



HAL
open science

Processus empiriques pour l'inférence dans le modèle de survie à risques non proportionnels

Cecile Chauvel

► **To cite this version:**

Cecile Chauvel. Processus empiriques pour l'inférence dans le modèle de survie à risques non proportionnels. Mathématiques générales [math.GM]. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : 2014PA066399 . tel-01127777v2

HAL Id: tel-01127777

<https://theses.hal.science/tel-01127777v2>

Submitted on 7 Mar 2015

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.



UNIVERSITÉ PIERRE ET MARIE CURIE – PARIS 6

École Doctorale 386
Sciences Mathématiques de Paris Centre

THÈSE DE DOCTORAT

Présentée en vue de l'obtention du titre de Docteur
de l'Université Pierre et Marie Curie - Paris 6

Discipline : Mathématiques Appliquées
Spécialité : Statistique

Processus Empiriques pour l'Inférence dans le Modèle de Survie à Risques Non Proportionnels

par

Cécile CHAUVEL

dirigée par John O'QUIGLEY et Philippe SAINT PIERRE

Soutenue le 1er décembre 2014 devant le jury composé de :

M. Antoine CHAMBAZ	Professeur, Université Paris 10 Nanterre	Examineur
Mme Agathe GUILLOUX	Maître de conférences, Université Paris 6	Examineur
M. Robin HENDERSON	Professeur, University of Newcastle	Rapporteur
M. Aurélien LATOUCHE	Professeur, CNAM	Examineur
M. John O'QUIGLEY	Professeur, Université Paris 6	Directeur
M. Philippe SAINT PIERRE	Maître de conférences, Université Paris 6	Co-directeur

Rapporteurs :

Mme. Tianxi CAI	Professeur, Harvard Public School of Health
M. Robin HENDERSON	Professeur, University of Newcastle

Laboratoire de Statistique
Théorique et Appliquée
4, place Jussieu
75252 Paris Cedex 05
Boîte courrier 158

UPMC
Ecole Doctorale 386 de Sciences
Mathématiques de Paris Centre
4, place Jussieu
75252 Paris Cedex 05
Boîte courrier 290

Remerciements

Mes premiers remerciements s'adressent à mes directeurs de thèse. John, merci pour la confiance que vous m'avez témoignée en acceptant d'encadrer mon doctorat. Merci d'avoir partagé avec moi vos idées, vos intuitions et votre culture statistique au cours de nos nombreuses discussions. Philippe, merci de m'avoir fait confiance en me proposant cette aventure alors que j'étais étudiante à l'ISUP. Merci pour nos discussions (pas toujours statistiques!), les encouragements, le soutien et les conseils que tu m'as prodigués tout au long de ce parcours. Je suis très touchée par l'attention que tu m'as portée.

I would like to thank Tianxi Cai and Robin Henderson for the interest they showed in my work by reviewing this thesis. It is an honour for me that they took time and commented my work. Je remercie également sincèrement Antoine Chambaz, Agathe Guilloux et Aurélien Latouche d'avoir accepté de participer à mon jury de thèse. En particulier, merci à Agathe pour les discussions que l'on a pu avoir au cours de ces trois années. Merci d'avoir pris le temps de me faire des commentaires très pertinents sur mon travail.

Je remercie le directeur du laboratoire, Gérard Biau, et l'ancien directeur, Paul Deheuvels, de m'avoir accueillie au LSTA. Merci à Louise et Corinne pour leur gentillesse et leur disponibilité. Merci aux autres membres permanents du laboratoire, toujours à l'écoute, prêts à nous aider en cas de besoin et à discuter autour d'un café!

Ce travail n'aurait pas été le même sans la présence de mes camarades et amis doctorants (docteurs pour certains!) : Agniezska, Alexis, Amadou, Assia, Baptiste, Benjamin, Boris, Erwan, Layal, Matthieu, Mokhtar, Nedjmeddine, Patricia, Roxane, Sarah L., Sarah O., Soumeya, Svetlana, Tarn et Virgile. Merci pour tous les bons moments passés ensemble. C'était toujours une joie de me rendre au laboratoire pour vous retrouver. Grâce à vous, je garderai un excellent souvenir de ces trois années. Je suis émue d'écrire ces mots qui marquent la fin d'une étape, mais je sais que nos chemins se recroiseront. Je remercie particulièrement Amadou, Benjamin, Erwan et Layal, qui ont partagé mon bureau pendant une grande partie de cette thèse. Merci à Baptiste, Benjamin, Erwan et Svetlana pour notre co-organisation du Groupe de Travail des Thésards. Merci également à toi Salim pour l'aide que tu m'as apportée et ta bonne humeur constante.

D'un point de vue plus personnel, je remercie mes parents de m'avoir toujours soutenue dans mes études et dans mes choix. Merci d'avoir été à l'écoute lorsque j'en avais besoin et d'avoir cru en mes capacités. Je remercie mes sœurs pour leurs encouragements et leur soutien sans faille. Je ne serai pas arrivée là sans vous quatre! Merci à mes amis, indispensables à la

réussite de cette thèse, et pour lesquels je n'ai pas toujours été disponible ces derniers mois. Eric, je n'ai pas de mots assez forts pour t'exprimer ma gratitude. Tu m'as toujours soutenue au quotidien et été patient dans les moments difficiles (et on sait que la fin de la thèse n'est pas un long fleuve tranquille). Ces trois années ont été douces à tes côtés. Ainsi s'achève notre année 2014, pleine d'évènements et d'émotions !

Résumé

Nous nous intéressons à des processus empiriques particuliers pour l'inférence dans le modèle à risques non proportionnels. Ce modèle permet au coefficient de régression de varier avec le temps et généralise le modèle à risques proportionnels très utilisé pour modéliser des données de survie. Le processus du score standardisé que nous étudions est une somme séquentielle des résidus standardisés du modèle. Le processus est considéré en présence d'une covariable dans le modèle, avant d'être étendu au cas de multiples covariables pouvant être corrélées. Le plan du manuscrit se décompose en trois parties. Dans un premier temps, nous établissons les propriétés limites du processus sous le modèle et sous un modèle mal spécifié. Dans une deuxième partie, nous utilisons les résultats de convergence du processus pour dériver des tests de la valeur du paramètre du modèle. Nous montrons qu'un des tests proposés est asymptotiquement équivalent au test de référence du log-rank pour comparer les fonctions de survie de plusieurs groupes de patients. Nous construisons des tests plus puissants que le test du log-rank sous certaines alternatives. Enfin, dans la dernière partie, nous étudions comment lier prédiction et adéquation dans le modèle à risques non proportionnels. Nous proposons une méthode de construction d'un modèle bien ajusté en maximisant sa capacité prédictive. Aussi, nous introduisons un test d'adéquation du modèle à risques proportionnels. Les performances des méthodes proposées, qu'il s'agisse des tests sur le paramètre ou de l'adéquation du modèle, sont comparées à des méthodes de référence par des simulations. Les méthodes sont illustrées sur des données réelles.

Mots-clés : Survie, Modèle à risques non proportionnels, Processus empirique, Adéquation, Prédiction, Tests d'hypothèse.

Abstract

Empirical Processes for Inference in the Non-Proportional Hazards Model

In this thesis, we focus on particular empirical processes on which we can base inference in the non-proportional hazards model. This time-varying coefficient model generalizes the widely used proportional hazards model in the field of survival analysis. Our focus is on the standardized score process that is a sequential sum of standardized model-based residuals. We consider first the process with one covariate in the model, before looking at its extension for multiple and possibly correlated covariates. The outline of the manuscript is composed of three parts. In the first part, we establish the limit properties of the process under the model as well as under a misspecified model. In the second part, we use these convergence results to derive tests for the value of the model parameter. We show that one proposed test is asymptotically equivalent to the log-rank test, which is a benchmark for comparing survival experience of two or more groups. We construct more powerful tests under some alternatives. Finally, in the last part, we propose a methodology linking prediction and goodness of fit in order to construct models. The resulting models will have a good fit and will optimize predictive ability. We also introduce a goodness-of-fit test of the proportional hazards model. The performances of our methods, either tests for the parameter value or goodness-of-fit tests, are compared to standard methods via simulations. The methods are illustrated on real life datasets.

Keywords : Survival, Non-proportional hazards model, Empirical Process, Goodness of fit, Prediction, Hypotheses testing.

Table des matières

Remerciements	3
Résumé	5
Abstract	6
Plan détaillé de la thèse	11
Liste des notations et abréviations	15
1 Introduction	17
1.1 Introduction à l'analyse de survie	17
1.1.1 Introduction	17
1.1.2 Notation	18
1.1.3 Estimateur de Kaplan–Meier de la survie	19
1.2 Modèle à risques proportionnels et extensions	22
1.2.1 Modèle à risques proportionnels	22
1.2.2 Modèle à risques non proportionnels	26
1.2.3 Alternatives au modèle à risques proportionnels	31
1.3 Test de la présence d'un effet des covariables sur la survie	34
1.4 Evaluation de l'adéquation et de la capacité prédictive du modèle à risques proportionnels	36
1.4.1 Adéquation du modèle	36
1.4.2 Mesures de capacité prédictive du modèle	38
2 Outils mathématiques	41
2.1 Espaces $D[0, 1]$ et $C[0, 1]$	41
2.1.1 Topologies des espaces $(C[0, 1], \mathbb{R})$ et $(D[0, 1], \mathbb{R})$	41
2.1.2 Topologies de $(D[0, 1], \mathbb{R}^p)$	42
2.2 Théorèmes limite	44
2.2.1 Loi faible des grands nombres pour des variables non corrélées	44
2.2.2 Théorèmes de la limite centrale fonctionnels	44
2.3 Inégalité de Taylor-Lagrange multidimensionnelle	47

3	Etude du processus du score standardisé	49
3.1	Introduction	49
3.2	Transformation de l'échelle du temps	51
3.3	Processus du score standardisé univarié	57
3.3.1	Définition du processus	57
3.3.2	Résultats asymptotiques	59
3.4	Processus du score standardisé multivarié	63
3.4.1	Définition du processus	64
3.4.2	Résultats asymptotiques	65
3.5	Preuves	68
3.5.1	Preuves de la Section 3.2	68
3.5.2	Preuves de la Section 3.3	69
3.5.3	Preuves de la Section 3.4	79
4	Tests de comparaison de courbes de survie	87
4.1	Introduction	87
4.1.1	Test du log-rank	88
4.1.2	Tests du log-rank pondéré	89
4.1.3	Combinaison de tests	90
4.2	Tests basés sur le processus du score standardisé	93
4.2.1	Distance depuis l'origine	93
4.2.2	Aire sous la courbe	95
4.2.3	Tests adaptatifs restreints	97
4.2.4	Tests multivariés	99
4.3	Simulations	101
4.4	Applications	105
4.4.1	Données de l'Institut Curie	105
4.4.2	Essai clinique	106
4.5	Discussion	108
5	Construction et évaluation du modèle à risques non proportionnels	115
5.1	Adéquation du modèle à risques proportionnels	116
5.1.1	Méthodes graphiques	116
5.1.2	Tests d'adéquation	119
5.2	Mesure de la capacité de prédiction du modèle à risques non proportionnels	121
5.2.1	Variance expliquée	121
5.2.2	Coefficient R^2	125
5.3	Adéquation et construction de modèles	133
5.3.1	Adéquation du modèle à risques proportionnels	133
5.3.2	Construction du modèle à risques non proportionnels	136
5.4	Simulations	139

5.4.1	Test d'adéquation	139
5.4.2	Construction du modèle à risques non proportionnels	144
5.5	Applications	150
5.5.1	Données de Freireich	150
5.5.2	Etude clinique	150
5.6	Discussion	153
Annexe - Simulations de données sous le modèle à risques non proportionnels		157
.1	Méthode 1 - Modèle de rupture	157
.1.1	Simulation des données	157
.1.2	Choix des points de rupture	160
.2	Méthode 2 - Modèle à risques non proportionnels	161
Bibliographie		163

Plan détaillé de la thèse

Chapitre 1 - Introduction

Le premier chapitre de cette thèse comporte une revue de littérature de certains aspects de l'analyse de survie. Dans la première partie de ce chapitre, nous définissons les notations qui seront adoptées dans la suite de la thèse et nous rappelons la définition de l'estimateur de Kaplan–Meier de la fonction de survie. Ensuite, nous décrivons le modèle à risques proportionnels (Cox, 1972) très utilisé en pratique pour représenter les durées de vie. Nous exposons la méthode d'estimation du paramètre du modèle, dont le logarithme représente le risque relatif de décès instantané entre deux individus différents. Ce modèle est un cas particulier du modèle dit "à risques non proportionnels". Ce modèle plus général n'impose aucune restriction sur le risque relatif qui peut varier avec le temps alors que, dans le modèle à risques proportionnels, le risque relatif est constant. Ces modèles vont nous intéresser dans les chapitres suivants. Nous présentons une revue de littérature sur les méthodes d'estimation du paramètre de régression du modèle à risques non proportionnels. Enfin, nous proposons quelques modèles alternatifs au modèle à risques proportionnels pour caractériser les données de survie. La troisième partie de ce chapitre introductif rappelle le test du log-rank pour comparer la survie de groupes de patients (Mantel, 1966). Ce test, très utilisé en pratique, est le test de référence pour comparer des courbes de survie. Nous présenterons des tests alternatifs de comparaison de courbes de survie au Chapitre 4. Enfin, la dernière partie du premier chapitre traite l'évaluation du modèle à risques proportionnels. Plus précisément, nous présentons des méthodes d'évaluation de l'adéquation et des mesures de la capacité de prédiction du modèle. Nous reviendrons sur la littérature des méthodes d'adéquation du modèle au Chapitre 5.

Chapitre 2 - Outils mathématiques

Dans le deuxième chapitre, nous énonçons des résultats mathématiques non usuels en analyse de survie qui seront utilisés pour l'obtention des résultats de convergence des chapitres suivants. Nous rappelons les particularités des espaces de fonctions continues à droite et limite à gauche de $[0, 1]$ dans \mathbb{R}^p . Nous distinguons les cas $p = 1$ et $p > 1$. Ceci nous amène à définir des topologies dans ces deux espaces. Ensuite, nous définissons le mouvement Brownien et le pont Brownien sur $[0, 1]$ avant d'énoncer des théorèmes de la limite centrale fonctionnels pour des variables indépendantes et centrées, mais dont les variances peuvent être différentes. Ces

résultats sont présentés dans les cas univariés de variables aléatoires réelles et multivariés de vecteurs aléatoires de \mathbb{R}^p . Nous introduisons une loi faible des grands nombres pour des variables aléatoires non indépendantes. Enfin, nous présentons une inégalité de Taylor–Lagrange multidimensionnelle.

Chapitre 3 - Etude du processus du score standardisé

Un processus empirique, appelé processus du score standardisé, est défini dans ce 3^{ème} chapitre. Ce processus est une somme pondérée séquentielle des résidus standardisés du modèle à risques non proportionnels avec un incrément à chaque temps de décès. Le processus est tracé en fonction des rangs des temps d'évènement renormalisés dans l'intervalle $[0, 1]$. Le processus dépend du paramètre du modèle à travers les résidus. En s'inspirant du processus de [Wei \(1984\)](#), [O'Quigley \(2003\)](#) a introduit le processus sous le modèle à risques proportionnels avec une unique covariable et un paramètre égal au maximum de vraisemblance partielle. La différence avec le processus de [Wei \(1984\)](#) réside dans une double standardisation. Chaque incrément du processus du score standardisé est standardisé différemment alors que le processus de Wei adopte une standardisation globale. La deuxième standardisation est celle de l'échelle des temps qui n'est pas effectuée avec le processus de Wei. Ceci permet l'obtention d'une loi limite simple et explicite en cas de mauvaise spécification du modèle.

L'originalité du travail présenté dans ce chapitre est d'étudier le processus avec n'importe quel paramètre pouvant éventuellement varier dans le temps. Nous formalisons la transformation à appliquer sur le temps avant d'établir des résultats de convergence pour ce processus. De plus, nous développons le processus avec plusieurs covariables dans le modèle. La standardisation des incréments que nous adoptons permet de prendre en compte la covariance entre les covariables. Nous étudions également le comportement limite du processus multivarié. Sous un modèle bien spécifié, la loi limite du processus univarié ou multivarié est un mouvement Brownien. Si le modèle est mal spécifié, le processus n'est plus centré et une dérive proportionnelle au coefficient de régression apparaît. Le graphique du processus du score standardisé en fonction du temps permet d'anticiper la forme du coefficient de régression du modèle. Les résultats de ce chapitre sont utilisés dans les Chapitres 4 et 5 pour effectuer des tests d'hypothèse ou pour estimer le paramètre du modèle à risques non proportionnels. Les résultats de ce chapitre ont fait l'objet de deux articles ([Chauvel et O'Quigley, 2014a,b](#)).

Chapitre 4 - Tests de comparaison de courbes de survie

Le processus du score standardisé et les résultats de convergence du Chapitre 3 permettent de dériver des tests asymptotiques de comparaison de courbes de survie. Il s'agit de tester s'il existe une différence entre les survies de différents groupes de patients. Ces tests peuvent également s'appliquer à des covariables continues, auquel cas il s'agit de tester si la covariable est liée à la survie. [O'Quigley \(2008, chap. 8\)](#) propose de s'intéresser à la distance depuis

l'origine ou à l'aire sous la courbe du processus du score standardisé. Sous l'hypothèse nulle du test, le processus limite du score standardisé est un mouvement Brownien standard. Les tests sont alors construits à partir des propriétés du mouvement Brownien.

Dans ce chapitre, nous étudions des propriétés théoriques de ces deux tests. Nous montrons que le test de la distance depuis l'origine est asymptotiquement équivalent au test du log-rank. Ainsi, comme le test du log-rank, le test de la distance depuis l'origine est le plus puissant lorsque les risques relatifs entre les groupes ne changent pas au cours du temps (Peto et Peto, 1972). En présence de risques relatifs non constants, le test peut perdre de la puissance. Nous montrons que, dans ces situations, le test de l'aire sous la courbe est plus puissant, mais il perd de la puissance par rapport au test du log-rank lorsque le risque relatif ne varie pas au cours du temps. Afin d'obtenir un test puissant sous les deux types d'alternative, nous avons introduit une classe de tests adaptatifs restreints. Les tests sont adaptatifs au sens où la forme de la statistique du test est déterminée en fonction des données. Cependant, les tests de cette classe dépendent d'un paramètre qui peut être délicat à fixer. Nous proposons alors un autre test adaptatif ne dépendant pas d'un paramètre. Nous étudions les comportements des tests sur des simulations et nous les comparons à des tests proposés dans la littérature. Les simulations montrent que les tests se comportent comme anticipé sous les différentes alternatives. Nous avons également développé les tests dans le cadre multivarié de plusieurs covariables dans le modèle afin de tester plusieurs paramètres du modèle à la fois. La prise en compte de la corrélation entre les covariables dans la construction du processus présente l'avantage de pouvoir appliquer le test sur des variables corrélées. De plus, le graphique du processus en fonction du temps permet de visualiser l'écart à la significativité des différents tests étudiés. Enfin, nous appliquons les tests présentés sur des données réelles. Les résultats de ce chapitre sont publiés dans le journal *Biometrika* (Chauvel et O'Quigley, 2014a).

Chapitre 5 - Construction et évaluation du modèle à risques non proportionnels

Le dernier chapitre de la thèse traite de la construction du modèle à risques non proportionnels en considérant à la fois l'adéquation et la capacité de prédiction du modèle. Les résultats limites du processus du score standardisé (Chapitre 3) permettent d'utiliser le processus pour estimer le paramètre de régression. Pour ce faire, le processus est évalué avec un paramètre de régression nul. Aucune estimation du paramètre n'est requise, comme c'est le cas en traçant un nuage de points en deux dimensions pour des données non censurées afin d'évaluer le lien entre les deux variables. Dans un premier temps, nous proposons une revue de littérature des méthodes graphiques et des tests d'adéquation du modèle à risques proportionnels. Ensuite, nous présentons la variance expliquée et son estimateur, le coefficient R^2 d'O'Quigley et Flandre (1994), qui permettent de mesurer la capacité prédictive du modèle à risques proportionnels.

Notre travail a d'abord consisté à étendre le coefficient R^2 au modèle à risques non proportionnels. Nous étudions les propriétés asymptotiques du coefficient R^2 dans un modèle à risques

non proportionnels mal spécifié. En présence d'une covariable dans le modèle, nous montrons que le maximum du coefficient R^2 est atteint lorsqu'il est calculé avec le vrai paramètre du modèle. Ensuite, nous proposons un test d'adéquation du modèle à risques proportionnels, ainsi que des bandes de confiance pour le processus à l'intérieur desquelles le processus évolue si les risques sont proportionnels. Ces bandes de confiance sont construites à partir de la loi limite du processus connue analytiquement. Souvent, à défaut de connaître la loi limite explicitement, les bandes de confiance des processus utilisées pour vérifier l'adéquation du modèle à risques proportionnels sont construites par simulation (Lin *et al.*, 1993, Scheike et Martinussen, 2004). Les performances du test d'adéquation développé dans cette thèse sont comparées par simulations numériques aux performances de tests de références. Les simulations indiquent que le test d'adéquation ne dépend pas de la corrélation entre les covariables, ce qui n'est pas le cas de tous les autres tests. Ceci permet de valider l'application du test en présence de covariables corrélées. De plus, les simulations permettent de conclure que notre test est plus puissant que les tests concurrents évalués dans les situations considérées.

Enfin, nous proposons une méthode de construction du modèle à risques non proportionnels à l'aide de l'adéquation et de la capacité de prédiction du modèle. L'adéquation est étudiée avec le graphique du processus du score standardisé en fonction du temps et la qualité de la prédiction est estimée par le coefficient R^2 . Ces outils sont utilisés pour développer une méthode de construction d'un modèle présentant une bonne adéquation et une prédiction optimale sur l'ensemble des modèles candidats. Nous montrons que le coefficient de régression du modèle sélectionné est le plus proche du vrai coefficient de régression au sens de la distance L^2 . Le coefficient de régression obtenu est un estimateur explicite et simple à interpréter du coefficient de régression. Le choix des modèles candidats est laissé à l'utilisateur afin d'obtenir un coefficient de régression facilement interprétable, même si une estimation non-paramétrique et plus automatique est envisageable. La méthode de construction de modèle est également illustrée par des simulations et des applications sur des données réelles.

Une partie des résultats de ce chapitre a été présentée dans Chauvel et O'Quigley (2014b). Nous avons développé le Package PHeval¹ programmé en langage R qui permet d'évaluer et de tracer le processus du score standardisé ainsi que d'estimer le coefficient R^2 .

Annexe

L'Annexe expose la méthode développée durant la thèse pour simuler numériquement des données sous le modèle à risques non proportionnels. Deux méthodes sont proposées. La première permet de simuler des données sous un modèle de rupture. La deuxième considère un modèle à risques non proportionnels plus général.

1. Disponible à l'adresse : <http://cran.r-project.org/web/packages/PHeval/index.html>

Liste des notations et abréviations

Notations

\mathcal{W}	Mouvement Brownien sur $[0, 1]$
\mathcal{B}	Pont Brownien sur $[0, 1]$
$\xrightarrow{\mathcal{L}}$	Convergence en distribution
$\xrightarrow{\mathbb{P}}$	Convergence en probabilité
$\xrightarrow{L^1}$	Convergence en moyenne
<i>p.s.</i>	Presque sûrement
$\mathbf{1}_A$	Fonction indicatrice de l'ensemble A
$\mathcal{M}_{p \times p}(\mathbb{R})$	Ensemble des matrices de dimension $p \times p$ à coefficients réels
$\ a\ $	Norme sup du vecteur $a \in \mathbb{R}^p$
$\ A\ $	Norme sup de la matrice $A \in \mathcal{M}_{p \times p}(\mathbb{R})$

Abréviations

\hat{F}	Estimateur de Kaplan–Meier de F
n	Taille de l'échantillon
p	Nombre de covariables

T_i	Temps de décès de l'individu i
X_i	Temps observé de l'individu i
δ_i	Indicateur de censure de l'individu i
Z_i	Vecteur de p covariables de l'individu i
$Z_i^{(l)}$	$l^{\text{ème}}$ covariable de l'individu i , $l = 1, \dots, p$
k_n	Nombre de décès observés sans épuisement des groupes d'individus
β	Coefficient de régression du modèle à risques proportionnels
$\beta(t)$	Coefficient de régression du modèle à risques non proportionnels
$Z(t)$	Covariable de l'individu qui décède au temps t
$\pi_i(\beta(t), t)$	Probabilité pour que l'individu i décède au temps de décès t sous le modèle à risques non proportionnels de paramètre $\beta(t)$
$\mathcal{E}_{\beta(t)}(Z t)$	Espérance de la covariable de l'individu qui décède au temps t sous le modèle à risques non proportionnels de paramètre $\beta(t)$
$\mathcal{V}_{\beta(t)}(Z t)$	Variance de la covariable de l'individu qui décède au temps t sous le modèle à risques non proportionnels de paramètre $\beta(t)$
$U_n^*(\beta(t), t)$	Processus du score standardisé évalué au paramètre $\beta(t)$ et au temps t

Chapitre 1

Introduction

1.1 Introduction à l'analyse de survie

1.1.1 Introduction

L'analyse de survie est une spécialité importante des biostatistiques qui consiste à étudier des durées. L'étude de ces durées peut, par exemple, permettre de comparer les temps jusqu'à la guérison, la rechute ou encore le décès de différents patients. Dans un essai clinique, l'efficacité d'un nouveau traitement peut être déterminée en évaluant si la durée de vie moyenne des patients a été rallongée après la prise du traitement. Dans les études épidémiologiques, l'analyse de survie permet d'évaluer si un ou plusieurs facteurs de risque sont liés à la durée de vie. Ceci peut amener à limiter l'exposition à des facteurs de risque environnementaux, lorsque cela est possible, afin d'allonger la durée de vie. On désigne communément l'évènement auquel on s'intéresse par le terme générique de décès et la durée est appelée durée de vie.

L'analyse de survie peut également permettre d'établir des prédictions sur les temps de survie des patients. Ceci peut se faire de manière non paramétrique, par exemple avec l'estimateur de Kaplan–Meier de la survie ([Kaplan et Meier, 1958](#)), ou de manière paramétrique ou semi-paramétrique par l'utilisation de modèles pronostiques. Ces prédictions ont une importance capitale dans le domaine de la santé. Ils permettent aux patients d'être informés sur leur état de santé futur (pronostic du médecin) à partir de leur état de santé actuel (diagnostic établi par le médecin). Parfois, le traitement du patient peut être adapté en fonction de ces prédictions. Par exemple, en oncologie, les effets secondaires de certains traitements sont lourds, amenant à réserver ces thérapies aux patients dont la probabilité de survivre après un temps donné est faible.

La particularité de l'analyse de survie, hormis le fait d'étudier des variables aléatoires positives, réside dans la présence de censure : certaines durées ne sont pas entièrement observées. Par exemple, lors d'une étude, si un patient ne se rend plus aux rendez-vous fixés avec le médecin, il sera désigné comme "perdu de vue". Après la dernière visite à laquelle il s'est rendu, aucune information supplémentaire n'est disponible pour ce patient. Aussi, l'étude peut se terminer alors que des patients n'ont pas encore subi l'évènement d'intérêt. Ces patients sont

appelés les "exclus vivants". Dans ces deux situations, on n'observe pas le temps de survie du patient, la seule information disponible est que le patient a survécu jusqu'à une date connue (la dernière visite ou la date de fin de l'étude). Les estimateurs usuels de la statistique ne peuvent pas être utilisés en présence de censure, comme nous allons le voir dans la Section 1.1.3 avec l'exemple de l'estimateur de la survie empirique qui est remplacé par l'estimateur de Kaplan–Meier. Définissons d'abord les notations qui sont utilisées dans cette thèse.

1.1.2 Notation

L'échantillon est composé de n observations. Pour le patient $i \in \{1, \dots, n\}$, la durée observée est notée X_i . Nous faisons l'hypothèse de censure à droite et considérons que cette durée est le minimum entre son temps de survie T_i et un temps de censure C_i : $X_i = \min(T_i, C_i)$. On observe également un indicateur de décès $\delta_i = \mathbf{1}_{T_i \leq C_i}$, où $\mathbf{1}$ est la fonction indicatrice : $\delta_i = 1$ si le décès du patient i a été observé et $\delta_i = 0$ si la durée de vie du patient i a été censurée. On dispose d'un vecteur de covariables Z_i pouvant correspondre à l'âge du patient, son sexe, un niveau de biomarqueur mesuré, le stade de sa maladie, etc. Formellement, l'échantillon est noté

$$\{(X_i, \delta_i, Z_i), i = 1, \dots, n\},$$

où les n triplets d'observations sont des réalisations indépendantes et identiquement distribuées du vecteur aléatoire (X, δ, Z) , avec $X = \min(T, C)$ et $\delta = \mathbf{1}_{T \leq C}$. On suppose que les supports des variables aléatoires C et T sont finis dans \mathbb{R}^+ et on note $[0, \mathcal{T}]$ le support de T et F sa fonction de répartition.

