, \ldots$.

For the estimate for $1 - F$, we are able to determine whether or not Y_i is in the past of any point t in the complement of the region $\rho_i^p := \{(t_1, t_2) : \kappa_{1,i} < t_1 < \kappa_{2,i},\ \kappa_{1,i} < t_2 < t_1 + c\} \cup \{(t_1, t_2) : \kappa_{2,i} \leq t_1,\ \kappa_{1,i} < t_2 < \kappa_{2,i} + c\}$ (see Figure 6.10), and so we make use of all the information outside ρ_i^p. It's easy to show that ρ_i^p are clouds with respect to $\mathcal{F}$, and therefore the projections of the censoring on the flow

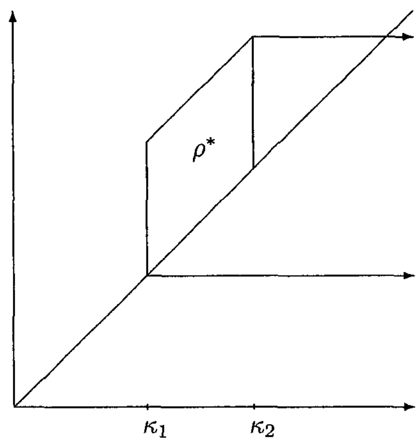

Figure 6.10: The region ρ_i^p.

are given by $\xi_i^\gamma := \{u \in [0,1] : \gamma(u) \in (\rho_i^p)^c\}$ for $i = 1, 2, \ldots$.

• Scenario 4: Consider two sequences of i.i.d. T-valued random variables (Y_i) and (C_i) independent of each other, and let $\xi_i := A_{C_i}$ be the uncensored region. Suppose we are allowed to observe $A_{Y_i} \cap \xi_i$, $i = 1, 2, \ldots$.

We first address the estimation of $1 - F$. Here, a point Y_i will be considered at risk at time t whenever both $t \in \xi_i$ and $Y_i \in A_t^c$, or, equivalently, if $t \in \xi_i \cap D_{Y_i}$; see Figure 6.11. If Y_i is no longer considered at risk at $\gamma(u) = t$, we would either have $t \in D_{Y_i}^c$ or $t \in \xi_i^c$. In case $t \in D_{Y_i}^c$ –which would mean Y_i is a failure time–, then $Y_i \in A_t \subseteq \xi_i$; this indicates that we may only count a jump in our estimate when Y_i is fully observed.

In the case of the estimate for the survival function S, Y_i will now be considered at risk at t only if $t \in \xi_i \cap A_{Y_i}$. If Y_i is no longer at risk at $\gamma(u) = t$, then we know that either $t \in A_{Y_i}^c$ or $t \in \xi_i^c$. If $t \in A_{Y_i}^c$, then $Y_i \in D_t$; however, this doesn't imply that $Y_i \in \xi_i$, since we only have that $A_t \subseteq \xi_i$. Hence, partial observations may actually be counted as jumps, which was a shortcoming of the Tsai-Crowley estimator. However, we can build a mechanism that allows us to differentiate between censored observations and actual failure times.

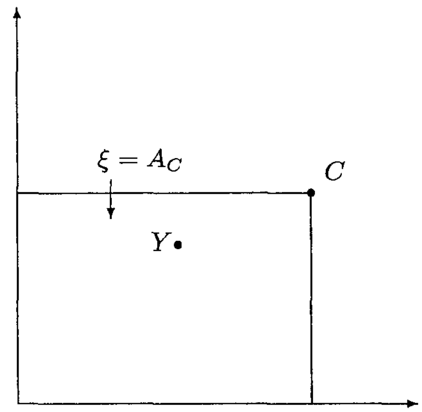

Figure 6.11: Only observations inside ξ will be counted as jump points.

Let $\mathcal{F}(t) = \sigma\{I_{\{s \le C_i\}}, I_{\{s \le Y_i\}}, t \le s\}$, and $\bar{\mathcal{F}}^{\gamma}(u) := \mathcal{F}(\gamma(u))$. Define $\bar{\tau}^{\gamma} := \inf\{u : \gamma(u) \in A_Y^c\}$ and $\eta^{\gamma} := \inf\{u : \gamma(u) \in \xi^c\}$; both are stopping times with respect to $\bar{\mathcal{F}}^{\gamma}$, since

$$\{\bar{\tau}^{\gamma} \le v\} = \{\gamma(v) \in A_Y^c\} = \{\gamma(v) \in A_Y\}^c \in \mathcal{F}(\gamma(v)) = \bar{\mathcal{F}}^{\gamma}(v)$$

and

$$\{\eta^{\gamma} \le v\} = \{\gamma(v) \in \xi^c\} = \{\gamma(v) \in \xi\}^c \in \mathcal{F}(\gamma(v)) = \bar{\mathcal{F}}^{\gamma}(v).$$

We are able to observe $\xi^{\gamma} := \bar{\tau}^{\gamma} \wedge \eta^{\gamma}$, which is also an $\bar{\mathcal{F}}^{\gamma}$-stopping time:

$$\begin{aligned}
\{\xi^{\gamma} \le v\} &= \{\xi^{\gamma} > v\}^c \\
&= \{\bar{\tau}^{\gamma} \cap \eta^{\gamma} > v\}^c \\
&= [\{\bar{\tau}^{\gamma} > v\} \cap \{\eta^{\gamma} > v\}]^c \\
&= \{\bar{\tau}^{\gamma} \le v\} \cup \{\eta^{\gamma} \le v\} \in \bar{\mathcal{F}}^{\gamma}(v).
\end{aligned}$$

We also know that $\{\xi^{\gamma} = \bar{\tau}^{\gamma}\} \in \bar{\mathcal{F}}_{\xi^{\gamma}}$, where $\bar{\mathcal{F}}_{\xi^{\gamma}} = \{A \in \bar{\mathcal{F}}^{\gamma} : A \cap \{\xi^{\gamma} \le v\} \in \bar{\mathcal{F}}^{\gamma}(v)$ $\forall\, v \in [0,1]\}$; see, for example [15]. We will then count a jump whenever $\{\xi^{\gamma} = \bar{\tau}^{\gamma}\}$, since we have a failure time only when the path exited A_Y. It is worth mentioning that the classical right censoring scheme (see, for example, [3] or [5]) is a special case of this Scenario; an illustration is provided in Figure 6.12.

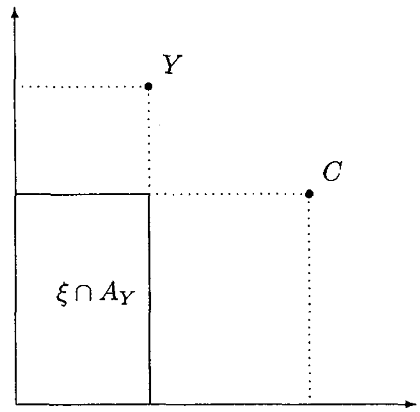

Figure 6.12: Partial observations such as Y may be counted as jump points.

6.2 Optimal Paths

At this point, we would like to address the way in which we choose a path. Ideally, as it was mentioned in [21], we want to pick a path that maximizes the number of uncensored projection points. The existence of an optimal path –that is, a path that generates more uncensored projection points than any other– is not guaranteed for the more general censoring models; however, there are some data structures that allow us to choose such a path. Before we go any further though, it would be prudent to define formally what we mean by an optimal path.

