Estimation non-paramétrique de données censurées dans un cadre multi-états

Ségolen Geffray

▶ **To cite this version:**

Ségolen Geffray. Estimation non-paramétrique de données censurées dans un cadre multi-états. Mathematics [math]. Université Pierre et Marie Curie - Paris VI, 2006. English. NNT : . tel-00138280

HAL Id: tel-00138280
https://theses.hal.science/tel-00138280

Submitted on 25 Mar 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE L'UNIVERSITE PARIS 6

Spécialité
Statistique

Présentée par
Ségolen Geffray

Pour obtenir le grade de
DOCTEUR de l'UNIVERSITE PARIS 6

Sujet de la thèse :
Estimation non-paramétrique de données censurées dans un cadre multi-états

soutenue le 3 novembre 2006

devant le jury composé de :

Paul DEHEUVELS	Directeur de thèse
Jean-Yves DAUXOIS	Rapporteur
Ingrid VAN KEILEGOM	Rapporteur
Gérard DERZKO	Examinateur
Agathe GUILLOUX	Examinatrice
Catherine HUBER	Examinatrice
Zhan SHI	Examinateur

Contents

1 Introduction to lifetime analysis **19**
1.1 Lifetime data and independent right-censoring 19
1.1.1 Failure times 19
1.1.2 Independent random right-censoring 20
1.1.3 The Nelson-Aalen and Kaplan-Meier estimators 22
1.1.4 Martingale-based inference 23
1.1.5 Empirical processes-based inference 25
1.2 Competing risks 28
1.2.1 Definitions 28
1.2.2 The problem of identifiability 30
1.2.3 Inference for independent and dependent competing risks 30
1.3 Dependent competing risks in presence of independent random right-censoring 31
1.3.1 The Aalen-Johansen estimator 31
1.3.2 Martingale-based inference 34
1.4 Some parametric models 38
1.4.1 Univariate exponential distribution 38
1.4.2 Bivariate exponential distribution 39
1.4.3 Univariate exponential Weibull distribution 40

2 Strong approximations and LIL for processes based on the Aalen-Johansen estimator for the subdistribution functions **43**
2.1 Introduction 43
2.2 Main results 43
2.2.1 Strong approximations 44
2.2.2 LIL-type results 46
2.3 Proofs 47
2.3.1 Preliminary results 47
2.3.2 A first step toward the proof of Theorem 10 50
2.3.3 Proof of Theorem 10 53
2.3.4 Proofs of Theorems 11 and 12 65
2.3.5 Proof of Theorem 13 and 14 70
2.3.6 Proof of Proposition 5 72

3 Construction of joint confidence bands for the subdistribution functions 77
3.1 Introduction . . . 77
3.2 Weak convergence . . . 77
3.2.1 The main result . . . 77
3.2.2 Convergence of a useful martingale . . . 78
3.2.3 Proof of Theorem 17 . . . 80
3.3 Confidence bands . . . 85
3.4 Simulations . . . 90
3.4.1 A parametric model . . . 90
3.4.2 Design of the simulations . . . 92
3.4.3 Results and discussion . . . 94
3.5 Graphical displays . . . 95
3.5.1 ACBVW examples . . . 95
3.5.2 Simulated confidence bands . . . 96

4 Proportional subdistribution functions 105
4.1 Introduction . . . 105
4.2 The estimator and its strong uniform consistency . . . 107
4.3 A LIL type result . . . 109
4.4 A strong approximation result . . . 110
4.5 Weak convergence . . . 112

5 Recurrent events in the presence of a terminal event and independent right-censoring 115
5.1 Introduction and background . . . 115
5.2 Notations . . . 119
5.3 Functions of interest . . . 121
5.4 Estimation . . . 122
5.4.1 Estimation of the censoring distribution function . . . 123
5.4.2 Estimation with the first duration . . . 123
5.4.3 Estimation with the second duration . . . 125
5.4.4 Estimation with the k-th duration . . . 127
5.5 Simulations . . . 128
5.6 Application to real data: the CAPRIE study . . . 131
5.6.1 Description of the CAPRIE study . . . 131
5.6.2 The results of the CAPRIE study . . . 132
5.6.3 Contribution of the preceding methodology . . . 133
5.7 Graphical displays . . . 135
5.7.1 Simulation results . . . 135
5.7.2 Analysis of the CAPRIE study . . . 135

A Mathematical background 149

B Some R programs 163

Introduction

Le présent travail traite de l'analyse statistique non-paramétrique des durées de vie. L'analyse classique des durées de vie s'intéresse à la durée jusqu'à l'apparition d'un évènement d'intérêt, comme la durée de vie avant un décès dû à une certaine cause (cancer, maladie infectieuse, accident de la route...), la durée de réponse à un traitement, la durée avant le développement d'une pathologie particulière, etc... La durée de vie d'intérêt est alors modélisée par une variable aléatoire positive X dont on veut estimer la loi. Dans la pratique, il est courant qu'on ne puisse pas observer X directement. C'est le cas, par exemple, quand un individu quitte l'étude en cours avant la survenue de l'évènement d'intérêt. Dans ce cas, on sait seulement que la durée X entre le début de l'étude et l'évènement d'intérêt est supérieure à la durée passée dans l'étude. On modélise ce phénomène en supposant que l'on observe le minimum $\min(X, C)$ entre la variable d'intérêt X et une variable positive que l'on note C et que l'on nomme censure aléatoire droite. Dans ce cas, on supposera toujours que l'on observe également l'indicatrice $I(X \leq C)$. L'analyse classique des durées de vie se concentre alors sur les méthodes qui permettent d'estimer la loi de X à partir d'un échantillon censuré de la forme $(\min(X_i, C_i), I(X_i \leq C_i))$ pour $i = 1, \ldots, n$.

Dans le chapitre 1, nous proposons un exposé des notions, modèles et résultats classiques de l'analyse des durées de vie.

Une première extension du cas simple précédent consiste à considérer des situations dans lesquelles il n'y a plus un seul évènement d'intérêt mais plusieurs types d'évènements, chacun étant dû à un risque donné. Soit $\mathcal{J}$ le nombre total de risques et, pour $j = 1, \ldots, \mathcal{J}$, soit X_j la durée jusqu'à l'apparition d'un évènement dû au risque j en l'absence des autres risques. Le modèle des risques concurrents postule que l'on n'observe pas toutes les variables X_j pour $j = 1, \ldots, \mathcal{J}$ mais seulement leur minimum $\min(X_1, \ldots, X_{\mathcal{J}})$. Dans ce cas, on supposera toujours que l'on observe en plus une variable $\mathcal{C}$ qui prend la valeur j lorsque le minimum observé correspond à un évènement dû au risque j. L'exemple usuel de risques concurrents est celui de la population humaine qui est soumise à plusieurs causes de mort : un individu meurt une seule fois et par une seule cause. Les fonctions de répartition spécifiques à une cause donnée correspondant au délai jusqu'à l'avènement d'un évènement d'un type donné permettent de décrire l'évolution d'un risque donné en présence de tous les autres risques.

Dans les chapitres 2 et 3, nous apportons une contribution à l'étude du modèle général des risques concurrents. Puis, dans le chapitre 4, nous nous intéressons au modèle des risques concurrents à fonctions de répartition spécifiques à une cause donnée proportionnelles.

Une seconde extension consiste à considérer des évènements qui peuvent se répéter au cours du temps chez un même sujet (infections opportunistes chez les sidéens, récurrence de tumeurs chez les cancéreux...). Pour chaque individu, les données consistent alors en une suite de temps d'arrivée correspondant aux évènements successifs. Dans la pratique, l'observation de ces durées est souvent définitivement stoppée soit par une censure droite indépendante soit par un évènement terminal dépendant des évènements récurrents (décès). Les fonctions de répartition correspondant à la durée entre deux évènements récurrents successifs conditionnellement à la survie aux évènements précédents permettent de décrire l'évolution du risque.

Dans le chapitre 5, nous proposons une étude du modèle d'évènement répétés terminés par un décès en présence de censure.

Modèle de risques concurrents dépendants en présence de censure droite indépendante (chapitres 2 et 3)

La première partie de cette thèse concerne l'étude du modèle des risques concurrents dépendants que nous présentons ici. Introduisons une population d'individus soumis à $\mathcal{J}$ ($\mathcal{J} \geq 2$) risques qui ne sont pas supposés indépendants. Chacun de ces risques entraîne la réalisation d'évènements d'un type donné. Pour un individu donné, on n'observe pas tous les types d'évènements susceptibles de survenir mais seulement le premier évènement qui se réalise au cours du temps. Cet évènement est dû à un risque et un seul. Autrement dit, nous supposons dans toute la suite que deux évènements de type différent ne peuvent se produire simultanément chez un même individu. Dans une telle situation, on dit que les différents risques sont en concurrence. Dans cette situation, un couple de variable aléatoire $(X, \mathcal{C})$ est associé à chaque individu de la population. La variable aléatoire X représente la durée au bout de laquelle se produit l'évènement d'intérêt, elle est positive de fonction de répartition F définie pour $t \geq 0$ par $F(t) = \mathbb{P}[X \leq t]$. La variable aléatoire à valeurs entières $\mathcal{C}$ prend la valeur j lorsque l'évènement est dû au $j^{\text{ème}}$ risque. Elle indique donc le type de l'évènement, le risque à l'origine de l'évènement.

Dans ce contexte, on s'intéresse aux différentes fonctions de répartition spécifiques à une cause donnée, elles sont définies pour $j = 1, \ldots, \mathcal{J}$ et $t \geq 0$ par

$$F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j].$$

Pour $t \geq 0$, la quantité $F^{(j)}(t)$ représente la probabilité qu'un évènement de type j se produise avant l'instant t et que les autres types d'évènement n'aient pas encore eu lieu à cet instant t. Les fonctions $F^{(j)}$ sont des fonctions de répartition impropres puisqu'elles ne valent pas 1 en l'infini. Nous supposons dans la suite que les différentes fonctions $F^{(j)}$ pour des indices j différents ont des points de discontinuité distincts.

Dans la réalité, le couple $(X, \mathcal{C})$ peut être censuré par une variable positive C de loi G. Dans toute la suite, nous supposerons l'indépendance de $(X, \mathcal{C})$ et C. Lorsqu'il y a censure, nous savons que la durée d'intérêt X est supérieure à l'instant de censure C

observé mais la valeur exacte de X est inconnue. Bien sûr, aucune information n'est alors disponible sur $\mathcal{C}$, le risque associé à la durée X. Pour chaque individu, on observe alors un couple de variables aléatoires

$$\big(T = \min(X, C),\ J = \mathcal{C} I(X \leq C)\big)\,.$$

Soit un échantillon (T_i, J_i) pour $i = 1, \ldots, n$ de couples indépendants distribués comme (T, J). Notons $T_{1,n}, T_{2,n}, \ldots, T_{n,n}$ les statistiques d'ordre associées à l'échantillon $T_1, \ldots, T_n$. Par indépendance de X et C, les variables aléatoires $T_1, \ldots, T_n$ sont indépendantes et identiquement distribuées de loi H donnée par la relation $1-H = (1-F)(1-G)$. Notons τ_H la borne supérieure du support de H (qui est le minimum des bornes supérieures des supports de F et G) définie par $\tau_H = \sup\{x\,:\,H(x) < 1\}$. Il est important de remarquer qu'aucune observation n'est possible au-delà de ce point.

Aalen (1978a, 1978b) a introduit des processus stochastiques qui permettent de décrire les données de l'analyse des durées de vie. En s'appuyant sur la théorie des martingales à temps continu et sur l'intégrale stochastique, Aalen (1978a, 1978b) a proposé des estimateurs pour les fonctions d'intérêt de la statistique des durées de vie et a obtenu différents résultats asymptotiques. Lorsque le processus des états successifs occupés par un individu au cours du temps est un processus de Markov inhomogène, Aalen et Johansen (1978) ont introduit un estimateur des probabilités de transition entre états en présence de censure aléatoire droite indépendante. Ils ont notamment obtenu la consistence faible de cet estimateur sur un compact ainsi qu'un résultat de convergence faible sous hypothèse de continuité. En fait, la situation qui nous occupe est un cas particulier du cas de Aalen et Johansen puisque nous sommes en présence d'un processus de Markov comportant un état transcient et plusieurs états absorbants ($\mathcal{J}$). Ainsi, les fonctions $F^{(j)}$ sont estimables au moyen de l'estimateur correspondant de Aalen-Johansen qui est défini pour $t \geq 0$ par :

$$\widehat{F}_n^{(j)}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1 - \widehat{F}_n^{-}(T_i)}{1 - H_n^{-}(T_i)} I(T_i \leq t, J_i = j)\,,$$

où H_n^- est la modification continue à gauche de la fonction de répartition empirique H_n définie pour $t \geq 0$ par :

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i \leq t)\,.$$

L'expression de l'estimateur de Aalen-Johansen implique également la modification continue à gauche $\widehat{F}_n^-$ de l'estimateur de Kaplan-Meier de F qui est défini pour $t \geq 0$ par :

$$\widehat{F}_n(t) = 1 - \prod_{i=1}^{n}\left(1 - \frac{I(T_i \leq t,\ J_i \neq 0)}{n(1 - H_n^{-}(T_i))}\right)\,.$$

Nous rappelons ci-dessous quelques-uns des nombreux résultats qui ont été proposés pour l'estimateur de Kaplan-Meier. Ils illustrent la difficulté que l'on rencontre dans l'établissement de résultats asymptotiques près de la borne supérieure τ_H du support de H. Nous distinguons principalement trois approches.

Certains auteurs, comme Breslow et Crowley (1974), Földes et Rejtó (1981b) ou Major et Rejtő (1988), ont obtenu des résultats asymptotiques pour $\widehat{F}_n$ sur un intervalle compact strictement inclus dans le support de H. On se référera à Deheuvels et Einmahl (1996, 2000) pour des résultats très fins du type loi du logarithme itéré fonctionnelle et valables en un point donné ou sur un compact strictement inclus dans le support de H.

D'autres auteurs, comme Földes et Rejtó (1981a), Gill (1983), Csörgő et Horváth (1983b), Ying (1989) ou Chen et Lo (1997) ont obtenu des résultats asymptotiques sur le support de H tout entier mais sous des hypothèses sur les queues de distribution de F et G restrictives et surtout invérifiables en pratique. Ou alors, en l'absence d'hypothèse, les vitesses de convergence fournies peuvent ne pas converger vers 0. Il apparaît donc qu'il n'est pas possible d'obtenir des résultats sur l'intervalle $[0, \tau_H]$ sans imposer des conditions sur les queues de distribution.

Les résultats de Stute (1994a) constituent un compromis entre les résultats obtenus sans hypothèse sur un compact et les résultats obtenus sous hypothèse sur le support de H tout entier. En effet, Stute (1994a) a obtenu des résultats sur des intervalles aléatoires croissants qui peuvent asymptotiquement recouvrir tout compact inclus dans le support de H, et ce, sans hypothèse sur les queues de distribution de F et G. Ces intervalles sont les intervalles $[0, T_{n-k_n,n}]$ s'arrêtant à la $(n-k_n)^{\text{ème}}$ statistique d'ordre. La suite (k_n) est une suite d'entiers compris entre 1 et $n-1$ et, si on la choisit négligeable devant n, alors $T_{n-k_n,n}$ converge en probabilité vers τ_H. Csörgő (1996) et Giné et Guillou (1999) ont également adopté cette approche.

Nous montrons dans les théorèmes 10, 11 et 13 que l'estimateur de Aalen-Johansen se comporte asymptotiquement de la même façon que l'estimateur de Kaplan-Meier. Nous nous alignons pour cela sur la troisième approche présentée. Autrement dit, on s'attachera à établir des résultats sur les intervalles aléatoires croissants $[0, T_{n-k_n,n}]$.

Nous fournissons une première approximation forte pour les processus $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ de façon jointe pour $j = 1, \ldots, \mathcal{J}$. L'obtention des théorèmes 10, 11 et 13 (ainsi que des théorèmes 21 et 22 au chapitre 4) requiert une hypothèse notée $(\mathcal{H})$ qui est explicitée au début de la section 2.2 du chapitre 2. Il s'agit de conditions de régularité et de croissance sur la suite d'entiers (k_n).

Pour la clarté de l'exposé, nous introduisons les notations suivantes. Pour une suite (ξ_n) de variables aléatoires et pour une suite (a_n) de constantes positives, nous écrivons que $\xi_n = O(a_n)$ ou $\xi_n = o(a_n)$ lorsque $\limsup_{n\to\infty} |\xi_n|/a_n \leq C$ presque sûrement pour une constante déterministe et finie $C > 0$ ou pour $C = 0$ respectivement. Nous écrivons $\xi_n = O_{\mathbb{P}}(a_n)$ lorsque $\lim_{x\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\xi_n| > x a_n] = 0$ et $\xi_n = o_{\mathbb{P}}(a_n)$ lorsque $\limsup_{n\to\infty} \mathbb{P}[|\xi_n| > x a_n] = 0$ pour tout $x > 0$.

Théorème 10. *Soit (k_n) une suite d'entiers telle que $1 \leq k_n < n$ pour tout n et satisfaisant la condition $(\mathcal{H})$. Si, pour $j = 1, \ldots, \mathcal{J}$, $F^{(j)}$ est continue, alors, pour n assez grand, sur un espace de probabilité convenablement élargi, on peut définir $\mathcal{J}$ suites de*

processus gaussiens $(\bar{L}_n^{(1)}), \ldots, (\bar{L}_n^{(\mathcal{J})})$ *telles que l'on ait de façon jointe pour* $j = 1, \ldots, \mathcal{J}$,

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right) - \bar{L}_n^{(j)}(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_n} \right).$$

Pour n *fixé, les processus* $\bar{L}_n^{(j)}$ *sont gaussiens de moyenne nulle et de covariance donnée pour* $k, j = 1, \ldots, \mathcal{J}$ *et pour* $s, t \geq 0$ *par*

$$Cov\Big(\bar{L}_n^{(j)}(t), \bar{L}_n^{(k)}(s) \Big) =$$
$$\int_0^{s \wedge t} \sum_{l=1}^{\mathcal{J}} \left(I(j = l) + \frac{F^{(j)} - F^{(j)}(t)}{1 - F} \right) \left(I(k = l) + \frac{F^{(k)} - F^{(k)}(s)}{1 - F} \right) \frac{(1 - F)^2}{(1 - H^-)^2} dH^{(1,l)}$$

où $H^{(1,j)}$ *est la fonction définie par* $H^{(1,j)}(t) = \mathbb{P}[T \leq t, J = j]$ *pour* $j = 1, \ldots, \mathcal{J}$ *et pour* $t \geq 0$.

Le processus approximant pour $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ est un processus gaussien qui n'est pas un processus de Wiener, alors que le processus de Kaplan-Meier

$$K_n^{(0)} = \sqrt{n} \, \frac{\widehat{F}_n - F}{1 - F}$$

s'approxime fortement par un processus de Wiener pris en une certaine fonction de variance. L'intérêt d'une approximation par un processus de Wiener réside dans le fait que de nombreux résultats ont été établis pour le processus de Wiener et que ce dernier peut facilement être simulé ou tabulé. En fait, il apparaît qu'il est possible d'approximer $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ par un processus de Wiener si une certaine quantité lui est retranchée. Ainsi, les processus

$$K_n^{(j)} = \sqrt{n} \left(\widehat{F}_n^{(j)} - F^{(j)} \frac{1 - \widehat{F}_n}{1 - F} \right)$$

peuvent être approximés, de façon jointe pour $j = 1, \ldots, \mathcal{J}$, par un processus de Wiener pris en une certaine fonction de variance.

Théorème 11. *Soit* (k_n) *une suite d'entiers telle que* $1 \leq k_n < n$ *pour tout* n *et satisfaisant la condition* $(\mathcal{H})$. *Si, pour* $j = 1, \ldots, \mathcal{J}$, $F^{(j)}$ *est continue, alors, pour* n *assez grand, sur un espace de probabilité convenablement élargi, on peut définir* $\mathcal{J}$ *suites de processus gaussiens* $(\check{L}_n^{(1)}), \ldots, (\check{L}_n^{(\mathcal{J})})$ *telles que l'on ait de façon jointe pour* $j = 1, \ldots, \mathcal{J}$:

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t) \frac{1 - \widehat{F}_n(t)}{1 - F(t)} \right) - \check{L}_n^{(j)}(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_n} \right).$$

Pour chaque n, *les processus* $\check{L}_n^{(j)}$ *sont de moyenne nulle et de covariance pour* $k, j = 1, \ldots, \mathcal{J}$ *et pour* $s, t \geq 0$ *donnée par*

$$Cov\left(\check{L}_n^{(j)}(t), \check{L}_n^{(k)}(s) \right) =$$
$$\sum_{l=1}^{\mathcal{J}} \int_0^{s \wedge t} \left(I(j = l) + \frac{F^{(j)}}{1 - F} \right) \left(I(k = l) + \frac{F^{(k)}}{1 - F} \right) \frac{(1 - F)^2}{(1 - H^-)^2} dH^{(1,l)}.$$

Ces deux approximations sont valides sur des intervalles aléatoires croissants $[0, T_{n,n-k_n}]$ allant de 0 à la $(n-k_n)^{\text{ème}}$ statistique d'ordre de l'échantillon des durées réellement observées, sous des hypothèses de régularité pour la suite d'entiers (k_n). Comme pour le processus de Kaplan-Meier, l'erreur d'approximation est en $O\left(\sqrt{n}\log n/k_n\right)$. Les différentes fonctions $F^{(j)}$ sont supposées continues afin de pouvoir utiliser les résultats analogues que Csörgő (1996) et Giné et Guillou (1999) ont obtenu pour le processus de Kaplan-Meier.

Nous fournissons ensuite une vitesse de convergence de $\widehat{F}_n^{(j)}$ vers $F^{(j)}$ uniformément sur les intervalles $[0, T_{n,n-k_n}]$.

Théorème 13. *Soit (k_n) une suite d'entiers telle que $1 \leq k_n < n$ pour tout n et vérifiant la condition $(\mathcal{H})$ pour la partie presque-sûre. Supposons que pour $j = 1, \ldots, \mathcal{J}$, $F^{(j)}$ est continue. Alors,*

$$\sup_{t \leq T_{n-k_n,n}} \left| \widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right| = \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_n}}\right) , \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) . \end{cases}$$

La vitesse presque-sûre est la même que celle obtenue pour l'estimateur de Kaplan-Meier par Giné et Guillou (1999), à savoir $O\left((\log\log n/k_n)^{1/2}\right)$. Ainsi, si la suite (k_n) est choisie trop petite, ce qui correspond au cas où $T_{n,n-k_n}$ se rapproche rapidement de τ_H lorsque n tend vers l'infini, la vitesse de convergence s'effondre. A l'inverse, si la suite (k_n) est choisie suffisamment grande, ce qui correspond au cas où $T_{n,n-k_n}$ se rapproche lentement de τ_H lorsque n tend vers l'infini, la vitesse de convergence est bonne. Cela illustre la difficulté voire l'impossibilité que l'on a à fournir de bonnes vitesses de convergence uniformément sur le support de H tout entier sans effectuer d'hypothèse sur les queues de distribution. Nous raffinons ce résultat par un résultat du type loi du logarithme itéré valide sur l'intervalle $[0, T_{n-k_n,n}]$.

Introduisons pour $t \geq 0$

$$C(t) = \int_0^t \frac{dH^{(1)}}{(1-H^-)^2}$$

et pour $j = 1, \ldots, \mathcal{J}$

$$C^{(j)}(t) = \int_0^{s\wedge t} \sum_{l=1}^{\mathcal{J}} \left(I(j=l) + \frac{F^{(j)}}{1-F} \right)^2 \frac{(1-F)^2}{(1-H^-)^2} dH^{(1,l)}.$$

Théorème 14. *Soit (k_n) une suite d'entiers vérifiant $1 \leq k_n < n$ pour tout n et $k_n \geq \log n$ pour n assez grand ainsi que*

$$k_n = o(n) \quad et \quad \frac{\sqrt{n}\log n}{k_n\sqrt{\log\log n}} \overset{n\to\infty}{\longrightarrow} 0\,.$$

Si $F^{(j)}$ est continue, si $C^{(j)}(\tau_H) < \infty$ pour $j = 1, \ldots, \mathcal{J}$, on a :

$$\limsup_n \sup_{t \leq T_{n-k_n,n}} \frac{\sqrt{n}}{\sqrt{2 \log\log n}} \left| \widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right| \leq \sqrt{C^{(j)}(\tau_H)} + \sqrt{C(\tau_H)}.$$

Les conditions requises pour la suite d'entiers (k_n) au théorème 14 sont plus restrictives que l'hypothèse $(\mathcal{H})$ (détaillée au chapitre 2). Le théorème 14 s'interprète de la façon suivante. Sous l'hypothèse que la suite (k_n) tend suffisamment vite vers l'infini tout en restant négligeable devant n, nous obtenons une vitesse de convergence de $\widehat{F}_n^{(j)}$ vers $F^{(j)}$ de l'ordre de $\sqrt{\log\log n}$. En fait, l'hypothèse sur la suite (k_n) signifie que l'on contraint $T_{n-k_n,n}$ à se rapprocher très lentement de τ_H. C'est le prix à payer pour la bonne vitesse de convergence obtenue.

Comme l'estimateur de Kaplan-Meier, l'estimateur de Aalen-Johansen possède une structure de martingale (à temps continu). L'intérêt de cette propriété est que la théorie de l'intégrale stochastique devient applicable. En particulier, sous des conditions d'intégrabilité appropriées, l'intégrale d'un processus prévisible par rapport à une martingale reste une martingale. L'intérêt des martingales réside dans le fait que de nombreux résultats limites sont disponibles. Aalen and Johansen (1978) ont obtenu la convergence faible des processus $K_n^{(j)}$ pour $j = 1, \ldots, \mathcal{J}$ sous l'hypothèse de continuité des $F^{(j)}$ en appliquant le théorème de Rebolledo qui devient difficile à utiliser si l'on ne fait pas cette hypothèse. Dauxois (2000) a obtenu la convergence faible du processus $K_n^{(0)}$ basé sur l'estimateur de Kaplan-Meier sans hypothèse de continuité. Il a utilisé pour cela un théorème de Jakubowski et al. (1989). Sa méthode se généralise à notre cas multi-dimensionnel. Définissons :

$$\widetilde{K}_n^{(0)} = \sqrt{n} \, \frac{\widehat{F}_n - F}{1 - \widehat{F}_n}$$

et pour $j = 1, \ldots, \mathcal{J}$:

$$\widetilde{K}_n^{(j)} = \sqrt{n} \left(\widehat{F}_n^{(j)} \frac{1 - F}{1 - \widehat{F}_n} - F^{(j)} \right).$$

Pour $\sigma < \tau_H$, on obtient la convergence jointe pour $j = 0, \ldots, \mathcal{J}$ des processus $K_n^{(j)}$ et $\widetilde{K}_n^{(j)}$ dans l'espace $D^{\mathcal{J}+1}[0, \sigma]$ des fonctions càdlàg (continues à droite et admettant des limites à gauche) de $[0, \sigma]$ dans $\mathbb{R}^{\mathcal{J}+1}$. Les processus limites sont des processus de Wiener corrélés, chacun étant pris en une certaine fonction de variance facilement estimable.

Théorème 17. *Soit $\sigma < \tau_H$. Dans $D^{\mathcal{J}+1}[0, \sigma]$, on a les convergences suivantes :*

$$\left(K_n^{(0)}, K_n^{(1)}, \ldots, K_n^{(\mathcal{J})}\right) \xrightarrow{\mathcal{D}} \left(K^{(0)}, K^{(1)}, \ldots, K^{(\mathcal{J})}\right),$$

$$\left(\widetilde{K}_n^{(0)}, \widetilde{K}_n^{(1)}, \ldots, \widetilde{K}_n^{(\mathcal{J})}\right) \xrightarrow{\mathcal{D}} \left(K^{(0)}, K^{(1)}, \ldots, K^{(\mathcal{J})}\right).$$

Les $K^{(j)}$ *sont des processus gaussiens de moyenne nulle et de covariance pour* $k, j = 0, \ldots, \mathcal{J}$ *et pour* $s, t \geq 0$ *:*

$$Cov\left(K^{(j)}(s), K^{(k)}(t)\right) = \int_0^{s\wedge t} \sum_{l=1}^{\mathcal{J}} \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)\left(I(k=l) + \frac{F^{(k)}}{1-F}\right) \frac{(1-F^-)(1-F)}{(1-H^-)^2} dH^{(1,l)}$$

avec $F^{(0)} \equiv F$.

Ce résultat de convergence faible permet d'obtenir des bandes de confiance asymptotiques sur tout compact inclus dans le support de H de façon jointe pour les $F^{(j)}$ avec $j = 1, \ldots, \mathcal{J}$ et pour F. Nous obtenons quatre types de bandes de confiance : les bandes de type Hall-Wellner et de type Aalen-Nair ainsi que les bandes modifiées de type Hall-Wellner et de type Aalen-Nair.

Afin de simuler ces différentes bandes de confiance, nous introduisons un modèle paramétrique appelé modèle Absolutely Continuous Bivariate Weibull (ACBVW). Il comporte deux paramètres de courbure α et β et trois paramètres d'aplatissement λ_0, λ_1 et λ_2. Ce modèle permet de générer des couples $(X, \mathcal{C})$ pour $\mathcal{J} = 2$ de telle sorte que les deux fonctions de répartition spécifiques à une cause donnée $F^{(1)}$ et $F^{(2)}$ sont absolument continues et sans perte de mémoire. Les deux risques concurrents sont dépendants à moins que $\lambda_0 = 0$ et les deux fonctions $F^{(1)}$ et $F^{(2)}$ ne sont pas proportionnelles à moins que $\alpha = \beta$. Les simulations ont révélé en particulier l'importance du choix du compact sur lequel les bandes de confiance sont tracées.

Modèle de risques concurrents dépendants à fonction de répartition spécifiques à la cause proportionnelles en présence de censure droite indépendante (chapitre 4)

Nous considérons ici un cas particulier du modèle précédent. Nous faisons l'hypothèse que les différentes fonctions de répartition spécifiques à la cause sont proportionnelles. Gather et Pawlitschko (1998) ainsi que Geffray et Guilloux (2005) ont étudié ce modèle sous l'hypothèse d'indépendance des différents risques concurrents.

Sans perte de généralité, nous supposons que les fonctions $F^{(j)}$ pour $j \geq 2$ sont proportionnelles à $F^{(1)}$. Cela entraîne que les fonctions $F^{(j)}$ pour $j = 1, \ldots, \mathcal{J}$ sont proportionnelles à F avec $\alpha_{1,j}$ comme coefficient de proportionnalité. Nous souhaitons alors proposer un estimateur de $F^{(j)}$ basé sur un estimateur de $\alpha_{1,j}$ et sur un estimateur de F. En ce qui concerne l'estimation de F, on dispose comme auparavant de l'estimateur de Kaplan-Meier $\widehat{F}_n$ de F. En ce qui concerne l'estimation de $\alpha_{1,j}$, on voit que $\alpha_{1,j}$ est la proportion théorique d'observations dues à la cause j et n'est donc pas directement estimable. Nous réexprimons alors $\alpha_{1,j}$ sous la forme du rapport de deux proportions théoriques estimables à savoir

$$\alpha_{1,j} = \frac{\mathbb{P}[J = j]}{\mathbb{P}[J \neq 0]}\,.$$

Nous proposons alors d'estimer $F^{(j)}$ au moyen d'un estimateur semi-paramétrique noté $\widetilde{F}_n^{(j)}$ défini comme le produit $\widetilde{F}_n^{(j)} = \widehat{\alpha}_{1,j}\,\widehat{F}_n$ où $\widehat{\alpha}_{1,j}$ est le rapport de deux proportions empiriques à savoir

$$\widehat{\alpha}_{1,j} = \frac{\sum_{i=1}^n I(J_i = j)}{\sum_{i=1}^n I(J_i \neq 0)}\,.$$

Nous allons voir que les propriétés des $\widetilde{F}_n^{(j)}$ sont très similaires à celles des $\widehat{F}_n^{(j)}$ pour $j = 1, \ldots, \mathcal{J}$.

Tout d'abord, la vitesse de convergence de $\widetilde{F}_n^{(j)}$ vers $F^{(j)}$ uniformément sur les intervalles aléatoires croissants $[0, T_{n,n-k_n}]$ est la même que la vitesse de convergence de $\widehat{F}_n^{(j)}$ vers $F^{(j)}$ en probabilité et presque-sûrement.

Théorème 21. *Soit (k_n) une suite d'entiers tels que $1 \leq k_n < n$ pour tout n et satisfaisant la condition $(\mathcal{H})$ pour la partie presque-sûre. Si $F^{(j)}$ est continue pour $j = 1, \ldots, \mathcal{J}$, alors*

$$\sup_{t \leq T_{n-k_n,n}} \left| \widetilde{F}_n^{(j)}(t) - F^{(j)}(t) \right| = \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_n}}\right)\,, \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\,. \end{cases}$$

D'autre part, les processus $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ peuvent être approximés de façon jointe pour $j = 1, \ldots, \mathcal{J}$ par des processus gaussiens corrélés.

Théorème 22. *Soit (k_n) une suite d'entiers tels que $1 \leq k_n < n$ pour tout n et satisfaisant la condition $(\mathcal{H})$. On suppose que $F^{(j)}$ est continue pour $j = 1, \ldots, \mathcal{J}$. Pour n assez grand, il existe un espace de probabilité convenablement élargi sur lequel on a presque sûrement et de façon jointe pour $j = 1 \ldots, \mathcal{J}$:*

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n}\left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right) - \widetilde{L}_n^{(j)}(t) \right| = O\left(\sqrt{n}\frac{\log n}{k_n}\right).$$

Pour $j = 1, \ldots, \mathcal{J}$, les processus $(\widetilde{L}_n^{(j)})$ sont, à n fixé, gaussiens de moyenne nulle et de covariance définie pour $k, j = 1, \ldots, \mathcal{J}$ et pour $s, t \geq 0$ par :

$$\begin{aligned} Cov\left(\widetilde{L}_n^{(j)}(s), \widetilde{L}_n^{(k)}(t)\right) &= F(s)F(t)\frac{\alpha_{1,j}\left(I(k=j) - \alpha_{1,k}\right)}{\mathbb{P}[J \neq 0]} \\ &+ \alpha_{1,j}\alpha_{1,k}(1 - F(s))(1 - F(t)) \int_0^{s \wedge t} \frac{dH^{(1)}}{(1 - H^-)^2}\,. \end{aligned}$$

Dans le cadre de ce modèle, notre dernier résultat consiste en la convergence faible des processus $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ de façon jointe pour $j = 1, \ldots, \mathcal{J}$. Cette convergence a lieu dans l'espace $D^{\mathcal{J}}[0, \sigma]$ des fonctions càdlàg de $[0, \sigma]$ dans $\mathbb{R}^{\mathcal{J}}$. Les processus limites sont des processus gaussiens corrélés de fonctions de covariance facilement estimables mais ce ne sont pas des processus connus (en particulier, les processus limites ne sont pas des processus de Wiener). Par ailleurs, la théorie des martingales n'est pas applicable à cause

du $\widehat{\alpha}_{1,j}$ qui n'est ni prévisible, ni une martingale. On utilise alors la théorie des processus empiriques.

Théorème 23. *Soit $\sigma < \tau_H$. Dans l'espace $D^{\mathcal{J}}[0,\sigma]$ des fonctions càdlàg sur $[0,\sigma]$ à valeurs dans $\mathbb{R}^{\mathcal{J}}$, on a :*

$$\left(\sqrt{n}\left(\widetilde{F}_n^{(1)} - F^{(1)}\right), \ldots, \sqrt{n}\left(\widetilde{F}_n^{(\mathcal{J})} - F^{(\mathcal{J})}\right)\right) \xrightarrow{\mathcal{D}} \left(\widetilde{K}^{(1)}, \ldots, \widetilde{K}^{(\mathcal{J})}\right),$$

où les $\widetilde{K}^{(j)}$ sont des processus gaussiens de moyenne nulle et de covariance donnée pour $k, j = 1, \ldots, \mathcal{J}$ et pour $s, t \geq 0$ par :

$$\begin{aligned} Cov\left(\widetilde{K}^{(j)}(s), \widetilde{K}^{(k)}(t)\right) &= F(s)F(t)\frac{\alpha_{1,j}\left(I(k=j) - \alpha_{1,k}\right)}{\mathbb{P}[J \neq 0]} \\ &\quad + \alpha_{1,j}\alpha_{1,k}(1-F(s))(1-F(t))\int_0^{s\wedge t} \frac{1-F^-}{1-F}\frac{dH^{(1)}}{(1-H^-)^2}. \end{aligned}$$

Ainsi, si l'hypothèse de proportionnalité des différentes fonctions de répartition spécifiques à la cause est vérifiée, la variance asymptotique de $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ est inférieure à la variance asymptotique de $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$. Ainsi, si l'hypothèse de proportionnalité des différentes fonctions de répartition spécifiques à une cause donnée est vérifiée, nous disposons d'un estimateur qui a le même comportement asymptotique, la même vitesse de convergence que l'estimateur de Aalen-Johansen, mais qui est asymptotiquement plus efficace. En revanche, il n'est pas possible de construire des bandes de confiance basées sur l'estimateur $\widetilde{F}_n^{(j)}$, du moins avec une méthodologie identique au cas précédent.

Modèle d'évènements récurrents en présence d'un évènement terminal et de censure droite indépendante (chapitre 5)

Dans de nombreuses études longitudinales, les individus peuvent subir des évènements de manière répétée. Dans ce travail, nous considérons le cas des patients souffrant d'athérosclérose. L'athérosclérose se caractérise par la formation de plaques dans la paroi interne des artères. L'évolution de l'athérosclérose peut aboutir à la formation d'un caillot au contact d'une plaque. Ce caillot obstrue l'artère, ce qui entraîne une diminution partielle ou totale de la vascularisation d'aval. Il en résulte alors une inadaptation entre les besoins et les apports en oxygène dans les tissus : c'est l'accident ischémique. Les manifestations cliniques de l'athérosclérose sont diverses. Les trois localisations majeures du problème correspondent aux artères coronaires (infarcus du myocarde), aux carotides (accident vasculaire cérébral), et aux artères des membres inférieures (artérite), voir Cambou (1999), Teigner et Castaigne (1999).

L'analyse des évènements récurrents de l'athérosclérose est effectuée en regroupant les différents accidents ischémiques en deux catégories : les accidents ischémiques non-fatals (AINF) et les accidents ischémiques fatals ou décès, et ce, quelque soit la localisation des accidents ischémiques en question. En effet, lorsqu'on s'intéresse à l'efficacité d'un

traitement pour la prévention des accidents ischémiques, il semble intéressant de prendre en compte tous les types d'accidents ischémiques afin d'émettre un avis sur les qualités globales du traitement, celui-ci devant prévenir au maximum les risques de récidive sans engendrer d'autres risques pour le patient. Les décès sont distingués pour prendre en compte leur caractère terminal. En effet, la survenue du décès du patient empêche bien sûr toute récurrence ultérieure des accidents ischémiques.

Au cours d'une étude longitudinale, l'observation des accidents ischémiques est parfois définitivement empêchée par la survenue d'un évènement de censure. Des causes fréquentes de censure dans ce contexte sont la fin de l'étude ou le décès pour une raison indépendante de la maladie et du traitement.

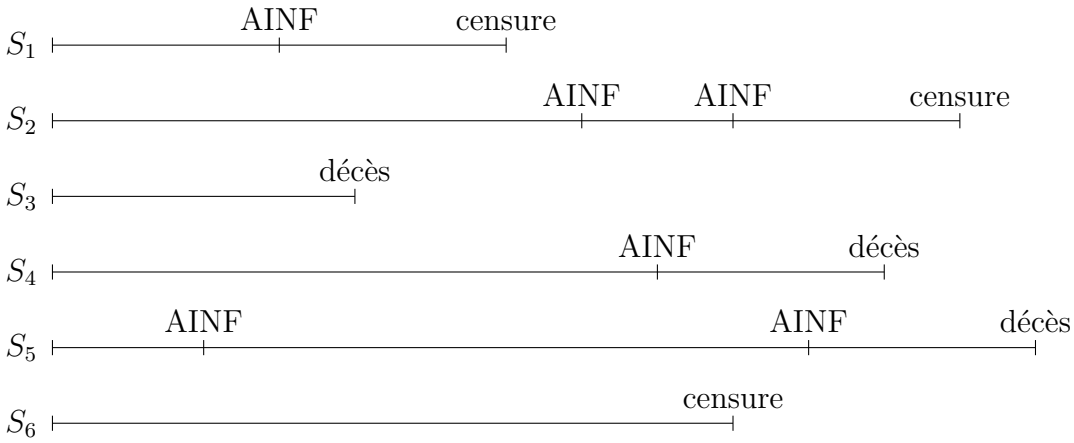

Figure 1: Exemple de données (AINF = Accident Ischémique Non-Fatal).

Les données recueillies au cours d'une étude longitudinale ressemblent typiquement à celles illustrées sur la figure . Il s'agit d'un genre particulier de données multivariées. En effet, notons tout d'abord que l'on ne dispose pas du même nombre de données pour les différents patients. Le nombre de patients encore vivants et encore présents dans l'étude diminue au fur et à mesure que les évènements se produisent. La survenue d'un AINF entraîne généralement une dégradation de l'état de santé et va ainsi de pair avec un risque accru de nouvel accident ischémique (fatal ou non). Ce mécanisme indique que les différents instants d'interarrivée d'accidents ischémiques (fatals ou non) pour un patient donné sont corrélés entre eux. En revanche, on suppose que la censure, qui survient au plus une fois pour un patient donné, est indépendante du processus des évènements récurrents et de décès. Il est important de noter que le dernier évènement que l'on enregistre pour un patient est soit une censure soit un décès.

On introduit $\widetilde{X}^{[k]}$ la v.a. positive représentant l'instant de survenue du $k^{\text{ème}}$ AINF. L'instant initial $\widetilde{X}^{[0]}$ est fixé à 0. Ces variables sont stochastiquement ordonnées : on a

presque-sûrement $\widetilde{X}^{[0]} < \widetilde{X}^{[1]} < \widetilde{X}^{[2]} < \dots$. On note Z la v.a. représentant l'instant de décès du patient (décès lié à l'athérosclérose) de fonction de répartition F_Z. Le nombre total d'accidents ischémiques (fatals ou non) que subit un patient donné est une v.a. que l'on note N.

La modélisation et l'estimation avec des données d'évènements récurrents, en présence ou non d'un évènement terminal et en présence ou non de censure, ont fait récemment l'objet de nombreux travaux. Dans l'analyse des données d'évènements récurrents, l'intérêt peut être porté sur la durée entre deux évènements successifs comme l'ont fait Wang et Chang (1999) et Lin et al. (1999), sur l'instant de survenue des évènements comme l'on fait Derzko et Leconte (2004) ou encore sur l'intensité du processus des évènements récurrents $N(t)$ qui compte le nombre d'évènements récurrents ayant eu lieu avant l'instant t. Derzko et Leconte (2004) ont développé un estimateur non-paramétrique pour la fonction de répartition de l'instant de survenue du $k^{\text{ème}}$ évènement récurrent en présence de mort. Lawless et Nadeau (1995) et Cook et Lawless (1997) se sont concentrés sur la fonction fréquence moyenne aussi appelée fonction moyenne cumulée définie pour $t \geq 0$ par $M(t) = \mathbb{E}[N(t)]$. Les deux approches généralement adoptées consistent à poser un modèle marginal ou un modèle de fragilité (dans lequel une variable latente est utilisée pour prendre en compte un effet aléatoire spécifique au patient). Des modèles de régression variés ont aussi été considérés pour répondre à différents types de questions, voir par exemple Prentice et al. (1981), Andersen et Gill (1982), Wei, Lin et Weissfeld (1989), Lin et al. (2000) ou Ghosh et Lin (2003). La pertinence dans le choix d'un modèle donné dépends de la nature des évènements récurrents ainsi que du centre d'intérêt de l'étude.

Ce chapitre s'appuye sur le travail de Lin et al. (1999) concernant la fonction de répartition associée à la durée entre deux évènements successifs conditionnellement au fait d'avoir expérimenté les évènements précédents avant un instant donné. Leur approche est adaptée pour prendre en compte le décès. Li et Lagakos (1997) et Derzko et Leconte (2004) ont considéré le décès comme un risque concurrent dépendant qui s'exerce à chaque récurrence de la maladie. C'est cette approche qui est adoptée ici par ce qu'elle correspond bien à la structure du problème. En effet, à chaque récurrence, à condition d'avoir survécu aux accidents ischémiques précédents et d'être encore dans l'étude, le patient est soumis à deux risques concurrents dépendants (le premier est "faire un accident ischémique", le second est "mourir d'un accident ischémique") en présence de censure aléatoire droite indépendante.

Pour prendre en compte les risques concurrents, on introduit les variables suivantes. Pour $k = 1, \dots, N$, considèrons $X^{[k]} = \widetilde{X}^{[k]} \wedge Z$ l'instant de survenue du $k^{\text{ème}}$ accident ischémique (fatal ou non). La nature du $k^{\text{ème}}$ accident ischémique est donnée par la v.a. à valeurs entières :

$$\mathcal{C}^{[k]} = \begin{cases} 1 & \text{si le } k^{\text{ème}} \text{ accident ischémique n'est pas fatal,} \\ 2 & \text{si le } k^{\text{ème}} \text{ accident ischémique est fatal.} \end{cases}$$

Pour $k = 1, \dots, N$, introduisons $Y^{[k]} = X^{[k]} - X^{[k-1]}$ la v.a. positive représentant la durée entre le $(k-1)^{\text{ème}}$ et le $k^{\text{ème}}$ accident ischémique. Bien sûr, on a $X^{[1]} = Y^{[1]}$ et

$X^{[k]} = \sum_{l=1}^{k} Y^{[l]}$. Les v.a. $Y^{[k]}$ sont corrélées entre elles et avec Z. Enfin, l'instant de censure est représenté par une v.a. positive C, de loi G, indépendante des v.a. $X^{[k]}$ (et donc des $Y^{[k]}$) et des v.a. $\mathcal{C}^{[k]}$.

Dans le présent contexte, les variables observables sont les suivantes.

- Le nombre total d'évènements observés pour un patient donné est représenté par une v.a. K positive à valeurs entières.
- L'instant de survenue du $k^{\text{ème}}$ évènement est représenté par une v.a. positive $T_X^{[k]}$.
- La durée entre le $(k-1)^{\text{ème}}$ et le $k^{\text{ème}}$ évènement est représenté par une v.a. positive $T_Y^{[k]}$.
- La v.a. $J^{[k]} = \mathcal{C}^{[k]} I\left(X^{[k]} \leq C\right) = \mathcal{C}^{[k]} I(\sum_{l=1}^{k} Y^{[l]} \leq C)$ indique la nature du $k^{\text{ème}}$ évènement.

On introduit enfin H la fonction de répartition définie par $1 - H = (1 - F_Z)(1 - G)$ ainsi que la borne supérieure droite du support de H notée par $\tau_H = \sup\{x : H(x) < 1\}$.

Lorsqu'on évalue l'efficacité d'un traitement pour la prévention des accidents ischémiques, il est souvent important d'évaluer si le traitement retarde le délai entre la déclaration de la maladie et le premier accident ischémique, celui entre le premier accident ischémique et le second et ainsi de suite. Nous préférons nous intéresser à la durée entre le $(k-1)^{\text{ème}}$ et le $k^{\text{ème}}$ accident ischémique plutôt qu'à la durée totale depuis le début de la prise du traitement jusqu'à la survenue du $k^{\text{ème}}$ accident ischémique. En effet, un traitement qui retarde la survenue du premier accident ischémique allongera inévitablement la durée totale depuis le début de la prise du traitement jusqu'à la survenue d'un second accident ischémique même s'il devient inefficace après la survenue du premier accident ischémique. De plus, dans certains cas, il peut exister un phénomène compensatoire entre les différents épisodes de la maladie. Par exemple, un traitement peut se révéler être plus efficace qu'un traitement de référence pour la première survenue d'accidents ischémiques mais moins efficace pour les récurrences ultérieures. Il est important de détecter un tel phénomène. Cela dit, les deux approches (qui consistent à s'intéresser respectivement aux durées inter-évènement et aux durées totales) peuvent être complémentaires.

Bien sûr, considérer la survenue d'un second accident ischémique n'a de sens que si le patient a survécu à l'accident ischémique précédent dont on sait qu'il a déjà eu lieu.

Enfin, un traitement peut diminuer l'incidence d'accidents ischémiques par une diminution de l'incidence des accidents ischémiques non-fatals ou par une diminution de l'incidence des décès. Comme ces deux types d'accidents ischémiques n'ont pas la même importance du point de vue du patient, il est intéressant de considérer ces deux aspects.

Par conséquent, on s'intéresse à la fonction de répartition associée à la durée entre le début de la prise du traitement et le premier accident ischémique, fonction que l'on décompose en deux sous-fonctions selon que l'accident ischémique est fatal ou non. Ce sont les fonctions de répartition spécifiques à une cause donnée dans le cadre des risques concurrents.

Nous considérons donc pour $t \geq 0$:

$$F^{[1]}(t) = \mathbb{P}\left[Y^{[1]} \leq t\right] ,$$

que l'on décompose en $F^{[1]}(t) = F^{[1(1)]}(t) + F^{[1(2)]}(t)$ où pour $t \geq 0$

$$F^{[1(1)]}(t) = \mathbb{P}\left[Y^{[1]} \leq t, \mathcal{C}^{[1]} = 1\right] ,$$

et

$$F^{[1(2)]}(t) = \mathbb{P}\left[Y^{[1]} \leq t, \mathcal{C}^{[1]} = 2\right] .$$

On s'intéresse aussi à la fonction de répartition associée à la durée entre le premier accident ischémique et le second conditionnellement au fait d'avoir eu le premier accident ischémique avant un instant donné et d'y avoir survécu. Cette fonction est décomposée à son tour en deux sous-fonctions selon que le second accident ischémique est fatal ou non. Nous considérons alors pour $t_1, t_2 \geq 0$:

$$F^{[2/1]}(t_2/t_1) = \mathbb{P}\left[Y^{[2]} \leq t_2 \middle| Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1\right] ,$$

que l'on décompose en $F^{[2/1]}(t_2/t_1) = F^{[2(1)/1]}(t_2/t_1) + F^{[2(2)/1]}(t_2/t_1)$ où

$$F^{[2(1)/1]}(t_2/t_1) = \mathbb{P}\left[Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 1 \middle| Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1\right] ,$$

et

$$F^{[2(2)/1]}(t_2/t_1) = \mathbb{P}\left[Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 2 \middle| Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1\right] .$$

On généralise facilement les résultats qui suivent aux récurrences ultérieures.

Les apports de ce chapitre sont les suivants. Tout d'abord, nous proposons des estimateurs fortement consistants des quantités d'intérêt exposées précédemment. Pour évaluer à distance finie le comportement de ces estimateurs, nous effectuons des simulations. Enfin, la méthodologie présentée est appliquée à un échantillon réel.

Supposons que l'on dispose d'un échantillon i.i.d. de n individus. Pour $i = 1, \ldots, n$, les données pour le $i^{\text{ème}}$ patient sont constituées d'un nombre K_i de couples où $K_i - 1$ est le nombre d'AINF observés. Pour $k = 1, \ldots, K_i$, le $k^{\text{ème}}$ couple est donné par $(T_{X,i}^{[k]}, J_i^{[k]})$ ou de manière équivalente par $(T_{Y,i}^{[k]}, J_i^{[k]})$ qui sont distribués respectivement comme $(T_X^{[k]}, J^{[k]})$ et $(T_Y^{[k]}, J^{[k]})$.

Avant d'estimer les quantités d'intérêt, notons que la fonction de répartition associée au décès (lié à l'athérosclérose) est directement estimable. En effet, le dernier évènement enregistré pour chaque patient est soit une censure soit un décès. Pour un patient donné, on n'observe pas à la fois décès et censure mais seulement le premier évènement qui survient au cours du temps et sa nature (décès ou censure). Comme il y a indépendance entre l'instant de décès et de censure, on se trouve exactement dans la situation où l'on peut estimer la fonction de répartition associée à la censure au moyen de l'estimateur de Kaplan-Meier calculé à partir des dernières observations de chacun des patients. Cet estimateur s'exprime donc sous la forme suivante pour $t \geq 0$:

$$\widehat{G}_n(t) = 1 - \prod_{i=1}^{n} \left(1 - \frac{I\left(T_{X,i}^{[K_i]} \leq t, J_i^{[K_i]} = 0\right)}{\sum_{l=1}^{n} I\left(T_{X,l}^{[K_l]} \geq T_{X,i}^{[K_i]}\right)}\right) .$$

Les fonctions d'intérêt précédentes ne sont pas directement estimables puisque les variables impliquées ne sont pas observables. En revanche, on peut les exprimer au moyen de fonctions directement estimables dont on injecte un estimateur pour finalement obtenir un estimateur des quantités d'intérêt. En ce qui concerne la première survenue d'évènements, cela mène à la forme suivante pour les estimateurs ($\widehat{G}_n^-$ est la modification continue à gauche de $\widehat{G}_n$). Ils sont définis pour $t \geq 0$ par :

$$\begin{aligned}
\widehat{F}_n^{[1]}(t) &= \int_0^t \frac{dH_n^{[1(1)]}}{1-\widehat{G}_n^-}\,, \\
\widehat{F}_n^{[1(1)]}(t) &= \int_0^t \frac{dH_n^{[1(1,1)]}}{1-\widehat{G}_n^-}\,, \\
\widehat{F}_n^{[1(2)]}(t) &= \int_0^t \frac{dH_n^{[1(1,2)]}}{1-\widehat{G}_n^-}\,.
\end{aligned}$$

Ces estimateurs sont exprimés en fonction de fonctions de répartition empiriques calculées à partir des observations de la première survenue d'évènements et définies pour $t \geq 0$ par :

$$\begin{aligned}
H_n^{[1(1)]}(t) &= \frac{1}{n}\sum_{i=1}^n I\left(T_{Y,i}^{[1]} \leq t, J_i^{[1]} \neq 0\right)\,, \\
H_n^{[1(1,1)]}(t) &= \frac{1}{n}\sum_{i=1}^n I\left(T_{Y,i}^{[1]} \leq t, J_i^{[1]} = 1\right)\,, \\
H_n^{[1(1,2)]}(t) &= \frac{1}{n}\sum_{i=1}^n I\left(T_{Y,i}^{[1]} \leq t, J_i^{[1]} = 2\right)\,.
\end{aligned}$$

Les trois estimateurs $\widehat{F}_n^{[1]}(t)$, $\widehat{F}_n^{[1(1)]}(t)$ et $\widehat{F}_n^{[1(2)]}(t)$ sont fortement consistents lorsque t reste strictement dans $[0, \tau_H]$.

En ce qui concerne la seconde durée d'intérêt, comme précédemment, on propose un estimateur "plug-in". Les estimateurs sont ainsi définis pour $t_1, t_2 \geq 0$ par :

$$\begin{aligned}
\widehat{F}_n^{[2/1]}(t_2/t_1) &= \frac{1}{\widehat{F}_n^{[1(1)]}(t_1)} \iint_{u\leq t_1, v\leq t_2} \frac{H_n^{[1(1,1),2(1)]}(du,dv)}{1-\widehat{G}_n^-(u+v)}\,, \\
\widehat{F}_n^{[2(1)/1]}(t_2/t_1) &= \frac{1}{\widehat{F}_n^{[1(1)]}(t_1)} \iint_{u\leq t_1, v\leq t_2} \frac{H_n^{[1(1,1),2(1,1)]}(du,dv)}{1-\widehat{G}_n^-(u+v)}\,, \\
\widehat{F}_n^{[2(2)/1]}(t_2/t_1) &= \frac{1}{\widehat{F}_n^{[1(1)]}(t_1)} \iint_{u\leq t_1, v\leq t_2} \frac{H_n^{[1(1,1),2(1,2)]}(du,dv)}{1-\widehat{G}_n^-(u+v)}\,.
\end{aligned}$$

Ces estimateurs sont exprimés en fonction de fonctions de répartition empiriques bivariées calculées à partir des observations des patients ayant eu au moins deux évènements et

définies pour $t \geq 0$ par :

$$H_n^{[1(1,1),2(1)]}(t_1,t_2) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t_1, T_{Y,i}^{[2]} \leq t_2, J_i^{[1]} = 1, J_i^{[2]} \neq 0\right),$$

$$H_n^{[1(1,1),2(1,1)]}(t_1,t_2) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t_1, T_{Y,i}^{[2]} \leq t_2, J_i^{[1]} = 1, J_i^{[2]} = 1\right),$$

$$H_n^{[1(1,1),2(1,2)]}(t_1,t_2) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t_1, T_{Y,i}^{[2]} \leq t_2, J_i^{[1]} = 1, J_i^{[2]} = 2\right).$$

Les trois estimateurs $\widehat{F}_n^{[2/1]}(t_2/t_1)$, $\widehat{F}_n^{[2(1)/1]}(t_2/t_1)$ and $\widehat{F}_n^{[2(2)/1]}(t_2/t_1)$ sont fortement consistents lorsque $t_1 + t_2$ reste strictement dans $[0, \tau_H]$.

Pour illustrer le comportement à distance finie de ces estimateurs, des simulations sont effectuées. Elles révèlent que la courbe estimée s'écarte de la courbe théorique dans la queue de distribution au fur et à mesure des récurrences. Cela est dû au poids de la censure qui augmente sur les queues de distribution au fur et à mesure des récurrences. En revanche, comme la censure privilégie l'observation des petites valeurs, le début des courbes est toujours très bien estimé.

L'étude CAPRIE a été mise en place pour évaluer l'efficacité relative du clopidogrel et de l'aspirine dans une population de patients présentant un risque vasculaire élevé. Il s'agit d'une étude multi-centrique de phase III randomisée en triple aveugle effectuée sur trois ans. Environ 20000 patients ont été randomisés. Le fichier de données de l'étude CAPRIE contient tous les évènements subis par chaque patient jusqu'à sa mort ou sa censure. A la première survenue d'un accident ischémique, on dispose, pour chaque groupe de traitement, d'environ 10000 patients. De même, pour l'estimation de la survie, on dispose également pour chaque groupe de traitement d'environ 10000 patients. En revanche, à la seconde survenue d'évènements, on ne dispose plus que de 10% des 10000 patients. A partir de la troisième survenue d'évènements, on ne dispose que moins de 3% des 10000 patients, ce qui est insuffisant pour obtenir des résultats convenables (si on trace les courbes d'incidence correspondant aux estimateurs précédants, elles sont indistinguables de l'axe des abscisses).

En terme de décès à la première survenue d'accidents ischémiques, les deux traitements semblent équivalents. En terme d'incidence d'un premier AINF, le clopidogrel semble légèrement plus efficace, ce qui est la conclusion à laquelle est arrivé le comité d'analyse de CAPRIE en s'appuyant sur le risque cumulé, voir Gent et al (1996). La deuxième survenue d'accidents ischémiques (que ce soit dans le cas d'AINF ou de décès) permet de mieux discriminer les deux traitements et donne l'avantage au clopidogrel.

Chapter 1

Introduction to lifetime analysis

1.1 Lifetime data and independent right-censoring

1.1.1 Failure times

Lifetime analysis (also referred to as survival analysis) is the area of statistics that focuses on analyzing the time duration between a given starting point and a specific event. This endpoint is often called *failure* and the corresponding length of time is called the *failure time* or *survival time* or *lifetime*.

Survival analysis finds applications in actuarial science, demography, epidemiology, medical research, reliability analysis and many other fields. Examples of failure times include the lifetimes of machine components in industrial reliability, the durations of strikes or periods of unemployment in economics, the times taken by individuals to complete a specific task in psychological experimentation, the lengths of tracks on a photographic plate in particle physics. In medical research, if the endpoint is the death of a patient, the resulting data are literally lifetimes. However, data of a similar form can be obtained when the endpoint is not fatal. Examples of lifetimes in clinical research include the time from start of treatment to the relief of a pain, the time from start of treatment to the recurrence of symptoms and, when studying an infectious disease, the time from onset to infection to onset of disease.

Formally, a failure time is a nonnegative random variable (r.v.) X that describes the length of time from a time origin until an event of interest occurs. We will suppose throughout that $\mathbb{P}[X < \infty] = 1$.

The most basic quantities used to summarize and describe the time elapsed from a starting point until the occurrence of an event of interest are the distribution function and the hazard function. The *cumulative distribution function* at time t, also called *lifetime distribution* or the *failure distribution*, is the probability that the failure time of an individual is less or equal than some value t. It is given for $t \geq 0$ by:

$$F(t) = \mathbb{P}[X \leq t] \,.$$

The function F is right-continuous, nondecreasing and satisfies $F(0) = 0$ and $F(\infty) = 1$. We denote by F^- the left-continuous function obtained from F in the following way:

$$F^-(t) = \lim_{u \uparrow t} F(u).$$

The distribution of X may equivalently be dealt with in terms of the *survival function* which is given, for $t \geq 0$, by:

$$1 - F(t) = \mathbb{P}[X > t]\,.$$

The *cumulative hazard function* is defined for $t \geq 0$ by:

$$\Lambda(t) = \int_0^t \frac{dF}{1 - F^-}\,. \tag{1.1}$$

When F is continuous, the relation $1 - F(t) = \exp(-\Lambda(t))$ is valid for all $t \geq 0$. We can then call Λ the *log-survival function.*

If F admits a derivative with respect to Lebesgue measure on $\mathbb{R}$, the *probability density function* exists and is defined for $t \geq 0$ by:

$$f(t) = \frac{dF(t)}{dt} = \lim_{h \to 0} \frac{\mathbb{P}[t \leq X < t + h]}{h}\,.$$

Heuristically, the function f may be seen as the instantaneous probability of experiencing the event.

With the same hypothesis of differentiability, the *hazard function* exists and is defined for $t \geq 0$ by:

$$\lambda(t) = \frac{f(t)}{1 - F^-(t)} = \lim_{h \to 0} \frac{1}{h} \frac{\mathbb{P}[t \leq X < t + h]}{\mathbb{P}[X \geq t]} = \lim_{h \to 0} \mathbb{P}[t \leq X < t + h | X \geq t].$$

The quantity $\lambda(t)$ can be interpreted as the instantaneous probability that an individual dies at time t, conditionally on he or she having survived until that time.

For an extensive introduction to lifetime analysis, the reader is referred e.g. to the books of Cox and Oakes (1984) and Kalbfleisch and Prentice (1980).

1.1.2 Independent random right-censoring

The main difficulty in the analysis of lifetime data lies in the fact that the actual failure times of some individuals may not be observed. An observation is *right-censored* if it is known to be greater than a certain value, but that the exact time is unknown. Let C be the nonnegative r.v. with distribution function G that stands for the censoring time of the individual. As before, the nonnegative r.v. X with distribution function F denotes the failure time of the individual. If X is censored, instead of X, we observe C which

gives the information that X is greater than C. In any case, the observable r.v. consist of

$$T = \min(X, C)\,,$$
$$D = I(X \leq C)\,,$$

where $I(.)$ denotes the indicator function. The nonnegative r.v. T stands for the observed duration of time which may correspond either to the event of interest ($D = 1$) or to a censoring time ($D = 0$).

To illustrate this, consider a clinical trial to compare therapies for lung cancer. Suppose that the time of interest X is the lifetime of patients having lung cancer. Reasons for censoring include the following examples. Failures may occur after the end of the study. Patients may experience a fatal road traffic accident. Patients may quit the study because of side-effects of the drug. Patients may move to another place in which case they can no longer be traced, they are lost to follow-up. All these events may prevent X from being observed in which case X is known to be greater than the observed censoring time.

An important assumption which will be made throughout is that the actual failure time of an individual X is independent of the censoring time C. For example, this is the case when censoring is due to end of study. However, this assumption cannot be made if the censoring time of an individual is the time of withdrawal from study due to side-effects of the drug under study. This type of censoring is known as dependent censoring. Great care should be taken to ensure that any censoring is independent, for otherwise the methods exposed hereafter are no longer valid.

In the sequel, it is assumed that X and C are independent. Consequently, the random variable T has distribution function H given by $1 - H = (1 - F)(1 - G)$. The following subdistribution functions of H will be needed:

$$H^{(0)}(t) = \mathbb{P}[T \leq t, D = 0]\,,$$
$$H^{(1)}(t) = \mathbb{P}[T \leq t, D = 1]\,.$$

The relation $H(t) = H^{(1)}(t) + H^{(0)}(t)$ is valid for any $t \geq 0$.

The relations that connect the distribution functions $H^{(0)}$, $H^{(1)}$ and to the distribution functions F and G are given by:

$$\begin{aligned} H^{(0)}(t) &= \int_0^t (1 - F)dG\,, \\ H^{(1)}(t) &= \int_0^t (1 - G^-)dF\,. \end{aligned}$$

The cumulative hazard function of X can be expressed as:

$$\Lambda(t) = \int_0^t \frac{dF}{1 - F^-} = \int_0^t \frac{dH^{(1)}}{1 - H^-}\,.$$

1.1.3 The Nelson-Aalen and Kaplan-Meier estimators

Initiated by Kaplan and Meier (1958), much effort has gone into the development of nonparametric methods for censored survival data. These methods are said to be nonparametric, since they do not require specific assumptions on the underlying distribution of the survival times. Kaplan and Meier (1958) introduced the product-limit estimator for the survival distribution function. The estimator of the cumulative hazard function is the Nelson-Aalen estimator introduced by Nelson (1969, 1972) and generalized by Aalen (1978a, 1978b). A vital assumption for the validity of the Kaplan-Meier estimator and the Nelson-Aalen estimator is that the censoring mechanism is independent of the failure mechanism as exposed in Subsection 1.1.2.

Let (T_i, D_i) for $i = 1, \ldots, n$ be n independent copies of the random vector (T, D). Let $T_{1,n} \leq T_{2,n} \leq \cdots \leq T_{n,n}$ be the order statistics associated to the sample $T_1, \ldots, T_n$. If there are ties between a failure time (or several failures times) and a censoring time, then the failure time(s) is (are) ranked ahead of the censoring time(s).

We define the empirical counterparts of H, $H^{(1)}$ and $H^{(0)}$ by:

$$\begin{aligned} H_n(t) &= \frac{1}{n}\sum_{i=1}^{n} I(T_i \leq t)\,, \\ H_n^{(1)}(t) &= \frac{1}{n}\sum_{i=1}^{n} I(T_i \leq t, D_i = 1)\,, \\ H_n^{(0)}(t) &= \frac{1}{n}\sum_{i=1}^{n} I(T_i \leq t, D_i = 0). \end{aligned}$$

The Kaplan-Meier product-limit estimator is defined for $t \geq 0$ by:

$$\widehat{F}_n(t) = 1 - \prod_{i=1}^{n}\left(1 - \frac{I(T_i \leq t,\, D_i = 1)}{n(1 - H_n^-(T_i))}\right). \tag{1.2}$$

The Nelson-Aalen estimator for Λ is then defined for $t \geq 0$ by:

$$\Lambda_n(t) = \int_0^t \frac{dH_n^{(1)}}{1 - H_n^-}\,.$$

The following relations are valid for $t \geq 0$:

$$\begin{aligned} H_n(t) &= H_n^{(1)}(t) + H_n^{(0)}(t)\,, \\ 1 - H_n(t) &= (1 - \widehat{F}_n(t))(1 - \widehat{G}_n(t))\,, \\ \Lambda_n(t) &= \int_0^t \frac{d\widehat{F}_n}{1 - \widehat{F}_n^-}\,, \end{aligned}$$

where $\widehat{G}_n$, the Kaplan-Meier estimator of G, is defined for $t \geq 0$ by:

$$\widehat{G}_n(t) = 1 - \prod_{i=1}^{n}\left(1 - \frac{I(T_i \leq t,\, D_i = 0)}{n(1 - H_n^-(T_i))}\right).$$

We will see hereafter that the Kaplan-Meier estimator has properties which generalize the properties of the usual empirical distribution function in presence of censoring. The asymptotic normality has been studied by Breslow and Crowley (1974), Gill (1980, 1983) and Dauxois (2000). The strong consistency has been obtained by Stute and Wang (1993), Gill (1994). Laws of the iterated logarithm have been obtained by Csörgő and Horvath (1983b), Csörgő (1996), Giné and Guillou (1999). At last, strong approximations by Gaussian processes have been obtained by Major and Rejtő (1988), Stute (1994a), Csörgő (1996) and Giné and Guillou (1999).

1.1.4 Martingale-based inference

Initiated by Aalen (1978b) and further developed by Gill (1980, 1983), the martingale approach has been applied successfully in the study of the Nelson-Aalen estimator for the cumulative hazard function and of the Kaplan-Meier estimator for the cumulative distribution function. The books of Shorack and Wellner (1986), Fleming and Harrington (1991) and Andersen et al. (1993) also provide a wealth of material. The role played by martingale theory in deriving asymptotic results is central here.