Le vecteur $Z = (Z^{(1)}, Z^{(2)}, \dots, Z^{(p)})$ est un vecteur de p covariables, appelées aussi variables explicatives. Lorsque la covariable $Z^{(j)}$ dépend du temps, $Z^{(j)}$ est une fonction de $[0, \mathcal{T}]$ dans \mathbb{R} , sinon $Z^{(j)} \in \mathbb{R}$, pour $j = 1, \dots, p$. On suppose que le vecteur aléatoire $Z^{(j)}$ est continu à gauche :

$$\lim_{\varepsilon \rightarrow 0, \varepsilon > 0} Z^{(j)}(t - \varepsilon) = Z^{(j)}(t), \quad \forall t \in [0, \mathcal{T}],$$

et est de variance non nulle :

$$V(Z^{(j)}(t)) > 0, \quad \forall t \in [0, \mathcal{T}].$$

Si le vecteur de covariables dépend du temps, on le note $Z = Z(t)$. Dans cette thèse, pour des raisons de clarté, nous notons $a(t)$ toute fonction a de $[0, \mathcal{T}]$ dans \mathbb{R}^p qui n'est pas constante (dans le Chapitre 3 et les suivants, on considèrera les fonctions de $[0, 1]$ dans \mathbb{R}^p). Les fonctions constantes seront dénotées a de manière à distinguer les deux cas aisément.

Pour tout $i = 1, \dots, n$, on note $Y_i(t) = \mathbf{1}_{X_i \geq t}$ l'indicateur que l'individu i soit à risque de décéder au temps t . Il vaut 1 si l'individu i n'a pas subi l'évènement d'intérêt avant t et n'a pas été censuré, et 0 sinon. On note

$$N_i(t) = \mathbf{1}_{T_i \leq t, T_i \leq C_i},$$

le processus de comptage valant 1 à partir du temps de décès de l'individu i et 0 avant. Ce processus est toujours nul si l'individu est censuré. La somme de ces processus de comptage individuels permet de définir le processus de comptage

$$\bar{N}(t) = \sum_{i=1}^n N_i(t),$$

qui a un saut de taille 1 à chaque temps de décès. Dans notre étude, on suppose qu'il n'y a pas d'ex aequo : chaque temps de décès correspond au décès d'un seul individu.

On note $\mathcal{M}_{p \times p}(\mathbb{R})$ l'ensemble des matrices de dimension $p \times p$ à coefficients réels. Soit $A \in \mathcal{M}_{p \times p}(\mathbb{R})$. Le produit $A^{\otimes 2}$ est la matrice dont l'élément (i, j) est $A_{i,j}^{\otimes 2} = A_i A_j$ pour $i, j = 1, \dots, p$. De plus, $\|A\| = \max\{|A_{i,j}|, i, j = 1, \dots, p\}$ est la norme sup de la matrice A et $\|a\| = \max\{|a_i|, i = 1, \dots, p\}$ la norme sup du vecteur $a \in \mathbb{R}^p$. Dans la suite du manuscrit, sauf mention du contraire, les normes des vecteurs et matrices seront leur norme sup.

1.1.3 Estimateur de Kaplan–Meier de la survie

En présence de censure, les durées complètes T_1, \dots, T_n ne sont pas toutes observées. Ainsi, l'estimateur usuel de la fonction de répartition empirique de T ,

$$\hat{F}_{\text{emp}}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{T_i \leq t}, \quad 0 \leq t \leq \mathcal{T},$$

ne peut pas être évalué. Sans données censurées, cet estimateur sans biais de $F(t) = P(T \leq t)$ peut être utilisé. Une solution pour estimer F pourrait être de considérer les temps censurés comme des décès et d'utiliser l'estimateur

$$1 - \frac{1}{n} \sum_{i=1}^n Y_i(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq t}, \quad 0 \leq t \leq \mathcal{T}.$$

Cet estimateur de la fonction de répartition de T est biaisé si l'échantillon comporte de la censure. Son espérance au temps t vaut $P(T \leq t, C \leq t)$, menant à une sous-estimation de $P(T \leq t)$. [Kaplan et Meier \(1958\)](#) ont introduit un estimateur convergent de F qui tient compte de la censure.

Définition 1.1.1. *L'estimateur de Kaplan–Meier de F , noté \hat{F} , est défini par*

$$\hat{F}(t) = 1 - \prod_{\substack{i=1, \dots, n \\ X_i \leq t}} \left(1 - \frac{\delta_i}{\sum_{j=1}^n Y_j(X_i)} \right), \quad 0 \leq t \leq \mathcal{T}.$$

Cet estimateur est une fonction constante par morceaux, continue à droite et limite à gauche, avec un saut à chaque temps de décès observé. Les sauts sont croissants ce qui permet de compenser le manque d'observations dans les queues de la distribution, c'est-à-dire pour les grandes valeurs du temps. Si aucune donnée n'est censurée, l'estimateur se réduit à l'estimateur

de la fonction empirique \hat{F}_{emp} . Notons que l'estimateur de Kaplan–Meier est convergent en faisant l'hypothèse d'indépendance entre la censure C et la durée jusqu'à l'évènement d'intérêt T . L'estimateur de Kaplan–Meier de la fonction de survie de T au temps t , $P(T \geq t)$, est défini par

$$\hat{S}(t) = 1 - \hat{F}(t). \quad (1.1)$$

Illustrons le calcul de l'estimateur sur un exemple. Supposons que l'on dispose des 7 observations de la Table 1.1.

X	2	4	7	9	13	16	22
δ	1	1	0	1	0	1	0

TABLE 1.1 – Données pour l'exemple du calcul de l'estimateur de Kaplan–Meier

Les temps 7, 13 et 22 sont des temps de censure et les autres temps correspondent à des temps de décès. La Table 1.2 illustre la construction de l'estimateur de Kaplan–Meier. Dans cette table sont recensées les valeurs de l'estimateur aux temps observés. Si $t \in [X_i, X_{i+1}[$ alors $\hat{S}(t) = \hat{S}(X_i)$. On remarque que la valeur de l'estimateur ne change pas aux temps censurés.

X_i	$\sum_{j=1}^n Y_j(X_i)$	$1 - \delta_i / \sum_{j=1}^n Y_j(X_i)$	$\hat{S}(X_i)$
2	7	$1 - 1/7$	$1 - 1/7 \simeq 0.857$
4	6	$1 - 1/6$	$(1 - 1/7)(1 - 1/6) \simeq 0.714$
7	5	1	0.714
9	4	$1 - 1/4$	$(1 - 1/7)(1 - 1/6)(1 - 1/4) \simeq 0.536$
13	3	1	0.536
16	2	$1 - 1/2$	$(1 - 1/7)(1 - 1/6)(1 - 1/4)(1 - 1/2) \simeq 0.268$
22	1	1	0.268

TABLE 1.2 – Exemple de calcul de l'estimateur de Kaplan–Meier

La Figure 1.1 représente l'estimateur de Kaplan–Meier \hat{S} en fonction du temps. Notons que l'estimateur n'atteint pas la valeur 0 car le dernier temps observé est un temps censuré.

L'amplitude du saut de l'estimateur de Kaplan–Meier au temps t est notée $d\hat{S}(t)$. Le résultat suivant permet de quantifier l'amplitude du saut de \hat{S} à chaque instant de décès.

Proposition 1.1.1. *Soit t un temps de décès observé. Nous avons la relation*

$$d\hat{S}(t) = \hat{S}(t) - \hat{S}(t^-) = \frac{\hat{S}(t^-)}{\sum_{j=1}^n Y_j(t)}, \quad (1.2)$$

où $\hat{S}(t^-) = \lim_{s \rightarrow t^-} \hat{S}(s)$.

Preuve de la Proposition 1.1.1. Soit $l \in \{1, \dots, n\}$. Supposons que l'individu l décède au temps t : on observe les réalisations $X_l = t$ et $\delta_l = 1$.

$$\hat{S}(t) - \hat{S}(t^-) = \prod_{\substack{i=1, \dots, n \\ X_i \leq t}} \left(1 - \frac{\delta_i}{\sum_{j=1}^n Y_j(X_i)}\right) - \prod_{\substack{i=1, \dots, n \\ X_i \leq t^-}} \left(1 - \frac{\delta_i}{\sum_{j=1}^n Y_j(X_i)}\right)$$

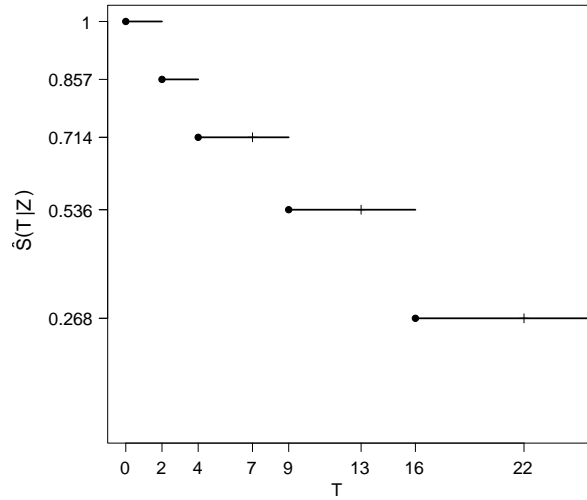


FIGURE 1.1 – Exemple d'estimateur de Kaplan–Meier en fonction du temps

$$\begin{aligned}
 &= \prod_{\substack{i=1, \dots, n \\ X_i \leq t^-}} \left(1 - \frac{\delta_i}{\sum_{j=1}^n Y_j(X_i)} \right) \left(1 - \frac{\delta_l}{\sum_{j=1}^n Y_j(X_l)} - 1 \right) \\
 &= \hat{S}(t^-) \frac{\delta_l}{\sum_{j=1}^n Y_j(t)} = \frac{\hat{S}(t^-)}{\sum_{j=1}^n Y_j(t)}.
 \end{aligned}$$

□

Les incréments de l'estimateur de Kaplan–Meier peuvent être utilisés pour pondérer des variables dépendantes du temps intégrées par rapport à \bar{N} . Les intégrales résultantes, dites intégrales de Kaplan–Meier, convergent alors vers l'espérance de fonctions de T et ne sont pas biaisées en présence de censure. [Stute \(1995\)](#) a montré la normalité asymptotique des intégrales de Kaplan–Meier sous de bonnes conditions. Pour ajuster un modèle à risques proportionnels de paramètre β alors que l'effet $\beta(t)$ inconnu change avec le temps, [Xu et O'Quigley \(2000\)](#) ont pondéré l'équation estimatrice du coefficient de régression du modèle à risques proportionnels par les incréments $d\hat{S}$. L'estimateur obtenu converge vers $E(\beta(T))$. Sans les pondérations, les auteurs montrent que l'estimateur de la vraisemblance partielle de Cox ne converge pas vers $E(\beta(T))$ lorsque l'effet $\beta(t)$ n'est pas fixe dans le temps.

Il est possible d'exprimer l'estimateur de Kaplan–Meier de la fonction de répartition de T comme une somme pondérée, avec des poids dépendants de l'estimateur de Kaplan–Meier de la fonction de répartition de la censure ([Satten et Datta, 2001](#)). On peut montrer que l'estimateur de Kaplan–Meier est uniformément convergent et asymptotiquement normal.

L'analyse de survie consiste à modéliser des durées, c'est-à-dire des variables aléatoires positives. De nombreux modèles ont été développés pour cette modélisation. Dans ce travail, nous considérons le cas d'un modèle semi-paramétrique très fréquemment utilisé en analyse de survie, le modèle à risques proportionnels ainsi que son extension à des risques non proportionnels. En Section 1.2, nous présentons les deux modèles et nous nous intéressons particulièrement aux

méthodes d'estimation qui ont été proposées dans ces modèles, avant de présenter quelques modèles alternatifs utilisés en analyse de survie. En Section 1.3, nous introduisons le test du log-rank. Ce test, très répandu dans les applications, permet de tester si les temps de survie de plusieurs groupes de patients sont différents. Une bibliographie sur des alternatives à ce test ainsi que ses améliorations seront présentées au Chapitre 4. Enfin, la Section 1.4 abordera les questions de l'adéquation et de la prédiction dans le modèle à risques proportionnels, qui seront approfondies au Chapitre 5.

1.2 Modèle à risques proportionnels et extensions

1.2.1 Modèle à risques proportionnels

Le modèle à risques proportionnels est défini à partir du risque instantané. Introduisons cette quantité avant d'expliquer les particularités du modèle.

Définition 1.2.1. *Le risque instantané, ou taux de hasard, de la variable aléatoire positive X est la fonction λ telle que*

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(X \in [t, t+h[\mid X \geq t)}{h}, \quad t \geq 0.$$

Si X représente la durée de vie d'un individu, le risque instantané de X calculé au temps t s'interprète comme la probabilité que l'individu décède juste après t , sachant qu'il a survécu jusqu'en t . Ce risque définit de manière unique chaque distribution, au même titre que la densité ou la fonction de répartition. Ceci justifie la définition d'un modèle à partir de cette quantité. On peut montrer que la fonction de survie et le risque instantané sont liés par la relation

$$S(t) = \exp\left(-\int_0^t \lambda(s) ds\right), \quad t \geq 0. \quad (1.3)$$

Nous rappelons que $Z(t)$ représente un vecteur de p covariables pouvant dépendre du temps. Par exemple, il peut s'agir du sexe du patient, de son âge lors de l'inclusion dans l'étude, de niveaux de biomarqueurs relevés à différentes dates, du stade de sa maladie, etc. Des modèles paramétriques ont été considérés pour relier les covariables au temps de survie, par exemple les modèles de vie accélérée avec une fonction de survie de base connue. L'avantage de ce modèle réside dans la simplicité d'estimation et d'interprétation des paramètres, mais il est très restrictif. Nous reviendrons plus en détail sur ce modèle dans la Section 1.2.3. Pour assouplir la relation entre Z et T , il est envisageable d'utiliser des méthodes non paramétriques. Par exemple, [Stute \(1993\)](#) a proposé un estimateur de type Kaplan–Meier pour estimer la fonction de répartition $P(T \leq t, Z \leq z)$. L'interprétation devient alors plus délicate. L'utilisation d'un modèle semi-paramétrique permet à la fois d'interpréter aisément les paramètres liant le temps de survie T aux covariables Z , tout en gardant un lien souple entre les deux. Le modèle semi-paramétrique que nous considérons est le modèle à risques proportionnels avec une fonction de lien exponentielle entre les covariables et le risque instantané.

Définition 1.2.2. *Le modèle à risques proportionnels est défini par*

$$\lambda(t | Z) = \lambda_0(t) \exp\{\beta^T Z(t)\}, \quad t \geq 0, \quad (1.4)$$

où λ_0 est un risque instantané de base et β^T est le vecteur transposé du vecteur $\beta \in \mathbb{R}^p$.

Le vecteur β est le vecteur de coefficients de régression associé à $Z(t)$. Ce modèle a été introduit par [Cox \(1972\)](#) qui a considéré des covariables Z constantes au cours du temps. Le modèle (1.4) est couramment appelé "modèle de Cox". L'extension à des covariables dépendantes du temps est immédiate. Avant de discuter des hypothèses du modèle, nous présentons la méthode d'estimation du coefficient de régression β . L'intérêt de ce modèle semi-paramétrique réside dans la séparation de la fonction temporelle λ_0 du coefficient de régression β . Ainsi, en maximisant la vraisemblance partielle de [Cox \(1972, 1975\)](#), l'estimation du paramètre β peut se faire sans spécifier ni estimer la fonction λ_0 . L'estimation du risque de base cumulé, $\int \lambda_0(s) ds$, peut s'effectuer notamment avec l'estimateur de Nelson–Aalen, mais nous ne donnons pas plus de détails ici ([Aalen, 1978](#), [Nelson, 1972](#)).

En conditionnant sur l'existence d'un décès au temps t et sur la connaissance des individus encore à risque de décéder à ce temps, la probabilité que l'individu i décède en t est

$$\pi_i(\beta, t) = \frac{Y_i(t) \exp\{\beta^T Z_i(t)\}}{\sum_{j=1}^n Y_j(t) \exp\{\beta^T Z_j(t)\}}, \quad i = 1, \dots, n. \quad (1.5)$$

Remarque 1.2.1. *En pratique, la covariable de l'individu i , $Z_i(\cdot)$, n'est pas observée aux temps supérieurs à X_i , ce qui implique que la covariable $Z_i(t)$ n'est pas définie pour $t > X_i$, $i = 1, \dots, n$. Nous choisissons d'associer la dernière valeur observée de la covariable pour les temps supérieurs au temps de décès. Dans la formule précédente, cela revient à remplacer $Z_i(t)$ par $Z_i(\min(t, X_i))$, mais nous ne l'écrivons pas pour des raisons de lisibilité. Cette définition suppose que les valeurs des covariables des individus sont connues à tous les temps de décès jusqu'à ce qu'ils soient censurés ou décèdent à leur tour. Cette hypothèse est courante pour traiter des covariables dépendantes du temps ([Andersen et Gill, 1982](#)). Une autre manière de contourner le problème serait d'annuler la valeur de la covariable pour les temps supérieurs au temps observé : $Z_i(t) = 0$ si $t > X_i$, $i = 1, \dots, n$ ([Arjas, 1988](#)). Cette remarque s'applique également aux quantités définies ci-après.*

Notons que l'on peut calculer l'espérance et la variance de $Z(t)$ par rapport aux probabilités $\pi_i(\beta, t)$, $i = 1, \dots, n$.

Définition 1.2.3. *Soit $t \in [0, \mathcal{T}]$. L'espérance et la variance de Z par rapport à la famille de probabilité $\{\pi_i(\beta, t), i = 1, \dots, n\}$ sont respectivement un vecteur de \mathbb{R}^p et une matrice de $\mathcal{M}_{p \times p}(\mathbb{R})$ tels que*

$$\mathcal{E}_\beta(Z | t) = \sum_{i=1}^n Z_i(t) \pi_i(\beta, t) = (\mathcal{E}_\beta(Z | t)_1, \dots, \mathcal{E}_\beta(Z | t)_p), \quad (1.6)$$

et

$$\mathcal{V}_\beta(Z | t)_{a,b} = \sum_{i=1}^n Z_i^{(a)}(t) Z_i^{(b)}(t) \pi_i(\beta, t) - \mathcal{E}_\beta(Z | t)_a \mathcal{E}_\beta(Z | t)_b, \quad (1.7)$$

où $\mathcal{V}_\beta(Z | t)_{a,b}$ est l'élément en a ème ligne et b ème colonne de la matrice $\mathcal{V}_\beta(Z | t)$, avec $a, b = 1, \dots, p$.

L'espérance $\mathcal{E}_\beta(Z | t)$ correspond à la valeur attendue de la covariable de l'individu qui décède en t sous le modèle à risques proportionnels de paramètre β , où t est un temps de décès. Cette quantité interviendra pour calculer des résidus et elle sera comparée à la valeur observée de la covariable de l'individu qui décède. Nous y reviendrons en Section 1.4.

Chaque contribution à la vraisemblance partielle de Cox sous le modèle à risques proportionnels (1.4) de paramètre β , notée $\mathcal{L}_n(\beta)$, est la probabilité $\pi_i(\beta, X_i)$ lorsque X_i correspond au temps de décès de l'individu i :

$$\mathcal{L}_n(\beta) = \prod_{i=1}^n \pi_i(\beta, X_i)^{\delta_i}.$$

L'estimation du paramètre β s'effectue en maximisant la log-vraisemblance partielle notée l_n telle que

$$l_n(\beta) = \log(\mathcal{L}_n(\beta)) = \sum_{i=1}^n \int_0^{\mathcal{T}} \left\{ \beta^T Z_i(t) - \log \left(\sum_{j=1}^n Y_j(t) \exp(\beta^T Z_j(t)) \right) \right\} dN_i(t).$$

Il s'agit alors de trouver le vecteur β qui annule le score $U(\beta)$, c'est-à-dire la dérivée de la log-vraisemblance partielle en β :

$$U(\beta) = \frac{\partial}{\partial \beta} l_n(\beta) = \sum_{i=1}^n \int_0^{\mathcal{T}} \{Z_i(t) - \mathcal{E}_\beta(Z | t)\} dN_i(t). \quad (1.8)$$

La dérivée seconde de la log-vraisemblance partielle par rapport au vecteur β est négative puisqu'il s'agit de l'opposé de $\mathcal{I}(\beta)$, où

$$\frac{\partial}{\partial \beta} U(\beta) = -\mathcal{I}(\beta) = -\sum_{i=1}^n \mathcal{V}_\beta(Z | X_i)^{\delta_i} < 0.$$

La log-vraisemblance partielle étant strictement concave, le vecteur β qui annule le score réalise le maximum de la vraisemblance partielle. On note ce vecteur $\hat{\beta}$ et on a $U(\hat{\beta}) = 0$. De nombreux ouvrages traitent de l'estimation du paramètre, nous renvoyons à l'un d'entre eux pour plus de détails ([Kalbfleisch et Prentice, 2002](#)).

Le modèle à risques proportionnels que nous considérons est particulier, car il repose sur l'hypothèse de log-linéarité entre les covariables et la fonction de risque. En relâchant cette hypothèse, nous obtenons encore un modèle à risques proportionnels défini par

$$\lambda(t | Z) = \lambda_0(t) r\{\beta^T Z(t)\},$$

où r est une fonction positive comme, par exemple, $r(x) = 1 + x$ ou $r(x) = 1/(1 + x)$ (Taulbee, 1979). Sous ce modèle, les risques sont toujours proportionnels mais le logarithme du risque instantané, s'il est fini, n'est pas linéaire en les paramètres. Le modèle à risques proportionnels défini équation (1.4), dit modèle à risques proportionnels de Cox, se retrouve en choisissant $r(x) = \exp(x)$. Prentice et Self (1983) ont étudié le comportement asymptotique de l'estimateur du maximum de vraisemblance partielle de ce modèle plus général. Des illustrations du modèle sont présentées dans Andersen *et al.* (1993). Taulbee (1979) a proposé une extension du modèle à risques proportionnels en considérant un risque instantané polynomial de la forme

$$\lambda(t | Z) = \sum_{k=0}^m \lambda_k t^k h(\beta_k, Z), \quad m \in \mathbb{N}^*,$$

où h est une fonction positive à choisir. Il a notamment considéré le cas $h(\beta_k, Z) = \exp(\beta_k^T Z)$. Il s'agit alors d'estimer paramétriquement $\beta_0, \beta_1, \dots, \beta_m$ et $\lambda_0, \lambda_1, \dots, \lambda_m$, dans un modèle très souple et moins restrictif que le modèle à risques proportionnels (1.4).

En pratique, le modèle à risques proportionnels (1.4) est très utilisé, de par sa simplicité d'interprétation. En effet, si $p = 1$ et Z est une variable de Bernoulli, alors

$$\frac{\lambda(t | Z = 1)}{\lambda(t | Z = 0)} = \exp(\beta), \quad t \geq 0. \quad (1.9)$$

Ainsi, le risque instantané de décès relatif entre les groupes correspondants à $Z = 1$ et $Z = 0$ est $\exp(\beta)$. Le risque relatif est une mesure très utilisée en épidémiologie, notamment dans les enquêtes de cohorte visant à trouver des facteurs de risque associés à une certaine maladie. Dans le modèle de Cox, le risque relatif est constant au cours du temps d'après l'équation (1.9), c'est pourquoi le modèle est aussi appelé modèle à risques proportionnels. Cette dernière appellation est plus globale, car elle autorise les covariables à dépendre du temps. L'équation (1.3) permet d'écrire le modèle à risques proportionnels (1.4) à partir de la fonction de survie conditionnelle

$$S(t | Z) = \exp \left\{ -\exp(\beta^T Z) \int_0^t \lambda_0(s) ds \right\}, \quad t \geq 0. \quad (1.10)$$

Notons $\Lambda(t | Z) = \int_0^t \lambda(s | Z) ds$ le risque instantané cumulé conditionnel aux covariables au temps t . La Figure 1.2 représente les courbes de survie $S(t | Z)$ et les log risques instantanés cumulés $\log(\Lambda(t | Z))$ sous le modèle à risques proportionnels en fonction du temps t . Dans cet exemple, Z est une variable catégorielle prenant les modalités 0 ou 1 et λ_0 est une fonction constante. L'hypothèse de proportionnalité des risques est visible sur le graphique du logarithme des risques cumulés puisque les deux courbes parallèles. L'écart entre les courbes de survie est croissant en fonction du temps, ce qui traduit graphiquement l'équation (1.10) où les deux fonctions de survie conditionnelles sont liées par une relation puissance.

L'hypothèse des risques proportionnels, appréciable pour l'interprétation, est très forte. Supposons que les deux groupes de patients représentent deux groupes de traitement : un groupe reçoit le placebo et l'autre un nouveau traitement. Si le modèle est respecté, alors

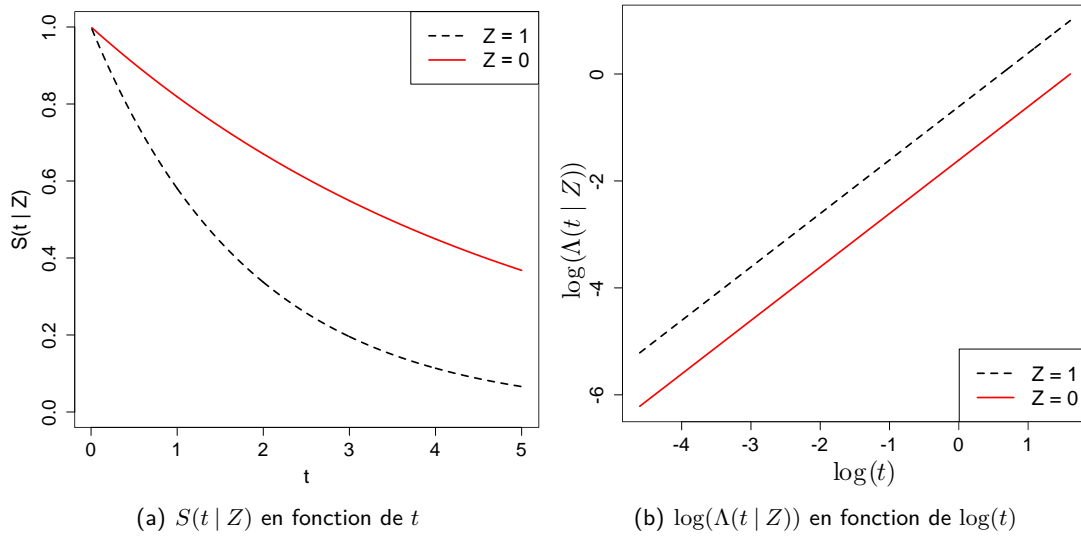


FIGURE 1.2 – Exemple de courbes de survie et de risque instantané sous le modèle à risques proportionnels.

l'effet entre les groupes est constant au cours du temps. Or, il a été souvent observé que l'effet d'un médicament peut s'estomper avec le temps à cause de l'accoutumance. A l'inverse, l'effet d'un traitement peut être tardif. Enfin, un cas plus particulier est celui d'un effet qui change de signe. Ce cas peut être illustré par l'exemple d'un traitement chirurgical, dont l'effet est négatif dans un premier temps puisque le risque de décès est élevé lors de l'opération. Par la suite, les patients ont survécu à l'opération et l'effet du traitement devient positif. Utiliser les résultats d'estimation du modèle à risques proportionnels alors que le vrai effet dépend du temps peut mener à des erreurs d'interprétation des données. L'hypothèse des risques proportionnels est très forte et peut être assouplie en considérant un effet qui dépend du temps.

1.2.2 Modèle à risques non proportionnels

Les modèles à risques non proportionnels sont plus souples que le modèle à risques proportionnels, car ils ne font pas l'hypothèse d'un risque relatif constant au cours du temps. Beaucoup de modèles font partie de la classe de modèles à risques non proportionnels. Dans cette thèse, nous appelons modèle à risques non proportionnels une extension du modèle à risques proportionnels de Cox dans lequel le coefficient de régression peut varier avec le temps.

Définition 1.2.4. *Le modèle à risques non proportionnels est défini par*

$$\lambda(t|Z) = \lambda_0(t) \exp\{\beta(t)^T Z(t)\}, \quad (1.11)$$

où λ_0 est un risque instantané de base et $\beta(t) = (\beta_1(t), \dots, \beta_p(t))$ un vecteur de fonctions de $[0, T]$ dans \mathbb{R}^p .