Definition 6.2.1 *A path γ joining the origin $\mathbf{0}$ to $t \in T$ is called optimal for ξ if and only if there exists $u \in [0,1]$ such that $A_{\gamma(u)} = A_t \cap \xi$. The path γ is jointly optimal for $\xi_1, \ldots, \xi_n$ if and only if it is optimal for each ξ_i, $i = 1, \ldots, n$.*

Note that, if γ is an optimal path, the one-dimensional points τ_i^γ will be observed uncensored whenever the actual jump points Y_i are; indeed, a point Y is observed before t only if $Y \in A_t \cap \xi$. But since γ is optimal, this means that there exists $u \in [0,1]$ such that $A_t \cap \xi = A_{\gamma(u)}$, so by the definition of τ^γ it follows that $\tau^\gamma \leq u$. On the other hand, since $A_t \cap \xi = A_{\gamma(u)}$, we have that $\gamma(u) \in \xi$, hence $u \in \xi^\gamma$ and so $\tau^\gamma \in \xi^\gamma$. Therefore, since a jointly optimal path will pick up every available jump point before t, it generates the maximum number of uncensored projection points possible.

The definition above points us towards the kind of data structure that will accommodate a jointly optimal path. The fact that the path is strictly increasing, plus the condition that it has to be optimal for each of the censoring sets, suggests that the anti-clouds have to be increasing in some sort of way. We will take a look at one such model.

In what follows, we will assume that $T = [0,1]^2$. Suppose we have a sequence of i.i.d. T-valued random variables (Y_i), as well as a sequence of anti-clouds (ξ_i) of the form $\xi_i = [0, C_i]^2$, where each C_i is a $[0,1]$-valued random variable with continuous distribution; as usual, Y_i is censored by ξ_i^c for each i. This setup amounts to bivariate data under univariate censoring, and if we let $\mathcal{S} = \{[0,u]^2 : u \in [0,1]\}$, it can be regarded as a particular case of the ordered censoring scheme of Section 5.2. In [21], Tsai and Crowley arrived at an estimate for the survival function of the (Y_i)'s at a point (t_1, t_2) by considering a path γ and then using their analogue of the projections $\bar{Y}_i^\gamma$ (see Definition 6.1.5). Moreover, they concluded that in this model, the path that creates the most uncensored projection points among all nondecreasing paths is the one that follows a straight line from $(0,0)$ to $(t_1 \wedge t_2, t_1 \wedge t_2)$ and a straight line from $(t_1 \wedge t_2, t_1 \wedge t_2)$ to (t_1, t_2). One problem with their estimate, as they themselves noted in [21], is that it is not a genuine survival function. Our objective is to avoid this undesirable property by estimating $1 - F(t)$ instead of the survival function. We intend to show that their path has the property of being jointly optimal for this model, and that our estimate of $1 - F(t)$, based on the projections Y_i^γ and given by (6.1.7), gives either negative mass or no mass at all to every point in T; in other words, our estimate of F is a measure. Furthermore, as we pointed out in Scenario 4, no mass is given to censoring times.

We want to look at the weights given by our estimate of $1-F(t)$ to each observation (Y_i), and the key to do that comes from a property of the usual one-dimensional

Kaplan-Meier estimate; therefore, we will step away from our analysis of the structure described above in order to have a quick look at the classic one-dimensional survival setup under right-censoring. Suppose we have sequences of i.i.d. $[0,1]$-valued random variables (X_i) and (T_i), where the (X_i)'s are failure times and the (T_i)'s act as censoring variables; we are able to observe $X_i \wedge T_i$ and $I_{\{X_i \le T_i\}}$ for $i = 1, 2, \ldots$. Let S^X, S^T denote the survival functions of (X_i) and (T_i) respectively, and let $\hat{S}_n^X$, $\hat{S}_n^T$ be their respective Kaplan-Meier estimates:

$$\hat{S}_n^X(u) = \prod_{i:X_i \wedge T_i \le u} \left(1 - \frac{I_{\{X_i \le T_i\}}}{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}}}\right),$$

$$\hat{S}_n^T(u) = \prod_{i:X_i \wedge T_i \le u} \left(1 - \frac{1 - I_{\{X_i \le T_i\}}}{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}}}\right).$$

Then we have that

$$\begin{aligned}
&\hat{S}_n^X(u)\hat{S}_n^T(u) \\
&= \prod_{i:X_i \wedge T_i \le u} \left(1 - \frac{I_{\{X_i \le T_i\}}}{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}}}\right) \prod_{i:X_i \wedge T_i \le u} \left(1 - \frac{1 - I_{\{X_i \le T_i\}}}{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}}}\right) \\
&= \prod_{i:X_i \wedge T_i \le u} \left(1 - \frac{1}{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}}}\right) \\
&= \prod_{i:X_i \wedge T_i \le u} \left(\frac{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}} - 1}{\sum_{j=1}^n I_{\{X_j \wedge T_j \ge X_i \wedge T_i\}}}\right) \\
&= n^{-1} \sum_{i=1}^n I_{\{X_i \wedge T_i > u\}};
\end{aligned}$$

in other words,

$$\hat{S}_n^X(u) = \frac{\sum_{i=1}^n I_{\{X_i \wedge T_i > u\}}}{n\hat{S}_n^T(u)}. \tag{6.2.1}$$

Equation (6.2.1) shows clearly that the weight given by the Kaplan-Meier estimate $\hat{S}_n^X$ to any failure time X is going to be $-1/n\hat{S}_n^T(X)$. Note that $\hat{S}_n^T(X)$ depends on the number of X_i's and T_i's found before T_X, where $T_X := \max\{T_i : T_i < X\}$; the more X_i's and T_i's inside $[0, T_X]$, the smaller the value of $\hat{S}_n^T(X)$. On the other hand, it does not depend on the number of observations X_i lying inside $[T_X, X]$, since the

Kaplan-Meier estimate only changes once we record a "success": for $\hat{S}_n^T$, that means observing a T_j.

Back to our two-dimensional model under univariate censoring, consider a map $\alpha : [0,1] \to \mathcal{S}$, and for any $t \in T$, let $\alpha^t : [0,1] \to \mathcal{S} \cap A_t$ be its truncated version: $\alpha^t(p) = \alpha(p) \cap A_t$. It's easy to see that γ, Tsai and Crowley's recommended path, is jointly optimal for our sequence of anti-clouds, by noting that $\alpha^t(p) = A_{\gamma(u)}$ for some $u \in [0,1]$. It remains to show that, under this kind of path, the estimator (6.1.7) gives no positive mass at any point in T.