We introduce the σ-field generated by the events observed before t denoted by $\mathcal{F}_n^{(1)}(t)$:

$$\mathcal{F}_n^{(1)}(t) = \sigma\left\{I(T_i \leq t), T_i I(T_i \leq t), D_i I(T_i \leq t) \,:\, i = 1, \ldots, n,\, 0 \leq s \leq t\right\}.$$

For fixed n, the family $\mathbb{F}_n^{(1)} = \{\mathcal{F}_n^{(1)}(t) \,:\, t \geq 0\}$ is a filtration.

We introduce the counting process $N_n^{(1)}$ and the processes Y_n and J_n. The quantity $N_n^{(1)}(t)$ stands for the number of failures observed before time t while $Y_n(t)$ stands for the number of individuals at risk at time t which is the number of individuals neither dead nor censored at time t^-. These processes are defined for $t \geq 0$ by:

$$\begin{aligned} N_n^{(1)}(t) &= \sum_{i=1}^{n} I(T_i \leq t, D_i = 1)\,, \\ Y_n(t) &= \sum_{i=1}^{n} I(T_i \geq t) = n(1 - H_n^-(t))\,, \\ J_n(t) &= I(Y_n(t) > 0) = I(H_n^-(t) < 1)\,. \end{aligned} \tag{1.3}$$

In addition, let τ_H denote the right-endpoint of H, that is $\tau_H = \sup\{t \,:\, H(t) < 1\}$. For any distribution function F, we define the discontinuous part of F as $\Delta F = F - F^-$. For any process Z, we denote by Z^* the process stopped at $T_{n,n}$ the last order statistic. It is defined for $t \geq 0$ by $Z^*(t) = Z(t \wedge T_{n,n}) = \int_0^t J_n dZ$.

It turns out that the Kaplan-Meier estimator has a martingale structure which is outlined in the next theorem.

Theorem 1. (Gill 1980)
1. *For $t \geq 0$, define*

$$M_n^{(1)}(t) = N_n^{(1)}(t) - \int_0^t Y_n d\Lambda \,. \tag{1.4}$$

Then $\{M_n^{(1)}(t), \mathcal{F}_n^{(1)}(t) : t \geq 0\}$ is a local square integrable martingale in t on $[0, \tau_H]$ with predictable variation process given, for $t \geq 0$, by:

$$\left\langle M_n^{(1)} \right\rangle (t) = \int_0^t Y_n(u)(1 - \Delta\Lambda(u))d\Lambda(u) \,.$$

2. *The following relation holds in $[0, \tau_H]$:*

$$\frac{\widehat{F}_n(t) - F^*(t)}{1 - F^*(t)} = \int_0^t \frac{1 - \widehat{F}_n^-}{1 - F} \frac{J_n}{Y_n} dM_n^{(1)} \,. \tag{1.5}$$

This entails that the process

$$\left\{ \frac{\widehat{F}_n(t) - F^*(t)}{1 - F^*(t)}, \, \mathcal{F}_n(t) : \, t \geq 0 \right\}$$

is a local square integrable martingale in t on $[0, \tau_H]$.

Gill (1994) showed that the Kaplan-Meier mean of a nonnegative function forms a (discrete) reverse supermartingale in n. Using this property, Gill obtained the strong consistency of the Kaplan-Meier estimator uniformly on a maximal time interval (without making assumptions on the underlying distribution).

Theorem 2. Gill (1994) *Let $\mathcal{I} = \{t : H^-(t) < 1\}$. The following convergence holds almost-surely.*

$$\sup_{t \in \mathcal{I}} \left| \widehat{F}_n(t) - F(t) \right| \to 0 \,.$$

A function is said to be *càdlàg* if it is right-continuous with left-hand limits at any point. The Skorohod space of càdlàg functions from $[0, \sigma]$ (resp. $[0, \tau_H]$) to $\mathbb{R}$ is denoted by $D[0, \sigma]$ (resp. $D[0, \tau_H]$).

Under the assumption of continuity of F, Breslow and Crowley (1974) obtained the weak convergence of $\sqrt{n}\,(\Lambda_n - \Lambda)$ (resp. $\sqrt{n}(\widehat{F}_n - F)/(1 - F)$) to a Gaussian process in $D[0, \sigma]$ for any $\sigma < \tau_H$. Using the martingale approach and, in particular, Rebolledo's theorem, Gill (1980, 1983) and Ying (1989) extended these results to $D[0, \tau_H]$ under the assumption of continuity of F and under the assumption that

$$\int_0^{\tau_H} \frac{dF}{1 - G^-} < \infty \,.$$

Dauxois (2000) obtained the weak convergence of the processes $\sqrt{n}\,(\Lambda_n - \Lambda)$ and $\sqrt{n}(\widehat{F}_n - F)/(1-F)$ in $D[0,\sigma]$ for $\sigma < \tau_H$ without continuity assumption. The main step of his method consists in applying Theorem 2.6 of Jakubowski et al. (1989) on the convergence in law for sequences of stochastic integrals on the space $D[0,\sigma]$ instead of Rebolledo's theorem which is hard to apply without the assumption of continuity.

Theorem 3. (Dauxois 2000) *Let W denote the Brownian motion on $[0,\infty[$. The following convergence results hold in $D[0,\sigma]$ for $\sigma < \tau_H$:*

$$\sqrt{n}(\Lambda_n - \Lambda) \xrightarrow{\mathcal{D}} WoA \quad \textit{and} \quad \sqrt{n}\left(\frac{\widehat{F}_n - F}{1-F}\right) \xrightarrow{\mathcal{D}} WoC\,,$$

where

$$A(t) = \int_0^t \frac{1-\Delta\Lambda}{1-H^-} d\Lambda = \int_0^t \frac{1-F}{(1-F^-)^3}\frac{dF}{1-G^-}\,, \tag{1.6}$$

and

$$C(t) = \int_0^t \frac{d\Lambda}{(1-\Delta\Lambda)(1-H^-)} = \int_0^t \frac{1}{(1-F^-)(1-F)}\frac{dF}{1-G^-}\,. \tag{1.7}$$

1.1.5 Empirical processes-based inference

Empirical processes-based inference for the product-limit estimator have drawn considerable attention from many authors. Before starting the exposition of the results, we recall some notations. For a sequence of random variables (ξ_n) and a sequence of positive constants (a_n), we say that $\xi_n = O(a_n)$ or $\xi_n = o(a_n)$ if $\limsup_{n\to\infty} |\xi_n|/a_n \leq C$ almost surely for some finite deterministic constant $C > 0$ or for $C = 0$ respectively. We write $\xi_n = O_{\mathbb{P}}(a_n)$ if $\lim_{x\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\xi_n| > x a_n] = 0$ and $\xi_n = o_{\mathbb{P}}(a_n)$ if $\limsup_{n\to\infty} \mathbb{P}[|\xi_n| > x a_n] = 0$ for any $x > 0$.

Some authors obtained results on a fixed compact interval $[0,\sigma]$ with $\sigma < \tau_H$. For example, Földes and Rejtó (1981b) obtained a LIL-type upper bound on $[0,\sigma]$. Major and Rejtő (1988) obtained a strong approximation result for $\sqrt{n}(\widehat{F}_n - F)$ on $[0,\sigma]$. Such a choice asymptotically excludes a fixed proportion of the data.

Földes and Rejtó (1981a) got a LIL-type result on the whole real line provided $\tau_F < \tau_H$ where τ_F is the right-endpoint of F ($\tau_F = \sup\{t : F(t) < 1\}$). But such an assumption guarantees that on the whole support of F there is no uncontrolled increase of the bias due to censoring effect. This seems very restrictive. Csörgő and Horváth (1983b) found a rate of convergence of $\widehat{F}_n$ to F which is given on the whole real line without assumptions on the tails. But the bound which is explicitly given in function of F and G may not

converge to zero at all. Relying on combinatorial and analytic calculations, Stute and Wang (1993) obtained the following result. It means that, when F and G do not have jumps in common, the estimator $\widehat{\widehat{F}}_n$ is consistent for F on $[0, \tau_H]$ if and only if either $\Delta F(\tau_H) = 0$ or $\Delta F(\tau_H) > 0$ but $G(\tau_H) < 1$.

Theorem 4. (Stute and Wang 1993) *The following result holds almost surely as n goes to infinity:*

$$\sup_{t \leq \tau_H} |\widehat{F}_n(t) - \tilde{F}(t)| \to 0 \quad \textit{with} \quad \tilde{F}(t) = \begin{cases} F(t) & \textit{if } t < \tau_H, \\ F(\tau_H^-) + I(\tau_H \in D_H)\Delta F(\tau_H) & \textit{if } t \geq \tau_H, \end{cases}$$

where D_H denotes the set of discontinuity points of H.

Chen and Lo (1997) obtained the following LIL-type result on $[0, \tau_H]$ under the assumption of continuity of F.

Theorem 5. (Chen and Lo 1997) *Assume that $F(\tau_H) < 1$ and that there exists a p in $]0, 1/2[$ such that $\int_0^{\tau_H} (1 - G^-)^{-p/(1-p)} dF < \infty$. If F is continuous, then*

$$\sup_{t \leq \tau_H} \left| \widehat{F}_n(t) - F(t) \right| =$$
$$\begin{cases} o(n^{-p}) & \textit{if and only if } \int_0^{\tau_H} (1 - G^-)^{-\frac{p}{1-p}} dF < \infty\,, \\ O_{\mathbb{P}}(n^{-p}) & \textit{if and only if } \limsup_{t \to \tau_H} \left(\int_t^{\tau_H} (1 - G^-) dF \right)^{1-p} / (1 - G^-(t)) < \infty\,, \\ o_{\mathbb{P}}(n^{-p}) & \textit{if and only if } \lim_{t \to \tau_H} \left(\int_t^{\tau_H} (1 - G^-) dF \right)^{1-p} / (1 - G^-(t)) = 0\,. \end{cases}$$

They noticed that this result holds for F and G possibly discontinuous on $[0, \tau_H[$ (but not on $[0, \tau_H]$) and that, for F and G possibly discontinuous, $F(t)$ is estimable only for $t < \tau_H$. Furthermore, $F(\tau_H)$ may not be estimable if $\mathbb{P}[X = \tau_H] > 0$ and $\mathbb{P}[C = \tau_H] = 0$. In general, the rate of uniform convergence for discontinuous F should be considered over the interval $[0, \tau_H[$ instead of $[0, \tau_H]$.

It appears that we can get neither the strong uniform consistency nor LIL-type results on the whole support $[0, \tau_H]$ without hypothesis on the tail distributions, hypothesis uncheckable in practice and restrictive.

Stute (1994a), followed by Csörgő (1996), Giné and Guillou (1999) turned to a compromise. They got asymptotic results for the Nelson-Aalen and the Kaplan-Meier estimators on increasing intervals that may asymptotically cover each $[0, \sigma]$ for $\sigma < \tau_H$. No assumption on the distribution tails of F and G is needed. These increasing intervals are determined by the data. Indeed, they consists of $[0, T_{n-k_n,n}]$ where $T_{n-k_n,n}$ is the $(n - k_n)$-th order statistic of the sample $T_1, \ldots, T_n$ with (k_n) being a sequence of integers such that $1 \leq k_n < n$. If k_n is chosen to be negligible with respect to n, then $T_{n-k_n,n}$ converges in probability to τ_H as n goes to ∞. In order to get almost-sure asymptotic results, (k_n) is assumed to fulfill some growing hypothesis. Throughout the sequel, we refer to the

following conditions:

($\mathcal{H}1$): *for n large enough, the sequence (k_n/n) is nonincreasing and $k_n \geq \log n$,*

($\mathcal{H}2$): *for n large enough, the sequence (k_n/n) is nonincreasing and there exists a constant $C > 0$ such that $k_n \geq C d_n \log n$ with (d_n) is a nonincreasing sequence such that:*

$$\sum \frac{1}{k d_{2^k} \log k} < \infty \,,$$

e.g. $d_n = (\log\log\log n)^{1+\varepsilon}$, $d_n = (\log\log\log n)(\log\log\log\log n)^{1+\varepsilon}$, etc...

Condition ($\mathcal{H}1$) is required when applying the results of Giné and Guillou (1999), while Condition ($\mathcal{H}2$) is required when applying the results of Csörgő (1996).

The following result formulates a LIL-type result on the mentioned increasing intervals.

Theorem 6. (Csörgő 1996; Giné and Guillou 1999) *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ and, for the almost sure results, satisfying ($\mathcal{H}2$). We have:*

$$\sup_{t \leq T_{n-k_n,n}} |\Lambda_n(t) - \Lambda(t)| \;=\; \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_{2n}}}\right) \,, \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) \,. \end{cases}$$

If, in addition, F is assumed continuous, then we also have:

$$\sup_{t \leq T_{n-k_n,n}} \left| \frac{\widehat{F}_n(t) - F(t)}{1 - F(t)} \right| \;=\; \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_{2n}}}\right) \,, \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) \,. \end{cases}$$

The condition of continuity of F is required to linearize the Kaplan-Meier process. Indeed, if F is continuous, then $(\widehat{F}_n - F)/(1-F)$ can be approximated by $\Lambda_n - \Lambda$ on the random interval $[0, T_{n-k_n,n}]$. Precisely, we have the following result.

Proposition 1. (Giné and Guillou 1999) *Let (k_n) be a sequence of integers satisfying $1 \leq k_n < n$ and Hypothesis ($\mathcal{H}1$). If F is continuous, then*

$$\sup_{t \leq T_{n-k_n,n}} \left| \frac{\widehat{F}_n(t) - F(t)}{1 - F(t)} - (\Lambda_n(t) - \Lambda(t)) \right| = O\left(\frac{\log\log n}{k_n}\right) \,.$$

At last, we quote a strong approximation result for the Nelson-Aalen and the Kaplan-Meier processes by means of Gaussian processes on the random increasing intervals $[0, T_{n-k_n,n}]$.

Theorem 7. (Csörgő 1996) *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ and fulfilling Hypothesis ($\mathcal{H}1$). On a suitably enlarged probability space, one can define a process L_n such that we have almost surely:*

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n}(\Lambda_n(t) - \Lambda(t)) - L_n(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_{2n}} \right).$$

The process L_n is Gaussian with mean 0 and has covariance function given for $s, t \geq 0$ by:

$$Cov\Big(L_n(t), L_n(s) \Big) = \int_0^{s \wedge t} (1 - \Delta\Lambda) \frac{d\Lambda}{1 - H^-}.$$

Furthermore, if F is continuous, then, on the same probability space, we also have almost surely:

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \, \frac{\widehat{F}_n(t) - F(t)}{1 - F(t)} - L_n(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_{2n}} \right).$$

The first part of this theorem relies on a decomposition of $\Lambda_n - \Lambda$ by means of empirical processes. This decomposition is given uniformly for t in $[0, T_{n-k_n,n}]$ by:

$$\sqrt{n}(\Lambda_n(t) - \Lambda(t)) = \int_0^t \frac{d\alpha_n^{(1)}}{(1 - H^-)} + \int_0^t \frac{\alpha_n^-}{(1 - H^-)^2} dH^{(1)} + O\left(\sqrt{n} \frac{\log n}{k_{2n}} \right), \qquad (1.8)$$

where $\alpha_n = \sqrt{n}(H_n - H)$ and $\alpha_n^{(1)} = \sqrt{n}(H_n^{(1)} - H^{(1)})$. The empirical processes α_n and $\alpha_n^{(1)}$ are then approximated by means of Brownian bridges.

1.2 Competing risks

1.2.1 Definitions

The model of competing risks have been widely studied in the literature, see e.g. Tsiatis (1975), Elandt-Johnson and Johnson (1980), Andersen et al. (1993), Crowder (2001). Competing risks problems are often formulated in terms of potential or latent failure times corresponding to the different failure types. Let $\mathcal{J}$ be the number of risks acting on the population. For $j = 1, \ldots, \mathcal{J}$, let X_j be the r.v. representing the lifetime of an individual exposed to the risk of death from cause j only. We assume that the X_j are proper, so that even in the absence of any other risk, an individual will experience an event of cause j eventually. The random variables $(X_1, \ldots, X_{\mathcal{J}})$ are called the *net lifetimes* corresponding to the $\mathcal{J}$ postulated possible causes of failure. We assume that each failure is due to a single cause and that the occurrence of an event of a given type precludes the other events from occurring. Consequently, for each individual, we cannot observe $(X_1, \ldots, X_{\mathcal{J}})$ jointly

but only the smallest X_j. For each individual, the observable r.v. consists of the overall failure time X and the cause of failure $\mathcal{C}$. They are given by:

$$\begin{aligned} X &= \min(X_1, \ldots, X_{\mathcal{J}}), \\ \mathcal{C} &= j \quad \text{if } X = X_j. \end{aligned}$$

The main example of competing risks is the human population which is at risk of dying from different causes (traffic accident, cancer, myocardial infarction, etc...). Each individual dies only once from a given cause. Consequently, a single failure time and a cause are recorded for each individual.

The vector $(X_1, \ldots, X_{\mathcal{J}})$ has marginal distribution functions given for $j = 1, \ldots, \mathcal{J}$ and $t \geq 0$ by:

$$F_j(t) = \mathbb{P}[X_j \leq t]. \tag{1.9}$$

In actuarial and demographic terminologies, the marginal distribution functions F_j are termed as the *net distribution functions*.

Prentice et al. (1978) have raised the analysis of competing risks in terms of the cause-specific functions. The subdistribution functions

$$F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j] \tag{1.10}$$

are called the *crude distribution functions* or *cause-specific distribution functions* or *sub-distribution functions* or *cumulative incidence function*. The quantity $F^{(j)}(t)$ describes the probability of experiencing an event from cause j before age t when all the causes are acting simultaneously.

The overall distribution function is given for $t \geq 0$ by:

$$F(t) = \mathbb{P}[X \leq t]. \tag{1.11}$$

As failures from different causes cannot occur simultaneously, the cause-specific distribution functions fulfill an additivity property, namely $F = \sum_{j=1}^{\mathcal{J}} F^{(j)}$.

The cumulative hazard rate pertaining to the latent variables X_j is called the *net cumulative hazard rate* and is defined for $t \geq 0$ by:

$$\Lambda_j(t) = \int_0^t \frac{dF_j(u)}{1 - F_j^-(u)}. \tag{1.12}$$

The cumulative hazard rate for cause j in the presence of all other causes acting simultaneously in a population is called the *crude cumulative hazard rate function* or *cause-specific cumulative hazard rate function* and is defined for $t \geq 0$ by:

$$\Lambda^{(j)}(t) = \int_0^t \frac{dF^{(j)}(u)}{1 - F^-(u)}\,. \tag{1.13}$$

The overall cumulative hazard rate function is given for $t \geq 0$ by:

$$\Lambda(t) = \int_0^t \frac{dF(u)}{1 - F^-(u)}. \tag{1.14}$$

As before, as failures from different causes are mutually exclusive events, the crude hazard rates fulfill an additivity property $\Lambda = \sum_{j=1}^{\mathcal{J}} \Lambda^{(j)}$.

Note that, in general, $\Lambda^{(j)} \neq \Lambda_j$. But if the risks are independent, as the overall survival time is defined by $X = \min(X_1, \ldots, X_{\mathcal{J}})$, the relation $1 - F = \prod_{j=1}^{\mathcal{J}} (1 - F_j)$ is satisfied. We also have the relation $F^{(j)} = \int_0^{\cdot} \prod_{k \neq j} (1 - F_k) dF_j$. Consequently, if the risks are independent, then $\Lambda^{(j)} = \Lambda_j$.

1.2.2 The problem of identifiability

The independent risks model postulates that the X_j are stochastically independent of each other. Several authors have shown that based on data from competing risks, the assumption of independence of the different risks is not testable because there is no way to distinguish between independent or dependent latent lifetimes. Tsiatis (1975) showed that without the hypothesis of independence of the different risks, the model of latent lifetimes is unidentifiable. Indeed, the set of crude distribution functions is consistent with an infinity of joint distributions of latent lifetimes. To each dependent-risks model, there corresponds a unique independent-risks model with the same subdistribution functions. But each independent-risks model has a whole class of satellite dependent-risks models. When the risks are dependent, Klein and Moeschberger (1984) showed that the product-limit estimator for the marginal distribution function pertaining to a given risk converges with probability one to a function which may not be the same as the marginal function of interest. We also refer to Crowder (2001) to see that deriving the F_j from the $F^{(j)}$ cannot be done unless assuming the independence of the different risks or specifying parametric assumptions for example.

The assumption that the risks act independently, i.e., that the latent failure times X_j are independent would appear to be especially dubious in medical applications, which is ironically the context in which the competing risks model was developed. The complex biochemical interactions between different disease processes and physiological conditions would seem to rule out any hope of independence at the outset, making such an assumption not even worthy of tentative consideration. To assume independence, one must be sure that a failure of one type has absolutely no bearing at all on the likelihood of failure of any other type, not even through some indirect link.

1.2.3 Inference for independent and dependent competing risks

Historically, the main aim of competing risks was seen as the estimation of the marginal distributions. This consists of taking data in which the risks act together and trying to

infer how some of them would act in isolation i.e. to make inference free of the "nuisance" aspects. This is an attempt to make inference about the net risks from observations on the crude risks. It amounts to deriving the F_j from the observations of $(X, \mathcal{C})$. This approach is justified if the independence of the risks can be assumed. Otherwise, the counter-argument is that the F_j do not describe events that physically occur, they only describe failures from some isolated cause in situations in which all the other risks have been removed somehow. If the risks are dependent, it is the cause-specific distribution function $F^{(j)}$, not the net distribution function F_j, that is relevant to the real situation describing failures from cause j that can occur in the presence of all the other risks. Prentice et al. (1978) emphasized these points in their criticism of the traditional approach of competing risks. Cause-specific subdistribution functions are the basic estimable quantities in the dependent competing risks framework. But this does provide no information about the joint distribution of the lifetimes.

Let $(X_i, \mathcal{C}_i)$ for $i = 1, \ldots, n$ be an i.i.d. sample distributed as $(X, \mathcal{C})$. If we are prepared to assume the independence of the risks, as explained above, the function of interest is the marginal distribution function F_j. The product-limit estimator given for $t \geq 0$ by

$$\widehat{F}_{n,j}(t) = 1 - \prod_{i=1}^{n} \left(1 - \frac{I(X_i \leq t, \mathcal{C}_i = j)}{n(1 - H_n^-(X_i))}\right)$$

is a consistent estimator of F_j, all the remaining risks being treated as independent right-censoring. The overall Kaplan-Meier estimator of F denoted by $\widehat{F}_n$ then fulfills the relation $\widehat{F}_n = 1 - \prod_{j=1}^{\mathcal{J}}(1 - \widehat{F}_{n,j})$.

If the independence of the risks cannot be assumed, the function of interest is $F^{(j)}$. The cause-specific distribution functions may be consistently estimated by their empirical counterparts defined for $t \geq 0$ and $j = 1, \ldots, \mathcal{J}$ by:

$$F_n^{(j)}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t, \mathcal{C}_i = j)\,.$$

1.3 Dependent competing risks in presence of independent random right-censoring

1.3.1 The Aalen-Johansen estimator

We consider a population in which each subject is exposed to $\mathcal{J}$ mutually exclusive competing risks which are dependent. In this set-up, to each individual is associated a couple $(X, \mathcal{C})$ of random variables. The r.v. X stands for the failure time of the individual, it is positive with distribution function F. The r.v. $\mathcal{C}$ indicates the cause of failure, i.e. $\mathcal{C}$ takes value j if the failure is due to cause j for a j in $\{1, \ldots, \mathcal{J}\}$.

We assume, in addition, that X is at risk of being independently right-censored by a non-negative r.v. C with distribution function G. Consequently, the observable r.v. are:

$$\big(T = \min(X, C),\ J = \mathcal{C}\, I(X \leq C)\big)\,.$$

As X and C are independent, the r.v. T has distribution function H given by $1 - H = (1 - F)(1 - G)$.

The interest focuses on the subdistribution functions which are given in (1.10) for $j = 1, \ldots, \mathcal{J}$ and $t \geq 0$ by:

$$F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j] \,.$$

As the different risks are mutually exclusive, the relation $F(t) = \sum_{j=1}^{\mathcal{J}} F^{(j)}(t)$ holds for $t \geq 0$.

The following subdistribution functions of H will be needed:

$$\begin{aligned} H^{(1)}(t) &= \mathbb{P}[T \leq t, J \neq 0] \,, \\ H^{(1,j)}(t) &= \mathbb{P}[T \leq t, J = j] \quad \text{for } j = 1, \ldots, \mathcal{J}. \end{aligned}$$

The relation $H^{(1)}(t) = \sum_{j=1}^{\mathcal{J}} H^{(1,j)}(t)$ is valid for $t \geq 0$.

The relations that connect the distribution functions $H^{(1,j)}$ to the subdistribution functions $F^{(j)}$ are given for $t \geq 0$ and $j = 1, \ldots, \mathcal{J}$ by:

$$H^{(1,j)}(t) = \int_0^t (1 - G^-) dF^{(j)} \,.$$

We introduce the cause-specific cumulative hazard function of X related to cause j for $t \geq 0$:

$$\Lambda^{(1,j)}(t) = \int_0^t \frac{dF^{(j)}}{1 - F^-} = \int_0^t \frac{dH^{(1,j)}}{1 - H^-} \,. \tag{1.15}$$

Let $(T_i, J_i)_{i=1,\ldots,n}$ be n independent copies of the random vector (T, J). Let $T_{1,n} \leq T_{2,n} \leq \cdots \leq T_{n,n}$ be the order statistics associated to the sample $T_1, \ldots, T_n$. It is assumed throughout that the functions $F^{(j)}$ have disjoint sets of discontinuity points. Consequently, failures from different causes can not be observed simultaneously with probability one. However, simultaneous failure times from the same cause can be observed with probability one. The simultaneous failure times are then ranked arbitrarily. The treatment of simultaneous censoring times is similar. If there are ties between a failure time (or several failures times from the same cause) and a censoring time, then the failure time(s) is (are) ranked ahead of the censoring time(s).

We define the empirical counterparts of $H^{(1)}$ and $H^{(1,j)}$ for $t \geq 0$ by:

$$\begin{aligned} H_n^{(1)}(t) &= \frac{1}{n} \sum_{i=1}^{n} I(T_i \leq t, J_i \neq 0) \,, \\ H_n^{(1,j)}(t) &= \frac{1}{n} \sum_{i=1}^{n} I(T_i \leq t, J_i = j) \,. \end{aligned}$$

As X is independently randomly right-censored by C, an estimator for F is the Kaplan-Meier estimator, defined in (1.2), which is now written from the (T_i, J_i) for $i = 1, \ldots, n$ as

$$\widehat{F}_n(t) = 1 - \prod_{i=1}^{n} \left(1 - \frac{I(T_i \leq t,\ J_i \neq 0)}{n(1 - H_n^-(T_i))}\right) . \tag{1.16}$$

The Nelson-Aalen estimators for the cause-specific cumulative hazards $\Lambda^{(1,j)}$ are defined for $t \geq 0$ by:

$$\Lambda_n^{(1,j)}(t) = \int_0^t \frac{dH_n^{(1,j)}}{1 - H_n^-} .$$

In the analysis of event history data, individual life histories are seen as independent sample paths of a stochastic process moving between states in a discrete state space. The states of the process correspond to various statuses for an individual, while transitions between the states correspond to occurrences of the events of interest. The origins go back to Aalen (1978a, 1978b) who reformulated the study of lifetime data in terms of stochastic processes, relying on the theory of time-continuous martingales and stochastic integrals. Aalen showed how the theory of multivariate counting processes provides a general framework in which both censored failure time data and censored observations from inhomogeneous Markov chains may be analysed. This methodology has gained great respect and wide application for its power in dealing with the theory in an unified way.

When the process of the states occupied by an individual in time is an inhomogeneous time continuous Markov process, Aalen and Johansen (1978) introduced an estimator of the transition probabilities between states in presence of independent random right-censoring. It has since been shown, see in particular Andersen et al. (1993), that the competing risks set-up can be considered as a special case of the Aalen-Johansen theory of estimation for inhomogeneous Markov processes. Indeed, the competing risks set-up corresponds to the case of one transcient state and several absorbing states (that can be labeled $1, \ldots, \mathcal{J}$).

The Aalen-Johansen estimator for $F^{(j)}$ is defined, for $t \geq 0$, by:

$$\widehat{F}_n^{(j)}(t) = \int_0^t \frac{1 - \widehat{F}_n^-}{1 - H_n^-} dH_n^{(1,j)} = \int_0^t (1 - \widehat{F}_n^-) d\Lambda_n^{(1,j)} ,$$

or equivalently, since $H_n^{(1,j)}$ may have jumps only when $J_n > 0$:

$$\widehat{F}_n^{(j)}(t) = \int_0^t \frac{J_n(1 - \widehat{F}_n^-)}{1 - H_n^-} dH_n^{(1,j)} = \int_0^t J_n(1 - \widehat{F}_n^-) d\Lambda_n^{(1,j)} .$$

The following relations are satisfied for $t \geq 0$:

$$\begin{aligned}
\Lambda_n^{(1,j)}(t) &= \int_0^t \frac{d\widehat{F}_n^{(j)}}{1-\widehat{F}_n^-}\,, \\
H_n^{(1)}(t) &= \sum_{j=1}^{\mathcal{J}} H_n^{(1,j)}(t)\,, \\
\widehat{F}_n(t) &= \sum_{j=1}^{\mathcal{J}} \widehat{F}_n^{(j)}(t)\,, \\
\Lambda_n(t) &= \sum_{j=1}^{\mathcal{J}} \Lambda_n^{(1,j)}(t)\,.
\end{aligned}$$

1.3.2 Martingale-based inference

Assuming the continuity of the underlying subdistribution functions $F^{(j)}$, Aalen and Johansen (1978) obtained the joint consistency of $\widehat{F}_n^{(j)}$ to $F^{(j)}$ for $j = 1, \ldots, \mathcal{J}$, uniformly over fixed compact intervals $[0, \sigma]$ for $\sigma < \tau_H$. They also obtained the joint weak convergence of the processes $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ for $j = 1, \ldots, \mathcal{J}$ on fixed compact intervals $[0, \sigma]$ for $\sigma < \tau_H$. For that, they made use of the properties of multivariate martingales and of the product-integral representation of the matrix of transition probabilities. Andersen et al. (1993) further developed these results.

In order to treat the structure of martingale of the Aalen-Johansen estimator, we introduce the $\mathcal{J}$-dimensional jump process $(N_n^{(1,1)}, \ldots, N_n^{(1,\mathcal{J})})$ where for $j = 1, \ldots, \mathcal{J}$ and $t \geq 0$,

$$N_n^{(1,j)}(t) = \sum_{i=1}^{n} I(T_i \leq t, J_i = j)\,,$$

counts the number of failures of cause j observed before t. Since we assumed that the subdistribution functions $F^{(j)}$ have disjoint sets of discontinuity points, then, with probability one, the different processes for $j = 1, \ldots, \mathcal{J}$ can not have simultaneous jumps. If, in addition, the different processes $N_n^{(1,j)}$ may not have jumps of size greater than one, then $(N_n^{(1,1)}, \ldots, N_n^{(1,\mathcal{J})})$ is a $\mathcal{J}$-dimensional counting process. This is in particular the case if the functions $F^{(j)}$ are continuous for $j = 1, \ldots, \mathcal{J}$.

The number of uncensored observations before t for $t \geq 0$, which is defined in (1.3), is now written from the (T_i, J_i) for $i = 1, \ldots, n$ as $N_n^{(1)}(t) = \sum_{i=1}^n I(T_i \leq t, J_i \neq 0)$. We have the relation $N_n^{(1)} = \sum_{j=1}^{\mathcal{J}} N_n^{(1,j)}$.

We introduce the σ-field generated by the events observed before t:

$$\mathcal{F}_n(t) = \sigma\big\{I(T_i \leq s),\, T_i I(T_i \leq s),\, J_i I(T_i \leq s) \,:\, i = 1, \ldots, n,\, 0 \leq s \leq t\big\}\,.$$

For fixed n, the family $\mathbb{F}_n = \{\mathcal{F}_n(t) \,:\, t \geq 0\}$ is a filtration.

Proposition 2. (Andersen et al. 1993) *Let* $\mathbf{M_n} = (M_n^{(1,1)}, \ldots, M_n^{(1,\mathcal{J})})$ *where* $M_n^{(1,j)}$ *is defined for* $t \geq 0$ *and* $j = 1, \ldots, \mathcal{J}$ *by:*

$$M_n^{(1,j)}(t) = N_n^{(1,j)}(t) - \int_0^t Y_n \, d\Lambda^{(1,j)} \,. \tag{1.17}$$

For fixed n, $\{\mathbf{M_n}(t), \mathcal{F}_n(t) \,:\, t \geq 0\}$ *is a square integrable* ($\mathcal{J}$*-dimensional) martingale in* t *on* $[0, \sigma]$ *for* $\sigma < \tau_H$ *with mean* 0 *and predictable covariation process given for* $k, j = 1, \ldots, \mathcal{J}$ *and* $t \geq 0$ *by:*

$$\left\langle M_n^{(1,j)}, M_n^{(1,k)} \right\rangle (t) = \, I(j = k) \int_0^t (1 - \Delta\Lambda) Y_n d\Lambda^{(1,j)} \,. \tag{1.18}$$

Note that the martingales $M_n^{(1,j)}$ and the martingale $M_n^{(1)}$ defined in (1.4) satisfy the relation:

$$M_n^{(1)} = \sum_{j=1}^{\mathcal{J}} M_n^{(1,j)} \,. \tag{1.19}$$

Note also that, for fixed n, $\{\mathbf{M_n}(t), \mathcal{F}_n(t) \,:\, t \geq 0\}$ is a local square integrable ($\mathcal{J}$-dimensional) martingale in t on $[0, \tau_H]$.

Proof. This proof is the analogue of the proof that Shorack and Wellner (1986) gave for the basic martingale associated to the Kaplan-Meier estimator. Firstly, the process $\{\mathbf{M_n}(t) : t \geq 0\}$ is adapted to the filtration $\mathbb{F}_n$ i.e. each $\mathbf{M_n}(t)$ is $\mathcal{F}_n(t)$-mesurable. Then, we carry out the following decomposition:

$$M_n^{(1,j)}(t) = \sum_{i=1}^{n} M_{n,i}^{(1,j)}(t)$$

with

$$M_{n,i}^{(1,j)}(t) = I(T_i \leq t, J_i = j) - \int_0^t I(T_i \geq u) \frac{dH^{(1,j)}(u)}{1 - H^-(u)}.$$

Let us prove that $\{M_{n,i}^{(1,j)}(t), \mathcal{F}_n(t) \,:\, t \geq 0\}$ is a martingale, for fixed n, j and i.

On the first hand, we have for $0 \leq s \leq t$:

$$\mathbb{E}\big[I(T_i \leq t, J_i = j) \big| \mathcal{F}_n(s)\big] = I(T_i \leq s, J_i = j) + \mathbb{E}\big[I(s < T_i \leq t, J_i = j) \big| \mathcal{F}_n(s)\big] \,,$$

with

$$\mathbb{E}\big[I(s < T_i \leq t, J_i = j) \big| \mathcal{F}_n(s)\big] = \frac{H^{(1,j)}(t) - H^{(1,j)}(s)}{1 - H(s)} \, I(T_i > s) \,.$$

On the other hand, applying Fubini's theorem, we get:

$$\begin{aligned}
&\mathbb{E}\left[\int_0^t I(T_i \geq u) \frac{dH^{(1,j)}}{1 - H^-(u)} \Big| \mathcal{F}_n(s)\right] \\
&= \int_0^s \mathbb{E}\big[I(T_i \geq u) \big| \mathcal{F}_n(s)\big] \frac{dH^{(1,j)}}{1 - H^-(u)} + \int_s^t \mathbb{E}\big[I(T_i \geq u) \big| \mathcal{F}_n(s)\big] \frac{dH^{(1,j)}}{1 - H^-(u)} \\
&= \int_0^s I(T_i \geq u) \frac{dH^{(1,j)}}{1 - H^-(u)} + \int_s^t \mathbb{E}\big[I(T_i \geq u) \big| \mathcal{F}_n(s)\big] \frac{dH^{(1,j)}}{1 - H^-(u)} \,,
\end{aligned}$$

with for $u > s$:

$$\mathbb{E}\big[I(T_i \geq u)\big|\mathcal{F}_n(s)\big] = \frac{1-H^-(u)}{1-H(s)} I(T_i > s)\,.$$

Finally, the equality

$$\begin{aligned}\mathbb{E}\big[M_{n,i}^{(1,j)}(t)\big|\mathcal{F}_n(s)\big] &= I(T_i \leq s, J_i = j) - \int_0^s I(T_i \geq u)\frac{dH^{(1,j)}}{1-H^-(u)} \\ &= M_{n,i}^{(1,j)}(s)\end{aligned}$$

shows that $\{M_{n,i}^{(1,j)}(t), \mathcal{F}_n(t) \,:\, t \geq 0\}$ is a martingale. It is straightforward to see that $M_{n,i}^{(1,j)}(t)$ has mean 0. To see that $M_{n,i}^{(1,j)}(t)$ is square integrable, we write for $t \geq 0$:

$$\mathbb{E}\left[M_{n,i}^{(1,j)}(t)^2\right] = Var\left(M_{n,i}^{(1,j)}(t)\right) = \int_0^t \left(1-\Delta\Lambda^{(1,j)}\right) dH^{(1,j)} < \infty\,.$$

Consequently, $\{M_{n,i}^{(1,j)}(t), \mathcal{F}_n(t) \,:\, t \geq 0\}$ is a square integrable martingale. It is then licit to determine its predictable variation process denoted by $\left\langle M_{n,i}^{(1,j)} \right\rangle$. Since

$$N_{n,i}^{(1,j)} = I(T_i \leq ., J_i = j)$$

is a counting process and since

$$A_{n,i}^{(1,j)} = \int_0^{\cdot} I(T_i \geq u) d\Lambda^{(1,j)}(u)$$

is its compensator, we can apply Theorem 2.3.1. of Gill (1980) (which is quoted in the Appendix) to

$$M_{n,i}^{(1,j)} = N_{n,i}^{(1,j)} - A_{n,i}^{(1,j)}\,.$$

We get for $t \geq 0$ and $k, j = 1, \ldots, \mathcal{J}$ such that $k \neq j$:

$$\begin{aligned}\left\langle M_{n,i}^{(1,j)}, M_{n,i}^{(1,k)} \right\rangle(t) &= -\int_0^t \Delta\Lambda^{(1,j)}(u) I(T_i \geq u) d\Lambda^{(1,k)}(u) \\ &= -\sum_{u \leq t} \Delta\Lambda^{(1,j)}(u) I(T_i \geq u) \Delta\Lambda^{(1,k)}(u) \\ &= 0\,,\end{aligned}$$

and

$$\begin{aligned}\left\langle M_{n,i}^{(1,j)} \right\rangle(t) &= \int_0^t \left(1-\Delta A_{n,i}^{(1,j)}\right) dA_{n,i}^{(1,j)} \\ &= \int_0^t \left(1-\Delta\Lambda^{(1,j)}(u)\right) I(T_i \geq u) d\Lambda^{(1,j)}(u) \\ &= \int_0^t \left(1-\Delta\Lambda(u)\right) I(T_i \geq u) d\Lambda^{(1,j)}(u)\,,\end{aligned}$$

since it is assumed that the different $F^{(j)}$ do not have common jumps.

For fixed n and j, $\{M_n^{(1,j)}(t), \mathcal{F}_n(t) \ : \ t \geq 0\}$ is a finite sum of square integrable martingales with mean 0, also is itself a square integrable martingale with mean 0. It is then licit to determine its predictable variation process. However, as the continuity of $F^{(j)}$ is not assumed, $N_n^{(1,j)}$ may not be a counting process and Theorem 2.3.1. of Gill (1980) can not be used directly. Nevertheless, one is able to calculate:

$$\begin{aligned}
& \mathbb{E}\left[M_n^{(1,j)}(t)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (t) \Big| \mathcal{F}_n(s) \right] \\
= \ & \mathbb{E}\left[\left(\sum_{i=1}^n M_{n,i}^{(1,j)}(t) \right)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (t) \Big| \mathcal{F}_n(s) \right] \\
= \ & \mathbb{E}\left[\sum_{i=1}^n M_{n,i}^{(1,j)}(t)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (t) - \sum_{i=1}^n \sum_{\substack{l=1 \\ i \neq l}}^n M_{n,i}^{(1,j)}(t) M_{n,l}^{(1,j)}(t) \Big| \mathcal{F}_n(s) \right] .
\end{aligned}$$

By definition of the predictable variation process,

$$\sum_{i=1}^n M_{n,i}^{(1,j)}(t)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (t)$$

is a martingale (in t). Using now the independence of the observations in the double sum, we get:

$$\begin{aligned}
& \mathbb{E}\left[M_n^{(1,j)}(t)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (t) \Big| \mathcal{F}_n(s) \right] \\
& = \sum_{i=1}^n M_{n,i}^{(1,j)}(s)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (s) - \sum_{i=1}^n \sum_{\substack{l=1 \\ i \neq l}}^n \mathbb{E}\left[M_{n,i}^{(1,j)}(t) \Big| \mathcal{F}_n(s) \right] . \mathbb{E}\left[M_{n,l}^{(1,j)}(t) \Big| \mathcal{F}_n(s) \right] \\
& = \sum_{i=1}^n M_{n,i}^{(1,j)}(s)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (s) - \sum_{i=1}^n \sum_{\substack{l=1 \\ i \neq l}}^n M_{n,i}^{(1,j)}(s) M_{n,l}^{(1,j)}(s) \\
& = \left(\sum_{i=1}^n M_{n,i}^{(1,j)}(s) \right)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (s) \\
& = M_n^{(1,j)}(s)^2 - \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (s) .
\end{aligned}$$

This means that the predictable variation process of $M_n^{(1,j)}$ is given for $t \geq 0$ by:

$$\left\langle M_n^{(1,j)} \right\rangle (t) = \sum_{i=1}^n \left\langle M_{n,i}^{(1,j)} \right\rangle (t) = \int_0^t (1 - \Delta\Lambda) Y_n d\Lambda^{(1,j)} .$$

The predictable covariation process $\left\langle M_n^{(1,j)}, M_n^{(1,k)} \right\rangle$ for $k \neq j$ is obtained in the same way. □

Theorem 8. (Aalen and Johansen 1978; Andersen et al. 1993) *Let $\sigma < \tau_H$. Assume that the subdistribution functions $F^{(j)}$ are continuous for $j = 1, \ldots, \mathcal{J}$. As $n \to \infty$, we have:*

$$\sup_{t \in [0,\sigma]} \left| \widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right| \xrightarrow{\mathbb{P}} 0 \,.$$

Theorem 9. (Aalen and Johansen 1978; Andersen et al. 1993) *Define*

$$Z_n^{(0)} = \sqrt{n}\left((1 - \widehat{F}_n) - (1 - F)\right) \quad \text{and} \quad Z_n^{(j)} = \sqrt{n}\left(\widehat{F}_n^{(j)} - F^{(j)}\right) \quad \text{for } j = 1, \ldots, \mathcal{J} \,.$$

Let $\sigma < \tau_H$. If the subdistribution functions $F^{(j)}$ are continuous for $j = 1, \ldots, \mathcal{J}$, the following convergence holds jointly, as $n \to \infty$, in $D^{\mathcal{J}+1}[0, \sigma]$, the space of càdlàg functions from $[0, \sigma]$ to $\mathbb{R}^{\mathcal{J}+1}$:

$$\left(Z_n^{(0)}, Z_n^{(1)}, \ldots, Z_n^{(\mathcal{J})}\right) \xrightarrow{\mathcal{D}} \mathbf{Z} = \left(Z^{(0)}, Z^{(1)}, \ldots, Z^{(\mathcal{J})}\right) \,,$$

where the process $\mathbf{Z}$ is Gaussian with mean 0. It is given for $t \geq 0$ by

$$\begin{aligned} Z^{(0)}(t) &= (1 - F(t))U^{(0)}(t) \\ Z^{(j)}(t) &= \int_0^t \left(F^{(j)}(t) - F^{(j)}\right) dU^{(0)} + \int_0^t (1 - F) dU^{(j)} \quad \text{for } j = 1, \ldots, \mathcal{J}. \end{aligned}$$

The processes $U^{(j)}$ for $j = 1, \ldots, \mathcal{J}$ are local square integrable martingales, they are null at time 0, orthogonal, Gaussian and have mean 0. They satisfy the relation $U^{(0)} = -\sum_{j=1}^{\mathcal{J}} U^{(j)}$. The covariance structure of the $U^{(j)}$ is given for $j = 1, \ldots, \mathcal{J}$ and $s, t \geq 0$ by:

$$Cov\left(U^{(j)}(t), U^{(j)}(s)\right) = \int_0^{s \wedge t} \frac{dH^{(1,j)}}{(1 - H^-)^2}.$$

1.4 Some parametric models

1.4.1 Univariate exponential distribution

In the analysis of random durations, the exponential distribution, which depends on a positive parameter λ, is a widely used model. This distribution will be denoted by $\mathcal{E}\mathrm{xp}(\lambda)$. The distribution function is given for $y \geq 0$ by:

$$F(y) = 1 - \exp(-\lambda y).$$

The corresponding probability density is given for $y \geq 0$ by:

$$f(y) = \lambda \exp(-\lambda y).$$

The expected value of a r.v. Y that has a $\mathcal{E}\text{xp}(\lambda)$ distribution is $\mathbb{E}[Y] = 1/\lambda$. An important property of the exponential distribution is the lack of memory property, that is, for $y, t \geq 0$

$$\mathbb{P}\left[Y > y + t \middle| Y > y\right] = \mathbb{P}\left[Y > t\right] .$$

This means that the distribution of the additional survival time is not affected by the knowledge that an individual has survived for a certain length of time.

1.4.2 Bivariate exponential distribution

There is an extensive literature on bivariate exponential models. There has been several formulations of bivariate exponential and related distributions, see e.g. Gumbel (1960), Marshall and Olkin (1967), Block and Basu (1974) and Ryu (1993).

Block and Basu (1974) introduced the absolutely continuous bivariate exponential model (ACBVE). A bivariate r.v. (Y_1, Y_2) is said to follow an ACBVE(λ_0,λ_1,λ_2) if, for $y_1, y_2 > 0$,

$$\begin{aligned}\mathbb{P}\left[Y_1 > y_1,\, Y_2 > y_2\right] = {} & \frac{\lambda}{\lambda_1 + \lambda_2} \exp\left(-\lambda_1 y_1 - \lambda_2 y_2 - \lambda_0 \max(y_1, y_2)\right) \\ & - \frac{\lambda_0}{\lambda_1 + \lambda_2} \exp\left(-\lambda \max(y_1, y_2)\right) ,\end{aligned}$$

where $\lambda = \lambda_0 + \lambda_1 + \lambda_2$. The corresponding density function is given by:

$$f(y_1, y_2) = \begin{cases} \frac{\lambda_1 \lambda(\lambda_2 + \lambda_0)}{\lambda_1 + \lambda_2} \exp\left(-\lambda_1 y_1 - (\lambda_2 + \lambda_0) y_2\right) & \text{if } y_1 \leq y_2, \\ \frac{\lambda_2 \lambda(\lambda_1 + \lambda_0)}{\lambda_1 + \lambda_2} \exp\left(-(\lambda_1 + \lambda_0) y_1 - \lambda_2 y_2\right) & \text{if } y_1 > y_2. \end{cases}$$

The marginal distribution functions are given for $y \geq 0$:

$$\mathbb{P}\left[Y_1 \leq y\right] = \frac{\lambda}{\lambda_1 + \lambda_2} \exp\left(-(\lambda_1 + \lambda_0) y\right) - \frac{\lambda_0}{\lambda_1 + \lambda_2} \exp(-\lambda y),$$

and

$$\mathbb{P}\left[Y_2 \leq y\right] = \frac{\lambda}{\lambda_1 + \lambda_2} \exp\left(-(\lambda_2 + \lambda_0) y\right) - \frac{\lambda_0}{\lambda_1 + \lambda_2} \exp(-\lambda y).$$

It is seen that Y_1 and Y_2 are independent if and only if $\lambda_0 = 0$, in which case the marginal distributions also are exponential.

Block and Basu (1974) showed that the bivariate r.v. (Y_1, Y_2) fulfills the bivariate loss of memory property, that is, satisfies, for $y_1, y_2, t \geq 0$, the following distributional equality:

$$\mathbb{P}\left[Y_1 > y_1 + t, Y_2 > y_2 + t \middle| Y_1 > y_1, Y_2 > y_2\right] = \mathbb{P}\left[Y_1 > t, Y_2 > t\right] .$$

Block and Basu (1974) gave a characterization of the ACBVE model in terms of univariate distributions that makes it practical for simulation purposes. This characterization is given in the next proposition.

Proposition 3. (Block and Basu 1974) *Let* (Y_1, Y_2) *be ACBVE(*λ_0*,*λ_1*,*λ_2*), then*

1. $\min(Y_1, Y_2)$ *is* $\mathcal{E}xp(\lambda)$ *distributed where* $\lambda = \lambda_1 + \lambda_2 + \lambda_0$
2. $Y_1 - Y_2$ *has distribution function:*

$$\varphi(z) = \begin{cases} \frac{\lambda_1}{\lambda_1+\lambda_2} \exp\left((\lambda_2 + \lambda_0)\, z\right) & \text{if } z \leq 0, \\ 1 - \frac{\lambda_2}{\lambda_1+\lambda_2} \exp\left((\lambda_1 + \lambda_0)\, z\right) & \text{if } z > 0. \end{cases} \tag{1.20}$$

3. $\min(Y_1, Y_2)$ *is independent of* $Y_1 - Y_2$.

It turns out that the ACBVE model has proportional cause-specific distribution functions. Namely, let us define

$$Y = \min(Y_1, Y_2) \quad \text{and} \quad \mathcal{C} = \begin{cases} 1 & \text{if } Y_1 \leq Y_2, \\ 2 & \text{otherwise,} \end{cases}$$

and set for $y \geq 0$:

$$F^{(1)}(y) = \mathbb{P}\left[Y \leq y, \mathcal{C} = 1\right] = \mathbb{P}\left[Y_1 \leq y, Y_1 \leq Y_2\right],$$

and

$$F^{(2)}(y) = \mathbb{P}\left[Y \leq y, \mathcal{C} = 2\right] = \mathbb{P}\left[Y_2 \leq y, Y_2 \leq Y_1\right].$$

Then, for $y \geq 0$, we have:

$$F^{(1)}(y) = \frac{\lambda_1}{\lambda_1 + \lambda_2}(1 - \exp(-\lambda y)),$$

and

$$F^{(2)}(y) = \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - \exp(-\lambda y)).$$

1.4.3 Univariate exponential Weibull distribution

In practice, the assumption of exponentially distributed lifetimes, or equivalently of a constant hazard function, is rarely tenable. A more general distribution widely used in reliability and life testing experiments is the Weibull distribution that depends on two

parameters λ and γ which are both greater then zero. This distribution will be denoted by $W(\alpha, \lambda)$. The distribution function is given for $x \geq 0$ by:

$$F(x) = 1 - \exp(-\lambda x^{\alpha}).$$

The corresponding probability density is given for $x \geq 0$ by:

$$f(x) = \lambda \alpha x^{\alpha-1} \exp(-\lambda x^{\alpha}).$$

The shape of the hazard function depends critically on the value of the shape parameter α, while λ is the scale parameter.

In the particular case $\alpha = 1$, the Weibull distribution reduces to exponential distribution with parameter λ and with constant hazard function. The Weibull distribution has increasing hazard function when the shape parameter α is greater than one and has decreasing hazard function when the shape parameter is less than one.

The expected value of a r.v. X which is drawn from a $W(\alpha, \lambda)$ distribution is $\mathbb{E}[X] = \lambda^{-1/\alpha}\Gamma(\alpha^{-1} + 1)$ where $\Gamma(x)$ is the gamma function defined by the integral $\Gamma(x) = \int_0^{\infty} u^{x-1}e^{-u}du$. For integer values of x, the value of this integral is $(x-1)!$ and so it can easily be computed. Otherwise, suitable computer software are needed to compute the mean.

The exponential and the Weibull distributions are related by a power transformation. Namely, Y is $\mathcal{E}\mathrm{xp}(\lambda)$-distributed if and only if $X = Y^{1/\alpha}$ is $W(\alpha, \lambda)$-distributed.

Chapter 2

Strong approximations and LIL for processes based on the Aalen-Johansen estimator for the subdistribution functions

2.1 Introduction

The aim of this chapter is to extend to the Aalen-Johansen estimators the laws of the iterated logarithm (LIL) and the strong approximations that Stute (1994a), Csörgő (1996) and Giné and Guillou (1999) obtained for the Kaplan-Meier estimator. With this aim in view, Section 2.2 deals with some asymptotic properties of these estimators. Namely, in Subsection 2.2.1, we get three strong approximation results which are valid up to a certain order statistic. First, we obtain a strong approximation result for the empirical processes $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ by means of Gaussian processes, jointly for $j = 1, \ldots, \mathcal{J}$. Then, we get a similar result for a transformation of these empirical processes leading to Brownian motions as approximating processes. We also derive a strong approximation result for a normalization of the aforementioned transformation. In Subsection 2.2.2, we obtain two laws of the iterated logarithm which are valid up to a certain order statistic. The first LIL-type result requires light assumptions on this order statistic, while the second LIL refines the first one but requires stronger assumptions on the order statistic. We do not require any condition on G and the $F^{(j)}$ as on their tails. For the sake of clarity, the proofs of the forthcoming results are postponed until Section 2.3.

The notations used in this chapter are defined in Chapter 1.

2.2 Main results

Following Giné and Guillou (2001), we say that a nonincreasing sequence (a_n) of numbers is regular if there exists a constant $M > 0$ such that for all n, $a_{2n} \geq M a_n$.

Throughout, we shall denote by hypothesis $(\mathcal{H})$ the following condition:

($\mathcal{H}$): *for n large enough, the sequence (k_n/n) is regular nonincreasing and there exists a constant $C > 0$ such that $k_n \geq C d_n \log n$ with (d_n) is a non-increasing sequence such that*

$$\sum \frac{1}{k d_{2^k} \log k} < \infty$$

(e.g. $d_n = (\log\log\log n)^{1+\varepsilon}$, $d_n = (\log\log\log n)(\log\log\log\log n)^{1+\varepsilon}$, etc).

If Hypothesis ($\mathcal{H}$) is fulfilled, Hypothesis ($\mathcal{H}1$) and ($\mathcal{H}2$) quoted in Chapter 1 are also fulfilled.

2.2.1 Strong approximations

We first provide an approximation of the empirical processes $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ jointly for $j = 1, \ldots, \mathcal{J}$ and uniformly over the random increasing intervals $[0, T_{n-k_n,n}]$.

Theorem 10. *Let (k_n) be a sequence of integers fulfilling $1 \leq k_n < n$ and Condition ($\mathcal{H}$). We assume that $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$. On a suitably enlarged probability space, one can define $\mathcal{J}$ sequences of Gaussian processes $(\bar{L}_n^{(1)}), \ldots, (\bar{L}_n^{(\mathcal{J})})$ such that the following approximations hold jointly for $j = 1, \ldots, \mathcal{J}$.*

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n}(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)) - \bar{L}_n^{(j)}(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_n} \right).$$

The Gaussian processes have mean 0 and satisfy for $k \neq j$ with $k, j = 1, \ldots, \mathcal{J}$ and for $s, t \geq 0$:

$$\begin{aligned}
Cov\Big(\bar{L}_n^{(j)}(t), \bar{L}_n^{(j)}(s)\Big) &= \int_0^{s\wedge t} \left(F^{(j)}(t) - F^{(j)}\right)\left(F^{(j)}(s) - F^{(j)}\right) \frac{dH^{(1)}}{(1-H^-)^2} \\
&\quad - \int_0^{s\wedge t} (1-F)\left(F^{(j)}(t) + F^{(j)}(s) - 2F^{(j)}\right) \frac{dH^{(1,j)}}{(1-H^-)^2} \\
&\quad + \int_0^{s\wedge t} (1-F)^2 \frac{dH^{(1,j)}}{(1-H^-)^2}\,, \\
Cov\Big(\bar{L}_n^{(j)}(t), \bar{L}_n^{(k)}(s)\Big) &= \int_0^{s\wedge t} \left(F^{(j)}(t) - F^{(j)}\right)\left(F^{(k)}(s) - F^{(k)}\right) \frac{dH^{(1)}}{(1-H^-)^2} \\
&\quad - \int_0^{s\wedge t} (1-F)\left(F^{(j)}(t) - F^{(j)}\right) \frac{dH^{(1,k)}}{(1-H^-)^2} \\
&\quad - \int_0^{s\wedge t} (1-F)\left(F^{(k)}(s) - F^{(j)}\right) \frac{dH^{(1,j)}}{(1-H^-)^2}.
\end{aligned}$$

We notice that the approximation of the processes $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ by Gaussian processes holds uniformly over $[0, T_{n-k_n,n}]$ with the same rate than the corresponding approximations of $\sqrt{n}(\Lambda_n - \Lambda)$ and $\sqrt{n}((\widehat{F}_n - F)/(1-F))$, see e.g. Csörgő (1996). But here, unlike the latter approximations, the Gaussian approximating processes are not Brownian motions. We could prefer a Brownian approximation since Brownian motions can

be easily simulated or tabulated and their properties have been carefully elaborated. For that, we carry out the following transformation of the empirical process $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ for $j = 1, \ldots, \mathcal{J}$:

$$K_n^{(j)}(t) = \sqrt{n}\left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\frac{1-\widehat{F}_n(t)}{1-F(t)}\right). \tag{2.1}$$

Theorem 11. *Let (k_n) be a sequence of integers fulfilling $1 \leq k_n < n$ and Condition $(\mathcal{H})$. We assume that $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$. On a suitably enlarged probability space, one can define $\mathcal{J}$ sequences of Gaussian processes $(\check{L}_n^{(1)}), \ldots, (\check{L}_n^{(\mathcal{J})})$ such that the following approximations hold jointly for $j = 1, \ldots, \mathcal{J}$:*

$$\sup_{t \leq T_{n-k_n,n}} \left|K_n^{(j)}(t) - \check{L}_n^{(j)}(t)\right| = O\left(\sqrt{n}\frac{\log n}{k_n}\right).$$

The Gaussian processes have mean 0 and satisfy for $k, j = 1, \ldots, \mathcal{J}$ such that $k \neq j$ and for $s, t \geq 0$:

$$\begin{aligned}
Cov\Big(\check{L}_n^{(j)}(t), \check{L}_n^{(j)}(s)\Big) &= \int_0^{s\wedge t} F^{(j)2}\frac{dH^{(1)}}{(1-H^-)^2} + 2\int_0^{s\wedge t}(1-F)F^{(j)}\frac{dH^{(1,j)}}{(1-H^-)^2} \\
&\quad + \int_0^{s\wedge t}(1-F)^2\frac{dH^{(1,j)}}{(1-H^-)^2}, \\
Cov\Big(\check{L}_n^{(j)}(t), \check{L}_n^{(k)}(s)\Big) &= \int_0^{s\wedge t} F^{(j)}F^{(k)}\frac{dH^{(1)}}{(1-H^-)^2} + \int_0^{s\wedge t}(1-F)F^{(j)}\frac{dH^{(1,k)}}{(1-H^-)^2} \\
&\quad - \int_0^{s\wedge t}(1-F)F^{(k)}\frac{dH^{(1,j)}}{(1-H^-)^2}.
\end{aligned}$$

This entails that the processes $\check{L}_n^{(j)}$ satisfy the distributional equality $\check{L}_n^{(j)} \overset{\mathcal{D}}{=} W_n^{(j)}(C^{(j)})$, for $j = 1, \ldots, \mathcal{J}$, where the $(W_n^{(j)})_n$ are correlated sequences of Brownian motions and where $C^{(j)}(t) = Var(\check{L}_n^{(j)}(t))$ for $t \geq 0$.

We are then interested in providing a normalized result. For that, we introduce estimators for the variances of the components of the approximating processes $\check{L}_n^{(j)}$. We propose the following estimator for $C^{(j)}(t)$:

$$\widehat{C}_n^{(j)}(t) = \int_0^t \frac{\widehat{F}_n^{(j)2}}{(1-H_n^-)^2}dH_n^{(1)} + 2\int_0^t \widehat{F}_n^{(j)}\frac{1-\widehat{F}_n}{(1-H_n^-)^2}dH_n^{(1,j)} + \int_0^t \frac{(1-\widehat{F}_n)^2}{(1-H_n^-)^2}dH_n^{(1,j)}. \tag{2.2}$$

It will be seen, in Lemma 9, that these estimators achieve the same rate of convergence to $C^{(j)}$ than the estimators of the variance of the approximating process of $\sqrt{n}(\Lambda_n - \Lambda)$ and $\sqrt{n}((\widehat{F}_n - F)/(1-F))$ respectively.

Theorem 12. *Let (k_n) be a sequence of integers fulfilling $1 \leq k_n < n$ and Hypothesis $(\mathcal{H})$. For $j = 1, \ldots, \mathcal{J}$, we assume that $F^{(j)}$ is continuous. On a suitably enlarged probability space, one can define $\mathcal{J}$ sequences of Gaussian processes such that the following approximations hold jointly for $j = 1, \ldots, \mathcal{J}$.*

$$\sup_{t \leq T_{n-k_n,n}} \left| \frac{K_n^{(j)}(t)}{1 + \widehat{C}_n^{(j)}(t)} - \frac{\check{L}_n^{\dagger(j)}(t)}{1 + C^{(j)}(t)} \right| = O\left(\sqrt{n} \frac{(\log n)^2}{k_n} + \frac{n \log\log n}{k_n^{3/2}} \right).$$

For $j = 1, \ldots, \mathcal{J}$, the Gaussian processes $\check{L}_n^{\dagger(j)}/(1+C^{(j)})$ satisfy the following distributional equality $\check{L}_n^{\dagger(j)}/(1 + C^{(j)}) \stackrel{\mathcal{D}}{=} K\left(C^{(j)}/(1 + C^{(j)}), n\right)/\sqrt{n}$ where $K(.,.)$ is a Kiefer sheet.

2.2.2 LIL-type results

Our first LIL-type result provides almost sure and in probability rates of convergence of $\widehat{F}_n^{(j)}$ to $F^{(j)}$, for $j = 1, \ldots, \mathcal{J}$, uniformly over the random increasing intervals $[0, T_{n-k_n,n}]$.

Theorem 13. *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ and, for the almost-sure results, such that Hypothesis $(\mathcal{H})$ is satisfied. For $j = 1, \ldots, \mathcal{J}$, $F^{(j)}$ is assumed to be continuous. Then,*

$$\sup_{t \leq T_{n-k_n,n}} \left| \widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right| = \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_n}} \right), \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}} \right). \end{cases}$$

One application for the strong approximation results of Theorems 11 and 12 is the achievement of the following LIL type result which refines the LIL type result of Theorem 13. For that, (k_n) is chosen large enough to force the approximation rates divided by $\sqrt{\log\log n}$ to go to zero.

Theorem 14. 1. *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$, $k_n = o(n)$, $k_n \geq \log n$ for n large enough and $\sqrt{n} \log n/(\sqrt{\log\log n} k_n) \stackrel{n\to\infty}{\longrightarrow} 0$. If $F^{(j)}$ is continuous and $C^{(j)}(\tau_H) < \infty$ for $j = 1, \ldots, \mathcal{J}$, we have:*

$$\limsup_n \sup_{t \leq T_{n-k_n,n}} \frac{1}{\sqrt{2 \log\log n}} \left| \sqrt{n} \left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right) \right| \leq \sqrt{C^{(j)}(\tau_H)} + \sqrt{C(\tau_H)},$$

where C given for $t \geq 0$ by $C(t) = \int_0^t dH^{(1)}/(1 - H^-)^2$ is the variance of the Nelson-Aalen estimator.

2. *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$, $k_n = o(n)$, for n large enough $k_n \geq \log n$, $n/(k_n^{3/2}) \stackrel{n\to\infty}{\longrightarrow} 0$. If $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$, we have:*

$$\limsup_n \sup_{t \leq T_{n-k_n,n}} \frac{1}{\sqrt{2 \log\log n}} \left| \frac{\sqrt{n} \left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t) \right)}{1 - \widehat{C}_n(t)} \right| \leq 2 \sup_{t \leq \tau_H} \frac{\sqrt{C^{(j)}(t)}}{1 + C^{(j)}(t)} + \sup_{t \leq \tau_H} \frac{\sqrt{C(t)}}{1 + C(t)},$$

where $\widehat{C}_n$ *given for* $t \geq 0$ *by* $\widehat{C}_n(t) = \int_0^t dH_n^{(1)}/(1-H_n^-)^2$ *is an estimator for the variance of the Nelson-Aalen estimator.*

2.3 Proofs

Strong approximation results have received much attention since the papers of Burke et al. (1981), Csörgő and Horváth (1982) and Csörgő and Horváth (1983a). In these papers, the authors carried out a decomposition of certain empirically transformed processes in a general scheme including in particular the independent competing risks model and the random right-censorship model. The underlying distribution functions are there assumed continuous. Horváth (1980) dropped these continuity conditions together with the independence of the risks in the competing risks submodel (under some necessary identifiability conditions). The decomposition in Burke et al. (1981), which is the basis for what would come later, involves leading terms plus remainder terms. In this decomposition, the leading terms contribute asymptotically to the approximating process while the remainder terms go to zero. One term among the remainder terms is particularly tedious to handle and may be called the Nasty Remainder Term (NRT). In Burke et al. (1981), the suboptimal rate at which the NRT goes to zero dominated the "real" approximation rate corresponding to the asymptotically contributing terms. Hence the overall rate was the suboptimal NRT rate. Horváth and Csörgő later noticed that they made a correctable error in the handling of the NRT. At the same time, they noticed an underlying U-statistic structure of the NRT. Using the corresponding technique, Burke et al. (1988) reduced substantially the rate of convergence of the NRT, albeit it was still dominant. The next step was made by Stute (1994a), who used a reverse martingale technique for the U-statistic in the NRT. The improvement was large but the NRT was still dominant. Then, Csörgő (1996) refined the results of Stute (1994a) so that the rate of NRT was pushed down under the "real" approximation rate and the latter was finally dominant. As statistical applications of these strong approximations, Csörgő (1996) also got LIL-type results in the random right-censorship model. It turns out that the LIL results could be refined and that is the work of Giné and Guillou (1999).

The next section gathers a serie of useful results coming from these works.

2.3.1 Preliminary results

Lemma 1. (Stute 1994a) *Let* L *be an arbitrary distribution function. We define the left-continuous generalized inverse of* L *by* $L^{-1}(u) = \inf\{s : L(s) \geq u\}$ *for* $0 < u < 1$. *For* $0 < u < 1$, *for* $k > 1$, *we have:*

$$\int_0^{L^{-1}(u)} \frac{dL}{(1-L^-)^k} \leq \frac{k}{(k-1)(1-u)^{k-1}}\,.$$

We recall that the following inequalities are always valid $L^-(L^{-1}(u)) \leq u \leq L(L^{-1}(u))$ for $0 < u < 1$ and $L^{-1}(L(u)) \leq t \leq L^{-1+}(L(u))$ for $-\infty < t < \infty$ where L^{-1+} denote the

right-continuous function obtained from L^{-1}.

Lemma 2. (Csörgő 1996; Wellner 1978; Csáki 1977) *Let (k_n) be a sequence of integers satisfying $1 \leq k_n < n$ and Hypothesis ($\mathcal{H}2$). For n large enough, the following inequalities hold almost surely.*
1.

$$T_{n-k_n,n} \leq H^{-1}\left(1-\frac{k_n}{7n}\right) \quad \text{and} \quad \frac{1}{1-H^-(T_{n-k_n,n})} \leq \frac{7n}{k_n}\,.$$

2.

$$\sup_{t\leq H^{-1}(1-\frac{k_n}{7n})} \frac{1-H_n^-(t)}{1-H^-(t)} = O(1)$$

and

$$\sup_{t\leq H^{-1}(1-\frac{k_n}{7n})} \frac{|H_n^-(t)-H^-(t)|}{(1-H^-(t))^{1/2}} = O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right)\,.$$

Theorem 15. *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ and, for the almost-sure results, such that ($\mathcal{H}1$) is satisfied. The rate of convergence of $\Lambda_n^{(1,j)}$ to $\Lambda^{(1,j)}$ is given by*

$$\sup_{t\leq T_{n-k_n,n}} \left|\Lambda_n^{(1,j)}(t)-\Lambda^{(1,j)}(t)\right| \;=\; \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_n}}\right)\,, \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\,. \end{cases}$$

The proof of this result follows straightforwardly from the proof of the first part of Theorem 6 concerning the supremum of $\Lambda_n - \Lambda$.

The developments to come require some elements of empirical processes theory. For the definition of a VC-class of functions, we refer to the Appendix. Further materials on VC-classes may be found e.g. in Van der Vaart and Wellner (1996).

As in Giné and Guillou (1999), the VC-classes we will use have a very simple structure and will all be obviously measurable. To show that they are VC, it will suffice to apply the next lemma.