Le modèle à risques proportionnels est un cas particulier de ce modèle plus général, qui se

retrouve lorsque $\beta(t)$ est constant. Plusieurs méthodes ont été proposées pour estimer $\beta(t)$. Nous les classons selon qu'elles consistent en une maximisation simple de la vraisemblance partielle, une maximisation d'une transformation de la vraisemblance ou qu'elles soient itératives.

Maximisation de la vraisemblance partielle

La première méthode consiste à transformer la fonction $\beta(t)$ en un vecteur de paramètres et à maximiser la vraisemblance partielle en ces paramètres. La transformation de $\beta(t)$ en un vecteur de paramètres peut se faire en considérant des coefficients de régression de la forme

$$\beta_i(t) = \beta_{0,i} h_i(t), \quad \text{pour } i = 1, \dots, p,$$

avec h_i une fonction déterministe, par exemple, linéaire (Cox, 1972), constante par morceaux (Brown, 1975) ou polynomiale (Stablein *et al.*, 1981). Le modèle s'écrit alors

$$\lambda(t | Z) = \lambda_0(t) \exp \left(\sum_{i=1}^p \beta_{0,i} h_i(t) Z^{(i)}(t) \right).$$

La partie temporelle du coefficient de régression, $h_i(t)$, peut être associée à la i ème covariable $Z^{(i)}(t)$, ce qui se ramène à estimer le vecteur de paramètres $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,p})$ associé au vecteur de covariables dépendantes du temps $(h_1(t)Z^{(1)}(t), \dots, h_p(t)Z^{(p)}(t))$.

Anderson et Senthilselvan (1982) ont étudié un effet de la forme $\beta(t) = \alpha \mathbf{1}_{t \leq B} + \gamma \mathbf{1}_{t > B}$, en estimant les paramètres α , γ et B par maximisation de la vraisemblance partielle. Plus généralement, les modèles de rupture sont des cas particuliers de modèles à risques non proportionnels dans lesquels le coefficient $\beta(t)$ est constant par morceaux, avec des temps de rupture spécifiés ou non (O'Quigley, 2008, chap. 12).

Moreau *et al.* (1985) ont considéré l'effet $\beta(t) = \beta_0 + G(t)$ où G est une fonction constante par morceaux avec des ruptures à des temps prédéfinis. Ils ont développé un test du score de l'hypothèse nulle $H_0 : G = 0$. Ce test permet de tester l'adéquation du modèle (1.4) puisque, sous H_0 , les risques sont proportionnels. Grambsch et Therneau (1994) ont étendu cette idée en considérant un effet de la forme $\beta(t) = \beta_0 + \theta g(t)$ et en proposant à la fois une classe de tests pour une fonction g déterministe ou aléatoire mais prévisible, ainsi qu'un estimateur du paramètre θ .

L'avantage des approches présentées jusqu'ici est d'obtenir un estimateur explicite de $\beta(t)$ et facile à interpréter, ce qui n'est pas toujours le cas avec les méthodes non paramétriques présentées dans la suite. Une autre possibilité pour transformer la fonction $\beta(t)$ en un vecteur de paramètres est de la projeter sur une base de fonctions connue avant de maximiser la vraisemblance partielle. Par exemple, Hess (1994) a utilisé une base de splines cubiques et Abrahamowicz *et al.* (1996) une base de splines quadratiques. Murphy et Sen (1991) ont projeté le coefficient sur une base d'histogrammes, l'estimateur résultant étant constant par morceaux sur des intervalles de temps prédéfinis. Murphy (1993) a développé un test du score basé sur cet estimateur, pour tester l'adéquation du modèle à risques proportionnels (1.4).

Autrement dit, tester l'égalité des différents coefficients projetés sur la base revient à tester si les données respectent bien l'hypothèse de proportionnalité des risques. Marzec et Marzec (1997) ont proposé des tests d'adéquation du modèle à risques non proportionnels basés sur un processus de résidus de martingales avec l'estimateur de $\beta(t)$ de Murphy et Sen (1991). Nous reviendrons sur ces deux tests d'adéquation au Chapitre 5.

Maximisation d'une transformation de la vraisemblance partielle

La seconde catégorie de méthodes d'estimation du coefficient $\beta(t)$ du modèle à risques non proportionnels consiste à maximiser la log-vraisemblance transformée, soit par pénalisation, soit par pondération. Le problème de maximisation en $\beta(t)$ de la log-vraisemblance pénalisée s'écrit

$$\max_{\beta(t) \in \mathcal{B}} \{l_n(\beta(t)) - \text{pen}(\beta(t))\}. \quad (1.12)$$

L'ensemble \mathcal{B} sur lequel est projeté $\beta(t)$ et la pénalisation pen changent selon les cas. L'utilisation d'une pénalisation permet de lisser l'estimateur résultant en imposant une certaine régularité sur sa structure.

Zucker et Karr (1990) ont considéré l'ensemble \mathcal{B} des fonctions m -différentiables par morceaux, pour m entier supérieur à 2 et une pénalisation dépendante de la m -différentielle de $\beta(t)$. Ils ont obtenu des résultats de convergence de l'estimateur résultant pour $m \geq 3$ et de normalité asymptotique pour $m \geq 4$.

Hastie et Tibshirani (1990) et Gray (1992) ont étudié la projection de $\beta(t)$ sur une base de splines cubiques avec une pénalisation de la log-vraisemblance partielle par la fonction

$$\text{pen}(\beta(t)) = \frac{1}{2} \sum_{i=1}^p \lambda_i \int \left(\frac{\partial^2}{\partial t^2} \beta_i(t) \right)^2 dt, \quad (1.13)$$

où $\lambda_1, \dots, \lambda_p$ sont des paramètres de lissage. Notons que Hastie et Tibshirani (1990) ont établi leur résultat pour d'autres liens que le produit scalaire entre $\beta(t)$ et Z . Ils ont notamment développé un algorithme pour la résolution de problème de maximisation (1.12). Gray (1992) a proposé un test de la nullité du projeté du paramètre $\beta(t)$ sur la base de splines. Hastie et Tibshirani (1993) ont étendu leur travail à une classe plus large de modèles avec des effets qui varient. Cette classe peut s'écrire par la relation

$$g(X) = \beta_0 + \beta_1(R_1)Z_1 + \dots + \beta_p(R_p)Z_p,$$

où X est la variable réponse, g est une fonction de lien, les variables Z_i et R_i sont des variables explicatives réelles et les β_i sont des fonctions de régression, $i = 1, \dots, p$. Cette classe comprend notamment les modèles linéaires généralisés et les modèles additifs généralisés. Par exemple, les effets peuvent être des fonctions du temps (si R_i représente la i ème mesure d'une expérience) ou des covariables Z_i . En choisissant $\beta_i(R_i)Z_i = \beta_i R_i Z_i$, cela revient à modéliser l'interaction entre les variables R_i et Z_i , $i = 1, \dots, p$.

Verweij et van Houwelingen (1995) ont considéré un modèle à risques proportionnels avec des coefficients différents pour chaque temps de décès. Les auteurs proposent de pénaliser la log-vraisemblance par la perte en variation totale avec la norme L^2 , qui correspond à la norme L^2 de la différence entre deux coefficients associés à des temps consécutifs. Ceci permet de lisser le coefficient $\beta(t)$ pour que deux valeurs successives du coefficient ne soient pas trop éloignées.

En utilisant un lissage de la log-vraisemblance et non une pénalisation, Cai et Sun (2003) proposent un lissage par un noyau pour des coefficients $\beta(t)$ localement linéaires, afin d'obtenir une estimation non paramétrique locale des coefficients autour de chaque temps de décès. Dans le modèle à risques proportionnels (1.4), plutôt que de pénaliser la log-vraisemblance partielle, il est possible de la pondérer en considérant le problème de maximisation suivant

$$\max_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \int H(u) \left(\beta^T Z_i(u) - \log \left(\sum_{j=1}^n Y_j(u) \exp(\beta^T Z_j(u)) \right) \right) dN_i(u),$$

où H est une fonction de poids prévisible. Dans une analyse de données avec le modèle à risques proportionnels de paramètre β , Valsecchi et Silvestri (1996) proposent notamment de pondérer la log-vraisemblance partielle par l'estimateur de Kaplan-Meier de la survie, qui décroît avec le temps, de manière à limiter l'influence des observations obtenues à la fin de l'étude. Les résultats asymptotiques de cet estimateur ont été étudiés par Sasieni (1993), dans le cas du modèle à risques proportionnels avec des conditions raisonnables sur les poids H . L'idée n'est pas d'estimer le vrai paramètre $\beta(t)$ mais de pondérer la log-vraisemblance afin de se ramener à un effet constant dans le temps. Xu et O'Quigley (2000) ont montré que choisir une pondération égale aux sauts de l'estimateur de Kaplan-Meier donnés en Proposition 1.1.1 permet d'estimer l'espérance de $\beta(t)$, $E(\beta(T)) = \int \beta(t) dF(t)$, lorsque le "vrai" modèle est à risques non proportionnels avec un coefficient de régression $\beta(t)$. En présence d'un coefficient de régression qui dépend du temps, l'estimateur de \mathbb{R}^p qui maximise la vraisemblance partielle non pondérée converge vers la solution d'une équation faisant intervenir la censure (Lin, 1991, Struthers et Kalbfleisch, 1986, Xu et O'Quigley, 2000). L'estimateur du maximum de vraisemblance est donc biaisé en présence de censure lorsque le modèle à risques non proportionnels est mal spécifié, comme l'ont illustré Xu et O'Quigley (2000) par un travail de simulation.

Méthodes itératives

La première méthode itérative que nous présentons a été proposée par Winnett et Sasieni (2003). En étendant les travaux de Grambsch et Therneau (1994), ils montrent que le coefficient $\beta(t)$ peut-être estimé par lissage des résidus du modèle à risques non proportionnels. Après avoir ajusté un modèle à risques proportionnels (1.4), un lissage des résidus de ce modèle permet d'obtenir une première estimation de $\beta(t)$. En itérant la combinaison du lissage des résidus du modèle avec le dernier coefficient $\beta(t)$ obtenu et d'une nouvelle estimation de $\beta(t)$,

les auteurs obtiennent un estimateur qui converge vers le paramètre $\beta(t)$.

[Gamerman \(1991\)](#) a proposé une approche bayésienne d'un modèle dynamique pour modéliser $\beta(t)$. Le temps est découpé en J intervalles I_1, \dots, I_J dont les extrémités ont été choisies indépendamment des données, comme proposé par [Kalbfleisch et Prentice \(1973\)](#). Ce choix pour les extrémités des intervalles est souvent mis en opposition avec celui de [Breslow \(1974\)](#) qui propose de considérer les temps de décès observés. Un modèle exponentiel par morceaux permet alors de modéliser les temps de décès, avec un coefficient β_j constant sur l'intervalle I_j ($j = 1, \dots, J$) :

$$\lambda(t | Z) = \exp(\beta_j^T Z), \quad t \in I_j, \quad j = 1, \dots, J.$$

Les coefficients β_j sont reliés par la série temporelle bayésienne

$$\beta_j = \beta_{j-1} + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \sigma_j^2). \quad (1.14)$$

Ce processus de Markov sur les paramètres β_j permet d'évaluer les termes de $\beta(t)$ séquentiellement dans le temps en utilisant la vraisemblance complète des observations. La loi de chaque β_j sachant le passé jusqu'au temps t_j devient la loi a priori de β_{j+1} sachant le passé jusqu'au temps t_j . Dans le même contexte, [Arjas et Gasbarra \(1994\)](#) ont choisi de relier les coefficients β_j par un processus autorégressif du premier ordre avec un bruit distribué selon une loi gamma. Plutôt que de considérer un risque instantané exponentiel par morceaux, [Sinha et al. \(1999\)](#) ont étudié un modèle à une covariable ($Z \in \mathbb{R}$) avec un risque instantané constant par morceaux :

$$\lambda(t | Z) = \lambda_j \exp(Z\beta_j), \quad t \in I_j, \quad j = 1, \dots, J,$$

où les λ_j sont des variables aléatoires de loi gamma, la série temporelle (1.14) relie les coefficients β_j et les β_j sont indépendants a priori des λ_k pour $j, k = 1, \dots, J$. Notons que des modèles bayésiens hiérarchiques ont également été proposés ([Gustafson, 1998](#)). Nous renvoyons à l'ouvrage d'[Ibrahim et al. \(2001\)](#) pour d'autres modélisations de données de survie à l'aide de méthodes bayésiennes.

Un algorithme d'estimation du paramètre cumulé $\int_0^t \beta(s) ds$ a été développé par [Scheike et Martinussen \(2004\)](#). Cet algorithme, de type Newton–Raphson, est basé sur l'équation estimatrice de $\beta(t)$ et fait intervenir un lissage par noyau. Les auteurs ont également développé un test de l'hypothèse de proportionnalité des risques pour une covariable, en laissant la possibilité aux autres covariables de varier dans le temps. Nous reviendrons sur ce test et cette propriété importante au Chapitre 5.

D'autres modèles alternatifs au modèle à risques proportionnels avec une fonction de lien exponentielle (1.4) et au modèle à risques non proportionnels (1.11) ont été proposés dans la littérature pour modéliser les durées de vie en prenant en compte des covariables. Nous en introduisons quelques-uns dans la partie suivante.

1.2.3 Alternatives au modèle à risques proportionnels

Modèles stratifiés

Une première possibilité pour étendre le modèle à risques proportionnels (1.4) est de le stratifier, en partitionnant la base de données en K groupes. Pour le groupe $k = 1, \dots, K$, on considère le modèle

$$\lambda_k(t | Z) = \lambda_{0,k}(t) \exp(\beta^T Z),$$

où $\lambda_k(t | Z)$ est le risque instantané de décès pour le groupe k au temps t conditionnellement à Z et $\lambda_{0,k}(t)$ est un risque de base pour le groupe k (Kalbfleisch et Prentice, 1973). Ainsi, l'hypothèse de proportionnalité globale des risques est assouplie en une hypothèse de proportionnalité des risques à l'intérieur de la strate k . Entre différentes strates, le risque relatif peut changer au cours du temps. Par exemple, le risque relatif entre l'individu i de la strate a et l'individu j et de la strate b au temps t est

$$\frac{\lambda_a(t | Z_i)}{\lambda_b(t | Z_j)} = \frac{\lambda_{0,a}(t)}{\lambda_{0,b}(t)} \exp(\beta^T (Z_i - Z_j)).$$

Ce risque relatif n'est pas constant au cours du temps si le rapport des risques de base $\lambda_{0,a}$ sur $\lambda_{0,b}$ ne l'est pas. Notons que le coefficient de régression β associé au vecteur Z est le même pour toutes les strates. Il peut être judicieux d'effectuer cette stratification selon les modalités d'une variable catégorielle qui ne respecte pas l'hypothèse de proportionnalité des risques (Therneau et Grambsch, 2000).

Modèles additifs et extensions

Au lieu de considérer un modèle à risques multiplicatifs, on peut utiliser le modèle à intensité additive proposé par Aalen (1980, 1989) et défini par

$$\lambda(t | Z) = \beta_0(t) + \beta^T(t)Z(t), \quad (1.15)$$

où le vecteur de coefficients de régression $\beta(t) = (\beta_0(t), \beta_1(t), \dots, \beta_p(t))$ dépend du temps. L'estimation des paramètres cumulés $\int_0^t \beta_i(s) ds$ pour $i = 1, \dots, p$ s'effectue par la méthode des moindres carrés pondérés. McKeague et Sasieni (1994) ont considéré le cas d'effets fixes pour certaines covariables et dépendants du temps pour les autres. Lin et Ying (1995) ont proposé un modèle à risques additifs et multiplicatifs de la forme

$$\lambda(t | X, Z) = Y(t)g\{\beta^T X(t)\} + \lambda_0(t)h\{\gamma^T Z(t)\},$$

où $(X(t), Z(t))$ est un vecteur de $p + q$ covariables dépendantes du temps, $\beta \in \mathbb{R}^p$ et $\gamma \in \mathbb{R}^q$. Nous rappelons que $Y(t)$ est l'indicateur d'être à risque au temps t . Les fonctions h et g sont des fonctions de lien. Par exemple, on peut choisir $h(x) = \exp(x)$ ou $1 + x$ et $g(x) = x$ ou $\exp(x)$. Cette classe de modèles englobe à la fois le modèle à risques proportionnels (1.4) et le modèle additif (1.15) pour des valeurs de h et g bien choisies. Scheike et Zhang (2002)

considèrent une autre combinaison des intensités multiplicatives et additives avec le modèle défini par

$$\lambda(t | X, Z) = Y(t)\beta(t)^T X(t) \exp(\gamma^T Z(t)),$$

où $\beta(t)$ dépend du temps. Les propriétés du modèle additif sont discutées dans l'ouvrage de [Aalen et al. \(2008\)](#). [Beran \(1981\)](#) ainsi que [McKeague et Utikal \(1990\)](#) ont étudié le modèle à intensité

$$\lambda(t | Z) = Y(t)\alpha(t, Z(t)), \quad (1.16)$$

sans spécifier la fonction α , qui est estimée par la méthode des noyaux.

Modèles transformés linéairement

Un calcul simple permet de montrer que le modèle à risques proportionnels (1.4) avec des covariables Z fixes au cours du temps peut s'écrire

$$\log\{-\log(S(t | Z))\} = \log\left(\int_0^t \lambda_0(s) ds\right) + \beta^T Z,$$

où nous rappelons que λ_0 est une fonction positive quelconque. On peut alors généraliser le modèle en considérant l'équation

$$h(S(t | Z)) = g(t) + \beta^T Z, \quad (1.17)$$

où h est une fonction de lien, g est une fonction inconnue et les deux fonctions sont strictement croissantes. Le modèle (1.17) peut également s'écrire

$$g(T) = -\beta^T Z + \varepsilon, \quad (1.18)$$

où ε est une variable aléatoire de fonction de répartition $1 - h^{-1}$. L'équation (1.18) décrit la classe de modèles de transformation de T linéairement en Z ([Cheng et al., 1995](#), [Dabrowska et Doksum, 1988](#), [Lin, 2013](#), [Wei, 1992](#)). Le modèle à odds proportionnels ([Bennett, 1983a](#)) est un cas particulier obtenu en fixant

$$h(x) = \text{logit}(x) = \log\left(\frac{x}{1-x}\right).$$

En considérant

$$h(x) = \log\left(\frac{1 - S(x | Z)}{S(x | Z)}\right) \quad \text{et} \quad g(t) = \phi \log(t),$$

on retrouve le modèle log-logistique ([Bennett, 1983b](#)). Une classe importante de modèles dits "modèles de vie accélérée" apparaît en considérant la fonction $g = \log$. Dans ce cas, la fonction de survie conditionnelle à Z s'écrit

$$S(t | Z) = S_0\left(t \exp(\beta^T Z)\right), \quad 0 \leq t \leq \mathcal{T}, \quad (1.19)$$

où S_0 est la fonction de survie associée à λ_0 . La formule (1.19) montre que les covariables ont un effet multiplicatif sur le temps, ce qui implique qu'elles accélèrent ou ralentissent le décès, selon le signe de $\beta^T Z$ (Martinussen et Peng, 2013). La classe de modèles transformés linéairement est plus générale que le modèle à risques proportionnels (1.4) mais elle est moins flexible, puisqu'elle ne prend pas en compte les covariables qui dépendent du temps et son extension à des événements multiples est délicate.

Modèle de Yang et Prentice

En présence d'une covariable binaire Z qui prend les modalités P (placebo) et T (traitement), Yang et Prentice (2005) ont proposé de modéliser le temps de survie dans les deux groupes par

$$\lambda_T(t) = \frac{\theta_1 \theta_0}{\theta_1 + (\theta_0 - \theta_1) S_P(t)} \lambda_P(t), \quad 0 \leq t \leq \mathcal{T}, \quad (1.20)$$

où θ_1 et θ_0 sont deux paramètres positifs, S_P est la fonction de survie dans le groupe placebo et λ_T (respectivement λ_P) est le risque instantané dans le groupe traitement (resp. groupe placebo). Le modèle à risques proportionnels ($\theta_1 = \theta_0$) et le modèle à odds proportionnels ($\theta_0 = 1$) en sont des cas particuliers. Ce modèle est très flexible et permet de bien modéliser des risques dont l'ordre s'inverse, c'est-à-dire qui se croisent. Cette dernière propriété sera particulièrement utile pour développer des tests, nous y reviendrons au Chapitre 4. Les paramètres du modèle sont tels que $\theta_1 = \lim_{t \downarrow 0} \lambda_T(t)/\lambda_P(t)$ et $\theta_0 = \lim_{t \uparrow \mathcal{T}} \lambda_T(t)/\lambda_P(t)$. Ainsi θ_1 représente le risque relatif à court terme et θ_0 celui à long terme. L'estimation dans le modèle est faite en maximisant un équivalent de la vraisemblance partielle spécifique au modèle. Les auteurs suggèrent l'introduction de covariables Z en remplaçant la fonction de survie $S_T(t)$ (respectivement $S_P(t)$) par $S_T(t)^{\exp(\beta^T Z)}$ (respectivement $S_P(t)^{\exp(\beta^T Z)}$).

Autres extensions du modèle à risques proportionnels

D'autres extensions du modèle à risques proportionnels (1.4) sont envisageables selon le type de données étudiées. Par exemple, les modèles de fragilité permettent de modéliser la dépendance entre plusieurs temps de décès par l'introduction d'un effet aléatoire. Il peut s'agir de relier différents temps pour un même individu (par exemple, si une covariable a été mesurée à différentes reprises) ou des individus différents (par exemple, des jumeaux). En revanche, pour modéliser les temps jusqu'à plusieurs types d'événements (par exemple, la rechute, la guérison et le décès d'un individu malade), il est plus judicieux d'utiliser un modèle multi-états ou un modèle à risques compétitifs. Des modèles pour événements récurrents sont aussi envisageables. Nous renvoyons à l'ouvrage très complet de Klein *et al.* (2013) pour des explications détaillées sur ces différents modèles.

Au temps t	Groupe $Z = 0$	Groupe $Z = 1$	Total
Décédés	$d_{0,t}$	$d_{1,t}$	D_t
A risque	$ar_{0,t}$	$ar_{1,t}$	AR_t
Total	$N_{0,t}$	$N_{1,t}$	N_t

TABLE 1.3 – Illustration du calcul de la statistique du log-rank

1.3 Test de la présence d'un effet des covariables sur la survie

Lors du traitement de données de survie, la question d'adopter une bonne modélisation n'est pas la seule qui se pose. Une autre question importante est de savoir si les covariables de la base de données ont un effet sur le temps de survie, observé ou non. Pour des données collectées lors d'un essai clinique dans lequel un groupe de patients a reçu le nouveau traitement à tester et un autre groupe a reçu un placebo, cela revient à se demander si le nouveau traitement a une influence sur la survie. Si les données sont collectées lors d'enquêtes épidémiologiques prospectives ou rétrospectives, on cherchera à savoir si les facteurs de risque recueillis (par exemple, l'âge, le fait de fumer ou non, le poids, etc) ont un lien avec la survie des patients. Dans ces deux exemples d'étude, les covariables sont catégorielles (groupe de traitement, le fait de fumer ou non) ou continues (le poids, l'âge).

Un test très populaire, le test du log-rank (Mantel, 1966), a été développé pour traiter le cas de covariables catégorielles. Ce test peut s'écrire comme un test du score dans le modèle à risques proportionnels (1.4), ce qui permet de l'utiliser en présence de covariables continues (Cox, 1972). Expliquons le principe du test avec une variable binaire Z prenant les modalités 0 ou 1 représentant deux groupes de patients. On veut tester l'hypothèse nulle H_0 contre l'alternative H_1 suivantes

$$\begin{cases} H_0 : \forall t, S_1(t) = S_0(t) \\ H_1 : \exists t_0, S_1(t_0) \neq S_0(t_0), \end{cases} \quad (1.21)$$

où S_0 (respectivement S_1) est la fonction de survie dans le groupe $Z = 0$ (respectivement $Z = 1$).

A chaque temps de décès observé t , on suppose que les nombres de décès, d'individus encore à risque de décéder et d'individus dans les groupes $Z = 0$ et $Z = 1$ sont fixés. On comptabilise le nombre de patients de chaque groupe qui sont décédés ou encore à risque de décéder comme représenté en Table 1.3. On considère donc que les marges du tableau sont fixes. On note $d_{1,t}$ la réalisation de la variable aléatoire $D_{1,t}$ égale au nombre d'individus décédés au temps t dans le groupe $Z = 1$. Cette variable aléatoire a une loi hypergéométrique de paramètres N_t , D_t/N_t et $N_{1,t}$. On peut ainsi calculer son espérance et sa variance :

$$\begin{aligned} E(D_{1,t}) &= \frac{D_t N_{1,t}}{N_t}, \\ V(D_{1,t}) &= \frac{N_{1,t} N_{0,t} D_t A R_t}{N_t^2 (N_t - 1)}. \end{aligned}$$

En supposant les tableaux indépendants pour les différents temps de décès, la statistique

$$\frac{(\sum_{i=1}^n \delta_i \{d_{1,X_i} - E(D_{1,X_i})\})^2}{\sum_{i=1}^n \delta_i V(D_{1,X_i})}$$

converge vers une loi du χ^2 à 1 degré de liberté sous l'hypothèse nulle, ce qui permet de construire un test asymptotique de niveau α . Nous présentons le test pour une population séparée en 2 groupes, mais une extension à K groupes est immédiate. La loi limite de la statistique est alors une loi du χ^2 à $K - 1$ degrés de liberté. Il est également possible de définir un test du log-rank stratifié, en séparant les individus en strates selon les modalités d'une variable catégorielle. Ceci permet de tester l'effet du médicament en tenant compte de la covariable selon laquelle la stratification a été effectuée. Aussi, lorsque $K > 2$, on peut effectuer des tests de tendance pour tester si les risques des K populations sont ordonnés. Nous renvoyons à l'ouvrage d'[Andersen et al. \(1993\)](#) pour plus de détails.

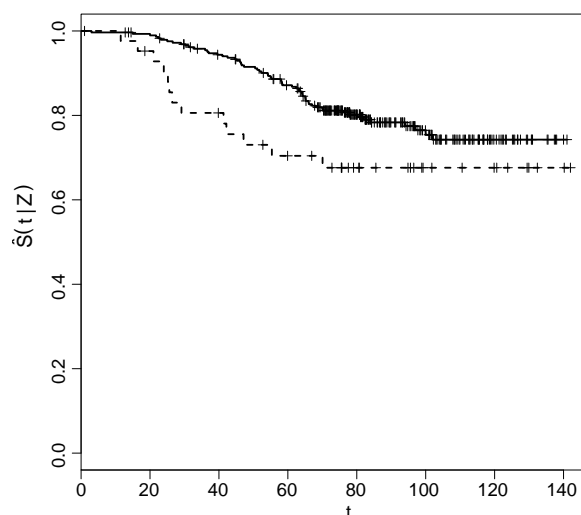


FIGURE 1.3 – Estimateurs de Kaplan–Meier de la survie pour les patientes ayant une tumeur de taille inférieure (ligne pleine) ou supérieure (ligne pointillée) à 60 mm.

Ce test, très utilisé en pratique, est le plus puissant sous le modèle à risques proportionnels ([Peto et Peto, 1972](#)). En revanche, si le rapport des risques n'est pas constant dans le temps, ce test peut subir de très fortes pertes de puissance ([Leurgans, 1983, 1984](#)). La Figure 1.3 représente les estimateurs de Kaplan–Meier de la survie de 2 groupes de patientes atteintes de cancer du sein et suivies à l'Institut Curie pendant plus de 12 ans. Dans un groupe, le diamètre de la tumeur des patientes est inférieur à 60 mm et, dans l'autre groupe, le diamètre est supérieur. Au total, 332 patientes ont été considérées. Graphiquement, on remarque une différence dans les survies des deux groupes puisque les estimateurs de Kaplan–Meier sont nettement distincts. Pendant les 60 premiers mois, les survies estimées sont très différentes, correspondant à un effet fort entre les groupes de patientes. Après 60 mois, les courbes se rapprochent, reflétant une diminution de l'effet. De plus, on remarque que 80% des patientes

dont la taille de la tumeur est inférieure à 60 mm survit au moins 75 mois. Cette durée est divisée par 3 et diminue à 28 mois pour les patientes dont la taille de la tumeur est supérieure à 60 mm. Pourtant, la p -value du test du log-rank vaut 0.06. Le test, non significatif au risque 5%, ne permet pas de conclure à une différence entre les survies des deux groupes. Ceci est dû au fait que les risques ne sont pas proportionnels, la différence entre les deux groupes s'estompe au cours du temps. Le test du log-rank n'est pas assez puissant pour détecter une différence.