Suppressing dependence on γ for a moment, let $\hat{S}_n^{Y,\gamma}(u) = \hat{S}_n^Y(u)$ denote the estimator (6.1.7), and let $\eta_i^\gamma = \sup\{u : u \in \xi_i^\gamma\}$. Note that we can write

$$\begin{aligned}
\hat{S}_n^Y(u) &= \prod_{i=1}^n \left(1 - \frac{I_{\{\tau_i^\gamma \le u,\, \tau_i^\gamma \in \xi_i^\gamma\}}}{Z_n^\gamma(\tau_i^\gamma)}\right) \\
&= \prod_{i=1}^n \left(1 - \frac{I_{\{\tau_i^\gamma \in \xi_i^\gamma\}} I_{\{\tau_i^\gamma \wedge \eta_i^\gamma \le u\}}}{\sum_{j=1}^n I_{\{\tau_j^\gamma \wedge \eta_j^\gamma \ge \tau_i^\gamma \wedge \eta_i^\gamma\}}}\right) \\
&= \prod_{i:\tau_i^\gamma \wedge \eta_i^\gamma \le u} \left(1 - \frac{I_{\{\tau_i^\gamma \in \xi_i^\gamma\}}}{\sum_{j=1}^n I_{\{\tau_j^\gamma \wedge \eta_j^\gamma \ge \tau_i^\gamma \wedge \eta_i^\gamma\}}}\right). \qquad (6.2.2)
\end{aligned}$$

If we reverse the roles of the observations Y_i and the censoring in the same way as we did in Chapter 5, we can come up with $\hat{S}_n^{\xi,\gamma}(u) = \hat{S}_n^\xi(u)$, the Kaplan-Meier estimate for the one dimensional survival function of the censoring variables η_i^γ:

$$\begin{aligned}
\hat{S}_n^\xi(u) &= \prod_{i=1}^n \left(1 - \frac{I_{\{\eta_i^\gamma \le u\}} I_{\{\eta_i^\gamma \le \tau_i^\gamma\}}}{\sum_{j=1}^n I_{\{\eta_j^\gamma \ge \eta_i^\gamma,\, \eta_i^\gamma \le \tau_j^\gamma\}}}\right) \\
&= \prod_{i=1}^n \left(1 - \frac{I_{\{\eta_i^\gamma \le \tau_i^\gamma\}} I_{\{\tau_i^\gamma \wedge \eta_i^\gamma \le u\}}}{\sum_{j=1}^n I_{\{\tau_j^\gamma \wedge \eta_j^\gamma \ge \tau_i^\gamma \wedge \eta_i^\gamma\}}}\right) \\
&= \prod_{i:\tau_i^\gamma \wedge \eta_i^\gamma \le u}^{n} \left(1 - \frac{1 - I_{\{\tau_i^\gamma \in \xi_i^\gamma\}}}{\sum_{j=1}^n I_{\{\tau_j^\gamma \wedge \eta_j^\gamma \ge \tau_i^\gamma \wedge \eta_i^\gamma\}}}\right). \qquad (6.2.3)
\end{aligned}$$

Then it is easy to see that

$$(6.1.7) = \hat{S}_n^Y(u) = \frac{\sum_{i=1}^n I_{\{\tau_i^\gamma \wedge \eta_i^\gamma > u\}}}{n\hat{S}_n^\xi(u)}, \qquad (6.2.4)$$

a result that was expected, given (6.2.1) and the fact that we reduced our two-dimensional data to a set of one-dimensional observations.

It is important to recall that in one-dimension, the weight at any single point u given by the usual Kaplan-Meier estimate is calculated by substracting the value of the estimate 'just before' that point from the actual value at u. Since (6.1.7) provides an estimate of the bivariate distribution $F(t)$ –or rather, of $1-F(t)$–, the weight that it assigns to a point $t=(t_1,t_2)\in T$ has to be computed by the usual two-dimensional formula $1-F(t)-(1-F(t_1^-,t_2))-(1-F(t_1,t_2^-))+(1-F(t_1^-,t_2^-))$, where t_j^- denotes a number arbitrarily close to, but strictly smaller than t_j, $j=1,2$; in other words, we will assume that $t^-=(t_1^-,t_2^-)$ is chosen so that there are no Y_i's or C_i's in the set $(t^-,t]\setminus\{t\}$.

Consider the mass given by $\hat{S}_n^\xi$ to any 'observation' τ^γ. This mass will depend on the number of η_i^γ's and τ_i^γ's that came before $\eta_{\tau^\gamma}^\gamma$, where $\eta_{\tau^\gamma}^\gamma:=\max\{\eta_i^\gamma:\eta_i^\gamma<\tau^\gamma\}$; just as important, it does not depend on the number of τ_i^γ's between $\eta_{\tau^\gamma}^\gamma$ and τ^γ.

Now we calculate the weight given to τ^γ by the estimate (6.1.7). Let $Y=(Y_1,Y_2)$ be the actual observation that generated τ^γ. We need to compute the values of the estimate at Y, $Y^-=(Y_1^-,Y_2^-)$, $Y^l=(Y_1^-,Y_2)$ and $Y^d=(Y_1,Y_2^-)$. Let γ, γ' be the paths that pass through Y and $Y^-=(Y_1^-,Y_2^-)$, respectively. Because of the last observation in the previous paragraph, the denominator in Equation (6.2.4) is $n\hat{S}_n^\xi(\tau^\gamma)$ for each of the four points being considered. However, the numerator is equal to the number of 'observations' still at risk; in the case of Y, we have removed τ^γ from that set, while we still consider τ^γ to be at risk for the rest of the points. Therefore, the weight given to τ^γ will be $-1/n\hat{S}_n^\xi(\tau^\gamma)$. This implies that for any jump point Y and $t>Y$ such that $\sum_{i=1}^n I_{\{Y_i\in[0,t]\setminus[0,Y]\}}=0$, the family of pathwise estimators give weight $-1/n\hat{S}_n^\xi(\tau^\gamma)$ to the point Y, and mass 0 to t.

On the other hand, there could be some points where there are no observations and yet the estimate assigns some mass; namely, on points of the form $Y\vee Y_k$, where Y_k is any observation such that $Y_k\notin E_Y\cup A_Y$, but $Y_k^\gamma<Y^\gamma$. The reason is that, as

we previously remarked, the weight generated by Y on $\hat{S}_n^\xi$ depends on the number of τ_i^γ's inside $[0, \eta_{\tau^\gamma}^\gamma]$. If both Y and Y_k are inside the same censoring set, there won't be any change in $\hat{S}_n^\xi$, since that number hasn't changed; however, if there is a censoring set ξ such that $Y_k \in \xi$ and $Y \notin \xi$ –which in turn would mean $\gamma^{-1}(Y_k^\gamma) = \tau_k^\gamma < \eta^\gamma < \tau^\gamma$ because of the optimality of the path $\gamma-$, the 'observation' τ_k^γ would now be in $[0, \eta_{\tau^\gamma}^\gamma]$, thereby affecting the weight given by $\hat{S}_n^\xi$. Now, the final question is raised: is this 'artificial' weight negative, as it should be?

To answer that question, we take a look at one such scenario. Let Y, Y_k and ξ be as described in the paragraph above. Consider the points $W = (W_1, W_2) := Y \vee Y_k$, $W^{(-)} = (W_1^-, W_2^-)$, $W^{(1-)} = (W_1^-, W_2)$ and $W^{(2-)} = (W_1, W_2^-)$, and let $\gamma_\vee$, $\gamma_{\vee-}$ be the paths that pass through W and $W^{(-)}$, respectively. In order to further complicate the notation, suppose that Y lies above the diagonal $t_1 = t_2$, and that the number of observations still at risk just after W is $n - j$. Then M, the mass given by the estimate (6.1.7) to the point $W = Y \vee Y_k$, is