Lemma 3. (Giné and Guillou 1999) 1. *If $\mathcal{E} = \{f_x : x \in \mathcal{J}\}$ where $\mathcal{J}$ is a subset of $\mathbb{R}$ satisfying $0 \leq f_x(s) \leq f_y(s)$ for $s \in S$ and for $x, y \in \mathcal{J}$ such that $x < y$, then $\mathcal{E}$ is a VC-class for $E = \sup\{|f| : f \in \mathcal{E}\}$.*
2. *If $\mathcal{E}$ is a measurable uniformly bounded VC-class and if $(Z_i)_{i\geq 1}$ is a sequence of i.i.d. random variables, then $\widetilde{\mathcal{F}} = \{f - \mathbb{E}[f(Z_1)] : f \in \mathcal{F}\}$ is a VC-class.*

Lemma 4. (Montgomery-Smith 1993) *Let $(S, \mathcal{S})$ be a measurable space. Let $Z_1, \ldots, Z_n$ be some i.i.d. S-valued random variables. Let $\mathcal{F}$ be a (countable) class of measurable real functions on S. Then, for $n \geq 1$, for $x > 0$,*

$$\mathbb{P}\left[\max_{k=1,\ldots,n}\left\|\sum_{i=1}^k f(Z_i)\right\|_{\mathcal{F}} > x\right] \leq 9\,\mathbb{P}\left[\left\|\sum_{i=1}^n f(Z_i)\right\|_{\mathcal{F}} > \frac{x}{30}\right]\,.$$

The following result which is used in Giné and Guillou (2001) is based on Talagrand (1996) exponential bound together with a result of Arconès and Giné (1995) for the expectation inequality.

Theorem 16. (Talagrand 1996) *Let $(S, \mathcal{S})$ be a measurable space and let $Z_1, \ldots, Z_n$ be some i.i.d. S-valued random variables. Let $\mathcal{F}$ be a (countable) class of measurable real functions on S. We choose $U \geq 2\sup_{f\in\mathcal{F}} \|f\|_\infty$ and $V_n \geq 4n\mathbb{E}\left[\sup_{f\in\mathcal{F}} f(Z_1)^2\right]$. Then, for $x > 0$, for some universal constant K, the following majorization is valid:*

$$\mathbb{P}\left[\left\|\sum_{i=1}^n (f(Z_i) - \mathbb{E}[Z_1])\right\|_{\mathcal{F}} > x + \mathbb{E}\left[\left\|\sum_{i=1}^n (f(Z_i) - \mathbb{E}[Z_1])\right\|_{\mathcal{F}}\right]\right] \leq K\exp\left(-\frac{1}{K}\frac{x}{U}\log\left(1 + \frac{xU}{V_n}\right)\right),$$

where we have, for $n \geq 1$, for a constant $C > 0$, provided that $\mathcal{F}$ is VC with respect to an envelop E such that $\mathbb{E}[E^2] < \infty$:

$$\mathbb{E}\left[\left\|\sum_{i=1}^n (f(Z_i) - \mathbb{E}[Z_1])\right\|_{\mathcal{F}}\right] \leq C\left(n\,\mathbb{E}\left[E^2\right]\right)^{1/2}.$$

At last, we will need the following result of Stute (1994b) the proof of which relies on U-statistics theory.

Proposition 4. (Stute 1994b) *Assume that $\{X_i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(r)}) : 1 \leq i \leq n\}$ is an $\mathbb{R}^r$-valued i.i.d. sample from some multivariate distribution function with marginals $H^{(m)}$ for $m = 1, \ldots, r$. Furthermore, let h be a function of r variables such that h in $L_p\left(H^{(1)} \otimes \cdots \otimes H^{(r)}\right)$ for a $p \geq 2$. For $r = 2$ and $r = 3$, we can write respectively:*

$$\begin{aligned}
&h(X_i^{(1)}, X_l^{(2)}, X_k^{(3)})I(X_i^{(1)} \leq x^{(1)}, X_l^{(2)} \leq x^{(2)}, X_k^{(3)} \leq x^{(3)}) \\
&= \int_0^{x^{(1)}}\int_0^{x^{(2)}}\int_0^{x^{(3)}} h(u_1,u_2,u_3)\,dH^{(1)}(u_1)\,dH^{(2)}(u_2)\,dH^{(3)}(u_3) \\
&- \int_0^{x^{(2)}}\int_0^{x^{(3)}} h(X_i^{(1)},u_2,u_3)\,dH^{(2)}(u_2)\,dH^{(3)}(u_3) \\
&- \int_0^{x^{(1)}}\int_0^{x^{(3)}} h(u_1,X_l^{(2)},u_3)\,dH^{(1)}(u_1)\,dH^{(3)}(u_3) \\
&- \int_0^{x^{(1)}}\int_0^{x^{(2)}} h(u_1,u_2,X_k^{(3)})\,dH^{(1)}(u_1)\,dH^{(2)}(u_2) \\
&+ \int_0^{x^{(1)}} h(u_1,X_l^{(2)},X_k^{(3)})\,dH^{(1)}(u_1) \\
&+ \int_0^{x^{(2)}} h(X_i^{(1)},u_2,X_k^{(3)})\,dH^{(2)}(u_2) \\
&+ \int_0^{x^{(3)}} h(X_i^{(1)},X_l^{(2)},u_3)\,dH^{(3)}(u_3) - R_3h(X_i^{(1)},X_l^{(2)},X_k^{(3)}),
\end{aligned}$$

and

$$
\begin{aligned}
&h(X_i^{(1)}, X_l^{(2)}) I(X_i^{(1)} \le x^{(1)}, X_l^{(2)} \le x^{(2)}) \\
&= -\int_0^{x^{(1)}} \int_0^{x^{(2)}} h(u_1, u_2)\, dH^{(1)}(u_1)\, dH^{(2)}(u_2) + \int_0^{x^{(1)}} h(u_1, X_l^{(2)})\, dH^{(1)}(u_1) \\
&+ \int_0^{x^{(2)}} h(X_i^{(1)}, u_2)\, dH^{(2)}(u_2) - R_2 h(X_i^{(1)}, X_l^{(2)})\,,
\end{aligned}
$$

where the rests satisfy respectively:

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{x^{(1)} \le x_0^{(1)},\, x^{(2)} \le x_0^{(2)},\, x^{(3)} \le x_0^{(3)}} \left| \sum_{(i,l,k) \in \{1,\dots,n\}^3 \,:\, i \ne l,\, l \ne k,\, k \ne i} R_3 h(X_i^{(1)}, X_l^{(2)}, X_k^{(3)}) \right|^p \right] \\
&\le C(p) n^{3p/2} \int_0^{x_0^{(1)}} \int_0^{x_0^{(2)}} \int_0^{x_0^{(3)}} |h(u_1, u_2, u_3)|^p\, dH^{(1)}(u_1) dH^{(2)}(u_2) dH^{(3)}(u_3)\,,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{x^{(1)} \le x_0^{(1)},\, x^{(2)} \le x_0^{(2)}} \left| \sum_{(i,l) \in \{1,\dots,n\}^2 \,:\, i \ne l} R_2 h(X_i^{(1)}, X_l^{(2)}) \right|^p \right] \\
&\le C(p) n^{p} \int_0^{x_0^{(1)}} \int_0^{x_0^{(2)}} |h(u_1, u_2)|^p\, dH^{(1)}(u_1) dH^{(2)}(u_2)\,.
\end{aligned}
$$

2.3.2 A first step toward the proof of Theorem 10

The following lemma is the root of the construction of the approximating processes.

Lemma 5. (Komlós, Major, and Tusnady 1975) *Let α_n^U be the uniform empirical process. On a suitably enlarged probability space, there exists a sequence of Brownian bridges $(B_n^U)_n$ such that:*

$$
\sup_{0 \le s \le 1} |\alpha_n^U(s) - B_n^U(s)| = O\left(\frac{\log n}{\sqrt{n}} \right).
$$

For $t \in \mathbb{R}$ and for $\mathbf{t} = (t_0, t_1, \dots, t_{\mathcal{J}+2}) \in \mathbb{R}^{\mathcal{J}+3}$, we introduce the following empirical processes:

$$
\begin{aligned}
\alpha_n(t) &= \sqrt{n}\,(H_n(t) - H(t))\,, \\
\alpha_n^{(0)}(t) &= \sqrt{n}\left(H_n^{(0)}(t) - H^{(0)}(t)\right)\,, \\
\alpha_n^{(1)}(t) &= \sqrt{n}\left(H_n^{(1)}(t) - H^{(1)}(t)\right)\,, \\
\alpha_n^{(1,j)}(t) &= \sqrt{n}\left(H_n^{(1,j)}(t) - H^{(1,j)}(t)\right) \quad \text{for } j = 1, \dots, \mathcal{J}\,, \\
\boldsymbol{\alpha}_n(\mathbf{t}) &= \left(\alpha_n^{(0)}(t_0), \alpha_n^{(1,1)}(t_1), \dots, \alpha_n^{(1,\mathcal{J})}(t_{\mathcal{J}}), \alpha_n^{(1)}(t_{\mathcal{J}+1}), \alpha_n(t_{\mathcal{J}+2})\right).
\end{aligned}
$$

Let $r_0 = 0$ and $r_j = H^{(1,1)}(\infty) + \cdots + H^{(1,j)}(\infty)$ for $j = 1, \ldots, \mathcal{J}$. Note that $r_{\mathcal{J}} = H^{(1)}(\infty)$. Let (B_n^U) be the sequence of Brownian bridges of the preceding lemma. We define for $t \in \mathbb{R}$ and for $\mathbf{t} = (t_0, t_1, \ldots, t_{\mathcal{J}+2}) \in \mathbb{R}^{\mathcal{J}+3}$:

$$\begin{aligned}
B_n^{(1,j)}(t) &= B_n^U(r_{j-1} + H^{(1,j)}(t)) - B_n^U(r_{j-1})\,, \\
B_n^{(1)}(t) &= \sum_{j=1}^{\mathcal{J}} B_n^{(1,j)}(t)\,, \\
B_n^{(0)}(t) &= B_n^U(r_{\mathcal{J}} + H^{(0)}(t)) - B_n^U(r_{\mathcal{J}})\,, \\
B_n(t) &= B_n^{(1)}(t) + B_n^{(0)}(t)\,, \\
\mathbf{B}_n(\mathbf{t}) &= \left(B_n^{(0)}(t_0), B_n^{(1,1)}(t_1), \ldots, B_n^{(1,\mathcal{J})}(t_{\mathcal{J}}), B_n^{(1)}(t_{\mathcal{J}+1}), B_n(t_{\mathcal{J}+2})\right).
\end{aligned}$$

In addition, we define the norm of a vector $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ by $\|\mathbf{x}\|_k := \max_{i=1,\ldots,k} |x_i|$.

The next step toward approximation of the Nelson-Aalen, the Kaplan-Meier and the Aalen-Johansen estimators is the following lemma.

Lemma 6. (Horváth 1980; Burke, Csörgő, and Horváth 1981) *The process $\boldsymbol{\alpha}_n(\mathbf{t})$ may be strongly approximated by $\mathbf{B}_n(\mathbf{t})$ on the probability space of Lemma 5. Namely,*

$$\sup_{\mathbf{t} \in [0, \tau_H]^{\mathcal{J}+3}} \|\boldsymbol{\alpha}_n(\mathbf{t}) - \mathbf{B}_n(\mathbf{t})\|_{\mathcal{J}+3} = O\left(\frac{\log n}{\sqrt{n}}\right).$$

The Gaussian processes $(B_n^{(0)})$, $(B_n^{(1,1)})$,..., $(B_n^{(1,\mathcal{J})})$, $(B_n^{(1)})$ and (B_n) have mean 0 and covariance structure defined for $k, j = 1, \ldots, \mathcal{J}$ such that $k \neq j$ and for $s, t \geq 0$ by:

$$\begin{aligned}
Cov\left(B_n^{(1,j)}(t), B_n^{(1,j)}(s)\right) &= H^{(1,j)}(s \wedge t) - H^{(1,j)}(t)H^{(1,j)}(s)\,, \\
Cov\left(B_n^{(1,j)}(t), B_n^{(1,k)}(s)\right) &= -H^{(1,j)}(t)H^{(1,k)}(s)\,, \\
Cov\left(B_n^{(1,j)}(t), B_n^{(1)}(s)\right) &= H^{(1,j)}(s \wedge t) - H^{(1,j)}(t)H^{(1)}(s)\,, \\
Cov\left(B_n^{(1,j)}(t), B_n^{(0)}(s)\right) &= -H^{(1,j)}(t)H^{(0)}(s)\,, \\
Cov\left(B_n^{(1,j)}(t), B_n(s)\right) &= H^{(1,j)}(s \wedge t) - H^{(1,j)}(t)H(s)\,, \\
Cov\left(B_n^{(1)}(t), B_n^{(1)}(s)\right) &= H^{(1)}(s \wedge t) - H^{(1)}(t)H^{(1)}(s)\,, \\
Cov\left(B_n^{(0)}(t), B_n^{(0)}(s)\right) &= H^{(0)}(s \wedge t) - H^{(0)}(t)H^{(0)}(s)\,, \\
Cov\left(B_n^{(1)}(t), B_n^{(0)}(s)\right) &= -H^{(1)}(t)H^{(0)}(s)\,, \\
Cov\left(B_n^{(1)}(t), B_n(s)\right) &= H^{(1)}(s \wedge t) - H^{(1)}(t)H(s)\,, \\
Cov\Big(B_n(t), B_n(s)\Big) &= H(s \wedge t) - H(t)H(s).
\end{aligned}$$

We will make use of the strong approximations of the Nelson-Aalen estimators of $\Lambda^{(1,j)}$ for $j = 1, \ldots, \mathcal{J}$ and of the Kaplan-Meier estimator of F on the intervals $[0, T_{n-k_n,n}]$. These approximations are displayed below.

Lemma 7. (Csörgő 1996) *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ and fulfilling hypothesis ($\mathcal{H}2$).*
1. On the suitably enlarged probability space of the former lemma, one can define $\mathcal{J}$ sequences of Gaussian processes $(L_n^{(1,1)}), \ldots, (L_n^{(1,\mathcal{J})})$ such that the following approximations hold jointly for $j = 1, \ldots, \mathcal{J}$.

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\Lambda_n^{(1,j)}(t) - \Lambda^{(1,j)}(t) \right) - L_n^{(1,j)}(t) \right| = O\Big(\sqrt{n} \frac{\log n}{k_n} \Big).$$

The Gaussian processes $L_n^{(1,1)}, \ldots, L_n^{(1,\mathcal{J})}$ have mean 0 and covariance structure given, for $s, t \geq 0$ and for $k, j = 1, \ldots, \mathcal{J}$ such that $k \neq j$, by:

$$\begin{aligned} Cov\Big(L_n^{(1,j)}(s), L_n^{(1,j)}(t) \Big) &= \int_0^{s \wedge t} (1 - \Delta\Lambda) \frac{d\Lambda^{(1,j)}}{1 - H^-}\,, \\ Cov\Big(L_n^{(1,j)}(s), L_n^{(1,k)}(t) \Big) &= 0\,. \end{aligned}$$

2. On the same probability space, letting $L_n = \sum_{j=1}^{\mathcal{J}} L_n^{(1,j)}$, we have:

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n}(\Lambda_n(t) - \Lambda(t)) - L_n(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_n} \right).$$

Furthermore, if F is continuous, then

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \frac{\widehat{F}_n(t) - F(t)}{1 - F(t)} - L_n(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_n} \right).$$

The process L_n is Gaussian with mean 0 and satisfies for $j = 1, \ldots, \mathcal{J}$ and $s, t \geq 0$:

$$\begin{aligned} Cov\Big(L_n(t), L_n(s) \Big) &= \int_0^{s \wedge t} \frac{dH^{(1)}}{(1 - H^-)^2}\,, \\ Cov\Big(L_n(t), L_n^{(1,j)}(s) \Big) &= \int_0^{s \wedge t} \frac{dH^{(1,j)}}{(1 - H^-)^2}\,. \end{aligned}$$

The reader is referred to the articles of Csörgő (1996) for detailed proofs. The approximating process $L_n^{(1,j)}$ is defined by:

$$L_n^{(1,j)}(t) = \int_0^t \frac{dB_n^{(1,j)}}{(1 - H^-)} + \int_0^t \frac{B_n^-}{(1 - H^-)^2} dH^{(1,j)}\,,$$

where the first integral is given by the integration by parts formula. This leads to the following expression

$$L_n^{(1,j)}(t) = \frac{B_n^{(1,j)}(t)}{1 - H(t)} - \int_0^t B_n^{(1,j)} d\left(\frac{1}{1 - H} \right) + \int_0^t \frac{B_n^-}{(1 - H^-)^2} dH^{(1,j)}\,, \tag{2.3}$$

which is used for the calculus of the covariance function.

The following proposition is needed in the sequel. Its proof is postponed at the end of this chapter, in Subsection 2.3.6.

Proposition 5. *Let (k_n) be a sequence of integers satisfying $1 \leq k_n < n$ together with Hypothesis ($\mathcal{H}$). Let us define:*

$$A_1 = \sup_{t \leq T_{n-k_n,n}} \left| \iint \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y < z \leq t)}{1-G^-(z)} dH^{(1)}(y) d(H_n^{(1,j)} - H^{(1,j)})(z) \right|, \tag{2.4}$$

$$A_2 = \sup_{t \leq T_{n-k_n,n}} \left| \int_0^t \frac{H^-(y)}{(1-H^-(y))^2} (F^{(j)}(t) - F^{(j)}(y)) d(H_n^{(1)} - H^{(1)})(y) \right|, \tag{2.5}$$

$$A_3 = \sup_{t \leq T_{n-k_n,n}} \left| \int_0^t \frac{(F^{(j)}(y))^2}{(1-H^-(y))^2} d(H_n^{(1)} - H^{(1)})(y) \right|. \tag{2.6}$$

Then, for $i = 1, 2, 3$, we have:

$$A_i = O\left(\frac{n(\log\log n)^{1/2}}{k_n^{3/2}} \right).$$

We are now in position to prove Theorem 10.

2.3.3 Proof of Theorem 10

The proof begins with the following decomposition:

$$\begin{aligned} \widehat{F}_n^{(j)}(t) - F^{(j)}(t) &= \int_0^t (1-\widehat{F}_n^-) d\Lambda_n^{(1,j)} - \int_0^t (1-F^-) d\Lambda^{(1,j)} \\ &= -\int_0^t \frac{(\widehat{F}_n^- - F^-)}{(1-H_n^-)} dH_n^{(1,j)} + \int_0^t (1-F^-) d(\Lambda_n^{(1,j)} - \Lambda^{(1,j)}) \\ &= I(t) + II(t). \end{aligned} \tag{2.7}$$

The second term of (2.7), $II(t)$, can be approximated straightforwardly using Lemma 7. Indeed, we have uniformly on $[0, T_{n-k_n,n}]$:

$$\begin{aligned} \sqrt{n} II(t) &= \sqrt{n} \int_0^t (1-F^-) d(\Lambda_n^{(1,j)} - \Lambda^{(1,j)}) \\ &= \sqrt{n}(1-F(t))(\Lambda_n^{(1,j)}(t) - \Lambda^{(1,j)}(t)) + \int_0^t \sqrt{n}(\Lambda_n^{(1,j)} - \Lambda^{(1,j)}) dF \\ &= (1-F(t)) L_n^{(1,j)}(t) + \int_0^t L_n^{(1,j)} dF + O\left(\sqrt{n} \frac{\log n}{k_n} \right), \end{aligned} \tag{2.8}$$

where $L_n^{(1,j)}$ is defined in (2.3). The first term of (2.7), $I(t)$, requires additional work. Indeed, we write:

$$\begin{aligned} I(t) &= -\int_0^t (\widehat{F}_n^- - F^-) \frac{dH_n^{(1,j)}}{1-H^-} - \int_0^t (\widehat{F}_n^- - F^-) \left(\frac{1}{1-H_n^-} - \frac{1}{1-H^-} \right) dH_n^{(1,j)} \\ &= -\int_0^t \frac{\widehat{F}_n^- - F^-}{1-F^-} \frac{dH_n^{(1,j)}}{1-G^-} - \int_0^t (\widehat{F}_n^- - F^-) \frac{H_n^- - H^-}{(1-H_n^-)(1-H^-)} dH_n^{(1,j)} \\ &= I_1(t) + I_2(t). \end{aligned} \tag{2.9}$$

The second term of (2.9), $I_2(t)$, is negligible. Indeed, for $t \leq T_{n-k_n,n}$, the majorization below holds almost-surely:

$$\begin{aligned} |I_2(t)| \;\leq\; & \sup_{t\leq T_{n-k_n,n}} \left| \frac{\widehat{F}_n(t)-F(t)}{1-F(t)} \right| \times \sup_{t\leq T_{n-k_n,n}} \left| \frac{H_n^-(t)-H^-(t)}{(1-H^-(t))^{1/2}} \right| \\ & \times \sup_{t\leq T_{n-k_n,n}} \left\{ \left(\frac{1-H_n^-(t)}{1-H^-(t)} \right)^{1/2} \right\} \times \int_0^{T_{n-k_n,n}} \frac{dH_n}{(1-H_n^-)^{3/2}} \,. \end{aligned} \tag{2.10}$$

Lemmas 1 and 2 and Theorem 15 entail that:

$$\sup_{t\leq T_{n-k_n,n}} |\sqrt{n} I_2(t)| = O\left(\frac{\sqrt{n}\log\log n}{k_n} \right).$$

It remains to treat $I_1(t)$. Proposition 1 implies:

$$\sup_{t\leq T_{n-k_n,n}} \left\{ \sqrt{n} \left| I_1(t) + \int_0^t (\Lambda_n^- - \Lambda^-) \frac{dH_n^{(1,j)}}{1-G^-} \right| \right\} = O\left(\sqrt{n} \frac{\log\log n}{k_n} \right).$$

Consequently, it suffices to consider the following quantity:

$$\begin{aligned} & -\int_0^t (\Lambda_n^- - \Lambda^-) \frac{dH_n^{(1,j)}}{1-G^-} \\ & = -\int_{s=0}^t \left(\int_{u=0}^{s^-} \frac{dH_n^{(1)}}{1-H_n^-} - \int_{u=0}^{s^-} \frac{dH^{(1)}}{1-H^-} \right) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \\ & = -\int_{s=0}^t \left(\int_{u=0}^{s^-} \frac{H_n^-(u)-H^-(u)}{(1-H_n^-(u))(1-H^-(u))} dH_n^{(1)}(u) + \int_{u=0}^{s^-} \frac{d(H_n^{(1)}-H^{(1)})(u)}{1-H^-(u)} \right) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \\ & = -\int_{s=0}^t \left(\int_{u=0}^{s^-} \frac{H_n^-(u)-H^-(u)}{(1-H_n^-(u))(1-H^-(u))} dH_n^{(1)}(u) \right) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \qquad (2.11) \\ & -\int_{s=0}^t \left(\frac{H_n^{(1)-}(s)-H^{(1)-}(s)}{1-H^-(s)} - \int_{u=0}^{s^-} (H_n^{(1)}-H^{(1)}) d\left(\frac{1}{1-H} \right)(u) \right) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \,. \qquad (2.12) \end{aligned}$$

Concerning the term in (2.11), we notice that:

$$\begin{aligned} & -\int_{s=0}^t \int_{u=0}^{s^-} \frac{H_n^-(u)-H^-(u)}{(1-H_n^-(u))(1-H^-(u))} dH_n^{(1)}(u) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \\ & = -\int_{s=0}^t \int_{u=0}^{s^-} \frac{H_n^-(u)-H^-(u)}{(1-H^-(u))^2} dH_n^{(1)}(u) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \\ & -\int_{s=0}^t \int_{u=0}^{s^-} \frac{(H_n^-(u)-H^-(u))^2}{(1-H_n^-(u)(1-H^-(u))^2} dH_n^{(1)}(u) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \,, \end{aligned} \tag{2.13}$$

where the following almost-sure majorization holds:

$$\left|\int_{s=0}^{t}\int_{u=0}^{s^-}\frac{(H_n^-(u)-H^-(u))^2}{(1-H_n^-(u)(1-H^-(u))^2}dH_n^{(1)}(u)\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\right|$$
$$\leq \sup_{t\leq T_{n-k_n,n}}\left|\left(\frac{H_n^-(u)-H^-(u)}{\sqrt{1-H^-(u)}}\right)^2\right|.\sup_{t\leq T_{n-k_n,n}}\left|\left(\frac{1-H_n^-(t)}{1-H^-(t)}\right)^2\right|$$
$$\times\int_{s=0}^{t}\int_{u=0}^{s^-}\frac{dH_n(u)}{(1-H_n^-(u))^2}\frac{dH_n^{(1,j)}(s)}{1-H_n^-(s)}. \tag{2.14}$$

With the help of Lemmas 1 and 2, we get:

$$\int_{s=0}^{T_{n-k_n,n}}\int_{u=0}^{s^-}\frac{dH_n^{(1)}(u)}{(1-H_n^-(u))^2}\frac{dH_n^{(1,j)}(s)}{1-H_n^-(s)}\leq 2\int_{s=0}^{t}\frac{dH_n(s)}{(1-H_n^-(s))^2}\leq 2\,\frac{n}{k_n}. \tag{2.15}$$

Then, we obtain from (2.14) and (2.15):

$$\sqrt{n}\left|\int_{s=0}^{t}\int_{u=0}^{s^-}\frac{(H_n^-(u)-H^-(u))^2}{(1-H_n^-(u)(1-H^-(u))^2}dH_n^{(1)}(u)\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\right|=O\left(\frac{\sqrt{n}\log\log n}{k_n}\right). \tag{2.16}$$

Consequently, combining (2.11), (2.12) with the decomposition in (2.13) and the majorization in (2.16), we only need to treat the three following terms:

$$\begin{aligned}
I_{11}(t) &= -\int_{s=0}^{t}\int_{u=0}^{s^-}\frac{H_n^-(u)-H^-(u)}{(1-H^-(u))^2}dH_n^{(1)}(u)\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\,,\\
I_{12}(t) &= -\int_{s=0}^{t}\frac{H_n^{(1)-}(s)-H^{(1)-}(s)}{1-H^-(s)}\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\,,\\
I_{13}(t) &= \int_{s=0}^{t}\int_{u=0}^{s^-}(H_n^{(1)}(u)-H^{(1)}(u))d\left(\frac{1}{1-H}\right)(u)\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\,.
\end{aligned}$$

Term $\sqrt{n}I_{12}(t)$ can directly be approximated by a Gaussian process, using Lemma 6. Namely,

$$\begin{aligned}
\sqrt{n}I_{12}(t) = &-\int_{s=0}^{t}\frac{B_n^{(1)-}(s)}{1-H^-(s)}\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\\
&-\int_{s=0}^{t}\frac{\sqrt{n}(H_n^{(1)-}(s)-H^{(1)-}(s))-B_n^{(1)-}(s)}{1-H^-(s)}\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\,,
\end{aligned}$$

where

$$\left|\int_{s=0}^{t}\frac{\sqrt{n}(H_n^{(1)-}(s)-H^{(1)-}(s))-B_n^{(1)-}(s)}{1-H^-(s)}\frac{dH_n^{(1,j)}(s)}{1-G^-(s)}\right|$$
$$\leq \sup_{t\leq T_{n-k_n,n}}\left|\sqrt{n}(H_n^{(1)-}(s)-H^{(1)-}(s))-B_n^{(1)-}(s)\right|$$
$$\times\sup_{t\leq T_{n-k_n,n}}\left|\left(\frac{1-H_n^-(t)}{1-H^-(t)}\right)^2\right|.\int_0^{T_{n-k_n,n}}\frac{dH_n}{(1-H_n^-)^2}.$$

Consequently, we have uniformly over $[0, T_{n-k_n,n}]$:

$$\sqrt{n} I_{12}(t) = -\int_{s=0}^{t} \frac{B_n^{(1)-}(s)}{1-H^-(s)} \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} + O\left(\frac{\sqrt{n}\log n}{k_n}\right).$$

For the treatment of $I_{11}(t)$ and $I_{13}(t)$, we will follow the method initiated by Stute (1994a) refined by Csörgő (1996). We decompose these terms in sums of U-statistics. First of all, we treat $I_{11}(t)$.

$$\begin{aligned}
I_{11}(t) &= -\int_{s=0}^{t}\int_{u=0}^{s^-} \frac{H_n^-(u)-H^-(u)}{(1-H^-(u))^2} dH_n^{(1)}(u) \frac{dH_n^{(1,j)}(s)}{1-G^-(s)} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int_{u=0}^{T_i^-} \frac{H_n^-(u)-H^-(u)}{(1-H^-(u))^2} dH_n^{(1)}(u) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&= -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n} \frac{H_n^-(T_l)-H^-(T_l)}{(1-H^-(T_l))^2} I(T_l<T_i, J_l\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&= -\frac{1}{n^3}\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n} \frac{I(T_k<T_l)-H^-(T_l)}{(1-H^-(T_l))^2} I(T_l<T_i, J_l\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)}.
\end{aligned}$$

So $I_{11}(t)$ is a V-statistic of degree 3 which is degenerate in one of the three coordinates. We decompose it into its U-statistic components.

$$\begin{aligned}
I_{11}(t) = &-\frac{1}{n^3}\sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} \frac{I(T_k<T_l)-H^-(T_l)}{(1-H^-(T_l))^2} I(T_l<T_i, J_l\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&-\frac{1}{n^3}\sum_{(i,l):\, i\neq l} \frac{I(T_i<T_l)-H^-(T_l)}{(1-H^-(T_l))^2} I(T_l<T_i, J_l\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&-\frac{1}{n^3}\sum_{(i,k):\, i\neq k} \frac{I(T_k<T_k)-H^-(T_k)}{(1-H^-(T_k))^2} I(T_k<T_i, J_k\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&-\frac{1}{n^3}\sum_{(i,k):\, i\neq k} \frac{I(T_k<T_i)-H^-(T_i)}{(1-H^-(T_i))^2} I(T_i<T_i, J_i\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&-\frac{1}{n^3}\sum_{i} \frac{I(T_i<T_i)-H^-(T_i)}{(1-H^-(T_i))^2} I(T_i<T_i, J_i\neq 0) \frac{I(T_i\le t, J_i=j)}{1-G^-(T_i)}.
\end{aligned}$$

Simplifications lead to:

$$\begin{aligned}
I_{11}(t) = &-\frac{1}{n^3}\sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} \frac{I(T_k<T_l)-H^-(T_l)}{(1-H^-(T_l))^2} \frac{I(T_l<T_i, J_l\neq 0, T_i\le t, J_i=j)}{1-G^-(T_i)} \\
&+\frac{2}{n^3}\sum_{(i,l):\, i\neq l} \frac{H^-(T_l)}{(1-H^-(T_l))^2} \frac{I(T_l<T_i, J_l\neq 0, T_i\le t, J_i=j)}{1-G^-(T_i)}.
\end{aligned}$$

For $i = 1, \ldots, n$, we introduce the random variables V_i and W_i defined by:

$$V_i = \begin{cases} T_i & \text{if } J_i\neq 0, \\ \infty & \text{if } J_i = 0, \end{cases} \qquad \text{and} \quad W_i = \begin{cases} T_i & \text{if } J_i = j, \\ \infty & \text{otherwise.} \end{cases}$$

Then $I_{11}(t)$ can be rewritten as:

$$\begin{aligned} I_{11}(t) =& -\frac{1}{n^3} \sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} \frac{I(T_k < V_l) - H^-(V_l)}{(1-H^-(V_l))^2} \frac{I(V_l < W_i, W_i \leq t)}{1-G^-(W_i)} \\ &+ \frac{2}{n^3} \sum_{(i,l):\, i\neq l} \frac{H^-(V_l)}{(1-H^-(V_l))^2} \frac{I(V_l < W_i, W_i \leq t)}{1-G^-(W_i)} \\ =& -\frac{1}{n^3} \sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} f_t(T_k, V_l, W_i) + \frac{2}{n^3} \sum_{(i,l):\, i\neq l} g_t(V_l, W_i)\,, \end{aligned}$$

where we used the notations

$$\begin{aligned} f_t(x,y,z) &= \frac{I(x<y) - H^-(y)}{(1-H^-(y))^2} \frac{I(y<z, z\leq t)}{1-G^-(z)}\,, \\ g_t(y,z) &= \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y<z, z\leq t)}{1-G^-(z)}\,. \end{aligned}$$

We make use of the Hoeffding-like decomposition of Stute (1994a,b) quoted in Proposition 4 to write $I_{11}(t)$ as a sum of canonical U-statistics of different orders:

$$\begin{aligned} I_{11}(t) =& -\frac{1}{n^3} \sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} \Bigg[- R_3 f_t(T_k, V_l, W_i) \\ &\qquad + \int f_t(x, V_l, W_i) dH(x) \\ &\qquad + \int f_t(T_k, y, W_i) dH^{(1)}(y) \\ &\qquad + \int f_t(T_k, V_l, z) dH^{(1,j)}(z) \\ &\qquad - \iint f_t(x, y, W_i) dH(x) dH^{(1)}(y) \\ &\qquad - \iint f_t(x, V_l, z) dH(x) dH^{(1,j)}(z) \\ &\qquad - \iint f_t(T_k, y, z) dH^{(1)}(y) dH^{(1,j)}(z) \\ &\qquad + \iiint f_t(x, y, z) dH(x) dH^{(1)}(y) dH^{(1,j)}(z) \Bigg] \\ &+ \frac{2}{n^3} \sum_{(i,l):\, i\neq l} \Bigg[- R_2 g_t(V_l, W_i) \\ &\qquad + \int g_t(y, W_i) dH^{(1)}(y) + \int g_t(V_l, z) dH^{(1,j)}(z) \\ &\qquad - \iint g_t(y, z) dH^{(1)}(y) dH^{(1,j)}(z) \Bigg]\,. \end{aligned}$$

We compute the different quantities involved in the above equality.

$$\begin{aligned}\int f_t(x, V_l, W_i)dH(x) &= \int \frac{I(x < V_l) - H^-(V_l)}{(1 - H^-(V_l))^2} \frac{I(V_l < W_i, W_i \leq t)}{1 - G^-(W_i)} dH(x) \\ &= 0\end{aligned}$$

$$\int f_t(T_k, y, W_i)dH^{(1)}(y) = \int \frac{I(T_k < y) - H^-(y)}{(1 - H^-(y))^2} \frac{I(y < W_i, W_i \leq t)}{1 - G^-(W_i)} dH^{(1)}(y)$$

$$\begin{aligned}\int f_t(T_k, V_l, z)dH^{(1,j)}(z) &= \int \frac{I(T_k < V_l) - H^-(V_l)}{(1 - H^-(V_l))^2} \frac{I(V_l < z, z \leq t)}{1 - G^-(z)} dH^{(1,j)}(z) \\ &= \frac{I(T_k < V_l) - H^-(V_l)}{(1 - H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t)\end{aligned}$$

$$\begin{aligned}&\iint f_t(x, y, W_i)dH(x)dH^{(1)}(y) \\ =&\iint \frac{I(x < y) - H^-(y)}{(1 - H^-(y))^2} \frac{I(y < W_i, W_i \leq t)}{1 - G^-(W_i)} dH(x)dH^{(1)}(y) \\ =&0\end{aligned}$$

$$\begin{aligned}&\iint f_t(x, V_l, z)dH(x)dH^{(1,j)}(z) \\ =&\iint \frac{I(x < V_l) - H^-(V_l)}{(1 - H^-(V_l))^2} \frac{I(V_l < z, z \leq t)}{1 - G^-(z)} dH(x)dH^{(1,j)}(z) \\ =&0\end{aligned}$$

$$\begin{aligned}&\iint f_t(T_k, y, z)dH^{(1)}(y)dH^{(1,j)}(z) \\ =&\iint \frac{I(T_k < y) - H^-(y)}{(1 - H^-(y))^2} \frac{I(y < z, z \leq t)}{1 - G^-(z)} dH^{(1)}(y)dH^{(1,j)}(z) \\ =&\int \frac{I(T_k < y) - H^-(y)}{(1 - H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y < t)dH^{(1)}(y)\end{aligned}$$

$$\begin{aligned}&\iiint f_t(x, y, z)dH(x)dH^{(1)}(y)dH^{(1,j)}(z) \\ =&\iiint \frac{I(x < y) - H^-(y)}{(1 - H^-(y))^2} \frac{I(y < z, z \leq t)}{1 - G^-(z)} dH(x)dH^{(1)}(y)dH^{(1,j)}(z) \\ =&0\end{aligned}$$

and

$$\int g_t(y, W_i) dH^{(1)}(y) = \int \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y < W_i, W_i \le t)}{1-G^-(W_i)} dH^{(1)}(y)$$

$$\begin{aligned}\int g_t(V_l, z) dH^{(1,j)}(z) &= \int \frac{H^-(V_l)}{(1-H^-(V_l))^2} \frac{I(V_l < z, z \le t)}{1-G^-(z)} dH^{(1,j)}(z) \\ &= \frac{H^-(V_l)}{(1-H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t)\end{aligned}$$

$$\begin{aligned}&\iint g_t(y, z) dH^{(1)}(y) dH^{(1,j)}(z) \\ =&\iint \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y < z, z \le t)}{1-G^-(z)} dH^{(1)}(y) dH^{(1,j)}(z) \\ =&\int \frac{H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y < t) dH^{(1)}(y).\end{aligned}$$

We deduce the following expression for $I_{11}(t)$:

$$\begin{aligned}I_{11}(t) = &- \frac{1}{n^3} \sum_{(i,l,k):\, i \neq l,\, l \neq k,\, i \neq k} \Bigg[- R_3 f_t(T_k, V_l, W_i) \\ &\qquad + \int \frac{I(T_k < y) - H^-(y)}{(1-H^-(y))^2} \frac{I(y < W_i, W_i \le t)}{1-G^-(W_i)} dH^{(1)}(y) \\ &\qquad + \frac{I(T_k < V_l) - H^-(V_l)}{(1-H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t) \\ &\qquad - \int \frac{I(T_k < y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y < t) dH^{(1)}(y) \Bigg] \\ &+ \frac{2}{n^3} \sum_{(i,l):\, i \neq l} \Bigg[- R_2 g_t(V_l, W_i) \\ &\qquad + \int \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y < W_i, W_i \le t)}{1-G^-(W_i)} dH^{(1)}(y) \\ &\qquad + \frac{H^-(V_l)}{(1-H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t) \\ &\qquad - \int \frac{H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y < t) dH^{(1)}(y) \Bigg].\end{aligned}$$

We get then the following expansion of the sums:

$$
\begin{aligned}
I_{11}(t) = & \frac{1}{n^3} \sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} R_3 f_t(T_k, V_l, W_i) \\
& - \frac{(n-2)}{n^3} \sum_{(i,k):\, i\neq k} \int \frac{I(T_k < y) - H^-(y)}{(1-H^-(y))^2} \frac{I(y < W_i, W_i \leq t)}{1-G^-(W_i)} dH^{(1)}(y) \\
& - \frac{(n-2)}{n^3} \sum_{(l,k):\, l\neq k} \frac{I(T_k < V_l) - H^-(V_l)}{(1-H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t) \\
& + \frac{(n-1)(n-2)}{n^2} \int \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y < t) dH^{(1)}(y) \\
& - \frac{2}{n^3} \sum_{(i,l):\, i\neq l} R_2 g_t(V_l, W_i) \\
& + \frac{2(n-1)}{n^3} \sum_{i=1}^{n} \int \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y < W_i, W_i \leq t)}{1-G^-(W_i)} dH^{(1)}(y) \\
& + \frac{2(n-1)}{n^3} \sum_{l=1}^{n} \frac{H^-(V_l)}{(1-H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t) \\
& - \frac{2n(n-1)}{n^3} \int \frac{H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y < t) dH^{(1)}(y) \,.
\end{aligned}
$$

Now, we decompose once again the second and the third term of the above equality, which are U-statistics of degree two. Let

$$
\begin{aligned}
\varphi_t(T_k, W_i) &= \int \frac{I(T_k < y) - H^-(y)}{(1-H^-(y))^2} \frac{I(y < W_i, W_i \leq t)}{1-G^-(W_i)} dH^{(1)}(y) \,, \\
\psi_t(T_k, V_l) &= \frac{I(T_k < V_l) - H^-(V_l)}{(1-H^-(V_l))^2} \left(F^{(j)}(t) - F^{(j)}(V_l)\right) I(V_l < t) \,.
\end{aligned}
$$

Exactly as before, we write:

$$
\begin{aligned}
& \sum_{(i,k):\, i\neq k} \varphi_t(T_k, W_i) \\
= & \sum_{(i,k):\, i\neq k} \int \frac{I(T_k < y) - H^-(y)}{(1-H^-(y))^2} (F^{(j)}(t) - F^{(j)}(y)) I(y < t) dH^{(1)}(y) \\
& - \sum_{(i,k):\, i\neq k} R_2 \varphi_t(T_k, W_i) \\
= & \, n(n-1) \int \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} (F^{(j)}(t) - F^{(j)}(y)) I(y < t) dH^{(1)}(y) \\
& - \sum_{(i,k):\, i\neq k} R_2 \varphi_t(T_k, W_i) \,,
\end{aligned}
$$

and

$$\begin{aligned}
&\sum_{(i,l):\, i\neq l} \psi_t(V_l, W_i) \\
=&\sum_{(i,l):\, i\neq l} \int \frac{I(T_k < y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y<t) dH^{(1)}(y) \\
&- \sum_{(i,l):\, i\neq l} R_2\psi_t(V_l, W_i) \\
=&\, n(n-1) \int \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y<t) dH^{(1)}(y) \\
&- \sum_{(i,l):\, i\neq l} R_2\psi_t(V_l, W_i)\,.
\end{aligned}$$

This leads to:

$$\begin{aligned}
I_{11}(t) =& \frac{1}{n^3} \sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} R_3 f_t(T_k, V_l, W_i) \\
&- \frac{(n-1)(n-2)}{n^2} \int \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y<t) dH^{(1)}(y) \\
&+ \frac{(n-2)}{n^3} \sum_{(i,k):\, i\neq k} R_2\varphi_t(T_k, W_i) \\
&+ \frac{(n-2)}{n^3} \sum_{(k,l):\, k\neq l} R_2\psi_t(T_k, V_l) \\
&- \frac{2}{n^3} \sum_{(i,l):\, i\neq l} R_2 g_t(V_l, W_i) \\
&+ \frac{2(n-1)}{n^2} \iint \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y<z\leq t)}{1-G^-(z)} dH^{(1)}(y) d(H_n^{(1,j)} - H^{(1,j)})(z) \\
&+ \frac{2(n-1)}{n^2} \iint \frac{H^-(y)}{(1-H^-(y))^2} \frac{I(y<z\leq t)}{1-G^-(z)} dH^{(1)}(y) dH^{(1,j)}(z) \\
&+ \frac{2(n-1)}{n^2} \int \frac{H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) I(y<t) d(H_n^{(1)} - H^{(1)})(y) \\
=& I_{111}(t) + I_{112}(t) + \cdots + I_{118}(t)\,.
\end{aligned}$$

First of all, our leading term is $I_{112}(t)$. Indeed,

$$\begin{aligned}
I_{112}(t) =& -\int_0^t \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) dH^{(1)}(y) \\
&+ \frac{3n-2}{n^2} \int_0^t \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) dH^{(1)}(y)
\end{aligned}$$

with uniformly in t in $[0, T_{n-k_n,n}]$

$$\sqrt{n}\frac{3n-2}{n^2} \left| \int_0^t \frac{H_n^-(y) - H^-(y)}{(1-H^-(y))^2} \left(F^{(j)}(t) - F^{(j)}(y)\right) dH^{(1)}(y) \right| = O\left(\sqrt{\frac{\log\log n}{nk_n}}\right).$$

Then, we treat $I_{116}(t)$ and $I_{118}(t)$ with the help of Proposition 5. We have uniformly on $[0, T_{n-k_n,n}]$:

$$\sqrt{n} I_{116}(t) = O\left(\frac{\sqrt{n}(\log\log n)^{1/2}}{k_n^{3/2}}\right) \quad \text{and} \quad \sqrt{n} I_{118}(t) = O\left(\frac{\sqrt{n}(\log\log n)^{1/2}}{k_n^{3/2}}\right).$$

We treat $I_{117}(t)$ with the help of Lemmas 1 and 2. It is straightforward to see that we have uniformly on $[0, T_{n-k_n,n}]$:

$$\sqrt{n} I_{117}(t) = O\left(\frac{\sqrt{n}}{k_n}\right).$$

We now bound terms $I_{111}(t)$, $I_{113}(t)$, $I_{114}(t)$ and $I_{115}(t)$ by means of Proposition 4. Indeed, the rests of degree two and three satisfy the inequalities below which will be used to find almost-sure bounds. Here, $T_{n-k_n,n}$ is denoted simply by t_n. The required expectations will be bounded successively. We also make repeated use of Lemma 1.

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\le t_n}\Big|\sum_{(i,l):\, i\neq l} R_2 g_t(V_l, W_i)\Big|^2\right] \\
&\le Cn^2 \iint \frac{H^-(y)^2}{(1-H^-(y))^4}\frac{I(y<z\le t_n)}{(1-G^-(z))^2} dH^{(1)}(y) dH^{(1,j)}(z) \\
&\le \frac{C\,n^2}{(1-G^-(t_n))}\int_0^{t_n}\frac{dH}{(1-H^-)^4} \\
&\le \frac{4C\,n^2}{3(1-H^-(H^{-1}(1-k_n/7n)))^4} \\
&\le \frac{4\times 49C\,n^6}{3k_n^4}\,.
\end{aligned}$$

The second expectation of use is given by

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\le t_n}\Big|\sum_{(k,l):\, k\neq l} R_2 \psi_t(T_k, V_l)\Big|^2\right] \\
&\le C\,n^2 \iint \frac{(I(x<y)-H^-(y))^2}{(1-H^-(y))^4}(F^{(j)}(t_n)-F^{(j)}(y))^2 I(y<t_n) dH(x) dH^{(1)}(y) \\
&\le C\,n^2 \int_0^{H^{-1}(1-\frac{k_n}{7n})}\frac{dH}{(1-H^-)^3} \\
&\le \frac{3Cn^2}{2(1-H^-(H^{-1}(1-\frac{k_n}{7n})))^2} \\
&\le \frac{4\times 49C\,n^4}{2k_n^2}\,.
\end{aligned}$$

The third expectation of use is:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\le t_n}\Big|\sum_{(i,l):\, i\neq l} R_2\varphi_t(T_k,W_i)\Big|^2\right]\\
&\le C\,n^2\iint\left(\int\frac{I(x<y)-H^-(y)}{(1-H^-(y))^2}\frac{I(y<z\le t_n)}{(1-G^-(z))}dH^{(1)}(y)\right)^2 dH(x)dH^{(1,j)}(z)\\
&= C\,n^2\iiint\frac{(I(x<y_1)-H^-(y_1))(I(x<y_2)-H^-(y_2))}{(1-H^-(y_1))^2(1-H^-(y_2))^2}\frac{I(y_1\vee y_2<z\le t_n)}{(1-G^-(z))^2}\\
&\qquad\qquad\times\, dH(x)dH^{(1)}(y_1)dH^{(1)}(y_2)dH^{(1,j)}(z)\\
&= C\,n^2\iiint\frac{H^-(y_1\wedge y_2)-H^-(y_1)H^-(y_2)}{(1-H^-(y_1))^2(1-H^-(y_2))^2}\frac{I(y_1\vee y_2<z\le t_n)}{(1-G^-(z))^2}dH^{(1)}(y_1)dH^{(1)}(y_2)dH^{(1,j)}(z)\\
&= C\,n^2\iiint\frac{H^-(y_1)}{(1-H^-(y_1))^2(1-H^-(y_2))}\frac{I(y_1\le y_2<z\le t_n)}{(1-G^-(z))^2}dH^{(1)}(y_1)dH^{(1)}(y_2)dH^{(1,j)}(z)\\
&+ C\,n^2\iiint\frac{H^-(y_2)}{(1-H^-(y_1))(1-H^-(y_2))^2}\frac{I(y_2<y_1<z\le t_n)}{(1-G^-(z))^2}dH^{(1)}(y_1)dH^{(1)}(y_2)dH^{(1,j)}(z)\\
&\le\frac{C\,n^2}{(1-G^-(t_n))}\iint\frac{I(y_1\le y_2<z\le t_n)}{(1-H^-(y_1))^2(1-H^-(y_2))}(F^{(j)}(t_n)-F^{(j)}(y_2))dH^{(1)}(y_1)dH^{(1)}(y_2)\\
&+\frac{C\,n^2}{(1-G^-(t_n))}\iint\frac{I(y_2<y_1<z\le t_n)}{(1-H^-(y_1))(1-H^-(y_2))^2}(F^{(j)}(t_n)-F^{(j)}(y_1))dH^{(1)}(y_1)dH^{(1)}(y_2)\\
&\le\frac{C\,n^2}{(1-G^-(t_n))}\iint\frac{I(y_1\le y_2<z\le t_n)}{(1-H^-(y_1))^2(1-H^-(y_2))}(1-F(y_2))(1-G^-(y_2))dF(y_2)dH^{(1)}(y_1)\\
&+\frac{C\,n^2}{(1-G^-(t_n))}\iint\frac{I(y_2<y_1<z\le t_n)}{(1-H^-(y_1))(1-H^-(y_2))^2}(1-F(y_1))(1-G^-(y_1))dF(y_1)dH^{(1)}(y_2)\\
&\le\frac{2C\,n^2}{(1-G^-(t_n))}\int_0^{H^{-1}(1-\frac{k_n}{7n})}\frac{dH}{(1-H^-)^2}\\
&\le\frac{4C\,n^2}{(1-H^-(H^{-1}(1-\frac{k_n}{7n})))^2}\\
&\le\frac{4\times 49C\,n^4}{k_n^2}\,.
\end{aligned}$$

At last, we bound

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\le t_n}\Big|\sum_{(i,l,k):\, i\neq l,\, l\neq k,\, i\neq k} R_3f_t(T_k,V_l,W_i)\Big|^2\right]\\
&\le C\,n^3\iiint\frac{(I(x<y)-H^-(y))^2}{(1-H^-(y))^4}\frac{I(y<z\le t_n)}{(1-G^-(z))^2}dH(x)dH^{(1)}(y)dH^{(1,j)}(z)\\
&\le\frac{3\times 7^3C\,n^6}{2k_n^3}\,.
\end{aligned}$$

We explain now how to bound $\sum_{(i,l,k)\in\{1,\dots,n\}^3,\, i\neq l,\, l\neq k,\, i\neq k} R_3f_t(T_k,V_l,W_i)$ which is denoted here simply by $R_n(t)$. For that, we notice that $\left\{\sup_{t\le t_{2^m}}\left|\frac{R_n(t)}{n(n-1)(n-2)}\right| : 2^{m-1}<n\le 2^m\right\}$,

for fixed m in $\{3,4,\dots\}$, is a reverse nonnegative submartingale in n. Consequently, it is licit to apply Doob's inequality on each block of the union $\cup_{m=3}^{\infty}\{n : 2^{m-1} < n \leq 2^m\}$. We deduce for an arbitrary $\varepsilon > 0$, for m large enough:

$$\begin{aligned}
& \mathbb{P}\left[\max_{2^{m-1}<n\leq 2^m} \sup_{t\leq t_{2^m}} \frac{k_n^{3/2}}{\log n}\left|\frac{R_n(t)}{n^3}\right| > \varepsilon\right] \\
\leq\ & \mathbb{P}\left[\frac{2(k_{2^m})^{3/2}}{M^{3/2}\log 2^m}\max_{2^{m-1}<n\leq 2^m}\sup_{t\leq t_{2^m}}\left|\frac{R_n(t)}{n(n-1)(n-2)}\right| > \varepsilon\right] \\
\leq\ & \mathbb{P}\left[\frac{2(k_{2^m})^{3/2}}{M^{3/2}\log 2^m}\sup_{t\leq t_{2^m}}\left|\frac{R_{2^m}(t)}{2^m(2^m-1)(2^m-2)}\right| > \varepsilon\right].
\end{aligned}$$

We apply Chebychev inequality and the bound for the involved expected value. We get for a $C' > 0$ large enough:

$$\begin{aligned}
& \mathbb{P}\left[\max_{2^{m-1}<n\leq 2^m} \sup_{t\leq t_{2^m}} \frac{k_n^{3/2}}{\log n}\left|\frac{R_n(t)}{n^3}\right| > \varepsilon\right] \\
\leq\ & \frac{4k_{2^m}^3}{\varepsilon^2 M^3(\log 2^m)^2}\mathbb{E}\left[\sup_{t\leq t_{2^m}}\left|\frac{R_{2^m}(t)}{2^m(2^m-1)(2^m-2)}\right|^2\right] \\
\leq\ & \frac{4k_{2^m}^3}{\varepsilon^2 M^3(\log 2^m)^2}\frac{1}{(2^m(2^m-1)(2^m-2))^2}\frac{3\times 7^3 C 2^{6m}}{2k_{2^m}^3} \leq \frac{C'}{m^2}.
\end{aligned}$$

Using Borel-Cantelli, we conclude that:

$$\sup_{t\leq T_{n-k_n,n}}\left|\frac{R_n(t)}{n^3}\right| = o\left(\frac{\log n}{k_n^{3/2}}\right).$$

The other terms may be dealt with in a similar way, leading to:

$$\begin{aligned}
\sup_{t\leq T_{n-k_n,n}}\left|\frac{1}{n}\sum_{(i,l):\, l\neq i} R_2 g_t(V_l, W_i)\right| &= o\left(\frac{\log n}{k_n}\right), \\
\sup_{t\leq T_{n-k_n,n}}\left|\frac{1}{n^2}\sum_{(i,l):\, l\neq i} R_2 \varphi_t(V_l, W_i)\right| &= o\left(\frac{\log n}{k_n}\right), \\
\sup_{t\leq T_{n-k_n,n}}\left|\frac{1}{n^2}\sum_{(k,l):\, l\neq k} R_2 \psi_t(T_k, V_l)\right| &= o\left(\frac{\log n}{k_n}\right).
\end{aligned}$$

The remaining term, $I_{13}(t)$, can be dealt with in the same way. To see this, we write:

$$I_{13}(t) = \frac{1}{n^2}\sum_{i=1}^n\sum_{l=1}^n \int \left(I(V_j \leq u) - H^{(1)}(u)\right) I(u < W_i) d\left(\frac{1}{1-H}\right)(u)\frac{I(W_i \leq t)}{1-G^-(W_i)}.$$

The same arguments as previously leads to:

$$I_{13}(t) = \int_{s=0}^{t} \int_{u=0}^{s^-} B_n^{(1)}(u) d\left(\frac{1}{1-H}\right)(u) dF^{(j)}(s) \; + O\left(\frac{\sqrt{n}\log n}{k_n}\right),$$

uniformly over $[0, T_{n-k_n,n}]$.

Finally, the Gaussian approximating processes of $\sqrt{n}(\widehat{F}_n^{(j)} - F^{(j)})$ are defined for $t \geq 0$ by:

$$\begin{aligned}
\bar{L}_n^{(j)}(t) &= -\int_{s=0}^{t} \int_{u=0}^{s^-} \frac{B_n^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) dF^{(j)}(s) \\
&\quad - \int_{s=0}^{t} \frac{B_n^{(1)-}(s)}{(1-H^-(s))} dF^{(j)}(s) \\
&\quad + \int_{s=0}^{t} \int_{u=0}^{s^-} B_n^{(1)}(u) d\left(\frac{1}{1-H}\right)(u) dF^{(j)}(s) \\
&\quad + (1-F(t)) L_n^{(1,j)}(t) \\
&\quad + \int_0^t L_n^{(1,j)} dF \\
&= -\int_0^t L_n^- dF^{(j)} + (1-F(t)) L_n^{(1,j)}(t) + \int_0^t L_n^{(1,j)} dF \,.
\end{aligned}$$

The latter expression is then used for the calculus of the covariances. □

Remark 1. *Our approximating processes satisfy the same relationships as the functions they estimate. Indeed, we have:*

$$\sum_{j=1}^{\mathcal{J}} \sqrt{n}\Big(\widehat{F}_n^{(j)} - F^{(j)}\Big) = \sqrt{n}\Big(\widehat{F}_n - F\Big),$$

$$\textit{and} \quad \sum_{j=1}^{\mathcal{J}} \bar{L}_n^{(j)} = -\int_0^t L_n^- \, dF + \int_0^t (1-F) \, dL_n = (1-F(t)) L_n(t) \,.$$

2.3.4 Proofs of Theorems 11 and 12

Proof of Theorem 11. The proof is straightforward with the following decomposition:

$$K_n^{(j)}(t) = \sqrt{n}(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)) + \sqrt{n} F^{(j)}(t) \frac{\widehat{F}_n(t) - F(t)}{1-F(t)}. \tag{2.17}$$

We then get the desired result letting:

$$\check{L}_n^{(j)}(t) = \bar{L}_n^{(j)}(t) + F^{(j)}(t) L_n(t). \quad \square$$

In order to prove Theorem 12, we need the following lemmas.

Lemma 8. *Let (k_n) be a sequence of integers fulfilling $1 \leq k_n < n$ and Condition $(\mathcal{H})$. For $j = 1, \ldots, \mathcal{J}$, $F^{(j)}$ is assumed to be continuous. On a suitably enlarged probability space, one can define $\mathcal{J}$ sequences of Gaussian processes $(\check{L}_n^{\dagger(1,1)}), \ldots, (\check{L}_n^{\dagger(1,\mathcal{J})})$ such that the following approximations hold jointly for $j = 1, \ldots, \mathcal{J}$:*

$$\sup_{t \leq T_{n-k_n,n}} \left| K_n^{(j)}(t) - \check{L}_n^{\dagger(j)}(t) \right| \;=\; O\left(\sqrt{n} \frac{(\log n)^2}{k_n} \right)$$

where $K_n^{(j)}$ has been defined in (2.1). The distributional equality $\check{L}_n^{\dagger(j)} \stackrel{\mathcal{D}}{=} W_j(C^{(j)}, n)/\sqrt{n}$ holds for $j = 1, \ldots, \mathcal{J}$ where $W_1(.,.)$,...,$W_{\mathcal{J}}(.,.)$ are correlated two-parameter Wiener processes.

Proof of Lemma 8. Following Csörgő (1996), let us redefine our Brownian bridges $B_n^{(1,j)}$, $B_n^{(1)}$, $B_n^{(0)}$ and B_n. Consider a Kiefer process $\{K(s,u) : 0 \leq s \leq 1, 0 \leq u < \infty\}$ such that for $n \in \mathbb{N}^*$, $K(.,n)/\sqrt{n}$ is a Brownian bridge. We let:

$$\begin{aligned}
B_n^{\dagger(1,j)}(t) &= \frac{1}{\sqrt{n}} \left(K\left(r_{j-1} + H^{(1,j)}(t),\, n \right) - K(r_{j-1}, n) \right) , \\
B_n^{\dagger(1)}(t) &= \sum_{i=1}^{n} B_n^{\dagger(1,j)}(t) \, , \\
B_n^{\dagger(0)}(t) &= \frac{1}{\sqrt{n}} \left(K\left(r_{\mathcal{J}} + H^{(0)}(t),\, n \right) - K(r_{\mathcal{J}}, n) \right) , \\
B_n^{\dagger}(t) &= B_n^{\dagger(1)}(t) + B_n^{\dagger(0)}(t) \, .
\end{aligned}$$

Then, letting for $\mathbf{t} = (t_0, t_1, \ldots, t_{\mathcal{J}+2}) \in \mathbb{R}^{\mathcal{J}+3}$:

$$\begin{aligned}
\boldsymbol{\alpha}_n(\mathbf{t}) &= (\alpha_n^{(0)}(t_0), \alpha_n^{(1,1)}(t_1), \ldots, \alpha_n^{(1,\mathcal{J})}(t_{\mathcal{J}}), \alpha_n^{(1)}(t_{\mathcal{J}+1}), \alpha_n(t_{\mathcal{J}+2})) \, , \\
\mathbf{B}_n^{\dagger}(\mathbf{t}) &= (B_n^{\dagger(0)}(t_0), B_n^{\dagger(1,1)}(t_1), \ldots, B_n^{\dagger(1,\mathcal{J})}(t_{\mathcal{J}}), B_n^{\dagger(1)}(t_{\mathcal{J}+1}), B_n^{\dagger}(t_{\mathcal{J}+2})) \, ,
\end{aligned}$$

the following strong approximation holds on a suitably enlarged probability space:

$$\sup_{\mathbf{t} \in [0, \tau_H]^{\mathcal{J}+3}} \left\| \boldsymbol{\alpha}_n(\mathbf{t}) - \mathbf{B}_n^{\dagger}(\mathbf{t}) \right\|_{\mathcal{J}+3} = O\left(\frac{(\log n)^2}{\sqrt{n}} \right) .$$

The Gaussian processes $(B_n^{\dagger})$, $(B_n^{\dagger(1,1)})$, ..., $(B_n^{\dagger(1,\mathcal{J})})$ have mean 0 and covariance structure

defined for $k, j = 1, \ldots, \mathcal{J}$ such that $k \neq j$ and for $s, t \geq 0$ by:

$$
\begin{aligned}
Cov\left(B_n^{\dagger(1,j)}(t), B_m^{\dagger(1,j)}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H^{(1,j)}(s \wedge t) - H^{(1,j)}(t) H^{(1,j)}(s)\right) , \\
Cov\left(B_n^{\dagger(1,j)}(t), B_m^{\dagger(1,k)}(s)\right) &= -\frac{n \wedge m}{\sqrt{nm}} H^{(1,j)}(t) H^{(1,k)}(s) , \\
Cov\left(B_n^{\dagger(1,j)}(t), B_m^{\dagger(1)}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H^{(1,j)}(s \wedge t) - H^{(1,j)}(t) H^{(1)}(s)\right) , \\
Cov\left(B_n^{\dagger(1,j)}(t), B_m^{\dagger(0)}(s)\right) &= -\frac{n \wedge m}{\sqrt{nm}} H^{(1,j)}(t) H^{(0)}(s) , \\
Cov\left(B_n^{\dagger(1,j)}(t), B_m^{\dagger}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H^{(1,j)}(s \wedge t) - H^{(1,j)}(t) H(s)\right) , \\
Cov\left(B_n^{\dagger(1)}(t), B_m^{\dagger(1)}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H^{(1)}(s \wedge t) - H^{(1)}(t) H^{(1)}(s)\right) , \\
Cov\left(B_n^{\dagger(0)}(t), B_m^{\dagger(0)}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H^{(0)}(s \wedge t) - H^{(0)}(t) H^{(0)}(s)\right) , \\
Cov\left(B_n^{\dagger(1)}(t), B_m^{\dagger(0)}(s)\right) &= -\frac{n \wedge m}{\sqrt{nm}} H^{(1)}(t) H^{(0)}(s) , \\
Cov\left(B_n^{\dagger(1)}(t), B_m^{\dagger}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H^{(1)}(s \wedge t) - H^{(1)}(t) H(s)\right) , \\
Cov\left(B_n^{\dagger}(t), B_m^{\dagger}(s)\right) &= \frac{n \wedge m}{\sqrt{nm}} \left(H(s \wedge t) - H(t) H(s)\right) .
\end{aligned}
$$

Let for $t \geq 0$:

$$
L_n^{\dagger(1,j)}(t) = \frac{B_n^{\dagger(1,j)}(t)}{1 - H(t)} - \int_0^t B_n^{\dagger(1,j)} d\left(\frac{1}{1 - H}\right) + \int_0^t \frac{B_n^{\dagger -}}{(1 - H^-)^2} dH^{(1,j)} .
$$

These $\mathcal{J}$ sequences of Gaussian processes approximate the different cause-specific cumulative hazard processes:

$$
\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\Lambda_n^{(1,j)}(t) - \Lambda^{(1,j)}(t)\right) - L_n^{\dagger(1,j)}(t) \right| = O\left(\sqrt{n} \frac{(\log n)^2}{k_n}\right) .
$$

Then letting

$$
\begin{aligned}
L_n^{\dagger}(t) &= \sum_{j=1}^{\mathcal{J}} L_n^{\dagger(1,j)}(t) , \\
\bar{L}_n^{\dagger(j)}(t) &= -\int_0^t L_n^{\dagger -} dF^{(j)} + (1 - F(t)) L_n^{\dagger(1,j)}(t) + \int_0^t L_n^{\dagger(1,j)} dF ,
\end{aligned}
$$

and imitating the proof of Theorem 10, we get the following strong approximations for $j = 1, \ldots, \mathcal{J}$:

$$
\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\right) - \bar{L}_n^{\dagger(j)}(t) \right| = O\left(\sqrt{n} \frac{(\log n)^2}{k_n}\right) .
$$

Finally, we let:

$$\check{L}_n^{\dagger(j)}(t) = \bar{L}_n^{\dagger(j)}(t) + F^{(j)}(t) L_n^{\dagger}(t)\,,$$

leading to the desired result along the lines of the proof of Theorem 11. □.

Lemma 9. *Let (k_n) be a sequence of integers fulfilling $1 \leq k_n < n$ and Condition $(\mathcal{H})$. We assume that $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$. For $j = 1, \ldots, \mathcal{J}$, we have:*

$$\sup_{t \leq T_{n-k_n,n}} \left| \widehat{C}_n^{(j)}(t) - C^{(j)}(t) \right| = O\left(\frac{n(\log\log n)^{1/2}}{k_n^{3/2}} \right),$$

where $\widehat{C}_n^{(j)}(t)$ is defined in (2.2).

Proof of Lemma 9. We write $\widehat{C}_n^{(j)}(t) - C^{(j)}(t)$ as the sum of three terms $D_1(t)$, $D_2(t)$ and $D_3(t)$ defined for $t \geq 0$ by:

$$\begin{aligned}
D_1(t) &= \int_0^t \frac{\widehat{F}_n^{(j)2}}{(1-H_n^-)^2} dH_n^{(1)} - \int_0^t \frac{F^{(j)2}}{(1-H^-)^2} dH^{(1)}\,,\\
D_2(t) &= \int_0^t \frac{\widehat{F}_n^{(j)}(1-\widehat{F}_n)}{(1-H_n^-)^2} dH_n^{(1,j)} - \int_0^t \frac{F^{(j)}(1-F)}{(1-H^-)^2} dH^{(1,j)}\,,\\
D_3(t) &= \int_0^t \frac{(1-\widehat{F}_n)^2}{(1-H_n^-)^2} dH_n^{(1,j)} - \int_0^t \frac{(1-F)^2}{(1-H^-)^2} dH^{(1,j)}\,.
\end{aligned}$$

Then, we carry out the following decomposition:

$$\begin{aligned}
D_1(t) &= \int_0^t \frac{\widehat{F}_n^{(j)2} - F^{(j)2}}{(1-H_n^-)^2} dH_n^{(1)} + \int_0^t F^{(j)2} \left(\frac{dH_n^{(1)}}{(1-H_n^-)^2} - \frac{dH^{(1)}}{(1-H^-)^2} \right)\\
&= \int_0^t \frac{\widehat{F}_n^{(j)2} - F^{(j)2}}{(1-H_n^-)^2} dH_n^{(1)} + \int_0^t F^{(j)2} \frac{(1-H^-)^2 - (1-H_n^-)^2}{(1-H_n^-)^2(1-H^-)^2} dH_n^{(1)}\\
&\quad + \int_0^t F^{(j)2} \frac{d(H_n^{(1)} - H^{(1)})}{(1-H^-)^2}\,. \qquad (2.18)
\end{aligned}$$

With the help of Lemmas 1, 2, Proposition 5 and Theorem 13 (which is to be proved in the next subsection), first and third terms of (2.18) are seen to be $O(n(\log\log n)^{1/2}/(k_n^{3/2}))$. Concerning the second term of (2.18), we have to work slightly more. We carry out the

following decomposition:

$$\begin{aligned}
&\int_0^t F^{(j)2}\frac{(1-H^-)^2-(1-H_n^-)^2}{(1-H_n^-)^2(1-H^-)^2}dH_n^{(1)}\\
&=\int_0^t F^{(j)2}\,\frac{(H_n^- - H^-)^2+2(1-H_n^-)(H_n^- - H^-)}{(1-H_n^-)^2(1-H^-)^2}dH_n^{(1)}\\
&=\int_0^t F^{(j)2}\,\frac{(H_n^- - H^-)^2}{(1-H_n^-)^2(1-H^-)^2}dH_n^{(1)}+2\int_0^t F^{(j)2}\,\frac{(H_n^- - H^-)}{(1-H_n^-)(1-H^-)^2}dH_n^{(1)}\\
&=\int_0^t F^{(j)2}\,\frac{(H_n^- - H^-)^2}{(1-H_n^-)^2(1-H^-)^2}dH_n^{(1)}+2\int_0^t F^{(j)2}\,\frac{(H_n^- - H^-)}{(1-H^-)^3}dH_n^{(1)}\\
&+2\int_0^t F^{(j)2}\,\frac{(H_n^- - H^-)^2}{(1-H_n^-)(1-H^-)^3}dH_n^{(1)}\,. \qquad (2.19)
\end{aligned}$$

First and third term of (2.19) are $O\left(n\log\log n/k_n^2\right)$. Concerning the second term of (2.19), we write it as a U-statistic plus a remainder term. This gives:

$$\begin{aligned}
&\int_0^t F^{(j)2}\frac{(H_n^- - H^-)}{(1-H^-)^3}dH_n^{(1)}\\
&=\frac{1}{n^2}\sum_{(i,k):\,i\neq k} F^{(j)}(V_i)^2\frac{I(T_k<V_i)-H^-(V_i)}{(1-H^-(V_i))^3}I(V_i\leq t)\\
&-\frac{1}{n}\int_0^t F^{(j)2}\frac{H^-}{(1-H^-)^3}dH_n^{(1)}\,.
\end{aligned}$$

Using the result of Stute (1994b), the first term is a $o\left(n\log n/k_n^2\right)$. The second term is a $O\left(n/k_n^2\right)$.