Plusieurs tests ont été proposés pour remédier à cette perte de puissance, comme les tests du log-rank pondérés (Fleming et Harrington, 1991, Gehan, 1965, Peto et Peto, 1972). Dans le Chapitre 4, nous présenterons le test du log-rank comme un test du score du modèle à risques non proportionnels. Nous exposerons aussi une revue de littérature sur des extensions et des alternatives au test du log-rank pour tester les hypothèses H_0 et H_1 définies équation (1.21). Nous proposerons également un nouveau test de comparaison de courbes de survie, et nous montrerons, sur le même exemple, que ce nouveau test détecte une différence dans la survie des deux groupes de patientes au risque 5%.

1.4 Evaluation de l'adéquation et de la capacité prédictive du modèle à risques proportionnels

1.4.1 Adéquation du modèle

Un modèle de régression a pour but de détecter des associations entre une variable réponse – dans notre cas le temps de survie – et une ou plusieurs covariables. Dans le cadre d'une analyse de données réelles, nous nous appuyons sur un modèle de régression afin d'étudier et d'interpréter les effets des covariables prises individuellement, conjointement ou d'une manière hiérarchique sur la variable à expliquer. Même si l'estimation des coefficients du modèle est possible, le modèle sélectionné ne s'ajuste pas nécessairement bien aux données. Il faut toujours vérifier son adéquation, c'est-à-dire vérifier que les hypothèses du modèle s'appliquent aux données. Ceci peut se faire par des méthodes graphiques ou des tests d'hypothèse.

Dans le cas du modèle linéaire, l'analyse des résidus permet de vérifier l'adéquation du modèle. En analyse de survie, deux variables aléatoires sont étudiées, Z et T . Ainsi, deux lois conditionnelles sont disponibles pour l'analyse des résidus : la loi de Z sachant T et celle de T sachant Z . En se basant sur la première loi, Schoenfeld (1982) a introduit des résidus qui ont été très utilisés par la suite pour vérifier l'adéquation du modèle à risques proportionnels (1.4). Il a considéré un vecteur de covariables Z fixes dans le temps mais leur extension à des covariables dépendantes du temps est immédiate et nous la présentons ici. Les résidus de Schoenfeld sont définis à chaque temps de décès t par

$$r_{\beta}(t) = Z_j(t) - \mathcal{E}_{\beta}(Z | t), \quad (1.22)$$

où j est l'individu qui décède en t . Ces résidus ont été utilisés par de nombreux auteurs dans

des méthodes graphiques ou des tests (Grambsch et Therneau, 1994, Lin, 1991, Lin et al., 1993, O'Quigley et Pessione, 1991, Wei, 1984). Un autre type de résidus, appelés résidus de martingales et basés sur la loi de T sachant Z , a été utilisé pour vérifier l'adéquation du modèle. Pour l'individu j et au temps t , ces résidus sont définis par

$$M_j(t) = N_j(t) - \hat{\Lambda}_j(t), \quad \text{avec} \quad \hat{\Lambda}_j(t) = \int_0^t \pi_j(\hat{\beta}, s) d\bar{N}(s).$$

Cette définition provient de la décomposition d'un processus de comptage en une martingale et un compensateur (Andersen et Gill, 1982). Dans le modèle à risques proportionnels de paramètre β , le compensateur du processus N_j au temps t est $\int_0^t \lambda(s | Z_j) ds$ et est estimé par $\hat{\Lambda}_j(t)$. Ainsi, si le modèle ajuste bien les données, $M_1(t), \dots, M_n(t)$ sont des réalisations de martingales indépendantes (Gill, 1980), ce qui implique qu'elles doivent être centrées et non corrélées. Ces propriétés doivent se retrouver sur le graphique des résidus des $M_j(\mathcal{T})$, $j = 1, \dots, n$. Barlow et Prentice (1988) ont défini une classe de résidus de martingales en considérant $\int \phi(t) M_j(t) dt$ où ϕ est une fonction prévisible. Ces résidus ont notamment été utilisés par Kay (1977), Lin (1991), Lin et al. (1993) et Therneau et al. (1990). Notons que le cas particulier $\phi = Z_j$ revient à considérer les résidus de Schoenfeld (1982). Une bibliographie plus complète sur l'adéquation du modèle à risques proportionnels (1.4) sera présentée au Chapitre 5.

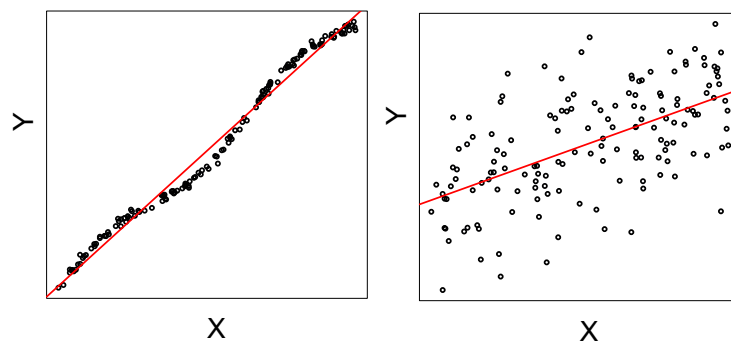


FIGURE 1.4 – Illustration de la différence entre qualités d'ajustement et de prédiction du modèle linéaire.

Après avoir vérifié la validité des hypothèses du modèle, une question centrale est de quantifier son pouvoir prédictif afin d'évaluer la qualité du modèle et éventuellement de comparer plusieurs modèles. L'adéquation du modèle et la capacité de prédiction sont deux aspects distincts, comme l'illustre la Figure 1.4 dans le contexte du modèle linéaire. Sur les deux figures, les points représentent les couples d'observations (X, Y) et les lignes sont les droites de régression $Y = \hat{a} + \hat{b}X$. Sur le graphique de gauche, la courbure des points indique que le lien entre les variables X et Y n'est probablement pas linéaire : l'adéquation du modèle n'est pas bonne. En revanche, les observations sont proches de la droite de régression ce qui signifie que la prédiction associée à une nouvelle observation X ne sera pas trop éloignée de la réalité. Pour le graphique de droite, les conclusions sont inversées : il ne paraît pas aberrant de relier

les variables par une fonction affine. Cependant, les valeurs prédites peuvent être éloignées des valeurs observées à cause d'un bruit fort. La capacité prédictive du modèle est donc moins bonne que dans la situation de gauche. Cette distinction entre prédiction et adéquation se retrouve dans tous les modèles statistiques.

1.4.2 Mesures de capacité prédictive du modèle

Les mesures de capacité prédictive permettent de répondre à plusieurs questions sur le modèle. Des questions typiques peuvent être : comment un ensemble de covariables influence la prédiction, comparé à un autre ensemble ? Combien d'information prédictive est contenue dans un biomarqueur lorsqu'on l'ajoute à un ensemble de facteurs pronostiques déjà connus ? Quand tous les facteurs pronostiques connus sont inclus dans un modèle, quelle quantité de variabilité de la variable réponse est prise en compte et, par conséquent, quelle part de variabilité reste à expliquer ? Comment l'assouplissement d'une hypothèse du modèle impacte la prédiction ? Cette dernière observation attire l'attention sur le fait que, bien qu'il s'agisse de différentes techniques avec différents buts, les objectifs de la vérification de l'adéquation du modèle et de la quantification de la capacité prédictive sont liés. Plusieurs catégories de mesures du pouvoir prédictif du modèle à risques proportionnels peuvent être utilisées, notamment des mesures basées sur la distance entre les observations et les valeurs prédites par le modèle, sur la vraisemblance du modèle ou des mesures de discrimination (Schmid *et al.*, 2010, Steyerberg *et al.*, 2010).

Schemper (1990) et Korn et Simon (1990) ont proposé des mesures de la distance L^2 entre la survie prédite par le modèle avec les covariables et celle prédite sans les covariables. La mesure de Schemper (1990) a été améliorée pour corriger sa dépendance en la censure (O'Quigley *et al.*, 1999, Schemper et Henderson, 2000) ou pour être plus robuste aux erreurs de spécification de modèle (Schmid *et al.*, 2010). Le Brier score, initialement utilisé pour évaluer les prévisions météorologiques, est une mesure de l'erreur quadratique entre l'indicatrice de survie au moins jusqu'au temps t , $\mathbf{1}_{T \geq t}$, et sa prédiction sous le modèle par $\pi(t, \hat{\beta})$ définie équation (1.5) (Brier, 1950, Graf *et al.*, 1999). Gerds et Schumacher (2006) ont proposé un estimateur du Brier score consistant en présence de données censurées. Schoop *et al.* (2008) ont considéré une estimation plus flexible actualisée avec le temps lorsque les covariables dépendent du temps. Cette mesure dépend de l'horizon temporel t auquel on veut évaluer la qualité de la prédiction, ce qui n'est pas toujours évident à établir dans les applications (Heagerty et Zheng, 2005). L'utilisation d'une mesure plus globale est souvent préférable.

La vraisemblance du modèle à risques proportionnels peut permettre de construire une mesure prédictive. La vraisemblance est alors évaluée avec l'estimation du paramètre qui la maximise et elle est associée à la vraisemblance sous le modèle dont le paramètre est nul. Kent et O'Quigley (1988) ont adapté le coefficient d'aléa expliqué ("explained randomness") au modèle à risques proportionnels (Kent, 1983). Ce coefficient, ainsi que celui de Xu et O'Quigley (1999), repose sur la divergence de Kullback–Leibler, encore appelée entropie relative, entre la vraisemblance du modèle avec des covariables et sans covariable. La vraisemblance peut

également être utilisée dans une mesure de rapport de vraisemblance (Allison, 1995, Maddala, 1983, Nagelkerke, 1991, Royston, 2006). Plus récemment, Heller (2012) a proposé une mesure de distance entre l'entropie du modèle à risques proportionnels et celle du modèle de paramètre nul. Pour plus de détails sur ces mesures d'aléa expliqué, on pourra consulter le chapitre 14 de l'ouvrage d'O'Quigley (2008).

La discrimination d'un modèle de survie mesure la séparation entre les individus qui ont un risque fort de présenter l'évènement d'intérêt têt de ceux dont le risque est plus faible. Plusieurs mesures de discrimination ont été introduites, comme l'indice de concordance défini par

$$C = P(T_1 \geq T_2 \mid P_1 \leq P_2),$$

où les couples (T_1, P_1) et (T_2, P_2) sont indépendants et P_i est le prédicteur de T_i ($i = 1, 2$), qui peut être une covariable (Z_i), l'indice pronostique ($\beta^T Z_i$) ou encore un score de réponse. Si l'effet de P sur la durée T est négatif (une forte valeur de P diminue la durée T) et si l'indice de concordance est proche de 1, cela signifie que l'ordre entre les temps de survie et les prédicteurs est respecté : un individu présentant une plus grande valeur de P qu'un autre individu aura une valeur de T plus faible avec grande probabilité. Plusieurs estimateurs de cet indice pronostique ont été proposés, notamment le très populaire C-index d'Harrell *et al.* (1982, 1984). Cet estimateur est basé sur le classement des paires (T_i, Z_i) selon l'ordre d'arrivée des T_i observés, pour tous les individus i non censurés. Les temps censurés ne sont pas utilisés, le rendant dépendant de la censure. Gönen et Heller (2005) ont proposé un estimateur de l'indice de concordance spécifique au modèle à risques proportionnels, dont un travail de simulation établit les bonnes propriétés.

D'autres méthodes telles que les courbes ROC (Receiver Operating Characteristic), notamment résumées par l'aire sous la courbe, permettent de mesurer la discrimination d'un modèle de survie lorsque la covariable est continue (Hanley et McNeil, 1982). Heagerty *et al.* (2000) ont adapté les courbes ROC aux données de survie. Les auteurs ont proposé deux estimateurs non paramétriques des courbes ROC, un reposant sur l'estimateur de Kaplan-Meier de la survie et l'autre étant un estimateur du plus proche voisin. Song et Zhou (2008) se sont intéressés à l'estimation des courbes ROC en présence de covariables, avec un modèle à risques proportionnels. Les courbes ROC sont évaluées à un horizon temporel t fixé à l'avance, par exemple en distinguant les individus qui ont subi l'évènement avant t des autres. Comme pour le Brier score, dans les applications, l'horizon temporel t pour lequel la qualité de la prédiction est évaluée n'est pas toujours identifié, une mesure plus globale est désirable. Heagerty et Zheng (2005) ont montré que l'indice de concordance C peut s'écrire comme une aire sous la courbe ROC pondérée. L'aire sous la courbe ROC, même si elle est largement utilisée, présente l'inconvénient de n'être pas très sensible à l'ajout d'une covariable importante dans le modèle (Pepe *et al.*, 2004, Ware, 2006). Malgré cette propriété qui rend la comparaison de plusieurs modèles à l'aide de l'aire sous la courbe ROC délicate, cette mesure reste très utilisée en pratique (Pepe et Janes, 2013). Une bibliographie des courbes ROC adaptée aux données de survie peut être consultée dans Pepe *et al.* (2008) et Blanche *et al.* (2013). Stare *et al.* (2011) ont proposé

une mesure de la discrimination des modèles de survie basée sur les différences entre les ordres d'arrivée des événements observés et ceux prédits par le modèle. L'Integrated Discrimination Improvement et le Net Reclassification Index (Pencina *et al.*, 2008) ont été également utilisés, mais de mauvaises propriétés de ces indices ont été démontrées récemment (Hilden et Gerds, 2013). Nous renvoyons à la thèse de Blanche (2013) et aux travaux de Gerds *et al.* (2008), Gu et Pepe (2009) et Steyerberg *et al.* (2010) pour plus de détails sur les mesures de capacité prédictive des modèles de survie.

Beaucoup d'auteurs définissent ces mesures de qualité de prédiction comme des mesures de type R^2 , étant donné qu'elles sont comprises entre 0 et 1. Pourtant, le coefficient R^2 est le rapport de la variance d'une prédiction sur la variance de la valeur observée. Le coefficient R^2 est un estimateur de la variance expliquée, que nous définirons au Chapitre 5. Certaines mesures citées ci-dessus n'ont pas de contrepartie théorique, il s'agit de mesures ad hoc construites à partir des données. De plus, leur comportement sous un modèle mal spécifié n'a généralement pas été étudié. Ceci peut être problématique puisque l'hypothèse de l'effet constant dans le temps du modèle à risques proportionnels est très restrictive. Les travaux de Graf *et al.* (1999), Hielscher *et al.* (2010), Müller *et al.* (2008), Rouam (2011), Schemper et Stare (1996) et Choodari-Oskoei *et al.* (2012) recensent et comparent un grand nombre de ces mesures dites de variance expliquée. A l'heure actuelle, aucun consensus n'a été trouvé quant à la mesure à privilégier dans les applications. Certaines mesures dépendent de la censure ou sont biaisées, d'autres ne reflètent pas vraiment la qualité de la prédiction, etc.

Notre étude va se concentrer sur un estimateur de la proportion de variance de Z expliquée par T , le coefficient R^2 introduit par O'Quigley et Flandre (1994). Nous présentons la variance expliquée, qui est un paramètre dans la population, et son estimateur, le coefficient R^2 étendu au modèle à risques non proportionnels au Chapitre 5. Dans ce chapitre, nous montrerons comment lier adéquation et prédiction dans ce modèle et nous proposerons une méthode pour sélectionner un modèle avec une bonne adéquation et une qualité de prédiction maximale (maximale dans un sens que nous définirons). De plus, nous présenterons les propriétés asymptotiques du coefficient R^2 en cas de mauvaise spécification du modèle à risques proportionnels. En d'autres termes, nous étudierons le comportement limite du coefficient R^2 lorsqu'un modèle à risques proportionnels a été ajusté, alors que le vrai coefficient $\beta(t)$ dépend du temps.

Chapitre 2

Outils mathématiques

Dans ce chapitre, nous mentionnons des résultats mathématiques utilisés dans les développements théoriques de la thèse. Nous commençons par énoncer quelques résultats sur les espaces des fonctions continues et discontinues de $[0, 1]$ dans \mathbb{R} ou \mathbb{R}^p . Ensuite, nous présentons une loi faible des grands nombres pour des variables non corrélées. Nous énonçons également des théorèmes de la limite centrale fonctionnels pour des variables centrées, indépendantes mais non identiquement distribuées. Enfin, nous définissons les normes vectorielles et matricielles utilisées dans cette thèse et nous présentons une inégalité de Taylor–Lagrange pour une fonction de \mathbb{R}^p dans \mathbb{R}^p , avec $p \in \mathbb{N}^*$.

2.1 Espaces $D[0, 1]$ et $C[0, 1]$

En reprenant les notations de [Whitt \(2002\)](#), notons $(C[0, 1], E)$ l'espace des fonctions continues définies sur l'intervalle $[0, 1]$ à valeurs dans l'espace E et $(D[0, 1], E)$ l'espace des fonctions continues à droite et limites à gauche (càdlàg) définies sur $[0, 1]$ à valeurs dans E . Dans les chapitres suivants de ce manuscrit, les processus étudiés seront des éléments aléatoires de ces deux espaces avec $E = \mathbb{R}^p$, $p \in \mathbb{N}^*$. Dans notre contexte, l'intervalle $[0, 1]$ représente le temps, c'est pourquoi tout point t de cet intervalle sera mentionné comme le temps t ou l'instant t .

2.1.1 Topologies des espaces $(C[0, 1], \mathbb{R})$ et $(D[0, 1], \mathbb{R})$

Dans les deux espaces de fonctions $(C[0, 1], \mathbb{R})$ et $(D[0, 1], \mathbb{R})$, des distances doivent être définies. Il est d'usage de munir $(C[0, 1], \mathbb{R})$ de la topologie de la convergence uniforme induite par la distance uniforme.

Définition 2.1.1. *La distance uniforme d entre deux éléments de $(C[0, 1], \mathbb{R})$ est définie par*

$$d(f, g) = \sup_{0 \leq t \leq 1} |f(t) - g(t)|, \quad f, g \in (C[0, 1], \mathbb{R}). \quad (2.1)$$

Cependant, deux fonctions f et g de $(D[0, 1], \mathbb{R})$ peuvent avoir des discontinuités à différents instants de $[0, 1]$. La topologie de la convergence uniforme ne prend pas en compte

les différences dans les instants de sauts, elle n'est pas assez fine pour l'espace des fonctions càdlàg. La topologie de Skorokhod induite par la distance δ définie ci-après répond à ce problème.

Définition 2.1.2. *La distance de Skorokhod δ est définie par*

$$\delta(f, g) = \inf_{\lambda \in \Lambda} \{d(f, g(\lambda)) \vee d(\lambda, I)\}, \quad f, g \in (D[0, 1], \mathbb{R}), \quad (2.2)$$

où I est la transformation identité sur $[0, 1]$, Λ est la classe d'homéomorphismes de $[0, 1]$ dans lui-même telle que $\lambda(0) = 0$ et $\lambda(1) = 1$ pour tout $\lambda \in \Lambda$ et $a \vee b = \max(a, b)$.

Nous renvoyons à la seconde édition de l'ouvrage de Billingsley (1999) pour plus de détails sur ces topologies. Notons que

$$\delta(f, g) \leq d(f, g), \quad f, g \in (D[0, 1], \mathbb{R}). \quad (2.3)$$

Cette relation implique le résultat suivant :

Proposition 2.1.1 (Billingsley, 1999). *Soit une suite de fonctions $(f_n)_{n \in \mathbb{N}}$ de $(D[0, 1], \mathbb{R})$ et $f \in (D[0, 1], \mathbb{R})$. Si*

$$\lim_{n \rightarrow +\infty} d(f_n, f) = 0,$$

alors

$$\lim_{n \rightarrow +\infty} \delta(f, g_n) = 0.$$

Autrement dit, dans l'espace $(D[0, 1], \mathbb{R})$, la convergence au sens de la topologie de la convergence uniforme entraîne la convergence au sens de la topologie de Skorokhod. Cependant, si la limite est une fonction continue, les deux convergences sont équivalentes.

Proposition 2.1.2 (Billingsley, 1999). *Pour tout $f \in (C[0, 1], \mathbb{R})$ et $(g_n)_{n \in \mathbb{N}} \in (D[0, 1], \mathbb{R})$,*

$$\lim_{n \rightarrow +\infty} d(f, g_n) = 0 \iff \lim_{n \rightarrow +\infty} \delta(f, g_n) = 0.$$

D'après cette proposition, il est équivalent de montrer la convergence d'une suite de fonctions càdlàg vers une fonction continue avec la topologie de la convergence uniforme ou la topologie de Skorokhod.

2.1.2 Topologies de $(D[0, 1], \mathbb{R}^p)$

Un autre espace dans lequel nous travaillons est l'espace des fonctions càdlàg de $[0, 1]$ dans \mathbb{R}^p où p est un entier strictement supérieur à 1. Dans ce cas, plusieurs distances peuvent être considérées. Nous choisissons de travailler avec la distance produit de Skorokhod.

Définition 2.1.3. *Soit $f = (f_1, \dots, f_p)$ et $g = (g_1, \dots, g_p)$ deux fonctions de $(D[0, 1], \mathbb{R}^p)$, avec f_i et g_i des fonctions de $(D[0, 1], \mathbb{R})$, pour $i = 1, \dots, p$. La distance produit de Skorokhod entre*

f et g est définie par

$$\delta_p(f, g) = \sum_{i=1}^p \delta(f_i, g_i).$$

Une autre distance convenable est de considérer la distance de Skorokhod (équation (2.2)) dans laquelle la valeur absolue dans la Définition 2.1 de la distance uniforme est remplacée par la norme sup telle que $\|a\| = \max_{i=1, \dots, p} |a_i|$, pour $a \in \mathbb{R}^p$. La topologie induite par cette distance est appelée topologie forte de Skorokhod. Ce terme provient du fait que la convergence au sens de cette topologie implique la convergence au sens de la topologie produit de Skorokhod (Whitt, 2002). Cependant, les théorèmes limites sur lesquels sont basés nos résultats utilisent la distance produit de Skorokhod. Nous utiliserons également le résultat de la proposition suivante.

Proposition 2.1.3. *Pour toute fonction $f \in (C[0,1], \mathbb{R}^p)$ et pour toute suite de fonctions $(g_n)_{n \in \mathbb{N}^*} \in (D[0,1], \mathbb{R}^p)$, si $(g_n)_{n \in \mathbb{N}^*}$ converge vers f au sens de la topologie de la convergence uniforme dans \mathbb{R}^p , c'est-à-dire si*

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq 1} \|f(t) - g_n(t)\| = 0,$$

alors

$$\lim_{n \rightarrow \infty} \delta_p(f, g_n) = 0.$$

Preuve de la Proposition 2.1.3. Soit $n \in \mathbb{N}^*$ et les fonctions $f = (f^1, \dots, f^p) \in (C[0,1], \mathbb{R}^p)$ et $g_n = (g_n^1, \dots, g_n^p) \in (D[0,1], \mathbb{R}^p)$. On a la relation

$$\begin{aligned} \sup_{0 \leq t \leq 1} \|f(t) - g_n(t)\| &= \max_{i=1, \dots, p} d(f^i, g_n^i) \\ &\geq \frac{1}{p} \sum_{i=1}^p d(f^i, g_n^i) \\ &\geq \frac{1}{p} \sum_{i=1}^p \delta(f^i, g_n^i) = \frac{1}{p} \delta_p(f, g_n), \end{aligned}$$

où la dernière inégalité est obtenue en appliquant l'équation (2.3). La conclusion est immédiate. \square

Ainsi, si une suite de fonctions de $(D[0,1], \mathbb{R}^p)$ converge vers une fonction de $(C[0,1], \mathbb{R}^p)$ au sens de la topologie de la convergence uniforme, alors la convergence a également lieu au sens de la topologie produit de Skorokhod. Les résultats présentés dans cette section seront utiles au Chapitre 3 pour étudier les comportements asymptotiques de processus empiriques à l'aide des théorèmes de la limite centrale énoncés dans la section suivante.

2.2 Théorèmes limite

2.2.1 Loi faible des grands nombres pour des variables non corrélées

Le théorème suivant donne des conditions pour établir la convergence en probabilité de la moyenne empirique de variables centrées et non corrélées vers 0.

Théorème 2.2.1. *Soit X_1, X_2, \dots, X_n des variables aléatoires centrées et non corrélées. On suppose qu'il existe une constante $C \in \mathbb{R}^+$ telle que $V(X_i) \leq C$, pour $i = 1, \dots, n$. Alors,*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} 0.$$

Preuve du Théorème 2.2.1. Soit $\varepsilon > 0$. Les variables X_1, X_2, \dots, X_n ne sont pas corrélées donc $V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i)$. L'inégalité de Bienaymé-Chebyshev implique

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2} V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2 \varepsilon^2} \sum_{i=1}^n V(X_i) \\ &\leq \frac{C}{n \varepsilon^2}. \end{aligned}$$

Par conséquent, $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i\right| > \varepsilon\right) = 0$ et le résultat est démontré. \square

2.2.2 Théorèmes de la limite centrale fonctionnels

Avant d'établir les théorèmes de la limite centrale fonctionnels, donnons la définition du mouvement Brownien sur $[0, 1]$. Soit (Ω, \mathcal{A}, P) un espace probabilisé. Un processus stochastique X sur l'intervalle $I \subset \mathbb{R}^+$ est une famille $\{X(t) = X(t, \cdot), t \in I\}$ de variables aléatoires telles que $X(t, \cdot)$ soit une variable aléatoire définie sur l'espace Ω pour tout temps $t \in I$. Le processus est continu si la fonction $t \rightarrow X(t, \omega)$ est continue pour tout $\omega \in \Omega$ fixé.

Définition 2.2.1. *Le mouvement Brownien standard, ou processus de Wiener, sur $[0, 1]$ noté \mathcal{W} est un processus continu tel que*

- $\mathcal{W}(t) \sim \mathcal{N}(0, t)$, $0 \leq t \leq 1$.
- Pour tout $0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq 1$, les variables

$$\mathcal{W}(t_1), \mathcal{W}(t_2) - \mathcal{W}(t_1), \mathcal{W}(t_3) - \mathcal{W}(t_2), \dots, \mathcal{W}(t_n) - \mathcal{W}(t_{n-1})$$

sont indépendantes.

Autrement dit, le mouvement Brownien est un processus de $(C[0, 1], \mathbb{R})$ gaussien à accroissements indépendants. La définition du processus engendre la stationnarité des accroissements

du processus :

$$\mathcal{W}(t+s) - \mathcal{W}(t) \sim \mathcal{N}(0, s), \quad 0 \leq t, s \leq 1.$$

La covariance du processus est

$$\text{cov}(\mathcal{W}(t), \mathcal{W}(s)) = E(\mathcal{W}(t)\mathcal{W}(s)) = \min(s, t), \quad 0 \leq t, s \leq 1.$$

Nous définissons le mouvement Brownien dans \mathbb{R}^p comme un vecteur de p mouvements Browniens réels indépendants. A partir du mouvement Brownien, on peut définir le processus du pont Brownien.

Définition 2.2.2. *Le pont Brownien sur $[0, 1]$, $\mathcal{B} = \{\mathcal{B}(t), 0 \leq t \leq 1\}$ est un processus défini par*

$$\mathcal{B}(t) = \mathcal{W}(t) - t\mathcal{W}(1), \quad 0 \leq t \leq 1. \quad (2.4)$$

Le pont Brownien est un processus gaussien centré, nul en 0 et en 1 presque sûrement. Sa fonction de covariance est

$$\text{cov}(\mathcal{B}(t), \mathcal{B}(s)) = s(1-t), \quad 0 \leq s \leq t \leq 1.$$

Le supremum du pont Brownien est distribué selon la loi de Kolmogorov de fonction de répartition

$$P\left(\sup_{0 \leq t \leq 1} |\mathcal{B}(t)| \leq a\right) = 1 - 2 \sum_{k \geq 1} (-1)^{k+1} \exp(-2k^2 a^2), \quad a \geq 0.$$

Cette distribution sera utile pour construire des tests et des bandes de confiance asymptotiques de processus convergeant vers un pont Brownien.

Le mouvement Brownien est le processus limite obtenu par application des théorèmes de la limite centrale fonctionnels que nous énonçons dans cette section. Ces théorèmes seront utilisés pour établir le comportement limite de processus empiriques définis dans les chapitres suivants de ce manuscrit. Un théorème central limite fonctionnel permet d'établir le comportement limite d'un processus empirique vérifiant de bonnes hypothèses. Le plus célèbre d'entre eux est le théorème de Donsker, qui étudie le cas de l'interpolation linéaire d'un processus dont les incréments sont des variables aléatoires indépendantes, identiquement distribuées, centrées et de même variance.

Théorème 2.2.2 (Donsker, 1951). *Si $(\xi_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes et identiquement distribuées sur un espace probabilisé (Ω, \mathcal{F}, P) telles que $E(\xi_n) = 0$ et $V(\xi_n) = \sigma^2$ pour tout $n \in \mathbb{N}$, alors*

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{W},$$

où, pour tout $t \in [0, 1]$,

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \xi_i + (nt - \lfloor nt \rfloor) \frac{1}{\sigma\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}.$$

En relâchant l'hypothèse de distribution identique et d'indépendance des variables aléatoires et en considérant des variables centrées, non identiquement distribuées et de variances différentes, on peut appliquer le théorème central limite fonctionnel introduit par Helland (1982). Ce théorème traite les suites de variables aléatoires comme des différences de martingales par rapport à une certaine famille de σ -algèbres, sans interpoler linéairement ces variables.