$$
\begin{aligned}
M &= \hat{S}_n^Y(\gamma_\vee^{-1}(W^{\gamma_\vee})) + \hat{S}_n^Y(\gamma_{\vee-}^{-1}(W^{(-)\gamma_{\vee-}})) \\
&\quad - \hat{S}_n^Y(\gamma_{\vee-}^{-1}(W^{(1-)\gamma_{\vee-}})) - \hat{S}_n^Y(\gamma_\vee^{-1}(W^{(2-)\gamma_\vee})) \\
&= \frac{n-j}{n \cdot \hat{S}_n^\xi(\gamma_\vee^{-1}(W^{\gamma_\vee}))} + \frac{n-j+2}{n \cdot \hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(-)\gamma_{\vee-}}))} \\
&\quad - \frac{n-j+1}{n \cdot \hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(1-)\gamma_{\vee-}}))} - \frac{n-j+1}{n \cdot \hat{S}_n^\xi(\gamma_\vee^{-1}(W^{(2-)\gamma_\vee}))} \\
&= \frac{n-j}{n \cdot \hat{S}_n^\xi(\gamma_\vee^{-1}(W^{\gamma_\vee}))} + \frac{n-j+2}{n \cdot \hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(-)\gamma_{\vee-}}))} \\
&\quad - \frac{n-j+1}{n \cdot \hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(-)\gamma_{\vee-}}))} - \frac{n-j+1}{n \cdot \hat{S}_n^\xi(\gamma_\vee^{-1}(W^{\gamma_\vee}))} \\
&= -\frac{1}{n \cdot \hat{S}_n^\xi(\gamma_\vee^{-1}(W^{\gamma_\vee}))} + \frac{1}{n \cdot \hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(-)\gamma_{\vee-}}))} \\
&= -\frac{1}{n}\left(\frac{1}{\hat{S}_n^\xi(\tau_W^{\gamma_\vee})} - \frac{1}{\hat{S}_n^\xi(\tau_{W^{(-)}}^{\gamma_{\vee-}})}\right);
\end{aligned}
$$

we get the second equality by applying Equation (6.2.4), the third equality comes from the fact that $\hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(1-)\gamma_{\vee-}})) = \hat{S}_n^\xi(\gamma_{\vee-}^{-1}(W^{(-)\gamma_{\vee-}}))$ and $\hat{S}_n^\xi(\gamma_\vee^{-1}(W^{(2-)\gamma_\vee})) =$

$\hat{S}_n^\xi(\gamma_\vee^{-1}(W^{\gamma_\vee}))$ –because in both cases no new observations have occurred inside the regions that could change the value of $\hat{S}_n^\xi-$, and the last one is obtained simply by defining $\tau_W^{\gamma_\vee} := \gamma_\vee^{-1}(W^{\gamma_\vee})$ and $\tau_{W(-)}^{\gamma_\vee-} := \gamma_{\vee-}^{-1}(W^{(-)\gamma_\vee-})$. Now it is easy to see that the weight M is negative; indeed, since $\hat{S}_n^\xi$ is a decreasing function and $\tau_{W(-)}^{\gamma_\vee-} < \tau_W^{\gamma_\vee}$, we have that $\hat{S}_n^\xi(\tau_W^{\gamma_\vee}) < \hat{S}_n^\xi(\tau_{W(-)}^{\gamma_\vee-})$, hence $\frac{1}{\hat{S}_n^\xi(\tau_W^{\gamma_\vee})} > \frac{1}{\hat{S}_n^\xi(\tau_{W(-)}^{\gamma_\vee-})}$. All other scenarios can be handled similarly, so we have proven that our estimate gives no positive mass at any point in T.

Functional Central Limit Theorem

The next topic in our discussion of this particular model is the existence of a functional central limit theorem for the estimator (6.1.7). It is worth mentioning that a CLT exists for this estimator at a fixed point u from the usual one-dimensional theory, but we are interested in the more general functional result. We are going to assume throughout this section that both the distribution of the failure times and that of the censoring variables are continuous with bounded derivatives; this will ensure that the limiting Gaussian processes are continuous. There are two obstacles that we need to overcome in order to prove such a result: first, we have to be able to express (6.1.7) in terms of the original observations Y_i and the anti-clouds $\xi_i = [0, C_i]^2$ at any two-dimensional point (s,t); the second is that after doing this, we will run into trouble while trying to write the Nelson-Aalen estimator corresponding to the Kaplan-Meier estimator of the Y_i's as an integral, where the integrand is the inverse of the at-risk process and the integrator is an empirical process of a subdistribution. This particular detail is crucial to us, because for us to use the functional Delta-Method we need a Hadamard-differentiable mapping from those two processes to the Kaplan-Meier estimator.

Our solution to these inconveniences consists of once again exploiting the relation (6.2.4). Note that the path γ passes through –and terminates at– $t \in [0,1]^2$. Therefore, the Kaplan-Meier estimate of the survival function of C in the denominator of

(6.2.4) only uses the information inside A_t, and so it actually depends on t even though the notation does not reflect this fact. We intend to show that the Kaplan-Meier estimates for the survival function of the censoring variables C_i follow a functional central limit theorem on $D[0,1]^2$. This, coupled with the fact that the numerator in (6.2.4) is simply a sum of i.i.d. random variables, implies the same result for our estimator (6.1.7): simply note that the map $(A, B) \to (A/B)$ is Hadamard-differentiable on a domain of the type $\{(A, B) : A \leq M, B \geq \epsilon\}$ for $\epsilon > 0$. If $P(u \in \xi^\gamma) > \epsilon$ for every u, the pair $(\frac{1}{n}\sum_{i=1}^n I_{\{\tau_i^\gamma \wedge \eta_i^\gamma > u\}}, \hat{S}_n^\xi(u))$ is contained in this domain with probability tending to 1 for $M \geq 1$ and sufficiently small ϵ.

We now consider the Kaplan-Meier estimator of the survival function of C, as evaluated at a point (s,t). Because of symmetry, we only consider points (s,t) above the diagonal, so the optimal path is a diagonal from $(0,0)$ to (s,s), and then a vertical line from (s,s) to $(s,1)$. Let $Z_i = C_i \cdot I_{\{Y_i \notin \xi_i\}}$ and $W_i = Y_i \cdot I_{\{Y_i \in \xi_i\}}$. Fix s and treat the variables C_i as if they were the observations. Then we would know that C_i is not at risk at (s,p) if $0 < Z_i < p$ or $(0,0) < W_i < (s \wedge p, p)$; therefore, the at-risk process for this kind of path can be written as

$$R_{n,1}(s,p) = \frac{1}{n}\sum_{i=1}^n \left(1 - I_{\{0<Z_i<p\}}\right)\left(1 - I_{\{(0,0)<W_i<(s\wedge p,p)\}}\right).$$

This process is centered by its expected value

$$\begin{aligned} r_1(s,p) &= E[R_{n,1}(s,p)] \\ &= 1 - P[0 < Z < p] - P[(0,0) < W < (s \wedge p, p)], \end{aligned}$$

where the last equality comes from the fact that we always have either $Z = 0$ or $W = 0$. Likewise, for points below the diagonal we would fix t and obtain the analogous process

$$R_{n,2}(p,t) = \frac{1}{n}\sum_{i=1}^n \left(1 - I_{\{0<Z_i<p\}}\right)\left(1 - I_{\{(0,0)<W_i<(p,p\wedge t)\}}\right),$$

as well as

$$\begin{aligned} r_2(p,t) &= E[R_{n,2}(p,t)] \\ &= 1 - P[0 < Z < p] - P[(0,0) < W < (p, p \wedge t)]. \end{aligned}$$