Terms $D_2(t)$ and $D_3(t)$ can then be handled in the same way. □

Proof of Theorem 12. The following inequalities lead to the desired result.

$$\begin{aligned}
&\sup_{t\leq T_{n-k_n,n}}\left|\frac{K_n^{(j)}(t)}{1+\widehat{C}_n^{(j)}(t)}-\frac{\check{L}_n^{\dagger(j)}}{(1+C^{(j)})}\right|\\
&\leq \sup_{t\leq T_{n-k_n,n}}\left\{\frac{1}{1+\widehat{C}_n^{(j)}(t)}\right\}.\sup_{t\leq T_{n-k_n,n}}\left|K_n^{(j)}(t)-\check{L}_n^{\dagger(j)}(t)\right|\\
&+\sup_{t\leq T_{n-k_n,n}}\left|\frac{\check{L}_n^{\dagger(j)}(t)}{1+C^{(j)}(t)}\right|.\sup_{t\leq T_{n-k_n,n}}\left|\frac{\widehat{C}_n^{(j)}(t)-C^{(j)}(t)}{1+\widehat{C}_n^{(j)}(t)}\right|\\
&\leq \sup_{t\leq T_{n-k_n,n}}\left|K_n^{(j)}(t)-\check{L}_n^{\dagger(j)}(t)\right|\\
&+\sup_{t\leq T_{n-k_n,n}}\left|\frac{\check{L}_n^{\dagger(j)}(t)}{1+C^{(j)}(t)}\right|.\sup_{t\leq T_{n-k_n,n}}\left|\widehat{C}_n^{(j)}(t)-C^{(j)}(t)\right|\,.
\end{aligned}$$

To conclude, we use Lemmas 8 and 9 together with Lemma 2 p.2759 in Csörgő (1996) for

any Kiefer sheet K. It gives:

$$\sup_{t\leq H^{-1}(1-k_n/7n)} \frac{\left|K\left(\frac{C^{(j)}(t)}{1+C^{(j)}(t)},n\right)\right|}{\sqrt{n}} = O\left(\sqrt{\log\log n}\right).$$

This ends the proof. □

2.3.5 Proof of Theorem 13 and 14

Proof of Theorem 13. The following decomposition is obtained for $t \geq 0$ by means of integration by parts:

$$\begin{aligned}
\widehat{F}_n^{(j)}(t) - F^{(j)}(t) &= \int_0^t (1-\widehat{F}_n^-)d\Lambda_n^{(1,j)} - \int_0^t (1-F^-)d\Lambda^{(1,j)} \\
&= \int_0^t (1-\widehat{F}_n^-)d(\Lambda_n^{(1,j)} - \Lambda^{(1,j)}) - \int_0^t (\widehat{F}_n^- - F^-)d\Lambda^{(1,j)} \\
&= (1-\widehat{F}_n^-(t))\left(\Lambda_n^{(1,j)}(t) - \Lambda^{(1,j)}(t)\right) \\
&+ \int_0^t (\Lambda_n^{(1,j)} - \Lambda^{(1,j)})d\widehat{F}_n - \int_0^t \frac{\widehat{F}_n^- - F^-}{1-F^-}dF^{(j)}. \qquad (2.20)
\end{aligned}$$

Equality (2.20) entails that:

$$\sup_{t\leq T_{n-k_n,n}} \left|\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\right| \leq 2 \sup_{u\leq T_{n-k_n,n}} \left|\Lambda_n^{(1,j)}(u) - \Lambda^{(1,j)}(u)\right| + \sup_{u\leq T_{n-k_n,n}} \left|\frac{\widehat{F}_n(u)-F(u)}{1-F(u)}\right|.$$

Notice that the assumption of continuity of $F^{(j)}$ for $j = 1,\ldots,\mathcal{J}$ ensures that F is continuous. We then conclude with Theorem 6 and Theorem 15. □

Proof of Theorem 14. Let $\rightsquigarrow$ denote the almost-sure relative compactness in $D[0,\tau_H]$, the space of functions that are right-continuous, have left-hands limits and $f(\tau_H^-) = f(\tau_H)$ is finite.

Let us deal with the first part of the theorem. Let $\mathcal{F}$ be the Strassen set:

$$\begin{aligned}
\mathcal{S} = \Big\{ & f | f : [0,1] \to \mathbb{R} \text{ is absolutely continuous with derivative f'}, \\
& f(0) = 0, \int_0^1 (f'(u))^2 du \leq 1 \Big\}.
\end{aligned}$$

We denote by $K_n^{(j)t_n}$ the process $K_n^{(j)}$ stopped at the $(n-k_n)$-th order statistic. It is defined for $t \geq 0$ by $K_n^{(j)t_n}(t) = K_n^{(j)}(t \wedge T_{n-k_n,n})$.
Using Theorem 11 and the functional log log laws which may be found e.g. in Csörgő and Révész (1981), we get the following result:

$$\frac{1}{\sqrt{2\log\log n}} K_n^{(j)t_n} \rightsquigarrow \{g \,:\, g = f(C^{(j)}), f \in \mathcal{S}\}.$$

This entails that:

$$\limsup_n \sup_{t\le T_{n-k_n,n}} \left| \frac{1}{\sqrt{2\log\log n}} K_n^{(j)}(t) \right| = \sqrt{C^{(j)}(\tau_H)}.$$

Then, in view of the decomposition (2.17), we deduce:

$$\begin{aligned}
&\limsup_n \sup_{t\le T_{n-k_n,n}} \frac{1}{\sqrt{2\log\log n}} \left| \sqrt{n}\left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\right)\right| \\
&\le \limsup_n \sup_{t\le T_{n-k_n,n}} \frac{1}{\sqrt{2\log\log n}} \left| K_n^{(j)}(t)\right| \\
&+ \limsup_n \sup_{t\le T_{n-k_n,n}} \frac{1}{\sqrt{2\log\log n}} \left| \sqrt{n}\frac{\widehat{F}_n(t) - F(t)}{1-F(t)}\right| \qquad (2.21)\\
&\le \sqrt{C^{(j)}(\tau_H)} + \sqrt{C(\tau_H)}.
\end{aligned}$$

We refer to Csörgő(1996) to bound the term in (2.21).

Let us deal now with the second part of the theorem. Let $\mathcal{F}$ be the Finkelstein set:

$$\begin{aligned}
\mathcal{F} = \Big\{ & f | f : [0,1] \to \mathbb{R} \text{ is absolutely continuous with derivative f'}, \\
& f(0) = f(1) = 0, \int_0^1 (f'(u))^2 du \le 1 \Big\}.
\end{aligned}$$

As before, $K_n^{(j)t_n}$ denotes the process $K_n^{(j)}$ stopped at the $(n-k_n)$-th order statistic. Notice that the conditions on (k_n) entail that $\sqrt{n}(\log n)^2/(\sqrt{\log\log n}k_n) \overset{n\to\infty}{\longrightarrow} 0$. Using Theorem 12 and the functional log log laws, we have:

$$\frac{1}{\sqrt{2\log\log n}} \frac{K_n^{(j)t_n}}{1+\widehat{C}_n^{(j)}} \rightsquigarrow \{g : g = f(C^{(j)}), f \in \mathcal{F}\}$$

with the convention that if $C^{(j)}(\tau_H) = \infty$, then we set $C^{(j)}(\tau_H)/(1+C^{(j)}(\tau_H)) = 1$. This entails that:

$$\limsup_n \sup_{t\le T_{n-k_n,n}} \left| \frac{1}{\sqrt{2\log\log n}} \frac{K_n^{(j)}(t)}{1+\widehat{C}_n^{(j)}(t)} \right| = \sup_{t\le\tau_H} \frac{\sqrt{C^{(j)}(t)}}{1+C^{(j)}(t)}.$$

Using the fact that $\widehat{C}_n^{(j)}(t) \le 2\widehat{C}_n(t)$ for $t \ge 0$, we then get:

$$\begin{aligned}
&\limsup_n \sup_{t\le t_n} \frac{1}{\sqrt{2\log\log n}} \left| \frac{\sqrt{n}\left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\right)}{1+\widehat{C}_n(t)} \right| \\
&\le 2 \limsup_n \sup_{t\le T_{n-k_n,n}} \left| \frac{1}{\sqrt{2\log\log n}} \frac{K_n^{(j)}(t)}{1+\widehat{C}_n^{(j)}(t)} \right| \\
&+ \limsup_n \sup_{t\le t_n} \frac{1}{\sqrt{2\log\log n}} \left| \frac{\sqrt{n}}{1+\widehat{C}_n(t)} \frac{\widehat{F}_n(t) - F(t)}{1-F(t)} \right| \qquad (2.22)\\
&\le 2 \sup_{t\le\tau_H} \frac{\sqrt{C^{(j)}(t)}}{1+C^{(j)}(t)} + \sup_{t\le\tau_H} \frac{\sqrt{C(t)}}{1+C(t)}.
\end{aligned}$$

We refer to Csörgő (1996) for the bound of the term in (2.22). □.

2.3.6 Proof of Proposition 5

Here $T_{n-k_n,n}$ will be denoted simply by t_n. The first step consists in rewriting the quantities to bound as a sum of i.i.d. random variables. We first treat A_1 in (2.4). For $i = 1, \ldots, n$, we let

$$W_i = \begin{cases} T_i & \text{if } J_i = j, \\ \infty & \text{otherwise.} \end{cases}$$

The W_i are (improper) i.i.d. random variables with distribution function $H^{(1,j)}$. We also introduce for $t \geq 0$ and $n \geq 1$

$$f_{t,n}(z) = \frac{1}{n} \frac{I(z \leq t)}{1 - G^-(z)} \int \frac{H^-(y)}{(1 - H^-(y))^2} I(y < z) dH^{(1)}(y).$$

We then write:

$$\begin{aligned} &\iint \frac{H^-(y)}{(1 - H^-(y))^2} \frac{I(y < z \leq t)}{1 - G^-(z)} dH^{(1)}(y) d(H_n^{(1,j)} - H^{(1,j)})(z) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(\int \frac{H^-(y)}{(1 - H^-(y))^2} \frac{I(y < W_i \leq t)}{1 - G^-(W_i)} dH^{(1)}(y) \right. \\ &\qquad \left. - \mathbb{E}\left[\int \frac{H^-(y)}{(1 - H^-(y))^2} \frac{I(y < W_i \leq t)}{1 - G^-(W_i)} dH^{(1)}(y) \right] \right) \qquad (2.23) \\ &= \sum_{i=1}^{n} \left(f_{t,n}(W_i) - \mathbb{E}[f_{t,n}(W_i)] \right) . \qquad (2.24) \end{aligned}$$

In order to bound (2.24), we work with the class of functions $\mathcal{F}_n = \{f_{t,n} : t \leq t_n\}$. For fixed n, this class of functions is bounded, measurable ($f_{t,n}(z)$ is jointly measurable in t and z) and, in view of Lemma 3, it is a VC-class with respect to the envelop $f_{t_n,n}$. Indeed, for fixed n, $f_{t,n}$ is a function that increases as t increases. So we can apply to $\mathcal{F}_n$ the exponential inequalities of Theorem 16. Let us determine the quantities involved in these inequalities.

$$\begin{aligned} \|f_{t,n}\|_\infty &= \sup_z \left\| \frac{1}{n} \frac{I(z \leq t)}{1 - G^-(z)} \int \frac{H^-(y)}{(1 - H^-(y))^2} I(y < z) dH^{(1)}(y) \right\|_\infty \\ &\leq \frac{2}{n(1 - H^-(t_n))^2} \leq \frac{2 \times 49n}{k_n^2} . \end{aligned}$$

So we may choose $U_n = 14^3\, n/k_n^2$.

$$\begin{aligned}
&\mathbb{E}\left[f_{t,n}(W_1)^2\right] = \mathbb{E}\left[\frac{1}{n^2}\frac{I(W_1 \le t)}{(1-G^-(W_1))^2}\left(\int \frac{H^-(y)}{(1-H^-(y))^2} I(y < W_1) dH^{(1)}(y)\right)^2\right] \\
&= \frac{1}{n^2}\iiint \frac{H^-(y_1)H^-(y_2)}{(1-H^-(y_1))^2(1-H^-(y_2))^2}\frac{I(y_1, y_2 \le z \le t)}{(1-G^-(z))^2} dH^{(1)}(y_1) dH^{(1)}(y_2) dH^{(1,j)}(z) \\
&\le \frac{1}{n^2}\left(\int_0^{t_n} \frac{H^-(y)}{(1-H^-(y))^2} dH^{(1)}(y)\right)^2 \int_0^{t_n} \frac{1}{(1-G^-(z))^2} dH^{(1,j)}(z) \\
&\le \frac{14^3 n}{k_n^3}.
\end{aligned}$$

We may choose $V_n = 4 \times 14^3\, n^2/(k_n^3)$. Then, we let $x = A\, n\sqrt{\log\log n}/(k_n^{3/2})$ for $A > 0$ with

$$\mathbb{E}\left[\left\|\sum_{i=1}^n \Big(f_{t,n}(W_i) - \mathbb{E}[f_{t,n}(W_1)]\Big)\right\|_{\mathcal{F}}\right] \le 14^{3/2} C\, \frac{n}{k_n^{3/2}}.$$

With the above choices, we apply Theorem 16. For n large enough, we get:

$$\begin{aligned}
&\mathbb{P}\left[\left\|\sum_{i=1}^n \Big(f_{t,n}(W_i) - \mathbb{E}[f_{t,n}(W_1)]\Big)\right\|_{\mathcal{F}} > 2A\, \frac{n(\log\log n)^{1/2}}{k_n^{3/2}}\right] \\
&\le K \exp\left(-\frac{A}{14^3 K}\sqrt{k_n \log\log n}\, \log\left(1 + \frac{A}{4}\sqrt{\frac{\log\log n}{k_n}}\right)\right).
\end{aligned}$$

Then, using the fact that, in view of assumption (H_2), for n large enough, $\log\log n/k_n \le 1$ and the fact that $\log(1+x) \ge (x/\lambda).\log(1+\lambda)$ for $x \le \lambda$, we obtain that:

$$\begin{aligned}
&\mathbb{P}\left[\left\|\sum_{i=1}^n \Big(f_{t,n}(W_i) - \mathbb{E}[f_{t,n}(W_1)]\Big)\right\|_{\mathcal{F}} > 2A\, \frac{n(\log\log n)^{1/2}}{k_n^{3/2}}\right] \\
&\le K \exp\left(-\frac{A}{14^3 K}\, \log\log n\, \log\left(1 + \frac{A}{4}\right)\right) \\
&\le K \exp\left(-2\log\log n\right),
\end{aligned}$$

where the last inequality holds for n and A large enough. We will use this inequality with n replaced by 2^j. We let $B_n = 2A\, n(\log\log n)^{1/2}/(k_n^{3/2})$. Notice that for $2^{j-1} < n \le 2^j$,

for a $d > 0$, we have $nB_n \geq d2^j B_{2^j}$. We make use of Lemma 4 and get:

$$\begin{aligned}
&\mathbb{P}\Bigg[\max_{2^{j-1}<n\leq 2^j} \sup_{t\leq t_n} \Bigg| \frac{1}{nB_n} \sum_{i=1}^{n} \Bigg(\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \\
&\qquad\qquad - \mathbb{E}\left[\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \right] \Bigg) \Bigg| > 1 \Bigg] \\
&\leq \mathbb{P}\Bigg[\frac{1}{2^j} \max_{2^{j-1}<n\leq 2^j} \sup_{t\leq t_n} \Bigg| \sum_{i=1}^{n} \Bigg(\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \\
&\qquad\qquad - \mathbb{E}\left[\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \right] \Bigg) \Bigg| > dB_{2^j} \Bigg] \\
&\leq \mathbb{P}\Bigg[\frac{1}{2^j} \max_{2^{j-1}<n\leq 2^j} \sup_{t\leq t_{2^j}} \Bigg| \sum_{i=1}^{n} \Bigg(\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \\
&\qquad\qquad - \mathbb{E}\left[\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \right] \Bigg) \Bigg| > dB_{2^j} \Bigg] \\
&\leq 9\mathbb{P}\Bigg[\frac{1}{2^j} \sup_{t\leq t_{2^j}} \Bigg| \sum_{i=1}^{2^j} \Bigg(\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \\
&\qquad\qquad - \mathbb{E}\left[\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \right] \Bigg) \Bigg| > \frac{d}{30} B_{2^j} \Bigg] \\
&\leq 9K \exp\left(-2\log\log(2^j)\right) = \frac{9K}{(\log 2)^2} \frac{1}{j^2}.
\end{aligned}$$

Using Borel-Cantelli, we conclude that:

$$\begin{aligned}
&\sup_{t\leq t_n} \Bigg| \frac{1}{n} \sum_{i=1}^{n} \Bigg(\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \\
&\qquad - \mathbb{E}\left[\int \frac{I(u < W_i \leq t)}{(1-G^-(W_i))} \frac{H^-(u)}{(1-H^-(u))^2} dH^{(1)}(u) \right] \Bigg) \Bigg| = O\left(\frac{n(\log\log n)^{1/2}}{k_n^{3/2}} \right),
\end{aligned}$$

which, in view of (2.24), proves (2.4).

In order to avoid repetitions, we only indicate the main step of the majorization of the expressions that remain to treat. We turn now to (2.5). We write:

$$\begin{aligned}
&\int_0^t \frac{H^-(y)}{(1-H^-(y))^2} (F^{(j)}(t) - F^{(j)}(y)) d(H_n^{(1)} - H^{(1)})(y) \\
&= \frac{1}{n} \sum_{i=1}^{n} \Bigg(\frac{H^-(V_i)}{(1-H^-(V_i))^2} (F^{(j)}(t) - F^{(j)}(V_i)) I(V_i \leq t) \\
&\qquad - \mathbb{E}\left[\frac{H^-(V_i)}{(1-H^-(V_i))^2} (F^{(j)}(t) - F^{(j)}(V_i)) I(V_i \leq t) \right] \Bigg),
\end{aligned}$$

where we let for $i = 1, \ldots, n$

$$V_i = \begin{cases} T_i & \text{if } J_i \neq 0, \\ \infty & \text{if } J_i = 0. \end{cases}$$

The V_i are (improper) i.i.d. random variables with distribution function $H^{(1)}$. It is convenient to consider the VC-class of functions $\mathcal{F}_n = \{f_{t,n} : t \leq t_n\}$ where $f_{t,n}$ is defined for $y \geq 0$ by

$$f_{t,n}(y) = \frac{1}{n} \frac{H^-(y)}{(1 - H^-(y))^2} (F^{(j)}(t) - F^{(j)}(y)) I(y \leq t).$$

The quantities of use (U_n, V_n and x) are exactly the same as previously and the proof runs as before.

Once again, the proof for (2.6) is identical to that of (2.4). In order to treat:

$$\int_0^t \frac{1}{(1 - H^-)^2} d(H_n^{(1)} - H^{(1)}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{I(V_i \leq t)}{(1 - H^-(V_i))^2} - \mathbb{E}\left[\frac{I(V_i \leq t)}{(1 - H^-(V_i))^2} \right] \right),$$

we work with the VC-class of functions $\mathcal{F}_n = \{f_{t,n} : t \leq t_n\}$ where $f_{t,n}$ is defined by:

$$f_{t,n}(y) = \frac{1}{n} \frac{I(y \leq t)}{(1 - H^-(y))^2}.$$

The proof is then identical. □

Chapter 3

Construction of joint confidence bands for the subdistribution functions

3.1 Introduction

The purpose of this chapter is to develop asymptotic confidence bands for the subdistribution functions $F^{(j)}$. With this aim in view, Section 3.2 is devoted to weak convergence results obtained by means of martingale methods without continuity assumption. These results enable us to construct, in Section 3.3, joint asymptotic confidence bands for the cause-specific subdistribution functions $F^{(j)}$. Then, in Section 3.4, we present a short simulation study to assess the accuracy of the bands.

The notations used in this chapter are defined in Chapter 1.

3.2 Weak convergence

In this section, we derive the joint weak convergence of the processes based on the Aalen-Johansen estimator of the different cause-specific subdistribution functions, correctly centered and normalized, without continuity assumption.

3.2.1 The main result

The next result is the key to construct the asymptotic bands of Section 3.3.

Theorem 17. *Let*

$$K_n^{(0)} = \sqrt{n}\left(\frac{\widehat{F}_n - F}{1-F}\right) \quad \textit{and} \quad K_n^{(j)} = \sqrt{n}\left(\widehat{F}_n^{(j)} - F^{(j)}\frac{1-\widehat{F}_n}{1-F}\right) \quad \textit{for } j = 1, \ldots, \mathcal{J}, \tag{3.1}$$

$$\widetilde{K}_n^{(0)} = \sqrt{n}\left(\frac{\widehat{F}_n - F}{1-\widehat{F}_n}\right) \quad \textit{and} \quad \widetilde{K}_n^{(j)} = \sqrt{n}\left(\widehat{F}_n^{(j)}\frac{1-F}{1-\widehat{F}_n} - F^{(j)}\right) \quad \textit{for } j = 1, \ldots, \mathcal{J}\,.$$

Fix $\sigma < \tau_H$. The following convergence holds jointly in $D^{\mathcal{J}+1}[0,\sigma]$:

$$\left(K_n^{(0)}, K_n^{(1)}, \ldots, K_n^{(\mathcal{J})}\right) \xrightarrow{\mathcal{D}} \mathbf{K} = \left(K^{(0)}, K^{(1)}, \ldots, K^{(\mathcal{J})}\right),$$

$$\left(\widetilde{K}_n^{(0)}, \widetilde{K}_n^{(1)}, \ldots, \widetilde{K}_n^{(\mathcal{J})}\right) \xrightarrow{\mathcal{D}} \mathbf{K} = \left(K^{(0)}, K^{(1)}, \ldots, K^{(\mathcal{J})}\right),$$

where the $K^{(j)}$ are Gaussian processes with mean 0 and covariance structure given for $k, j = 0, \ldots, \mathcal{J}$ and $s, t \geq 0$ by:

$$Cov\left(K^{(j)}(s), K^{(k)}(t)\right) = \sum_{l=1}^{\mathcal{J}} \int_0^{s\wedge t} \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)\left(I(k=l) + \frac{F^{(k)}}{1-F}\right)(1-F^-)(1-F)\frac{dH^{(1,l)}}{(1-H^-)^2}.$$

This entails that the processes $K^{(j)}$ for $j = 0, \ldots, \mathcal{J}$ satisfy the following distributional equality $K^{(j)} \stackrel{\mathcal{D}}{=} W^{(j)}(C^{(j)})$ where the $W^{(j)}$ are correlated Brownian motions and where $C^{(j)}(t) = Var\left(K^{(j)}(t)\right)$.

The proof of this theorem is postponed until Subsection 3.2.3.

3.2.2 Convergence of a useful martingale

Proposition 6. *Define for $j = 0, \ldots, \mathcal{J}$ and $t \geq 0$*

$$\check{M}_n^{(1,j)}(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\mathcal{J}} \int_0^t \left(I(j=k) + \frac{F^{(j)}}{1-F}\right)(1-F^-)dM_n^{(1,k)}\,, \tag{3.2}$$

where the local square integrable martingale $M_n^{(1,k)}$ is defined in (1.17) and where $F^{(0)} \equiv 1$. Let

$$\check{\mathbf{M}}_\mathbf{n} = \left(\check{M}_n^{(1,0)}, \check{M}_n^{(1,1)}, \ldots, \check{M}_n^{(1,\mathcal{J})}\right).$$

Then $\check{\mathbf{M}}_\mathbf{n}$ is a $(\mathcal{J}+1)$-dimensional local square integrable martingale (in t) with mean 0 and predictable covariation process given for $t \geq 0$ and $k, j = 0, \ldots, \mathcal{J}$ by:

$$\left\langle \check{M}_n^{(1,j)}, \check{M}_n^{(1,k)} \right\rangle(t) = \sum_{l=1}^{\mathcal{J}} \int_0^t \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)\left(I(k=l) + \frac{F^{(k)}}{1-F}\right)\frac{(1-H_n^-)}{(1-H^-)}(1-F^-)(1-F)dH^{(1,l)}\,. \tag{3.3}$$

The proof of Proposition 6 is omitted since it is a simple consequence of the construction of the stochastic integral (see the Appendix).

The weak convergence of the $\mathcal{J}$-dimensional martingale $\check{\mathbf{M}}_n$ is required in order to prove Theorem 17. With this aim in view, we quote a criterion of tightness for d-dimensional martingales that Jacod et al. (1987) gave for semimartingales.

Proposition 7. (Jacod and Shiryayev 1987) *Let $\mathcal{B}_n = (\Omega_n, \mathcal{F}_n, \mathbb{P}_n, (\mathcal{F}_n(t))_t)$ be a stochastic basis. Let $\mathbf{X_n} = (X_n^{(1)}, \ldots, X_n^{(d)})$ be a locally square integrable d-dimensional local martingale null at time $t = 0$. We associate with $\mathbf{X_n}$ the following predictable process $\widetilde{Y}_n = \sum_{k=1}^{d} \left\langle X_n^{(k)} \right\rangle$. For $(\mathbf{X_n})$ to be tight, it suffices that, for each n, there exists an increasing (i.e. non-negative, non-decreasing, right-continuous and null at time $t = 0$) predictable process Y_n on $\mathcal{B}_n$ such that $Y_n - \widetilde{Y}_n$ is increasing and such that (Y_n) converges in law to a deterministic process.*

Proposition 8. *Fix $\sigma < \tau_H$. The $(\mathcal{J}+1)$-dimensional martingale $\check{\mathbf{M}}_\mathbf{n}$ converge weakly in $D^{\mathcal{J}+1}[0,\sigma]$ (the space of càdlàg functions from $[0,\sigma]$ to $\mathbb{R}^{\mathcal{J}+1}$ equipped with the Skorohod topology) to a Gaussian martingale denoted by $\check{\mathbf{M}} = \left(\check{M}^{(1,0)}, \check{M}^{(1,1)}, \ldots, \check{M}^{(1,\mathcal{J})}\right)$ with mean 0 and covariance function given for $k, j = 0, \ldots, \mathcal{J}$ and $s, t \geq 0$:*

$$
\begin{aligned}
&Cov\left(\check{M}^{(1,j)}(t), \check{M}^{(1,k)}(s)\right) \\
&= \sum_{l=1}^{\mathcal{J}} \int_0^{s\wedge t} \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)\left(I(k=l) + \frac{F^{(k)}}{1-F}\right)(1-F^-)(1-F)dH^{(1,l)}\,. \qquad (3.4)
\end{aligned}
$$

Proof. As, for $j = 0, \ldots, \mathcal{J}$, $\check{M}_n^{(1,j)}$ is a sum of bounded independent random variables, the finite-dimensional laws of $\check{\mathbf{M}}_\mathbf{n}$ converge weakly by the multidimensional central limit theorem.

To get the weak convergence in $D^{\mathcal{J}+1}[0,\sigma]$, it remains to prove the tightness, what we will obtain applying Proposition 7. Using (3.3), we see that:

$$
\begin{aligned}
& \sup_{s\in[0,\sigma]} \left| \sum_{j=0}^{\mathcal{J}} \left\langle \check{M}_n^{(1,j)} \right\rangle (s) - \sum_{j=0}^{\mathcal{J}} \sum_{l=1}^{\mathcal{J}} \int_0^s \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)^2 (1-F^-)(1-F)dH^{(1,l)} \right| \\
= & \sup_{s\in[0,\sigma]} \left| \sum_{j=0}^{\mathcal{J}} \sum_{l=1}^{\mathcal{J}} \int_0^s \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)^2 (1-F^-)(1-F)(H_n^- - H^-)d\Lambda^{(1,l)} \right| \\
\leq & \sup_{s\in[0,\sigma]} \left| H_n^-(s) - H^-(s) \right| . \sum_{j=0}^{\mathcal{J}} \sum_{l=1}^{\mathcal{J}} \int_0^\sigma \left(I(j=l) + \frac{F^{(j)}}{1-F}\right)^2 (1-F^-)(1-F)d\Lambda^{(1,l)} \\
\leq & \sup_{s\in[0,\sigma]} \left| H_n^-(s) - H^-(s) \right| . 4(\mathcal{J}+1). \int_0^\sigma \frac{dF}{1-F}\,,
\end{aligned}
$$

with, by Glivenko-Cantelli's theorem, $\sup_{s\in[0,\sigma]}|H_n^-(s)-H^-(s)| \overset{n\to\infty}{\longrightarrow} 0$ a.s. Consequently, $\sum_{j=0}^{\mathcal{J}}\left\langle \check{M}_n^{(1,j)}\right\rangle$ converge in probability in $D[0,\sigma]$ to a deterministic process. Since the process $\sum_{j=0}^{\mathcal{J}}\left\langle \check{M}_n^{(1,j)}\right\rangle$ is itself predictable, we get the result. It is then straightforward to see that the limiting process of $\check{\mathbf{M}}_\mathbf{n}$ is Gaussian with mean 0 and covariance function as announced. □

3.2.3 Proof of Theorem 17

Dauxois (2000) proved weak convergence results for the process $\sqrt{n}(\widehat{F}_n-F)/(1-F)$ allowing discontinuities for the distribution function F. The main step of his method consists in applying a theorem of Jakubowski et al. (1989). The proof of Theorem 17 is an adaptation of Dauxois (2000) to the present set-up. It consists in applying a multivariate generalization of the aforementioned theorem of Jakubowski et al. (1989). Indeed, Theorem 2.6. of Jakubowski et al. (1989) on the convergence in law for sequences of stochastic integrals on the space D^1 of Skorohod can straightforwardly be generalized for a sequence of local d-dimensional martingales and a sequence of real càdlàg processes (L_n) in the following way. In the next theorem, the d-dimensional Skorohod space is denoted by D^d.

Theorem 18. (Jakubowski et al. 1989) *For each n, let $\mathcal{B}_n=(\Omega_n,\mathcal{F}_n,\mathbb{P}_n,(\mathcal{F}_n(t))_t)$ be a stochastic basis and let L_n be a real càdlàg process. Let $\mathbf{X_n}=(X_n^{(1)},\ldots,X_n^{(d)})$ be a local d-dimensional martingale. Assume that $\sup_n\mathbb{E}\left[\sup_s\sum_{j=1}^d\left|\Delta X_n^{(j)}(s)\right|\right]<\infty$ and that $\mathbf{X_n}\overset{\mathcal{D}}{\longrightarrow}X$ in D^d. Then the convergence $(L_n,\mathbf{X_n})\overset{\mathcal{D}}{\longrightarrow}(L,\mathbf{X})$ in D^{d+1} entails the convergence $(\int L_n^-\,dX_n^{(1)},\ldots,\int L_n^-\,dX_n^{(d)})\overset{\mathcal{D}}{\longrightarrow}(\int L^-\,dX^{(1)},\ldots,\int L^-\,dX^{(d)})$ in D^d.*

We are now in position to prove Theorem 17.

Proof of Theorem 17. We denote by $K_n^{(j)*}$ the process $K_n^{(j)}$ stopped at the last order statistic i.e. $K_n^{(j)*}(.)=K_n^{(j)}(.\wedge T_{n,n})$. In the same way, we denote by $F^{(j)*}$ (respectively F^*) the function $F^{(j)}$ (respectively F) stopped at the last order statistic.

First, notice that we only need to prove the joint weak convergence of the stopped processes $K_n^{(j)*}$ since $K_n^{(j)*}-K_n^{(j)}$ converges to zero in probability.

The next step consists in deriving a relation between the $(\mathcal{J}+1)$-dimensional martingale $\check{\mathbf{M}}_n$ defined in (3.2) and the processes $K_n^{(j)*}$ for $j=0,\ldots,\mathcal{J}$ defined in (3.1). We carry

out the following decomposition for $j = 1, \ldots, \mathcal{J}$, then get by integration by parts:

$$\begin{aligned}\widehat{F}_n^{(j)}(t) - F^{(j)*}(t) &= \int_0^t J_n \frac{(1-\widehat{F}_n^-)}{n(1-H_n^-)} dM_n^{(1,j)} - \int_0^t \frac{\widehat{F}_n^- - F^{*-}}{1-F^{*-}} dF^{(j)*} \\ &= \int_0^t J_n \frac{1-\widehat{F}_n^-}{n(1-H_n^-)} dM_n^{(1,j)} - F^{(j)*}(t) \frac{\widehat{F}_n(t) - F^*(t)}{1-F^*(t)} \\ &\quad + \int_0^t F^{(j)*} d\left(\frac{\widehat{F}_n - F^*}{1-F^*}\right).\end{aligned}$$

We get:

$$\widehat{F}_n^{(j)}(t) - \frac{1-\widehat{F}_n(t)}{1-F^*(t)} F^{(j)*}(t) = \int_0^t J_n \frac{1-\widehat{F}_n^-}{n(1-H_n^-)} dM_n^{(1,j)} + \int_0^t F^{(j)*} d\left(\frac{\widehat{F}_n - F^*}{1-F^*}\right).$$

Combining Equations (1.5) and (1.19), the following relation holds on $[0, \tau_H]$:

$$\frac{\widehat{F}_n(t) - F^*(t)}{1-F^*(t)} = \sum_{k=1}^{\mathcal{J}} \int_0^t \frac{1-\widehat{F}_n^-}{1-F} \frac{J_n}{n(1-H_n^-)} dM_n^{(1,k)}. \tag{3.5}$$

We deduce:

$$\begin{aligned}&\widehat{F}_n^{(j)}(t) - \frac{1-\widehat{F}_n(t)}{1-F^*(t)} F^{(j)*}(t) \qquad (3.6) \\ &= \int_0^t J_n \frac{1-\widehat{F}_n^-}{n(1-H_n^-)} dM_n^{(1,j)} + \sum_{k=1}^{\mathcal{J}} \int_0^t \frac{J_n F^{(j)}}{n(1-H_n^-)} \frac{1-\widehat{F}_n^-}{1-F} dM_n^{(1,k)} \\ &= \sum_{k=1}^{\mathcal{J}} \int_0^t J_n \frac{1-\widehat{F}_n^-}{n(1-H_n^-)} \left(I(j=k) + \frac{F^{(j)}}{1-F}\right) dM_n^{(1,k)}. \qquad (3.7)\end{aligned}$$

As $F^{(0)} \equiv 1$ and in view of Equations (3.7) and (3.5), we get for $j = 0, \ldots, \mathcal{J}$ and $t \geq 0$:

$$\begin{aligned}K_n^{(j)*}(t) &= \sum_{k=1}^{\mathcal{J}} \int_0^t J_n \frac{1-\widehat{F}_n^-}{1-H_n^-} \left(I(j=k) + \frac{F^{(j)}}{1-F}\right) \frac{dM_n^{(1,k)}}{\sqrt{n}} \\ &= \int_0^t \frac{J_n}{1-H_n^-} \frac{1-\widehat{F}_n^-}{1-F^-} d\check{M}_n^{(1,j)}.\end{aligned}$$

Note that this relation entails that $\{\mathbf{K_n}^*(t), \mathcal{F}_n(t) : t \geq 0\}$ is, for fixed n, a local square integrable $(\mathcal{J}+1)$-dimensional martingale in t with mean 0. Considering the real càdlàg process defined by

$$\zeta_n = \frac{J_n}{1-H_n} \frac{1-\widehat{F}_n}{1-F},$$

each $K_n^{(j)*}$ can be rewritten as a convenient stochastic integral for $j = 0, \ldots, \mathcal{J}$:

$$K_n^{(j)*} = \int_0^{\cdot} \zeta_n^- \, d\check{M}_n^{(1,j)}.$$

The process ζ_n converge in probability to $1/(1-H)$ in $D[0,\sigma]$ and according to Proposition 8, the process $\check{\mathbf{M}}_\mathbf{n}$ converge weakly in $D^{\mathcal{J}+1}[0,\sigma]$ to $\check{\mathbf{M}}$. Careful consideration of the jump times of ζ_n and of the jump times of $\check{\mathbf{M}}_\mathbf{n}$ yields the joint convergence in $D^{\mathcal{J}+2}[0,\sigma]$

$$(\zeta_n, \check{\mathbf{M}}_\mathbf{n}) \xrightarrow{\mathcal{D}} (\zeta, \check{\mathbf{M}})\,.$$

In order to apply Theorem 18, in view of the weak convergence of $\check{\mathbf{M}}_\mathbf{n}$ in Proposition 8 and since $\check{\mathbf{M}}_\mathbf{n}$ is a $(\mathcal{J}+1)$-dimensional martingale, it suffices to show that

$$\sup_n \mathbb{E}\left[\sup_{0\le s\le\sigma}\sum_{j=0}^{\mathcal{J}}\left|\Delta\check{M}_n^{(1,j)}(s)\right|\right] < \infty\,.$$

Now,

$$\begin{aligned}
\left(\mathbb{E}\left[\sup_{0\le s\le\sigma}\sum_{j=0}^{\mathcal{J}}\left|\Delta\check{M}_n^{(1,j)}(s)\right|\right]\right)^2 &\le \mathbb{E}\left[\sup_{0\le s\le\sigma}\left(\sum_{j=0}^{\mathcal{J}}\left|\Delta\check{M}_n^{(1,j)}(s)\right|\right)^2\right]\\
&\le (\mathcal{J}+1)^2\mathbb{E}\left[\sup_{0\le s\le\sigma}\sum_{j=0}^{\mathcal{J}}\left|\Delta\check{M}_n^{(1,j)}(s)\right|^2\right]\\
&\le (\mathcal{J}+1)^2\mathbb{E}\left[\sum_{s\le\sigma}\sum_{j=0}^{\mathcal{J}}\left(\Delta\check{M}_n^{(1,j)}(s)\right)^2\right]\\
&\le (\mathcal{J}+1)^2\sum_{j=0}^{\mathcal{J}}\mathbb{E}\left[\left\langle\check{M}_n^{(1,j)}\right\rangle(\sigma)\right]\,.
\end{aligned}$$

Then, using (3.3), we get:

$$\begin{aligned}
&\left(\mathbb{E}\left[\sup_{0\le s\le\sigma}\sum_{j=0}^{\mathcal{J}}\left|\Delta\check{M}_n^{(1,j)}(s)\right|\right]\right)^2\\
&\le (\mathcal{J}+1)^2\sum_{j=0}^{\mathcal{J}}\sum_{l=1}^{\mathcal{J}}\int_0^\sigma\left(I(j=l)+\frac{F^{(j)}}{1-F}\right)^2(1-F^-)^2(1-\Delta\Lambda^{(1,l)})(1-H^-)d\Lambda^{(1,l)}\\
&\le (\mathcal{J}+1)^2\sum_{j=0}^{\mathcal{J}}\sum_{l=1}^{\mathcal{J}}\int_0^\sigma\left((1-F)I(j=l)+F^{(j)}\right)^2\frac{dF^{(1,l)}}{1-F}\\
&\le 4(\mathcal{J}+1)^3\int_0^\sigma\frac{dF}{1-F} < \infty\,.
\end{aligned}$$

Applying Theorem 18, we get the weak convergence in $D^{\mathcal{J}+1}[0,\sigma]$:

$$\mathbf{K}_n = \left(K_n^{(0)*}, K_n^{(1)*}, \ldots, K_n^{(\mathcal{J})*}\right) \xrightarrow{\mathcal{D}} \mathbf{K} = \left(K^{(0)}, K^{(1)}, \ldots, K^{(\mathcal{J})}\right)\,.$$

It is then straightforward to determine the covariance structure of $\mathbf{K}$. As $K^{(1,j)}$ has mean 0 and using the definition of the predictable variation process, we have:

$$\begin{aligned}\mathrm{Cov}(K^{(1,j)}(t), K^{(1,k)}(s)) &= \mathbb{E}\left[K^{(1,j)}(t)K^{(1,k)}(s)\right] \\ &= \left\langle K^{(1,j)}(t), K^{(1,k)}(s)\right\rangle \\ &= \left\langle \int_0^t \frac{d\check{M}^{(1,j)}}{1-H^-}, \int_0^s \frac{d\check{M}^{(1,k)}}{1-H^-}\right\rangle \\ &= \int_0^{s\wedge t} \frac{d\left\langle \check{M}^{(1,j)}, \check{M}^{(1,k)}\right\rangle}{(1-H^-)^2}\,.\end{aligned}$$

The joint convergence of the processes $\widetilde{K}_n^{(j)}$ is then straightforward. □

Remark 2. *These martingale properties enable us to derive inequalities for the bias of $\widehat{F}_n^{(j)}$ for $j = 1, \ldots, \mathcal{J}$. Fix a σ such that $H^-(\sigma) < 1$. We will make repeated use of the following facts. For $t \geq 0$, the inequality $F^{(j)*}(t) \leq F^{(j)}(t)$ holds and according to Andersen et al. (1993),*

$$\mathbb{E}\left[\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right] = 1\,.$$

For $j = 1, \ldots, \mathcal{J}$ and t in $[0, \sigma]$, as $K_n^{(j)}$ has mean 0, we have:

$$0 \leq \mathbb{E}\left[\widehat{F}_n^{(j)}(t)\right] = \mathbb{E}\left[F^{(j)*}(t)\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right] \leq F^{(j)}(t)\mathbb{E}\left[\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right] = F^{(j)}(t)\,,$$

so that

$$\mathbb{E}\left[\widehat{F}_n^{(j)}(t)\right] - F^{(j)}(t) \leq 0\,.$$

On the other hand, we write:

$$\begin{aligned}\mathbb{E}\left[\widehat{F}_n^{(j)}(t)\right] &= \mathbb{E}\left[F^{(j)*}(t)\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right] \\ &= \mathbb{E}\left[F^{(j)*}(t)\right] - \mathbb{E}\left[F^{(j)*}(t)\frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right]\,. \qquad (3.8)\end{aligned}$$

Concerning the first term of (3.8)*, we compute:*

$$\mathbb{E}\left[F^{(j)*}(t)\right] = \mathbb{E}\left[\int_0^t I(u \leq T_{n,n})dF^{(j)}\right] = \int_0^t (1-(1-H^-(u))^n)dF^{(j)}.$$

Concerning the second term of (3.8)*, we use the Cauchy-Schwarz inequality and get:*

$$\begin{aligned}\mathbb{E}\left[F^{(j)*}(t)\frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right] &\leq \left(\mathbb{E}\left[F^{(j)*}(t)^2\right]\mathbb{E}\left[\left(\frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right)^2\right]\right)^{1/2}\\ &= \left(\mathbb{E}\left[F^{(j)*}(t)^2\right]\right)^{1/2}.\left(\mathbb{E}\left[\left\langle\frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right\rangle\right]\right)^{1/2}\\ &\leq \frac{F^{(j)}(t)}{\sqrt{n}}\left(\int_0^t\mathbb{E}\left[(1-\widehat{F}_n^-)^2\frac{J_n}{1-H_n^-}\right](1-\Delta\Lambda)\frac{d\Lambda}{(1-F)^2}\right)^{1/2}.\end{aligned}$$

We get the following lower bound for the bias:

$$\begin{aligned}&-\int_0^t(1-H^-(u))^n dF^{(j)}-\frac{F^{(j)}(t)}{\sqrt{n}}\left(\int_0^t\mathbb{E}\left[\frac{J_n(1-\widehat{F}_n^-)^2}{1-H_n^-}\right]\frac{dF}{(1-F)(1-F^-)^2}\right)^{1/2}\\ &\leq \mathbb{E}\left[\widehat{F}_n^{(j)}(t)\right]-F^{(j)}(t)\,.\end{aligned}$$

Consequently, for $j=1,\ldots,\mathcal{J}$*,* $\widehat{F}_n^{(j)}$ *is an asymptotically unbiased estimator for* $F^{(j)}$ *on* $[0,\sigma]$*.*

Remark 3. *The martingale properties enable us to prove that* $\widehat{F}_n^{(j)}$ *converge in quadratic mean to* $F^{(j)}$ *on* $[0,\sigma]$ *for* σ *such that* $H^-(\sigma)<1$*. We consider the mean squared error function (MSE) of the Aalen-Johansen estimator which is defined by:*

$$MSE\left(\widehat{F}_n^{(j)}(t)\right)=\mathbb{E}\left[\left(\widehat{F}_n^{(j)}(t)-F^{(j)}(t)\right)^2\right].$$

Using the fact that, for any (a,b)*,*

$$(a+b)^2\leq 2a^2+2b^2\,, \tag{3.9}$$

the MSE is bounded by:

$$MSE\left(\widehat{F}_n^{(j)}(t)\right)\leq 2\mathbb{E}\left[\left(\widehat{F}_n^{(j)}(t)-F^{(j)*}(t)\right)^2\right]+2\mathbb{E}\left[\left(F^{(j)*}(t)-F^{(j)}(t)\right)^2\right]. \tag{3.10}$$

As the equality $F^{(j)*}(t)-F^{(j)}(t)=\int_0^t(J_n-1)dF^{(j)}$ *holds for* $t\geq 0$*, there exists a positive constant* c *such that* $|F^{(j)*}(t)-F^{(j)}(t)|\leq c$*, for* n *large enough. Then the dominated convergence theorem entails that* $\mathbb{E}\left[\left(F^{(j)*}(t)-F^{(j)}(t)\right)^2\right]\overset{n\to\infty}{\longrightarrow}0$*. Consequently, if suffices to consider the second term of* (3.10)*. Making use of* (3.9) *again and of the fact that, for*

$t \geq 0$, we have $F^{(j)*}(t) \leq F^{(j)}(t)$, the second term of (3.10) *is seen to be bounded by:*

$$\begin{aligned}
&\mathbb{E}\left[\left(\widehat{F}_n^{(j)}(t) - F^{(j)*}(t)\right)^2\right] \\
&\leq 2\mathbb{E}\left[\left(\widehat{F}_n^{(j)}(t) - F^{(j)*}(t)\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right)^2\right] + 2\mathbb{E}\left[\left(F^{(j)*}(t)\frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right)^2\right] \\
&\leq 2\mathbb{E}\left[\left\langle \widehat{F}_n^{(j)}(t) - F^{(j)*}(t)\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right\rangle\right] + 2F^{(j)}(t)^2\mathbb{E}\left[\left\langle \frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right\rangle\right].
\end{aligned}$$

Now,

$$\mathbb{E}\left[\left\langle \frac{\widehat{F}_n(t)-F^*(t)}{1-F^*(t)}\right\rangle\right] = \frac{1}{n}\int_0^t \mathbb{E}\left[(1-\widehat{F}_n^-)^2\frac{J_n}{1-H_n^-}\right](1-\Delta\Lambda)\frac{d\Lambda}{(1-F)^2} \overset{n\to\infty}{\longrightarrow} 0.$$

By the same way, we have:

$$\mathbb{E}\left[\left\langle \widehat{F}_n^{(j)}(t) - F^{(j)*}(t)\frac{1-\widehat{F}_n(t)}{1-F^*(t)}\right\rangle\right] \overset{n\to\infty}{\longrightarrow} 0.$$

Finally, we get, for $j = 1, \ldots, \mathcal{J}$, that $MSE(F^{(j)}(t)) \overset{n\to\infty}{\longrightarrow} 0$ meaning that $\widehat{F}_n^{(j)}$ converge in quadratic mean to $F^{(j)}$ on $[0, \sigma]$.

3.3 Confidence bands

Several types of asymptotic confidence bands have been proposed in the literature for the distribution function F in presence of right random censorship. The reader is referred to Gillespie and Fisher (1979), Hall and Wellner (1980), Csörgő and Horváth (1986) for an extensive study on this subject. This section is devoted to the construction of asymptotic simultaneous confidence bands for the cause-specific distribution functions $F^{(j)}$ for $j = 1, \ldots, \mathcal{J}$. We will follow the so-called approach of Hall-Wellner (HW) and Aalen-Nair (AN) respectively.

An estimator for the variances of the components of the limiting process is needed. We propose the following estimator for $C^{(j)}(t) = \text{Var}\,(K^{(j)}(t))$:

$$\begin{aligned}
\widehat{C}_n^{(j)}(t) = &\int_0^t \frac{(1-\widehat{F}_n^-)(1-\widehat{F}_n)}{(1-H_n^-)^2} dH_n^{(1,j)} + 2\int_0^t \widehat{F}_n^{(j)}\frac{1-\widehat{F}_n^-}{(1-H_n^-)^2} dH_n^{(1,j)} \\
&+ \int_0^t \frac{(\widehat{F}_n^{(j)})^2}{(1-H_n^-)^2}\frac{(1-\widehat{F}_n^-)}{(1-\widehat{F}_n)} dH_n^{(1)}.
\end{aligned}$$

The strong uniform convergence of the empirical functions H_n and $H_n^{(1,j)}$, of the Kaplan-Meier estimator $\widehat{F}_n$ and of the Aalen-Johansen estimator $\widehat{F}_n^{(j)}$ implies the strong uniform convergence on $[0, \tau_H[$ of $\widehat{C}_n^{(j)}(t)$ for $j = 1, \ldots, \mathcal{J}$.

We also introduce the following quantities

$$K_W(\lambda, a) = \mathbb{P}\left[\sup_{0\le t\le a} |W(t)| \le \lambda\right]$$

where W is the Brownian motion on $[0,1]$ and

$$K_B(\lambda, a) = \mathbb{P}\left[\sup_{0\le s\le a} |B(s)| \le \lambda\right]$$

where B is the Brownian bridge on $[0,1]$. These quantities have been tabulated for example in Koziol and Byar (1975) or in Hall and Wellner (1980). Their exact expressions may be found in Anderson (1960) or in Gillespie and Fisher (1979) in terms of ϕ the distribution function of the $\mathcal{N}(0,1)$. For $0 < a \le 1$, the quantity $K_W(\lambda, a)$ is given by:

$$K_W(\lambda, a) = 1 - 2\phi\left(-\frac{\lambda}{\sqrt{a}}\right) + 2\sum_{k=1}^{\infty}(-1)^k \left(\phi\left(\frac{\lambda(1+2k)}{\sqrt{a}}\right) - \phi\left(\frac{\lambda(2k-1)}{\sqrt{a}}\right)\right).$$

For $0 < a < 1$, the quantity $K_B(\lambda, a)$ is given by:

$$K_B(\lambda, a) = 1 - 2\phi\left(\frac{-\lambda}{\sqrt{a(1-a)}}\right) + 2\sum_{k=1}^{\infty}(-1)^k e^{-2\lambda^2 k^2} \left(\phi\left(\lambda\,\frac{2k(1-a)+1}{\sqrt{a(1-a)}}\right) - \phi\left(\lambda\,\frac{2k(1-a)-1}{\sqrt{a(1-a)}}\right)\right),$$

and for $a = 1$, the quantity $K_B(\lambda, 1)$ is given by:

$$K_B(\lambda, 1) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2\lambda^2 k^2}.$$

Theorem 19. *Fix $\sigma < \tau_H$. For $(\lambda_1, \ldots, \lambda_\jmath)$ in $]0,\infty[^\jmath$, we introduce*

$$\mathcal{K}_j(t) = \begin{cases} \frac{\lambda_j}{\sqrt{n}}\sqrt{\widehat{C}_n^{(j)}(\sigma)} & \textit{for the AN and the modified AN bands,} \\ \frac{\lambda_j}{\sqrt{n}}\left(1 + \widehat{C}_n^{(j)}(t)\right) & \textit{for the HW and the modified HW bands,} \end{cases}$$

and $\mathcal{K}_0(t) = \sum_{j=1}^{\jmath} \mathcal{K}_j(t)$. We also introduce:

$$\alpha(\sigma) = \begin{cases} 1 - \sum_{j=1}^{\jmath}\left(1 - K_W(\lambda_j, 1)\right) & \textit{for the AN and the modified AN bands,} \\ 1 - \sum_{j=1}^{\jmath}\left(1 - K_B\left(\lambda_j, \frac{C^{(j)}(\sigma)}{1+C^{(j)}(\sigma)}\right)\right) & \textit{for the HW and the modified HW bands.} \end{cases}$$

The asymptotic AN and HW confidence bands are given by:

$$\lim_{n\to\infty} \mathbb{P}\left[\bigcap_{j=1}^{\jmath}\left\{r_n^{(j)}(t) \le F^{(j)}(t) \le R_n^{(j)}(t), t \le \sigma\right\} \bigcap \left\{r_n^{(0)}(t) \le F(t) \le R_n^{(0)}(t), t \le \sigma\right\}\right] \ge \alpha(\sigma),$$

where for $j = 1, \ldots, \mathcal{J}$

$$r_n^{(j)}(t) = \frac{\widehat{F}_n^{(j)}(t) - \mathcal{K}_j(t)}{1 + \mathcal{K}_0(t)}\,,\ R_n^{(j)}(t) = \frac{\widehat{F}_n^{(j)}(t) + \mathcal{K}_j(t)}{1 - \mathcal{K}_0(t)}\,,$$

and

$$r_n^{(0)}(t) = 1 - \frac{1 - \widehat{F}_n(t)}{1 - \mathcal{K}_0(t)}\,,\ R_n^{(0)}(t) = 1 - \frac{1 - \widehat{F}_n(t)}{1 + \mathcal{K}_0(t)}\,.$$

The asymptotic modified AN and modified HW confidence bands are given by:

$$\lim_{n\to\infty} \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}} \left\{\widetilde{r}_n^{(j)}(t) \le F^{(j)}(t) \le \widetilde{R}_n^{(j)}(t), t \le \sigma\right\} \bigcap \left\{\widetilde{r}_n^{(0)}(t) \le F(t) \le \widetilde{R}_n^{(0)}(t), t \le \sigma\right\}\right] \ge \alpha(\sigma)\,,$$

where for $j = 1, \ldots, \mathcal{J}$

$$\begin{aligned}\widetilde{r}_n^{(j)}(t) &= \widehat{F}_n^{(j)}(t)\,(1 - \mathcal{K}_0(t)) - \mathcal{K}_j(t)\,,\\ \widetilde{R}_n^{(j)}(t) &= \widehat{F}_n^{(j)}(t)\,(1 + \mathcal{K}_0(t)) + \mathcal{K}_j(t)\,,\end{aligned}$$

and

$$\begin{aligned}\widetilde{r}_n^{(0)}(t) &= 1 - (1 - \widehat{F}_n(t))\,(1 + \mathcal{K}_0(t))\ ,\\ \widetilde{R}_n^{(0)}(t) &= 1 - (1 - \widehat{F}_n(t)\,(1 - \mathcal{K}_0(t))\ .\end{aligned}$$

Proof of Theorem 19. First of all, we quote the following distributional equalities which are used in the sequel. For any real-valued Wiener process W and any fixed $r > 0$, for any real-valued Brownian bridge B, we have the distributional equalities which may be found e.g. in Shorack and Wellner (1986):

$$W(.) \ \stackrel{\mathcal{D}}{=}\ \sqrt{r}\, W\left(\frac{\cdot}{r}\right)\,, \tag{3.11}$$

$$W(.) \ \stackrel{\mathcal{D}}{=}\ (1 + .)\, B\left(\frac{\cdot}{1 + .}\right)\,. \tag{3.12}$$

For the Aalen-Nair bands, we consider the probability defined by:

$$P_n^{AN} = \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}} \left\{\left|K_n^{(j)}(t)\right| \le \lambda_j \sqrt{\widehat{C}_n^{(j)}(\sigma)}, t \le \sigma\right\}\right]\,. \tag{3.13}$$

Using the Bonferroni inequality, we get the bound:

$$P_n^{AN} \ge 1 - \sum_{j=1}^{\mathcal{J}} \left(1 - \mathbb{P}\left[\left|K_n^{(j)}(t)\right| \le \lambda_j \sqrt{\widehat{C}_n^{(j)}(\sigma)}, t \le \sigma\right]\right)\,.$$

We apply Theorem 17 and get:

$$\lim_{n\to\infty} P_n^{AN} \ge 1 - \sum_{j=1}^{\mathcal{J}} \left(1 - \mathbb{P}\left[\left|W_j(C^{(j)}(t))\right| \le \lambda_j \sqrt{C^{(j)}(\sigma)}, t \le \sigma\right]\right)\,.$$

Now, we apply (3.11) and get:

$$\begin{aligned}\lim_{n\to\infty} P_n^{AN} &\geq 1-\sum_{j=1}^{\mathcal{J}}\left(1-\mathbb{P}\left[\sup_{0\leq u\leq 1}|W_j(u)|\leq\lambda_j\right]\right)\\ &= 1-\sum_{j=1}^{\mathcal{J}}\left(1-K_W(\lambda_j,1)\right).\end{aligned}$$

Now, for the Hall-Wellner bands, instead of (3.13), we consider the probability defined by:

$$P_n^{HW}=\mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\left|K_n^{(j)}(t)\right|\leq\lambda_j\left(1+\widehat{C}_n^{(j)}(t)\right),t\leq\sigma\right\}\right]. \tag{3.14}$$

Then using the Bonferroni inequality, we get the bound:

$$P_n^{HW}\geq 1-\sum_{j=1}^{\mathcal{J}}\left(1-\mathbb{P}\left[\left|K_n^{(j)}(t)\right|\leq\lambda_j\left(1+\widehat{C}_n^{(j)}(t)\right),t\leq\sigma\right]\right).$$

Applying Theorem 17, we obtain:

$$\begin{aligned}\lim_{n\to\infty} P_n^{HW} &\geq 1-\sum_{j=1}^{\mathcal{J}}\left(1-\mathbb{P}\left[\left|W_j(C^{(j)}(t))\right|\leq\lambda_j\left(1+C^{(j)}(t)\right),t\leq\sigma\right]\right)\\ &= 1-\sum_{j=1}^{\mathcal{J}}\left(1-\mathbb{P}\left[|W_j(u)|\leq\lambda_j\left(1+u\right),t\leq C^{(j)}(\sigma)\right]\right).\end{aligned}$$

Now, applying (3.12), we obtain:

$$\begin{aligned}\lim_{n\to\infty} P_n^{HW} &\geq 1-\sum_{j=1}^{\mathcal{J}}\left(1-\mathbb{P}\left[\sup_{0\leq t\leq C^{(j)}(\sigma)}\left|B_j\left(\frac{u}{1+u}\right)\right|\leq\lambda_j\right]\right)\\ &= 1-\sum_{j=1}^{\mathcal{J}}\left(1-K_B\left(\lambda_j,\frac{C^{(j)}(\sigma)}{1+C^{(j)}(\sigma)}\right)\right).\end{aligned}$$

Here, P_n will denote either P_n^{AN} or P_n^{HW}. Since the different risks are mutually exclusive, the relations $F=\sum_{j=1}^{\mathcal{J}}F^{(j)}$, $\widehat{F}_n=\sum_{j=1}^{\mathcal{J}}\widehat{F}_n^{(j)}$ and $K_n^{(0)}=\sum_{j=1}^{\mathcal{J}}K_n^{(j)}$ hold. We then

write:

$$\begin{aligned}
P_n &= \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\left|K_n^{(j)}(t)\right| \le \sqrt{n}\mathcal{K}_j(t), t \le \sigma\right\} \bigcap \left\{\left|K_n^{(0)}(t)\right| \le \sqrt{n}\mathcal{K}_0(t), t \le \sigma\right\}\right] \\
&= \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\frac{1-F(t)}{1-\widehat{F}_n(t)}\left(\widehat{F}_n^{(j)}(t) - \mathcal{K}_j(t)\right) \le F^{(j)}(t) \le \frac{1-F(t)}{1-\widehat{F}_n(t)}\left(\widehat{F}_n^{(j)}(t) + \mathcal{K}_j(t)\right), t \le \sigma\right\}\right. \\
&\qquad\qquad \left.\bigcap\left\{\frac{1}{1+\mathcal{K}_0(t)} \le \frac{1-F(t)}{1-\widehat{F}_n(t)} \le \frac{1}{1-\mathcal{K}_0(t)}, t \le \sigma\right\}\right] \\
&\le \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\frac{\widehat{F}_n^{(j)}(t) - \mathcal{K}_j(t)}{1+\mathcal{K}_0(t)} \le F^{(j)}(t) \le \frac{\widehat{F}_n^{(j)}(t) + \mathcal{K}_j(t)}{1-\mathcal{K}_0(t)}, t \le \sigma\right\}\right. \\
&\qquad\qquad \left.\bigcap\left\{1 - \frac{1-\widehat{F}_n(t)}{1-\mathcal{K}_0(t)} \le F(t) \le 1 - \frac{1-\widehat{F}_n(t)}{1+\mathcal{K}_0(t)}, t \le \sigma\right\}\right]
\end{aligned}$$

which is a part of the desired result.

We can repeat the methods for the construction of the Hall-Wellner and Aalen-Nair bands but with the processes $\widetilde{K}_n^{(j)}$ instead of the $K_n^{(j)}$. Instead of (3.13) and (3.14), we consider now the following probabilities:

$$\begin{aligned}
\widetilde{P}_n^{AN} &= \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\left|\widetilde{K}_n^{(j)}(t)\right| \le \lambda_j\sqrt{\widehat{C}_n^{(j)}(\sigma)}, t \le \sigma\right\}\right], \\
\widetilde{P}_n^{HW} &= \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\left|\widetilde{K}_n^{(j)}(t)\right| \le \lambda_j\left(1+\widehat{C}_n^{(j)}(t)\right), t \le \sigma\right\}\right].
\end{aligned}$$

As before, we get the bounds:

$$\begin{aligned}
\lim_{n\to\infty}\widetilde{P}_n^{AN} &\ge 1 - \sum_{j=1}^{\mathcal{J}}\left(1 - K_W(\lambda_j, 1)\right), \\
\lim_{n\to\infty}\widetilde{P}_n^{HW} &\ge 1 - \sum_{j=1}^{\mathcal{J}}\left(1 - K_B\left(\lambda_j, \frac{C^{(j)}(\sigma)}{1+C^{(j)}(\sigma)}\right)\right).
\end{aligned}$$

Here, $\widetilde{P}_n$ will denote either $\widetilde{P}_n^{AN}$ or $\widetilde{P}_n^{HW}$. The relation $\sum_{j=1}^{\mathcal{J}}\widetilde{K}_n^{(j)} = \widetilde{K}_n^{(0)}$ is satisfied. As

before, we may then write:

$$\begin{aligned}
\widetilde{P}_n &= \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\left|\widetilde{K}_n^{(j)}(t)\right| \leq \sqrt{n}\mathcal{K}_j(t), t \leq \sigma\right\}\bigcap\left\{\left|\widetilde{K}_n^{(0)}(t)\right| \leq \sqrt{n}\mathcal{K}_0(t), t \leq \sigma\right\}\right] \\
&= \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\widehat{F}_n^{(j)}(t)\frac{1-F(t)}{1-\widehat{F}_n(t)} - \mathcal{K}_j(t) \leq F^{(j)}(t) \leq \widehat{F}_n^{(j)}(t)\frac{1-F(t)}{1-\widehat{F}_n(t)} + \mathcal{K}_j(t), t \leq \sigma\right\}\right. \\
&\qquad \left.\bigcap\left\{1-\mathcal{K}_0(t) \leq \frac{1-F(t)}{1-\widehat{F}_n(t)} \leq 1+\mathcal{K}_0(t), t \leq \sigma\right\}\right] \\
&\leq \mathbb{P}\left[\bigcap_{j=1}^{\mathcal{J}}\left\{\widehat{F}_n^{(j)}(t)\left(1-\mathcal{K}_0(t)\right) - \mathcal{K}_j(t) \leq F^{(j)}(t) \leq \widehat{F}_n^{(j)}(t)\left(1+\mathcal{K}_0(t)\right) + \mathcal{K}_j(t), t \leq \sigma\right\}\right. \\
&\qquad \left.\bigcap\left\{1-\left(1-\widehat{F}_n(t)\right)\left(1+\mathcal{K}_0(t)\right) \leq F(t) \leq 1-\left(1-\widehat{F}_n(t)\right)\left(1-\mathcal{K}_0(t)\right), t \leq \sigma\right\}\right],
\end{aligned}$$

which ends the proof. □

Remark 4. *For n large enough, Csörgő and Horváth (1986) noticed that the inequalities $r_n^{(0)}(t) \leq \widetilde{r}_n^{(0)}(t)$ and $R_n^{(0)}(t) \leq \widetilde{R}_n^{(0)}(t)$ hold uniformly on $[0,\sigma]$. For $j = 1,\ldots,\mathcal{J}$, the inequalities $\widetilde{r}_n^{(j)}(t) \leq r_n^{(j)}(t)$ and $\widetilde{R}_n^{(j)}(t) \leq R_n^{(j)}(t)$ hold uniformly on $[0,\sigma]$. Consequently, following Csörgő and Horváth (1986), we can also consider mixed confidence bands.*

3.4 Simulations

In this section, we present a simulation study to illustrate, for the case of finite samples, the most significant results for the confidence bands proposed in Section 3.3. This section is organized as follows. First, in Subsection 3.4.1, we introduce the bivariate parametric model used to generate two dependent variables. Then, in Subsection 3.4.2, we expose the elements needed for the simulations (e.g. the reference parameters, the different proportions of censored observations considered,...). We discuss the results in Subsection 3.4.3. Finally, the main graphics are exposed in Section 3.5.

3.4.1 A parametric model

In this subsection, we introduce the bivariate distribution appropriate to generate data according to a model with two dependent competing risks.

In the context of modeling dependent lifetimes, the bivariate exponential model of Block and Basu (1974), exposed in Subsection 1.4.2, received much attention. But, as outlined in Subsection 1.4.2, this model has the bivariate loss of memory property. In the

context of lifetime analysis, this property is not realistic. It would be more relevant to generate data according to a bivariate Weibull distribution. Indeed, the Weibull distribution plays a crucial role in building models for various failure time distribution.

We introduce a bivariate Weibull distribution which is obtained from a power-transformation of the ACBVE model of Block and Basu (1974). A bivariate r.v. (X_1, X_2) is said to follow an absolutely continuous bivariate Weibull distribution with two positive parameters α and β and three non-negative parameters λ_0, λ_1 and λ_1 if the bivariate r.v. given by $(Y_1, Y_2) = (X_1^\alpha, X_2^\beta)$ follows an ACBVE(λ_0,λ_1,λ_2). We write for short that (X_1, X_2) follows an ACBVW(α,β,λ_0,λ_1,λ_2) distribution. Without loss of generality, we assume in the sequel that $\alpha \leq \beta$.