Théorème 2.2.3 (Helland, 1982). Soit $\xi_{j,n}$ une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{F}, P) , pour $j = 1, \dots, n$. Soit $\mathcal{F}_{j,n}$ une σ -algèbre telle que $\xi_{j,n}$ est $\mathcal{F}_{j,n}$ -mesurable et $\mathcal{F}_{j-1,n} \subset \mathcal{F}_{j,n} \subset \mathcal{F}$ pour tout $j = 2, \dots, n$. Soit r_n une fonction définie sur \mathbb{R}^+ telle que $r_n(t)$ est un temps d'arrêt par rapport à $\mathcal{F}_{j,n}, j = 1, \dots, n$. Supposons que les trajectoires de r_n sont continues à droite, à valeurs entières et croissantes, avec $r_n(0) = 0$. Notons

$$X_n(t) = \sum_{j=1}^{r_n(t)} \xi_{j,n}. \quad (2.5)$$

Soit f une fonction mesurable et positive telle que $\int_0^t f^2(s) ds < \infty, \forall t > 0$. Si les conditions suivantes sont vérifiées :

a. $\xi_{j,n}$ est une différence de martingales, c'est-à-dire

$$E(\xi_{j,n} | \mathcal{F}_{j-1,n}) = 0, \quad j = 1, \dots, n,$$

b. $\sum_{j=1}^{r_n(t)} E(\xi_{j,n}^2 | \mathcal{F}_{j-1,n}) \xrightarrow[n \rightarrow \infty]{P} \int_0^t f^2(s) ds,$

c. $\sum_{j=1}^{r_n(t)} E(\xi_{j,n}^2 I(|\xi_{j,n}| > \varepsilon) | \mathcal{F}_{j-1,n}) \xrightarrow[n \rightarrow \infty]{P} 0, \forall \varepsilon > 0,$

Alors,

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} f\mathcal{W},$$

où $f\mathcal{W}(t) = \int_0^t f(s) d\mathcal{W}(s)$ pour $t \in \mathbb{R}^+$.

Dans le théorème précédent, si f est la fonction constante égale à 1, le processus X_n converge en loi vers un mouvement Brownien standard. La condition (c) est entraînée par la condition de Lyapunov notée (c') telle que

$$\exists \delta > 0, \sum_{j=1}^{r_n(t)} E(\xi_{j,n}^{2+\delta} | \mathcal{F}_{j-1,n}) \xrightarrow[n \rightarrow \infty]{P} 0. \quad (2.6)$$

C'est cette dernière condition que nous utilisons dans les démonstrations de ce manuscrit. Enfin, le théorème central limite fonctionnel et multivarié suivant donne le comportement asymptotique de vecteurs aléatoires centrés, de lois et de matrices de variance-covariance différentes.

Théorème 2.2.4 (Helland, 1982). Soit $n, m \in \mathbb{N}^*$. Soit $\xi_{j,n}^{(l)}$ une variable aléatoire définie sur un espace probabilisé (Ω, \mathcal{F}, P) , pour $j = 1, \dots, n$ et $l = 1, \dots, m$. On suppose que les ensembles $\{\xi_{j,n}^{(l)}, j = 1, \dots, n, n = 1, 2, \dots\}$ sont des tableaux de différences de martingales par rapport à la suite croissante de σ -algèbres $(\mathcal{F}_{j,n})_{j=1,2,\dots,n}$, pour $l = 1, \dots, m$. Soit r_n une fonction définie sur \mathbb{R}^+ telle que $r_n(t)$ est un temps d'arrêt par rapport à $\mathcal{F}_{j,n}, j = 1, \dots, n$. Supposons que les trajectoires de r_n sont continues à droite, à valeurs entières et croissantes, avec $r_n(0) = 0$. Notons

$$X_n(t) = \left(X_n^{(1)}(t), \dots, X_n^{(m)}(t) \right), \quad X_n^{(l)}(t) = \sum_{j=1}^{r_n(t)} \xi_{j,n}^{(l)}, \quad l = 1, \dots, m. \quad (2.7)$$

Soit f_1, \dots, f_m , m fonctions mesurables et positives telles que $\int_0^t f_l^2(s) ds < \infty, \forall t > 0$, pour $l = 1, \dots, m$. Si les conditions suivantes sont vérifiées, pour tout $i, l = 1, \dots, m$:

- a. $\sum_{j=1}^{r_n(t)} E \left(\xi_{j,n}^{(l)} \xi_{j,n}^{(i)} \middle| \mathcal{F}_{j-1,n} \right) \xrightarrow[n \rightarrow \infty]{P} 0$, pour $t > 0$ et $l \neq i$.
- b. $\sum_{j=1}^{r_n(t)} E \left(\left(\xi_{j,n}^{(l)} \right)^2 \middle| \mathcal{F}_{j-1,n} \right) \xrightarrow[n \rightarrow \infty]{P} \int_0^t f_l^2(s) ds$,
- c. $\sum_{j=1}^{r_n(t)} E \left(\left(\xi_{j,n}^{(l)} \right)^2 I(|\xi_{j,n}^{(l)}| > \varepsilon) \middle| \mathcal{F}_{j-1,n} \right) \xrightarrow[n \rightarrow \infty]{P} 0$, $\forall \varepsilon > 0$,

Alors, $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} (\int f_1 d\mathcal{W}_1, \dots, \int f_m d\mathcal{W}_m)$ par rapport à la topologie produit de Skorokhod, où $\int_0^t f_l d\mathcal{W}_l = \int_0^t f_l(s) d\mathcal{W}_l(s)$ pour $t \in \mathbb{R}^+$, $l = 1, \dots, m$ et $\mathcal{W}_1, \dots, \mathcal{W}_m$ sont m mouvements Browniens indépendants.

2.3 Inégalité de Taylor-Lagrange multidimensionnelle

On note $\mathcal{M}_{p \times p}(\mathbb{R})$ l'ensemble des matrices de dimension $p \times p$ à coefficients réels. Soit A une matrice de $\mathcal{M}_{p \times p}(\mathbb{R})$ et a un vecteur de \mathbb{R}^p . Notons $\|A\| = \max_{i,j=1,\dots,p} |A_{i,j}|$ la norme sup de la matrice A et $\|a\| = \max_{i=1,\dots,p} |a_i|$ la norme sup du vecteur a . Dans la suite de notre travail, sauf mention du contraire, nous utilisons ces normes matricielles et vectorielles. Remarquons que la norme matricielle du sup n'est pas sous-multiplicative, mais on a la relation

$$\|AB\| \leq p \|A\| \|B\|, \quad A, B \in \mathcal{M}_{p \times p}(\mathbb{R}). \quad (2.8)$$

La norme matricielle du sup n'étant pas sous-multiplicative, elle n'est pas subordonnée à la norme vectorielle du sup. En revanche, on a l'inégalité suivante

$$\|Ax\| \leq p \|A\| \|x\|, \quad A \in \mathcal{M}_{p \times p}(\mathbb{R}), \quad x \in \mathbb{R}^p. \quad (2.9)$$

Introduisons une inégalité de type Taylor-Lagrange pour une fonction de \mathbb{R}^p à valeurs dans \mathbb{R}^p .

Théorème 2.3.1. Soit f une fonction d'un ouvert $U \subset \mathbb{R}^p$ à valeurs dans \mathbb{R}^p de classe C^2 sur U et $[a, a+h]$ un segment inclus dans U . Si f admet en tout point de $]a, a+h[$ une différentielle d'ordre 2 et s'il existe un réel M tel que $\sup_{x \in]a, a+h[} \left\| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right\| \leq M$, alors

$$\|f(a+h) - f(a) - Jf(a)h\| \leq \frac{1}{2}M\|h\|^2,$$

où $Jf(x)$ est la matrice jacobienne de $f = (f_1, \dots, f_p)$ calculée au point $x \in]a, a+h[$ dont l'élément (i, j) est $Jf(x)_{i,j} = \frac{\partial}{\partial x_j} f_i(x)$, pour $i, j = 1, \dots, p$.

A la différence des fonctions de \mathbb{R} dans \mathbb{R} , il n'existe pas d'égalité de Taylor–Lagrange pour les fonctions de \mathbb{R}^p dans \mathbb{R}^p . Ceci est dû au fait que les dérivations de chaque fonction de \mathbb{R}^p dans \mathbb{R} ne s'effectuent pas dans la même direction.

Chapitre 3

Etude du processus du score standardisé

Dans ce chapitre, nous étudions le processus du score standardisé initialement développé par [O'Quigley \(2003\)](#). Ce processus a été défini dans le cas univarié, c'est-à-dire lorsqu'une seule covariable est présente dans le modèle. Dans cette thèse, nous étendons la construction du processus au cas de plusieurs covariables et nous obtenons les comportements limites des processus univariés et multivariés. Les résultats de convergence des processus seront utilisés dans les chapitres suivants pour construire une classe de tests de la présence d'un effet des covariables sur la survie. Aussi, ils permettront d'évaluer l'adéquation des modèles à risques proportionnels et non proportionnels. Les résultats de ce chapitre ont fait l'objet de deux articles ([Chauvel et O'Quigley, 2014a,b](#)).

3.1 Introduction

Dans le modèle à risques proportionnels (1.4) de paramètre β , plusieurs processus empiriques ont été étudiés afin de vérifier l'adéquation du modèle aux données ou d'effectuer des tests d'hypothèses sur la valeur du paramètre de régression. Ces processus dépendent d'un paramètre à fixer. Lorsque le paramètre choisi est l'estimateur du maximum de vraisemblance partiel $\hat{\beta}$, le processus permet de vérifier l'adéquation du modèle à risques proportionnels aux données, par des méthodes graphiques ou des tests de l'hypothèse nulle $H_0 : \beta = \hat{\beta}$ contre son alternative ([Barlow et Prentice, 1988](#), [Lin et al., 1993](#), [Therneau et al., 1990](#), [Wei, 1984](#)). Une autre possibilité est de fixer le paramètre à une valeur β_0 particulière afin de développer des tests de l'hypothèse nulle $H_0 : \beta = \beta_0$ contre l'alternative ([Cox, 1972](#), [Mantel, 1966](#)).

Le processus du score a été introduit par [Wei \(1984\)](#) dans le cas d'une covariable Z binaire, par exemple représentant deux groupes de traitement. Le processus du score évalué au temps t et avec un paramètre de régression β est défini par

$$U(\beta, t) = \sum_{i=1}^n \int_0^t \{Z_i(s) - \mathcal{E}_\beta(Z | s)\} dN_i(s), \quad 0 \leq t \leq \mathcal{T}. \quad (3.1)$$

A chaque temps de décès, le processus est incrémenté de la différence entre la valeur de la covariable de l'individu qui décède et son espérance sous le modèle. Au dernier temps de décès, le processus est égal à la dérivée de la log-vraisemblance, aussi appelé score (équation (1.8)), ce qui lui confère sa dénomination.

L'intérêt du processus (3.1) réside dans le fait d'évaluer les contributions successives à la vraisemblance. [Wei \(1984\)](#) a démontré la convergence du processus vers un pont Brownien lorsque le paramètre est l'estimateur du maximum de vraisemblance partielle. [Haara \(1987\)](#) a étendu ce résultat aux cas plus généraux de covariables non binaires. De nombreux travaux sont basés sur ce processus. Par exemple, pour vérifier l'adéquation du modèle à risques proportionnels (1.4), [Therneau et al. \(1990\)](#) ont proposé de tracer le processus du score avec une standardisation globale en fonction du temps et de comparer le processus résultant avec un pont Brownien. La convergence du processus vers un pont Brownien n'étant plus vraie en présence de covariables corrélées, [Lin et al. \(1993\)](#) ont proposé de comparer le processus à une enveloppe simulée de processus gaussiens. [Lin et al. \(1996\)](#) ont également étendu l'utilisation de ce processus dans le cadre de censure dépendante des covariables, par des méthodes graphiques et des tests.

D'autres processus empiriques existent pour évaluer l'adéquation du modèle à risques proportionnels. [Arjas \(1988\)](#) a proposé un processus assez proche du processus du score. Il s'agit, à chaque temps de décès, d'ajouter au processus la différence entre le nombre de décès courants et le nombre de décès attendus sous le modèle. [Barlow et Prentice \(1988\)](#) ont développé une classe de résidus dits de martingales. Ces résidus sont donnés pour chaque individu dans l'étude en fonction du temps et dépendent d'une fonction de poids. Une fois sommés sur les individus, les résidus permettent de retrouver plusieurs processus empiriques selon la fonction de poids considérée et notamment le processus du score ([Lin et al., 1993](#), [Therneau et Grambsch, 2000](#)). Nous renvoyons au Chapitre 5 pour plus de détails sur ces processus. Remarquons que tous les processus cités jusqu'à présent ne s'appliquent qu'au modèle à risques proportionnels (1.4).

Le processus empirique que nous construisons dans les sections suivantes permet à la fois d'effectuer des tests d'hypothèses sur la valeur du paramètre et d'étudier l'adéquation du modèle à risques non proportionnels sans avoir à estimer le coefficient de régression. En plus de cet aspect, les incréments du processus sont standardisés à chaque temps, plutôt que d'adopter une standardisation globale pour tous les incréments comme la construction adoptée par [Wei \(1984\)](#). Cette standardisation permet de considérer des covariables multiples et corrélées, car, à la différence du processus du score non standardisé (3.1), la corrélation entre les covariables n'est pas supposée constante au cours du temps. Les variables sont décorréélées avant d'être utilisées dans le processus. Aussi, nous proposons de tracer le processus selon les rangs des temps de décès, à la manière d'[Arjas \(1988\)](#), pour obtenir une loi asymptotique simple et explicite. Avant de construire le processus, nous devons procéder à un changement de l'échelle du temps. Cette transformation sera cruciale pour l'obtention des résultats asymptotiques du processus, sans avoir à utiliser le théorème de [Rebolledo \(1980\)](#) ni l'inégalité de [Lenglart \(1977\)](#) qui sont des outils classiques de l'analyse de survie. Dans la section suivante, nous présentons

la transformation du temps utilisée. En Section 3.3, nous définissons le processus du score standardisé univarié et étudions son comportement asymptotique. En Section 3.4, nous nous intéressons au cas multivarié. La Section 3.5 regroupe les preuves des théorèmes de ce chapitre.

3.2 Transformation de l'échelle du temps

Dans le modèle à risques proportionnels, une transformation strictement monotone croissante sur les temps ne change pas l'inférence sur le coefficient de régression β . En effet, les valeurs des temps de décès et de censure ne sont pas utilisées pour l'inférence, seul leur ordre d'arrivée compte. Nous allons donc appliquer une transformation conservant les rangs des temps observés. En ce qui concerne le modèle à risques non proportionnels, la transformation va être captée dans la fonction de régression $\beta(t)$, mais en appliquant la transformation inverse, on pourra se ramener à l'échelle des temps initiale.

Notons $\#E$ le nombre d'éléments de l'ensemble E . On considère une constante $C_V > 0$ très petite et choisie arbitrairement. Dans le cas d'une unique covariable ($p = 1$), soit

$$\widehat{k}_n(\beta) = \# \left\{ i \in \{1, \dots, n\}; \delta_i = 1; \mathcal{V}_{\beta(X_i)}(Z | X_i) > C_V \right\},$$

le nombre de décès observés tels que les variances conditionnelles définies équation (1.7) et évaluées à ces temps soient strictement supérieures à C_V . En pratique, C_V est choisie comme étant la précision machine, qui est le plus petit réel positif que l'ordinateur distingue de 0. De cette manière, $\widehat{k}_n(\beta)$ est égal au nombre de temps de décès pour lesquels les variances conditionnelles calculées avec le paramètre $\beta(t)$ sont strictement positives. La stricte positivité de la constante C_V est utile pour la théorie mais est peut-être non nécessaire. L'introduction de la constante permet de minorer toutes les variances conditionnelles par une constante qui ne dépend pas de la taille de l'échantillon n . Dans les simulations et les applications, nous avons considéré toutes les variances conditionnelles non nulles.

Si la variance est nulle à un temps de décès, cela signifie que l'ensemble des individus à risque est composé d'une population homogène partageant la même valeur de la covariable. Cela se traduit mathématiquement par une contribution nulle de ces temps de décès au score de l'équation (1.8) étant donné que la valeur de la covariable de l'individu qui décède est égale à son espérance (équation (1.6)). Nous voulons éviter ces variances d'un point de vue technique, car nous allons standardiser par la racine carrée de la variance dans la suite. Par exemple, dans le cas de la comparaison de deux groupes de traitement, une variance conditionnelle nulle correspond à une situation dans laquelle un groupe est épuisé, mais des individus du second groupe sont encore présents dans l'étude. Intuitivement, on comprend que ces temps de décès ne contribuent pas à l'estimation du coefficient de régression β puisqu'aucune information pour comparer les deux groupes n'est présente. Remarquons que la nullité d'une variance conditionnelle à un temps donné implique la nullité des variances aux temps suivants. Ainsi, d'après la définition de $\widehat{k}_n(\beta)$, les variances conditionnelles $\mathcal{V}_{\beta(t)}(Z | t)$ évaluées pour les $\widehat{k}_n(\beta)$ premiers temps de décès t sont strictement supérieures à C_V . Pour une covariable continue

et si le dernier temps observé est un décès, on a l'égalité presque-sûre $\widehat{k}_n(\beta) = \sum_{i=1}^n \delta_i - 1$. En effet, la variance conditionnelle $\mathcal{V}_{\beta(t)}(Z | t)$ est nulle au dernier temps observé t et c'est presque-sûrement le seul temps pour lequel l'ensemble des individus à risque est composé d'individus présentant la même valeur de la covariable. Si le dernier temps observé est une censure, alors $\widehat{k}_n(\beta) = \sum_{i=1}^n \delta_i$ presque-sûrement.

Dans le cas de plusieurs covariables ($p > 1$), on note

$$\widehat{k}_n(\beta) = \# \left\{ i \in \{1, \dots, n\}; \delta_i = 1; \quad x^T \mathcal{V}_{\beta(X_i)}(Z | X_i) x > 0, \forall x \in \mathbb{R}^p \setminus \{0\}; \right. \\ \left. \left\| \mathcal{V}_{\beta(X_i)}(Z | X_i)^{-1} \right\| \leq C_V^{-1} \right\},$$

le nombre de décès observés tels que les matrices de variance-covariance conditionnelles évaluées à ces temps soient définies positives et que la norme de leur inverse soit majorée par une constante finie. De la même manière que dans le cas univarié, nous avons uniquement considéré les variances définies positives dans les simulations et les applications. La constante C_V est introduite pour l'obtention des résultats théoriques, mais les simulations indiquent qu'elle n'est peut-être pas nécessaire. Il serait intéressant de trouver une autre approche théorique qui n'utilise pas cette constante.

On suppose qu'il existe $\alpha_0 \in]0, 1]$ tel que

$$\frac{\widehat{k}_n(\beta)}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \alpha_0.$$

De plus, on suppose qu'il existe une suite $(a_n)_n$ convergeant vers 1 et $\widetilde{N} \in \mathbb{N}^*$ tel que pour tout $n \geq \widetilde{N}$,

$$a_n \geq 1, \quad \frac{\widehat{k}_n(\beta)}{n} a_n \geq \alpha_0 \quad \text{p.s.} \quad (3.2)$$

Ceci implique en particulier que $\widehat{k}_n(\beta)$ tende vers l'infini lorsque n tend vers l'infini, c'est-à-dire que le nombre de décès augmente sans borne avec la taille de l'échantillon. Si Z est une variable discrète, cela signifie que pour un échantillon de taille infinie, il n'y a pas d'épuisement des individus dans un groupe. Si Z est continue, $\widehat{k}_n(\beta) = \sum_{i=1}^n \delta_i - 1$ ou $\widehat{k}_n(\beta) = \sum_{i=1}^n \delta_i$ presque-sûrement. Dans ce cas, $n^{-1} \widehat{k}_n(\beta)$ est un estimateur convergeant de $P(T \leq C)$ et $\alpha_0 = P(T \leq C)$. La loi du logarithme itéré permet alors de construire une suite (a_n) vérifiant l'équation (3.2).

Remarque 3.2.1. *On se place dans le cas d'une covariable continue avec une censure au dernier temps observé, ce qui implique que $\widehat{k}_n(\beta) = \sum_{i=1}^n \delta_i = \sum_{i=1}^n \mathbf{1}_{X_i \leq C_i}$. Les variables $\delta_1, \dots, \delta_n$ sont indépendantes et identiquement distribuées et admettent un moment d'ordre 2 fini. Leur espérance est $\alpha_0 = P(T \leq C)$ et leur variance est $\alpha_0(1 - \alpha_0)$. D'après la loi du logarithme itéré,*

$$\liminf_{n \rightarrow \infty} \frac{\sqrt{n} (n^{-1} \widehat{k}_n(\beta) - \alpha_0)}{\sqrt{2\alpha_0(1 - \alpha_0) \log \log(n)}} = -1 \quad \text{p.s.}$$

Ainsi,

$$\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N}^*, \forall n \geq N_\varepsilon, \frac{\sqrt{n} \left(n^{-1} \widehat{k}_n(\beta) - \alpha_0 \right)}{\sqrt{2\alpha_0(1-\alpha_0) \log \log(n)}} \geq -1 - \varepsilon \quad \text{p.s.}$$

En posant $\varepsilon = 1/2$, $\exists N_{1/2} \in \mathbb{N}^*$, $\forall n \geq N_{1/2}$,

$$\frac{\sqrt{n} \left(n^{-1} \widehat{k}_n(\beta) - \alpha_0 \right)}{\sqrt{2\alpha_0(1-\alpha_0) \log \log(n)}} \geq -\frac{3}{2} \quad \text{p.s.}, \quad (3.3)$$

$$\text{et } \sqrt{\frac{\log \log(n)}{n}} \leq \frac{1}{3} \sqrt{\frac{2\alpha_0}{1-\alpha_0}}. \quad (3.4)$$

Soit $n \geq N_{1/2}$. Alors, d'après l'équation (3.3),

$$\frac{\widehat{k}_n(\beta)}{n} \geq \alpha_0 - \frac{3}{\sqrt{2n}} \sqrt{\alpha_0(1-\alpha_0) \log \log(n)},$$

et en posant

$$a_n = \left(1 - \frac{3}{\sqrt{2\alpha_0 n}} \sqrt{(1-\alpha_0) \log \log(n)} \right)^{-1},$$

on obtient

$$\frac{\widehat{k}_n(\beta)}{n} a_n \geq \alpha_0.$$

D'après l'équation (3.4), $a_n \geq 1$ et on a bien $\lim_{n \rightarrow \infty} a_n = 1$. On a ainsi construit une suite (a_n) qui vérifie l'équation (3.2).

La variable aléatoire $\widehat{k}_n(\beta)$ n'est connue qu'à la fin du recueil des données. Elle ne peut donc pas être utilisée au cours de l'expérience pour construire un processus prévisible (au sens d'une famille croissante de σ -algèbres définie ci-après). C'est pourquoi nous définissons la quantité déterministe

$$k_n(\beta) = n \frac{\alpha_0}{a_n},$$

pour obtenir les résultats théoriques. Nous avons la convergence presque sûre $\frac{\widehat{k}_n(\beta)}{k_n(\beta)} \xrightarrow[n \rightarrow \infty]{} 1$, et l'équation (3.2) implique que

$$\forall n \geq \tilde{N}, \quad \widehat{k}_n(\beta) \geq k_n(\beta) \quad \text{p.s.} \quad (3.5)$$

Dans les applications, $k_n(\beta)$ est inconnu et il est remplacé par son estimateur consistant $\widehat{k}_n(\beta)$. Dans la suite du mémoire de thèse, on suppose que la taille de l'échantillon n est supérieure à \tilde{N} . Pour simplifier les notations et jusqu'à mention du contraire, on pose $\widehat{k}_n = \widehat{k}_n(\beta)$ et $k_n = k_n(\beta)$.

Considérons la transformation ϕ_n sur les temps observés X_1, \dots, X_n suivante

$$\phi_n(X_i) = \frac{\bar{N}(X_i)}{k_n} \left[1 + (1 - \delta_i) \frac{\#\{j : j = 1, \dots, n, X_j < X_i, \bar{N}(X_j) = \bar{N}(X_i)\}}{\#\{j : j = 1, \dots, n, \bar{N}(X_j) = \bar{N}(X_i)\}} \right], \quad (3.6)$$

et on pose $\phi_n(0) = 0$. On définit l'inverse de ϕ_n sur $[0, 1]$ par ϕ_n^{-1} tel que

$$\phi_n^{-1}(t) = \inf \{X_i, \phi_n(X_i) \geq t, i = 1, \dots, n\}, \quad 0 \leq t \leq 1.$$

Rappelons que le processus de comptage $\{\bar{N}(t)\}_{t \in [0, \mathcal{T}]}$ présente un saut de taille 1 à chaque temps de décès. Ainsi, sur la nouvelle échelle du temps qui est l'image des temps observés $\{X_1, \dots, X_n\}$ par ϕ_n , les valeurs de l'ensemble $\{1/k_n, 2/k_n, \dots, 1\}$ correspondent à des temps de décès et le i ème temps de décès, noté t_i , est tel que

$$t_i = i/k_n, \quad i = 1, \dots, k_n.$$

On pose $t_0 = 0$. L'ensemble $\{1/k_n, 2/k_n, \dots, 1\}$ est inclus, mais pas nécessairement égal à l'ensemble image de tous les temps de décès. Travaillant avec les k_n premiers temps de décès, nous allons nous restreindre ainsi aux valeurs inférieures à l'unité. L'importance dans la transformation appliquée aux temps observés étant de conserver l'ordre, plusieurs transformations traitant différemment les temps de censure pourraient être utilisées. Dans la formule (3.6), nous choisissons de répartir uniformément les temps de censure entre les temps de décès adjacents. Le temps τ_0 sur cette nouvelle échelle correspond au $100 \times \tau_0$ ème percentile de décès dans l'échantillon. Par exemple, au temps $\tau_0 = 0.5$, la moitié des décès a été observée. Au temps $\tau_0 = 0.2$, 20% des décès se sont produits. La transformation inverse de ϕ_n peut être obtenue facilement et permet d'interpréter les résultats sur l'échelle du temps initiale. Dans la suite, nous travaillons avec l'échelle du temps standardisée dans $[0, 1]$.

Sur l'échelle du temps transformée, pour $t \in [0, 1]$, nous définissons l'indicateur d'être à risque $Y_i^*(t)$ et le processus de comptage individuel $N_i^*(t)$, pour l'individu $i = 1, \dots, n$, par

$$\begin{aligned} Y_i^*(t) &= \mathbf{1}_{\phi_n(X_i) \geq t}, \\ N_i^*(t) &= \mathbf{1}_{\phi_n(X_i) \leq t, \delta_i = 1}. \end{aligned}$$

Chaque temps de décès t_i est un $(\mathcal{F}_t^*)_{t \in [0, 1]}$ -temps d'arrêt, où, pour $t \in [0, 1]$, la σ -algèbre \mathcal{F}_t^* est définie par

$$\mathcal{F}_t^* = \sigma \left\{ N_i^*(u), Y_i^*(u^+), Z_i; i = 1, \dots, n; u = 0, 1, \dots, [tk_n] \right\},$$

$[\cdot]$ est la fonction partie entière et $Y_i^*(t^+) = \lim_{s \rightarrow t^+} Y_i^*(s)$. On remarque que si $0 \leq s < t \leq 1$, $\mathcal{F}_s^* \subset \mathcal{F}_t^*$.

Remarque 3.2.2. Notons $a \wedge b = \min(a, b)$, pour $a, b \in \mathbb{R}$. Si les covariables dépendent du

temps, pour $t \in [0, 1]$, la σ -algèbre \mathcal{F}_t^* est définie par

$$\mathcal{F}_t^* = \sigma \left\{ N_i^*(u), Y_i^*(u^+), Z_i \left(\phi_n^{-1} \left(\frac{u}{k_n} \right)^+ \wedge X_i \right); i = 1, \dots, n; u = 0, 1, \dots, \lfloor tk_n \rfloor \right\},$$

puisque nous rappelons que la covariable $Z_i(\cdot)$ n'est pas observée après le temps X_i , $i = 1, \dots, n$. Comme à la Remarque 1.2.1, dans la suite du travail et pour simplifier les notations, nous écrivons $Z_i(\phi_n^{-1}(t))$ à la place de

$$Z_i \left(\phi_n^{-1}(t)^+ \wedge X_i \right), \quad t \in [0, 1].$$

Nous définissons le processus de comptage associé aux temps transformés qui a un saut de taille 1 aux temps de décès sur la nouvelle échelle par

$$\bar{N}^*(t) = \sum_{i=1}^n \mathbf{1}_{\phi_n(X_i) \leq t, \delta_i = 1}, \quad 0 \leq t \leq 1.$$

Ce processus de comptage satisfait le résultat suivant :

Proposition 3.2.1. $\forall t \in [0, 1], \bar{N}^*(t) = \lfloor k_n t \rfloor$.