Let

$$G_n^C(p) = \frac{1}{n}\sum_{i=1}^n I_{\{0<Z_i\le p\}} = \frac{1}{n}\sum_{i=1}^n I_{\{0<C_i\le p\}} I_{\{Y_i\notin\xi_i\}},$$

the empirical distribution of the "uncensored" C_i's. Then $E[G_n^C(p)] := G^C(p) = P[0 < Z \le p]$. It is easy to see that the classes of functions $\mathcal{G} := \{g_{s,t} : s,t \in [0,1]\}$, $\mathcal{H}_1 := \{h^1_{s,t} : s,t \in [0,1]\}$ and $\mathcal{H}_2 := \{h^2_{s,t} : s,t \in [0,1]\}$, where $g_{s,t}(x,y) = 1 - I_{\{0<x<t\}}$, $h^1_{s,t}(x,y) = 1 - I_{\{(0,0)<y<(s\wedge t,t)\}}$ and $h^2_{s,t}(x,y) = 1 - I_{\{(0,0)<y<(s,s\wedge t)\}}$ for $x \in \mathbf{R}_+$, $y \in \mathbf{R}^2_+$, and their products are actually VC-classes (see [22], Section 2.6), since each of those functions consists of either an indicator of a 1-dimensional cell or a 2-dimensional rectangle; hence they are also Donsker (again, refer to Section 2.6.2 in [22]). Therefore, if we restrict our attention to a set $[0,\tau]^2$ such that $r_i(s,t) > 0$ for every $(s,t) \in [0,\tau]^2$, $i = 1,2$, we have that

$$\begin{aligned} &\sqrt{n}(R_{n,1}(s,p) - r_1(s,p), R_{n,2}(p,t) - r_2(p,t), G_n^C(p) - G^C(p)) \\ &\quad \Rightarrow \quad (Q^{r_1}(s,p), Q^{r_2}(p,t), Q^C(p)) \end{aligned}$$

in $D[0,\tau]^2 \times D[0,\tau]^2 \times D[0,\tau]$, where Q^{r_1}, Q^{r_2} and Q^C are all mean-zero continuous Gaussian processes. That restriction is required in order for the composition

$$\begin{aligned} &(R_{n,1}(\cdot,\cdot), R_{n,2}(\cdot,\cdot), G_n^C(\cdot)) \\ \longrightarrow\ & \left(\frac{1}{R_{n,1}(\cdot,\cdot)}, \frac{1}{R_{n,2}(\cdot,\cdot)}, G_n^C(\cdot)\right) \\ \longrightarrow\ & \left(\int_0^{\cdot} \frac{1}{R_{n,1}(\cdot,u)} dG_n^C(u), \int_0^{\cdot} \frac{1}{R_{n,1}(u,\cdot)} dG_n^C(u)\right) \\ :=\ & (H^C_{n,1}(\cdot,\cdot), H^C_{n,2}(\cdot,\cdot)) \\ \longrightarrow\ & \left(\tilde{\prod}[1 - dH^C_{n,1}(s,\cdot)], \tilde{\prod}[1 - dH^C_{n,2}(\cdot,t)]\right) \\ :=\ & (S^C_{n,1}(\cdot,\cdot), S^C_{n,2}(\cdot,\cdot)) \\ \longrightarrow\ & S^C_{n,1}(s,t) I_{\{s\le t\}} + S^C_{n,2}(s,t) I_{\{s>t\}} = S^C_n(s,t), \end{aligned}$$

where $\tilde{\prod}$ stands for the usual product integral, to be Hadamard-differentiable tangentially to $C[0,\tau]^2 \times C[0,\tau]^2 \times D[0,\tau]$ on a certain domain; then we will have the functional CLT. This, however, is not as straightforward as in previous cases: we now have processes depending on two parameters, so we have to tread carefully. Note that we have switched from $\hat{S}_n^{\xi,\gamma}$ to the notation $S_{n,i}^C$ now that the paths have been specified.

We first consider the integration part of the composition. Let $A \in D[0,1]^2$, $B \in BV_M[0,1]^2$, as well as $\alpha \in C[0,1]^2$ and $\beta \in D[0,1]^2$. We will keep s fixed, and define $\psi(A,B)(s,p) = \int_0^p A(s,u)B(s,du)$. We now show that the derivative of the map ψ is given by

$$\psi'_{A,B}(\alpha,\beta)(s,p) = \int_0^p A(s,u)\beta(s,du) + \int_0^p \alpha(s,u)B(s,du).$$

Let $\alpha_t, \beta_t \in D[0,1]^2$ such that both

$$\sup_{s,u} \|\alpha_t(s,u) - \alpha(s,u)\| \to 0, \qquad \sup_{s,u} \|\beta_t(s,u) - \beta(s,u)\| \to 0 \tag{6.2.5}$$

as $t \to 0$. Write $A_t = A + t\alpha_t$ and $B_t = B + t\beta_t$. We also need that $\int |A(s,dp)| < \infty$ for every s. We use the same approach found in the proof of Lemma 3.9.17 in [22]. We have that

$$\begin{aligned}
&\frac{1}{t}\left[\int_0^p A_t(s,u)B_t(s,du) - \int_0^p A(s,u)B(s,du)\right] - \psi'_{A,B}(\alpha_t,\beta_t)(s,p) \\
&= t\int_0^p \alpha_t(s,u)\beta_t(s,du) \\
&= \int_0^p \alpha_t(s,u)(B_t - B)(s,du) \\
&= \int_0^p \alpha(s,u)(B_t - B)(s,du) + \int_0^p (\alpha_t - \alpha)(s,u)(B_t - B)(s,du).
\end{aligned} \tag{6.2.6}$$

Since $A(s,\cdot)$ is of bounded variation for every s, the derivative map is continuous with respect to the uniform norm, and so $\psi'_{A,B}(\alpha_t,\beta_t) \to \psi'_{A,B}(\alpha,\beta)$ as $t \to 0$. Hence, it suffices to show that (6.2.6) converges to 0. Together, the condition in (6.2.5) and the fact that $\beta_t(s,\cdot) \in BV_M[0,1]^2$ for all $s \in [0,1]$ imply that the second term on the right converges to zero as $t \to 0$. On the other hand, note that for any arbitrary ϵ we can find a grid $0 = p_0 < p_1 < \ldots < p_m = 1$ and $0 = s_0 < s_1 < \ldots < s_n = 1$ such that $\sup\{|\alpha(u) - \alpha(v)| : u, v \in (s_{i-1}, s_i] \times (p_{j-1}, p_j]\} < \epsilon$. But since we are holding s fixed, we are basically only looking at lines, just as in Lemma 3.9.17 in [22]; moreover, the same partition can be used for every s, so we are able to take supremums over s and p at the same time. The rest of the proof of Lemma 3.9.17 in [22] follows from here without problems.

Next, we consider the differentiability of the product integral in two parameters. Our proof follows closely that of Lemma 3.9.30 in [22]. Consider $A \in D[0,1]^2$ such that $A(s,\cdot)$ is of bounded variation for every s, and define $A_s(t) := A(s,t)$. Let

$$\tilde{\phi}(A)(s,t) = \tilde{\prod}_{0 \le u \le t}(1 + dA_s(u)) = \phi(A)_s(t).$$