The density of the couple (X_1, X_2) is given for $x_1, x_2 \geq 0$ by:

$$h(x_1, x_2) = \begin{cases} \frac{\lambda_1\lambda(\lambda_2+\lambda_0)}{\lambda_1+\lambda_2}\alpha\beta x_1^{\alpha-1}x_2^{\beta-1}\exp\left(-\lambda_1 x_1^\alpha - (\lambda_2+\lambda_0)x_2^\beta\right) & \text{if } x_1^\alpha < x_2^\beta, \\ \frac{\lambda_2\lambda(\lambda_1+\lambda_0)}{\lambda_1+\lambda_2}\alpha\beta x_1^{\alpha-1}x_2^{\beta-1}\exp\left(-(\lambda_1+\lambda_0)x_1^\alpha - \lambda_2 x_2^\beta\right) & \text{if } x_1^\alpha \geq x_2^\beta. \end{cases}$$

We define

$$X = \min(X_1, X_2) \quad \text{and} \quad \mathcal{C} = \begin{cases} 1 & \text{if } X_1 \leq X_2, \\ 2 & \text{otherwise.} \end{cases}$$

The cause-specific distribution functions are given for $x \geq 0$ by:

$$F^{(1)}(x) = \mathbb{P}\left[X \leq x, \mathcal{C} = 1\right] = \mathbb{P}\left[X_1 \leq x, X_1 \leq X_2\right],$$

and

$$F^{(2)}(x) = \mathbb{P}\left[X \leq x, \mathcal{C} = 2\right] = \mathbb{P}\left[X_2 \leq x, X_2 \leq X_1\right].$$

Then, for $x \geq 0$, we have:

$$F^{(1)}(x) = \begin{cases} \frac{\lambda(\lambda_1+\lambda_0)}{\lambda_1+\lambda_2}\alpha\int_0^x u^{\alpha-1}\exp\left(-(\lambda_1+\lambda_0)u^\alpha - \lambda_2 u^\beta\right)du \\ +\frac{\lambda_0}{\lambda_1+\lambda_2}\left(\exp\left(-\lambda x^\alpha\right) - 1\right) & \text{if } 0 \leq x \leq 1, \\ F^{(1)}(1) + \frac{\lambda\lambda_1}{\lambda_1+\lambda_2}\alpha\int_1^x u^{\alpha-1}\exp\left(-\lambda_1 u^\alpha - (\lambda_2+\lambda_0)u^\beta\right)du & \text{if } 1 < x, \end{cases}$$

and

$$F^{(2)}(x) = \begin{cases} \frac{\lambda\lambda_2}{\lambda_1+\lambda_2}\beta\int_0^x u^{\beta-1}\exp\left(-(\lambda_1+\lambda_0)u^\alpha - \lambda_2 u^\beta\right)du & \text{if } 0 \leq x \leq 1, \\ F^{(2)}(1) + \frac{\lambda(\lambda_2+\lambda_0)}{\lambda_1+\lambda_2}\beta\int_1^x u^{\beta-1}\exp\left(-\lambda_1 u^\alpha - (\lambda_2+\lambda_0)u^\beta\right)du \\ +\frac{\lambda_0}{\lambda_1+\lambda_2}\left(\exp(-\lambda x^\beta) - \exp(-\lambda)\right) & \text{if } 1 < x, \end{cases}$$

The distribution function F is given for $x \geq 0$ by $F(x) = \mathbb{P}[X \leq x] = F^{(1)}(x) + F^{(2)}(x)$. This leads to the following expression for $x \geq 0$:

$$F(x) = \begin{cases} 1 + \frac{\lambda_0}{\lambda_1+\lambda_2} \exp\left(-\lambda x^\alpha\right) - \frac{\lambda}{\lambda_1+\lambda_2} \exp\left(-(\lambda_1+\lambda_0)x^\alpha - \lambda_2 x^\beta\right) & \text{if } 0 \leq x \leq 1, \\ 1 + \frac{\lambda_0}{\lambda_1+\lambda_2} \exp\left(-\lambda x^\beta\right) - \frac{\lambda}{\lambda_1+\lambda_2} \exp\left(-\lambda_1 x^\alpha - (\lambda_2+\lambda_0)x^\beta\right) & \text{if } 1 < x. \end{cases} \tag{3.15}$$

The corresponding probability density function is given, for $x \geq 0$, by:

$$f(x) = \begin{cases} \frac{\lambda}{\lambda_1+\lambda_2}\left(\alpha(\lambda_0+\lambda_1)x^{\alpha-1} + \beta\lambda_2 x^{\beta-1}\right) \exp\left(-(\lambda_0+\lambda_1)x^\alpha - \lambda_2 x^\beta\right) \\ -\frac{\lambda\lambda_0}{\lambda_1+\lambda_2}\alpha x^{\alpha-1}\exp\left(-\lambda x^\alpha\right) & \text{if } 0 \leq x \leq 1, \\ \frac{\lambda}{\lambda_1+\lambda_2}\left(\alpha\lambda_1 x^{\alpha-1} + \beta(\lambda_0+\lambda_2)x^{\beta-1}\right) \exp\left(-\lambda_1 x^\alpha - (\lambda_0+\lambda_2)x^\beta\right) \\ -\frac{\lambda\lambda_0}{\lambda_1+\lambda_2}\beta x^{\beta-1}\exp\left(-\lambda x^\beta\right) & \text{if } 1 < x. \end{cases}$$

The ACBVW model does not have proportional cause-specific distribution functions unless $\alpha = \beta$ in which case, we get for $x \geq 0$:

$$F^{(1)}(x) = \frac{\lambda_1}{\lambda_1+\lambda_2}\left(1 - \exp\left(-\lambda x^\alpha\right)\right),$$

and

$$F^{(2)}(x) = \frac{\lambda_2}{\lambda_1+\lambda_2}\left(1 - \exp\left(-\lambda x^\alpha\right)\right).$$

The ACBVW model has the following properties.
1. The distribution is absolutely continuous.
2. It has no lack of memory property (neither in marginal nor jointly) unless $\alpha = \beta = 1$.
3. The two competing risks are dependent unless $\lambda_0 = 0$.
4. The cause-specific distribution functions of the two competing risks are not proportional unless $\alpha = \beta$.

Examples of the cause-specific distribution functions are drawn in Subsection 3.5.1 for different values of the parameters. The value of α and β is seen to affect the concavity of the curves while the value of λ_0, λ_1 and λ_2 is seen to affect the flattening of the curves.

3.4.2 Design of the simulations

We deal here with the design of the simulation. First, we choose the reference distributions. We consider a model with $\mathcal{J} = 2$ competing risks. To generate $(X, \mathcal{C})$, we use the ACBVW model so that the two competing risks are neither independent nor proportional. For that, we set the parameters of the ACBVW model as $\alpha = 1.5$, $\beta = 3$, $\lambda_0 = 0.1$,

$\lambda_1 = 0.05$, $\lambda_2 = 0.01$. With this choice, the theoretical proportion of observations of X due to cause 1 is $F^{(1)}(\infty) = 0.3262$ while the theoretical proportion of observations of X due to cause 2 is $F^{(2)}(\infty) = 0.6738$. Since the censoring time and the lifetime are independent, the theoretical proportion of censoring can be computed from the expression:

$$\begin{aligned}\mathbb{P}[J=0] &= \mathbb{P}[X > C] \\ &= \iint I(x > c) dF(x) dG(c) \\ &= \int G^{-}(x) f(x)\, dx\,.\end{aligned}$$

We take the censoring mechanism as an exponential variable with parameter a. The theoretical proportion of censoring is then given by:

$$\mathbb{P}[J=0] = \int_0^\infty (1 - \exp(-ax))\, f(x) dx\,. \tag{3.16}$$

The value $a = 0.055$ (resp. 0.155, resp 0.385) yields a theoretical proportion of censored observations equal to 10% (resp. 25%, resp. 50%).

In order to avoid the usual problem of variance explosion at the end of the observation interval, the endpoint of the bands σ is chosen equal to the 90-th or 80-th observed percentile.

To obtain one realization of the sample $(T_i, J_i)_{i=1,\dots,n}$, we do as follows. For $i = 1, \dots, n$, we generate $(X_{1,i}, X_{2,i})$ according to an ACBVW(1.5,3,0.1,0.05,0.01) model. For that, making use of Proposition 3, we generate independently a r.v. Z_i according to an $\mathcal{E}$xp(1.6) law and a r.v. δ_i with distribution function φ given in (1.20) and set:

$$\begin{aligned}X_{1,i} &= (Z_i I(\delta_i < 0) + (Z_i + \delta_i) I(\delta_i \geq 0))^{1/\alpha}\,, \\ X_{2,i} &= ((Z_i - \delta_i) I(\delta_i < 0) + Z_i I(\delta_i \geq 0))^{1/\beta}\,.\end{aligned}$$

Then, we get:

$$X_i = \min(X_{1,i}, X_{2,i}) \quad \text{and} \quad \mathcal{C}_i = \begin{cases} 1 & \text{if } X_{1,i} \leq X_{2,i}, \\ 2 & \text{otherwise.}\end{cases}$$

For $i = 1, \dots, n$, we generate C_i according to a $\mathcal{E}$xp(a) law, then we determine the observed lifetimes

$$T_i = \min(X_i, C_i),\ J_i = \mathcal{C}_i I(X_i \leq C_i).$$

We compute the different useful estimators. Let $(T_{1,n}, \dots, T_{n,n})$ denote the order statistics of the sample $(T_1, \dots, T_n)$. Let $(J_{1,n}, \dots, J_{n,n})$ be the concomitant statistics of $(J_1, \dots, J_n)$, i.e. if $T_{i,n} = T_j$ then $J_{i,n} = J_j$. The following algorithm is used to compute the product-limit estimators. Under the initial conditions $\widehat{G}_n(0) = 0$ and $\widehat{F}_n^{(j)}(0) = 0$, for

$j = 1, \ldots, \mathcal{J}$, we have for $i = 1, \ldots, n$:

$$\begin{aligned}
\widehat{F}_n(T_{i,n}) &= \widehat{F}_n(T_{i-1,n}) + \frac{I(J_{i,n} \neq 0)}{n\left(1 - \widehat{G}_n(T_{i-1,n})\right)}, \\
\widehat{G}_n(T_{i,n}) &= \widehat{G}_n(T_{i-1,n}) + \frac{I(J_{i,n} = 0)}{n\left(1 - \widehat{F}_n(T_{i-1,n})\right)}, \\
\widehat{F}_n^{(j)}(T_{i,n}) &= \widehat{F}_n^{(j)}(T_{i-1,n}) + \frac{I(J_{i,n} = j)}{n\left(1 - \widehat{G}_n(T_{i-1,n})\right)}.
\end{aligned}$$

We choose 0.95% of confidence level and determine the confidence bands.

In order to generate Monte-Carlo confidence bands, we generate $N = 1000$ realizations of the sample $(T_i, J_i)_{i=1,\ldots,n}$ for a given scenario. We compute the pertaining confidence bands. Then we draw the mean confidence bands computed from the 1000 obtained confidence bands.

The simulation study has been implemented in `R` and the main functions are listed and commented in Appendix.

3.4.3 Results and discussion

The first point is that the proposed bands are asymptotic. Consequently, they should not be used for small sample sizes as $n = 50$. For $n = 100$, the bands are quite crude. The bands computed with $n = 200$, 500 and 1000 illustrate the convergence of the bands as n goes to infinity. The four types of bands are seen to contain the true distribution function $F^{(1)}$, $F^{(2)}$ and F respectively. But, in the case of $F^{(1)}$ and $F^{(2)}$, they are quite wide especially in the neighborhood of σ. This comes from the fact that the Kaplan-Meier estimator grows to one as t grows to ∞ while the Aalen-Johansen estimator grows to an unknown proportion.

As noticed in Remark 4, the non-modified bands for $F^{(1)}$ and $F^{(2)}$ are uniformly above the modified bands while the modified bands for F are uniformly above the non-modified bands. The Hall-Wellner and the modified Hall-Wellner bands are more narrow than the Aalen-Nair and the modified Aalen-Nair bands for values of t distant of σ. But the Hall-Wellner and the modified Hall-Wellner bands have a trumpet form so that for values of t closed to σ, they are wider than the Aalen-Nair and the modified Aalen-Nair bands.

It appears that the width of the bands is affected the value of σ. Namely, the four types of bands widen as σ is closed to the last order statistic $T_{n,n}$. It turns out that, for $j = 1, \ldots, \mathcal{J}$, the value of the variance of $\widehat{F}_n^{(j)}$ in σ which is $\widehat{C}_n^{(j)}(\sigma)$ grows very large as σ goes to the last order statistic and this affects the width of the whole band whatever the type.

Finally, notice that the proportion of censored observations affects the accuracy of the bands in a surprising way. It turns out that the bands widen as the proportion of censored observations decreases. This is explained by the fact that the censoring mechanism favors the observation of small values with the consequence that the 90-th, resp. 80-th, observed percentile (up to which the bands are drawn) gets smaller when the proportion of censored observations increases. Hence the number of data points used to make inference remains the same but the data points are located in a shorter time interval (that goes from 0 to the 90-th, resp 80-th, observed percentile) when the proportion of censored observations increases. The fact that the 90-th, resp. 80-th, observed percentile gets smaller when the proportion of censored observations increases can be shown analytically. For that, we use the two following facts. On the one hand, we make use of the law of the k-th order statistic of the sample $(T_1, \ldots, T_n)$ which may be found e.g. in Capéraà and Van Cutsem (1988). We have for $t \geq 0$:

$$\mathbb{P}[T_{k,n} \leq t] = \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} H(t)^i (1 - H(t))^{n-i}.$$

In our set-up, the d.f. H is given by the relation $H = 1 - (1 - F)(1 - G)$ where G is the exponential distribution with parameter a. This leads to:

$$\begin{aligned}
\mathbb{P}[T_{k,n} \leq t] &= \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} \left(1 - (1 - F(t))(1 - G(t))\right)^i (1 - F(t))^{n-i} (1 - G(t))^{n-i} \\
&= \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} \left(1 - (1 - F(t))e^{-at}\right)^i (1 - F(t))^{n-i} e^{(n-i)at}.
\end{aligned}$$

Consider a sample $(T'_1, \ldots, T'_n)$ distributed according to a distribution $1 - (1 - F)(1 - G')$ where G' is an exponential distribution with parameter a' and let $T'_{k,n}$ be the k-th order statistic of this sample. From the above formula, we see that if $a' \geq a$, then we have $\mathbb{P}[T_{k,n} \leq t] \leq \mathbb{P}[T'_{k,n} \leq t]$ for $t \geq 0$. On the other hand, we see from (3.16) that the proportion of censored observations is greater in the sample $(T'_1, \ldots, T'_n)$ than in the sample $(T_1, \ldots, T_n)$ if $a' \geq a$. This leads to the desired result.

3.5 Graphical displays

3.5.1 ACBVW examples

On Figures 3.1, 3.2 and 3.3, we display some examples of the cause-specific distribution functions obtained from an ACBVW model with different values of the parameters. The curve pertaining to F is drawn in black, the curve pertaining to $F^{(1)}$ is drawn in red while the curve pertaining to $F^{(2)}$ is drawn in blue.

3.5.2 Simulated confidence bands

On Figures 3.4, 3.5 and 3.6, we trace the Monte-Carlo confidence bands for the sample size $n = 200$ with a theoretical proportion of censored observations equal respectively to 10%, 25% and 50%. The value of σ is chosen equal to the 90-th observed percentile.

On Figures 3.8, 3.9 and 3.10, we trace the Monte-Carlo confidence bands for the sample size $n = 500$ with a theoretical proportion of censored observations equal respectively to 10%, 25% and 50%. The value of σ is chosen equal to the 90-th observed percentile.

On Figures 3.12, 3.13 and 3.14, we trace the Monte-Carlo confidence bands for the sample size $n = 1000$ with a theoretical proportion of censored observations equal respectively to 10%, 25% and 50%. The value of σ is chosen equal to the 90-th observed percentile.

On Figures 3.7, 3.11 and 3.15, we trace the Monte-Carlo confidence bands for $n = 200$, $n = 500$ and $n = 1000$ with a theoretical proportion of censored observations equal respectively to 25%. The value of σ is equal to the 80-th observed percentile.

The theoretical distribution is drawn in black. The Aalen-Nair bands are drawn in red. The modified Aalen-Nair bands are drawn in blue. The Hall-Wellner bands are drawn in purple. The modified Hall-Wellner bands are drawn in light green.

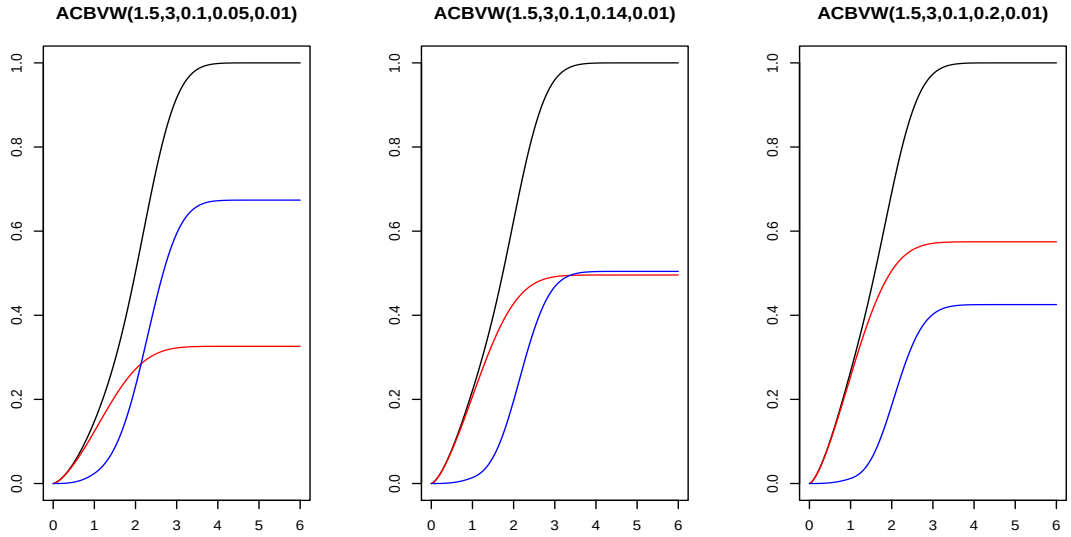

Figure 3.1: Representation of $F^{(1)}$, $F^{(2)}$ and F generated from an ACBVW model.

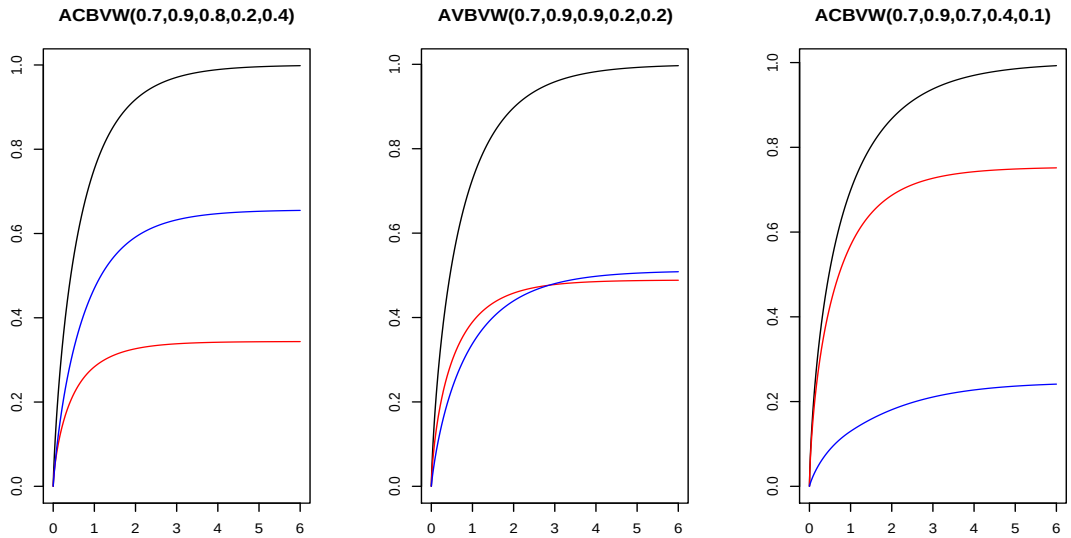

Figure 3.2: Representation of $F^{(1)}$, $F^{(2)}$ and F generated from an ACBVW model (continued).

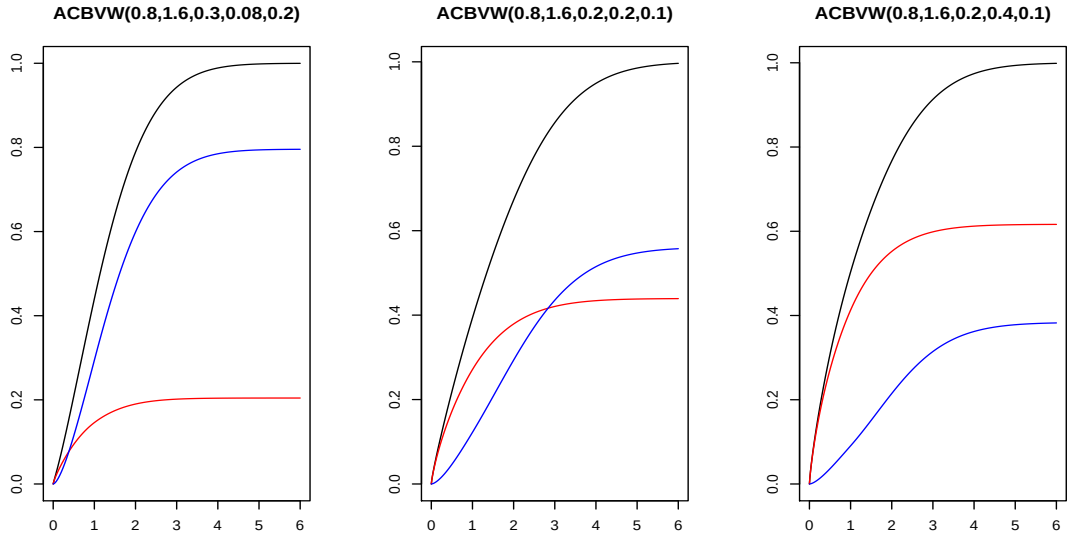

Figure 3.3: Representation of $F^{(1)}$, $F^{(2)}$ and F generated from an ACBVW model (continued).

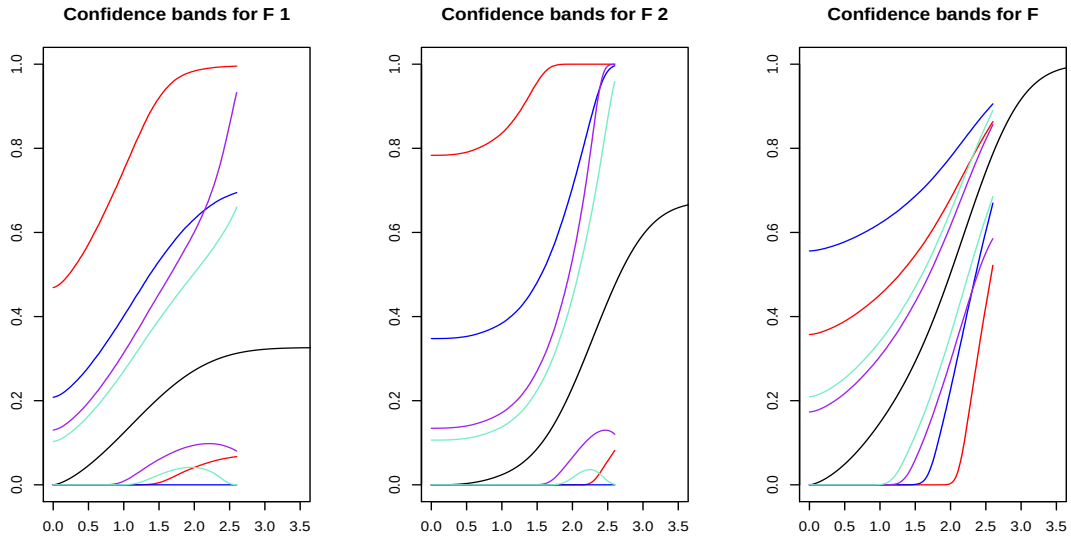

Figure 3.4: Confidence bands with n=200, 10% of censored observations and σ equal to the 90-th observed percentile.

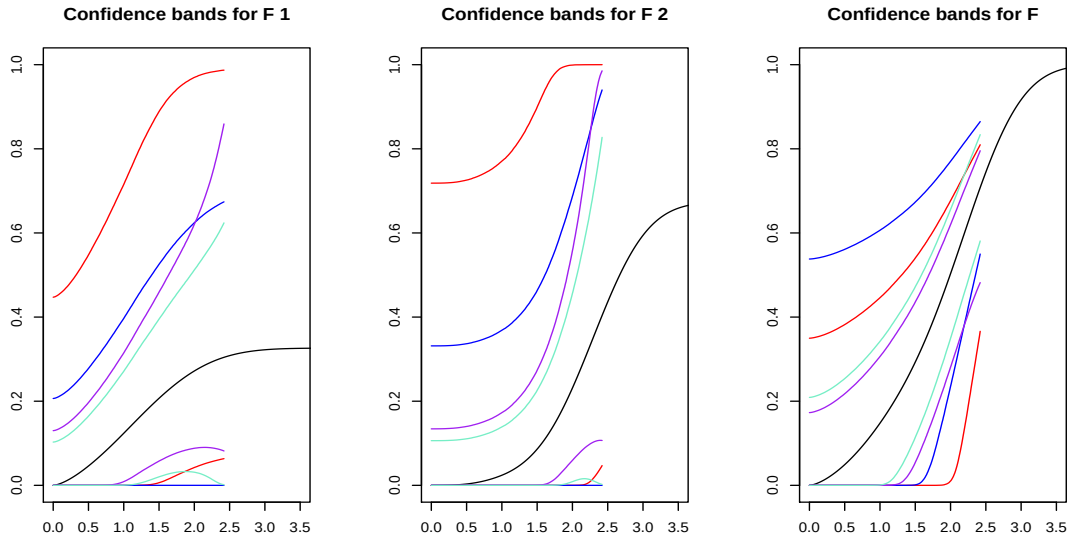

Figure 3.5: Confidence bands with n=200, 25% of censored observations and σ equal to the 90-th observed percentile.

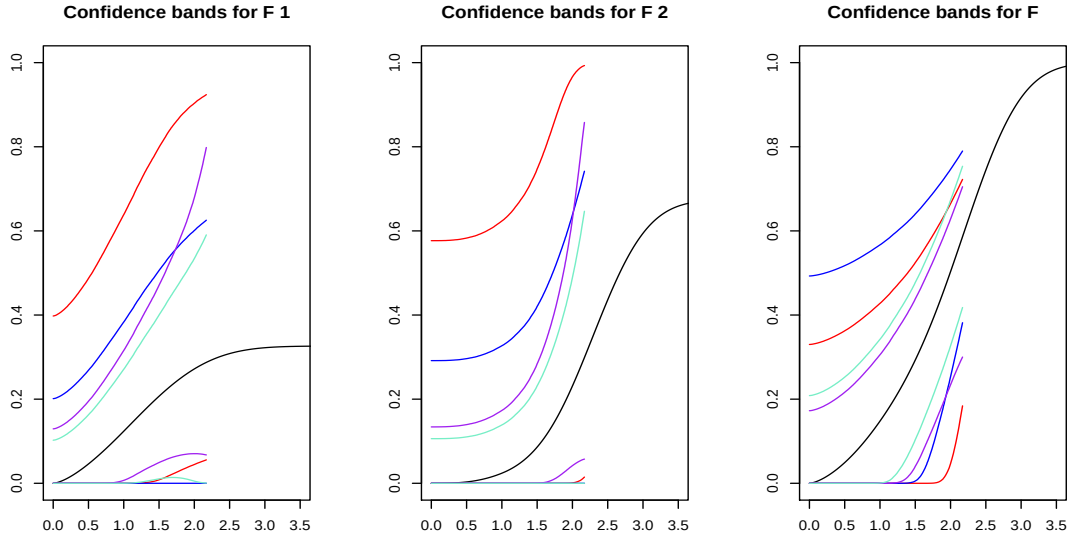

Figure 3.6: Confidence bands with n=200, 50% of censored observations and σ equal to the 90-th observed percentile.

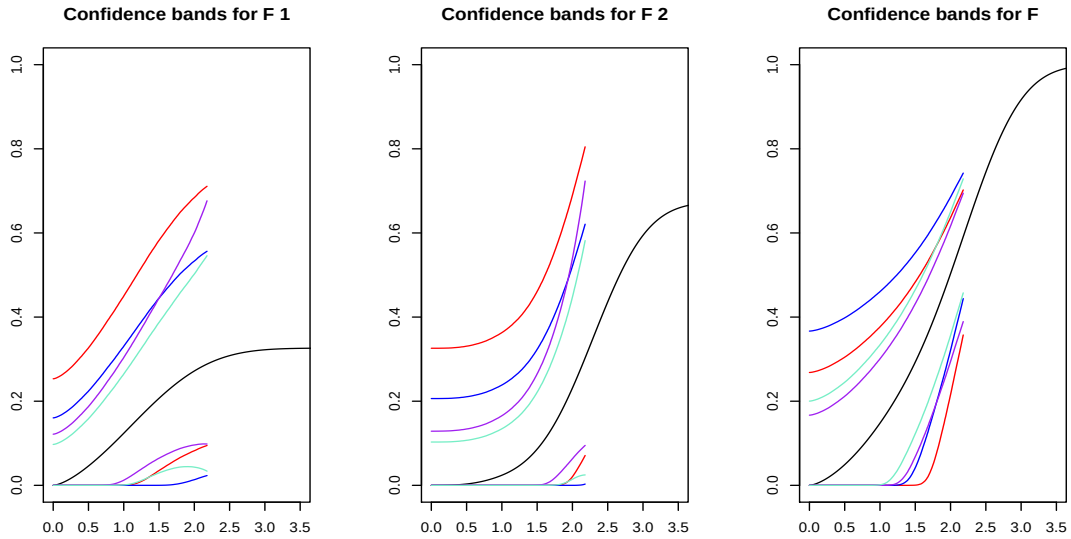

Figure 3.7: Confidence bands with n=200, 25% of censored observations and σ equal to the 80-th observed percentile.

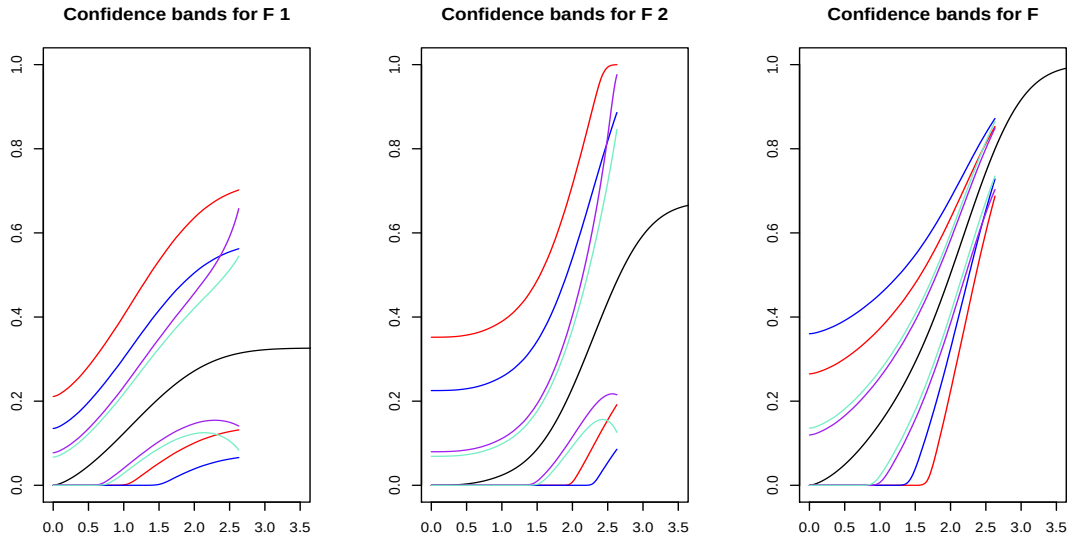

Figure 3.8: Confidence bands with n=500, 10% of censored observations and σ equal to the 90-th observed percentile.

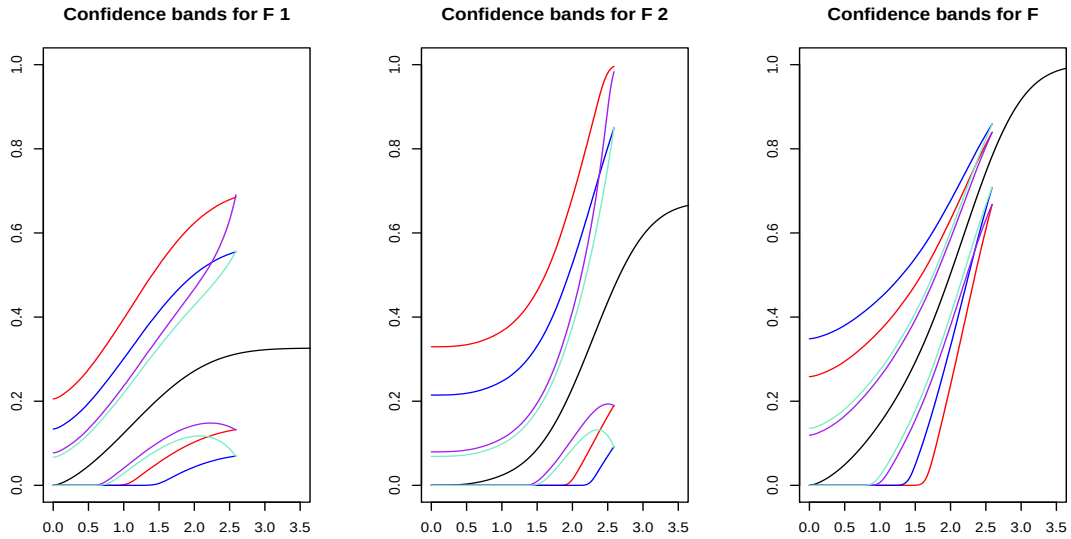

Figure 3.9: Confidence bands with n=500, 25% of censored observations and σ equal to the 90-th observed percentile.

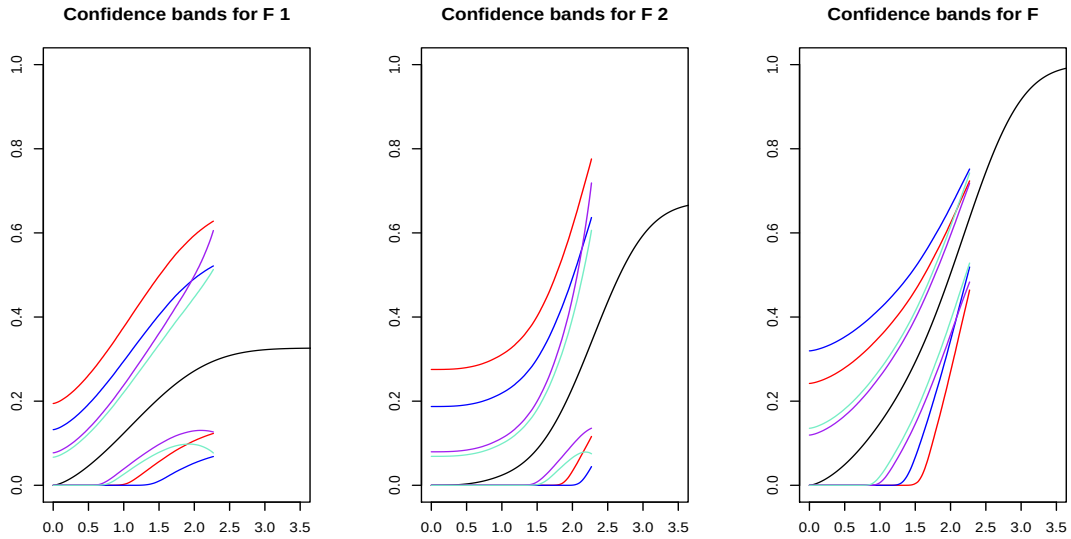

Figure 3.10: Confidence bands with n=500, 50% of censored observations and σ equal to the 90-th observed percentile.

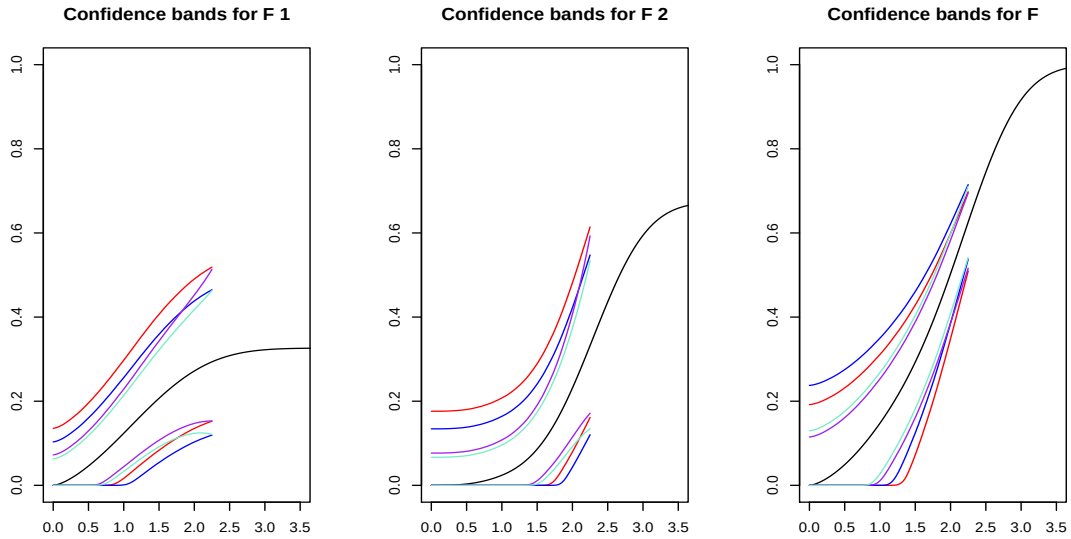

Figure 3.11: Confidence bands with n=500, 25% of censored observations and σ equal to the 80-th observed percentile.

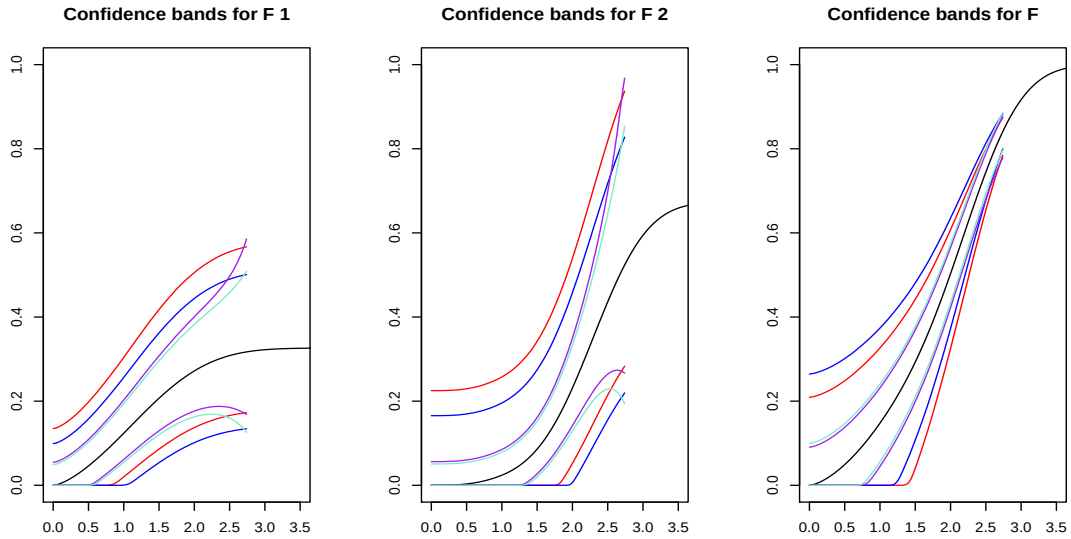

Figure 3.12: Confidence bands with n=1000, 10% of censored observations and σ equal to the 90-th observed percentile.

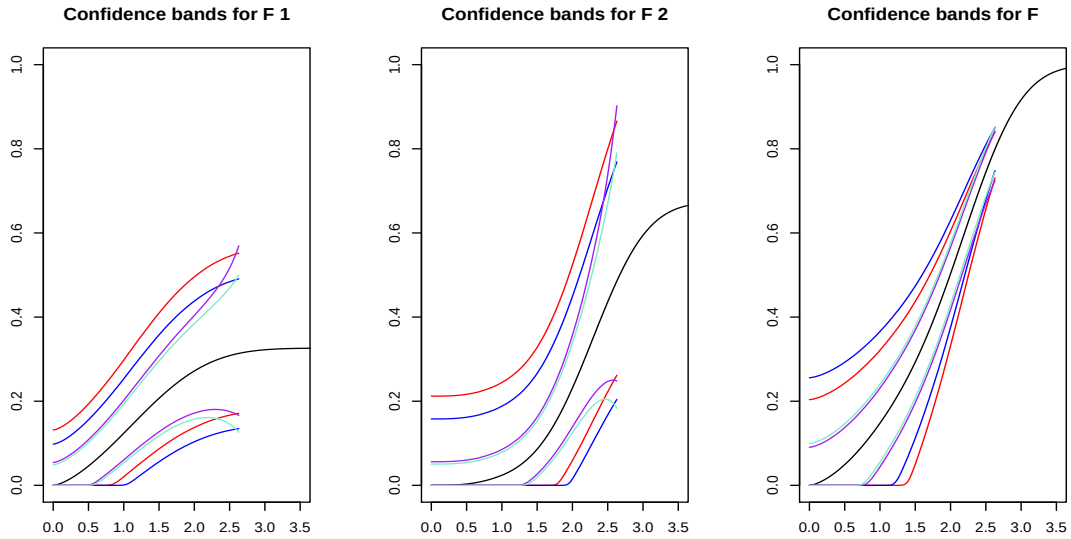

Figure 3.13: Confidence bands with n=1000, 25% of censored observations and σ equal to the 90-th observed percentile.

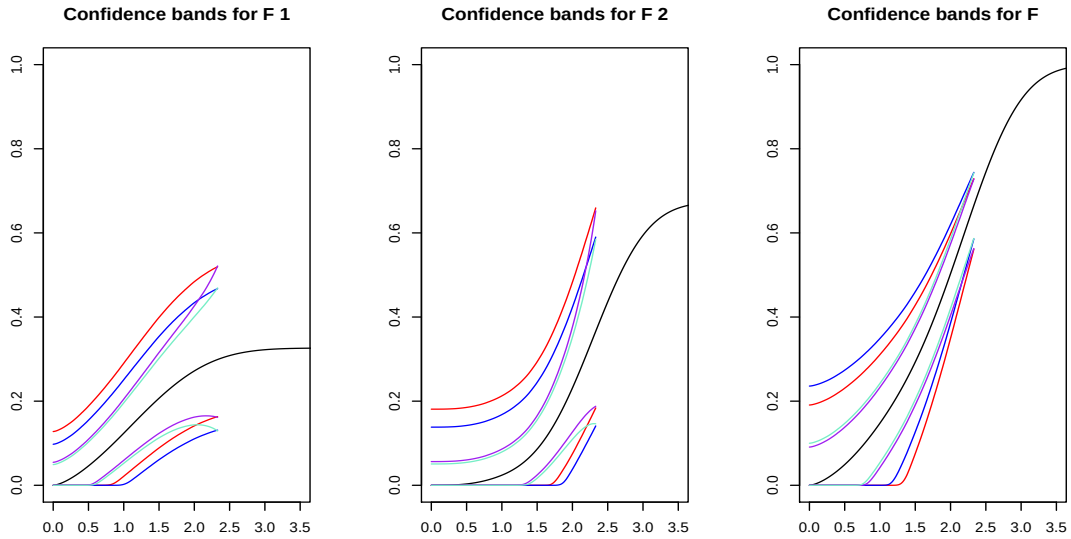

Figure 3.14: Confidence bands with n=1000, 50% of censored observations and σ equal to the 90-th observed percentile.

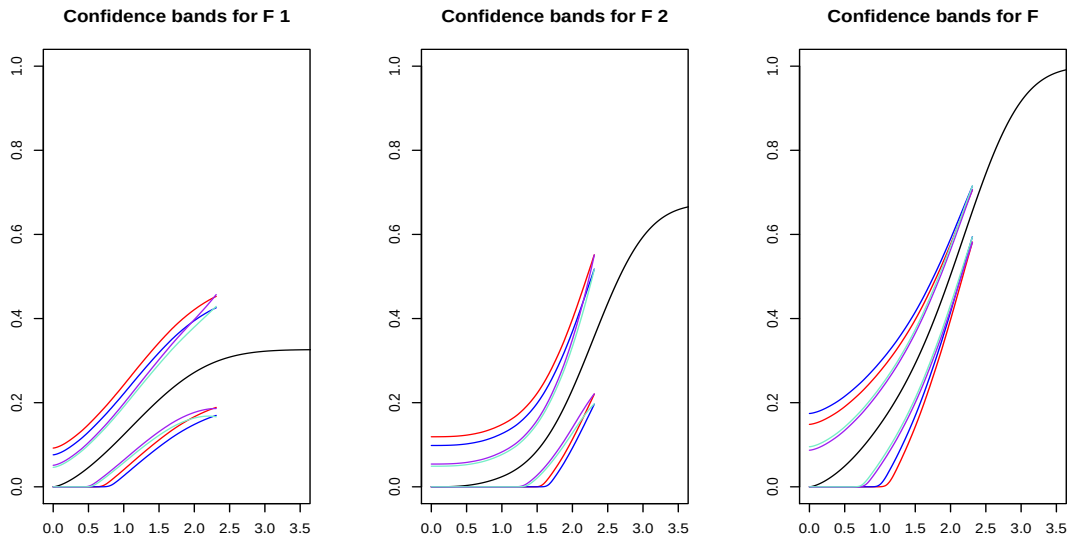

Figure 3.15: Confidence bands with n=1000, 25% of censored observations and σ equal to the 80-th observed percentile.

Chapter 4

Proportional subdistribution functions

4.1 Introduction

In the setup of independent right-censoring, a simple model of informative censoring is the Koziol-Green model. Let X_1 be a non-negative r.v. with distribution function F_1 that stands for the failure time of an individual. Let X_2 be a non-negative r.v. with distribution function F_2 that stands for the censoring time of an individual. Instead of the failure time X_1, we observe

$$\big(X = \min(X_1, X_2), I(X_1 \leq X_2)\big) .$$

We assume that X_1 and X_2 are independent. The r.v. X has distribution function F given by $1 - F = (1 - F_1)(1 - F_2)$.

The Koziol-Green model is defined by the existence of a positive constant β such that for $t \geq 0$:

$$\int_0^t \frac{dF_2}{1 - F_2^-} = \beta \int_0^t \frac{dF_1}{1 - F_1^-} .$$

This proportional hazards condition is equivalent to the existence of a constant $\beta > 0$ such that $1 - F_2 = (1 - F_1)^\beta$. In the Koziol-Green model, Cheng and Lin (1987) have pointed out that the maximum likelihood estimator of the lifetime distribution function F_1, denoted by $\widetilde{F}_{1,n}$, is of a much simpler form than the usual product-limit estimator $\widehat{F}_{1,n}$. Indeed, the relation $1 - F_1 = (1 - F)^\alpha$ with $\alpha = 1/(1 + \beta) = \mathbb{P}[X_1 \leq X_2]$ allows easy calculations. Moreover, the estimator $\widetilde{F}_{1,n}$ is asymptotically more efficient than $\widehat{F}_{1,n}$. Since this model is easily interpretable and intuitively appealing, it has been popular for theoretical purposes. We refer to Csörgő and Horváth (1981, 1983b), Csörgő (1988), Stute (1992), Dikta (1995) and Ghorai and Schmitter (1999) for an extensive review of the properties of the estimator $\widetilde{F}_{1,n}$.

Gather and Pawlitschko (1998) proposed a generalization of the Koziol-Green model where only part of the censored observations are supposed to be informatively censored. For $j = 1, 2$, we introduce independent non-negative r.v. X_j with distribution function F_j

that stands for the failure time of an individual from the j-th cause. Let C be the nonnegative r.v. with distribution function G that stands for the censoring time of an individual. We assume that C is independent from both X_1 and X_2. We only observe

$$T = \min(X_1, X_2, C) \text{ and } \delta = \begin{cases} 0 \text{ if } C \leq X_1 \wedge X_2 \\ 1 \text{ if } X_1 \leq X_1 \wedge C \\ 2 \text{ if } X_2 \leq C \wedge X_2 \end{cases}$$

where T has distribution function H given by $1 - H = (1 - F_1)(1 - F_2)(1 - G)$. The r.v. $\min(X_1, X_2)$ has distribution function F given by $1 - F = (1 - F_1)(1 - F_2)$. Gather and Pawlitschko (1998) assumed the existence of positive constants α_j for $j = 1, 2$ such that:

$$1 - F_j = (1 - F)^{\alpha_j} .$$

Their assumption turns out to be equivalent to assume the proportionality of the net cumulative hazard functions Λ_j for $j = 1, 2$ defined in (1.12) and the overall hazard function Λ pertaining to the r.v. $\min(X_1, X_2)$:

$$\Lambda_j = \alpha_j \Lambda.$$

Under independent right-censoring, Gather and Pawlitschko (1998) proposed a semi-parametric estimator of F_j and investigated its asymptotic properties while Geffray and Guilloux (2005) proposed an estimator for Λ_j and derived its asymptotic properties.

In this chapter, we consider failure times from different causes which are not independent but which have proportional cause-specific cumulative hazards, in the presence of independent right-censoring. For $j = 1, \ldots, \mathcal{J}$, we introduce non-independent non-negative r.v. X_j that stands for the failure time of an individual from the j-th cause. Let $X = \min(X_1, \ldots, X_{\mathcal{J}})$ be the r.v. that stands for the overall failure time of the individual with distribution function F. Let $\mathcal{C}$ indicate the cause of failure i.e. $\mathcal{C}$ takes value j if the failure is due to cause j for a j in $\{1, \ldots, \mathcal{J}\}$ i.e. if $X = X_j$. The cause-specific distribution function pertaining to the j-th risk is given for $t \geq 0$ by:

$$F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j].$$

We assume that X is at risk of being independently right-censored by a non-negative r.v. C with distribution function G. Consequently, the observable r.v. are:

$$\big(T = \min(X, C),\ J = \mathcal{C}\, I(X \leq C)\big) .$$

Since X and C are independent, the r.v. T has distribution function H given by the relation $1 - H = (1 - F)(1 - G)$. We assume the proportionality of the cause-specific proportional hazards defined in (1.15). Namely, we assume that there exists a $\beta_{1,j} > 0$ for $j = 1, \ldots, \mathcal{J}$ such that:

$$\Lambda^{(1,j)} = \beta_{1,j}\, \Lambda^{(1,1)} . \tag{4.1}$$

Of course, $\beta_{1,1} = 1$. This assumption is equivalent to the assumption of proportionality of the cause-specific distribution functions i.e. for $j = 1, \ldots, \mathcal{J}$ there exists a $\beta_{1,j} > 0$ such that:

$$F^{(j)} = \beta_{1,j}\, F^{(1)}\,. \tag{4.2}$$

To avoid confusion with Cox's proportional hazards model, this model is referred to as the proportional subdistribution functions model. Dauxois and Kirmani (2004) developed an adequacy test for this model.

In this set-up, we introduce an alternative estimator $\widetilde{F}_n^{(j)}$ of the subdistribution function $F^{(j)}$ for $j = 1, \ldots, \mathcal{J}$ and study its asymptotic properties. The chapter is organized as follows. In Section 4.2, we introduce our estimator and show its consistency uniformly on a maximal time-interval. In Section 4.3, we give a law-of-the-iterated-logarithm type result valid up to the $(n-k_n)$-th order statistic. Section 4.5 deals with the joint weak convergence of the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ for $j = 1, \ldots, \mathcal{J}$, on a compact time-interval. Finally, in Section 4.4, we establish a joint strong approximation for the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ for $j = 1, \ldots, \mathcal{J}$ valid up to the $(n - k_n)$-th order statistic.

4.2 The estimator and its strong uniform consistency

Assumption (4.2) together with the fact that $F = \sum_{j=1}^{\mathcal{J}} F^{(j)}$ entail that

$$F = F^{(1)} \sum_{j=1}^{\mathcal{J}} \beta_{1,j}\,.$$

Consequently, letting

$$\alpha_{1,j} = \frac{\beta_{1,j}}{1 + \beta_{1,2} + \cdots + \beta_{1,\mathcal{J}}}\,, \tag{4.3}$$

the subdistribution functions $F^{(j)}$ for $j = 1, \ldots, \mathcal{J}$ are proportional to F:

$$F^{(j)} = \alpha_{1,j} F\,. \tag{4.4}$$

The relation (4.4) suggests to propose a semiparametric estimator for $F^{(j)}$ based on an estimator for $\alpha_{1,j}$ and an estimator for F. In order to estimate $\alpha_{1,j}$, some additional work is needed.

In view of (4.2), the quantity $\beta_{1,j}$ may be written for $j = 1, \ldots, \mathcal{J}$ as:

$$\beta_{1,j} = \frac{\mathbb{P}[\mathcal{C} = j]}{\mathbb{P}[\mathcal{C} = 1]}.$$

This fact combined with Equality (4.3) entails that:

$$\alpha_{1,j} = \frac{\mathbb{P}[\mathcal{C} = j]}{\sum_{k=1}^{\mathcal{J}} \mathbb{P}[\mathcal{C} = k]} = \mathbb{P}[\mathcal{C} = j]\,.$$

This means that $\alpha_{1,j}$ is the theoretical proportion of observations of X which are due to cause j. As this quantity is not directly estimable (the implied variables are not observable), we find another expression for $\alpha_{1,j}$. For that, notice that:

$$\begin{aligned}
\mathbb{P}\left[J = j\right] &= \mathbb{P}\left[X \leq C, \mathcal{C} = j\right] \\
&= \iint I(x \leq c) dF^{(j)}(x) dG(c) \\
&= \int_0^\infty (1 - G^-) dF^{(j)} \\
&= \alpha_{1,j} \int_0^\infty (1 - G^-) dF.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{P}\left[J \neq 0\right] &= \mathbb{P}\left[X \leq C\right] \\
&= \iint I(x \leq c) dF(x) dG(c) \\
&= \int_0^\infty (1 - G^-) dF.
\end{aligned}$$

Consequently, $\alpha_{1,j}$ can be rewritten:

$$\alpha_{1,j} = \frac{\mathbb{P}[J = j]}{\mathbb{P}[J \neq 0]}. \tag{4.5}$$

We let $p^{(1,j)} = \mathbb{P}[J = j]$ and $p^{(1)} = \mathbb{P}[J \neq 0]$. Throughout the sequel, we will assume that $p^{(1,j)} \neq 0$ for $j = 1, \ldots, \mathcal{J}$. As $p^{(1)} = \sum_{j=1}^{\mathcal{J}} p^{(1,j)}$, this entails that $p^{(1)} \neq 0$.

Let $(T_i, J_i)_{i=1,\ldots,n}$ be n independent copies of the random vector (T, J) and let $T_{1,n} \leq T_{2,n} \leq \cdots \leq T_{n,n}$ be the order statistics associated with the sample $T_1, \ldots, T_n$.

The probabilities $p^{(1)}$ and $p^{(1,j)}$ for $j = 1, \ldots, \mathcal{J}$ may be estimated by their empirical counterparts defined by:

$$p_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} I(J_i \neq 0)\,,$$

and for $j = 1, \ldots, \mathcal{J}$

$$p_n^{(1,j)} = \frac{1}{n} \sum_{i=1}^{n} I(J_i = j)\,.$$

We then suggest the following estimator for $\alpha_{1,j}$:

$$\widehat{\alpha}_{1,j} = \frac{p_n^{(1,j)}}{p_n^{(1)}}, \tag{4.6}$$

which is the empirical analogue of (4.5).

Consequently, we propose the following estimator for $F^{(j)}$ defined for $t \geq 0$ by:

$$\widetilde{F}_n^{(j)}(t) = \widehat{\alpha}_{1,j}\widehat{F}_n(t) \tag{4.7}$$

where $\widehat{F}_n$ is the Kaplan-Meier estimator of F written in (1.16).

Our first result consists in the strong consistency of the estimator uniformly on a maximal-time interval.

Theorem 20. *Let $\mathcal{I} = \{t : H(t^-) < 1\}$. As n goes to ∞, we have almost-surely:*

$$\sup_{t\in\mathcal{I}} \left|\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right| \to 0\,.$$

Lemma 10. *The following convergence holds almost-surely as n goes to ∞.*

$$\widehat{\alpha}_{1,j} \to \alpha_{1,j}\,.$$

Proof. The proof relies on the strong law of large numbers applied separately to the numerator and to the denominator of (4.6). Namely,

$$\widehat{\alpha}_{1,j} = \frac{\frac{1}{n}\sum_{i=1}^n I(J_i = j)}{\frac{1}{n}\sum_{i=1}^n I(J_i \neq 0)} \to \frac{p^{(1,j)}}{p^{(1)}} = \alpha_{1,j}\,. \quad \square \tag{4.8}$$

Proof of Theorem 20. We carry out the following decomposition for $t > 0$

$$\begin{aligned}\widetilde{F}_n^{(j)}(t) - F^{(j)}(t) =& \widehat{\alpha}_{1,j}\widehat{F}_n(t) - \alpha_{1,j}F(t)\\ =& \alpha_{1,j}\left(\widehat{F}_n(t) - F(t)\right) + \left(\widehat{\alpha}_{1,j} - \alpha_{1,j}\right)F(t)\\ &+ \left(\widehat{\alpha}_{1,j} - \alpha_{1,j}\right)\left(\widehat{F}_n(t) - F(t)\right).\end{aligned}$$

Theorem 2 and Lemma 10 lead to the desired result. $\square$

4.3 A LIL type result

Theorem 21. *Let (k_n) is a sequence of integers such that $1 \leq k_n < n$ and satisfying Condition ($\mathcal{H}$) for the almost-sure result. If $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$, then*

$$\sup_{t\leq T_{n-k_n,n}} \left|\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right| = \begin{cases} O\left(\sqrt{\frac{\log\log n}{k_n}}\right), \\ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right). \end{cases}$$

The proof of Theorem 21 requires the following lemma.

Lemma 11. *The following equality holds almost-surely.*

$$|\widehat{\alpha}_{1,j} - \alpha_{1,j}| = O\left(\sqrt{\frac{\log\log n}{n}}\right) . \tag{4.9}$$

Proof. We have to be careful because $\widehat{\alpha}_{1,j}$ is not a sum of r.v. but the rapport of two sums of r.v. We write $\widehat{\alpha}_{1,j} - \alpha_{1,j}$ as:

$$\begin{aligned}
\widehat{\alpha}_{1,j} - \alpha_{1,j} &= \frac{p_n^{(1,j)}}{p_n^{(1)}} - \frac{p^{(1,j)}}{p^{(1)}} \\
&= p_n^{(1,j)} \left(\frac{1}{p_n^{(1)}} - \frac{1}{p^{(1)}}\right) + \frac{p_n^{(1,j)} - p^{(1,j)}}{p^{(1)}} \\
&= p_n^{(1,j)} \frac{p^{(1)} - p_n^{(1)}}{p^{(1)}\, p_n^{(1)}} + \frac{p_n^{(1,j)} - p^{(1,j)}}{p^{(1)}}
\end{aligned}$$

Applying the strong law of large number, we get, for n large enough, that the inequalities $1/p_n^{(1)} \leq 2/p^{(1)}$ and $p_n^{(1,j)} \leq 2\, p^{(1,j)}$ hold almost-surely. Then, the classical law of the iterated logarithm for i.i.d. r.v. of Hartman and Wintner (1941) entails that we have almost-surely

$$\left|p^{(1)} - p_n^{(1)}\right| = O\left(\sqrt{\frac{\log\log n}{n}}\right)$$

and for $j = 1, \ldots, \mathcal{J}$

$$\left|p^{(1,j)} - p_n^{(1,j)}\right| = O\left(\sqrt{\frac{\log\log n}{n}}\right) .$$

This leads to the desired result. □

Proof of Theorem 21. Consider the following decomposition:

$$\begin{aligned}
\widehat{\alpha}_{1,j}\, \widehat{F}_n(t) - \alpha_{1,j}\, F(t) &= \alpha_{1,j} \left(\widehat{F}_n(t) - F(t)\right) + (\widehat{\alpha}_{1,j} - \alpha_{1,j})\, F(t) \\
&\quad + (\widehat{\alpha}_{1,j} - \alpha_{1,j}) \left(\widehat{F}_n(t) - F(t)\right) .
\end{aligned}$$

Applying Theorem 6 and Lemma 11, we obtain straightly the result. □

4.4 A strong approximation result

We provide a strong approximation result for the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ jointly for $j = 1, \ldots, \mathcal{J}$. This result holds uniformly on the random increasing intervals $[0, T_{n-k_n,n}]$. The conditions under which this result holds are the same than the hypothesis required in Subsection 2.2.1.

Theorem 22. *Let (k_n) be a sequence of integers such that $1 \leq k_n < n$ for any n and fulfilling Hypothesis ($\mathcal{H}$). Assume that $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$. For n large enough, there exists a suitably enlarged probability space on which the following approximation result holds almost-surely, jointly for $j = 1 \ldots, \mathcal{J}$:*

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t) \right) - \widetilde{L}_n^{(j)}(t) \right| = O\left(\sqrt{n} \frac{\log n}{k_n} \right).$$

For $j = 1, \ldots, \mathcal{J}$, for fixed n, the processes $(\widetilde{L}_n^{(j)})$ zero-mean Gaussian processes with covariance function given for $k, j = 1, \ldots, \mathcal{J}$ and for $s, t \geq 0$ by:

$$\begin{aligned} Cov\left(\widetilde{L}_n^{(j)}(s), \widetilde{L}_n^{(k)}(t) \right) &= F(s)F(t) \frac{\alpha_{1,j} \left(I(k=j) - \alpha_{1,k} \right)}{p^{(1)}} \\ &+ \alpha_{1,j}\alpha_{1,k}(1 - F(s))(1 - F(t)) \int_0^{s \wedge t} \frac{dH^{(1)}}{(1 - H^-)^2}. \end{aligned}$$

The following decomposition will be of use for the strong approximation and the weak convergence.

$$\begin{aligned} \widehat{\alpha}_{1,j} - \alpha_{1,j} &= \widehat{\alpha}_{1,j} \frac{p^{(1)} - p_n^{(1)}}{p^{(1)}} + \frac{p_n^{(1,j)} - p^{(1,j)}}{p^{(1)}} \\ &= -\alpha_{1,j} \frac{p_n^{(1)} - p^{(1)}}{p^{(1)}} + \frac{p_n^{(1,j)} - p^{(1,j)}}{p^{(1)}} - \left(\widehat{\alpha}_{1,j} - \alpha_{1,j} \right) \frac{p_n^{(1)} - p^{(1)}}{p^{(1)}} \\ &= -\int \frac{\alpha_{1,j}}{p^{(1)}} d\left(H_n^{(1)} - H^{(1)} \right) + \int \frac{1}{p^{(1)}} d\left(H_n^{(1,j)} - H^{(1,j)} \right) + O\left(\frac{\log \log n}{n} \right). \end{aligned} \tag{4.10}$$

Proof of Theorem 22. We write:

$$\begin{aligned} \sqrt{n} \left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t) \right) &= \sqrt{n} \left(\widehat{\alpha}_{1,j} \widehat{F}_n(t) - \alpha_{1,j} F(t) \right) \\ &= \alpha_{1,j} \sqrt{n} \left(\widehat{F}_n(t) - F(t) \right) + \sqrt{n} \left(\widehat{\alpha}_{1,j} - \alpha_{1,j} \right) F(t) \\ &+ \sqrt{n} \left(\widehat{\alpha}_{1,j} - \alpha_{1,j} \right) \left(\widehat{F}_n(t) - F(t) \right). \end{aligned}$$

In view of Theorem 6 and Lemma 11, we have almost-surely:

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n} \left(\widehat{\alpha}_{1,j} - \alpha_{1,j} \right) \left(\widehat{F}_n(t) - F(t) \right) \right| = O\left(\frac{\log \log n}{(k_n)^{1/2}} \right).$$

Consequently, it suffices to consider:

$$I(t) = \alpha_{1,j} \sqrt{n} \left(\widehat{F}_n(t) - F(t) \right) + \sqrt{n} \left(\widehat{\alpha}_{1,j} - \alpha_{1,j} \right) F(t).$$

The next step consists in linearizing $\widehat{F}_n - F$ by means of Proposition 1. We get uniformly in t in $[0, T_{n-k_n,n}]$ that:

$$I(t) = \alpha_{1,j}\sqrt{n}(1-F(t))\left(\Lambda_n(t)-\Lambda(t)\right) + \sqrt{n}\left(\widehat{\alpha}_{1,j}-\alpha_{1,j}\right)F(t) + O\left(\sqrt{n}\frac{\log\log n}{k_n}\right).$$

We make use of the decomposition of $\widehat{\alpha}_{1,j}-\alpha_{1,j}$ obtained in (4.10) and of the decomposition of $\Lambda_n - \Lambda$ outlined in (1.8). Using the construction of Subsection 2.3.2, we define the approximating processes for $t \geq 0$ by:

$$\begin{aligned} II^{(j)}(t) = \alpha_{1,j}\left(1-F(t)\right)\left(\int_0^t \frac{dB_n^{(1)}}{(1-H^-)} + \int_0^t \frac{B_n^-}{(1-H^-)^2}dH^{(1)}\right) \\ + F(t)\left(-\int \frac{\alpha_{1,j}}{p^{(1)}}\,dB_n^{(1)} + \int \frac{1}{p^{(1)}}\,dB_n^{(1,j)}\right), \end{aligned}$$

where B_n, $B_n^{(1)}$ and $B_n^{(1,j)}$ for $j = 1, \ldots, \mathcal{J}$ are the correlated Brownian bridges of Lemma 6. The construction of Subsection 2.3.2 guaranties that the approximation holds jointly for $j = 1, \ldots, \mathcal{J}$ on a suitably enlarged probability space. □

4.5 Weak convergence

In this section, we derive the joint weak convergence of the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ for $j = 1, \ldots, \mathcal{J}$. The result is given on $[0, \sigma]$ for $\sigma < \tau_H$ but also holds on $[0, T_{n-k_n,n}]$ when the sequence of integers (k_n) satisfies Hypothesis $(\mathcal{H})$. Note that here we cannot use the martingale theory, as in Theorems 3 and 17, because of $\widehat{\alpha}_{1,j}$ which is neither a martingale nor predictable.

Theorem 23. *Let $\sigma < \tau_H$. Assume that $F^{(j)}$ is continuous for $j = 1, \ldots, \mathcal{J}$. The following convergence holds in the space $D^{\mathcal{J}}[0,\sigma]$ of càdlàg functions from $[0,\sigma]$ to $\mathbb{R}^{\mathcal{J}}$:*

$$\left(\sqrt{n}\left(\widetilde{F}_n^{(1)} - F^{(1)}\right), \ldots, \sqrt{n}\left(\widetilde{F}_n^{(\mathcal{J})} - F^{(\mathcal{J})}\right)\right) \xrightarrow{\mathcal{D}} \left(\widetilde{K}^{(1)}, \ldots, \widetilde{K}^{(\mathcal{J})}\right),$$

where the $\widetilde{K}^{(j)}$ are mean-zero Gaussian processes with covariance function given for $s, t \geq 0$ and for $j = 1, \ldots, \mathcal{J}$ by:

$$\begin{aligned} Cov\left(\widetilde{K}^{(j)}(s), \widetilde{K}^{(k)}(t)\right) = F(s)F(t)\,\frac{\alpha_{1,j}\left(I(k=j)-\alpha_{1,k}\right)}{p^{(1)}} \\ + \alpha_{1,j}\alpha_{1,k}\left(1-F(s)\right)\left(1-F(t)\right)\int_0^{s\wedge t}\frac{dH^{(1)}}{(1-H^-)^2}. \end{aligned}$$

Here, the assumption of continuity of the $F^{(j)}$ is required to apply the lemma to come which is a linearization of the product-limit quantity $\widehat{F}_n - F$ on any compact interval included in the support of H.

Lemma 12. (Giné and Guillou 1999) *Let $\sigma < \tau_H$. If F is continuous, then*

$$\sup_{t\leq\sigma}\left|\frac{\widehat{F}_n(t)-F(t)}{1-F(t)}-(\Lambda_n(t)-\Lambda(t))\right| = O\left(\frac{\log\log n}{n}\right).$$

Lemma 13. (Giné and Guillou 1999) *Let $\sigma < \tau_H$. If F is continuous, then*

$$\sup_{t\leq\sigma}\left|\frac{\widehat{F}_n(t)-F(t)}{1-F(t)}\right| = O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right).$$

Lemma 14. (Giné and Guillou 1999) *Let $\sigma < \tau_H$. Then the following decomposition holds uniformly for t in $[0,\sigma]$:*

$$\sqrt{n}(\Lambda_n(t)-\Lambda(t)) = \int_0^t \frac{d\alpha_n^{(1)}}{(1-H^-)} + \int_0^t \frac{\alpha_n^-}{(1-H^-)^2} dH^{(1)} + O\left(\sqrt{n}\,\frac{\log\log n}{n}\right), \quad (4.11)$$

where $\alpha_n = \sqrt{n}(H_n - H)$ and $\alpha_n^{(1)} = \sqrt{n}(H_n^{(1)} - H^{(1)})$.