Preuve de la Proposition 3.2.1. Le processus de comptage \bar{N} sur l'échelle du temps initiale est à valeurs dans \mathbb{N} . Ainsi, pour tout $u \in [0, \mathcal{T}]$,

$$\{x \in \mathbb{R} : \bar{N}(u) \leq x\} = \{x \in \mathbb{R} : \bar{N}(u) \leq \lfloor x \rfloor\}.$$

Le processus inverse est tel que $\bar{N}^{-1}(j)$ est le j ème temps de décès ordonné, $j = 1, \dots, k_n$. Remarquons que $\bar{N}(\bar{N}^{-1}(j)) = j$. Donc, par définition de \bar{N}^* , nous avons, pour $0 \leq t \leq 1$,

$$\begin{aligned} \bar{N}^*(t) &= \sum_{i=1}^n \mathbf{1}_{\{\bar{N}(X_i) \leq \lfloor k_n t \rfloor, \delta_i = 1\}} \\ &= \sum_{i=1}^n \mathbf{1}_{\{X_i \leq \bar{N}^{-1}(\lfloor k_n t \rfloor), \delta_i = 1\}} \\ &= \bar{N}(\bar{N}^{-1}(\lfloor k_n t \rfloor)) = \lfloor k_n t \rfloor. \end{aligned}$$

□

Autrement dit, \bar{N}^* présente un saut de taille 1 aux temps $t_j = j/k_n$ pour $j = 1, \dots, k_n$. Ceci implique le lemme suivant qui sera utilisé pour prouver le Théorème 3.3.2 :

Lemme 3.2.1. Soit $n \in \mathbb{N}$ et $\{A_n(t), t \in \{t_1, \dots, t_{k_n}\}\}$ un processus à valeurs dans \mathbb{R} . Supposons que

$$\sup_{i=1, \dots, k_n} |A_n(t_i) - a(t_i)| \xrightarrow[n \rightarrow \infty]{P} 0,$$

où a est une fonction définie et bornée sur $[0, 1]$. Alors,

$$\sup_{s \in [0, 1]} \left| \frac{1}{k_n} \int_0^s A_n(t) d\bar{N}^*(t) - \int_0^s a(t) dt \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

La preuve du lemme est donnée en Section 3.5. Pour faciliter les notations, nous définissons un processus, noté \mathcal{Z} , qui vaut 0 partout sauf aux temps de décès. A chaque temps de décès, ce processus prend la valeur de la covariable à ce temps de l'individu qui décède.

Définition 3.2.1. On note $\mathcal{Z} = \{\mathcal{Z}(t), t \in [0, 1]\}$ un processus tel que

$$\mathcal{Z}(t) = \sum_{i=1}^n Z_i(X_i) \mathbf{1}_{\phi_n(X_i) = t, \delta_i = 1}, \quad t \in [0, 1]. \quad (3.7)$$

La famille de probabilités $\{\pi_i(\beta(t), t), i = 1, \dots, n\}$ avec $t \in [0, \mathcal{T}]$ définie en (1.5) peut être étendue au modèle à risques non proportionnels et sur l'échelle transformée comme suit.

Définition 3.2.2. Pour $i = 1, \dots, n$ et $t \in [0, 1]$, la probabilité pour que l'individu i décède au temps t sous le modèle à risques non proportionnels de paramètre $\beta(t)$ conditionnellement à l'ensemble des individus à risque au temps de décès t est définie par

$$\pi_i(\beta(t), t) = \begin{cases} \frac{Y_i^*(t) \exp\left(\beta(t)^T Z_i(\phi_n^{-1}(t))\right)}{\sum_{j=1}^n Y_j^*(t) \exp\left(\beta(t)^T Z_j(\phi_n^{-1}(t))\right)} & \text{si } k_n t \in \mathbb{N}, \\ 0 & \text{sinon.} \end{cases}$$

Remarque 3.2.3. Les seules valeurs pour lesquelles ces quantités seront calculées par la suite sont les temps de décès t_1, \dots, t_{k_n} . On définit pourtant $\mathcal{Z}(t)$ et $\{\pi_i(\beta(t), t), i = 1, \dots, n\}$ pour tout $t \in [0, 1]$ de manière à pouvoir écrire leurs sommes respectives comme des intégrales par rapport au processus de comptage \bar{N}^* . Les valeurs 0 que prennent ces quantités en dehors des temps de décès sont choisies arbitrairement.

Les espérances et variances par rapport à la famille de probabilité $\{\pi_i(\beta(t), t), i = 1, \dots, n\}$ peuvent maintenant être définies.

Définition 3.2.3. Soit $t \in [0, 1]$. L'espérance et la variance de Z par rapport à la famille de probabilité $\{\pi_i(\beta(t), t), i = 1, \dots, n\}$ sont respectivement un vecteur de \mathbb{R}^p et une matrice de $\mathcal{M}_{p \times p}(\mathbb{R})$ tels que

$$\mathcal{E}_{\beta(t)}(Z | t) = \sum_{i=1}^n Z_i(\phi_n^{-1}(t)) \pi_i(\beta(t), t), \quad (3.8)$$

et

$$\mathcal{V}_{\beta(t)}(Z | t) = \sum_{i=1}^n Z_i(\phi_n^{-1}(t))^{\otimes 2} \pi_i(\beta(t), t) - \mathcal{E}_{\beta(t)}(Z | t)^{\otimes 2}. \quad (3.9)$$

On peut montrer que la matrice jacobienne de $\mathcal{E}_{\beta(t)}(Z | t)$ est $\frac{\partial}{\partial \beta} \mathcal{E}_{\beta(t)}(Z | t) = \mathcal{V}_{\beta(t)}(Z | t)$, pour $t \in [0, 1]$.

Proposition 3.2.2. Soit $t \in [0, 1]$. Sous le modèle à risques non proportionnels de paramètre $\beta(t)$, l'espérance $E_{\beta(t)}(\mathcal{Z}(t) \mid \mathcal{F}_{t^-}^*)$ et la variance $V_{\beta(t)}(\mathcal{Z}(t) \mid \mathcal{F}_{t^-}^*)$ de la variable aléatoire $\mathcal{Z}(t)$ conditionnellement à la σ -algèbre $\mathcal{F}_{t^-}^*$ sont

$$\begin{aligned} E_{\beta(t)}(\mathcal{Z}(t) \mid \mathcal{F}_{t^-}^*) &= \mathcal{E}_{\beta(t)}(Z \mid t), \\ V_{\beta(t)}(\mathcal{Z}(t) \mid \mathcal{F}_{t^-}^*) &= \mathcal{V}_{\beta(t)}(Z \mid t). \end{aligned}$$

Preuve de la Proposition 3.2.2. Soit $t \in [0, 1]$. L'espérance de $\mathcal{Z}(t)$ (défini en (3.7)) conditionnellement à la σ -algèbre $\mathcal{F}_{t^-}^*$ est

$$\begin{aligned} E_{\beta(t)}(\mathcal{Z}(t) \mid \mathcal{F}_{t^-}^*) &= E_{\beta(t)}\left(\sum_{i=1}^n Z_i(X_i) \mathbf{1}_{\phi_n(X_i) = t, \delta_i = 1} \mid \mathcal{F}_{t^-}^*\right) \\ &= E_{\beta(t)}\left(\sum_{i=1}^n Z_i(\phi_n^{-1}(t)) \mathbf{1}_{\phi_n(X_i) = t, \delta_i = 1} \mid \mathcal{F}_{t^-}^*\right) \\ &= \sum_{i=1}^n Z_i(\phi_n^{-1}(t)) E_{\beta(t)}\left(\mathbf{1}_{\phi_n(X_i) = t, \delta_i = 1} \mid \mathcal{F}_{t^-}^*\right) \\ &= \sum_{i=1}^n Z_i(\phi_n^{-1}(t)) P_{\beta(t)}(\phi_n(X_i) = t, \delta_i = 1 \mid \mathcal{F}_{t^-}^*). \end{aligned}$$

En effet, Z et ϕ_n^{-1} sont continues à gauche, donc $Z_i(\phi_n^{-1}(t)) = Z_i(\phi_n^{-1}(t^-))$ est $\mathcal{F}_{t^-}^*$ mesurable. Par définition, la probabilité pour que l'individu i décède au temps de décès t sous le modèle de paramètre $\beta(t)$ conditionnellement à l'ensemble des individus à risque au temps t est $P_{\beta(t)}(\phi_n(X_i) = t, \delta_i = 1 \mid \mathcal{F}_{t^-}^*) = \pi_i(\beta(t), t)$. Ainsi,

$$E_{\beta(t)}(\mathcal{Z}(t) \mid \mathcal{F}_{t^-}^*) = \mathcal{E}_{\beta(t)}(Z \mid t).$$

Le même raisonnement pour le moment d'ordre 2 implique le résultat pour la variance. \square

En résumé, pour $j = 1, \dots, k_n$, au temps de décès t_j , en conditionnant sur les ensembles d'individus à risque, l'espérance de $\mathcal{Z}(t_j)$ est $\mathcal{E}_{\beta(t_j)}(Z \mid t_j)$ et sa matrice de variance-covariance est $\mathcal{V}_{\beta(t_j)}(Z \mid t_j)$. La connaissance de ces moments permet de définir le processus du score standardisé que nous étudions séparément dans les cas univarié ($p = 1$) et multivarié ($p > 1$).

3.3 Processus du score standardisé univarié

Dans cette section, nous nous plaçons dans le cadre univarié ($p = 1$), c'est-à-dire qu'on ne considère qu'une covariable $Z(t)$ dans le modèle.

3.3.1 Définition du processus

L'inférence sur le coefficient de régression β dans le modèle à risques proportionnels est faite en se plaçant à la fin de l'expérience et en considérant tous les temps de décès à la fois.

Ici, nous allons plutôt adopter un point de vue séquentiel en conditionnant sur les temps de décès et les ensembles des individus à risque, comme l'a proposé Cox (1972). L'individu i qui décède en t_j est tiré aléatoirement dans l'ensemble $\{i : Y_i^*(t_j) = 1; i = 1, \dots, n\}$ des individus à risque de décéder en t_j avec une probabilité $\pi_i(\beta(t_j), t_j)$. Avec ces deux conditionnements, au temps de décès t_j , les informations nécessaires au calcul de l'espérance et de la variance de $\mathcal{Z}(t_j)$ par rapport à la loi $\{\pi_i(\beta(t_j), t_j); i = 1, \dots, n\}$ sont disponibles, pour $j = 1, \dots, k_n$. Par définition de $\widehat{k}_n = \widehat{k}_n(\beta)$ et comme, par hypothèse, $k_n \leq \widehat{k}_n$ presque-sûrement (équation (3.5)), les variances conditionnelles $\mathcal{V}_{\beta(t)}(\mathcal{Z} | t)$ évaluées pour les k_n premiers temps de décès t sont strictement supérieures à $C_{\mathcal{V}}$. Ainsi,

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \mathcal{V}_{\beta(t)}(\mathcal{Z} | t) > C_{\mathcal{V}} \quad \text{p.s.}$$

Les variables $\mathcal{Z}(t_j)$ peuvent être centrées et réduites. En sommant séquentiellement toutes ces variables centrées réduites, on définit le du score standardisé.

Définition 3.3.1. Soit $j = 0, 1, \dots, k_n$. Le score standardisé évalué au paramètre $\beta(t_j)$ et au temps t_j est défini par

$$U_n^*(\beta(t_j), t_j) = \frac{1}{\sqrt{k_n}} \int_0^{t_j} \mathcal{V}_{\beta(s)}(\mathcal{Z} | s)^{-1/2} \left\{ \mathcal{Z}(s) - \mathcal{E}_{\beta(s)}(\mathcal{Z} | s) \right\} d\bar{N}^*(s). \quad (3.10)$$

Remarque 3.3.1. Pour $j = 1, \dots, k_n$, la variable aléatoire $U_n^*(\beta(t_j), t_j)$ peut également s'écrire comme une somme :

$$U_n^*(\beta(t_j), t_j) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^j \mathcal{V}_{\beta(t_i)}(\mathcal{Z} | t_i)^{-1/2} \left\{ \mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(\mathcal{Z} | t_i) \right\}. \quad (3.11)$$

Remarque 3.3.2. Soit w une fonction déterministe ou aléatoire et $(\mathcal{F}_t^*)_{t \in [0,1]}$ -prévisible de $[0, 1]$ dans \mathbb{R} . D'après la Proposition 3.2.2, sous le modèle à risques non proportionnels de paramètre $\beta(t)$, conditionnellement à l'ensemble des individus à risque au temps de décès t_j , l'espérance de $w(t_j)\mathcal{Z}(t_j)$ est $w(t_j)\mathcal{E}_{\beta(t_j)}(\mathcal{Z} | t_j)$ et sa variance est $w(t_j)^2\mathcal{V}_{\beta(t_j)}(\mathcal{Z} | t_j)$ par rapport à la famille de probabilités $\{\pi_i(\beta(t_j), t_j); i = 1, \dots, n\}$. En sommant les variables pondérées centrées et réduites, on obtient la même définition du processus qu'en formule (3.10). Nous reviendrons sur cette remarque en Section 4.2.1 à propos du test du log-rank pondéré.

Nous pouvons maintenant définir le processus U_n^* sur $[0, 1]$ en interpolant linéairement les $k_n + 1$ variables aléatoires $\{U_n^*(\beta(t_j), t_j), j = 0, 1, \dots, k_n\}$.

Définition 3.3.2. Le processus du score standardisé évalué au paramètre $\beta(t)$ est défini par $\{U_n^*(\beta(t), t), t \in [0, 1]\}$ où, pour $j = 0, \dots, k_n$ et $t \in [t_j, t_{j+1}[$,

$$U_n^*(\beta(t), t) = U_n^*(\beta(t_j), t_j) + (tk_n - j) \{U_n^*(\beta(t_{j+1}), t_{j+1}) - U_n^*(\beta(t_j), t_j)\}.$$

Par définition, le processus $U_n^*(\beta(\cdot), \cdot)$ est continu sur l'intervalle $[0, 1]$. Le processus dépend

de 2 paramètres, le temps t et le coefficient de régression $\beta(t)$. Par souci de clarté, nous rappelons que la fonction temporelle $\beta : t \rightarrow \beta(t)$ est notée $\beta(t)$, ce qui permettra de distinguer plus facilement dans la suite le modèle à risques proportionnels de paramètre β du modèle à risques non proportionnels de paramètre $\beta(t)$. Sous le modèle à risques non proportionnels de paramètre $\beta(t)$, les incréments du processus U_n^* sont centrés et de variance égale à l'unité. De plus, les incréments sont non corrélés, comme le montre la proposition suivante.

Proposition 3.3.1 (Cox, 1975). *Sous le modèle à risques non proportionnels de paramètre $\beta(t)$, les variables aléatoires $U_n^*(\beta(t_j), t_j)$ pour $j = 1, 2, \dots, k_n$ ne sont pas corrélées.*

Preuve. Soit $i, j = 1, 2, \dots, k_n$, $i < j$. Par définition, la variable aléatoire

$$\mathcal{V}_{\beta(t_i)}(Z | t_i)^{-1/2} \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right),$$

est $\mathcal{F}_{t_i}^*$ mesurable. L'inclusion des σ -algèbres $\mathcal{F}_{t_i}^* \subset \mathcal{F}_{t_j}^*$ implique que cette variable aléatoire est également $\mathcal{F}_{t_j}^*$ mesurable. De plus, d'après la Proposition 3.2.2,

$$\begin{aligned} & E \left(\mathcal{V}_{\beta(t_j)}(Z | t_j)^{-1/2} \left(\mathcal{Z}(t_j) - \mathcal{E}_{\beta(t_j)}(Z | t_j) \right) \middle| \mathcal{F}_{t_j}^* \right) \\ &= \mathcal{V}_{\beta(t_j)}(Z | t_j)^{-1/2} E \left(\mathcal{Z}(t_j) - \mathcal{E}_{\beta(t_j)}(Z | t_j) \middle| \mathcal{F}_{t_j}^* \right) = 0. \end{aligned}$$

Ainsi,

$$\begin{aligned} & E \left(\mathcal{V}_{\beta(t_i)}(Z | t_i)^{-1/2} \left\{ \mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right\} \mathcal{V}_{\beta(t_j)}(Z | t_j)^{-1/2} \left\{ \mathcal{Z}(t_j) - \mathcal{E}_{\beta(t_j)}(Z | t_j) \right\} \right) \\ &= E \left[E \left(\mathcal{V}_{\beta(t_i)}(Z | t_i)^{-1/2} \left\{ \mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right\} \mathcal{V}_{\beta(t_j)}(Z | t_j)^{-1/2} \left\{ \mathcal{Z}(t_j) - \mathcal{E}_{\beta(t_j)}(Z | t_j) \right\} \middle| \mathcal{F}_{t_j}^* \right) \right] \\ &= E \left[\mathcal{V}_{\beta(t_i)}(Z | t_i)^{-1/2} \left\{ \mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right\} E \left(\mathcal{V}_{\beta(t_j)}(Z | t_j)^{-1/2} \left\{ \mathcal{Z}(t_j) - \mathcal{E}_{\beta(t_j)}(Z | t_j) \right\} \middle| \mathcal{F}_{t_j}^* \right) \right] \\ &= 0. \end{aligned}$$

□

La propriété de non-corrélation des incréments est également utilisée dans le calcul de la statistique du log-rank (voir Section 1.3 du chapitre introductif). Toutes ces propriétés permettent d'obtenir la convergence du processus U_n^* .

3.3.2 Résultats asymptotiques

L'espace $(D[0, 1], \mathbb{R})$ est muni de la topologie de la convergence uniforme. Définissons des hypothèses suffisantes pour étudier le comportement asymptotique du processus. Soit

$$S^{(r)} \{ \beta(t), t \} = \frac{1}{n} \sum_{i=1}^n Y_i^*(t) Z_i \left(\phi_n^{-1}(t) \right)^r \exp \left(\beta(t) Z_i \left(\phi_n^{-1}(t) \right) \right), \quad 0 \leq t \leq 1, \quad r = 0, 1, 2.$$

Remarquons que pour $t \in \{t_1, t_2, \dots, t_{k_n}\}$,

$$\mathcal{E}_{\beta(t)}(Z | t) = \left. \frac{\partial}{\partial b} \log \{S^{(0)}(b, t)\} \right|_{b=\beta(t)} = \frac{S^{(1)}(\beta(t), t)}{S^{(0)}(\beta(t), t)},$$

et

$$\mathcal{V}_{\beta(t)}(Z | t) = \left. \frac{\partial^2}{\partial b^2} \log \{S^{(0)}(b, t)\} \right|_{b=\beta(t)} = \frac{S^{(2)}(\beta(t), t)}{S^{(0)}(\beta(t), t)} - \left(\frac{S^{(1)}(\beta(t), t)}{S^{(0)}(\beta(t), t)} \right)^2.$$

Plaçons-nous sous les conditions suivantes :

H1. (Stabilité asymptotique). Il existe $\delta_1 > 0$, un voisinage de β de rayon δ_1 contenant la fonction nulle, noté $\mathbb{B} = \{\gamma, \sup_{t \in [0,1]} |\gamma(t) - \beta(t)| < \delta_1\}$, et des fonctions $s^{(r)}$ définies sur $\mathbb{B} \times [0, 1]$ pour $r = 0, 1, 2$, telles que

$$\sqrt{n} \sup_{t \in [0,1], \gamma \in \mathbb{B}} |S^{(r)}(\gamma(t), t) - s^{(r)}(\gamma(t), t)| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.12)$$

H2. (Régularité asymptotique). Les fonctions déterministes $s^{(r)}$, $r = 0, 1, 2$, définies en **H1.** sont uniformément continues en $t \in [0, 1]$ et bornées sur $\mathbb{B} \times [0, 1]$. De plus, pour $r = 0, 1, 2$, et $t \in [0, 1]$, $s^{(r)}(\cdot, t)$ est une fonction continue sur \mathbb{B} . La fonction $s^{(0)}$ est minorée par une constante strictement positive.

Par analogie avec les quantités empiriques, pour $t \in [0, 1]$ et $\gamma \in \mathbb{B}$, notons

$$e(\gamma(t), t) = \frac{s^{(1)}(\gamma(t), t)}{s^{(0)}(\gamma(t), t)},$$

et

$$v(\gamma(t), t) = \frac{s^{(2)}(\gamma(t), t)}{s^{(0)}(\gamma(t), t)} - e(\gamma(t), t)^2. \quad (3.13)$$

H3. (Homoscedasticité). Pour tout $t \in [0, 1]$ et $\gamma \in \mathbb{B}$, on a $\frac{\partial}{\partial t} v(\gamma(t), t) = 0$.

H4. (Covariables bornées uniformément). Il existe $L \in \mathbb{R}^{*+}$ tel que

$$\sup_{i=1, \dots, n} \sup_{u \in [0, T]} |Z_i(u)| \leq L.$$

Les deux premières conditions sont classiques en analyse de survie. Elles ont été introduites par [Andersen et Gill \(1982\)](#) pour utiliser la théorie des processus de comptage et la théorie des martingales, comme l'inégalité de [Lenglart \(1977\)](#) ou le théorème de [Rebolledo \(1980\)](#). La variance $\mathcal{V}_{\beta(t)}(Z | t)$ est, par définition, un estimateur de la variance de Z sachant $T = t$ sous le modèle à risques non proportionnels de paramètre $\beta(t)$. Ainsi, la condition **H3.** d'homoscedasticité signifie que la variance asymptotique ne dépend pas du temps, pour des paramètres proches du vrai coefficient de régression $\beta(t)$. Cette condition est souvent rencontrée dans l'utilisation du modèle à risques proportionnels de manière implicite. C'est le cas par

exemple dans l'estimation de la variance du paramètre β ou dans l'expression de la statistique du log-rank où la contribution à la variance globale est la même à chaque temps de décès. En effet, la variance globale est une somme non pondérée des variances conditionnelles. La stabilité temporelle de la variance a été soulignée par plusieurs auteurs, notamment Grambsch et Therneau (1994), Xu (1996) et Xu et O'Quigley (2000).

Théorème 3.3.1. *Sous le modèle à risques non proportionnels de paramètre $\beta(t)$,*

$$U_n^*(\beta(\cdot), \cdot) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{W},$$

où \mathcal{W} est un mouvement Brownien standard.

La preuve, donnée en Section 3.5.2, repose sur l'utilisation du théorème central limite fonctionnel pour des différences de martingales introduit par Helland (1982). Ce théorème peut être vu comme étant une extension du théorème central limite fonctionnel de Donsker (1951) au cas de variables centrées, réduites, mais non nécessairement indépendantes et de lois différentes. La Figure 3.1(a) représente le résultat d'une simulation du processus du score standardisé $U_n^*(0, t)$ en fonction du temps t , lorsque le vrai paramètre est $\beta(t) = 0$. L'approximation du processus par un mouvement Brownien semble être convenable. Nous étudierons comment évaluer cette approximation sur un échantillon particulier aux Chapitres 4 et 5. Nous rappelons que le paramètre k_n est inconnu dans les applications et est remplacé par son estimateur consistant \widehat{k}_n .

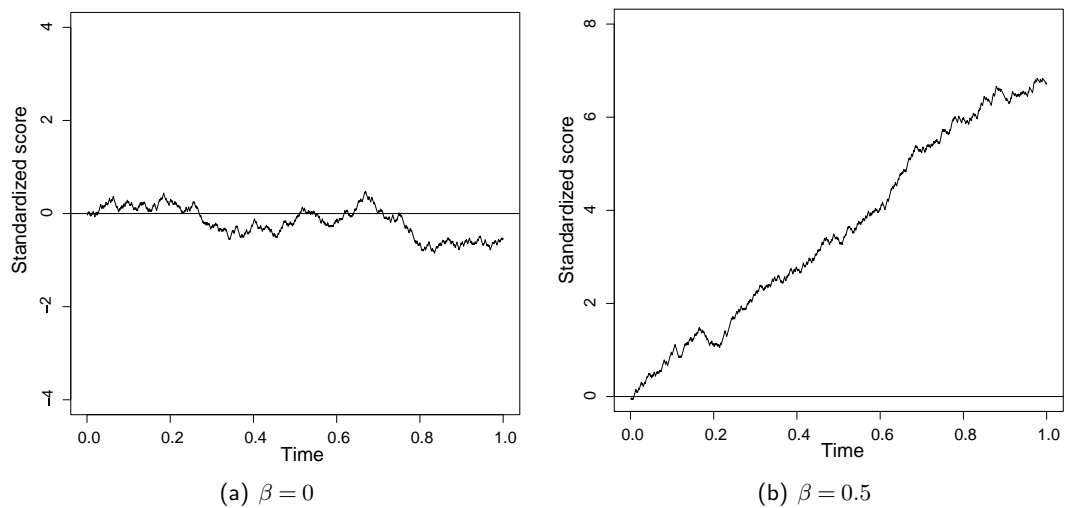


FIGURE 3.1 – Processus $U_n^*(0, \cdot)$ en fonction du temps issus de données simulées sous le modèle à risques proportionnels (1.4) avec différents paramètres β .

Nous allons maintenant nous intéresser au comportement asymptotique du processus du score standardisé évalué en un paramètre différent de $\beta(t)$ qui est la vraie fonction de régression. En pratique, c'est cette situation que nous rencontrons étant donné que la valeur de $\beta(t)$ n'est

pas connue. Pour tout $\gamma \in \mathbb{B}$, on notera $U_n^*(\gamma, \cdot)$ le processus du score standardisé évalué au paramètre γ . Pour construire ce processus de paramètre γ , on a considéré $k_n = k_n(\gamma)$ qui est estimé par $\widehat{k}_n(\gamma)$ et qui vérifie la relation presque sûre $k_n(\gamma) \leq \widehat{k}_n(\gamma)$ (équation (3.5)). Par définition de $\widehat{k}_n(\gamma)$, nous avons alors

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \mathcal{V}_{\gamma(t)}(Z | t) > C_{\mathcal{V}} \quad \text{p.s.} \quad (3.14)$$

Théorème 3.3.2. *Soit β_0 une fonction constante de \mathbb{B} . Sous le modèle à risques non proportionnels de paramètre $\beta(t)$, on considère*

$$A_n(t) = \frac{1}{k_n} \int_0^t \{\beta(s) - \beta_0\} \frac{\mathcal{V}_{\tilde{\beta}(s)}(Z | s)}{\mathcal{V}_{\beta_0}(Z | s)^{1/2}} d\bar{N}^*(s), \quad 0 \leq t \leq 1, \quad (3.15)$$

où $\tilde{\beta}$ est dans la boule de centre β et de rayon $\sup_{t \in [0,1]} |\beta(t) - \beta_0|$. Alors il existe deux constantes $C_1(\beta, \beta_0)$ et $C_2 \in \mathbb{R}^{*+}$ telles que $C_1(\beta, \beta) = 1$ et

$$U_n^*(\beta_0, \cdot) - \sqrt{k_n} A_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} C_1(\beta, \beta_0) \mathcal{W}. \quad (3.16)$$

De plus,

$$\sup_{0 \leq t \leq 1} \left| A_n(t) - C_2 \int_0^t \{\beta(s) - \beta_0\} ds \right| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.17)$$

Remarque 3.3.3. *Le Théorème 3.3.1 est un cas particulier du Théorème 3.3.2 dans le cas où $\beta(t) = \beta_0$. En effet, dans le Théorème 3.3.2, lorsque $\beta(t) = \beta_0$, la constante $C_1(\beta, \beta) = 1$, la dérive s'annule et le processus limite de $U_n^*(\beta_0, \cdot)$ est un mouvement Brownien standard.*

Remarque 3.3.4. *L'équation (3.14) appliquée en $\gamma = \beta_0$ et $k_n = k_n(\beta_0)$ devient*

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \mathcal{V}_{\beta_0}(Z | t) > C_{\mathcal{V}} \quad \text{p.s.} \quad (3.18)$$

La démonstration du théorème se trouve en Section 3.5.2. La dérive du processus du score standardisé au temps t tend vers l'infini lorsque $n \rightarrow \infty$ d'après l'équation (3.16). En pratique, nous travaillons à n fixé et assez grand et nous considérons $\beta_0 = 0$. Le théorème indique que l'allure de la dérive du processus $U_n^*(0, \cdot)$ est proportionnelle au coefficient cumulé $\int \beta(s) ds$. Distinguons deux cas dans lesquels le coefficient de régression $\beta(t)$ est non nul. Tout d'abord, plaçons-nous sous le modèle à risques proportionnels (1.4), c'est-à-dire que $\beta(t)$ est constant au cours du temps. D'après le Théorème 3.3.2, le processus $U_n^*(0, \cdot)$ peut être approché par un mouvement Brownien avec une dérive linéaire. Une illustration est donnée en Figure 3.1(b) dans laquelle un processus a été simulé sous le modèle à risques proportionnels de paramètre $\beta = 0.5$. La dérive peut être approchée par une droite, ce qui indique que le coefficient de régression est constant dans le temps. Le coefficient β , du même signe que la pente de la droite d'après le Théorème 3.3.2, est positif.