Write $A_n = A + t_n\alpha_n$, where $\alpha_n \to \alpha \in D[0,1]^2$ and $t_n \to 0$. We would like to show that

$$\begin{aligned}
\frac{\left(\tilde{\phi}(A_n) - \tilde{\phi}(A)\right)(s,t)}{t_n} &\longrightarrow \tilde{\phi}'_A(\alpha)(s,t) = \tilde{\phi}'_{A_s}(\alpha)(t) \\
&= \int_{(0,t]} \tilde{\phi}(A)(s,(0,u))\tilde{\phi}(A)(s,(u,t])\alpha(s,du) \\
&= \int_{(0,t]} \phi(A)_s(0,u)\phi(A)_s(u,t]\alpha_s(du);
\end{aligned}$$

just as in [22], $\phi(A)_s(u,t] := \phi(A)_s(t)/\phi(A)_s(u)$. Fix s. By the Duhamel equation (See Appendix),

$$\begin{aligned}
\frac{\left(\tilde{\phi}(A_n) - \tilde{\phi}(A)\right)(s,t)}{t_n} &= \frac{\left(\phi(A_n)_s - \phi(A)_s\right)(t)}{t_n} \\
&= \int_{(0,t]} \phi(A)_s(0,u)\phi(A_n)_s(u,t](\alpha_n)_s(du);
\end{aligned}$$

hence it is enough to show that

$$\int_{(0,t]} \phi(A)_s(0,u)\phi(A_n)_s(u,t](\alpha_n)_s(du) \longrightarrow \int_{(0,t]} \phi(A)_s(0,u)\phi(A)_s(u,t]\alpha_s(du)$$

uniformly in both s and t. For s fixed, we know by Lemma 3.9.30 in [22] that the convergence in t is uniform. Additionally, if $A(s,\cdot)$ is decreasing and $A(\cdot,0)$ is bounded, then we have that $\phi(A)_s(t) \le \sup_s \exp\{A_s(0)\}$, so $\tilde{\phi}$ is uniformly bounded in both s and t; by the Duhamel equation, $\tilde{\phi}$ is also uniformly continuous in both components, and then the product integrals $\tilde{\phi}(A_n)$ converge uniformly to their limit $\tilde{\phi}(A)$ by another application of the Duhamel equation. From here, the proof of Lemma 3.9.30 found in [22] follows. Note that the requirements that $A(s,\cdot)$ be decreasing and $A(\cdot,0)$ bounded are met by $A_n = -H^C_{n,i}$ and $A = -H^C_i$, $i = 1,2$, where H^C_1 and H^C_2 are the actual cumulative hazard functions for the censoring variable C for points above and below the diagonal, respectively.

Finally, we can conclude that the composition map $\left(R_{n,1}, R_{n,2}, G^C_n\right) \to S^C_n$ described before is Hadamard-differentiable tangentially to $C[0,\tau]^2 \times C[0,\tau]^2 \times D[0,\tau]$ on a domain of the type $\left\{(A,B,C) : A, B \ge \epsilon, \int |dC| \le M\right\}$ for given M and $\epsilon > 0$ at every (A,B,C) such that $1/A$ and $1/B$ are of bounded variation. If we restrict our attention to points (s,t) in a set $[0,\tau]^2$ such that $r_i(s,t) > 0$ for every $(s,t) \in [0,\tau]^2$, $i = 1,2$, then $\left(R_{n,1}, R_{n,2}, G^C_n\right)$ is contained in this domain with probability tending to 1 for $M \ge 1$ and sufficiently small ϵ. The central limit theorem follows from the Delta-Method.

This discussion leads us to the Functional Central Limit Theorem for the estimator (6.1.7):

Theorem 6.2.2 *Let γ_t denote the optimal path from 0 to t, and $u_t = \gamma_t^{-1}(t)$. Then let $\hat{F}_n(t) = 1 - \hat{S}^{Y,\gamma_t}_n(u_t)$, and define $\varpi^{(n)}(t) := \sqrt{n}(\hat{F}_n(t) - F(t))$. If $P(t \in \xi) > 0$ for every t, then $\varpi^{(n)} \Rightarrow \Theta$ in $D[0,\tau]^2$, where Θ is a continuous mean-zero Gaussian process and τ is such that $r_i(s,t) > 0$ for every $(s,t) \in [0,\tau]^2$, $i = 1,2$. The variance of $\Theta(t)$ is given by*

$$(S^{\gamma_t}(u_t))^2 \int_0^{u_t} \frac{1}{P(\tau^\gamma \ge u, u \in \xi^{\gamma_t})} dH^{\gamma_t}(u)$$

Proof: We have shown in the preceeding paragraphs that if Φ denotes the map $(R_{n,1}, R_{n,2}, G^C_n) \to S^C_n$, then $\sqrt{n}(S^C_n - \Phi(r_1, r_2, G^C))$ converges in distribution to a mean-zero Gaussian process in $D[0,\tau]^2$. We also made a comment at the beginning of the Section about the need for $P(u_t \in \xi^{\gamma_t}) = P(t \in \xi)$ to be bounded away from zero in regards to the Hadamard-differentiability of the map $\zeta(A, B) = A/B$. Next, (6.2.4) shows that $\hat{S}_n^{Y,\gamma_t}(u_t)$ depends on the pair $(\frac{1}{n}\sum_{i=1}^n I_{\{\tau_i^\gamma \wedge \eta_i^\gamma > u_t\}}, \hat{S}_n^{C,\gamma_t}(u_t))$ through the map ζ; apply the Delta-Method to conclude that $\varpi^{(n)} \Rightarrow \Theta$. The fact that the distributions of both the failure times and the censoring variables are continuous with bounded derivatives ensures that the limiting process is continuous by an application of Condition (1.15) of Theorem 1.4 in [2] (As we mentioned before, refer to the discussion in p. 20 of [2] for an extension of that Theorem to multidimensional Euclidean spaces). The variance of Θ comes from the usual one-dimensional Kaplan-Meier estimator (See the Appendix). $\square$

Remark 6.2.3 We now have two approaches to estimating $1 - F$ in the case of univariate censoring. The reverse probability estimator of Chapter 5 puts mass only on observed points Y_i; the estimator proposed in this Section has the advantage of being dynamic and will spread mass over a larger number of points. We present a simple example in Figure 6.13 to highlight the different way these estimators assign mass. In the case of the reverse probability estimator (5.2.2), we take $\alpha(u) = [0, u]^2$ for every $u \in [0, 1]$; for the path-dependent estimator (6.1.7), we let $\gamma(u) = (u, u)$, so that for each $t \in [0, 1]^2$, $\gamma^t(u) = (t_1 \wedge u,\ t_2 \wedge u)$ describes the optimal path suggested by Tsai and Crowley. Figure 6.14 illstrates the estimated value of the distribution function using $\hat{\mu}_\alpha^{(n)}$, while Figure 6.15 does the same, only this time using $1 - \hat{S}_n^{Y,\gamma}$. Note that under the reverse probability estimator, Y_1 and Y_2 are assigned masses of $\frac{2}{3}$ and $\frac{1}{3}$, respectively, whereas $1 - \hat{S}_n^{Y,\gamma}$ puts mass on three points: Y_1 has mass $\frac{1}{2}$, Y_2 has mass $\frac{1}{3}$, and $Y_1 \vee Y_2$ has mass $\frac{1}{6}$.

Remark 6.2.4 The bootstrapping principles employed in the previous chapters to obtain confidence bands for different statistics remain valid. We have shown that our estimator $\hat{S}_n^Y(u)$ depends on $(\frac{1}{n}\sum_{i=1}^n I_{\{\tau_i^\gamma \wedge \eta_i^\gamma > u\}}, \hat{S}_n^C(u))$ through a Hadamard-differentiable map on a certain domain; just as in previous cases, this pair in turn

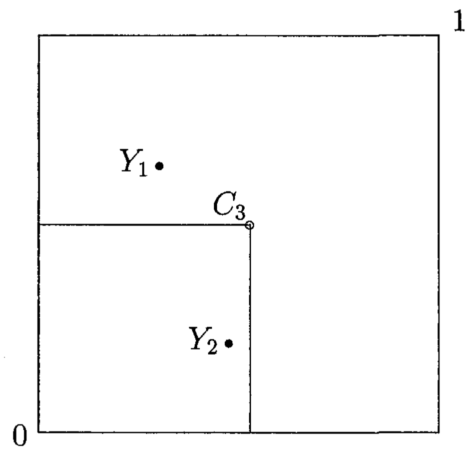

Figure 6.13: Univariate censoring: two observations, one censoring.