Proof of Theorem 23. We write:

$$\sqrt{n}\left(\widetilde{F}_n^{(j)}(t)-F^{(j)}(t)\right) = \alpha_{1,j}\sqrt{n}\left(\widehat{F}_n(t)-F(t)\right) + \sqrt{n}\left(\widehat{\alpha}_{1,j}-\alpha_{1,j}\right)F(t) + \sqrt{n}\left(\widehat{\alpha}_{1,j}-\alpha_{1,j}\right)\left(\widehat{F}_n(t)-F(t)\right).$$

In view of Lemma 13 and Lemma 11, we have almost-surely:

$$\sup_{t\leq\sigma}\left|\sqrt{n}\left(\widehat{\alpha}_{1,j}-\alpha_{1,j}\right)\left(\widehat{F}_n(t)-F(t)\right)\right| = o(1).$$

Consequently, it suffices to consider:

$$I_n^{(j)}(t) = \alpha_{1,j}\sqrt{n}\left(\widehat{F}_n(t)-F(t)\right) + \sqrt{n}\left(\widehat{\alpha}_{1,j}-\alpha_{1,j}\right)F(t).$$

The next step consists in linearizing $\widehat{F}_n - F$ by means of Lemma 12. We get uniformly in t in $[0,\sigma]$ that:

$$I_n^{(j)}(t) = \alpha_{1,j}(1-F(t))\sqrt{n}\left(\Lambda_n(t)-\Lambda(t)\right) + \sqrt{n}\left(\widehat{\alpha}_{1,j}-\alpha_{1,j}\right)F(t) + O\left(\frac{\log\log n}{\sqrt{n}}\right).$$

Then we make use of the decomposition of $\widehat{\alpha}_{1,j} - \alpha_{1,j}$ obtained in (4.10) and of the decomposition of $\Lambda_n - \Lambda$ outlined in Lemma 14. We get that the limiting process of $I_n^{(j)}(t)$ is the same than the limiting process of

$$II_n^{(j)}(t) = \alpha_{1,j}(1-F(t))\left(\int_0^t \frac{d\alpha_n^{(1)}}{(1-H^-)} + \int_0^t \frac{\alpha_n^-}{(1-H^-)^2} dH^{(1)}\right) + F(t)\left(-\int \frac{\alpha_{1,j}}{p^{(1)}}\, d\alpha_n^{(1)} + \int \frac{1}{p^{(1)}}\, d\alpha_n^{(1,j)}\right).$$

Lemma 6 guaranties that the weak convergence of the processes $\alpha_n^{(1)}$ and $\alpha_n^{(1,j)}$ for $j = 1, \ldots, \mathcal{J}$ holds jointly in $D^{\mathcal{J}+2}[0, \sigma]$. It also guaranties that the processes $II_n^{(j)}$ for $j = 1, \ldots, \mathcal{J}$ converge jointly in $D^{\mathcal{J}}[0, \sigma]$ as n goes to infinity to the limiting processes defined for $t \geq 0$ and $j = 1, \ldots, \mathcal{J}$ by:

$$\widetilde{II}^{(j)}(t) = \alpha_{1,j}\,(1 - F(t)) \left(\int_0^t \frac{dB^{(1)}}{(1 - H^-)} + \int_0^t \frac{B^-}{(1 - H^-)^2} dH^{(1)} \right) + F(t) \left(- \int \frac{\alpha_{1,j}}{p^{(1)}}\, dB^{(1)} + \int \frac{1}{p^{(1)}}\, dB^{(1,j)} \right),$$

where B, $B^{(1)}$ and $B^{(1,j)}$ for $j = 1, \ldots, \mathcal{J}$ are correlated Brownian bridges with the same covariance structure as in Lemma 6. □

Remark 5. *The structure of the covariance function above is very similar to the structure obtained by Gather and Pawlitschko (1998). In addition, under the assumption of continuity of F, the term $(1 - F(s))(1 - F(t)) \int_0^{s \wedge t} dH^{(1)}/(1 - H^-)^2$ is exactly the variance function of the limiting process of the Kaplan-Meier estimator. Consequently, the variance is exactly the same as if $\widehat{F}_n$ and $\widehat{\alpha}_{1,j}$ were independent, which corresponds to the uncensored case.*

Remark 6. *The transformation $\widehat{F}_n^{(j)} - F^{(j)}(1 - \widehat{F}_n)/(1 - F)$ of the process $\widehat{F}_n^{(j)} - F^{(j)}$ led, in Section 3.2, to a weak convergence result that permitted to construct confidence bands. Here, the same transformation does not lead to the same interesting result. Indeed, the same transformation*

$$\widetilde{F}_n^{(j)} - F^{(j)} \frac{1 - \widehat{F}_n}{1 - F} = \frac{\widehat{\alpha}_{1,j} \widehat{F}_n - \alpha_{1,j} F}{1 - F}$$

does not change the nature of the limiting process.

Remark 7. *If the model given by (4.2) is valid, then we have seen in the present chapter that the subdistribution function $F^{(j)}$ may be consistenly estimated by the estimator $\widetilde{F}_n^{(j)}$ for $j = 1, \ldots, \mathcal{J}$. This estimator has very similar properties to $\widehat{F}_n^{(j)}$ for $j = 1, \ldots, \mathcal{J}$ in the sense that it achieve the same rates of convergence in probability and almost-surely, it may be approximated by a Gaussian process with the same error uniformly on $[0, T_{n-k_n,n}]$ and converges weakly to a Gaussian process. But, the estimator $\widetilde{F}_n^{(j)}$ is asymptotically more efficient than $\widehat{F}_n^{(j)}$ under the model given by (4.2). On the other hand, we can not construct confidence bands for $F^{(j)}$ based on $\widetilde{F}_n^{(j)}$, at least with the methodology of Subsection 3.3 of Chapter 3.*

Chapter 5

Recurrent events in the presence of a terminal event and independent right-censoring

In clinical and epidemiological studies, subjects may experience recurrent events. Examples include repeated seizures in epileptic patients, tumor recurrences in cancer patients and recurrent opportunistic infections in HIV patients. The recurrence of such serious events is often associated with a high risk of death with the consequence that subjects may die during the study. For each recurrence, death and non-fatal events are treated as competing risks. We focus on the cause-specific distribution function of the duration time between two successive non-fatal events and on the cause-specific distribution function of the duration time between a non-fatal event and successive death, both conditionally on having survived the previous occurrences. We propose a nonparametric estimator for each of these functions in the presence of independent right-censoring and show their consistency. We also investigate through simulations the finite sample properties of the proposed estimators. Finally, we apply our estimators to the data of the study named Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events (CAPRIE) which is described in the sequel.

5.1 Introduction and background

Recurrence time data are commonly encountered in longitudinal studies when failure events can occur repeatedly over time for each subject. As written above, in clinical and epidemiological studies, classical examples of recurrent events are repeated seizures in epileptic patients, tumor recurrences in cancer patients and recurrent opportunistic infections in HIV patients. In the sequel, we consider a population of patients suffering from atherosclerosis.

Atherosclerosis is the name of the process in which deposits of fatty substances, cholesterol, cellular waste products, calcium and other substances build up in the inner lining

of an artery. This buildup is called plaque. It usually affects large and medium-sized arteries and leads to the hardening and/or to the stricture of the concerned arteries. Indeed, plaques can grow large enough to significantly reduce the blood's flow through an artery. But most of the damage occurs when they become fragile and rupture. Plaques that rupture cause the forming of blood clots. These clots can block blood flow partially or totally. The consequence is a failure to adjust the oxygen supply to the needs of the tissues: it is the ischaemic event. The clinical manifestations of atherosclerosis are various. The three main localizations correspond to the coronary arteries (myocardial infarction), to the carotid arteries (stroke) and to the lower limbs arteries (arteritis). See Cambou (1999), Teigner and Castaigne (1999) or the American Heart Association (www.americanheart.org) for additional materials.

In order to analyze the recurrent events, we distinguish two categories of ischaemic events which are respectively non-fatal ischaemic events (NFIE) and fatal ischaemic events (death), whatever the localization of the ischaemic events. Indeed, when assessing the efficacy of a treatment in prevention of ischaemic events, it seems interesting to take into account all types of ischaemic events in order to put forward an opinion on the global qualities of the treatment, the latter having to prevent from further recurrence of ischaemic events without generating other risks for the patient. The deaths are distinguished because of their "terminal" nature which should be taken into account. Indeed, the death of a patient obviously stops any subsequent recurrence of ischaemic events (hence the name "terminal event" for death).

The observation of recurrent events may be right-censored. Censoring occurs at most once for a given patient and prevents any further ischaemic events from being observed. Possible reasons for independent censoring are loss to follow-up (non-related to side-effects of the treatment under study) or end of study.

The kind of data collected during such a trial is illustrated on Figure 5.1 for six subjects $S_1, \ldots, S_6$. The follow-up time for a given patient is represented by a straight line along which the different events are indicated. These data can be regarded as multivariate data that have specific characteristics. First of all, the different patients do not experience the same number of events and the number of patients still alive and still under study decreases as the events occur. In addition, note that the last observation of each subject is either a censoring time or a death time. Secondly, for a given patient, the occurrence of a NFIE is often followed by a deterioration of the general state of health and therefore increases the risk of novel NFIE and even of death (fatal ischaemic event). Thus, for a given patient, the recurrence times are mutually correlated and are correlated with the death time. This dependence should be accounted for. On the opposite, the censoring process is assumed in the sequel to be independent of both recurrent and terminal processes.

Throughout, the recurrent, terminal and censoring processes are all assumed to have continuous distribution so that NFIE and death cannot happen at the same time.

Point process modeling and estimation with recurrent event data has received much interest lately. In the analysis of recurrent event data, the focus can be placed on the

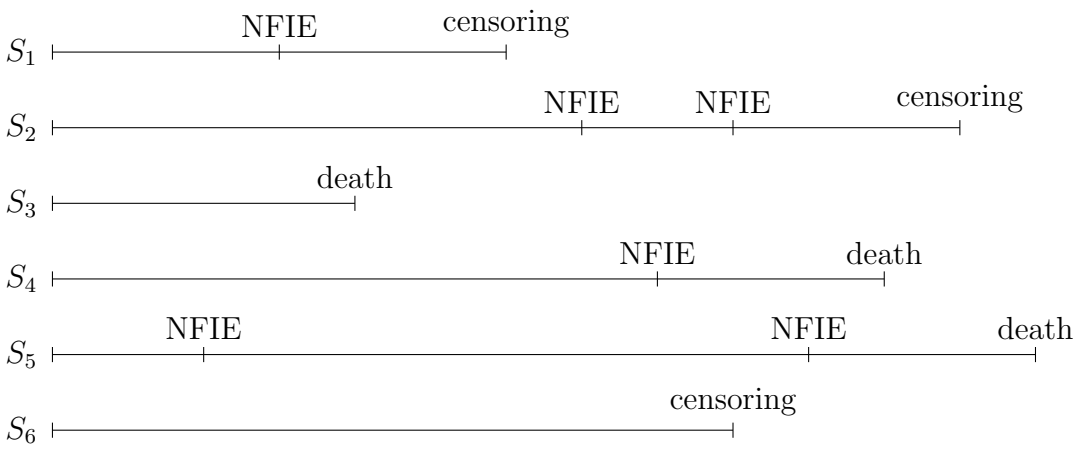

Figure 5.1: Example of data (NFIE = non-fatal ischaemic event).

duration times between successive events, on the times to recurrent events, on the recurrent event process intensity or on the *cumulative mean function* also termed as the *mean frequency function* defined, for $t \geq 0$, by $M(t) = \mathbb{E}[N(t)]$ where, for $t \geq 0$, the recurrent event process $N(t)$ records the number of recurrent events occurring before t. Two modeling approaches often adopted are marginal models and frailty models (in which a latent variable is used to take into consideration a subject-specific random effect). Various regression models have also been considered to address different types of questions. The appropriateness of a selected model depends on the nature of the recurrent events data as well as on the interest of the study.

In the absence of a terminal event, an abundant literature on the analysis of recurrent events has been developed. We present here a brief review.

Prentice et al. (1981) considered Cox-type regression models which allow the intensity of the recurrent event process to depend on the individual's prior failure history through stratification. Allowing for independent censoring and time-dependent covariates, Andersen and Gill (1982) carried out Cox-type regression analysis for the intensity of the recurrent process which is assumed to be a time-transformed Poisson process. Andersen et al. (1993) also adopted modeling techniques based on the intensity process in the presence of censoring. Lin et al. (2000) provided asymptotic distribution theory for the fitting of Cox-type mean models without the Poisson assumption.

Wei, Lin and Weissfeld (1989) emphasized times to recurrent events. Within this framework, the k-th event is modeled as the k-th failure time variable. Each marginal distribution of the different recurrence times is formulated by a Cox proportional hazards model.

Lawless and Nadeau (1995) proposed a nonparametric estimator for the cumulative mean function. They also proposed a semiparametric regression model in which the intensity of the recurrent event process is expressed as a function of regression variables

unconditional on the individual's preceding failure history.

Lin, Wei and Ying (1998) studied an accelerated failure time model for the cumulative mean function. Chang (2004) considered an accelerated failure time model for the gap times between successive recurrent events with an additional frailty. In both models, covariate effect remains the same for all recurrences.

Wang and Wells (1998) considered joint nonparametric estimation for two successive duration times in the presence of independent right-censoring. Wang and Chang (1999) also focused on the gap times between successive events and developed inference procedures under a nonparametric frailty model assuming that, given the frailty, the successive duration times are i.i.d.. Lin et al. (1999) proposed a nonparametric estimator for the distribution function of the duration time between two successive events conditional on having experienced the previous events before a given time. They also proposed a nonparametric estimator for the joint distribution function of the gap times.

All these methods, however, assume that recurrent events are not terminated by death during the study. In many situations, a terminal event (death) may happen during the follow-up period precluding further occurrences of the recurrent events. Furthermore, as noticed before, the terminal event is often dependent of the recurrent event history. Therefore it is also termed as dependent censoring or informative censoring. Some efforts have been made recently to account for death in the analysis of recurrent events.

Li and Lagakos (1997) and Derzko and Leconte (2004) treated death as a competing risk. Li and Lagakos (1997) estimated nonparametrically the cause-specific hazard function for each recurrence while Derzko and Leconte (2004) proposed a nonparametric estimator of the marginal distribution of the times to recurrent events.

Adapting the work of Lawless and Nadeau (1995) to take the presence of a terminal event into consideration, Cook and Lawless (1997) proposed a nonparametric estimator of the mean frequency function conditional on survival. Ghosh and Lin (2000) studied the one and two-sample asymptotic properties of this estimator.

Ghosh and Lin (2003) performed semiparametric regression with a scale-change model that formulates jointly the marginal distributions of the recurrent event process and death, while leaving the dependence structure unspecified.

Ghosh and Lin (2002) developed semiparametric regression models using IPCW (Inverse Probability of Censoring Weighting) and a similar technique IPSW (Inverse Probability of Survival Weighting) in the presence of death. Miloslavsky et al. (2004) also presented an IPCW estimator for the regression parameters in the Andersen-Gill model which is consistent under dependent censoring when the censoring mechanism is estimated consistently. But the application of these kind of models may be limited due to the required conditions which may not be satisfied in practice.

Frailty models have also received attention in the presence of a terminal event. Wang, Qin and Chiang (2001) modeled the occurrence of recurrent events by a non-stationary Poisson process in which a frailty characterizes the correlation structure between the recurrent event process and the terminal event. Conditionally on frailty, recurrent events and terminal events are independent. They treated the distribution of both death and frailty as a nuisance parameter. Thus, their model cannot be applied to situations where modeling the recurrent and terminal events are both of interest. On the opposite, Huang

and Wang (2004) considered a joint semi-parametric regression model for both the recurrent event process and the death. A frailty is used to model the association between the intensity of the recurrent event process (which is assumed to be non-stationary Poisson) and the hazard of the terminal event. Liu et al. (2004) considered a proportional hazards model for the recurrent and terminal event processes with an additional frailty. The dependence is modeled by conditioning on the frailty which is included in both the intensity of the recurrent event process and the hazard of the terminal event. The terminal and recurrent events are not independent even conditional on frailty.

As in Lin et al. (1999), the present work focuses on conditional distribution functions. Here interest is centred on the nonparametric estimation of the distribution function of the duration time between two successive events conditional on having survived the previous events which have occured before a given time. The methodology of Lin et al. (1999) is adapted to take death into consideration by treating death as a competing risk as in Li and Lagakos (1997) and Derzko and Leconte (2004).

5.2 Notations

Let $j = 0$ denote the index of the initial time point and set the initial time point as 0. Let $\widetilde{X}^{[0]} = 0$ and, for $k \geq 1$, let $\widetilde{X}^{[k]}$ be a non-negative r.v. representing the occurrence time of the k-th NFIE. These r.v. are stochastically ordered, they satisfy $\widetilde{X}^{[0]} < \widetilde{X}^{[1]} < \widetilde{X}^{[2]} < \dots$. Let Z be a non-negative r.v. with distribution function F_Z that stands for the death time. The r.v. Z is observed at most once and its realization precludes the further realization of $\widetilde{X}^{[k]}$. The r.v. $\widetilde{X}^{[k]}$ and Z are correlated. Let $N = \sum_{k=1}^{\infty} I(\widetilde{X}^{[k]} < Z) + 1$ be an integer-valued r.v. representing the total number of ischaemic events (including non-fatal and fatal events) experienced by a given individual.

For a given patient, after the occurrence of the first event, there is three possibilities. The patient may have died, precluding the occurrence of any subsequent event. The patient may be definitively censored precluding the observation of any subsequent event. The patient may have experienced a first NFIE, in which case the patient is still alive and still under study. The patient is then at risk of having a novel NFIE and of dying. He may also be right-censored. Thus, for each recurrence, conditional on having survived the previous occurrences, the patient is submitted to two dependent competing risks (NFIE and death) in the presence of independent right-censoring as displayed on Figure 5.2.

To account for dependent competing risks, we introduce the following r.v.

- For $k = 1, \dots, N$, let $X^{[k]} = \widetilde{X}^{[k]} \wedge Z$ be the non-negative r.v. representing the occurrence time of the k-th ischaemic event (fatal or non-fatal). These r.v. satisfy $X^{[1]} < \dots < X^{[N]}$.

- For $k = 1, \dots, N$, let $Y^{[k]} = X^{[k]} - X^{[k-1]}$ be the duration time between the $(k-1)$-th and the k-th ischaemic event. Of course, we have $X^{[1]} = Y^{[1]}$ and $X^{[k]} = \sum_{l=1}^{k} Y^{[l]}$.

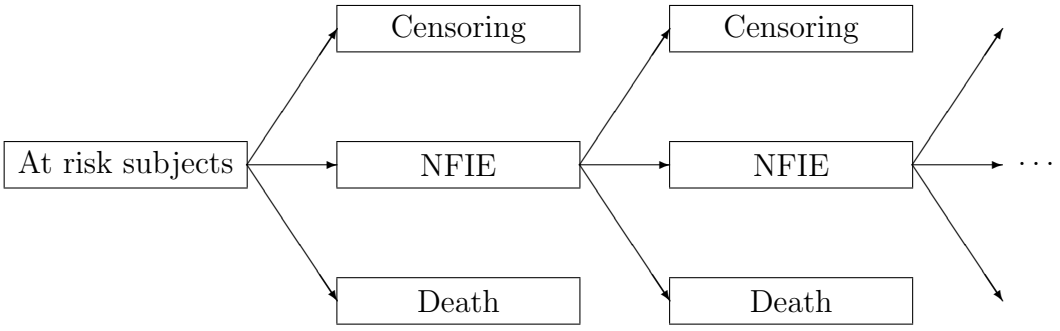

Figure 5.2: Competing risks at each recurrence in the presence of independent censoring (NFIE = non-fatal ischaemic event).

- For $k = 1, \ldots, N$, let $\mathcal{C}^{[k]}$ be the r.v. indicating the type of the k-th ischaemic event:

$$\mathcal{C}^{[k]} = \begin{cases} 1 & \text{if the observed event is a NFIE,} \\ 2 & \text{if the observed event is a death (fatal ischaemic event).} \end{cases}$$

- Let C be the r.v. with distribution function G that independently right-censors the $X^{[k]}$ for $k = 1, \ldots, N$ or equivalently the $Y^{[k]} = X^{[k]} - X^{[k-1]}$ for $k = 1, \ldots, N$. It is observed at most once. No further realization of the $X^{[k]}$ can be observed after the realization of C.

The following r.v. are observable.

- Let $K = \sum_{k=1}^{\infty} I(X^{[k]} < C) + I(C < Z)$ be the total number of observed events for a given individual.

- For $k = 1, \ldots, K$, let $T_X^{[k]} = \min(X^{[k]}, C)$ be the time of occurrence of the k-th event (which is either a NFIE or a death or a censoring). These r.v. satisfy $T_X^{[1]} < \cdots < T_X^{[K]}$.

- For $k = 1, \ldots, K$, let $T_Y^{[k]} = \min(Y^{[k]}, C - \sum_{l=1}^{k-1} Y^{[l]})$ be the r.v. that stand for the duration time between the $(k-1)$-th and the k-th event.

- For $k = 1, \ldots, K$, the r.v. $J^{[k]} = \mathcal{C}^{[k]} I\left(X^{[k]} \leq C\right) = \mathcal{C}^{[k]} I(\sum_{l=1}^{k} Y^{[l]} \leq C)$ indicates the nature of the k-th observed event.

At last, let H be the distribution function defined by $1 - H = (1 - F_Z)(1 - G)$ and let $\tau_H = \sup\{x : H(x) < 1\}$ be the right-endpoint of the distribution function H.

5.3 Functions of interest

When evaluating the efficacy of a treatment on an episodic illness, it is often important to assess whether or not the treatment delays the time from treatment start to the first episode, the time from the first episode to the second episode and so on. We prefer to focus on the gap time between the $(k-1)^{\text{th}}$ event and the k^{th} NFIE rather than on the total time from treatment start to the k^{th} NFIE. Indeed, several phenomenons may affect the understanding of the illness mechanism. On the one hand, a treatment which delays the first episode will inevitably lengthen the total time to the second episode even if it becomes ineffective after the first episode. On the other hand, in some cases, a compensating phenomenon between the different stages of a disease or of a treatment may exist. For example, a treatment may delay the occurrence of the first ischaemic event but have the reverse effect on the occurrence of the subsequent ischaemic events. It is important to detect such a phenomenon. Considering the total time from treatment start to the k^{th} NFIE, as in Derzko and Leconte (2004), would be interesting as a complementary approach.

In addition to these remarks, it is straightforward that the duration times between an ischaemic event and the next are relevant only among patients who are still alive after the previous occurrences of ischaemic events. Consequently, relevant quantities of interest are the duration times between an ischaemic event and the next conditionally on having experienced the preceding events before a given time and on having survived these previous occurrences.

To refine this conditional analysis, we consider separately NFIE and death. Indeed, a new treatment may decrease the ischaemic-event incidence through a decrease in the incidence of NFIE or through a decrease in the incidence of death (fatal ischaemic event). These endpoints do not have the same seriousness for the patient. Of course, a NFIE is of lesser concern than death. Moreover, death has a terminal nature that should be taken into consideration. A prudent analysis would consider both aspects of ischaemic event occurrence. Thus, for each recurrence, the incidence function of ischaemic events is decomposed into two cause-specific distribution functions pertaining to the two dependent competing risks that are respectively experiencing a NFIE and dying.

With these remarks in view, the functions that serve as useful descriptive devices are the following. Concerning the first duration time, we consider for $t \geq 0$:

$$F^{[1]}(t) = \mathbb{P}\left[Y^{[1]} \leq t\right] ,$$

which we further decompose into $F^{[1]}(t) = F^{[1(1)]}(t) + F^{[1(2)]}(t)$ where for $t \geq 0$

$$F^{[1(1)]}(t) = \mathbb{P}\left[Y^{[1]} \leq t, \mathcal{C}^{[1]} = 1\right] ,$$

and

$$F^{[1(2)]}(t) = \mathbb{P}\left[Y^{[1]} \leq t, \mathcal{C}^{[1]} = 2\right] .$$

Concerning the second duration time, we consider for $t_1, t_2 \geq 0$:

$$F^{[2/1]}(t_2/t_1) = \mathbb{P}\left[Y^{[2]} \leq t_2 \middle| Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1\right] ,$$

which we further decompose into $F^{[2/1]}(t_2/t_1) = F^{[2(1)/1]}(t_2/t_1) + F^{[2(2)/1]}(t_2/t_1)$ where

$$F^{[2(1)/1]}(t_2/t_1) = \mathbb{P}\left[Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 1 \middle| Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1\right] ,$$

and

$$F^{[2(2)/1]}(t_2/t_1) = \mathbb{P}\left[Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 2 \middle| Y^{[1]} \leq t_1, \mathcal{C}^{[1]} = 1\right] .$$

More generally, we consider for $t_1, \ldots, t_k \geq 0$:

$$\begin{aligned} &F^{[k/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) \\ &= \mathbb{P}\left[Y^{[k]} \leq t_k \middle| Y^{[1]} \leq t_1, \ldots, Y^{[k-1]} \leq t_{k-1}, \mathcal{C}^{[1]} = 1, \ldots, \mathcal{C}^{[k-1]} = 1\right] , \end{aligned}$$

which we further decompose into

$$\begin{aligned} &F^{[k/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) \\ &= F^{[k(1)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) + F^{[k(2)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) \end{aligned}$$

where

$$\begin{aligned} &F^{[k(1)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) \\ &= \mathbb{P}\left[Y^{[k]} \leq t_k, \mathcal{C}^{[k]} = 1 \middle| Y^{[1]} \leq t_1, \ldots, Y^{[k-1]} \leq t_{k-1}, \mathcal{C}^{[1]} = 1, \ldots, \mathcal{C}^{[k-1]} = 1\right] , \end{aligned}$$

and

$$\begin{aligned} &F^{[k(2)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) \\ &= \mathbb{P}\left[Y^{[k]} \leq t_k, \mathcal{C}^{[k]} = 2 \middle| Y^{[1]} \leq t_1, \ldots, Y^{[k-1]} \leq t_{k-1}, \mathcal{C}^{[1]} = 1, \ldots, \mathcal{C}^{[k-1]} = 1\right] . \end{aligned}$$

Notice that $F^{[1]}(t)$ is estimable only if $t < \tau_H$, that $F^{[2/1]}(t_2/t_1)$ is estimable only if $t_1 + t_2 < \tau_H$ and so on.

5.4 Estimation

The objective of this section is to develop nonparametric estimators for the functions displayed above without any assumption regarding either the dependence structure among the multiple endpoints or an eventual Poisson structure of the recurrent process intensity. The strong consistency on compact intervals of these estimators is stated in Theorems 24, 25 and 26.

We suppose that observations are taken on an i.i.d. sample of n individuals. For $i = 1, \ldots, n$, the data for the i-th individual consists of K_i couples where $K_i - 1$ is the number of observed NFIE. For $k = 1, \ldots, K_i$, the k-th couple is given by $(T_{X,i}^{[k]}, J_i^{[k]})$ or equivalently $(T_{Y,i}^{[k]}, J_i^{[k]})$ which are distributed as $(T_X^{[k]}, J^{[k]})$ and $(T_Y^{[k]}, J^{[k]})$ respectively.

5.4.1 Estimation of the censoring distribution function

As seen in Section 5.1, the last observation for a given patient is either a censoring time or a death time. For a given patient, we do not observe both the censoring event and the death but only the first event that occurs. Since the death and the censoring processes are independent, the censoring distribution function (and the distribution function of the death) may be estimated by the Kaplan-Meier estimator based on the last observations for each patient i.e. on the data $(T_{X,i}^{[K_i]}, J_i^{[K_i]})$ for $i = 1, \ldots, n$.

The Kaplan-Meier estimator of the censoring distribution function G is given for $t \geq 0$ by:

$$\widehat{G}_n(t) = 1 - \prod_{i=1}^{n} \left(1 - \frac{I\left(T_{X,i}^{[K_i]} \leq t, J_i^{[K_i]} = 0\right)}{\sum_{l=1}^{n} I\left(T_{X,l}^{[K_l]} \geq T_{X,i}^{[K_i]}\right)} \right). \tag{5.1}$$

5.4.2 Estimation with the first duration

We derive an estimator for the functions $F^{[1(1)]}$, $F^{[1(2)]}$ and $F^{[1]}$. Let $H^{[1(1,1)]}$ be the empirical function defined for $t \geq 0$ by:

$$H^{[1(1,1)]}(t) = \mathbb{P}\left[T_Y^{[1]} \leq t, J^{[1]} = 1\right].$$

For $t \geq 0$, we obtain the following relation:

$$\begin{aligned} H^{[1(1,1)]}(t) &= \mathbb{P}\left[Y^{[1]} \leq t, Y^{[1]} \leq C, \mathcal{C}^{[1]} = 1\right] \\ &= \iint I\left(u \leq t, u \leq c\right) G(dc) F^{[1(1)]}(du) \\ &= \int_{u \leq t} \left(1 - G^{-}(u)\right) F^{[1(1)]}(du). \end{aligned}$$

We deduce that $F^{[1(1)]}(t)$ can be written in terms of the estimable functions G and $H^{[1(1,1)]}$:

$$F^{[1(1)]}(t) = \int_{u \leq t} \frac{H^{[1(1,1)]}(du)}{1 - G^{-}(u)}.$$

We can get in the same way that for $t \geq 0$:

$$F^{[1]}(t) = \int_0^t \frac{dH^{[1(1)]}}{1 - G^{-}},$$

and

$$F^{[1(2)]}(t) = \int_0^t \frac{dH^{[1(1,2)]}}{1 - G^{-}},$$

where

$$H^{[1(1)]}(t) = \mathbb{P}\left[T_Y^{[1]} \leq t, J^{[1]} \neq 0\right],$$

and

$$H^{[1(1,2)]}(t) = \mathbb{P}\left[T_Y^{[1]} \leq t, J^{[1]} = 2\right] .$$

Consequently, we propose to estimate the functions $F^{[1]}$, $F^{[1(1)]}$ and $F^{[1(2)]}$ by means of "plug-in" estimators denoted respectively by $\widehat{F}_n^{[1]}$, $\widehat{F}_n^{[1(1)]}$ and $\widehat{F}_n^{[1(2)]}$. These estimators are obtained by replacing $H^{[1(1)]}$, $H^{[1(1,1)]}$, $H^{[1(1,2)]}$ and G by their corresponding estimators which are respectively the empirical distribution functions $H_n^{[1(1)]}$, $H_n^{[1(1,1)]}$ and $H_n^{[1(1,2)]}$ and the Kaplan-Meier estimator $\widehat{G}_n$. The Kaplan-Meier estimator $\widehat{G}_n$ is defined in (5.1) while the empirical distribution functions $H_n^{[1(1)]}$, $H_n^{[1(1,1)]}$ and $H_n^{[1(1,2)]}$ are defined for $t \geq 0$ by:

$$H_n^{[1(1)]}(t) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t, J_i^{[1]} \neq 0\right) ,$$

$$H_n^{[1(1,1)]}(t) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t, J_i^{[1]} = 1\right) ,$$

$$H_n^{[1(1,2)]}(t) = \frac{1}{n}\sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t, J_i^{[1]} = 2\right) .$$

Consequently, we let for $t \geq 0$

$$\widehat{F}_n^{[1]}(t) = \int_0^t \frac{dH_n^{[1(1)]}}{1-\widehat{G}_n^-} , \quad \widehat{F}_n^{[1(1)]}(t) = \int_0^t \frac{dH_n^{[1(1,1)]}}{1-\widehat{G}_n^-} \quad \text{and} \quad \widehat{F}_n^{[1(2)]}(t) = \int_0^t \frac{dH_n^{[1(1,2)]}}{1-\widehat{G}_n^-} .$$

Theorem 24. *For any $\sigma < \tau_H$, the estimators $\widehat{F}_n^{[1]}$, $\widehat{F}_n^{[1(1)]}$ and $\widehat{F}_n^{[1(2)]}$ defined above are strongly consistent on $[0,\sigma]$ for $F^{[1]}$, $F^{[1(1)]}$ and $F^{[1(2)]}$ respectively.*

Proof. We write:

$$\begin{aligned} F^{[1(1)]}(t) - \widehat{F}_n^{[1(1)]}(t) &= \int_0^t \frac{dH^{[1(1,1)]}}{1-G^-} - \int_0^t \frac{dH_n^{[1(1,1)]}}{1-\widehat{G}_n^-} \\ &= \int_0^t \frac{1}{1-G^-}\, d\left(H^{[1(1,1)]} - H_n^{[1(1,1)]}\right) \\ &+ \int_0^t \left(\frac{1}{1-G^-} - \frac{1}{1-\widehat{G}_n^-}\right) dH_n^{[1(1,1)]} . \end{aligned}$$

We carry out integration by parts on the first term in the above equality and get straightforwardly:

$$\begin{aligned} \sup_{t\leq\sigma}\left|\widehat{F}_n^{[1(1)]}(t) - F^{[1(1)]}(t)\right| &\leq \frac{2}{1-G(\sigma)} \sup_{t\leq\sigma}\left|H^{[1(1,1)]}(t) - H_n^{[1(1,1)]}(t)\right| \\ &+ \sup_{t\leq\sigma}\left|\frac{1}{1-G(t)} - \frac{1}{1-\widehat{G}_n(t)}\right| . \end{aligned}$$

The fact that $G(\sigma) < 1$ together with Glivenko-Cantelli's theorem and Theorem 4 entails the almost sure convergence of $\widehat{F}_n^{[1(1)]}$ to $F^{[1(1)]}$ on $[0, \sigma]$. The proof is identical for $\widehat{F}_n^{[1(2)]}$ and $\widehat{F}_n^{[1]}$. $\square$

5.4.3 Estimation with the second duration

We derive here an estimator for each of the conditional functions $F^{[2(1)/1]}$, $F^{[2(2)/1]}$ and $F^{[2/1]}$. For that, consider the following bivariate distribution functions defined for $t_1, t_2 \geq 0$ by:

$$\begin{aligned}
H^{[1(1,1),2(1)]}(t_1, t_2) &= \mathbb{P}\left[T_Y^{[1]} \leq t_1, T_Y^{[2]} \leq t_2, J^{[1]} = 1, J^{[2]} \neq 0\right], \\
H^{[1(1,1),2(1,1)]}(t_1, t_2) &= \mathbb{P}\left[T_Y^{[1]} \leq t_1, T_Y^{[2]} \leq t_2, J^{[1]} = 1, J^{[2]} = 1\right], \\
H^{[1(1,1),2(1,2)]}(t_1, t_2) &= \mathbb{P}\left[T_Y^{[1]} \leq t_1, T_Y^{[2]} \leq t_2, J^{[1]} = 1, J^{[2]} = 2\right].
\end{aligned}$$

By definition of $F^{[2/1]}$, we have:

$$F^{[2/1]}(t_2/t_1) = \frac{F^{[1(1),2]}(t_1, t_2)}{F^{[1(1)]}(t_1)},$$

where

$$F^{[1(1),2]}(t_1, t_2) = \mathbb{P}\left[Y^{[1]} \leq t_1, Y^{[2]} \leq t_2, \mathcal{C}^{[1]} = 1\right].$$

We derive the following relation:

$$\begin{aligned}
H^{[1(1,1),2(1)]}(t_1, t_2) &= \mathbb{P}\left[Y^{[1]} \leq t_1, Y^{[1]} \leq C, Y^{[2]} \leq C - Y^{[1]}, Y^{[2]} \leq t_2, \mathcal{C}^{[1]} = 1\right] \\
&= \mathbb{P}\left[Y^{[1]} \leq t_1, Y^{[2]} \leq t_2, Y^{[1]} + Y^{[2]} \leq C, \mathcal{C}^{[1]} = 1\right] \\
&= \iiint I\left(u \leq t_1, v \leq t_2, c \geq u + v\right) G(dc) F^{[1(1),2]}(du, dv) \\
&= \iint_{u \leq t_1, v \leq t_2} \left(1 - G^-(u + v)\right) F^{[1(1),2]}(du, dv)
\end{aligned}$$

We deduce that

$$F^{[1(1),2]}(t_1, t_2) = \iint_{u \leq t_1, v \leq t_2} \frac{H^{[1(1,1),2(1)]}(du, dv)}{\left(1 - G^-(u + v)\right)}.$$

Finally, we get the relation:

$$F^{[2/1]}(t_2/t_1) = \frac{1}{F^{[1(1)]}(t_1)} \iint_{u \leq t_1, v \leq t_2} \frac{H^{[1(1,1),2(1)]}(du, dv)}{1 - G^-(u + v)}.$$

We can obtain in the same way

$$F^{[2(1)/1]}(t_2/t_1) = \frac{1}{F^{[1(1)]}(t_1)} \iint_{u \leq t_1, v \leq t_2} \frac{H^{[1(1,1),2(1,1)]}(du, dv)}{1 - G^-(u + v)},$$

and

$$F^{[2(2)/1]}(t_2/t_1) = \frac{1}{F^{[1(1)]}(t_1)} \iint_{u \le t_1, v \le t_2} \frac{H^{[1(1,1),2(1,2)]}(du, dv)}{1 - G^-(u+v)}.$$

As before, we propose "plug-in" estimators for $F^{[2/1]}$, $F^{[2(1)/1]}$ and $F^{[2(2)/1]}$. These estimators are obtained by replacing $H^{[1(1,1),2(1,1)]}$, $H^{[1(1,1),2(1,2)]}$ and $H^{[1(1,1),2(1)]}$ by their corresponding bivariate empirical distribution functions $H_n^{[1(1,1),2(1,1)]}$, $H_n^{[1(1,1),2(1,2)]}$ and $H_n^{[1(1,1),2(1)]}$ defined for $t_1, t_2 \ge 0$ by:

$$H_n^{[1(1,1),2(1)]}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \le t_1, T_{Y,i}^{[2]} \le t_2, J_i^{[1]} = 1, J_i^{[2]} \ne 0\right),$$

$$H_n^{[1(1,1),2(1,1)]}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \le t_1, T_{Y,i}^{[2]} \le t_2, J_i^{[1]} = 1, J_i^{[2]} = 1\right),$$

$$H_n^{[1(1,1),2(1,2)]}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \le t_1, T_{Y,i}^{[2]} \le t_2, J_i^{[1]} = 1, J_i^{[2]} = 2\right).$$

Moreover, the function G is replaced by the Kaplan-Meier estimator $\widehat{G}_n$ defined in (5.1) while $F^{[1(1)]}$ is replaced $\widehat{F}_n^{[1(1)]}$ which was seen to be consistent in Theorem 24. Consequently, we let for $t_1, t_2 \ge 0$:

$$\widehat{F}_n^{[2/1]}(t_2/t_1) = \frac{1}{\widehat{F}_n^{[1(1)]}(t_1)} \iint_{u \le t_1, v \le t_2} \frac{H_n^{[1(1,1),2(1)]}(du, dv)}{1 - \widehat{G}_n^-(u+v)},$$

$$\widehat{F}_n^{[2(1)/1]}(t_2/t_1) = \frac{1}{\widehat{F}_n^{[1(1)]}(t_1)} \iint_{u \le t_1, v \le t_2} \frac{H_n^{[1(1,1),2(1,1)]}(du, dv)}{1 - \widehat{G}_n^-(u+v)},$$

$$\widehat{F}_n^{[2(2)/1]}(t_2/t_1) = \frac{1}{\widehat{F}_n^{[1(1)]}(t_1)} \iint_{u \le t_1, v \le t_2} \frac{H_n^{[1(1,1),2(1,2)]}(du, dv)}{1 - \widehat{G}_n^-(u+v)}.$$

Theorem 25. *The estimators $\widehat{F}_n^{[2/1]}(t_2/t_1)$, $\widehat{F}_n^{[2(1)/1]}(t_2/t_1)$ and $\widehat{F}_n^{[2(2)/1]}(t_2/t_1)$ defined above are consistent respectively for $F^{[2/1]}(t_2/t_1)$, $F^{[2(1)/1]}(t_2/t_1)$ and $F^{[2(2)/1]}(t_2/t_1)$ for $0 \le t_1 \le \sigma_1$ and $0 \le t_2 \le \sigma_2$ such that $\sigma_1, \sigma_2 \ge 0$ and $\sigma_1 + \sigma_2 < \tau_H$.*

The proof runs as the proof of Theorem 24 and is therefore omitted.

Remark 8. *If t_1 is taken equal to ∞, by definition of the estimator $\widehat{F}_n^{[1(1)]}$, one would have $\widehat{F}_n^{[1(1)]}(\infty)$ equal to $\widehat{F}_n^{[1(1)]}(T_{Y,n,n}^{[1]})$ where $T_{Y,n,n}^{[1]}$ is the last order statistic of the sample $(T_{Y,1}^{[1]}, \ldots, T_{Y,n}^{[1]})$. The convergence $\widehat{F}_n^{[1(1)]}(T_{Y,n,n}^{[1]}) \to F^{[1(1)]}(\tau_{1(1)} \wedge \tau_G)$ holds in probability where $\tau_{1(1)}$ is the right-endpoint of $F^{[1(1)]}$ and where τ_G is the right-endpoint of G. But $F^{[1(1)]}(\tau_{1(1)} \wedge \tau_G)$ may be strictly inferior to $F^{[1(1)]}(\infty)$. This is fulfilled in particular if $\tau_G < \tau_{1(1)}$ which is the case in a clinical trial for example where this is to hope that many patients will survive beyond the end of study. This situation would lead to a biased estimation of $F^{[1(1)]}(\infty)$. As the quantity $\widehat{F}_n^{[1(1)]}(\infty)$ normalizes the estimators $\widehat{F}_n^{[2/1]}(./\infty)$,*

$\widehat{F}_n^{[2(1)/1]}(./\infty)$ and $\widehat{F}_n^{[2(2)/1]}(./\infty)$, a bad estimation of $F^{[1(1)]}(\infty)$ leads to bad estimations of $F^{[2/1]}(./\infty)$, $F^{[2(1)/1]}(./\infty)$ and $F^{[2(2)/1]}(./\infty)$. Consequently, this situation should be avoid.

5.4.4 Estimation with the k-th duration

The expression of the preceding estimators can be straightforwardly generalized to obtain an estimator for the distribution function of the k-th duration. The preceding method can be straightforwardly transposed to the present situation. Therefore, details are omitted.

Consider the multivariate empirical distribution functions defined for $t_1, \ldots, t_k \geq 0$ by:

$$
\begin{aligned}
&H_n^{[1(1,1),\ldots,k-1(1,1),k(1)]}(t_1, \ldots, t_k) \\
&\qquad = \frac{1}{n} \sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t_1, \ldots, T_{Y,i}^{[k]} \leq t_k, J_i^{[1]} = \cdots = J_i^{[k-1]} = 1, J_i^{[k]} \neq 0\right), \\
&H_n^{[1(1,1),\ldots,k-1(1,1),k(1,1)]}(t_1, \ldots, t_k) \\
&\qquad = \frac{1}{n} \sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t_1, \ldots, T_{Y,i}^{[k]} \leq t_k, J_i^{[1]} = \cdots = J_i^{[k]} = 1\right), \\
&H_n^{[1(1,1),\ldots,k-1(1,1),k(1,2)]}(t_1, \ldots, t_k) \\
&\qquad = \frac{1}{n} \sum_{i=1}^{n} I\left(T_{Y,i}^{[1]} \leq t_1, \ldots, T_{Y,i}^{[k]} \leq t_k, J_i^{[1]} = \cdots = J_i^{[k-1]} = 1, J_i^{[k]} = 2\right).
\end{aligned}
$$

For $t_1, \ldots, t_k \geq 0$, let:

$$
\widehat{p}_1(t_1) = \widehat{F}_n^{[1(1)]}(t_1),
$$

and for $t_1, \ldots, t_k \geq 0$ and $k \geq 2$, let:

$$
\widehat{p}_k(t_1, \ldots, t_k) = \widehat{F}_n^{[1(1)]}(t_1) \times \widehat{F}_n^{[2(1)/1]}(t_2/t_1) \times \cdots \times \widehat{F}_n^{[k(1)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}).
$$

We propose the estimators $\widehat{F}_n^{[k/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1})$, $\widehat{F}_n^{[k(1)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1})$ and $\widehat{F}_n^{[k(2)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1})$ defined for $t_1, \ldots, t_k \geq 0$ by:

$$
\begin{aligned}
&\widehat{F}_n^{[k/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) = \\
&\qquad \frac{1}{\widehat{p}_{k-1}(t_1, \ldots, t_{k-1})} \int \cdots \int_{u_1 \leq t_1, \ldots, u_k \leq t_k} \frac{H_n^{[1(1,1),\ldots,k-1(1,1),k(1)]}(du_1, \ldots, du_k)}{1 - \widehat{G}_n^-(u_1 + \cdots + u_k)}, \\
&\widehat{F}_n^{[k(1)/1,\ldots,k-1]}(t_k/t_1, \ldots, t_{k-1}) = \\
&\qquad \frac{1}{\widehat{p}_{k-1}(t_1, \ldots, t_{k-1})} \int \cdots \int_{u_1 \leq t_1, \ldots, u_k \leq t_k} \frac{H_n^{[1(1,1),\ldots,k-1(1,1),k(1,1)]}(du_1, \ldots, du_k)}{1 - \widehat{G}_n^-(u_1 + \cdots + u_k)},
\end{aligned}
$$

and

$$\widehat{F}_n^{[k(2)/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1}) = \frac{1}{\widehat{p}_{k-1}(t_1,\ldots,t_{k-1})}\int\cdots\int_{u_1\leq t_1,\ldots,u_k\leq t_k}\frac{H_n^{[1(1,1),\ldots,k-1(1,1),k(1,2)]}(du_1,\ldots,du_k)}{1-\widehat{G}_n^-(u_1+\cdots+u_k)}.$$

Theorem 26. *The estimators* $\widehat{F}_n^{[k/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1})$, $\widehat{F}_n^{[k(1)/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1})$ *and* $\widehat{F}_n^{[k(2)/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1})$ *are consistent respectively for* $F^{[k/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1})$, $F^{[k(1)/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1})$ *and* $F^{[k(2)/1,\ldots,k-1]}(t_k/t_1,\ldots,t_{k-1})$ *for* $(t_1\ldots,t_k)$ *in* $\prod_{i=1}^k[0,\sigma_i]$ *such that for* $i=1,\ldots,k$, $\sigma_i\geq 0$ *and* $\sum_{i=1}^k\sigma_i<\tau_H$.

The proof of is omitted.

5.5 Simulations

To investigate the behaviour of the proposed estimators, we conduct a Monte-Carlo study. We generate 1000 replications of the sample

$$\left(\left(T_{Y,i}^{[k]},J_i^{[k]}\right)_{i=1,\ldots,K_i}\right)_{i=1,\ldots,n}.$$

For each replication of this sample, we compute for the first three occurrences the estimators derived above. Figures 5.5, 5.6, 5.7 and 5.8 in Section 5.7 represent the curves obtained by taking the mean estimators over the 1000 replications. We also display for each of these estimators a pointwise empirical confidence bands with a 95% confidence level. For that, at each point, we compute the value of the estimator for each replication of the sample. The 2.5% of lower values and the 2.5% of higher values are dropped away. The remaining lower and higher values give empirical bounds.

To obtain each realization of the above sample, we do as follows. For the i-th patient, we generate the censoring time C_i either according to an uniform law or according to an exponential law. Then, we generate i.i.d. couples of r.v. $(Y_{1,i}^{[k]},Y_{2,i}^{[k]})$ according to an ACBVW(α,β,λ_0,λ_1,λ_2) model (see Chapter 3, Subsection 3.4.1) and set

$$Y_i^{[k]}=\min\left(Y_{1,i}^{[k]},Y_{2,i}^{[k]}\right)\quad\text{and}\quad \mathcal{C}_i^{[k]}=\begin{cases}1 & \text{if } Y_{1,i}^{[k]}\leq Y_{2,i}^{[k]},\\ 2 & \text{otherwise,}\end{cases}$$

until $\mathcal{C}_i^{[k]}=2$ since the occurrence of a terminal event stops the further occurrence of the recurrent events. Finally, the generated data for the i-th individual consists of $(T_{Y,i}^{[k]},J_i^{[k]})$ for $k=1,\ldots,K_i$ where

$$\left(T_{Y,i}^{[k]}=\min\left(Y_i^{[k]},C-\sum_{l=0}^{k-1}Y_i^{[l]}\right),J_i^{[k]}=\mathcal{C}_i^{[k]}I\left(Y_i^{[k]}\leq C-\sum_{l=0}^{k-1}Y_i^{[l]}\right)\right)$$

until $J_i^{[k]} = 0$ or $J_i^{[k]} = 2$ since the occurrence of a censoring event or of a terminal event stops the further occurrence of the recurrent events. For the i-th patient, the r.v. $T_{X,i}^{[K_i]}$ is given by the sum $T_{X,i}^{[K_i]} = \sum_{j=1}^{K_i} T_{Y,i}^{[j]}$. With this method, the successive marked duration times are independent. But, for each recurrence, the risk of experiencing a NFIE is neither independent (unless $\lambda_0 = 0$) nor proportional (unless $\alpha = \beta$) to the risk of dying.

Let $(T_{Y,1,n}^{[1]}, \ldots, T_{Y,n,n}^{[1]})$ be the order statistics of the sample $(T_{Y,1}^{[1]}, \ldots, T_{Y,n}^{[1]})$ with concomitant statistics $(J_{1,n}^{[1]}, \ldots, J_{n,n}^{[1]})$ i.e. if $T_{Y,i,n}^{[1]} = T_{Y,j}^{[1]}$ then $J_{i,n}^{[1]} = J_j^{[1]}$. In the same way, let $(T_{X,1,n}^{[K_{1,n}]}, \ldots, T_{X,n,n}^{[K_{n,n}]})$ be the order statistics of the sample $(T_{X,1}^{[K_1]}, \ldots, T_{X,n}^{[K_n]})$ with concomitant statistics $(J_{1,n}^{[K_{1,n}]}, \ldots, J_{n,n}^{[K_{n,n}]})$.

Let n_2 denote the number of patients still alive and still in the study after the occurrence of the first event. Let $(T_{Y,i_1}^{[2]}, \ldots, T_{Y,i_{n_2}}^{[2]})$ denote the observed second duration times (among the patients still alive and still in the study after the first event) with $1 \leq n_2 \leq n$ and $1 \leq i_1 < \cdots < i_{n_2} \leq n$. Let $(T_{Y,i_1,i_{n_2}}^{[2]}, \ldots, T_{Y,i_{n_2},i_{n_2}}^{[2]})$ denote the corresponding order statistics and let $(J_{i_1,i_{n_2}}^{[2]}, \ldots, J_{i_{n_2},i_{n_2}}^{[2]})$ denote the corresponding concomitant statistics.

For the first occurrence, we compute

$$\widehat{F}_n^{[1(1)]},\ \widehat{F}_n^{[1(2)]} \text{ and } \widehat{F}_n^{[1]}.$$

For the second occurrence, we compute

$$\widehat{F}_n^{[2(1)/1]}\left(./T_{Y,n,n}^{[1]}\right),\ \widehat{F}_n^{[2(2)/1]}\left(./T_{Y,n,n}^{[1]}\right) \text{ and } \widehat{F}_n^{[2/1]}\left(./T_{Y,n,n}^{[1]}\right).$$

At last, for the third occurrence, we compute

$$\widehat{F}_n^{[3(1)/1,2]}\left(./T_{Y,n,n}^{[1]}, T_{Y,i_{n_2},i_{n_2}}^{[2]}\right),\ \widehat{F}_n^{[3(2)/1,2]}\left(./T_{Y,n,n}^{[1]}, T_{Y,i_{n_2},i_{n_2}}^{[2]}\right) \text{ and } \widehat{F}_n^{[3/1,2]}\left(./T_{Y,n,n}^{[1]}, T_{Y,i_{n_2},i_{n_2}}^{[2]}\right).$$

We outline here to compute $\widehat{G}_n$, $\widehat{F}_n^{[1(1)]}$ and $\widehat{F}_n^{[2/1]}$. The first step consists in computing, under the initial condition $\widehat{G}_n(0) = 0$, for $i = 1, \ldots, n$:

$$\widehat{G}_n\left(T_{X,i,n}^{[K_i]}\right) = \widehat{G}_n\left(T_{X,i-1,n}^{[K_i]}\right) + \frac{I\left(J_{i,n}^{[K_{i,n}]} = 0\right)}{n\left(1 - \widehat{F}_{n,Z}\left(T_{X,i-1,n}^{[K_i]}\right)\right)},$$

where $\widehat{F}_{n,Z}$ the Kaplan-Meier estimator of the lifetime distribution function F_Z is computed from $\widehat{F}_{n,Z}(0) = 0$ and

$$\widehat{F}_{n,Z}\left(T_{X,i,n}^{[K_i]}\right) = \widehat{F}_{n,Z}\left(T_{X,i-1,n}^{[K_i]}\right) + \frac{I\left(J_{i,n}^{[K_{i,n}]} \neq 0\right)}{n\left(1 - \widehat{G}_n\left(T_{X,i-1,n}^{[K_i]}\right)\right)}.$$

Then, under the initial condition $\widehat{F}_n^{[1(1)]}(0) = 0$, we compute for $i = 1, \ldots, n$:

$$\widehat{F}_n^{[1(1)]}\left(T_{Y,i,n}^{[1]}\right) = \widehat{F}_n^{[1(1)]}\left(T_{Y,i-1,n}^{[1]}\right) + \frac{I\left(J_{i,n}^{[1]} = 1\right)}{n\left(1 - \widehat{G}_n^{-}\left(T_{Y,i,n}^{[1]}\right)\right)}.$$

Define $T^{[2]}_{Y,i_0,i_{n_2}}$ as 0. Under the initial condition $\widehat{F}^{[2/1]}_n(0/T^{[1]}_{Y,n,n}) = 0$, we compute for $j = 1, \ldots, n_2$:

$$\widehat{F}^{[2/1]}_n\left(T^{[2]}_{Y,i_j,i_{n_2}}/T^{[1]}_{Y,n,n}\right) = \widehat{F}^{[2/1]}_n\left(T^{[2]}_{Y,i_{j-1},i_{n_2}}/T^{[1]}_{Y,n,n}\right) + \frac{I\left(J^{[1]}_{i_j,i_n} = 1, J^{[2]}_{i_j,i_{n_2}} \neq 0\right)}{n\,\widehat{F}^{[1(1)]}_n\left(T^{[1]}_{Y,n,n}\right)\left(1 - \widehat{G}^{-}_n\left(T^{[1]}_{Y,i_j,i_n} + T^{[2]}_{Y,i_j,i_{n_2}}\right)\right)}.$$

All the simulation study has been implemented in `R` and the main functions are listed and commented in Appendix.

The behavior of the preceding estimators is investigated for $n = 1000$ under different scenarios. We choose the following values for the parameters of the ACBVW model and two censoring distributions, uniform and exponential leading to four cases:

- case 1: ACBVW(1.5,3,0.1,0.05,0.1) and $\mathcal{U}(0,5)$
- case 2: ACBVW(1.5,3,0.1,0.2,0.04) and $\mathcal{U}(0,3)$
- case 3: ACBVW(1.5,3,0.1,0.05,0.1) and $\mathcal{E}\text{xp}(0.2)$
- case 4: ACBVW(1.5,3,0.1,0.2,0.04) and $\mathcal{E}\text{xp}(1.2)$

The main characteristics of the corresponding simulated data are summarized on Table 5.1. In this table, $nb^{[k(1,1)]}$, $nb^{[k(1,2)]}$ and $nb^{[k(0)]}$ are respectively the mean number of patients experiencing a NFIE as k-th event, respectively dying and being censored. The maximum number of observed events is greater than three but the data at an occurrence greater than three are too parsimonious to be used. The obtained curves are displayed in Section 5.7 on Figures 5.5 (case 1), 5.6 (case 2), 5.7 (case 3) and 5.8 (case 4).

The data generated from an ACBVW(1.5,3,0.1,0.05,0.1) model for the marked duration times correspond to a high risk of death at each occurrence while the data generated from an ACBVW(1.5,3,0.1,0.2,0.04) model for the marked duration times correspond to a moderate risk of death at each occurrence.

The data generated with a $\mathcal{U}(0,5)$ law as censoring distribution correspond to a high level of censoring while the data generated with a $\mathcal{U}(0,3)$ law as censoring distribution correspond to a very high level of censoring. The data generated with a $\mathcal{E}\text{xp}(0.2)$ law as censoring distribution correspond to a very mild level of censoring while the data generated with a $\mathcal{E}\text{xp}(1.2)$ law as censoring distribution correspond to a moderate level of censoring. Note that the censoring level should not be confused with the proportion of censored observations. The notion of censoring level is outlined for example in Hall and Wellner (1980). The censoring level may be high and the proportion of censored observations moderate (as in case 1). On the opposite, the censoring level may be moderate and the proportion of censored observations high (as in case 4).

A first thing to notice is the fact that, when the number of occurrences grows, the pointwise empirical interval widen. This reflects the fact that the variability in the generated samples increases as the number of occurrences grows.

Secondly, we see that the different curves are very well estimated for small time values. But, if the censoring level is too high, the estimations are biased in the tail distributions. The higher the censoring level is, the more biased the estimators are. This is explained by the fact that the small observations are favored with respect to the large observations by the censoring phenomenon. On the opposite, even if the proportion of censored observations is high, the estimators are correct if the censoring level is moderate (case 4). When the censoring level is high (case 2), the estimators perform poorly even if the number of patients still alive and still under study is correct.

5.6 Application to real data: the CAPRIE study

We apply the method described in this chapter to the Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events study (CAPRIE). The data have been provided by Gérard Derzko (Sanofi-Aventis).

5.6.1 Description of the CAPRIE study

CAPRIE was a large, randomized, triple-blinded, international, multicenter trial designed to assess the relative efficacy of clopidogrel (an inhibitor of platelet aggregation) and aspirin (an anticoagulant) in reducing the risk of a composite outcome cluster of ischaemic stroke, myocardial infarction, or vascular death. The relative safety of clopidogrel and aspirin was also assessed. The studied population comprised 3 subgroups of patients with atherothrombotic disease manifested as either recent ischaemic stroke (IS), recent myocardial infarction (MI) or symptomatic peripheral arterial disease (PAD). 19185 patients, with more than 6300 in each of the clinical subgroups, were recruited over 3 years, with a mean follow-up of 1.91 years. The treatment groups were well matched with respect to age, sex, race and cardiovascular risk factors.

CAPRIE is the first study of antiplatelet drug to include patients from the clinical subgroups of ischaemic cerebrovascular, cardiac and peripheral arterial disease under a common protocol. Indeed, in a study on prevention, separations within and amongst clinical subgroups are not necessary because the underlying condition is atherothrombosis which can become clinically manifest in different ways. This approach can be justified by the common aetiology because many patients have experienced one manifestation when they present to medical attention with another, because of the consistency of the effect of the antiplatelet drugs across clinical subgroups and because in necropsy studies, many patients who have atherosclerosis in one part of the body are found to have it in others.

Different outcome events were observed during the study. NFIE were non-fatal IS, non-fatal MI, non-fatal primary intracranial hemorrhage and leg amputation. Deaths

were classified as fatal IS, fatal MI, hemorrhagic death, other vascular death or non-vascular death. The classification of fatal IS or MI was based on either death within 28 days after the onset of signs or symptoms of the acute outcome event, in the absence of other clear causes or on necropsy findings. Other vascular deaths were any deaths that were not clearly non-vascular and did not meet the criteria for fatal stroke, fatal myocardial infarction or hemorrhage. Deaths considered by the Central Validation Committee to be directly related to the qualifying event were classified as other vascular.

After randomization, 16 patients, 10 in the clopidogrel group and 6 in the aspirin group, were found to not to have the qualifying disease. Most were entered as having ischaemic stroke but were subsequently found to be misdiagnosed (e.g. as multiple sclerosis or primary intracranial hemorrhage). The study drug was terminated within 4 months of randomization for 13 of these 16 patients but the other 3 patients were continued on study drug. All 16 continued to be followed as per protocol and included in the analyses.

During the study, 42 patients (0.22%) were lost to follow-up, 22 in the clopidogrel group and 20 in the aspirin group. These 42 patients were included in the analyses with their follow-up censored at the last time of contact.

4059 (21.2%) patients had study drug permanently discontinued early, for reasons other than the occurrence of an outcome event, 21.3% in the clopidogrel and 21.1% in the aspirin group. Reasons for stopping study drug early were similar in the two groups: adverse events (11.4%), withdrawn consent (4.7%), contraindicated medications (2.4%), non-compliance (1.8%) and other (0.8%).

With exclusion of follow-up after any early permanent discontinuation of study drug, mean compliance with clopidogrel and aspirin was similar at 91%. 46 patients in the clopidogrel group and 40 in the aspirin group never took any study drug.

5.6.2 The results of the CAPRIE study

The results and conclusions obtained from CAPRIE study are recalled here. Further information may be found in Gent et al. (1996).

Two analytical strategies were planned: an *intention-to-treat* analysis in which all patients randomized were considered at risk to their planned end of study, irrespective of their compliance with study protocol and an *on-treatment* analysis in which an at risk patient's time was censored 28 days after early permanent discontinuation of study drug.

The primary analysis of efficacy was by intention-to-treat and was based on the first occurrence of an event in the outcome cluster of IS, MI or vascular death among all patients randomized.

Four secondary analyses were carried out. The first secondary outcome cluster included amputation in addition to IS, MI or vascular death. A further analysis was based on vascular death only. The third secondary analysis assessed an overall net benefit with the outcome cluster of any IS, MI or death from any cause. Indeed, even if the main focus was on events presumed to be due to atherosclerotic disease, primary intracranial hemorrhage

and fatal bleeds were possible adverse events, so these events were also included in the comparison. The fourth secondary analysis assessed all-cause mortality.

Safety assessments were based on the proportion of patients experiencing one or more episodes of a specific adverse event.

Similar analyses were carried out for each of the clinical subgroups even if CAPRIE was powered to detect a realistic treatment effect in the whole study cohort but not in each of the three clinical subgroups.

Assessments of relative efficacy were based on a comparison between the two treatments groups of the cumulative risk over time of each of the five prespecified outcomes. Survival curves based on the proportion of patients remaining event-free were estimated by the Kaplan-Meier method and compared by a two-sided Mantel-Haenzel test, stratified by clinical subgroups. In addition, to take into account any imbalances between the two treatments groups in baseline prognostic variables, analyses were repeated with adjustment procedures based on Cox's proportional hazards model.

The intention-to-treat analysis showed that patients treated with clopidogrel had an annual 5.32% risk of IS, MI or vascular death compared with 5.83% with aspirin. These rates reflect a statistically significant (p=0.043) relative-risk reduction of 8.7% in favour of clopigrel (95% CI: 0.3-16.5). Corresponding on-treatment analysis yielded a relative-risk reduction of 9.4%. There were no major differences in terms of safety. Reported adverse experiences in the clopidogrel and aspirin groups judged to be severe included rash (0.26% vs 0.10%), diarrhoa (0.23% vs 0.11%), upper gastrointestinal discomfort (0.97% vs 1.22%), intracranial haemorrhage (0.33% vs 0.47%) and gastrointestinal haemorrhage (0.52% vs 0.72%), respectively.

But the direction and the size of the effect was not homogeneous with the respect to three predefined stratified clinical subgroups (which are labeled as IS, MI or PAD). In fact, in the MI patients who were treated with clopidogrel, the risk of an outcome event was larger than in those treated with aspirin, although this is not statistically significant. Clopidogrel may be slightly better in preventing major ischaemic events in high-risk patients, but the results of CAPRIE suggest that there is a room for doubt, as there is no direct proof of an effect in patients with recent MI. However, aspirin may still be of value in the early treatment of acute stroke.

Long-term administration of clopidogrel to patients with atherosclerotic vascular disease is more effective than aspirin in reducing the combined risk of IS, MI or vascular death. The overall safety profile of clopidogrel is at least as good as that of medium-dose aspirin.

5.6.3 Contribution of the preceding methodology

Relying on the preceding results, we propose a novel analysis of the CAPRIE data. We suggest to compare the relative efficacy of clopidigrel and aspirin at the first and second occurrences of an ischaemic event by means of the estimators derived previously

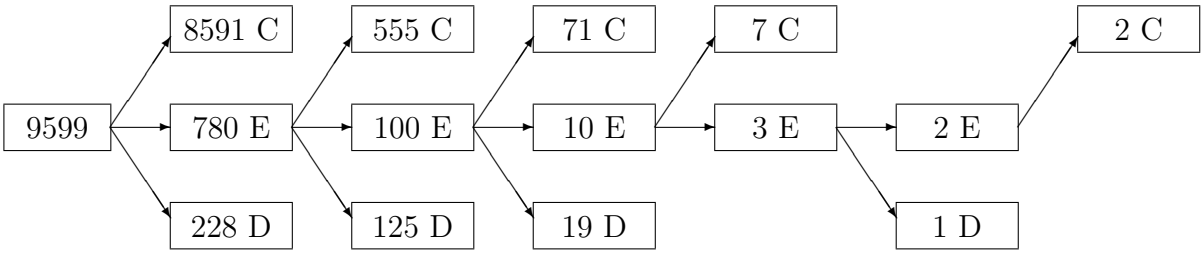

Figure 5.3: Number of patients randomized with clopidogrel experiencing a given event at a given occurrence (C: censored, E: having experienced a NFIE, D: dead).

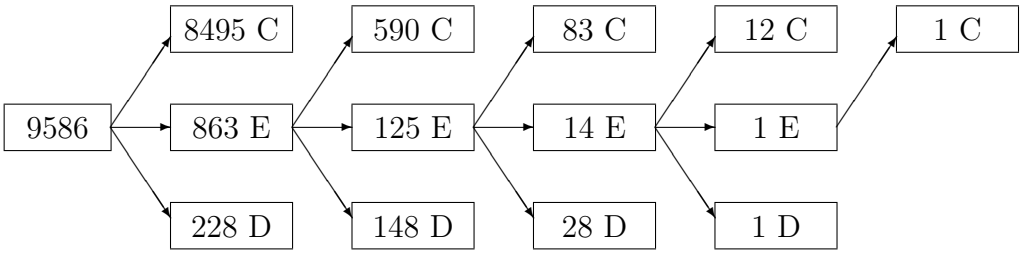

Figure 5.4: Number of patients randomized with aspirin experiencing a given event at a given occurrence (C: censored, E: having experienced a NFIE, D: dead).

for relevant cause-specific distribution functions. It turns out that the database records all the events experienced by each patient until death or censoring. The structure of the data is illustrated on Figure 5.3 for the patients randomized with clopidogrel and on Figure 5.4 for the patients randomized with aspirin. The data concerning third and further occurrences of ischaemic event are too parsimonious to be used.

Among the available information in the database, the following variables are used for each patient:

- the identification number of the patient,
- the treatment group (Clopidogrel: 1, Aspirin: 2),
- the state of health at the inclusion (MI: 1, IS: 2, PAD: 3),

and for each observed event:

- the duration time from inclusion to the occurrence of the event,
- the type of the observed event (MI: 1, IS: 2, Amputation: 3, Hemorrhage: 4, Vascular death: 5, Hemorrhagic death: 6, Non-vascular death: 7, End of study: 0).

NFIE is taken to be the cluster outcome of MI, IS, amputation and hemorrhage. Death (as fatal ischaemic event) is taken to be the outcome cluster of vascular death and hemorrhagic death. Censoring is taken to be the outcome cluster of non-vascular death and end of study.

We trace the curves pertaining to the estimators developed in this chapter for the first two occurrences of an event pertaining to NFIE or death or both. We also trace the (overall) mortality distribution curve (outcome cluster of any vascular death and hemorrhagic death) for each treatments by means of Kaplan-Meier estimator. Similar analyses are carried out for each of the clinical subgroups defined by the state at the inclusion (IS or MI or PAD). The curves are displayed in Section 5.7.

The cause-specific distribution functions pertaining respectively to the third occurrence of NFIE or death or both are undistinguishable from the horizontal axis and therefore are not displayed.

As found by the CAPRIE Steering Committee in Gent et al. (1996), the mortality incidence is very slightly inferior in patients treated with clopidogrel although this seems not significant. This result seems more significant in patients included in the strate PAD. In the same way, at the first occurrence, the incidence of an ischaemic event (fatal or not) is slightly inferior in patients treated with clopidogrel. But this seems not significant, except in the strate PAD where the result seems more significant. At the second occurrence, the incidence of an ischaemic event (fatal or not) is markedly inferior in patients treated with clopidogrel.

5.7 Graphical displays

5.7.1 Simulation results

We present here the graphics obtained under the four cases recalled below:

- case 1: ACBVW(1.5,3,0.1,0.05,0.1) and $\mathcal{U}(0,5)$
- case 2: ACBVW(1.5,3,0.1,0.2,0.04) and $\mathcal{U}(0,3)$
- case 3: ACBVW(1.5,3,0.1,0.05,0.1) and $\mathcal{E}\mathrm{xp}(0.2)$
- case 4: ACBVW(1.5,3,0.1,0.2,0.04) and $\mathcal{E}\mathrm{xp}(1.2)$

The main characteristics of the corresponding simulated data are summarized on Table 5.1. The curves are displayed on Figures 5.5 (case 1), 5.6 (case 2), 5.7 (case 3) and 5.8 (case 4). On these curves, the true distribution is drawn in black, the Monte-Carlo estimator is drawn in light blue and the pointwise empirical intervals are drawn in medium blue.

5.7.2 Analysis of the CAPRIE study

We present the analysis of the CAPRIE data on Figures 5.9 to 5.13. We show the censoring and the lifetime distributions together with the estimators developed previously for the first two occurrences. The curves are displayed for all patients in each treatment

	case 1	case 2	case 3	case 4
$nb^{[1(1,1)]}$	123	321	126	172
$nb^{[1(1,2)]}$	581	196	623	72
$nb^{[1(0)]}$	295	483	251	756
$nb^{[2(1,1)]}$	14	77	16	30
$nb^{[2(1,2)]}$	63	32	78	12
$nb^{[2(0)]}$	46	212	32	130
$nb^{[3(1,1)]}$	1	13	2	5
$nb^{[3(1,2)]}$	6	4	10	2
$nb^{[3(0)]}$	7	60	4	23

Table 5.1: Mean number of patients experiencing a given event for each of the first three occurrences, namely $nb^{[k(1,1)]}$ = mean number of patients experiencing a NFIE at the k-th occurrence, $nb^{[k(1,2)]}$ = mean number of patients dying at the k-th occurrence, $nb^{[k(0)]}$ = mean number of patients being censored at the k-th occurrence.

group and for each clinical subgroup (MI, IS or PAD) within each each treatment group. The curves pertaining to clopidogrel are drawn in black while the curves pertaining to aspirin are drawn in blue.