Le second cas correspond à la situation d'un coefficient de régression non constant au cours du temps. Plusieurs possibilités sont envisageables. Par exemple, l'effet peut être constant puis

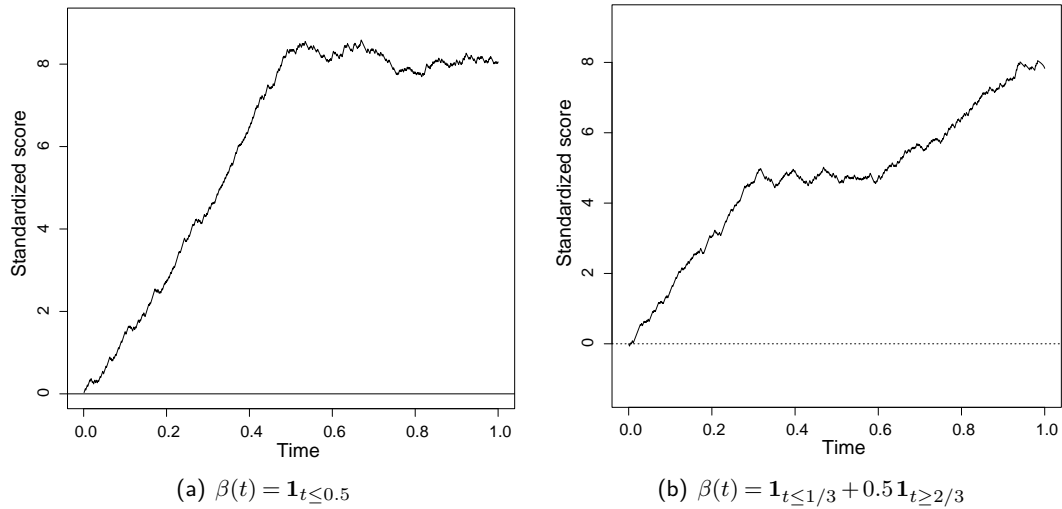


FIGURE 3.2 – Processus $U_n^*(0, \cdot)$ en fonction du temps issus de données simulées sous le modèle à risques non proportionnels avec différents paramètres $\beta(t)$.

décroître avec un certain temps τ , l'effet peut être constant par morceaux au cours du temps, augmenter ou diminuer de manière continue, etc. Dans toutes ces situations, l'effet cumulé $\int \beta(s) ds$ sera reflété dans la dérive du processus (Théorème 3.3.2). Par exemple, la Figure 3.2(a) représente un processus simulé avec un effet constant par morceaux $\beta(t) = \mathbf{1}_{t \leq 0.5}$. Avant $t = 0.5$, on observe une tendance linéaire positive correspondant à $\beta = 1$, et, pour $t > 0.5$, le coefficient s'annule et la dérive du processus $U_n^*(0, \cdot)$ est parallèle à l'axe du temps. La Figure 3.2(b) représente le processus du score standardisé évalué pour un jeu de données simulé sous un modèle de rupture avec un paramètre de régression $\beta(t) = \mathbf{1}_{t \leq 1/3} + 0.5 \mathbf{1}_{t \geq 2/3}$. La dérive du processus peut être séparée en trois parties linéaires reflétant la force de l'effet, chaque partie pouvant être approchée par une droite. La pente de la première droite semble être deux fois plus élevée que la pente de la dernière droite et la pente de la deuxième droite est nulle.

Finalement, que l'effet corresponde à des risques proportionnels ou non proportionnels, toute l'information concernant l'effet de régression $\beta(t)$ est captée dans le processus $U_n^*(0, \cdot)$. Nous pouvons alors utiliser ce processus pour tester la valeur du coefficient de régression et évaluer la qualité de l'adéquation du modèle à risques proportionnels. Avant de développer ces aspects dans les Chapitres 4 et 5, définissons le processus du score standardisé dans le cadre multivarié.

3.4 Processus du score standardisé multivarié

Dans cette section, nous étendons le processus du score standardisé de la section précédente au cas de plusieurs covariables regroupées dans le vecteur $Z(t)$ de dimension p , avec $p > 1$.

3.4.1 Définition du processus

Supposons le modèle à risques non proportionnels de fonction de régression $\beta(t)$ et de vecteur de covariables $Z(t)$ qui sont des fonctions de $[0, 1]$ à valeurs dans \mathbb{R}^p . Soit $t \in [0, 1]$ et

$$\begin{aligned} S^{(0)}(\beta(t), t) &= n^{-1} \sum_{i=1}^n Y_i^*(t) \exp\left(\beta(t)^T Z_i(\phi_n^{-1}(t))\right) \in \mathbb{R}, \\ S^{(1)}(\beta(t), t) &= n^{-1} \sum_{i=1}^n Y_i^*(t) Z_i(\phi_n^{-1}(t)) \exp\left(\beta(t)^T Z_i(\phi_n^{-1}(t))\right) \in \mathbb{R}^p, \\ S^{(2)}(\beta(t), t) &= n^{-1} \sum_{i=1}^n Y_i^*(t) Z_i(\phi_n^{-1}(t))^{\otimes 2} \exp\left(\beta(t)^T Z_i(\phi_n^{-1}(t))\right) \in \mathcal{M}_{p \times p}(\mathbb{R}). \end{aligned}$$

Pour $t \in \{t_1, \dots, t_{k_n}\}$, l'espérance conditionnelle $\mathcal{E}_{\beta(t)}(Z | t)$ et la matrice de variance-covariance conditionnelle $\mathcal{V}_{\beta(t)}(Z | t)$ définies en (3.8) et (3.9) peuvent être écrites en fonction des quantités $S^{(r)}(\beta(t), t)$ ($r = 0, 1, 2$) par les relations

$$\begin{aligned} \mathcal{E}_{\beta(t)}(Z | t) &= \frac{S^{(1)}(\beta(t), t)}{S^{(0)}(\beta(t), t)}, \\ \mathcal{V}_{\beta(t)}(Z | t) &= \frac{S^{(2)}(\beta(t), t)}{S^{(0)}(\beta(t), t)} - \mathcal{E}_{\beta(t)}(Z | t)^{\otimes 2} = \frac{\partial}{\partial \beta} \mathcal{E}_{\beta(t)}(Z | t). \end{aligned}$$

La matrice $\mathcal{V}_{\beta(t)}(Z | t)$ étant symétrique et définie positive, il existe une matrice orthogonale $P_{\beta(t)}(t)$ et une matrice diagonale $D_{\beta(t)}(t)$ telles que

$$\mathcal{V}_{\beta(t)}(Z | t) = P_{\beta(t)}(t) D_{\beta(t)}(t) P_{\beta(t)}(t)^T.$$

Cela mène à définir la matrice symétrique $\mathcal{V}_{\beta(t)}(Z | t)^{-1/2}$ par

$$\mathcal{V}_{\beta(t)}(Z | t)^{-1/2} = P_{\beta(t)}(t) D_{\beta(t)}(t)^{-1/2} P_{\beta(t)}(t)^T.$$

Par définition de $\widehat{k}_n = \widehat{k}_n(\beta)$ pour le cas multivarié et comme $k_n \leq \widehat{k}_n$ presque-sûrement (équation (3.5)), les matrices de variance-covariances conditionnelles $\mathcal{V}_{\beta(t)}(Z | t)$ évaluées pour les k_n premiers temps de décès t sont définies positives et

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \left\| \mathcal{V}_{\beta(t)}(Z | t)^{-1} \right\| \leq C_V^{-1} \quad \text{p.s.}$$

Définition 3.4.1. Soit $j = 0, 1, \dots, k_n$. Le score standardisé multivarié évalué au temps t_j et au paramètre $\beta(t_j)$ est défini par

$$U_n^*(\beta(t_j), t_j) = \frac{1}{\sqrt{k_n}} \int_0^{t_j} \mathcal{V}_{\beta(s)}(Z | s)^{-1/2} \left\{ Z(s) - \mathcal{E}_{\beta(s)}(Z | s) \right\} d\bar{N}^*(s),$$

où

$$\int_0^{t_j} a(s) d\bar{N}^*(s) = \left(\int_0^{t_j} a_1(s) d\bar{N}^*(s), \dots, \int_0^{t_j} a_p(s) d\bar{N}^*(s) \right),$$

pour $a = (a_1, \dots, a_p)$ où a_i est une fonction de $[0, 1]$ dans \mathbb{R} , $i = 1, \dots, p$.

Remarque 3.4.1. Comme dans le cas univarié, la variable aléatoire $U_n^*(\beta(t_j), t_j)$ peut également s'écrire comme une somme :

$$U_n^*(\beta(t_j), t_j) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^j \mathcal{V}_{\beta(t_i)}(Z | t_i)^{-1/2} \left\{ \mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(\mathcal{Z} | t_i) \right\}.$$

Le score standardisé est une fonction aléatoire définie aux temps t_1, \dots, t_{k_n} et à valeurs dans \mathbb{R}^p , dont nous pouvons étendre la définition sur $[0, 1]$ par une interpolation linéaire.

Définition 3.4.2. Le processus du score standardisé multivarié $\{U_n^*(\beta(t), t), t \in [0, 1]\}$ évalué au paramètre $\beta(t)$ est tel que, pour $j = 0, 1, \dots, k_n$ et $t \in [t_j, t_{j+1}[$,

$$U_n^*(\beta(t), t) = U_n^*(\beta(t_j), t_j) + (tk_n - j) \{U_n^*(\beta(t_{j+1}), t_{j+1}) - U_n^*(\beta(t_j), t_j)\}.$$

Remarque 3.4.2. Une autre définition du processus du score standardisé multivarié pourrait être de regrouper les processus univariés associés à chaque covariable dans un même vecteur. Cependant, la Définition 3.4.2 est plus intéressante, car elle permet de prendre en compte des covariances entre les covariables.

3.4.2 Résultats asymptotiques

Nous rappelons que les normes des vecteurs de \mathbb{R}^p et matrices de $\mathcal{M}_{p \times p}(\mathbb{R})$ sont leurs normes sup, définies en Section 2.3. L'espace des fonctions $(D[0, 1], \mathbb{R}^p)$ est muni de la topologie produit de Skorokhod. Pour établir le comportement limite du processus, nous nous plaçons sous les conditions suivantes :

H1'. (Stabilité asymptotique). Il existe $\delta_2 > 0$, un voisinage borné \mathbb{B}' de β de rayon δ_2 contenant la fonction nulle, noté $\mathbb{B}' = \left\{ \gamma, \sup_{t \in [0, 1]} \|\gamma(t) - \beta(t)\| < \delta_2 \right\}$, et des fonctions $s^{(r)}$ définies sur $\mathbb{B}' \times [0, 1]$ pour $r = 0, 1$ et 2 et à valeurs respectivement dans \mathbb{R} , \mathbb{R}^p et l'espace des matrices $\mathcal{M}_{p \times p}(\mathbb{R})$, telles que

$$\sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}'} \left\| S^{(r)}(\gamma(t), t) - s^{(r)}(\gamma(t), t) \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

H2'. (Régularité asymptotique). Les fonctions déterministes $s^{(r)}$, $r = 0, 1, 2$, définies en **H1'** sont uniformément continues en $t \in [0, 1]$ et bornées sur $\mathbb{B}' \times [0, 1]$. De plus, pour $r = 0, 1, 2$ et $t \in [0, 1]$, $s^{(r)}(\cdot, t)$ est une fonction continue sur \mathbb{B}' . La fonction $s^{(0)}$ est minorée par une constante strictement positive.

H3'. (Homoscedasticité). Il existe une matrice symétrique et définie positive $\Sigma \in \mathcal{M}_{p \times p}(\mathbb{R})$ et une suite de constantes positives $(M_n)_n$ telle que $\lim_{n \rightarrow \infty} M_n = 0$ et qui vérifient pour

tout $\gamma \in \mathbb{B}'$, $t \in [0, 1]$ et $i, j = 1, \dots, p$,

$$\left\| \frac{\partial^2}{\partial \beta_i \partial \beta_j} \mathcal{E}_\beta(Z | t) \Big|_{\beta=\gamma(t)} \right\| \leq M_n \quad \text{p.s.}, \quad (3.19)$$

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} - \Sigma^{1/2} \right\| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.20)$$

H4'. (Covariables bornées uniformément). Il existe $L' \in \mathbb{R}^{*+}$ tel que

$$\sup_{i=1, \dots, n} \sup_{u \in [0, T]} \sup_{j=1, \dots, p} \left| Z_i^{(j)}(u) \right| \leq L'.$$

Ces conditions sont les extensions des hypothèses **H1.**, **H2.**, **H3.** et **H4.** de la Section 3.3 au cas de plusieurs covariables. Nous faisons l'hypothèse supplémentaire que les matrices de variance-covariance conditionnelles ne dépendent pas du paramètre lorsqu'il est proche du vrai paramètre de régression. Il s'agit d'une hypothèse technique due à l'utilisation de l'inégalité de Taylor-Lagrange multidimensionnelle (2.3.1) dans les preuves. Dans le cas univarié, nous n'avons pas fait cette hypothèse grâce à l'existence d'une égalité de Taylor-Lagrange.

Proposition 3.4.1. *Sous les hypothèses **H1'**, **H3'** et **H4'**, nous avons la convergence*

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t) - \Sigma \right\| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.21)$$

La proposition est démontrée en Section 3.5.3. Le théorème suivant donne le comportement limite du processus du score standardisé univarié. Soit $a = (a_1, \dots, a_p)$ une fonction de \mathbb{R}^p dans $[0, 1]$. Notons

$$\int_0^t a(s) ds = \left(\int_0^t a_1(s) ds, \dots, \int_0^t a_p(s) ds \right), \quad 0 \leq t \leq 1.$$

Soit $\gamma \in \mathbb{B}'$. On note $U_n^*(\gamma, \cdot)$ le processus du score standardisé multivarié évalué au paramètre γ . Pour construire ce processus, on a considéré $k_n = k_n(\gamma)$ qui est estimé par $\widehat{k}_n(\gamma)$. On rappelle que, $k_n(\gamma) \leq \widehat{k}_n(\gamma)$ presque-sûrement (équation (3.5)) et par définition de $\widehat{k}_n(\gamma)$ dans le cas multivarié, nous avons

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{-1} \right\| \leq C_V^{-1} \quad \text{p.s.} \quad (3.22)$$

Théorème 3.4.1. *Soit β_0 une fonction constante de \mathbb{B}' . Sous le modèle à risques non proportionnels de coefficient de régression $\beta(t)$, si le processus du score standardisé multivarié U_n^* est évalué au paramètre β_0 , nous avons la convergence en loi*

$$U_n^*(\beta_0, \cdot) - \sqrt{k_n} B_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{W}_p, \quad (3.23)$$

où \mathcal{W}_p est un mouvement Brownien standard de dimension p et, pour tout $t \in [0, 1]$,

$$B_n(t) = \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \left\{ \mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\beta_0}(Z | t_i) \right\}.$$

De plus,

$$\sup_{t \in [0, 1]} \left\| B_n(t) - \Sigma^{1/2} \int_0^t \{\beta(s) - \beta_0\} ds \right\| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.24)$$

Nous rappelons qu'un mouvement Brownien de dimension p est un vecteur de p mouvements Brownien à valeurs dans \mathbb{R} indépendants.

Remarque 3.4.3. La constante $C_1(\beta, \beta_0)$ du Théorème 3.3.2 qui établit la convergence du processus U_n^* univarié n'a pas d'équivalent dans le cas multivarié. C'est une conséquence de l'équation (3.19) de l'hypothèse **H3'**, qui est utilisée dans la preuve dans une inégalité de Taylor–Lagrange multidimensionnelle, mais qui n'est pas nécessaire dans le cadre univarié. La constante C_2 du Théorème 3.3.2 est équivalente à la matrice $\Sigma^{1/2}$.

En pratique, à n fixé et assez grand, le processus du score standardisé multivarié $U_n^*(0, \cdot)$ peut donc être approché par un mouvement Brownien avec une dérive qui dépend du vrai paramètre $\beta(t)$ du modèle. Chaque composante de la dérive $\sqrt{k_n} \Sigma^{1/2} \int_0^t \beta(s) ds$ reflète une combinaison linéaire des effets cumulés $\int_0^t \beta_1(s) ds, \dots, \int_0^t \beta_p(s) ds$, sauf si la matrice de variance-covariance Σ est diagonale. Sous les hypothèses du théorème, une conséquence immédiate de l'équation (3.21) est

$$\hat{\Sigma} := \frac{1}{k_n} \sum_{i=1}^{k_n} \mathcal{V}_{\beta_0}(Z | t_i) \xrightarrow[n \rightarrow \infty]{P} \Sigma, \quad (3.25)$$

ce qui signifie que $\hat{\Sigma}$ est un estimateur convergeant de Σ . Le corollaire suivant permet alors de décorréler les composantes du processus U_n^* afin que les dérivées des composantes reflètent les effets $\beta_1(t), \dots, \beta_p(t)$ séparément.

Corollaire 3.4.1. Sous les conditions du Théorème 3.4.1,

$$\hat{\Sigma}^{-1/2} U_n^*(\beta_0, \cdot) - \sqrt{k_n} \hat{\Sigma}^{-1/2} B_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Sigma^{-1/2} \mathcal{W}_p,$$

et

$$\sup_{t \in [0, 1]} \left\| \hat{\Sigma}^{-1/2} B_n(t) - \int_0^t \{\beta(s) - \beta_0\} ds \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Ainsi, en présence de plusieurs covariables, les dérivées du processus $\hat{\Sigma}^{-1/2} U_n^*(\beta_0, \cdot)$ permettent de faire de l'inférence sur les coefficients de régression. Chaque dérive représente un coefficient cumulé et l'inférence se fait séparément pour chaque dérive comme dans le cas univarié. Ce processus sera utile au Chapitre 5 pour construire le modèle à risques non proportionnels. Nous l'illustrerons sur des simulations.

3.5 Preuves

3.5.1 Preuves de la Section 3.2

Preuve du Lemme 3.2.1. Soit $\varepsilon > 0$. Nous avons

$$\begin{aligned} & P \left\{ \sup_{0 \leq s \leq 1} \left| \frac{1}{k_n} \int_0^s A_n(t) d\bar{N}^*(t) - \int_0^s a(t) dt \right| > \varepsilon \right\} \\ & \leq P \left\{ \sup_{0 \leq s \leq 1} \frac{1}{k_n} \left| \int_0^s \{A_n(t) - a(t)\} d\bar{N}^*(t) \right| > \frac{\varepsilon}{2} \right\} \\ & \quad + P \left\{ \sup_{0 \leq s \leq 1} \left| \frac{1}{k_n} \int_0^s a(t) d\bar{N}^*(t) - \int_0^s a(t) dt \right| > \frac{\varepsilon}{2} \right\} \end{aligned} \quad (3.26)$$

Étudions les deux termes de droite séparément. Soit $s \in [0, 1]$. Pour le premier terme, nous avons

$$\begin{aligned} \frac{1}{k_n} \left| \int_0^s \{A_n(t) - a(t)\} d\bar{N}^*(t) \right| &= \frac{1}{k_n} \left| \sum_{i=1}^{\lfloor k_n s \rfloor} (A_n(t_i) - a(t_i)) \right| \\ &\leq \frac{\lfloor k_n s \rfloor}{k_n} \sup_{i=1, \dots, \lfloor k_n s \rfloor} |A_n(t_i) - a(t_i)| \\ &\leq \sup_{i=1, \dots, k_n} |A_n(t_i) - a(t_i)|. \end{aligned}$$

Donc

$$\sup_{0 \leq s \leq 1} \frac{1}{k_n} \left| \int_0^s \{A_n(t) - a(t)\} d\bar{N}^*(t) \right| \leq \sup_{i=1, \dots, k_n} |A_n(t_i) - a(t_i)|,$$

et, par hypothèse, $\sup_{i=1, \dots, k_n} |A_n(t_i) - a(t_i)| \xrightarrow[n \rightarrow \infty]{P} 0$. Par conséquent,

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{0 \leq s \leq 1} \frac{1}{k_n} \left| \int_0^s \{A_n(t) - a(t)\} d\bar{N}^*(t) \right| > \frac{\varepsilon}{2} \right\} = 0.$$

Pour le second terme de droite de l'équation (3.26), nous avons

$$\sup_{0 \leq s \leq 1} \left| \frac{1}{k_n} \int_0^s a(t) d\bar{N}^*(t) - \int_0^s a(t) dt \right| = \sup_{0 \leq s \leq 1} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor k_n s \rfloor} a(t_i) - \int_0^s a(t) dt \right|.$$

On note a_n la fonction constante par morceaux définie sur $[0, 1]$ telle que $a_n(t) = a(t_i)$ pour $t \in [t_i, t_{i+1}[$, $i = 0, \dots, k_n$. On a alors

$$\frac{1}{k_n} \sum_{i=1}^{\lfloor k_n s \rfloor} a(t_i) = \int_0^s a_n(t) dt - \left(s - \frac{\lfloor k_n s \rfloor}{k_n} \right) a \left(\frac{\lfloor k_n s \rfloor}{k_n} \right).$$

Par conséquent,

$$\begin{aligned} & \sup_{0 \leq s \leq 1} \left| \frac{1}{k_n} \int_0^s a(t) d\bar{N}^*(t) - \int_0^s a(t) dt \right| \\ & \leq \sup_{0 \leq s \leq 1} \left| \int_0^s a_n(t) dt - \int_0^s a(t) dt \right| + \sup_{0 \leq s \leq 1} \left| \left(s - \frac{\lfloor k_n s \rfloor}{k_n} \right) a \left(\frac{\lfloor k_n s \rfloor}{k_n} \right) \right| \\ & \leq \sup_{0 \leq s \leq 1} \left| \int_0^s a_n(t) dt - \int_0^s a(t) dt \right| + \frac{1}{k_n} \sup_{0 \leq s \leq 1} \left| a \left(\frac{\lfloor k_n s \rfloor}{k_n} \right) \right|, \end{aligned}$$

Remarquons que $\lim_{n \rightarrow \infty} \sup_{s \in [0,1]} |a_n(s) - a(s)| = 0$ par uniforme continuité de la fonction a . Ceci implique la convergence uniforme

$$\lim_{n \rightarrow \infty} \sup_{s \in [0,1]} \left| \int_0^s a_n(t) dt - \int_0^s a(t) dt \right| = 0.$$

De plus, a est bornée donc $\sup_{0 \leq s \leq 1} \left| a \left(\frac{\lfloor k_n s \rfloor}{k_n} \right) \right| < \infty$. En conclusion,

$$\lim_{n \rightarrow \infty} \sup_{0 \leq s \leq 1} \left| \frac{1}{k_n} \int_0^s a(t) d\bar{N}^*(t) - \int_0^s a(t) dt \right| = 0, \quad (3.27)$$

ce qui termine la démonstration. □

3.5.2 Preuves de la Section 3.3

Lemmes préliminaires

Avant de donner la preuve des Théorèmes 3.3.1 et 3.3.2, nous fournissons quelques résultats intermédiaires. Le premier lemme indique que sous les conditions des Théorèmes 3.3.1 et 3.3.2, la variance asymptotique v définie à l'équation (3.13) est strictement positive.

Lemme 3.5.1. *Sous les conditions **H1.** et **H3.**, pour tout $t \in [0, 1]$ et $\gamma \in \mathbb{B}$, $v(\gamma(t), t) > 0$.*

Preuve. D'après la loi des grands nombres, la condition **H1.** implique que pour $r = 0, 1, 2$, on a

$$\begin{aligned} s^{(r)}(\gamma(0), 0) &= E \left(S^{(r)}(\gamma(0), 0) \right) \\ &= E \{ Y(0) Z(0)^r \exp(\gamma(0) Z(0)) \} \\ &= E \{ Z(0)^r \exp(\gamma(0) Z(0)) \}, \end{aligned}$$

car $Y(0) = \mathbf{1}_{X \geq 0} = 1$. Ainsi, par définition de v ,

$$\begin{aligned} v(\gamma(0), 0) &= \frac{E \{ Z(0)^2 \exp(\gamma(0) Z(0)) \}}{E \{ \exp(\gamma(0) Z(0)) \}} - \left(\frac{E \{ Z(0) \exp(\gamma(0) Z(0)) \}}{E \{ \exp(\gamma(0) Z(0)) \}} \right)^2 \\ &= E \{ Z(0)^2 \Phi_{\gamma(0)}(Z(0)) \} - \left(E \{ Z(0) \Phi_{\gamma(0)}(Z(0)) \} \right)^2, \end{aligned}$$

où

$$\Phi_{\gamma(0)}(z) = \frac{\exp(\gamma(0)z)}{E(\exp(\gamma(0)Z(0)))}, \quad z \in \mathbb{R}.$$

Ainsi, en notant $P_{Z(0)}$ la fonction de répartition de la variable aléatoire $Z(0)$ et X la variable aléatoire de fonction de répartition P_X telle que $dP_X(x) = \Phi_{\gamma(0)}(x)dP_{Z(0)}(x)$, $x \in \mathbb{R}$, on a

$$v(\gamma(0), 0) = E(X^2) - E(X)^2 = V(X).$$

La fonction $\Phi_{\gamma(0)}$ étant inversible sur \mathbb{R} , X est distribuée selon une loi de Dirac si et seulement si $Z(0)$ suit une loi de Dirac. Par hypothèse, $V(Z(0)) > 0$ ce qui équivaut à dire que $Z(0)$ ne suit pas une distribution de Dirac. Donc X ne suit pas une loi de Dirac et $v(\gamma(0), 0) = V(X) > 0$. L'hypothèse **H3**. d'homoscédasticité indique que, pour tout $t \in [0, 1]$, $v(\gamma(t), t) = v(\gamma(0), 0)$, ce qui termine la démonstration. \square

Lemme 3.5.2. *Sous les hypothèses **H1.**, **H2.** et **H3.**, et si $k_n = k_n(\beta_0)$, pour tout $\gamma_1 \in \mathbb{B}$, il existe des constantes $C(\gamma_1)$ et $C(\beta_0) \in \mathbb{R}^{*+}$ telles que*

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \mathcal{V}_{\gamma_1(t)}(Z | t) - C(\gamma_1) \right| \xrightarrow[n \rightarrow \infty]{P} 0, \quad (3.28)$$

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{\mathcal{V}_{\gamma_1(t)}(Z | t)}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{C(\gamma_1)}{C(\beta_0)} \right| \xrightarrow[n \rightarrow \infty]{P} 0, \quad (3.29)$$

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{\mathcal{V}_{\gamma_1(t)}(Z | t)}{\mathcal{V}_{\beta_0}(Z | t)^{1/2}} - \frac{C(\gamma_1)}{C(\beta_0)^{1/2}} \right| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.30)$$

Preuve. Preuve de l'équation (3.28). Plaçons-nous sous les conditions **H1.** et **H2.**. Soit $\gamma_1 \in \mathbb{B}$. Nous rappelons que $v(\gamma_1(\cdot), \cdot)$ est défini à l'équation (3.13). Nous avons l'inégalité presque-sûre

$$\begin{aligned} & \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \mathcal{V}_{\gamma_1(t)}(Z | t) - v(\gamma_1(t), t) \right| \\ &= \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{S^{(2)}(\gamma_1(t), t)}{S^{(0)}(\gamma_1(t), t)} - \left(\frac{S^{(1)}(\gamma_1(t), t)}{S^{(0)}(\gamma_1(t), t)} \right)^2 - \left[\frac{s^{(2)}(\gamma_1(t), t)}{s^{(0)}(\gamma_1(t), t)} - \left(\frac{s^{(1)}(\gamma_1(t), t)}{s^{(0)}(\gamma_1(t), t)} \right)^2 \right] \right| \\ &\leq \sqrt{n} \sup_{t \in [0, 1]} \left| \frac{S^{(2)}(\gamma_1(t), t)}{S^{(0)}(\gamma_1(t), t)} - \left(\frac{S^{(1)}(\gamma_1(t), t)}{S^{(0)}(\gamma_1(t), t)} \right)^2 - \left[\frac{s^{(2)}(\gamma_1(t), t)}{s^{(0)}(\gamma_1(t), t)} - \left(\frac{s^{(1)}(\gamma_1(t), t)}{s^{(0)}(\gamma_1(t), t)} \right)^2 \right] \right| \\ &\leq V_n + W_n, \end{aligned}$$

où

$$V_n = \sqrt{n} \sup_{t \in [0, 1]} \left| \frac{S^{(2)}(\gamma_1(t), t)}{S^{(0)}(\gamma_1(t), t)} - \frac{s^{(2)}(\gamma_1(t), t)}{s^{(0)}(\gamma_1(t), t)} \right|,$$

et

$$W_n = \sqrt{n} \sup_{t \in [0, 1]} \left| \left(\frac{S^{(1)}(\gamma_1(t), t)}{S^{(0)}(\gamma_1(t), t)} \right)^2 - \left(\frac{s^{(1)}(\gamma_1(t), t)}{s^{(0)}(\gamma_1(t), t)} \right)^2 \right|.$$