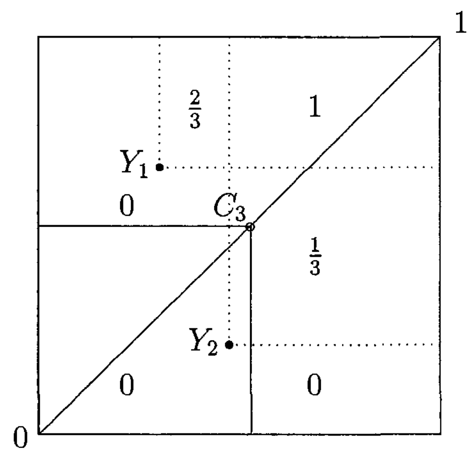

Figure 6.14: Values of the reverse probability estimator $\hat{\mu}_\alpha^{(n)}$ in each of the enclosed regions.

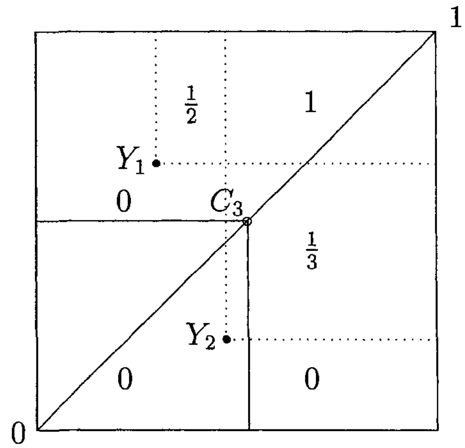

Figure 6.15: Values of the path-dependent estimator $1 - \hat{S}_n^{Y,\gamma}$ in each of the enclosed regions.

depends on indicator functions that form Donsker classes themselves. Letting $\gamma(u) = t = (s,p) \in [0,1]^2$, those functions are $I_{\{\eta_i^\gamma > u\}} = I_{\{t \in \xi_i\}}$, $I_{\{\tau_i^\gamma > u\}} = I_{\{Y_i \notin A_t^\circ\}}$, $I_{\{Y_i \in \xi_i\}}$, $I_{\{C_i < p\}}$ and $I_{\{Y_i \leq (s \wedge p, p)\}}$; hence we can bootstrap directly from the observed values of these indicators. Theorem 3.9.11 in [22] then guarantees the conditional convergence given (Y_i, ξ_i) of the bootstrapped estimator to the limit Θ found in Theorem 6.2.2.

Remark 6.2.5 It is possible to use the path-dependent estimator (6.1.7) for the applications explored in Chapter 5, such as the test of independence of the components of $Y = (Y_1, Y_2)$ or the copula estimation, with one caveat: just as in Section 5.2, (6.1.7) is restricted to a set $[0, \tau]^2$, so we would have to use the one-dimensional Kaplan-Meier estimates for both marginals. This presents no troubles since, as we mentioned in that very Section, the data structure being considered lends itself to such calculations. Remark 6.2.4 takes care of the bootstrap mechanism.

Remark 6.2.6 We conjecture that the techniques in this Section can be extended to more general ordered censoring models as defined in Section 5.2, using paths on increasing sets instead of paths on T.

Chapter 7

Conclusion

The modifications to Ivanoff and Merzbach's censoring model by clouds introduced in this thesis have been successful, in the sense that the observability problem found in the construction of the Nelson-Aalen estimator has been corrected. Moreover, the asymptotic results for this estimate, together with our proposed bootstrap techniques, allow for a number of interesting applications, some of which have been explored in this thesis.

As for the estimates for the distribution function of Chapters 5 and 6, they each have their advantages, as well as certain drawbacks. The reverse probability estimators of Chapter 5 feature a very simple construction, and they also possess the quality of actually being measures. These two characteristics have proven to be quite elusive for previous multidimensional estimators. However, these estimators do come with important restrictions on the censoring mechanism they allow, which limits their applicability. Another disadvantage is that neither of them is dynamic, as they require the entire data set to yield results.

The path-dependent estimator of Chapter 6, as we have pointed out, avoids these inconveniences. It can be used when dealing with a very general framework, both in terms of the censoring scheme and the space where the variables lie. In some cases, however, it may not give rise to measures; moreover, when the choice of path is not clear-cut, there could be uniqueness issues, as two different paths may not give the same estimate. Fortunately, when an optimal path exists these problems disappear.

The last point we want to highlight is that, with the possible exception of some path-dependent estimators, all the estimators presented in this thesis follow functional central limit theorems. This is a crucial feature that allows us to automatically establish asymptotic results for estimators of continuous functionals of the distribution function or of the cumulative hazard.

Chapter 8

Appendix

This Appendix includes some basic information about the construction, asymptotic behavior and relation between the one-dimensional versions of the Nelson-Aalen estimator for the cumulative hazard function and the Kaplan-Meier estimator for the survival function. We also elected to include a few lines about the product integral, since it actually links both estimators. There are many books covering these subjects extensively, but we liked the exposition given by Van der Vaart and Wellner in [22], from which the following developments are taken. As a last comment before introducing these concepts, the one-dimensional versions of the Nelson-Aalen estimator and the Kaplan-Meier estimator can be extended to more general censoring, but we will present them in the context of right censoring, as this is all we need from them for the purposes of this thesis.

8.1 The Product Integral

Let $A \in D(0,b]$, the space of cadlag functions in $(0,b]$. The product limit integral is defined as

$$\phi(A)(t) := \tilde{\prod_{0<s\leq t}} (1 + dA(s)) = \lim_{\max|t_i - t_{i-1}|} \prod_i [1 + A(t_i) - A(t_{i-1})],$$

where the limit is taken over a sequence of progressively finer partitions $0 = t_0 < t_1 < \ldots < t_n = t$. If we let $\Delta A(t) = A(t) - A(t^-)$ and $A^c(t) := A(t) - \sum_{0<s\leq t} \Delta A(s)$ be the jump part and the continuous parts of A, respectively, we can also write

$$\phi(A)(t) = \prod_{0<s\leq t} (1 + \Delta A(s)) \exp(A^c(t)).$$

The product integral map is Hadamard-differentiable as long as the function A is of bounded variation (See Lemma 3.9.30 in [22]). The Duhamel equation, which asserts that

$$(\phi(B) - \phi(A))(t) = \int_{(0,t]} \phi(A)(0,u)\phi(B)(u,t](B - A)(du)$$

for $A, B \in D(0, b]$, is essential to prove that result. Here, $\phi(B)(u, t]$ represents

$$\phi(B)(u,t] = \tilde{\prod_{u<s\leq t}} (1 + dB(s)) = \frac{\phi(B)(t)}{\phi(B)(u)}, \quad u < t.$$