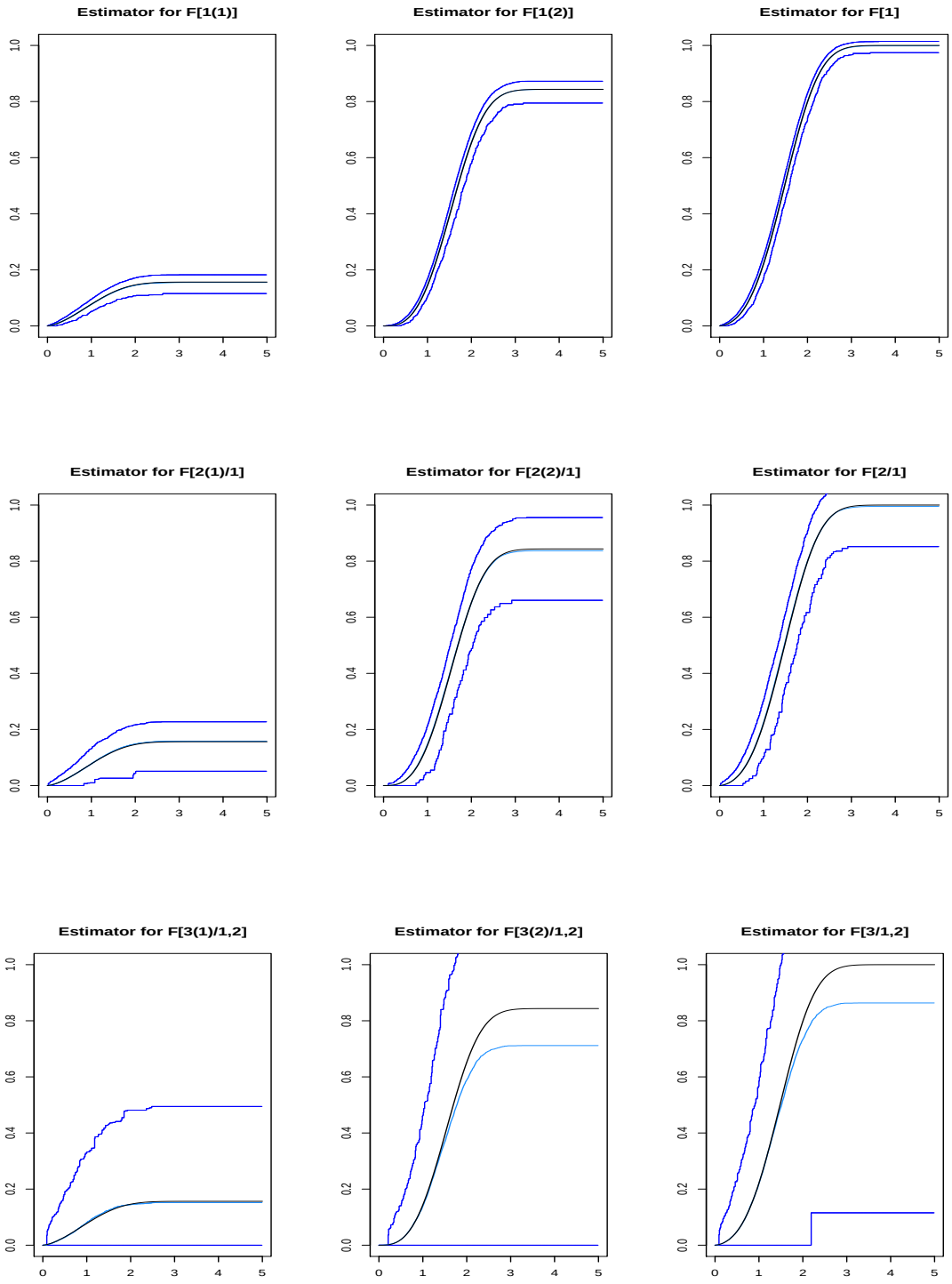

Figure 5.5: Monte-Carlo estimators and pointwise empirical confidence bands obtained in case 1 for the first three event occurrences.

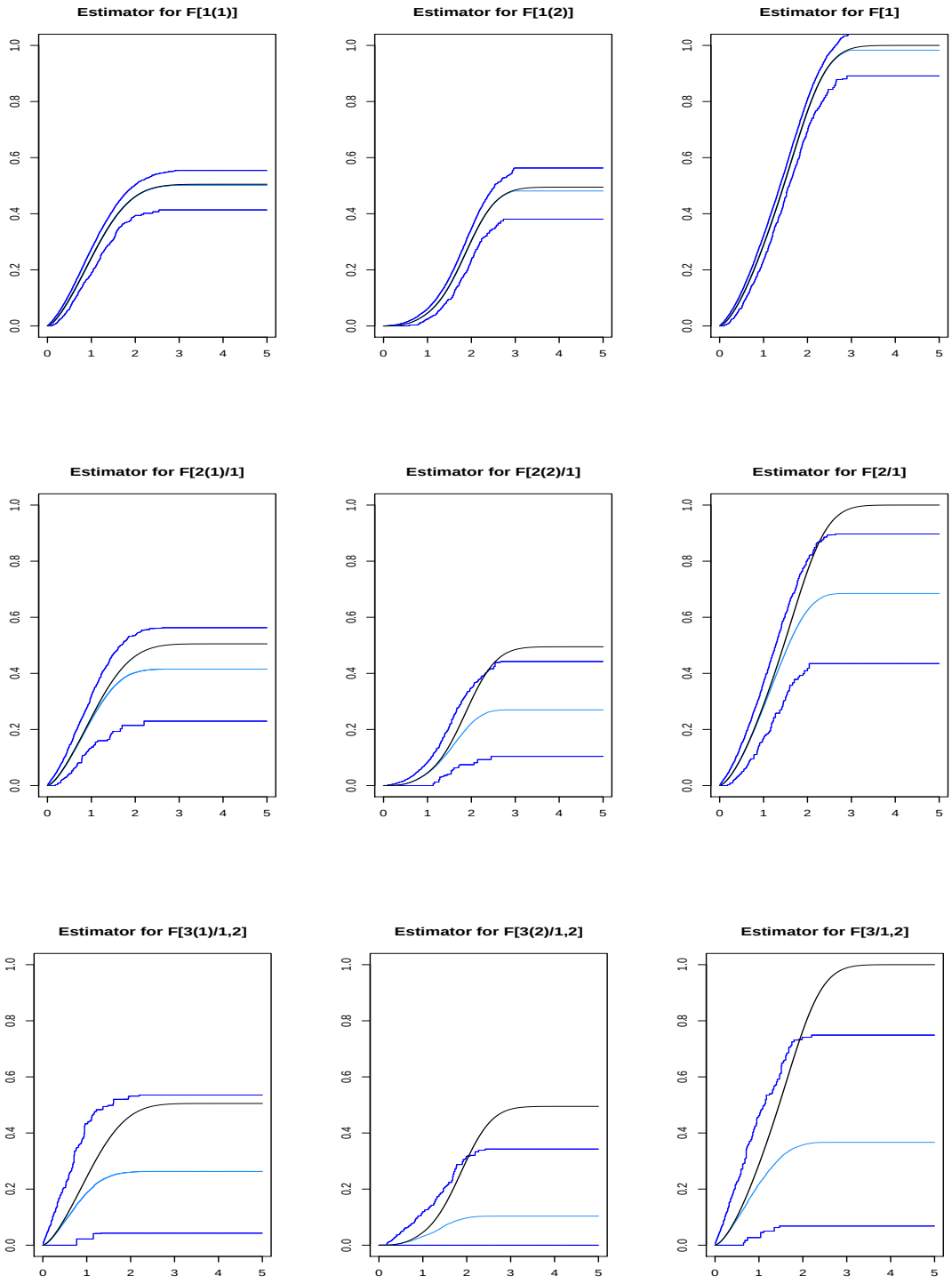

Figure 5.6: Monte-Carlo estimators and pointwise empirical confidence bands obtained in case 2 for the first three event occurrences.

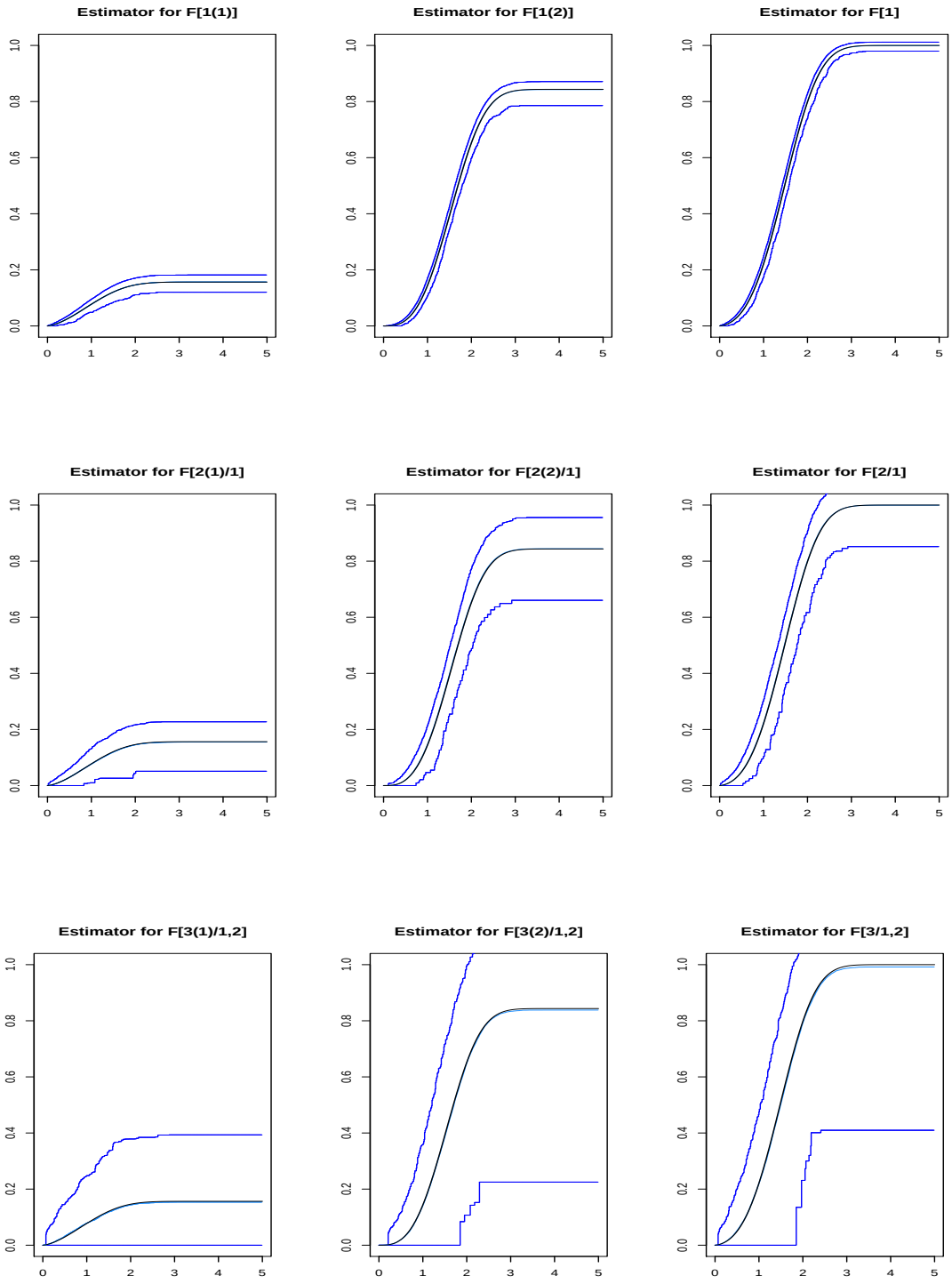

Figure 5.7: Monte-Carlo estimators and pointwise empirical confidence bands obtained in case 3 for the first three event occurrences.

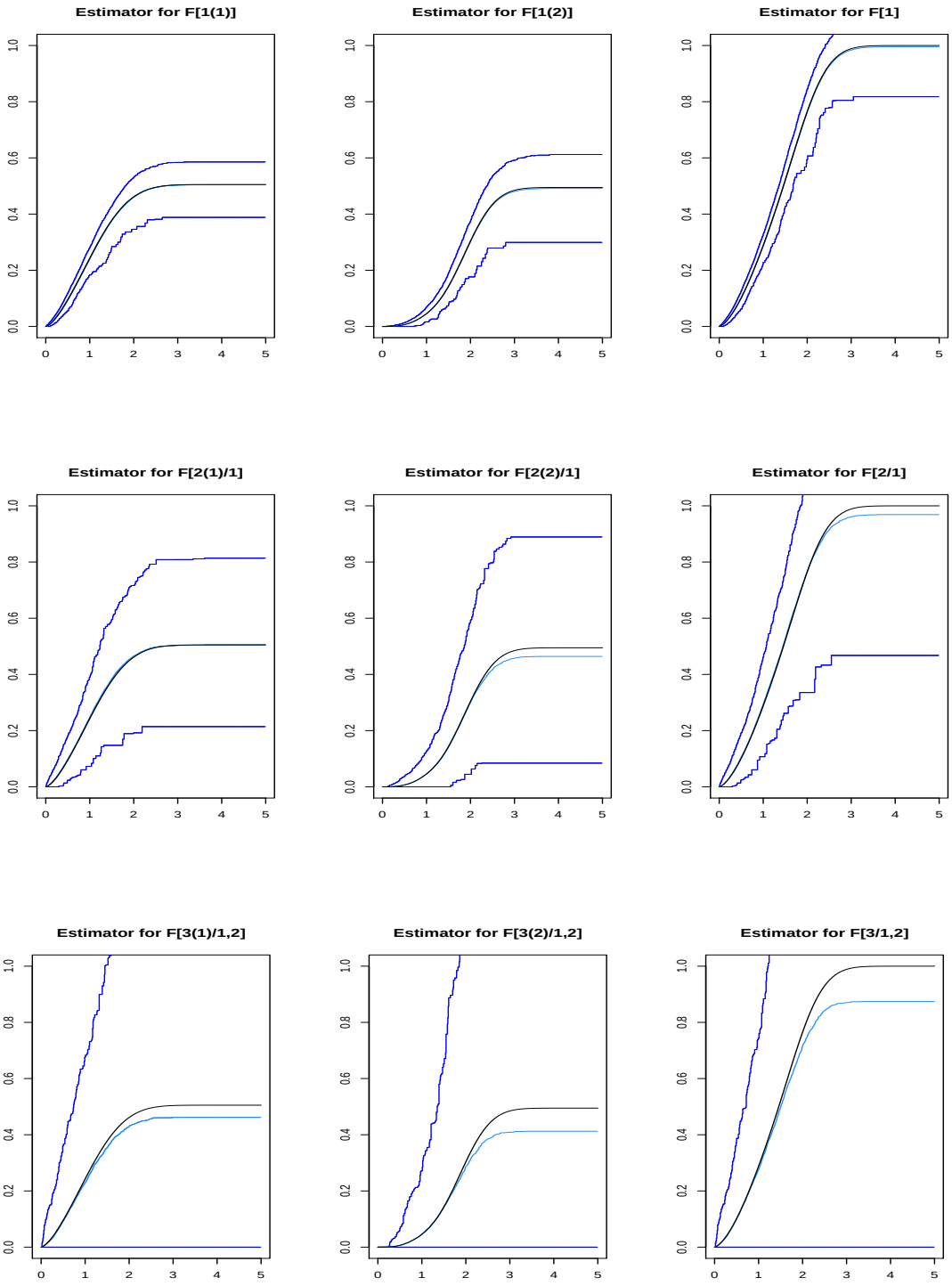

Figure 5.8: Monte-Carlo estimators and pointwise empirical confidence bands obtained in case 4 for the first three event occurrences.

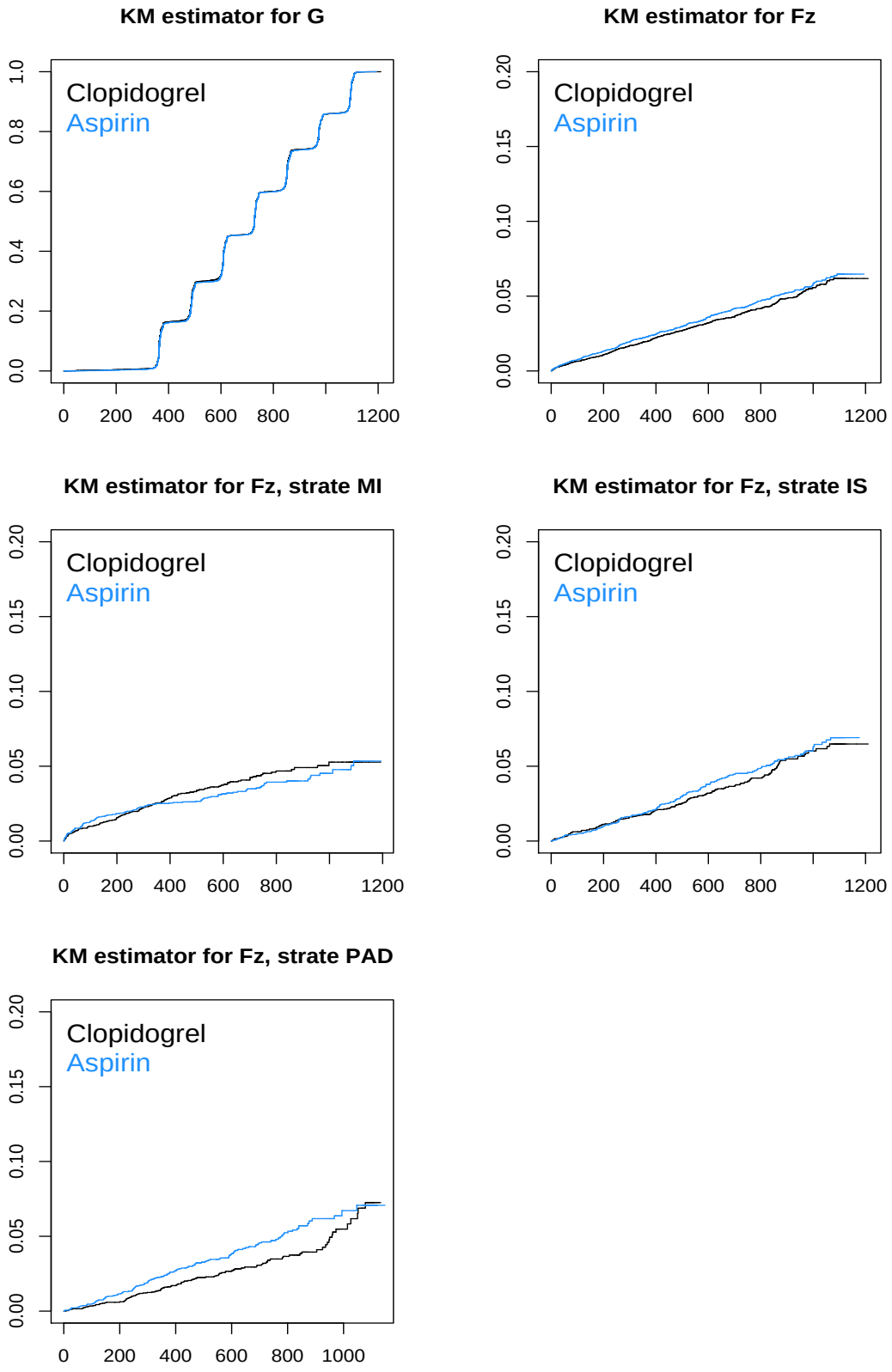

Figure 5.9: Representation of $\widehat{G}_n$ and $\widehat{F}_{n,Z}$.

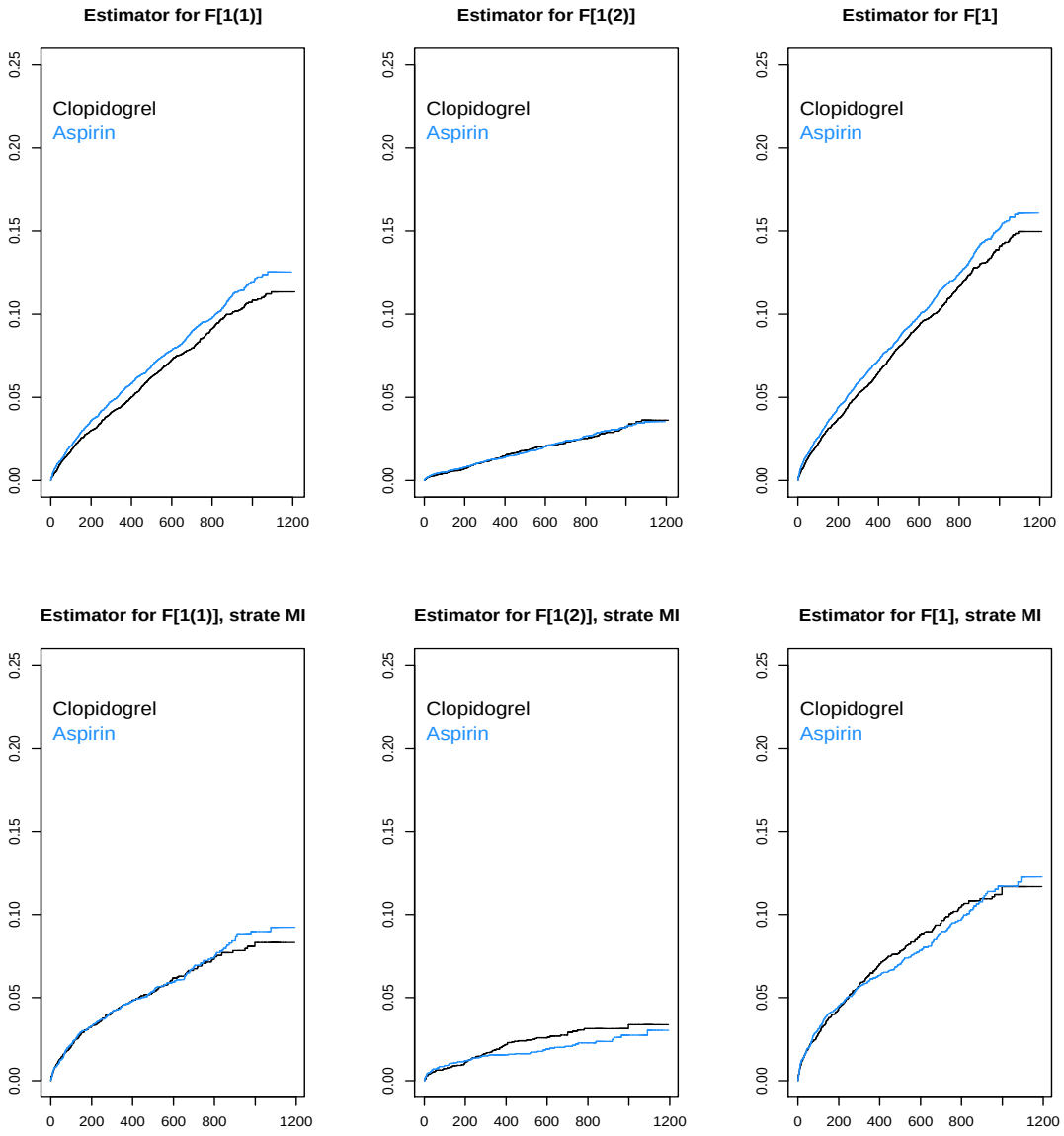

Figure 5.10: Representation of $\widehat{F}_n^{[1(1)]}$, $\widehat{F}_n^{[1(2)]}$ and $\widehat{F}_n^{[1]}$.

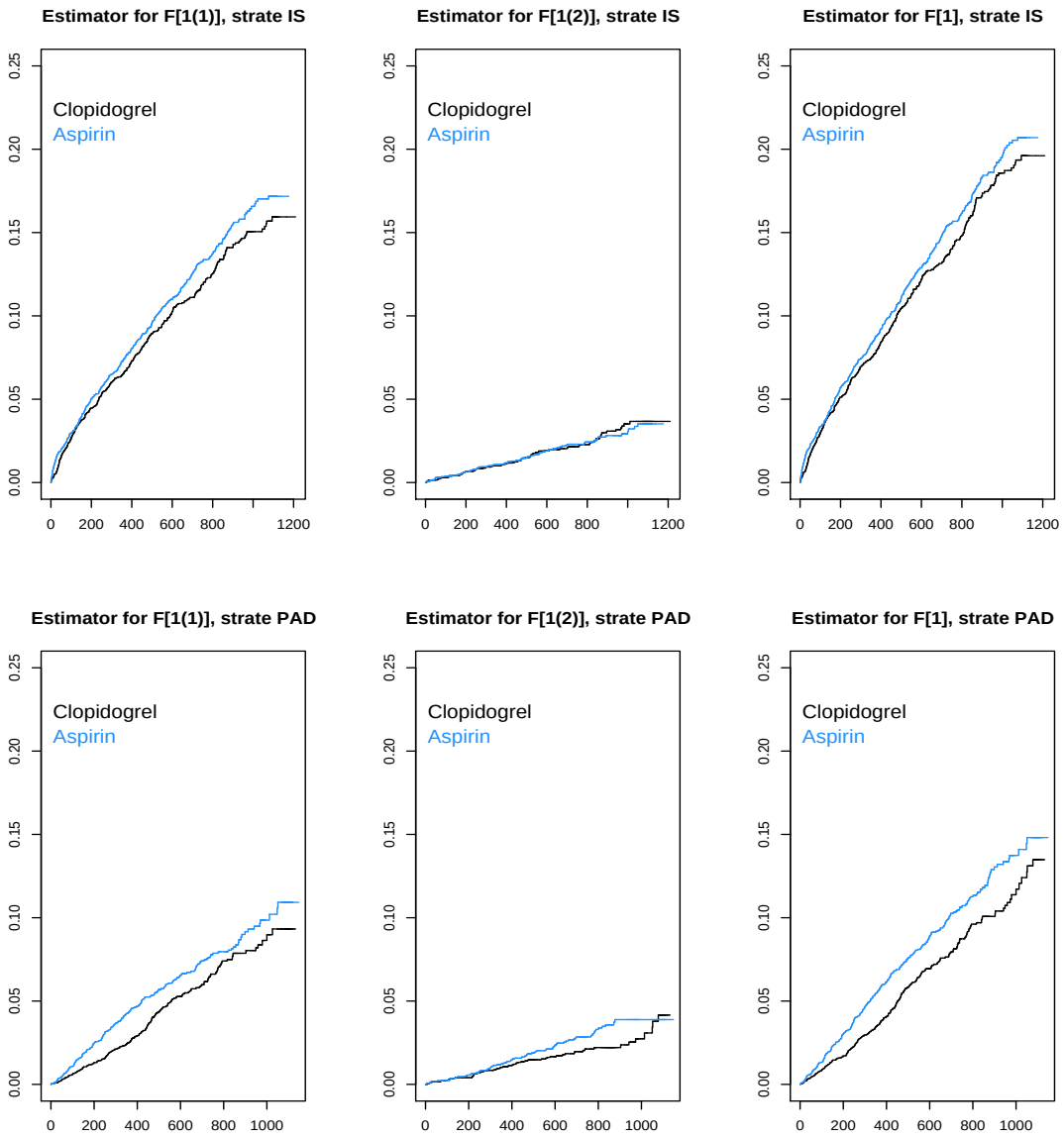

Figure 5.11: Representation of $\widehat{F}_n^{[1(1)]}$, $\widehat{F}_n^{[1(2)]}$ and $\widehat{F}_n^{[1]}$ (continued).

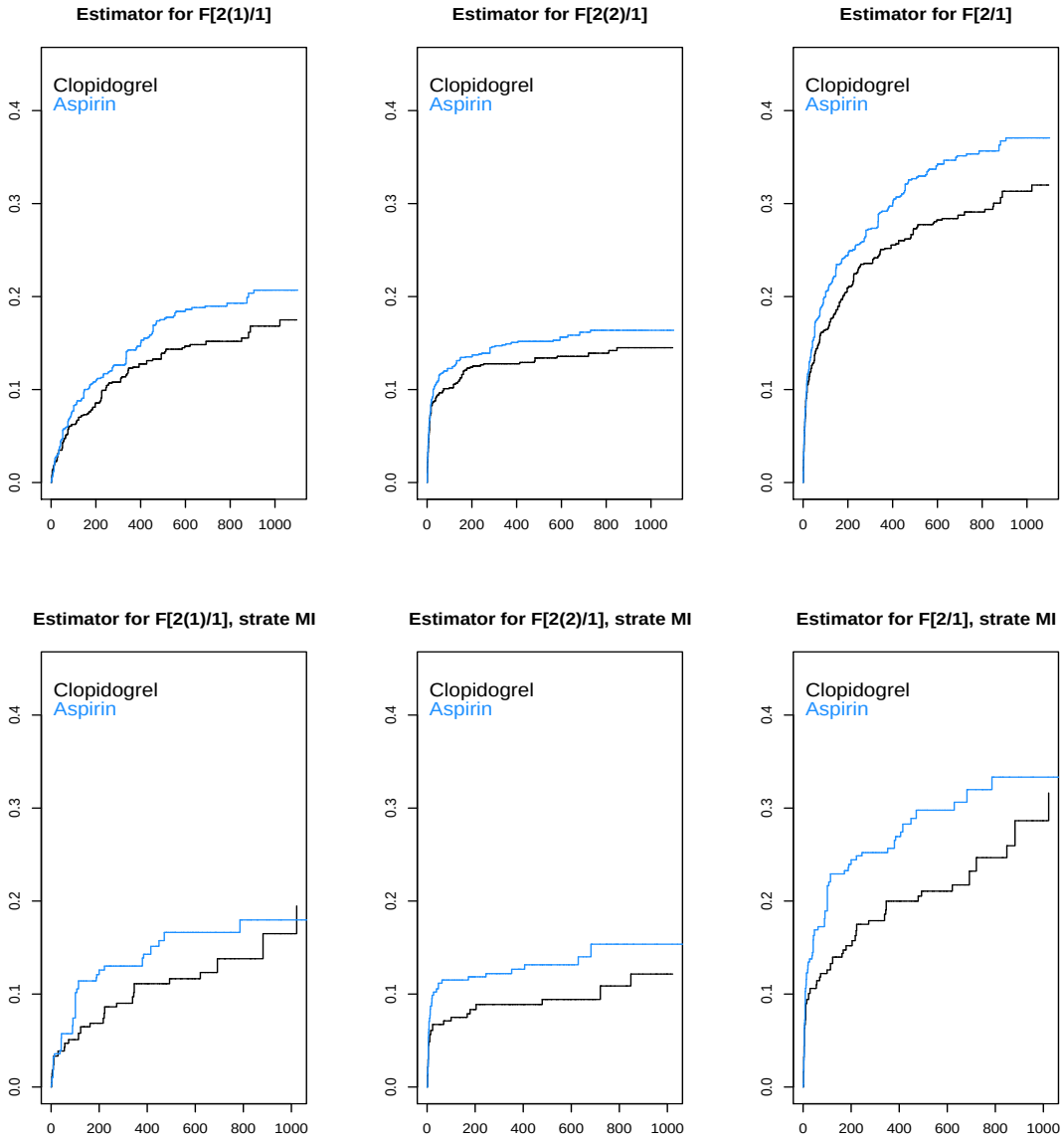

Figure 5.12: Representation of $\widehat{F}_n^{[2(1)/1]}(./a)$, $\widehat{F}_n^{[2(2)/1]}(./a)$ and $\widehat{F}_n^{[2/1]}(./a)$ where a is the minimun of the respective last order statistics of the two treatments.

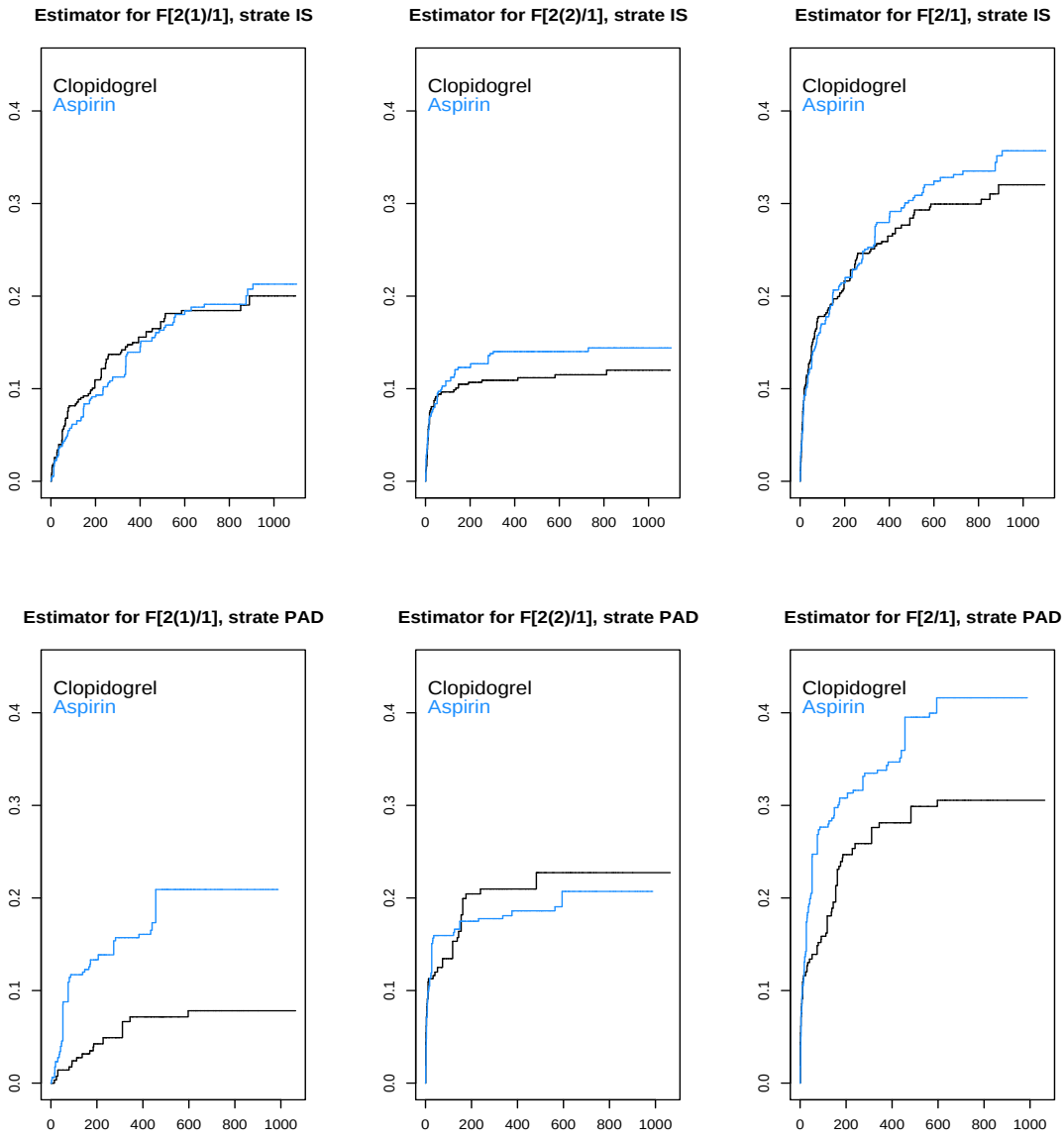

Figure 5.13: Representation of $\widehat{F}_n^{[2(1)/1]}(./a)$, $\widehat{F}_n^{[2(2)/1]}(./a)$ and $\widehat{F}_n^{[2/1]}(./a)$ where a is the minimun of the respective last order statistics of the two treatments (continued).

Conclusion et perspectives

Dans cette thèse, nous avons abordé différents aspects, théoriques comme appliqués, de l'analyse des durées de vie. Nous résumons ci-dessous les résultats obtenus.

Dans les chapitres 2 et 3, nous avons étudié les propriétés asymptotiques de l'estimateur non-paramétrique de Aalen et Johansen (1978) pour les fonctions de répartition spécifiques à une cause donnée dans le cadre de risques concurrents non indépendants en présence de censure aléatoire droite indépendante. Des approximations fortes pour des processus basés sur les estimateurs de Aalen et Johansen (1978) ont été développées ainsi que des résultats du type loi du logarithme itéré. Des bandes de confiance asymptotiques pour les fonctions de répartition spécifiques à une cause donnée ont été construites à partir d'un résultat de convergence faible pour des processus basés sur les estimateurs de Aalen et Johansen (1978). Ces bandes de confiance ont été simulées. A cette occasion, le modèle paramétrique ACBVW a été utilisé.

Dans le chapitre 4, nous avons supposé la proportionnalité entre les différentes fonctions de répartition spécifiques à une cause donnée. Nous avons alors proposé un nouvel estimateur non-paramétrique pour les fonctions de répartition spécifiques à une cause donnée en présence de censure aléatoire droite indépendante. Cet estimateur a des propriétés asymptotiques très similaires à celui de Aalen et Johansen (1978) comme le montrent des résultats du type loi du logarithme itéré, approximation forte et convergence faible. La différence essentielle réside dans le fait que notre nouvel estimateur est asymptotiquement plus efficace que l'estimateur de Aalen et Johansen (1978) lorsque la proportionnalité est respectée.

Dans le chapitre 5, nous nous sommes placés dans la situation où des évènements peuvent se produire de manière répétée chez un même individu en présence d'un évènement terminal et de censure aléatoire droite indépendante. Nous avons proposé des estimateurs non-paramétriques pour les fonctions de répartition associées à la durée entre deux évènements successifs conditionnellement à la survie aux évènements précédents. Nous avons montré que ces estimateurs sont fortement consistents sur le support de la censure. Le comportement à distance finie de ces estimateurs a été évalué par simulation. La méthodologie proposée a ensuite été mise en oeuvre sur les données de l'étude CAPRIE.

Les modèles étudiés et les résultats obtenus dans ce travail permettent d'envisager de nombreuses perspectives théoriques mais également pratiques. Nous citons ci-dessous différentes pistes de recherche à développer dans le futur.

Tout d'abord, dans le chapitre 3, nous avons vu les limites des bandes de confiance construites à partir de l'estimateur de Aalen et Johansen (1978). En effet, ces bandes de confiance nécessitent un grand nombre d'individus et sont larges. On pourra s'attacher à développer des bandes de confiance non-asymptotiques en employant une méthodologie différente. Par exemple, on pourra s'intéresser à une approximation en probabilité à distance finie pour les processus $K_n^{(j)}$ en se basant sur l'approximation poissonnienne à distance finie du processus empirique obtenue par Bretagnolle et Massart (1989).

La fonction quantile est une quantité utilisée dans de nombreuses applications : médecine, écologie, risques environnementaux... Il serait également intéressant de développer des résultats à distance finie pour le quantile des fonctions de répartition spécifiques à une cause donnée.

La grande flexibilité que permet le modèle paramétrique ACBVW introduit au chapitre 3 pour la modélisation de deux risques concurrents, dépendants ou non, proportionnels ou non, rend ce modèle très largement utilisable en pratique. Il serait alors intéressant d'écrire un test d'ajustement à ce modèle et de développer des estimateurs paramétriques pour les fonctions de répartition spécifiques à une cause donnée en présence de censure aléatoire droite.

Dans le chapitre 5, nous nous sommes intéressés, pour la seconde survenue d'accidents ischémiques, à l'estimation des fonctions $F^{[2/1]}$, $F^{[2(1)/1]}$ et $F^{[2(2)/1]}$. Ces fonctions présentent deux intérêts majeurs. D'une part, elles sont estimables au moyen d'un estimateur empirique simple à calculer. D'autre part, elles sont facilement représentables graphiquement. Cependant, du point de vue médical, ces fonctions ne sont peut-être pas les plus pertinentes. Il semblerait préférable de s'intéresser aux fonctions

$$\mathbb{P}\left[Y^{[2]} \leq t_2 \middle| Y^{[1]} = t_1, \mathcal{C}^{[1]} = 1\right],$$
$$\mathbb{P}\left[Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 1 \middle| Y^{[1]} = t_1, \mathcal{C}^{[1]} = 1\right],$$
$$\mathbb{P}\left[Y^{[2]} \leq t_2, \mathcal{C}^{[2]} = 2 \middle| Y^{[1]} = t_1, \mathcal{C}^{[1]} = 1\right].$$

Ces nouvelles fonctions prennent mieux en compte l'histoire médicale, déjà connue, du patient. Une fois obtenus des estimateurs pour ces nouvelles fonctions, on pourra bien sûr s'intéresser au développement d'un test non-paramétrique de comparaison de deux échantillons ainsi qu'à l'introduction de covariables.

Appendix A

Mathematical background

In this appendix, we describe briefly the main mathematical notions used in this work. For an extensive exposition, the reader is referred to the books of Billingsley (1968), Serfling (1980), Shorack and Wellner (1986), Andersen et al. (1993), Fleming and Harrington (1991), Kosorok (2005).

Probability spaces

Assume that Ω is the set of all outcomes of a random experiment with each outcome denoted generically by ω. The set Ω is called the *sample space*. Subsets of Ω are called *events*.

A collection of sets that comprises Ω itself and the primary events of interest together with unions and complements (and intersections too since the intersection of two sets is the complement of the union of their complements) forms a *field* or an *algebra*. A core concept in probability theory is that of *σ-field* (also called *σ-algebra*). A *σ-field* is a field in which unions are allowed to extend to countable number of events. If $\mathcal{G}$ is a family of subsets of Ω, we denote by $\sigma(\mathcal{G})$ the smallest σ-field containing $\mathcal{G}$ obtained by bringing in all unions and complements.

Suppose that a σ-field $\mathcal{F}$ is defined on Ω. A *probability measure* $\mathbb{P}$ assigns a number between 0 and 1 to each set in $\mathcal{F}$ according to the following rules $\mathbb{P}[\Omega] = 1$, $\mathbb{P}[\emptyset] = 0$ and $\mathbb{P}[\cup_i A_i] = \sum \mathbb{P}[A_i]$ for any countable collection of disjoint sets A_i in $\mathcal{F}$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ then constitutes a *probability space*.

A real-valued *random variable* (r.v.) X defined on a sample space Ω assigns a real number to each outcome ω in Ω i.e. $X(\omega) = x$ for some real x. Thus, the r.v. X imposes a structure on Ω corresponding to the different values that it takes. If X is discrete, the primary sets can be taken of the form $\{\omega \in \Omega : X(\omega) = x\}$ or just $\{X = x\}$ meaning the set of outcomes ω in Ω at which X takes the real value x. If X is continuous, the primary sets are usually taken to be of the form $\{X \leq x\}$.

The *σ-field generated by* the r.v. X is denoted $\sigma(X)$ and defined as the smallest σ-field containing the primary sets described above. It is obtained by bringing in all unions and complements.

Suppose that Y is a r.v. defined on Ω whose value is determined by that of X i.e. Y is a function of X. The subset $\{Y \leq y\}$ of Ω belongs then to $\sigma(X)$ for each y. This is because $\{Y \leq y\}$ can be expressed as $\{\omega \in \Omega : X(\omega) \in A_y\}$ where A_y is the set of values yielding $Y \leq y$. In this case, we say that the r.v. Y is *measurable with respect to* $\sigma(X)$ or equivalently that $\sigma(Y)$ is contained in $\sigma(X)$ meaning that $\sigma(X)$ defines a finer partition of Ω than $\sigma(Y)$, each set of $\sigma(Y)$ being a union of sets of $\sigma(X)$. Conversely, if Y is measurable with respect to $\sigma(X)$, the r.v. Y is constant on ω-sets on which X is constant and so is a function of X.

More generally, if $\{Y \leq y\}$ is in $\mathcal{F}$ for each y, where $\mathcal{F}$ is a σ-field defined on Ω, we say that Y is measurable with respect to $\mathcal{F}$ or just $\mathcal{F}$-measurable. If a probability measure $\mathbb{P}$ is defined on $\mathcal{F}$, then, provided that Y is $\mathcal{F}$-measurable, the quantity $\mathbb{P}[Y \leq y]$ is defined because $\{Y \leq y\}$ belongs to $\mathcal{F}$. In this way, the probability of any Y-event expressible in terms of basic $\{Y \leq y\}$-sets (by countable unions and complements) is computable.

Filtration

We must be able to talk mathematically about past and future. This is done through the concept of filtration or history. Let us model the occurrence in time of random events, in fact, discrete events occurring in continuous time. We fix a continuous-time interval $\mathcal{I} = [0, \tau[$ or $[0, \tau]$ for a given terminal time τ with $0 < \tau \leq \infty$. Note that the terminal time point τ may or may not be included.

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space where $\mathcal{F}$ and $\mathbb{P}$ are respectively a σ-field and a probability measure on Ω. The following definition allows the rigorous formulation of the concept of information accruing over time. A family of sub-σ-fields $(\mathcal{F}_t)_{t\in\mathcal{I}}$ of a σ-field $\mathcal{F}$ is a *filtration* if

1. it is *increasing* that is $s \leq t$ implies $\mathcal{F}_s \subseteq \mathcal{F}_t$ (i.e. $A \in \mathcal{F}_s$ implies $A \in \mathcal{F}_t$).
2. it is *right-continuous* i.e. for any t in $\mathcal{I}$, the σ-field $\mathcal{F}_{t^+}$ defined as $\cap_{s>t}\mathcal{F}_s$ fulfills $\mathcal{F}_{t^+} = \mathcal{F}_t$.

When $(\mathcal{F}_t)_{t\in\mathcal{I}}$ is a filtration, the σ-field $\mathcal{F}_t$ is interpreted as follows. It contains all events (up to null sets) whose occurrence is fixed by time t. Sometimes, we also consider the limit from the left $\mathcal{F}_{t^-}$ defined as $\sigma\left(\cup_{s<t}\mathcal{F}_s\right)$ (sometimes also denoted by $\bigvee_{s<t}\mathcal{F}_s$). It contains events fixed strictly before t.

A *stochastic basis* is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$ and is denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{I}}, \mathbb{P})$. A stochastic basis is called *complete* if $\mathcal{F}$ contains any subset of a $\mathbb{P}$-null set (so $\mathcal{F}$ is complete) and if $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{F}$.

Stochastic process

A *stochastic process* X is a mapping from $\mathcal{I} \times \Omega$ to $\mathbb{R}^d$. The set $\mathcal{I}$ indexes time. It is usually either $\{0, 1, 2, \dots\}$ for discrete-time processes or, for continuous-time process, a continuous-time interval $\mathcal{I} = [0, \tau[$ or $[0, \tau]$ for a given terminal time τ with $0 < \tau \leq \infty$. We restrict our attention to the continuous case. We write $X(t, \omega)$ for the realized value of the random variable $X(t)$ for the outcome $\omega \in \Omega$. The stochastic process X can be regarded as a time-indexed collection of random variables $(X(t))_{t\in\mathcal{I}}$ all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ i.e. as a function of ω for fixed t. But we can also think of X as a function of t for fixed ω. For fixed $\omega \in \Omega$, the random functions $X(., \omega) : \mathcal{I} \to \mathbb{R}$ are called the *sample paths* or *trajectories* of X.

If $X(.,.) : \mathcal{I}\times\Omega \to \mathbb{R}^d$ is a measurable mapping when $\mathbb{R}^d$ is equipped with the product Borel σ-field, then the process X is said to be *measurable*. In such cases, Fubini's theorem implies that the sample paths of X are measurable functions from $\mathcal{I}$ to $\mathbb{R}^d$. It can be shown that any process whose paths are right or left-continuous with probability one are measurable.

A process X is called *càdlàg* (continu à droite, limite à gauche) if its sample paths $(X(t, \omega))_{t\in\mathcal{I}}$ are right-continuous with left-hand limits for almost every ω. The set of càdlàg functions, often denoted by $D(\mathcal{I})$, is called the *Skorohod space* of weak convergence topology.

A stochastic process is called *right* or *left-continuous*, of *bounded variation*, *increasing* or said to *have limits from left or right* when the set of sample paths with the corresponding property has probability one. In addition, a stochastic process X is

- *bounded* if there exists a finite constant c such that $\mathbb{P}\left[\sup_{t\in\mathcal{I}} |X(t)| < c\right] = 1$.
- *integrable* if $\sup_{t\in\mathcal{I}} \mathbb{E}\left[|X(t)|\right] < \infty$.
- *square integrable* if $\sup_{t\in\mathcal{I}} \mathbb{E}\left[X(t)^2\right] < \infty$.
- *adapted* to the filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$ if $X(t)$ is $(\mathcal{F}_t)$-measurable for any t in $\mathcal{I}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$. A process is called *predictable* for the filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$ if it is measurable with respect to the σ-field on $\mathcal{I} \times \Omega$ generated by the adapted processes with left-continuous paths on $\mathcal{I}$. Equivalently, it can be shown that X is *predictable* if, as a mapping from $\mathcal{I} \times \Omega$ to $\mathbb{R}^d$, it it measurable with respect to the predictable σ-field defined as the σ-field generated by the sets of the form $]s, t] \times A$ with A in $\mathcal{F}_s$, s, t in $\mathcal{I}$ with $s < t$. This means that a predictable process is a process whose value at time t is fixed from the moment that we know what happened until time t^-.

The following properties are satisfied. Adapted processes with left-continuous paths on $\mathcal{I}$ are predictable. Deterministic measurable functions considered as a stochastic processes are predictable. If X is a $(\mathcal{F}_t)_{t\in\mathcal{I}}$-predictable process, then, for any t in $\mathcal{I}$, the r.v. $X(t)$ is $\mathcal{F}_{t^-}$-measurable.

A process X is called a *finite variation process* if its paths are of bounded variation i.e. $\int_{[0,t]} |dX(s)|$ is finite for t in $\mathcal{I}$ and for almost any ω in Ω. The process $\int |dX|$ is called its *(total) variation process.*

Filtration generated by a stochastic process

The most natural filtrations are histories of stochastic processes. Suppose that we are observing some stochastic process $\{X(s) : s \geq 0\}$. Define $\mathcal{F}_t$ as the σ-field generated by X over the time period $[0,t]$ i.e. $\mathcal{F}_t = \sigma(\{X(s) : 0 \leq s \leq t\})$. For $t \geq 0$, $\mathcal{F}_t$ is a partition of the sample space into sets each representing a possible sample path of X up to time t. Thus, given a set in $\mathcal{F}_t$, we know the values taken by $X(s)$ for s in $[0,t]$ i.e. the sample path. In this case, $\mathcal{F}_t$ contains the information generated by the process X on $[0,t]$ and is measurable. Assume that, as time goes on, no information is lost so that all that is present in $\mathcal{F}_s$ is included in $\mathcal{F}_t$ for $s \leq t$. Then $(\mathcal{F}_t)_{t\geq 0}$ is clearly a filtration. Obviously, any process is adapted to its history.

Frequently, we add to $\mathcal{F}_t$ events supposed to be fixed at time zero or equivalently add to the random variables generating $\mathcal{F}_t$ certain other random variables whose values are supposed to be fixed at time zero.

An important result of Courrège and Priouret (1965) states that the family of σ-fields generated by a right-continuous jump process is right-continuous, a jump process being a process such that, on intervals of the form $[t, t+\epsilon[$, the function $X(.,\omega)$ is a constant (depending on t and ϵ in general). It can be shown from this that filtrations generated naturally when considering discrete events occurring in continuous time are automatically right-continuous. Extending the filtration at time zero as just described leaves it right-continuous. In particular, completion of a filtration by adding null events to every $\mathcal{F}_t$ preserves right-continuity.

Stopping time, localization of stochastic processes

A *stopping time* T with respect to a filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$ is a r.v. taking values in $\bar{\mathcal{I}} = [0,\tau]$ such that $\{T \leq t\}$ is in $\mathcal{F}_t$ for t in $\mathcal{I}$. An increasing sequence of random times $(T_n)_n$ is called a *localizing sequence* with respect to the filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$ if T_n is a stopping time for each n and if $\lim_{n\to\infty} T_n = \tau$ a.s.. A property is said to hold *locally* for a stochastic process if the property is satisfied by the stopped process $X_n = \{X(t \wedge T_n) : t \in \mathcal{I}\}$ for each n where $(T_n)_n$ forms a localizing sequence. Let T be a stopping time. The σ-field of events prior to T, which is denoted by $\mathcal{F}_T$, consists of all sets A in $\bigvee_{t\in\mathcal{I}} \mathcal{F}_t$ such that $A \cap \{T \leq t\}$ is in $\mathcal{F}_t$ for t in $\mathcal{I}$.

A process X adapted to a filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$ with $X(0) = 0$ a.s. is locally bounded if and only if there exists an adapted left-continuous process H with right-hand limits such

that $|X(t)| \leq H(t)$ a.s. for any t in $\mathcal{I}$. Any $(\mathcal{F}_t)_{t\in\mathcal{I}}$-adapted left-continuous process X with right-hand limits with $X(0) = 0$ a.s. is locally bounded. Any càdlàg predictable process is locally bounded. Indeed, by predictability, the process can be stopped strictly before it exceeds a given value. In general, by localization, integrability conditions can be avoid altogether.

Martingales, local martingales and stochastic integrals with respect to a martingale

All the processes considered here are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

A continuous-time stochastic process $\{M(t) : t \in \mathcal{I}\}$ is a *martingale* (respectively a *submartingale*) with respect to the filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$, if M is adapted to $(\mathcal{F}_t)_{t\in\mathcal{I}}$ and càdlàg and if

1. M is integrable i.e. $\mathbb{E}[|M(t)|] < \infty$ for any t in $\mathcal{I}$,
2. $\mathbb{E}[M(t)|\mathcal{F}_s] = M(s)$ (respectively $\mathbb{E}[X(t)|\mathcal{F}_s] \geq X(s)$) for any s, t in $\mathcal{I}$ such that $s \leq t$.

Martingales can be considered as a kind of pure random noise process. A square integrable martingale may be extended to $\bar{\mathcal{I}} = [0, \tau_H]$ by taking $M(\tau_H) = \lim_{n\to\tau_H} M(t)$ which exists almost-surely.

A *local martingale* is a process M such that there exists a localizing sequence such that, for each n, the stopped process $M(. \wedge T_n)$ is a martingale. A local martingale is a càdlàg and adapted process. Similarly, a *local square integrable martingale* is a process M such that a localizing sequence can be chosen so that $M(. \wedge T_n)$ is a square integrable martingale for each n.

While a martingale is a pure random noise, a predictable process has a kind of regularity or semideterministic behavior. Many stochastic processes can be written as the sum of a (local) martingale and a finite variation predictable process. In such a decomposition (the celebrated Doob-Meyer decomposition), we are splitting the process into a random and a systematic part. The latter is called the compensator of the process because substracted of the process, a local martingale (unsystematic noise) is left. Written out in full, the Doob-Meyer decomposition states that, for a càdlàg adapted process X, the compensator $\tilde{X}$ of X, if it exists, is defined as the unique predictable, càdlàg and finite variation process such that $X - \tilde{X}$ is a local martingale, null at time zero. It turns out that X has a compensator if and only if X is the difference of two local submartingales.

Processes having a compensator are called *special semimartingales*. By definition, they can be written as the sum of a local martingale and a finite variation predictable process. A *semimartingale* (a more general object still) is the sum of a local martingale and a càdlàg adapted finite variation process (i.e. the finite variation part is not necessary predictable). Semimartingales have become a most important class of processes, but are not so relevant for our special purpose.

All (local) submartingales have compensators. In particular, nondecreasing, nonnegative locally integrable càdlàg processes have compensators (which are also nondecreasing) because such processes are trivially local submartingales.

Suppose that M is a (local) square integrable martingale on $\mathcal{I}$ with respect to a filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$. By Jensen's inequality, the process M^2 is a (local) submartingale. Consequently, there exists a unique increasing predictable process (the compensator of M^2) denoted by $\langle M\rangle$ with $\langle M\rangle(0) = 0$ a.s. and $\mathbb{E}[\langle M\rangle(t)] < \infty$ for any t such that $\{M(t)^2 - \langle M\rangle(t) : t \in \mathcal{I}\}$ is a (local) martingale. Usually, $\langle M\rangle$ is called the *predictable quadratic variation* of M or the *predictable variation process* of M. Here are two useful properties.

1. If (T_n) is a localizing sequence for M, then $\langle M\rangle(t) = \lim_{n\to\infty}\langle M(.\wedge T_n)\rangle(t)$.
2. If M is a martingale with $M(0) = 0$ a.s., then $\mathbb{E}[M^2(t)] = \mathbb{E}[\langle M\rangle(t)]$ for any t.

Assume that M_1 and M_2 are (local) square integrable martingales with respect to a filtration $(\mathcal{F}_t)_{t\in\mathcal{I}}$. The product of two (local) square integrable martingales can be written as a difference of two (local) square integrable martingales in the following way

$$M_1M_2 = \frac{1}{4}(M_1 + M_2)^2 - \frac{1}{4}(M_1 - M_2)^2$$

and therefore has a compensator of finite variation. Then there exists a unique predictable process $\langle M_1, M_2\rangle$ (locally) integrable with bounded variation with $\langle M_1, M_2\rangle(0) = 0$ a.s. such that $M_1M_2 - \langle M_1, M_2\rangle$ is a (local) martingale, null at zero. The process $\langle M_1, M_2\rangle$ is called the *predictable covariation process* of M_1 and M_2. The process $\langle M_1, M_2\rangle$ is càdlàg. If $M_1 = M_2 = M$, then $\langle M, M\rangle$ is denoted more simply by $\langle M\rangle$.

Stochastic integral

Stochastic integration is the forming of the integral of one stochastic process X with respect to another one, say Y.

When X is a locally bounded predictable process and Y is a càdlàg finite variation process, stochastic integration may be defined as a pathwise operation. For a given ω in Ω, one forms an ordinary Lebesgue-Stieltjes integral over a given time interval. The stochastic process $\int XdY$ is then given by

$$t \to \int_0^t X(s)dY(s) = \int_{[0,t]} X(s)dY(s)\,.$$

By convention, $Y(0^-) = 0$ so that, for $t = 0$, we get $\int_0^t XdY = X(0)Y(0)$.

If the process Y is not finite variation, stochastic integrals can not be defined pathwise. However, stochastic integrals with respect to processes which are not finite variation

processes can be defined as the limit in probability or in norm L^2 as n goes to infinity of Riemann sums defined by

$$\sum_{t_i \in \Delta_n} X_{t_i}(Y_{t_{i+1}} - Y_{t_i}) \,,$$

where Δ_n is a partition of the interval $[0,t]$ such that $\sup\{|t_{i+1} - t_i| : t_i \in \Delta_n\}$ goes to zero as n goes to infinity.

The stochastic integral $\int X dY$ has special and valuable properties when the integrand X is a predictable process H and when we integrate with respect to a process Y which is a (local) martingale M. Under the appropriate local integrability conditions, the resulting process $\int H dM$ is a (local) martingale in the sense L^2 or in probability. Moreover, its predictable variation process can be obtained from those of M.

Precisely, suppose that M is a square integrable martingale, that H is a predictable process and that

$$\mathbb{E}\left[\int_0^\tau H^2 d\langle M\rangle\right] < \infty \,.$$

Then, the process $\int H dM$ can be defined as the limit in the sense L^2 of the Riemann sums defined above. The process $\int H dM$ is a square integrable martingale and we have:

$$\left\langle \int H\, dM \right\rangle = \int H^2 d\langle M\rangle \ .$$

This stochastic integral is often called the *Itô integral*.

In the same way, suppose that M_1 and M_2 are square integrable martingales, that H_1 and H_2 are predictable processes and that

$$\mathbb{E}\left[\int_0^\tau H_1 d\langle M_1\rangle\right] < \infty,\ \mathbb{E}\left[\int_0^\tau H_2 d\langle M_2\rangle\right] < \infty \text{ and } \mathbb{E}\left[\int_0^\tau H_1 H_2 d\langle M_1, M_2\rangle\right] < \infty \,.$$

Then, $\int H_1 dM_1$ and $\int H_2 dM_2$ are square integrable martingales defined as limits of Riemann sums in the sense L^2 and we have:

$$\left\langle \int H_1\, dM_1, \int H_2 dM_2 \right\rangle = \int H_1 H_2 d\langle M_1, M_2\rangle \ .$$

Now, suppose that M is a local square integrable martingale, that H is a predictable process such that there exists a localizing sequence (T_n) satisfying

$$\mathbb{E}\left[\int_0^{T_n} H^2 d\langle M\rangle\right] < \infty \,.$$

Then, the process $\int H dM$ is defined as the limit in probability of the Riemann sums. The process $\int H dM$ is a local square integrable martingale and we have:

$$\left\langle \int H\, dM \right\rangle = \int H^2 d\langle M\rangle \ .$$

Now, suppose that M_1 and M_2 are local square integrable martingales, that H_1 and H_2 are predictable processes such that there exists a localizing sequence (T_n) satisfying

$$\mathbb{E}\left[\int_0^{T_n} H_1 d\langle M_1\rangle\right] < \infty,\ \mathbb{E}\left[\int_0^{T_n} H_2 d\langle M_2\rangle\right] < \infty \text{ and } \mathbb{E}\left[\int_0^{T_n} H_1 H_2 d\langle M_1, M_2\rangle\right] < \infty\,.$$

Then, $\int H_1 dM_1$ and $\int H_2 dM_2$ are local square integrable martingales defined as in probability limits of Riemann sums and we have:

$$\left\langle \int H_1\, dM_1, \int H_2 dM_2 \right\rangle = \int H_1 H_2 d\langle M_1, M_2\rangle\ .$$

Counting processes

A *counting process* N is an adapted, non-decreasing, right-continuous, integer-valued process with $N(0) = 0$ a.s. and with jumps of size $+1$ only. A *multivariate counting process* $(N^{(1)}, \ldots, N^{(d)})$ is a finite collection of counting processes, $N^{(k)}$ for $k = 1, \ldots, d$ with the additional restriction that two processes $N^{(k)}$ and $N^{(l)}$ for $k \neq l$ can not jump at the same time.

Let N be an arbitrary counting process. Then there exists a unique increasing predictable process A with $A(0) = 0$ a.s. such that the process $M = N - A$ is a local martingale which turns out to be a local square integrable martingale with $M(0) = 0$. The process A is the compensator or the *dual predictable projection* of N. In fact, the compensator of a counting process is its systematic part. It is a smoothly varying and predictable process which, subtracted from the counting process, leaves unpredictable zero-mean noise. The predictable variation process $\langle M\rangle$ of the local square integrable martingale M can be easily computed from the compensator

$$\langle M\rangle(t) = \int_0^t (1 - \Delta A) dA\,.$$

The same decomposition holds for multivariate counting processes. A multivariate counting process can be thought of as registering the occurrences in time of a number of types of disjoint, discrete events. If $\mathbf{N} = (N^{(1)}, \ldots, N^{(d)})$ is a d-dimensional counting process, then there exists increasing predictable processes $A^{(k)}$ null at time $t = 0$ and such that the $M^{(k)} = N^{(k)} - A^{(k)}$ for $k = 1, \ldots, d$ are local square integrable martingales. The $A^{(k)}$ are the respective compensators of the $N^{(k)}$.

Theorem 2.3.1. **of Gill (1980)** *If $M^{(k)} = N^{(k)} - A^{(k)}$ for $k = 1, \ldots, d$ is a sequence of square integrable martingales where $N^{(k)}$ is a counting process and $A^{(k)}$ is its compensator, then*

$$\begin{aligned}\left\langle M^{(k)}, M^{(l)}\right\rangle &= -\int \Delta A^{(k)} dA^{(l)} \quad \text{if } k \neq l\\ \left\langle M^{(k)}\right\rangle &= \int (1 - \Delta A^{(k)}) dA^{(k)} \quad \text{if } k = l\,.\end{aligned}$$

Vapnik-Červonenkis classes of functions

Empirical processes theory addresses the basic question of how well frequency (or sample mean) approaches probability (or expected value). The classical period, from the 1920s to the 1960s, considers the empirical distribution function in $\mathbb{R}$ and also in $\mathbb{R}^J$ (Glivenko and Cantelli, Kolmogorov and Smirnov, Cramér, Doob and Donsker, Kiefer, etc...) Vapnik and Červonenkis (1971) and Dudley (1978) sparked a revival by considerably broadening the scope of the theory.

The major distinction between empirical process theory and more standard asymptotics is that the random quantities studied have realizations as functions rather real numbers or vectors. An empirical process is a stochastic process based on a random sample. Consider a random sample $X_1, \ldots, X_n$ of i.i.d. real r.v. with distribution function F. The empirical distribution function is defined for t in $\mathbb{R}$ by:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t).$$

More generally, we can consider a random sample $X_1, \ldots, X_n$ of independent draws from a probability measure P on an arbitrary measure space $(S, \mathcal{S})$. We define the empirical measure as

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

where δ_x is the Dirac measure which assigns mass 1 at x and zero elsewhere. For a measurable function $f : S \to \mathbb{R}$, we define

$$P_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i)\,.$$

For any class $\mathcal{F}$ of measurable functions $f : S \to \mathbb{R}$, the set $\{P_n f, f \in \mathcal{F}\}$ is called the empirical process indexed by the class $\mathcal{F}$. This simple approach can generate a surprising variety of empirical processes.

Setting $S = \mathbb{R}$, we can re-express F_n as the empirical process $\{P_n f, f \in \mathcal{F}\}$ where $\mathcal{F} = \{I(x \leq t), t \in \mathbb{R}\}$. Thus one can view the stochastic process F_n as indexed either by t in $\mathbb{R}$ or $f \in \mathcal{F}$.

Vapnik and Červonenkis (1971) considered i.i.d. random variables $X_1, \ldots, X_n$ with values in a measurable space $(S, \mathcal{S})$ (which can be $\mathbb{R}$ but also $\mathbb{R}^J$, the sphere or a space of functions) and common law P. Then, they asked (and answered to a large extent) the following natural question. For what families $\mathcal{F}$ of measurable functions do we have

$$\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \to 0$$

in probability or almost surely ? If $S = \mathbb{R}$, the collection of all half-lines $]-\infty, x]$ is an example of a family of sets $\mathcal{C}$ with this property, by Glivenko-Cantelli theorem. So, Vapnik and Červonenkis (1971) replaced $(\mathbb{R}, \mathcal{B})$ by a general measurable space $(S, \mathcal{S})$. Similar

questions can be asked about the central limit theorem, the law of the iterated logarithm, exponential bounds, etc.

Let (T, d) be a metric space. For $\varepsilon > 0$, the *number of ε-covering* of (T, d) denoted by $N(T, d, \varepsilon)$ is the minimal number of open d-balls of radius ε and centers in T required to cover T. A class $\mathcal{E}$ of measurable functions on a measure space $(S, \mathcal{S})$ is a *Vapnik-Červonenkis (VC) class* of functions with respect to an *envelope* E if there exists a measurable function E everywhere finite such that $|f| \leq E$, for all f in $\mathcal{E}$ and if there exists finite numbers A and v such that for ε in $]0, 1[$ and for all P probability measure on $(S, \mathcal{S})$ such that $\int E^2 dP < \infty$ we have:

$$N\left(\mathcal{E}, \|.\|_{L^2(P)}, \varepsilon \|E\|_{L^2(P)}\right) \leq \left(\frac{A}{\varepsilon}\right)^v .$$

A class $\mathcal{E}$ is called *measurable* if it can be parametrized by a complete separable metric space Θ and if the map $(\theta, x) \to f_\theta(x)$ is measurable.

We refer the reader to Van der Waart and Wellner (1996), for instance, for an exposition of the relevant results.

U-statistics

We present here some definitions and refer to Serfling (1980), Lee (1990) or Stute (1994b) for a detailed presentation about U-statistics and U-processes.

Let us first consider one-sample U-statistics. Let $\mathcal{P}$ be a family of probability measures on an arbitrary measurable space. By $\theta : P \to \theta(P)$, we denote a real-valued functional defined on $\mathcal{P}$. The functional $\theta(P)$ is called a *regular* or *estimable parameter* if there exists an unbiased estimate of it i.e. $\theta(P)$ may be represented as:

$$\theta(P) = \mathbb{E}_P[h(X_1, \ldots, X_m)] = \int h(x_1, \ldots, x_m) dP(x_1) \ldots dP(x_m)$$

for i.i.d. r.v. $X_1, \ldots, X_m$ with distribution function P and for some measurable function $h : X^m \to \mathbb{R}$ of m variables. The function h is often named *kernel*. The smallest integer m with the above property is called the *degree* of $\theta(P)$ or of the kernel h.

Without loss of generality, we may assume that h is symmetric. For, if not, it may be replaced by the symmetric kernel

$$\frac{1}{m!} \sum_p h(x_{i_1}, \ldots, x_{i_m})$$

where $\sum_p$ denotes the summation over the $m!$ permutations $(i_1, \ldots, i_m)$ of $(1, \ldots, m)$.

For any kernel h, the corresponding U-statistic for estimation of θ on the basis of the sample $X_1, \ldots, X_n$ of size $n \geq m$ is obtained by averaging the kernel h symmetrically over

the observations:

$$U_n = \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1}, \ldots, X_{i_m})$$

where $\sum_c$ denotes summation over the combinations of m distinct elements $\{i_1, \ldots, i_m\}$ among $\{1, \ldots, n\}$. The function h is called the *kernel* of the U-statistic and the integer m is called the *degree* of the U-statistic.

Clearly, U_n is an unbiased estimate of $\theta(P)$. Moreover it has the optimality property of being a best unbiased estimate of $\theta(P)$ if $\mathcal{P}$ is large enough. For example, if $\mathcal{P}$ contains all the distributions P for which $\theta(P)$ is finite, then the order statistics form a complete sufficient statistic from P in $\mathcal{P}$. As U_n is a symmetric function of $X_1, \ldots, X_n$, it is a function of the order statistics and so is a best unbiased estimate of $\theta(P)$ due to the Hodges-Lehmann theorem. This means that no other unbiased estimate of $\theta(P)$ based on $X_1, \ldots, X_n$ can have a smaller variance than U_n.

Let us present a few examples of U-statistics. Let $X_1, \ldots, X_n$ be i.i.d. r.v. with distribution function F.

Consider the functional $\theta(F) = \mu(F) = \int x dF(x)$. For the kernel $h(x) = x$, the corresponding U-statistic is the sample mean

$$U_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n \,.$$

Consider the functional $\theta(F) = \sigma^2(F) = \int (x - \mu)^2 dF(x)$. For the kernel given by $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$, the corresponding U-statistic is the sample variance

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j) = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2 \right) .$$

Consider the functional $\theta(F) = F(t_0) = \mathbb{P}_F[X_1 \le t_0] = \int_{-\infty}^{t_0} dF(x)$. For the kernel $h(x) = I(x \le t_0)$, the corresponding U-statistic is where F_n denotes the sample distribution function

$$U_n = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le t_0) = F_n(t_0) \,.$$

An appealing feature of a U-statistic is its simple structure as a sum of identically distributed random variables. However, except in the case of a kernel of degree $m = 1$, the summands are not all independent, so that a direct application of the abundant theory for sums of *independent* r.v. is not possible. But, by the special device of "projection", a U-statistic may be approximated within a sufficient degree of accuracy by a sum of i.i.d. random variables. In this way, classical limit theory for sums does carry over to U-statistics and yields the relevant asymptotic results.

The *projection* of the U-statistic U_n is defined as

$$\hat{U}_n = \sum_{i=1}^{n} \mathbb{E}_F[U_n|X_i] - (n-1)\theta\,.$$

Note that it is exactly a sum of i.i.d. random variables. This decomposition is therefore used to study asymptotic theory of U-statistics.

Corresponding to a U-statistic

$$U_n = \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1}, \ldots, X_{i_m})$$

for the estimation of $\theta(F) = \mathbb{E}_F[h]$, the associated *Von Mises statistic* is

$$V_n = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(X_{i_1}, \ldots, X_{i_m}) = \theta(F_n)\,.$$

We denote the empirical measure by P_n. Then, we can write V_n in the integral form $V_n = \int h(x_1, \ldots, x_m) dP_n(x_1) \ldots dP_n(x_m)$. Though P_n is an unbiased estimate of P, the functional V_n is not an unbiased estimate of $\theta(P)$. U-statistics often appear as components of higher order terms when expanding a Von Mises statistic.

Let X_i for $i \in \mathbb{N}$ be i.i.d. r.v. and common law P and let $\mathcal{F}$ be a collection of measurable functions $f : S^m \to \mathbb{R}$. The *U-process* of order m indexed by $\mathcal{F}$ is defined, for $n \in \mathbb{N}$, as:

$$U_n(f) = \frac{(n-m)!}{n!} \sum_{(i_1,\ldots,i_m)\in I_m^n} f(X_{i_1}, \ldots, X_{i_m}),\ f \in \mathcal{F},$$

where $I_m^n = \{(i_1, \ldots, i_m) : 1 \le i_j \le n \text{ and } i_j \ne i_k \text{ if } j \ne k\}$.

The extension of the preceding definitions to the case of several samples is straightforward. Consider k independent collections of independent observations $\{X_1^{(1)}, X_2^{(1)}, \ldots\}$,..., $\{X_1^{(k)}, X_2^{(k)}, \ldots\}$ taken from distributions $F^{(1)}, \ldots, F^{(k)}$, respectively. Let

$$\theta : (F^{(1)}, \ldots, F^{(k)}) \to \theta(F^{(1)}, \ldots, F^{(k)})$$

be a real-valued function such that there exists an unbiased estimator of $\theta(F^{(1)}, \ldots, F^{(k)})$. That is

$$\theta = \mathbb{E}[h(X_1^{(1)}, \ldots, X_{m_1}^{(1)}; \ldots; X_1^{(k)}, \ldots, X_{m_k}^{(k)})]$$

where h is assumed, without loss of generality, to be symmetric within each of its k blocks of arguments.

Corresponding to the kernel h, and assuming $n_1 \ge m_1, \ldots, n_k \ge m_k$, the U-statistic for the estimation of θ is defined as:

$$U_n = \frac{1}{\prod_{j=1}^{k} \binom{n_j}{m_j}} \sum_c h\left(X_{i_{11}}^{(1)}, \ldots, X_{i_{1m_1}}^{(1)}; \ldots; X_{i_{k1}}^{(k)}, \ldots, X_{i_{km_k}}^{(k)}\right).$$

Here, for $j = 1, \ldots, k$, the set $\{i_{j1}, \ldots, i_{jm_j}\}$ denotes a set of m_j distinct elements among $\{1, \ldots, n_j\}$ and $\sum_c$ denotes summation over all such combinations.

For example, let $\{X_1, \ldots, X_{n_1}\}$ and $\{Y_1, \ldots, Y_{n_2}\}$ be two independent i.i.d. samples drawn from a continuous distributions F and G respectively. Then, for $\theta = \theta(F, G) = \mathbb{P}[X \leq Y] = \int_{X(\Omega)\times Y(\Omega)} F dG$, an unbiased estimator is:

$$U_{n_1,n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_i \leq Y_j)\,.$$

In the case of k samplings, the multi-sample Von Mises functional has the form

$$V_{\bar{n}} = \frac{1}{\prod_{j=1}^{k} n_j^{m_j}} \sum h(X_{jl}, l = i_{j1}, \ldots, i_{jm_j}, j = 1, \ldots, k)$$

where summation is carried out over all indexes $1 \leq i_{jk} \leq n$ for $1 \leq k \leq m_j$ and $j = 1, \ldots, k$. With the help of the empirical measure, we can write $V_{\bar{n}}$ in the integral form

$$V_{\bar{n}} = \int h(x_{ij}, i = 1, \ldots, m_j, j = 1, \ldots, k) \prod_{j=1}^{k} \prod_{i=1}^{m_j} dP_{n_j}(x_{ij})$$

where P_{n_j} is the empirical measure corresponding to the j-th sampling.

Appendix B

Some R programs

In this appendix, we describe briefly the main programs used in the simulations. They have been implemented in R. We first present the programs used for the simulations of Chapter 3.

```
#-------------------------------------------------------------------------#
# Tracé des fonctions de répartition spécifiques à la cause et globale    #
# obtenues par modèle ACBVW                                               #
# lorsque l0=0, il y a indépendance entre les deux risques                #
# lorsque a=b=1, on retrouve un modèle ACBVE(l0,l1,l2)                    #
#-------------------------------------------------------------------------#

Fj<-function(j,x,a,b,L0,L1,L2)
{
 L<-L0+L1+L2
 aux1inf<-function(x)
  {
   aux<-x^(a-1)*exp(-(L1+L0)*x^a-L2*x^b)
   aux
   }
 aux1sup<-function(x)
  {
   aux<-x^(a-1)*exp(-L1*x^a-(L2+L0)*x^b)
   aux
   }
 aux2inf<-function(x)
  {
   aux<-x^(b-1)*exp(-(L1+L0)*x^a-L2*x^b)
   aux
   }
 aux2sup<-function(x)
  {
   aux<-x^(b-1)*exp(-L1*x^a-(L2+L0)*x^b)
   aux
```

```
   }
 if (j==1)
 {
 if ((0<=x)&(x<=1)) y<- L*(L1+L0)/(L1+L2)*a*integrate(aux1inf,0.000001,x)$value
                        +L0/(L1+L2)*(exp(-L*x^a)-1)
 if (x>1) y<- L*(L1+L0)/(L1+L2)*a*integrate(aux1inf,0.000001,1)$value
              +L0/(L1+L2)*(exp(-L)-1)+L1*L/(L1+L2)*a*integrate(aux1sup,1,x)$value
 }
 if (j==2)
 {
 if ((0<=x)&(x<=1)) y<-L*L2/(L1+L2)*b*integrate(aux2inf,0.000001,x)$value
 if (x>1) y<- L*L2/(L1+L2)*b*integrate(aux2inf,0.000001,1)$value
              +L*(L2+L0)/(L1+L2)*b*integrate(aux2sup,1,x)$value
              +L0/(L1+L2)*(exp(-L*x^b)-exp(-L))
 }
 y
}

F<-function(x,a,b,L0,L1,L2)
{
L<-L0+L1+L2
if (x<=1) y<-1-L/(L1+L2)*exp(-(L0+L1)*x^a-L2*x^b)+L0/(L1+L2)*exp(-L*x^a)
if (x>1) y<-1-L/(L1+L2)*exp(-L1*x^a-(L0+L2)*x^b)+L0/(L1+L2)*exp(-L*x^b)
y
}

aaa<-seq(0,6,0.01)
bb1<-numeric(length(aaa))
bb2<-numeric(length(aaa))
bb3<-numeric(length(aaa))
bb1[1]<-0
bb2[1]<-0
bb3[1]<-0
for (i in 2:length(aaa))
{ bb1[i]<-Fj(1,aaa[i],1.5,3,0.1,0.05,0.01)
  bb2[i]<-Fj(2,aaa[i],1.5,3,0.1,0.05,0.01)
  bb3[i]<-F(aaa[i],1.5,3,0.1,0.05,0.01)
 }
plot(aaa,bb3,type="l",xlab=" ",ylab=" ",main="ACBVW(1.5,3,0.1,0.05,0.01)")
lines(aaa,bb1,type="l",col="red")
lines(aaa,bb2,type="l",col="blue")


#-------------------------------------------------------------------------------#
# Détermination de la proportion théorique d'observations non censurées         #
# pour des données générées avec un modèle ACBVW et une censure exponentielle   #
#-------------------------------------------------------------------------------#
```

```
prop<-function(mu,a,b,L0,L1,L2)
{
 finf<-function(x)
    {
     L<-L0+L1+L2
     y <- (-L0)*L/(L1+L2)*a*x^(a-1)*exp(-L*x^a-mu*x)+L/(L1+L2)*((L0+L1)*a*x^(a-1)
          +L2*b*x^(b-1))*exp(-(L0+L1)*x^a-L2*x^b-mu*x)
     y
     }

 fsup<-function(x)
   {
    L <- L0+L1+L2
    y <- (-L0)*L/(L1+L2)*b*x^(b-1)*exp(-L*x^b-mu*x)+L/(L1+L2)*(L1*a*x^(a-1)
         +(L0+L2)*b*x^(b-1))*exp(-L1*x^a-(L0+L2)*x^b-mu*x)
    y
    }

  L<-L0+L1+L2
  prop<- integrate(finf,0.000001,1)$value+integrate(fsup,1,300)$value

  prop
}



#-------------------------------------------------------------------------------------#
# Tabulation de KB(lambda,a) et KW(lambda,1)                                          #
#-------------------------------------------------------------------------------------#

calculK<-function(lambda)
{
calcul<-numeric(301)
calcul[1]<- 1-2*pnorm(-lambda)
for (k in 2:301)
  { calcul[k]<- 2*(-1)^(k-1)*(pnorm(lambda*(1+2*(k-1)))-pnorm(lambda*(2*(k-1)-1)))}
calculfin<-sum(calcul)
calculfin
}

tabulationW<-function(ntot,unite)
{
Wtab<-matrix(0,unite*ntot+1,2)
for (i in 0:(unite*ntot))
 { Wtab[i+1,1]<- 1+i/ntot
   Wtab[i+1,2]<- calculK(1+i/ntot)
   }
Wtab
```

```
}


calculB<-function(lambda,a)
{
r<-lambda*sqrt((1-a)/a)
calcul<-numeric(301)
if (a<1)
 { calcul[1]<- 1-2*pnorm(-lambda/sqrt(a*(1-a)))
   for (k in 2:301)
      { calcul[k]<- 2*(-1)^{k-1}*exp(-2*lambda^2*(k-1)^2)
                    *(pnorm(r*(2*(k-1)+1/(1-a)))-pnorm(r*(2*(k-1)-1/(1-a)))) }
   calculfin<-sum(calcul)
 }
if (a==1)
 { calcul[1]<- 1
for (k in 2:301)
      { calcul[k]<- 2*(-1)^{k-1}*exp(-2*lambda^2*(k-1)^2)   }
   calculfin<-sum(calcul)
 }
calculfin
}

tabulationB<-function(ntot,unite)
{
Btab<-matrix(0,unite*ntot+1,92)
for (i in 1:(unite*ntot+1))
 { Btab[i,1]<- 0.150+(i-1)/ntot
   for (k in 2:92)
      { Btab[i,k]<- calculB(Btab[i,1],0.10+(k-2)/100) }
   }
Btab
}

matW<-tabulationW(1000,3)
matB<-tabulationB(1000,1.850)


#-------------------------------------------------------------------------------#
# Détermination des lambda pour un niveau de confiance donné                    #
#-------------------------------------------------------------------------------#

LambdaAN<-function(niveau,J,matW)
{
lambdaAN<-0
conf<-round(1-(1-niveau)/J,4)
for (i in 1:(dim(matW)[1]-1))
        { if ((matW[i,2]<conf)&(matW[i+1,2]>=conf)) lambdaAN<-matW[i+1,1]}
```

```
lambdaAN
}

LambdaHW<-function(niveau,J,sigma,esti,matB)
{
lambdaHW<-numeric(J)
borne<-numeric(J)
col<- numeric(J)
conf<-round(1-(1-niveau)/J,4)
kn<-0
L<-dim(esti)[1]
for (i in 1:L) {if (esti[i,1]<=sigma) kn<-(kn+1) }
for (j in 1:J)
   { borne[j]<-round(esti[kn,2*J+6+j]/(1+esti[kn,2*J+6+j]),2)
     col[j]<-100*(borne[j]-0.08)
     for (i in 1:(dim(matB)[1]-1))
        { if ((matB[i,col[j]]<conf)&(matB[i+1,col[j]]>=conf))
           lambdaHW[j]<-matB[i+1,1]}
     }
lambdaHW
}


#-------------------------------------------------------------------------------#
# tirage des échantillons                                                       #
# J=nombre total de risques dépendants                                          #
# n=nombre de patients                                                          #
# on arrondit au (np)ème chiffre après la virgule                               #
# on tire x1 et x2 selon un ACBVW(a,b,l0,l1,l2)                                 #
#-------------------------------------------------------------------------------#

tirACBVWC<-function(n,np,a,b,l0,l1,l2,poids)
{
dbrutes<-matrix(0,n,3)
inf<-rexp(n,l1+l2+l0)
uni<-runif(n,0,1)
diff<-rep(0,n)
y1<-rep(0,n)
y2<-rep(0,n)
x1<-rep(0,n)
x2<-rep(0,n)
x<-rep(0,n)
c<-runif(n,0,poids)
t<-rep(0,n)
d<-rep(0,n)
jj<-rep(0,n)

for (i in 1:n)
```

```
  {  if (uni[i]< (l1/(l1+l2))) { diff[i]<- log((1+l2/l1)*uni[i])/(l0+l2)}
     if (uni[i]>=(l1/(l1+l2))) { diff[i]<- -log((1+l1/l2)*(1-uni[i]))/(l0+l1)}
     if (diff[i]< 0) { y1[i]<-inf[i]
                       y2[i]<-inf[i]-diff[i] }
     if (diff[i]>=0) { y1[i]<-diff[i]+inf[i]
                       y2[i]<-inf[i] }
     x1[i]<-(y1[i])^(1/a)
     x2[i]<-(y2[i])^(1/b)
     x[i]<-min(x1[i],x2[i])
     if (x[i]<=c[i]) { t[i]<-round(x[i],np)
                       d[i]<-1
                       if (x1[i]<=x2[i]) {jj[i]<-1}
                       else {jj[i]<-2}
                       }
    else { t[i]<-round(c[i],np)
           d[i]<-0
           jj[i]<-0
          }
    }
dbrutes<-cbind(t,d,jj)
dbrutes
}


#-----------------------------------------#
# Etape d'ordonnencement des donnees      #
#-----------------------------------------#

donnee<-function(z)
{
z<-z[order(z[,1]), ]
z
}


#-------------------------------------------------------------------------#
# Calcul de toutes les quantités utiles, dans l'ordre :                   #
# T_{i,n},H_n^{(1,1)},...,H_n^{(1,J)},H_n^{(1)},H_n^{(0)},H_n,            #
# ^F,^G,^F^{(1)},...,^F^{(J)},^C^{(1,1)},...^C^{(1,J)}                    #
#-------------------------------------------------------------------------#

estim<-function(z)
{
L<-dim(z)[1]
J<-max(z[,3])
diff<-matrix(0,L,2+J)
esti<-matrix(0,L+1,3*J+6)
```

```
for (i in 1:L) {if (z[i,2]==1) {diff[i,J+1]<-(1/L)
                                for (j in 1:J) {if (z[i,3]==j) diff[i,j]<-(1/L)} }
               else diff[i,J+2]<-(1/L)}

for (i in 1:L)
  {esti[i+1,1]<-z[i,1]
   for (j in 1:J) { esti[i+1,1+j]<-esti[i,1+j]+diff[i,j] }
   esti[i+1,J+2]<-esti[i,J+2]+diff[i,J+1]
   esti[i+1,J+3]<-esti[i,J+3]+diff[i,J+2]
   esti[i+1,J+4]<-esti[i+1,J+2]+esti[i+1,J+3]
   esti[i+1,J+5]<-esti[i,J+5]+(diff[i,J+1]/(1-esti[i,J+6]))
   esti[i+1,J+6]<-esti[i,J+6]+(diff[i,J+2]/(1-esti[i,J+5]))
   for (j in 1:J) {esti[i+1,J+6+j]<-esti[i,J+6+j]+diff[i,j]/(1-esti[i,J+6])}
   for (j in 1:J)
   {esti[i+1,2*J+6+j]<- esti[i,2*J+6+j]
                        +(1-esti[i+1,J+5])*diff[i,j]/((1-esti[i,J+4])*(1-esti[i,J+6]))
                        +2*esti[i+1,J+6+j]*diff[i,j]/((1-esti[i,J+6])*(1-esti[i,J+4]))
                        +esti[i+1,J+6+j]^2*diff[i,J+1]/((1-esti[i,J+4])
                           *(1-esti[i+1,J+5])*(1-esti[i,J+6]))}
     }
count<-0
for (i in 1:n)
  { if (esti[i-count,1]==esti[i+1-count]) { esti<-esti[-(i-count),]
                                           count<-count+1 }
    }
esti
}


# ------------------------------------------------------------------------------#
# Calcul des bornes inf et sup des bandes de confiance, dans l'ordre :          #
# borneinf de F(1),...,borneinf de F(J),borneinf de F,                          #
# bornesup de F(1),...,bornesup de F(J),bornesup de F                           #
# AN puis HW puis ANmodif puis HWmodif                                          #
# lambda permet de determiner la largeur de la bande et le niveau de confiance  #
# sigma est le point ou l'on stoppe l'estimation                                #
# kn est l'indice tel que sigma est égal au (n-kn)ème percentile                #
#-------------------------------------------------------------------------------#

bandconf<-function(esti,lAN,lHW,sigma)
{
J<-(dim(esti)[2]/3-2)
L<-dim(esti)[1]     #L=n+1 sauf si il y a des répétitions dans les données
Kn<-0
for (i in 1:L) {if (esti[i,1]<=sigma) Kn<-(Kn+1)}

kAN<-rep(0,J+1)
kHW<-matrix(0,Kn,J+1)
```

```
for (j in 1:J)
    { kAN[j]<- lAN*sqrt(esti[Kn,2*J+6+j])/sqrt(L-1)
      for (i in 1:Kn)  {kHW[i,j]<- lHW[j]*(1+esti[i,2*J+6+j])/sqrt(L-1)}
     }
kAN[J+1]<-sum(kAN)
for (i in 1:Kn)   {kHW[i,J+1]<-sum(kHW[i,])}

ANband<-matrix(0,Kn,2*(J+1))
ANmodifband<-matrix(0,Kn,2*(J+1))
HWband<-matrix(0,Kn,2*(J+1))
HWmodifband<-matrix(0,Kn,2*(J+1))

for (i in 1:Kn)
   { for (j in 1:J)
         { ANband[i,j]<-max((esti[i,J+6+j]-kAN[j])/(1+kAN[J+1]),0)
           ANband[i,1+J+j]<-min((esti[i,J+6+j]+kAN[j])/(1-kAN[J+1]),1)
           HWband[i,j]<-max((esti[i,J+6+j]-kHW[i,j])/(1+kHW[i,J+1]),0)
           HWband[i,1+J+j]<-min((esti[i,J+6+j]+kHW[i,j])/(1-kHW[i,J+1]),1)
           ANmodifband[i,j]<-max(esti[i,J+6+j]*(1-kAN[J+1])-kAN[j],0)
           ANmodifband[i,1+J+j]<-min(esti[i,J+6+j]*(1+kAN[J+1])+kAN[j],1)
           HWmodifband[i,j]<-max(esti[i,J+6+j]*(1-kHW[i,J+1])-kHW[i,j],0)
           HWmodifband[i,1+J+j]<-min(esti[i,J+6+j]*(1+kHW[i,J+1])+kHW[i,j],1)
          }
       ANband[i,J+1]<-max(1-(1-esti[i,J+5])/(1-kAN[J+1]),0)
       ANband[i,2*J+2]<-min(1-(1-esti[i,J+5])/(1+kAN[J+1]),1)
       HWband[i,J+1]<-max(1-(1-esti[i,J+5])/(1-kHW[i,J+1]),0)
       HWband[i,2*J+2]<-min(1-(1-esti[i,J+5])/(1+kHW[i,J+1]),1)
       ANmodifband[i,J+1]<-max(1-(1-esti[i,J+5])*(1+kAN[J+1]),0)
       ANmodifband[i,2*J+2]<-min(1-(1-esti[i,J+5])*(1-kAN[J+1]),1)
       HWmodifband[i,J+1]<-max(1-(1-esti[i,J+5])*(1+kHW[i,J+1]),0)
       HWmodifband[i,2*J+2]<-min(1-(1-esti[i,J+5])*(1-kHW[i,J+1]),1)
    }

statordre<-rep(0,Kn)
for (i in 1:Kn) statordre[i]<-esti[i,1]
band<-cbind(statordre,ANband,ANmodifband,HWband,HWmodifband)
band
}


#-------------------------------------------------------------------------#
# Simulations de Monte-Carlo                                              #
#-------------------------------------------------------------------------#

J<-2
n<-200
kn<-20
np<-3
```

```
p<-10^(-np)
a<-1.5
b<-3
l0<-0.1
l1<-0.05
l2<-0.01
poids<-3.9
nivo<-0.95
N<-1000

bandMC<-matrix(0,1001,8*(J+1)+1)
sigmaMC<-rep(0,N)
lastMC<-rep(0,N)

for (m in 1:N)
{ zwMC<-tirACBVWC(n,np,a,b,l0,l1,l2,poids)
  zzwMC<-donnee(zwMC)
  sigmaMC[m]<-zzwMC[n-kn,1]
  lastMC[m]<-zzwMC[n,1]
  sigmaaux<-sigmaMC[m]
  estiaux<-estim(zzwMC)
  lAN<-LambdaAN(nivo,J,matW)
  lHW<-LambdaHW(nivo,J,sigmaaux,estiaux,matB)
  bandaux<-bandconf(estiaux,lAN,lHW,sigmaaux)

  for (k in 1:1001)
    { bandMC[k,1]<-(k-1)*5/1000
      for (i in 1:(dim(bandaux)[1]-1))
          { if ((bandaux[i,1]<=bandMC[k,1])&(bandMC[k,1]<bandaux[i+1,1]))
              { for (j in 2:(8*(J+1)+1))  bandMC[k,j]<-bandMC[k,j]+bandaux[i,j]/N
                }
            }
      if (bandMC[k,1]>=bandaux[dim(bandaux)[1],1])
          { for (j in 2:(8*(J+1)+1))
            bandMC[k,j]<-bandMC[k,j]+bandaux[dim(bandaux)[1],j]/N
            }
      }
}

sigmaMCfin<-min(sigmaMC)
count<-0
for (k in 1:1001)
   { if ((k-1)*5/1000>sigmaMCfin)
          { bandMC<-bandMC[-(k-count),]
            count<-count+1 }
     }
lastMCfin<-min(lastMC)
```

```
bandrecup2<-bandMC
sigmarecup2<-sigmaMC
lastrecup2<-lastMC

aaa<- seq(0,5,0.01)
bbb<- numeric(length(aaa))
couleur<-c("red","blue","purple","aquamarine2","cyan","green","gray","pink","orange")
par(mfrow=c(1,J+1))
for (j in 1:(J+1))
  { if (j<J+1) {plot(c(0,0),c(1,0),type="l",xlab=" ",ylab=" ",xlim=c(0,3.5),
                     main=paste("Confidence bands for F",j))
                for (i in 1:length(aaa)) bbb[i]<-Fj(j,aaa[i],a,b,l0,l1,l2)
                lines(aaa,bbb,type="l") }
    if (j==J+1) {plot(c(0,0),c(1,0),type="l",xlab=" ",ylab=" ",xlim=c(0,3.5),
                     main=paste("Confidence bands for F"))
                 for (i in 1:length(aaa)) bbb[i]<-F(aaa[i],a,b,l0,l1,l2)
                 lines(aaa,bbb,type="l") }
    for (k in 1:4)
       { lines(bandMC[,1],bandMC[,1+j+(k-1)*2*(J+1)],type="l",col=couleur[k])
         lines(bandMC[,1],bandMC[,1+j+J+1+(k-1)*2*(J+1)],type="l",col=couleur[k])
         }
}
```

We now present the functions used to carry out the simulations of Chapter 5.

```
#-------------------------------------------------------------------------------#
# Calcul de ^G et ^F_Z (KM basé sur les derniers évènements de chaque individu)  #
#-------------------------------------------------------------------------------#

estiG<-function(z)
{
L<-dim(z)[1]
lastobs<-matrix(0,L,2)
for (i in 1:L)
    { last<-0
      k<-1
      total<-0
      while (last==0)
          { total<-total+z[i,1+3*(k-1)]
            if ((z[i,1+3*(k-1)]!=0)&(z[i,2+3*(k-1)]==0))
               { lastobs[i,1]<-total
                 lastobs[i,2]<-0
                 last<-1 }
            if (z[i,3+3*(k-1)]==2)
               { lastobs[i,1]<-total
```

```
                lastobs[i,2]<-2
                last<-1 }
            k<-k+1
            }
      }
lastobs<-lastobs[order(lastobs[,1]), ]

diff<-matrix(0,L,2)
for (i in 1:L)
      { if (lastobs[i,2]==0)  diff[i,1]<-(1/L)
        else diff[i,2]<-(1/L)
        }

GFZ<-matrix(0,L+1,3)
for (i in 1:L)
  {GFZ[i+1,1]<-lastobs[i,1]
   GFZ[i+1,2]<-GFZ[i,2]+(diff[i,1]/(1-GFZ[i,3]))
   GFZ[i+1,3]<-GFZ[i,3]+(diff[i,2]/(1-GFZ[i,2]))
   }

GFZ
}


#-------------------------------------------------------------------------------#
# Inference avec le premier évènement, dans l'ordre :                           #
# T_{i,n}[1], ^F[1], ^F[1(1)], ^F[1(2)]                                         #
#-------------------------------------------------------------------------------#

estim1<-function(z,GFZ)
{
L<-dim(z)[1]
aux<-matrix(0,L,3)
for (i in 1:3) { aux[,i]<-z[,i] }
aux<-aux[order(aux[,1]), ]
esti<-matrix(0,L+1,4)

for (i in 1:L)
  { esti[i+1,1]<-aux[i,1]
    for (j in 2:4) esti[i+1,j]<-esti[i,j]
    if (aux[i,2]==1)
        { for (j in 1:(dim(GFZ)[1]-1))
            {if ((GFZ[j,1]<aux[i,1])&(aux[i,1]<=GFZ[j+1,1])) div<- 1/(1-GFZ[j,2])
             if (aux[i,1]>GFZ[j+1,1]) div<- 1/(1-GFZ[dim(GFZ)[1],2])
             }
          esti[i+1,4]<-esti[i+1,4]+div/L
          if (aux[i,3]==1)  esti[i+1,2]<-esti[i+1,2]+div/L
          if (aux[i,3]==2)  esti[i+1,3]<-esti[i+1,3]+div/L
```

```
          }
    }
count<-0
for (i in 1:L)
  { if (esti[i-count,1]==esti[i+1-count]) { esti<-esti[-(i-count),]
                                           count<-count+1
                                           }
    }
esti
}


#-------------------------------------------------------------------------------#
# Inference avec le deuxième évènement, dans l'ordre :                          #
# T_{i,n}[1], ^F[1], ^F[1(1)], ^F[1(2)]                                         #
#-------------------------------------------------------------------------------#

estim2<-function(z,GFZ,cond)
{
L<-dim(z)[1]
aux<-matrix(0,L,6)
for (i in 1:6) { aux[,i]<-z[,i] }
aux<-aux[order(aux[,4]), ]
count<-0
for (i in 1:L)
  { if (aux[i-count,4]==0) { aux<-aux[-(i-count),]
                             count<-count+1 }
    }

esti<-matrix(0,dim(aux)[1]+1,4)

for (i in 1:dim(aux)[1])
  { esti[i+1,1]<- aux[i,4]
    for (j in 2:4) esti[i+1,j]<-esti[i,j]
    if (aux[i,5]==1)
       { test<- aux[i,1]+aux[i,4]
         for (j in 1:(dim(GFZ)[1]-1))
            { if ((GFZ[j,1]<test)&(test<=GFZ[j+1,1])) div<- 1/(1-GFZ[j,2])
              if (test>GFZ[dim(GFZ)[1],1]) div<- 1/(1-GFZ[dim(GFZ)[1],2])
              }
         esti[i+1,4]<-esti[i+1,4]+div/(L*cond)
         if (aux[i,6]==1) { esti[i+1,2]<-esti[i+1,2]+div/(L*cond) }
         if (aux[i,6]==2) { esti[i+1,3]<-esti[i+1,3]+div/(L*cond) }
         }
    }
count<-0
for (i in 1:dim(aux)[1])
  { if (esti[i-count,1]==esti[i+1-count]) { esti<-esti[-(i-count),]
```

```
                                                count<-count+1
                                                }
    }
esti
}


#-------------------------------------------------------------------------------#
# Inference avec le troisième évènement, dans l'ordre :                         #
# T_{i,n}[1], ^F[1], ^F[1(1)], ^F[1(2)]                                         #
#-------------------------------------------------------------------------------#

estim3<-function(z,GFZ,cond1,cond2)
{
L<-dim(z)[1]
cond<-cond1*cond2

if (sum(z[,7])!=0)
 {
   aux<-matrix(0,L,9)
   for (i in 1:9) { aux[,i]<-z[,i] }
   aux<-aux[order(aux[,7]), ]
   count<-0
   for (i in 1:L)
      { if (aux[i-count,7]==0) { aux<-aux[-(i-count),]
                                 count<-count+1 }
        }

   esti<-matrix(0,dim(aux)[1]+1,4)

   for (i in 1:dim(aux)[1])
     { esti[i+1,1]<- aux[i,7]
       for (j in 2:4) esti[i+1,j]<-esti[i,j]
       if (aux[i,8]==1)
          { test<- aux[i,1]+aux[i,4]+aux[i,7]
            for (j in 1:(dim(GFZ)[1]-1))
               { if ((GFZ[j,1]<test)&(test<=GFZ[j+1,1])) div<- 1/(1-GFZ[j,2])
                 if (test>GFZ[dim(GFZ)[1],1]) div<- 1/(1-GFZ[dim(GFZ)[1],2])
                }
            esti[i+1,4]<-esti[i+1,4]+div/(L*cond)
            if (aux[i,9]==1) { esti[i+1,2]<-esti[i+1,2]+div/(L*cond) }
            if (aux[i,9]==2) { esti[i+1,3]<-esti[i+1,3]+div/(L*cond) }
            }
       }
   count<-0
   for (i in 1:dim(aux)[1])
     { if (esti[i-count,1]==esti[i+1-count]) { esti<-esti[-(i-count),]
                                              count<-count+1
```

```
                                                  }
       }
 }
if (sum(z[,7])==0)
 { esti<-matrix(0,dim(z)[1]+1,4)
   }
 esti
}


#-------------------------------------------------------------#
# Tirage de (x1,x2) selon un ACBVW(a,b,l0,l1,l2)              #
#-------------------------------------------------------------#

tirsimpleACBVWC<-function(a,b,l0,l1,l2)
{
inf<-rexp(1,l1+l2+l0)
uni<-runif(1,0,1)
diff<-0
y1<-0
y2<-0
x1<-0
x2<-0
x<-0

if (uni< (l1/(l1+l2))) { diff<- log((1+l2/l1)*uni)/(l0+l2) }
if (uni>=(l1/(l1+l2))) { diff<- -log((1+l1/l2)*(1-uni))/(l0+l1) }
if (diff< 0) { y1<-inf
               y2<-inf-diff }
if (diff>=0) { y1<-diff+inf
               y2<-inf }
x1<-(y1)^(1/a)
x2<-(y2)^(1/b)
x<-min(x1,x2)
if (x==0) x<- 2*p
if (x1<x2) cronde<-1
else cronde<-2

xc<-c(x,cronde)
xc
}


#-------------------------------------------------------------------------------------#
# Monte-Carlo des estimateurs                                                         #
#-------------------------------------------------------------------------------------#

#-------------------------------------------------------------------------------------#
```

```
# Pour simuler les données avec une censure uniforme                              #
#--------------------------------------------------------------------------------#

#--------------------------------------------------------------------------------#
# tirage des échantillons                                                        #
# ce sont les Y ie les temps d'interarrivées qui sont reportés dans le tableau   #
# censure=uniforme                                                               #
# on suppose les temps d'interarrivée indépendants                               #
# n=nombre de patients                                                           #
# unique censure pour toute la séquence d'évènements d'un individu               #
# np=on arrondit au npˆème chiffre après la virgule                              #
#--------------------------------------------------------------------------------#

tirrepeated<-function(n,np,a,b,l0,l1,l2,poids)
{
dbrutes<-matrix(0,n,24)
for (i in 1:n)
  { mort<-0
    censure<-0
    c<-runif(1,0,poids)
    if (c==0) c<-p
    total<-0
    k<-1
    while ((mort==0)&(censure==0)&(k<=8))
       { xc<-tirsimpleACBVWC(a,b,l0,l1,l2)
         ycens<-c-total
         total<-total+xc[1]
         if (total<=c) { dbrutes[i,1+(k-1)*3]<- round(xc[1],np)
                         dbrutes[i,2+(k-1)*3]<- 1
                         dbrutes[i,3+(k-1)*3]<- xc[2]
                         if (xc[2]==2) mort<-1
                         }
         if (total>c) { dbrutes[i,1+(k-1)*3]<- round(ycens,np)
                        dbrutes[i,2+(k-1)*3]<- 0
                        dbrutes[i,3+(k-1)*3]<- 0
                        censure<-1
                        }
         if (dbrutes[i,1+(k-1)*3]==0) dbrutes[i,1+(k-1)*3]<-10^(-np)
         if ((k==8)&(dbrutes[i,3+(k-1)*3]==1))  dbrutes[i,3+(k-1)*3]<- 2
             #je fais mourir ceux qui dépasseraient 8 AINF (très rare)
         k<- k+1
         }
   }
dbrutes
}


#--------------------------------------------------------------------------------#
```

```
# Pour simuler les données avec une censure exponentielle                    #
#-------------------------------------------------------------------------------#

tirrepeated<-function(n,np,a,b,l0,l1,l2,poids)
{
dbrutes<-matrix(0,n,24)
for (i in 1:n)
  { mort<-0
    censure<-0
    c<-rexp(1,poids)
    if (c==0) c<-p
    total<-0
    k<-1
    while ((mort==0)&(censure==0)&(k<=8))
       { xc<-tirsimpleACBVWC(a,b,l0,l1,l2)
         ycens<-c-total
         total<-total+xc[1]
         if (total<=c) { dbrutes[i,1+(k-1)*3]<- round(xc[1],np)
                         dbrutes[i,2+(k-1)*3]<- 1
                         dbrutes[i,3+(k-1)*3]<- xc[2]
                         if (xc[2]==2) mort<-1
                         }
         if (total>c) { dbrutes[i,1+(k-1)*3]<- round(ycens,np)
                        dbrutes[i,2+(k-1)*3]<- 0
                        dbrutes[i,3+(k-1)*3]<- 0
                        censure<-1
                        }
         if (dbrutes[i,1+(k-1)*3]==0) dbrutes[i,1+(k-1)*3]<-10^(-np)
         if ((k==8)&(dbrutes[i,3+(k-1)*3]==1))  dbrutes[i,3+(k-1)*3]<- 2
            #je fais mourir ceux qui dépasseraient 8 AINF (très rare)
         k<- k+1
         }
   }
dbrutes
}


#-------------------------------------------------------------------------------#
#-------------------------------------------------------------------------------#

n<-1000
np<-2
p<-10^(-np)
a<-1.5
b<-3
l0<-0.1
l1<-0.05
l2<-0.1
```

```
poids<-5
N<-1000

estiMC <- matrix(0,poids/p+1,10)
dMC <- dim(estiMC)[1]
minMC11 <- matrix(0,poids/p+1,25)
maxMC11 <- matrix(0,poids/p+1,25)
minMC12 <- matrix(0,poids/p+1,25)
maxMC12 <- matrix(0,poids/p+1,25)
minMC13 <- matrix(0,poids/p+1,25)
maxMC13 <- matrix(0,poids/p+1,25)
minMC21 <- matrix(0,poids/p+1,25)
maxMC21 <- matrix(0,poids/p+1,25)
minMC22 <- matrix(0,poids/p+1,25)
maxMC22 <- matrix(0,poids/p+1,25)
minMC23 <- matrix(0,poids/p+1,25)
maxMC23 <- matrix(0,poids/p+1,25)
minMC31 <- matrix(0,poids/p+1,25)
maxMC31 <- matrix(0,poids/p+1,25)
minMC32 <- matrix(0,poids/p+1,25)
maxMC32 <- matrix(0,poids/p+1,25)
minMC33 <- matrix(0,poids/p+1,25)
maxMC33 <- matrix(0,poids/p+1,25)

for (m in 1:N)
{ zw<-tirrepeated(n,np,a,b,l0,l1,l2,poids)

  GFZ<-estiG(zw)
  esti1<-estim1(zw,GFZ)
  cond1<-esti1[dim(esti1)[1],2]
  esti2<-estim2(zw,GFZ,cond1)
  cond2<-esti2[dim(esti2)[1],2]
  esti3<-estim3(zw,GFZ,cond1,cond2)

  for (k in 1:dMC)
    { estiMC[k,1]<-(k-1)*p
      for (i in 1:(dim(esti1)[1]-1))
          { if ((esti1[i,1]<=estiMC[k,1])&(estiMC[k,1]<esti1[i+1,1]))
               { estiMC[k,2]<-estiMC[k,2]+esti1[i,2]/N
                 estiMC[k,3]<-estiMC[k,3]+esti1[i,3]/N
                 estiMC[k,4]<-estiMC[k,4]+esti1[i,4]/N
                 if ((1<=m)&(m<=25))
                       { minMC11[k,m]<- maxMC11[k,m]<- esti1[i,2]
                         minMC12[k,m]<- maxMC12[k,m]<- esti1[i,3]
                         minMC13[k,m]<- maxMC13[k,m]<- esti1[i,4]
                         }
                 if (m>25)
                       { if (max(minMC11[k,])>esti1[i,2])
```

```
                        minMC11[k,which.max(minMC11[k,])]<-esti1[i,2]
                  if (max(minMC12[k,])>esti1[i,3])
                        minMC12[k,which.max(minMC12[k,])]<-esti1[i,3]
                  if (max(minMC13[k,])>esti1[i,4])
                        minMC13[k,which.max(minMC13[k,])]<-esti1[i,4]
                  if (min(maxMC11[k,])<esti1[i,2])
                        maxMC11[k,which.min(maxMC11[k,])]<-esti1[i,2]
                  if (min(maxMC12[k,])<esti1[i,3])
                        maxMC12[k,which.min(maxMC12[k,])]<-esti1[i,3]
                  if (min(maxMC13[k,])<esti1[i,4])
                        maxMC13[k,which.min(maxMC13[k,])]<-esti1[i,4]
                  }
            }
      }
for (i in 1:(dim(esti2)[1]-1))
    { if ((esti2[i,1]<=estiMC[k,1])&(estiMC[k,1]<esti2[i+1,1]))
        { estiMC[k,5]<-estiMC[k,5]+esti2[i,2]/N
          estiMC[k,6]<-estiMC[k,6]+esti2[i,3]/N
          estiMC[k,7]<-estiMC[k,7]+esti2[i,4]/N
          if ((1<=m)&(m<=25))
                { minMC21[k,m]<- maxMC21[k,m]<- esti2[i,2]
                  minMC22[k,m]<- maxMC22[k,m]<- esti2[i,3]
                  minMC23[k,m]<- maxMC23[k,m]<- esti2[i,4]
                  }
          if (m>25)
                { if (max(minMC21[k,])>esti2[i,2])
                        minMC21[k,which.max(minMC21[k,])]<-esti2[i,2]
                  if (max(minMC22[k,])>esti2[i,3])
                        minMC22[k,which.max(minMC22[k,])]<-esti2[i,3]
                  if (max(minMC23[k,])>esti2[i,4])
                        minMC23[k,which.max(minMC23[k,])]<-esti2[i,4]
                  if (min(maxMC21[k,])<esti2[i,2])
                        maxMC21[k,which.min(maxMC21[k,])]<-esti2[i,2]
                  if (min(maxMC22[k,])<esti2[i,3])
                        maxMC22[k,which.min(maxMC22[k,])]<-esti2[i,3]
                  if (min(maxMC23[k,])<esti2[i,4])
                        maxMC23[k,which.min(maxMC23[k,])]<-esti2[i,4]
                  }
          }
      }
for (i in 1:(dim(esti3)[1]-1))
    { if ((esti3[i,1]<=estiMC[k,1])&(estiMC[k,1]<esti3[i+1,1]))
        { estiMC[k,8]<-estiMC[k,8]+esti3[i,2]/N
          estiMC[k,9]<-estiMC[k,9]+esti3[i,3]/N
          estiMC[k,10]<-estiMC[k,10]+esti3[i,4]/N
          if ((1<=m)&(m<=25))
                { minMC31[k,m]<- maxMC31[k,m]<- esti3[i,2]
                  minMC32[k,m]<- maxMC32[k,m]<- esti3[i,3]
```

```
                    minMC33[k,m]<- maxMC33[k,m]<- esti3[i,4]
                    }
              if (m>25)
                    { if (max(minMC31[k,])>esti3[i,2])
                          minMC31[k,which.max(minMC31[k,])]<-esti3[i,2]
                      if (max(minMC32[k,])>esti3[i,3])
                          minMC32[k,which.max(minMC32[k,])]<-esti3[i,3]
                      if (max(minMC33[k,])>esti3[i,4])
                          minMC33[k,which.max(minMC33[k,])]<-esti3[i,4]
                      if (min(maxMC31[k,])<esti3[i,2])
                          maxMC31[k,which.min(maxMC31[k,])]<-esti3[i,2]
                      if (min(maxMC32[k,])<esti3[i,3])
                          maxMC32[k,which.min(maxMC32[k,])]<-esti3[i,3]
                      if (min(maxMC33[k,])<esti3[i,4])
                          maxMC33[k,which.min(maxMC33[k,])]<-esti3[i,4]
                      }
              }
       }
if (estiMC[k,1]>=esti1[dim(esti1)[1],1])
    { estiMC[k,2]<-estiMC[k,2]+esti1[dim(esti1)[1],2]/N
      estiMC[k,3]<-estiMC[k,3]+esti1[dim(esti1)[1],3]/N
      estiMC[k,4]<-estiMC[k,4]+esti1[dim(esti1)[1],4]/N
      if ((1<=m)&(m<=25))
            { minMC11[k,m]<- maxMC11[k,m]<- esti1[dim(esti1)[1],2]
              minMC12[k,m]<- maxMC12[k,m]<- esti1[dim(esti1)[1],3]
              minMC13[k,m]<- maxMC13[k,m]<- esti1[dim(esti1)[1],4]
              }
      if (m>25)
            { if (max(minMC11[k,])>esti1[dim(esti1)[1],2])
                    minMC11[k,which.max(minMC11[k,])]<-esti1[dim(esti1)[1],2]
              if (max(minMC12[k,])>esti1[dim(esti1)[1],3])
                    minMC12[k,which.max(minMC12[k,])]<-esti1[dim(esti1)[1],3]
              if (max(minMC13[k,])>esti1[dim(esti1)[1],4])
                    minMC13[k,which.max(minMC13[k,])]<-esti1[dim(esti1)[1],4]
              if (min(maxMC11[k,])<esti1[dim(esti1)[1],2])
                    maxMC11[k,which.min(maxMC11[k,])]<-esti1[dim(esti1)[1],2]
              if (min(maxMC12[k,])<esti1[dim(esti1)[1],3])
                    maxMC12[k,which.min(maxMC12[k,])]<-esti1[dim(esti1)[1],3]
              if (min(maxMC13[k,])<esti1[dim(esti1)[1],4])
                    maxMC13[k,which.min(maxMC13[k,])]<-esti1[dim(esti1)[1],4]
              }
     }
if (estiMC[k,1]>=esti2[dim(esti2)[1],1])
    { estiMC[k,5]<-estiMC[k,5]+esti2[dim(esti2)[1],2]/N
      estiMC[k,6]<-estiMC[k,6]+esti2[dim(esti2)[1],3]/N
      estiMC[k,7]<-estiMC[k,7]+esti2[dim(esti2)[1],4]/N
      if ((1<=m)&(m<=25))
            { minMC21[k,m]<- maxMC21[k,m]<- esti2[dim(esti2)[1],2]
```

```
                    minMC22[k,m]<- maxMC22[k,m]<- esti2[dim(esti2)[1],3]
                    minMC23[k,m]<- maxMC23[k,m]<- esti2[dim(esti2)[1],4]
                    }
            if (m>25)
                  { if (max(minMC21[k,])>esti2[dim(esti2)[1],2])
                          minMC21[k,which.max(minMC21[k,])]<-esti2[dim(esti2)[1],2]
                    if (max(minMC22[k,])>esti2[dim(esti2)[1],3])
                          minMC22[k,which.max(minMC22[k,])]<-esti2[dim(esti2)[1],3]
                    if (max(minMC23[k,])>esti2[dim(esti2)[1],4])
                          minMC23[k,which.max(minMC23[k,])]<-esti2[dim(esti2)[1],4]
                    if (min(maxMC21[k,])<esti2[dim(esti2)[1],2])
                          maxMC21[k,which.min(maxMC21[k,])]<-esti2[dim(esti2)[1],2]
                    if (min(maxMC22[k,])<esti2[dim(esti2)[1],3])
                          maxMC22[k,which.min(maxMC22[k,])]<-esti2[dim(esti2)[1],3]
                    if (min(maxMC23[k,])<esti2[dim(esti2)[1],4])
                          maxMC23[k,which.min(maxMC23[k,])]<-esti2[dim(esti2)[1],4]
                    }
           }
      if (estiMC[k,1]>=esti3[dim(esti3)[1],1])
          { estiMC[k,8]<-estiMC[k,8]+esti3[dim(esti3)[1],2]/N
            estiMC[k,9]<-estiMC[k,9]+esti3[dim(esti3)[1],3]/N
            estiMC[k,10]<-estiMC[k,10]+esti3[dim(esti3)[1],4]/N
            if ((1<=m)&(m<=25))
                  { minMC31[k,m]<- maxMC31[k,m]<- esti3[dim(esti3)[1],2]
                    minMC32[k,m]<- maxMC32[k,m]<- esti3[dim(esti3)[1],3]
                    minMC33[k,m]<- maxMC33[k,m]<- esti3[dim(esti3)[1],4]
                        }
            if (m>25)
                  { if (max(minMC31[k,])>esti3[dim(esti3)[1],2])
                          minMC31[k,which.max(minMC31[k,])]<-esti3[dim(esti3)[1],2]
                    if (max(minMC32[k,])>esti3[dim(esti3)[1],3])
                          minMC32[k,which.max(minMC32[k,])]<-esti3[dim(esti3)[1],3]
                    if (max(minMC33[k,])>esti3[dim(esti3)[1],4])
                          minMC33[k,which.max(minMC33[k,])]<-esti3[dim(esti3)[1],4]
                    if (min(maxMC31[k,])<esti3[dim(esti3)[1],2])
                          maxMC31[k,which.min(maxMC31[k,])]<-esti3[dim(esti3)[1],2]
                    if (min(maxMC32[k,])<esti3[dim(esti3)[1],3])
                          maxMC32[k,which.min(maxMC32[k,])]<-esti3[dim(esti3)[1],3]
                    if (min(maxMC33[k,])<esti3[dim(esti3)[1],4])
                          maxMC33[k,which.min(maxMC33[k,])]<-esti3[dim(esti3)[1],4]
                    }
            }
      }
}



#-------------------------------------------------------------------------------#
```

```
# Exploitation graphique des résultats                                          #
#--------------------------------------------------------------------------------#

F1<-function(x,a,b,L0,L1,L2)
{
 L<-L0+L1+L2
aux1inf<-function(x)
  {
   aux<-x^(a-1)*exp(-(L1+L0)*x^a-L2*x^b)
   aux
   }
 aux1sup<-function(x)
  {
   aux<-x^(a-1)*exp(-L1*x^a-(L2+L0)*x^b)
   aux
   }
 if (x<=0.000001) F1<- 0
 if ((0.000001<x)&(x<=1)) F1<- L*(L1+L0)/(L1+L2)*a*integrate(aux1inf,0.000001,x)$value
                              +L0/(L1+L2)*(exp(-L*x^a)-1)
 if (x>1) F1<- L*(L1+L0)/(L1+L2)*a*integrate(aux1inf,0.000001,1)$value
              +L0/(L1+L2)*(exp(-L)-1)+L1*L/(L1+L2)*a*integrate(aux1sup,1,x)$value
 F1
}

F2<-function(x,a,b,L0,L1,L2)
{
 L<-L0+L1+L2
 aux2inf<-function(x)
  {
   aux<-x^(b-1)*exp(-(L1+L0)*x^a-L2*x^b)
   aux
   }
 aux2sup<-function(x)
  {
   aux<-x^(b-1)*exp(-L1*x^a-(L2+L0)*x^b)
   aux
   }

 if (x<=0.000001) F2<- 0
 if ((0.000001<x)&(x<=1)) F2<-L*L2/(L1+L2)*b*integrate(aux2inf,0.000001,x)$value
 if (x>1) F2<- L*L2/(L1+L2)*b*integrate(aux2inf,0.000001,1)$value
              +L*(L2+L0)/(L1+L2)*b*integrate(aux2sup,1,x)$value
              +L0/(L1+L2)*(exp(-L*x^b)-exp(-L))
 F2
}

F<-function(x,a,b,L0,L1,L2)
{
```

```
L<-L0+L1+L2
if (x==0) y<-0
if ((0<x)&(x<=1)) y<-1-L/(L1+L2)*exp(-(L0+L1)*x^a-L2*x^b)+L0/(L1+L2)*exp(-L*x^a)
if (x>1) y<-1-L/(L1+L2)*exp(-L1*x^a-(L0+L2)*x^b)+L0/(L1+L2)*exp(-L*x^b)
y
}


#-------------------------------------------------------------------------------#
#-------------------------------------------------------------------------------#

a<-1.5
b<-3
l0<-0.1
l1<-0.05
l2<-0.1

aaa<-seq(0,5,0.01)
bbb<-numeric(length(aaa))
ccc<-numeric(length(aaa))
ddd<-numeric(length(aaa))
for (i in 1:length(aaa))
{ bbb[i]<-F1(aaa[i],a,b,l0,l1,l2)
  ccc[i]<-F2(aaa[i],a,b,l0,l1,l2)
  ddd[i]<-F(aaa[i],a,b,l0,l1,l2)
  }

mini11 <- numeric(dim(minMC11)[1])
maxi11 <- numeric(dim(maxMC11)[1])
for (i in 1:length(mini11)) { mini11[i]<- min(minMC11[i,])
                              maxi11[i]<- min(maxMC11[i,])
                              }
mini12 <- numeric(dim(minMC12)[1])
maxi12 <- numeric(dim(maxMC12)[1])
for (i in 1:length(mini12)) { mini12[i]<- min(minMC12[i,])
                              maxi12[i]<- min(maxMC12[i,])
                              }
mini13 <- numeric(dim(minMC13)[1])
maxi13 <- numeric(dim(maxMC13)[1])
for (i in 1:length(mini13)) { mini13[i]<- min(minMC13[i,])
                              maxi13[i]<- min(maxMC13[i,])
                              }

par(mfrow=c(1,3))
plot(estiMC[,1],mini11,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[1(1)]",col="blue")
lines(estiMC[,1],maxi11,type="s",col="blue")
lines(estiMC[,1],estiMC[,2],type="s",col="dodgerblue")
```

```
lines(aaa,bbb,type="l")
plot(estiMC[,1],mini12,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[1(2)]",col="blue")
lines(estiMC[,1],maxi12,type="s",col="blue")
lines(estiMC[,1],estiMC[,3],type="l",col="dodgerblue")
lines(aaa,ccc,type="l")
plot(estiMC[,1],mini13,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[1]",col="blue")
lines(estiMC[,1],maxi13,type="s",col="blue")
lines(estiMC[,1],estiMC[,4],type="l",col="dodgerblue")
lines(aaa,ddd,type="l")

mini21 <- numeric(dim(minMC21)[1])
maxi21 <- numeric(dim(maxMC21)[1])
for (i in 1:length(mini21)) { mini21[i]<- min(minMC21[i,])
                              maxi21[i]<- min(maxMC21[i,])
                              }
mini22 <- numeric(dim(xminMC22)[1])
maxi22 <- numeric(dim(maxMC22)[1])
for (i in 1:length(mini22)) { mini22[i]<- min(minMC22[i,])
                              maxi22[i]<- min(maxMC22[i,])
                              }
mini23 <- numeric(dim(minMC23)[1])
maxi23 <- numeric(dim(maxMC23)[1])
for (i in 1:length(mini23)) { mini23[i]<- min(minMC23[i,])
                              maxi23[i]<- min(maxMC23[i,])
                              }

par(mfrow=c(1,3))
plot(estiMC[,1],mini21,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[2(1)/1]",col="blue")
lines(estiMC[,1],maxi21,type="s",col="blue")
lines(estiMC[,1],estiMC[,5],type="s",col="dodgerblue")
lines(aaa,bbb,type="l")
plot(estiMC[,1],mini22,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[2(2)/1]",col="blue")
lines(estiMC[,1],maxi22,type="s",col="blue")
lines(estiMC[,1],estiMC[,6],type="l",col="dodgerblue")
lines(aaa,ccc,type="l")
plot(estiMC[,1],mini23,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[2/1]",col="blue")
lines(estiMC[,1],maxi23,type="s",col="blue")
lines(estiMC[,1],estiMC[,7],type="l",col="dodgerblue")
lines(aaa,ddd,type="l")

mini31 <- numeric(dim(minMC31)[1])
maxi31 <- numeric(dim(maxMC31)[1])
for (i in 1:length(mini31)) { mini31[i]<- min(minMC31[i,])
```

```
                              maxi31[i]<- min(maxMC31[i,])
                              }
mini32 <- numeric(dim(minMC32)[1])
maxi32 <- numeric(dim(maxMC32)[1])
for (i in 1:length(mini32)) { mini32[i]<- min(minMC32[i,])
                              maxi32[i]<- min(maxMC32[i,])
                              }
mini33 <- numeric(dim(minMC33)[1])
maxi33 <- numeric(dim(maxMC33)[1])
for (i in 1:length(mini33)) { mini33[i]<- min(minMC33[i,])
                              maxi33[i]<- min(maxMC33[i,])
                              }

par(mfrow=c(1,3))
plot(estiMC[,1],mini31,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
      main="Estimator for F[3(1)/1,2]",col="blue")
lines(estiMC[,1],maxi31,type="s",col="blue")
lines(estiMC[,1],estiMC[,8],type="s",col="dodgerblue")
lines(aaa,bbb,type="l")
plot(estiMC[,1],mini32,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
       main="Estimator for F[3(2)/1,2]",col="blue")
lines(estiMC[,1],maxi32,type="s",col="blue")
lines(estiMC[,1],estiMC[,9],type="l",col="dodgerblue")
lines(aaa,ccc,type="l")
plot(estiMC[,1],mini33,type="s",xlab=" ",ylab=" ",xlim=c(0,5),ylim=c(0,1),
       main="Estimator for F[3/1,2]",col="blue")
lines(estiMC[,1],maxi33,type="s",col="blue")
lines(estiMC[,1],estiMC[,10],type="l",col="dodgerblue")
lines(aaa,ddd,type="l")
```

References

Aalen, O. (1978a). Nonparametric estimation of partial transition probabilities in multiple decrement model. *Ann. Statist. 6*, 534–545.

Aalen, O. (1978b). Nonparametric inference for a family of counting processes. *Ann. Statist. 6*, 701–726.

Aalen, O. and S. Johansen (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist. 5*, 141–150.

Andersen, P., O. Borgan, R. Gill, and N. Keiding (1993). *Statistical models based on counting processes*. New-York: Springer-Verlag.

Andersen, P. and R. Gill (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist. 10*, 1100–1120.

Anderson, T. (1960). A modification of the sequential probability ratio test to reduce sample size. *Ann. Math. Statist. 31*, 165–197.

Arconès, M. and E. Giné (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic Process. Appl. 58*, 217–245.

Billingsley, P. (1968). *Convergence of probability measures*. New-York: Wiley.

Block, H. and A. Basu (1974). A continuous bivariate exponential extension. *J. Amer. Statist. Assoc. 69*, 1031–1037.

Breslow, N. and J. Crowley (1974). A large sample study of the life table and product-limit estimates under random censorship. *Ann. Statist. 2*, 437–453.

Bretagnolle, J. and P. Massart (1989). Hungarian construction from the nonasymptotic viewpoint. *Ann. Probab. 17*, 239–256.

Burke, M., S. Csörgő, and L. Horváth (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrsch. Verw. Gebiete 56*, 87–112.

Burke, M., S. Csörgő, and L. Horváth (1988). A correction to and an improvement of "strong approximations of some biometric estimates under random censorship". *Probab. Theory and Rel. Fields 79*, 51–57.

Cambou, J. (1999). Epidémiologie de l'athérosclérose : risques et paradoxes. *La Revue du Praticien 49*, 2094–2099.

Capéraà, P. and B. Van Custem (1988). *Méthodes et modèles en statistique non-paramétrique*. Paris: Dunod.

Chang, S.-H. (2004). Estimating marginal effects in accelerated failure time models for serial sejourn times among repeated events. *Lifetime Data Analysis 10*, 175–190.

Chen, K. and S.-H. Lo (1997). On the rate of uniform convergence of the product-limit estimator: strong and weak laws. *Ann. Statist. 25*, 1050–1087.

Cheng, P. and G. Lin (1987). Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model. *Statist. Probab. Letters 5*, 75–80.

Cook, R. and J. Lawless (1997). Marginal analysis of recurrent events and a terminating event. *Statist. Med. 16*, 911–924.

Courrège, P. and P. Priouret (1965). Temps d'arrêt d'une fonction aléatoire. *Publ. Inst. Statist. Univ. Paris 14*, 245–274.

Cox, D. and O. D. (1984). *Analysis of survival data*. London: Chapman and Hall.

Crowder, M. (2001). *Classical competing risks*. London: Chapman and Hall.

Csáki, E. (1977). The law of the iterated logarithm for normalized empirical distribution function. *Z. Wahrsch. Verw. Gebiete 38*, 147–167.

Csörgő, M. and P. Révész (1981). *Strong approximations in probability and statistics*. New-York: Academic Press.

Csörgő, S. (1988). Estimation in the proportional hazards model of random censorship. *Statistics 19*, 437–463.

Csörgő, S. (1996). Universal Gaussian approximations under random censorship. *Ann. Statist. 24*, 2744–2778.

Csörgő, S. and L. Horváth (1981). On the Koziol-Green model for random censorship. *Biometrika 68*, 391–401.

Csörgő, S. and L. Horváth (1982). On random censorship from the right. *Acta Scientarum Mathematicarum (Szeged) 44*, 23–34.

Csörgő, S. and L. Horváth (1983a). On cumulative hazard processes under random censorship. *Scand. J. Statist. 9*, 13–21.

Csörgő, S. and L. Horváth (1983b). On the Koziol-Green model for random censorship ii. *Studia Sci. Math. Hungar. 18*, 195–203.

Csörgő, S. and L. Horváth (1986). Confidence bands for censored samples. *Canad. J. Statist. 14*, 131–144.

Dauxois, J.-Y. (2000). A new method for proving weak convergence results applied to nonparametric estimators in survival analysis. *Stochastic Process. Appl. 90*, 327–334.

Dauxois, J.-Y. and S. Kirmani (2004). On testing the proportionality of two cumulative incidence functions in a competing risks setup. *J. Nonparametric Statist. 16*, 479–492.

Deheuvels, P. and J. Einmahl (1996). On the strong limiting behavior of local functionals of empirical processes based upon censored data. *Ann. Probab. 24*, 504–525.

Deheuvels, P. and J. Einmahl (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. *Ann. Probab. 28*, 1301–1335.

Derzko, G. and E. Leconte (2004). Estimation non-paramétrique pour des évènements répétés tronqués et censurés aléatoirement à droite - Applications biomédicales. *J. SFDS 145*, 77–101.

Dikta, G. (1995). Asymptotic normality under the Koziol-Green model. *Comm. Stat. Theory Meth. 24*, 1537–1549.

Dudley, R. (1978). Central limit theorems for empirical measures. *Ann. Probab. 6*, 899–929.

Elandt-Johnson, R. and N. Johnson (1980). *Survival models and data analysis*. New-York: Wiley.

Földes, A. and L. Rejtő (1981a). A LIL-type result for the product-limit estimator. *Z.W.V.G. 56*, 75–86.

Földes, A. and L. Rejtő (1981b). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist. 9*, 122–129.

Fleming, T. and D. Harrington (1991). *Counting processes and survival analysis*. New-York: Wiley.

Gather, U. and J. Pawlitschko (1998). Estimating the survival function under a generalized Koziol-Green model with partially informative censoring. *Metrika 48*, 189–207.

Geffray, S. and A. Guilloux (2005). Estimation dans un modèle de Koziol-Green généralisé. *C. R. Acad. Sci. Paris, Ser. I 341*, 49–52.

Gent, S. (1996). A randomised, blinded trial of clopidogrel versus apirin in patients at risk of ischaemic events (CAPRIE). *The Lancet 348*, 1329–1339.

Ghorai, J. and J. Schmitter (1999). The asymptotic distribution of the suprema of the standardized empirical processes under the Koziol-Green model. *Statist. Probab. Letters 41*, 303–313.

Ghosh, D. and D. Lin (2000). Nonparametric analysis of recurrent events and death. *Biometrics 56*, 554–562.

Ghosh, D. and D. Lin (2002). Marginal regression models for recurrent and terminal events. *Statist. Sinica 12*, 663–688.

Ghosh, D. and D. Lin (2003). Semiparametric analysis of recurrent events in the presence of dependent censoring. *Biometrics 59*, 877–885.

Gill, R. (1980). *Censoring and stochastic integrals*. Mathematisch Centrum, Amsterdam: Mathematical Centre Tracts 124.

Gill, R. (1983). Large sample behavior of the product-limit estimator in the whole line. *Ann. Statist. 11*, 49–58.

Gill, R. (1994). Glivenko-Cantelli for Kaplan-Meier. *Math. Methods Statist. 3*, 76–87.

Gillespie, M. (1979). Confidence bands for the Kaplan-Meier survival curve estimate. *Ann. Statist. 7*, 920–924.

Giné, E. and A. Guillou (1999). Laws of the iterated logarithm for censored data. *Ann. Probab. 27*, 2042–2067.

Giné, E. and A. Guillou (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptative intervals. *Ann. Inst. H. Poincaré Probab. Statist. 27*, 2042–2067.

Gumbel, E. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc. 55*, 698–707.

Hall, W. and J. Wellner (1980). Confidence bands for a survival curve from censored data. *Biometrika 67*, 133–143.

Hartman, P. and A. Wintner (1941). On the law of the iterated logarithm. *American Journal of Mathematics 63*, 169–176.

Horváth, L. (1980). Dropping continuity and independence assumptions in random censorship models. *Studia Scient. Math. Hungarica 15*, 381–389.

Huang, C.-Y. and M.-C. Wang (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc. 99*(468), 1153–1165.

Jacod, J. and A. Shiryayev (1987). *Limit theorems for stochastic processes*. Berlin: Springer-Verlag.

Jakubowski, A., J. Mémin, and G. Pagès (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace D^1 de Skorokhod. *Probab. Theory and Rel. Fields 81*, 111–137.

Kalbfleisch, J. and R. Prentice (1980). *The statistical analysis of failure time data*. New-York: Wiley.

Kaplan, E. and P. Meier (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc. 53*, 457–481.

Klein, E. and P. Moeschberger (1984). Nonparametric estimation from incomplete observations. *IAPQR Trans. 9*(1), 1–7.

Komlós, J., P. Major, and G. Tusnady (1975). An approximation of partial sums of independent r.v.'s. and the sample d.f. 1. *Z. Wahrsch. Verw. Gebiete 32*, 111–132.

Kosorok, M. (2005). *Introduction to empirical processes and semi-parametric inference*. New-York: Springer.

Koziol, J. (1975). Percentage points of the asymptotic distributions of one and two-samples K.-S. statistics for truncated and censored data. *Technometrics 17*, 507–510.

Lawless, J. and C. Nadeau (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics 37*, 158–168.

Lee, A. (1990). *U-statistics, theory and practice*. New-York: Marcel Dekker, Inc.

Li, Q. and S. Lagakos (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statist. Med. 16*, 925–940.

Lin, D., W. Sun, and Z. Ying (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika 86*, 59–70.

Lin, D., L. Wei, I. Yang, and Z. Ying (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. B 62*, 711–730.

Lin, D., L. Wei, and Z. Ying (1998). Accelerated failure time models for counting processes. *Biometrika 85*, 605–618.

Liu, L., R. Wolfe, and X. Huang (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics 60*, 747–756.

Major, P. and L. Rejtő (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist. 16*(3), 1113–1132.

Marshall, A. and I. Olkin (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc. 62*, 30–44.

Miloslavsky, M., S. Keles, M. Van der Laan, and S. Butler (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J. Roy. Statist. Soc. B 66*, 239–257.

Montgomery-Smith, S. (1993). Comparisons of sums of independent identically distributed random vectors. *Probab. Math. Statist. 14*, 281–285.

Nelson, W. (1969). Hazard plotting for incomplete observations. *J. Qual. Tech. 1*, 27–52.

Nelson, W. (1972). A short life test for comparing a sample with previous accelerated test results. *Technometrics 14*, 175–185.

Prentice, L., J. Kalbfleisch, A. Peterson, N. Flournoy, V. Farewell, and N. Breslow (1978). The analysis of failure times in the presence of competing risks. *Biometrics 34*, 541–554.

Prentice, L., B. Williams, and A. Peterson (1981). On the regression analysis of multivariate time data. *Biometrika 68*, 373–379.

Ryu, K. (1993). An extension of Marshall and Olkin's bivariate exponential distribution. *J. Amer. Statist. Assoc. 88*, 1458–1465.

Serfling, R. (1980). *Approximation theorems of mathematical statistics*. New-York: Wiley.

Shorack, G. and J. Wellner (1986). *Empirical processes with statistical application*. New-York: Wiley.

Stute, W. (1992). Strong consistency under the Koziol-Green model. *Statist. Probab. Letters 14*, 313–320.

Stute, W. (1994a). Strong and weak representations of cumulative hazard function and Kaplan-Meier estimators on increasing sets. *J. Statist. Plann. Inference 42*, 315–329.

Stute, W. (1994b). U-statistic processes: a martingale approach. *Ann. Probab. 22*, 1725–1744.

Stute, W. and J.-L. Wang (1993). The strong law under random censorship. *Ann. Statist. 21*, 1591–1607.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math. 126*, 505–563.

Teigner, E. and A. Castaigne (1999). Description et mécanismes de l'ischémie d'origine athéroscléreuse. *La Revue du Praticien 49*, 2110–2116.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. USA 72*, 20–22.

Van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes with applications to statistics*. New-York: Springer-Verlag.

Vapnik, V. and A. Červonenkis (1971). The uniform convergence of frequencies of the appearance of events to their probabilities. *Theoret. Probab. Appl. 16*, 264–280.

Wang, M. and S.-H. Chang (1999). Nonparametric estimation of a recurrent survival function. *J. Amer. Statist. Assoc. 94*, 146–153.

Wang, M.-C., J. Qin, and C.-T. Chiang (2001). Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc. 96*, 1057–1065.

Wang, W. and M. Wells (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika 85*, 561–572.

Wei, L., D. Lin, and L. Weissfeld (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc. 84*, 1065–1073.

Wellner, J. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete 45*, 73–88.

Ying, Z. (1989). A note on the asymptotic properties of the product-limit estimator on the whole line. *Stat. and Probab. Lett. 26*, 265–275.