8.2 The One-dimensional Nelson-Aalen Estimator

Let $X_1, \ldots, X_n$ and $T_1, \ldots, T_n$ be sequences of i.i.d. one-dimensional random variables with distribution function F and G, respectively, and independent of each other; moreover, let $S = 1 - F$ denote the survival function of the first sequence. The X_i's represent failure times and the T_i's act as censoring variables, and we are able to observe $Z_i = X_i \wedge T_i$ and $\Delta_i = I_{\{X_i \leq T_i\}}$ for $i = 1, 2, \ldots$. The cumulative hazard function of the failure times at time t, $H(t)$, can be written as

$$H(t) = \int_0^t \frac{1}{S}\, dF = \int_0^t \frac{1}{\bar{J}}\, dJ^u,$$

where $\bar{J}(t) = P(Z \geq t)$ and $J^u(t) = P(Z \leq t, \Delta = 1)$. Let $\bar{\mathcal{J}}_n(t) = \frac{1}{n}\sum_{i=1}^n I_{\{Z_i \geq t\}}$ and $\mathcal{J}_n^u(t) = \frac{1}{n}\sum_{i=1}^n \Delta_i \cdot I_{\{Z_i \leq t\}}$ be the empirical counterparts of $\bar{J}$ and J^u, respectively. The Nelson-Aalen estimator is given by

$$\hat{H}_n(t) = \int_0^t \frac{1}{\bar{\mathcal{J}}_n}\, d\mathcal{J}_n^u = \sum_{i:Z_i\leq t} \frac{\Delta_i}{\sum_{j=i}^n I_{\{Z_j \geq Z_i\}}}.$$

Since the pair $(\mathcal{J}_n^u, \bar{\mathcal{J}}_n)$ can be identified with the empirical distribution of the observations indexed by the functions $\delta I_{\{z\leq t\}}$ and $I_{\{z>t\}}$ when t ranges over $\mathbf{R}$, any of the main empirical central limit theorems yields

$$\sqrt{n}(\mathcal{J}_n^u - J^u, \bar{\mathcal{J}}_n - \bar{J}) \Rightarrow (W^u, \bar{W})$$

in $(D[0,\tau])^2$, where $(W^u, \bar{W})$ is a mean-zero Gaussian process, and τ is such that $\bar{J}(\tau) > 0$. Moreover, the Nelson-Aalen estimator depends on that pair $(\mathcal{J}_n^u, \bar{\mathcal{J}}_n)$ through the Hadamard-differentiable composition map

$$(A, B) \longrightarrow (A, \frac{1}{B}) \longrightarrow \int_0^t \frac{1}{B}\, dA,$$

so an application of the Delta-Method gives

$$\sqrt{n}(\hat{H}_n - H) \Rightarrow \int_0^t \frac{1}{\bar{J}}\, dW^u - \int_0^t \frac{\bar{W}}{\bar{J}^2}\, dJ^u =: U, \tag{8.2.1}$$

in $D[0,\tau]$ for every τ such that $\bar{J}(\tau) > 0$. The limit U is a mean-zero Gaussian process with its first term defined via integration by parts. Its covariance function is given by

$$\mathrm{Cov}(U(s), U(t)) = \int_{[0,s\wedge t]} \frac{1-\Delta H}{\bar{J}}\, dH.$$

8.3 The One-dimensional Kaplan-Meier Estimator

Consider the exact same set-up again. It is easy to derive the Kaplan-Meier estimator from the Nelson-Aalen estimator by noting that

$$S(t) = \tilde{\prod_{s\leq t}}(1 - dH(s)). \tag{8.3.1}$$

If we replace the cumulative hazard function H in (8.3.1) by the Nelson-Aalen estimator $\hat{H}_n$, we arrive at the Kaplan-Meier estimator $\hat{S}_n$ for the survival function:

$$\hat{S}_n(t) = \tilde{\prod_{s\leq t}}(1 - d\hat{H}_n(s)) = \prod_{i:Z_i\leq t}\left(1 - \frac{\Delta_i}{\sum_{j=i}^n I_{\{Z_j\geq Z_i\}}}\right).$$

Since the product integral is a Hadamard differentiable mapping, the asymptotic normality of the Nelson-Aalen estimator carries over to the Kaplan-Meier estimator; indeed, if the estimators $\hat{H}_n$ are of uniformly bounded total variation with probability tending to 1, then the Delta-Method gives

$$\sqrt{n}(\hat{S}_n - S) \Rightarrow \int_{(0,t]} S(s-)\frac{S(t)}{S(s)}\, dU(s) = S(t)\int_{(0,t]} \frac{1}{1-\Delta H(u)}\, dU(s)$$

in $D[0,\tau]$, where τ is such that $\bar{J}(\tau) > 0$ and U is the Gaussian process in (8.2.1). Its covariance function, for $s, t \leq \tau$, is given by

$$S(s)S(t)\int_{[0,s\wedge t]} \frac{1}{(1-\Delta H)\bar{J}}\, dH.$$

Bibliography

[1] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N., *Statistical Models Based on Counting Processes*, Springer Series in Statistics, Springer-Verlag, New York, 1993.

[2] Adler, Robert J., *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics Lecture Notes-Monograph Series, 1990.

[3] Burke, Murray D., Estimation of a Bivariate Distribution Function Under Random Censorship, *Biometrika* Vol. 75, No.2, Jun. 1988, 379-382.

[4] Burke, M. D., Csőrgö, S. and Horváth, L. Strong Approximations of some Biometric Estimates under Censorship, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* Vol. 56, 1981, 87-112.

[5] Campbell, G. and Földes, A., Large-Sample Properties of Nonparametric Bivariate Estimators with Censored Data, In *Nonparametric Statistical Inference, Colloquia Mathematica Societatis János Bolyai*, Ed. B. V. Gnedenko, M. L. Puri and I. Vincze, pp. 103-22, 1982. Amsterdam: North Holland.

[6] Efron, B., Censored data and the bootstrap, *J. Amer. Statist. Assoc.* Vol. 76, 1981, 312-319.

[7] Hildebrandt, T. H., *Introduction to the Theory of Integration*, Academic Press Inc., New York, 1963.

[8] Hougaard, Philip, *Analysis of Multivariate Survival Data*, Springer, New York, 2000.

[9] Ivanoff, B. G. and Merzbach, E., *Set-Indexed Martingales*, Chapman and Hall/CRC Press, Boca Raton, 2000.

[10] Ivanoff, B. G. and Merzbach, E., Set-Indexed Markov Processes, *Canadian Mathematical Society Conference Proceedings* 26, 2000, 217-232.

[11] Ivanoff, B. G. and Merzbach, E., Random censoring in set-indexed survival analysis, *The Annals of Applied Probability* 12, 2002, 944-971.

[12] Ivanoff, B. G. and Merzbach, E., Random clouds and an application to censoring in survival analysis, *Stochastic Processes and their Applications* 111, 2004, 259-279.

[13] Ivanoff, B. G, Merzbach, E. and Plante, M., A Compensator Characterization of Point Processes on Topological Lattices, *Electronic Journal of Probability* 12, 2007, 47-74.

[14] Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New Jersey, 2002.

[15] Kopp, P. E., *Martingales and stochastic integrals*, Cambridge University Press, London, 1984.

[16] Lin, D. Y. and Ying, Z., A Simple Nonparametric Estimator of the Bivariate Survival Function Under Univariate Censoring, *Biometrika* 80, 1993, 573-581.

[17] Lu, X. and Burke, M. D., Nonparametric Estimation of Linear Functionals of a Bivariate Distribution Under Univariate Censoring, *Journal of Statistical Planning and Inference*: in press, Jan. 2008.

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

[19] Pons, O. and Turckheim, E. de, Tests of independence for bivariate censored data based on the empirical joint hazard function, *Scand. J. Statist.* 18, 21-37 (1989).

[20] Shorack, G. R. and Wellner, J. A., *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.

[21] Tsai, W-Y and Crowley, J., A Note on Nonparametric Estimators of the Bivariate Survival Function Under Univariate Censoring, *Biometrika* Vol. 85, No.3, Sep. 1998, 573-580.

[22] van der Vaart, Aad W. and Wellner, Jon A., *Weak Convergence and Empirical Processes*, Springer series in statistics, 1996.