Etudions le comportement asymptotique de ces deux termes séparément. Notons m_0 , M_1 et M_2 les constantes strictement positives telles que $s^{(0)}(\gamma(t), t) \geq m_0$ et $|s^{(i)}(\gamma(t), t)| \leq M_i$, ($i = 1, 2$) pour tout $t \in [0, 1]$ et $\gamma \in \mathbb{B}$. Leur existence est entraînée par la condition **H2**. Nous avons

$$\begin{aligned}
V_n &\leq \sqrt{n} \sup_{t \in [0, 1]} \left| \frac{1}{S^{(0)}(\gamma_1(t), t)} \left(S^{(2)}(\gamma_1(t), t) - s^{(2)}(\gamma_1(t), t) \right) \right| \\
&\quad + \sqrt{n} \sup_{t \in [0, 1]} \left| s^{(2)}(\gamma_1(t), t) \left(\frac{1}{S^{(0)}(\gamma_1(t), t)} - \frac{1}{s^{(0)}(\gamma_1(t), t)} \right) \right| \\
&\leq \sqrt{n} \sup_{t \in [0, 1]} \frac{1}{S^{(0)}(\gamma_1(t), t)} \sup_{t \in [0, 1]} \left| S^{(2)}(\gamma_1(t), t) - s^{(2)}(\gamma_1(t), t) \right| \\
&\quad + M_2 \sqrt{n} \sup_{t \in [0, 1]} \left| \frac{1}{S^{(0)}(\gamma_1(t), t)} - \frac{1}{s^{(0)}(\gamma_1(t), t)} \right|. \tag{3.31}
\end{aligned}$$

De plus,

$$\begin{aligned}
&\sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}} \left| \frac{1}{S^{(0)}(\gamma(t), t)} - \frac{1}{s^{(0)}(\gamma(t), t)} \right| \\
&= \sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}} \frac{|s^{(0)}(\gamma(t), t) - S^{(0)}(\gamma(t), t)|}{S^{(0)}(\gamma(t), t) s^{(0)}(\gamma(t), t)} \\
&\leq m_0^{-1} \sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}} |s^{(0)}(\gamma(t), t) - S^{(0)}(\gamma(t), t)| \sup_{t \in [0, 1], \gamma \in \mathbb{B}} \frac{1}{S^{(0)}(\gamma(t), t)}.
\end{aligned}$$

Or, d'après l'équation (3.12), $\sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}} |s^{(0)}(\gamma(t), t) - S^{(0)}(\gamma(t), t)| = o_P(1)$, et comme $s^{(0)}$ est minoré uniformément par m_0 , on a $\sup_{t \in [0, 1], \gamma \in \mathbb{B}} S^{(0)}(\gamma(t), t)^{-1} = O_P(1)$, ce qui signifie que ce terme est borné à partir d'un certain rang avec grande probabilité. Par conséquent,

$$\sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}} \left| \frac{1}{S^{(0)}(\gamma(t), t)} - \frac{1}{s^{(0)}(\gamma(t), t)} \right| \xrightarrow[n \rightarrow \infty]{P} 0. \tag{3.32}$$

De plus, sous la condition **H1**,

$$\sqrt{n} \sup_{t \in [0, 1], \gamma \in \mathbb{B}} |S^{(2)}(\gamma(t), t) - s^{(2)}(\gamma(t), t)| \xrightarrow[n \rightarrow \infty]{P} 0,$$

et comme $\sup_{t \in [0, 1], \gamma \in \mathbb{B}} S^{(0)}(\gamma(t), t)^{-1} = O_P(1)$, on a la convergence

$$\sqrt{n} \sup_{t \in [0, 1]} \frac{1}{S^{(0)}(\gamma_1(t), t)} \sup_{t \in [0, 1]} |S^{(2)}(\gamma_1(t), t) - s^{(2)}(\gamma_1(t), t)| \xrightarrow[n \rightarrow \infty]{P} 0.$$

L'équation (3.31) permet alors de conclure que V_n converge vers 0 en probabilité lorsque $n \rightarrow \infty$. En utilisant le même raisonnement, nous pouvons majorer le terme W_n par

$$W_n \leq \sqrt{n} \sup_{t \in [0, 1]} \left| \frac{1}{S^{(0)}(\gamma_1(t), t)^2} \left(S^{(1)}(\gamma_1(t), t)^2 - s^{(1)}(\gamma_1(t), t)^2 \right) \right|$$

$$\begin{aligned}
& + \sqrt{n} \sup_{t \in [0,1]} \left| s^{(1)}(\gamma_1(t), t)^2 \left(\frac{1}{S^{(0)}(\gamma_1(t), t)^2} - \frac{1}{s^{(0)}(\gamma_1(t), t)^2} \right) \right| \\
& \leq \sqrt{n} \sup_{t \in [0,1]} \frac{1}{S^{(0)}(\gamma_1(t), t)^2} \sup_{t \in [0,1]} \left| S^{(1)}(\gamma_1(t), t)^2 - s^{(1)}(\gamma_1(t), t)^2 \right| \\
& + M_1^2 \sqrt{n} \sup_{t \in [0,1]} \left| \frac{1}{S^{(0)}(\gamma_1(t), t)^2} - \frac{1}{s^{(0)}(\gamma_1(t), t)^2} \right|. \tag{3.33}
\end{aligned}$$

Nous avons

$$\begin{aligned}
& \sqrt{n} \sup_{t \in [0,1], \gamma \in \mathbb{B}} \left| \frac{1}{S^{(0)}(\gamma(t), t)^2} - \frac{1}{s^{(0)}(\gamma(t), t)^2} \right| \\
& = \sqrt{n} \sup_{t \in [0,1], \gamma \in \mathbb{B}} \left| \left(\frac{1}{S^{(0)}(\gamma(t), t)} - \frac{1}{s^{(0)}(\gamma(t), t)} \right) \left(\frac{1}{S^{(0)}(\gamma(t), t)} + \frac{1}{s^{(0)}(\gamma(t), t)} \right) \right| \\
& \leq \sqrt{n} \sup_{t \in [0,1], \gamma \in \mathbb{B}} \left| \frac{1}{S^{(0)}(\gamma(t), t)} - \frac{1}{s^{(0)}(\gamma(t), t)} \right| \left(\sup_{t \in [0,1], \gamma \in \mathbb{B}} \frac{1}{S^{(0)}(\gamma(t), t)} + m_0^{-1} \right).
\end{aligned}$$

L'équation (3.32) et le fait que $\sup_{t \in [0,1], \gamma \in \mathbb{B}} S^{(0)}(\gamma(t), t)^{-1} = O_P(1)$ impliquent que

$$\sqrt{n} \sup_{t \in [0,1], \gamma \in \mathbb{B}} \left| \frac{1}{S^{(0)}(\gamma(t), t)^2} - \frac{1}{s^{(0)}(\gamma(t), t)^2} \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Avec des arguments similaires, on peut montrer que

$$\sqrt{n} \sup_{t \in [0,1], \gamma \in \mathbb{B}} \left| S^{(1)}(\gamma(t), t)^2 - s^{(1)}(\gamma(t), t)^2 \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Comme $\sup_{t \in [0,1], \gamma \in \mathbb{B}} S^{(0)}(\gamma(t), t)^{-2} = O_P(1)$, on a la convergence

$$\sqrt{n} \sup_{t \in [0,1]} \frac{1}{S^{(0)}(\gamma_1(t), t)^2} \sup_{t \in [0,1]} \left| S^{(1)}(\gamma_1(t), t)^2 - s^{(1)}(\gamma_1(t), t)^2 \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Par conséquent, d'après l'équation (3.33), W_n converge en probabilité vers 0 lorsque $n \rightarrow \infty$.

En conclusion,

$$\sqrt{n} \sup_{t \in \{t_1, \dots, t_{k_n}\}} \left| \mathcal{V}_{\gamma_1(t)}(Z | t) - v(\gamma_1(t), t) \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Sous la condition **H3.**, il existe une constante positive $C(\gamma_1)$ telle que $v(\gamma_1(t), t) = C(\gamma_1)$ pour tout $t \in [0, 1]$. Le Lemme 3.5.1 entraîne la stricte positivité de $C(\gamma_1)$, et l'équation (3.28) est démontrée.

Preuve des équations (3.29) et (3.30). Soit $\gamma_1 \in \mathbb{B}$. D'après l'équation (3.28), sous les conditions **H1.**, **H2.** et **H3.**, il existe deux constantes $C(\gamma_1)$, $C(\beta_0) > 0$ telles que

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \mathcal{V}_{\gamma_1(t)}(Z | t) - C(\gamma_1) \right| \xrightarrow[n \rightarrow \infty]{P} 0, \tag{3.34}$$

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \mathcal{V}_{\beta_0}(Z | t) - C(\beta_0) \right| \xrightarrow[n \rightarrow \infty]{P} 0. \tag{3.35}$$

Puisque $k_n = k_n(\beta_0)$, l'équation (3.18) est vérifiée :

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \mathcal{V}_{\beta_0}(Z | t) > C_{\mathcal{V}} \quad \text{p.s.}$$

Nous avons alors

$$\begin{aligned} & \sup_{t \in \{t_1, \dots, t_{k_n}\}} \left| \frac{\mathcal{V}_{\gamma_1(t)}(Z | t)}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{C(\gamma_1)}{C(\beta_0)} \right| \\ & \leq \sup_{t \in \{t_1, \dots, t_{k_n}\}} \left| \frac{1}{\mathcal{V}_{\beta_0}(Z | t)} (\mathcal{V}_{\gamma_1(t)}(Z | t) - C(\gamma_1)) \right| + C(\gamma_1) \sup_{t \in \{t_1, \dots, t_{k_n}\}} \left| \frac{1}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{1}{C(\beta_0)} \right| \\ & \leq C_{\mathcal{V}}^{-1} \sup_{t \in \{t_1, \dots, t_{k_n}\}} |\mathcal{V}_{\gamma_1(t)}(Z | t) - C(\gamma_1)| + C(\gamma_1) \sup_{t \in \{t_1, \dots, t_{k_n}\}} \left| \frac{1}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{1}{C(\beta_0)} \right|. \end{aligned}$$

De plus,

$$\begin{aligned} \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{1}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{1}{C(\beta_0)} \right| &= \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \frac{|\mathcal{V}_{\beta_0}(Z | t) - C(\beta_0)|}{C(\beta_0)\mathcal{V}_{\beta_0}(Z | t)} \\ &\leq C(\beta_0)^{-1} C_{\mathcal{V}}^{-1} \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} |\mathcal{V}_{\beta_0}(Z | t) - C(\beta_0)|. \end{aligned}$$

Le résultat (3.35) implique alors

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{1}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{1}{C(\beta_0)} \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Cette convergence ainsi que celle de l'équation (3.34) permettent de conclure que

$$\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{\mathcal{V}_{\gamma_1(t)}(Z | t)}{\mathcal{V}_{\beta_0}(Z | t)} - \frac{C(\gamma_1)}{C(\beta_0)} \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Nous omettons la démonstration de l'équation (3.30) qui est basée sur les mêmes arguments. \square

Preuves des théorèmes

Preuve du Théorème 3.3.1. Nous renvoyons à la preuve du Théorème 3.3.2 qui est plus générale. \square

Preuve du Théorème 3.3.2. Soit $t \in [0, 1]$. Sous les hypothèses **H1.**, **H2.**, **H3.** et **H4.**, supposons que $\beta(t)$ soit le vrai paramètre du modèle à risques non proportionnels et que le processus du score standardisé U_n^* soit évalué en $\beta_0 \in \mathbb{B}$. Rappelons les égalités de la Proposition 3.2.2 :

$$\mathcal{E}_{\beta(t)}(Z | t) = E_{\beta(t)}(\mathcal{Z}(t) | \mathcal{F}_t^*),$$

et

$$\mathcal{V}_{\beta_0}(Z | t) = V_{\beta_0}(\mathcal{Z}(t) | \mathcal{F}_t^*).$$

Définissons le processus $\{U_n^{**}(\beta_0, t), t \in [0, 1]\}$ par

$$U_n^{**}(\beta_0, t) = \int_0^t \frac{\mathcal{Z}(s) - E_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{\left(k_n V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)\right)^{1/2}} d\bar{N}^*(s), \quad 0 \leq t \leq 1. \quad (3.36)$$

Ce processus de $(D[0, 1], \mathbb{R})$ est constant par morceaux avec un saut à chaque temps de décès. Le schéma de la preuve est le suivant. Dans un premier temps, nous allons décomposer $U_n^{**}(\beta_0, \cdot)$ en deux processus de $(D[0, 1], \mathbb{R})$. Nous allons montrer que le premier processus converge vers un mouvement Brownien à l'aide d'un théorème de convergence de différences de martingales. La convergence du deuxième processus sera obtenue par application du Lemme 3.2.1. Finalement, le comportement asymptotique du processus $U_n^*(\beta_0, \cdot)$ de $(C[0, 1], \mathbb{R})$, qui est l'interpolation linéaire des variables $\{U_n^{**}(\beta_0, t_i), i = 1, \dots, k_n\}$, sera obtenu en montrant que la différence entre U_n^* et U_n^{**} tend vers 0 en probabilité lorsque $n \rightarrow \infty$.

Un développement de Taylor-Lagrange de l'espérance conditionnelle $E_{\beta(t)}\{\mathcal{Z}(t) | \mathcal{F}_{t^-}^*\}$ au point $\beta(t) = \beta_0$ donne

$$E_{\beta_0}(\mathcal{Z}(t) | \mathcal{F}_{t^-}^*) = E_{\beta(t)}(\mathcal{Z}(t) | \mathcal{F}_{t^-}^*) + (\beta_0 - \beta(t)) \left. \frac{\partial}{\partial \beta} E_{\beta}(\mathcal{Z}(t) | \mathcal{F}_{t^-}^*) \right|_{\beta = \tilde{\beta}(t)}, \quad (3.37)$$

où $\tilde{\beta}$ appartient à la boule de centre β et de rayon $\sup_{t \in [0, 1]} |\beta(t) - \beta_0|$. Rappelons que

$$\frac{\partial}{\partial \beta} E_{\beta}(\mathcal{Z}(t) | \mathcal{F}_{t^-}^*) = V_{\beta}(\mathcal{Z}(t) | \mathcal{F}_{t^-}^*).$$

Ainsi,

$$\begin{aligned} U_n^{**}(\beta_0, t) &= \int_0^t \left(\frac{\mathcal{Z}(s) - E_{\beta(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{\left[k_n V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)\right]^{1/2}} - (\beta_0 - \beta(s)) \frac{V_{\tilde{\beta}(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{\left[k_n V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)\right]^{1/2}} \right) d\bar{N}^*(s) \\ &= X_n(t) + \sqrt{k_n} A_n(t), \end{aligned}$$

où

$$\begin{aligned} A_n(t) &= \frac{1}{k_n} \int_0^t \{\beta(s) - \beta_0\} \frac{V_{\tilde{\beta}(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)^{1/2}} d\bar{N}^*(s), \\ X_n(t) &= \frac{1}{k_n^{1/2}} \int_0^t \frac{\mathcal{Z}(s) - E_{\beta(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)^{1/2}} d\bar{N}^*(s). \end{aligned}$$

Étudions le comportement limite de A_n . Nous avons

$$\sup_{t \in [0, 1]} \left| A_n(t) - C_2 \int_0^t \{\beta(s) - \beta_0\} ds \right|$$

$$\begin{aligned}
&\leq \sup_{t \in [0,1]} |\beta(t) - \beta_0| \sup_{t \in [0,1]} \left| \frac{1}{k_n} \int_0^t \frac{V_{\tilde{\beta}(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)^{1/2}} d\bar{N}^*(s) - C_2 t \right| \\
&\leq \delta_1 \sup_{t \in [0,1]} \left| \frac{1}{k_n} \int_0^t \frac{V_{\tilde{\beta}(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)^{1/2}} d\bar{N}^*(s) - C_2 t \right|, \tag{3.38}
\end{aligned}$$

car $\beta_0 \in \mathbb{B}$ de rayon δ_1 (condition **H1.**). Sous les conditions **H1.**, **H2.** et **H3.**, l'équation (3.30) du Lemme 3.5.2 indique l'existence d'une constante $C_2 > 0$ telle que

$$\sup_{s \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \frac{V_{\tilde{\beta}(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)^{1/2}} - C_2 \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Le Lemme 3.2.1 implique alors

$$\sup_{t \in [0,1]} \left| \frac{1}{k_n} \int_0^t \frac{V_{\tilde{\beta}(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)^{1/2}} d\bar{N}^*(s) - C_2 t \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Finalement, d'après l'équation (3.38), la convergence de l'équation (3.17) est démontrée.

Montrons que X_n converge vers un mouvement Brownien à l'aide du théorème central limite fonctionnel pour les différences de martingales introduit par [Helland \(1982\)](#) et rappelé en Théorème 2.2.3. Soit $j = 1, \dots, k_n$. Notons ξ_{j,k_n} le j ème incrément de X_n :

$$\xi_{j,k_n} = \frac{\mathcal{Z}(t_j) - E_{\beta(t_j)}(\mathcal{Z}(t_j) | \mathcal{F}_{t_j^-}^*)}{\left[k_n V_{\beta_0}(\mathcal{Z}(t_j) | \mathcal{F}_{t_j^-}^*) \right]^{1/2}}.$$

L'incrément ξ_{j,k_n} est $\mathcal{F}_{t_j}^*$ -mesurable et $E_{\beta(t_j)}(\xi_{j,k_n} | \mathcal{F}_{t_j^-}^*) = 0$. Le processus \bar{N}^* présentant un saut à chaque temps de décès $t_j = j/k_n$ (Proposition 3.2.1), nous avons

$$X_n(t) = \int_0^{\lfloor tk_n \rfloor / k_n} \frac{\mathcal{Z}(s) - E_{\beta(s)}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*)}{\left\{ k_n V_{\beta_0}(\mathcal{Z}(s) | \mathcal{F}_{s^-}^*) \right\}^{1/2}} d\bar{N}^*(s) = \sum_{j=1}^{\lfloor tk_n \rfloor} \xi_{j,k_n},$$

où la fonction $t \rightarrow \lfloor tk_n \rfloor$ est positive, croissante, à valeurs entières et continue à droite. Vérifions les hypothèses du Théorème 2.2.3 :

a. (Espérance). L'inclusion des σ -algèbres $\mathcal{F}_{t_{j-1}}^* \subset \mathcal{F}_{t_j}^*$ entraîne

$$E_{\beta(t_{j-1})}(\xi_{j,k_n} | \mathcal{F}_{t_{j-1}}^*) = E_{\beta(t_{j-1})} \left(E_{\beta(t_j)}(\xi_{j,k_n} | \mathcal{F}_{t_j^-}^*) \middle| \mathcal{F}_{t_{j-1}}^* \right) = 0. \tag{3.39}$$

b. (Variance). Sous les hypothèses **H1.**, **H2.** et **H3.**, d'après le Lemme 3.5.2, il existe une

constante $C_1(\beta, \beta_0) > 0$ telle que

$$\sup_{j=1, \dots, k_n} \left| \frac{V_{\beta(t_j)} \left(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j^-}^* \right)}{V_{\beta_0} \left(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j^-}^* \right)} - C_1(\beta, \beta_0)^2 \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

En particulier, $C_1(\beta, \beta_0) = 1$ lorsque $\beta(t) = \beta_0$. Soit $s \in \{t_1, t_2, \dots, t_{k_n}\}$. D'après l'équation (3.18),

$$V_{\beta_0}(\mathcal{Z}(s) \mid \mathcal{F}_{s^-}^*) > C_V \quad \text{p.s.}$$

D'autre part, d'après l'hypothèse **H4.**, nous avons la majoration

$$\begin{aligned} |\mathcal{Z}(s)| &= \left| \sum_{i=1}^n Z_i(X_i) \mathbf{1}_{\phi_n(X_i) = s, \delta_i = 1} \right| \\ &\leq \sup_{i=1, \dots, n} |Z_i(X_i)| \sum_{i=1}^n \mathbf{1}_{\phi_n(X_i) = s, \delta_i = 1} \\ &\leq L, \end{aligned} \tag{3.40}$$

car $s \in \{t_1, t_2, \dots, t_{k_n}\}$ donc $\sum_{i=1}^n \mathbf{1}_{\phi_n(X_i) = s, \delta_i = 1} = 1$. Par conséquent,

$$V_{\beta(s)}(\mathcal{Z}(s) \mid \mathcal{F}_{s^-}^*) \leq E_{\beta(s)}(\mathcal{Z}(s)^2 \mid \mathcal{F}_{s^-}^*) \leq L^2.$$

En conclusion, nous avons la majoration presque sûre

$$\sup_{j=1, \dots, k_n} \frac{V_{\beta(t_j)} \left(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j^-}^* \right)}{V_{\beta_0} \left(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j^-}^* \right)} \leq \frac{L^2}{C_V},$$

et une application du théorème de convergence dominée implique que

$$\sup_{j=1, \dots, k_n} \left| \frac{V_{\beta(t_j)} \left(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j^-}^* \right)}{V_{\beta_0} \left(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j^-}^* \right)} - C_1(\beta, \beta_0)^2 \right| \xrightarrow[n \rightarrow \infty]{L^1} 0. \tag{3.41}$$

Montrons que

$$\sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\xi_{j, k_n}^2 \mid \mathcal{F}_{t_{j-1}}^* \right) \xrightarrow[n \rightarrow \infty]{L^1} C_1(\beta, \beta_0)^2 t.$$

Nous avons les relations

$$\begin{aligned} &\left| \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\xi_{j, k_n}^2 \mid \mathcal{F}_{t_{j-1}}^* \right) - C_1(\beta, \beta_0)^2 t \right| \\ &= \left| \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(E_{\beta(t_j)} \left(\xi_{j, k_n}^2 \mid \mathcal{F}_{t_j^-}^* \right) \mid \mathcal{F}_{t_{j-1}}^* \right) - C_1(\beta, \beta_0)^2 t \right| \end{aligned}$$

$$\begin{aligned}
&= \left| \frac{1}{k_n} \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\frac{V_{\beta(t_j)}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)}{V_{\beta_0}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)} \middle| \mathcal{F}_{t_{j-1}}^* \right) - C_1(\beta, \beta_0)t \right| \\
&\leq \frac{1}{k_n} \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\left| \frac{V_{\beta(t_j)}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)}{V_{\beta_0}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)} - C_1(\beta, \beta_0)^2 \right| \middle| \mathcal{F}_{t_{j-1}}^* \right) + C_1(\beta, \beta_0)^2 \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right|.
\end{aligned}$$

Par conséquent,

$$\begin{aligned}
&E \left(\left| \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\xi_{j,k_n}^2 \middle| \mathcal{F}_{t_{j-1}}^* \right) - C_1(\beta, \beta_0)^2 t \right| \right) \\
&\leq \frac{1}{k_n} \sum_{j=1}^{\lfloor tk_n \rfloor} E \left(\left| \frac{V_{\beta(t_j)}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)}{V_{\beta_0}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)} - C_1(\beta, \beta_0)^2 \right| \right) + C_1(\beta, \beta_0)^2 \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right| \\
&\leq \frac{\lfloor tk_n \rfloor}{k_n} \sup_{j=1, \dots, \lfloor tk_n \rfloor} E \left(\left| \frac{V_{\beta(t_j)}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)}{V_{\beta_0}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)} - C_1(\beta, \beta_0)^2 \right| \right) + C_1(\beta, \beta_0)^2 \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right| \\
&\leq \frac{\lfloor tk_n \rfloor}{k_n} E \left(\sup_{j=1, \dots, \lfloor tk_n \rfloor} \left| \frac{V_{\beta(t_j)}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)}{V_{\beta_0}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)} - C_1(\beta, \beta_0)^2 \right| \right) + C_1(\beta, \beta_0)^2 \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right| \\
&\leq \frac{\lfloor tk_n \rfloor}{k_n} E \left(\sup_{j=1, \dots, k_n} \left| \frac{V_{\beta(t_j)}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)}{V_{\beta_0}(\mathcal{Z}(t_j) \mid \mathcal{F}_{t_j}^*)} - C_1(\beta, \beta_0)^2 \right| \right) + C_1(\beta, \beta_0)^2 \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right|.
\end{aligned}$$

L'équation (3.41) permet de conclure que $\sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta} \left(\xi_{j,k_n}^2 \middle| \mathcal{F}_{t_{j-1}}^* \right)$ converge en moyenne et donc en probabilité vers $C_1(\beta, \beta_0)^2 t$ lorsque $n \rightarrow \infty$.

- c. (Condition de Lyapunov). Montrons la convergence en moyenne de $\sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\xi_{j,k_n}^3 \middle| \mathcal{F}_{t_{j-1}}^* \right)$ vers 0 lorsque $n \rightarrow \infty$. Nous avons

$$\begin{aligned}
E \left(\left| \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})} \left(\xi_{j,k_n}^3 \middle| \mathcal{F}_{t_{j-1}}^* \right) \right| \right) &\leq \sum_{j=1}^{\lfloor tk_n \rfloor} E \left(E_{\beta(t_{j-1})} \left(\left| \xi_{j,k_n}^3 \right| \middle| \mathcal{F}_{t_{j-1}}^* \right) \right) \\
&= \sum_{j=1}^{\lfloor tk_n \rfloor} E \left(\left| \xi_{j,k_n}^3 \right| \right), \tag{3.42}
\end{aligned}$$

et nous rappelons que pour $j = 1, \dots, k_n$,

$$\xi_{j,k_n} = \frac{\mathcal{Z}(t_j) - E_{\beta(t_j)} \left(\mathcal{Z}(t_j) \middle| \mathcal{F}_{t_j}^* \right)}{\left\{ k_n V_{\beta_0} \left(\mathcal{Z}(t_j) \middle| \mathcal{F}_{t_j}^* \right) \right\}^{1/2}}.$$

L'équation (3.40) implique que

$$E \left(\left| Z(t) - E_{\beta(t)}(Z(t) | \mathcal{F}_{t^-}^*) \right|^3 \right) \leq 8L^3.$$

De plus, d'après l'équation (3.18),

$$\forall t \in \{t_1, t_2, \dots, t_{k_n}\}, \quad V_{\beta_0}(Z(t) | \mathcal{F}_{t^-}^*) > C_V \quad \text{p.s.} \quad (3.43)$$

Ainsi, l'équation (3.42) devient

$$\begin{aligned} E \left(\left| \sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})}(\xi_{j,k_n}^3 | \mathcal{F}_{t_{j-1}}^*) \right|^3 \right) &\leq \sum_{j=1}^{k_n} \frac{1}{(k_n C_V)^{3/2}} E \left(\left| Z(t_j) - E_{\beta(t_j)}(Z(t_j) | \mathcal{F}_{t_j^-}^*) \right|^3 \right) \\ &\leq \frac{8L^3}{k_n^{1/2} C_V^{3/2}}, \end{aligned} \quad (3.44)$$

et $\sum_{j=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{j-1})}(\xi_{j,k_n}^3 | \mathcal{F}_{t_{j-1}}^*)$ converge en moyenne vers 0 lorsque $n \rightarrow \infty$ et donc en probabilité. La condition de Lyapunov est vérifiée.

Nous pouvons à présent appliquer le Théorème 2.2.3 et conclure que

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} C_1(\beta, \beta_0) \mathcal{W},$$

où la constante $C_1(\beta, \beta_0)^2$ est le ratio de deux variances conditionnelles asymptotiques v évaluées aux paramètres β et β_0 . Lorsque $\beta = \beta_0$, on a $C_1(\beta, \beta_0) = 1$. Nous avons la relation

$$\begin{aligned} U_n^*(\beta_0, \cdot) - \sqrt{k_n} A_n &= U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot) + U_n^{**}(\beta_0, \cdot) - \sqrt{k_n} A_n \\ &= U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot) + X_n. \end{aligned} \quad (3.45)$$

Nous venons de donner le comportement limite de X_n lorsque $n \rightarrow \infty$. Montrons la convergence en probabilité de $U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot)$ vers 0 lorsque $n \rightarrow \infty$ par rapport à la norme uniforme. Soit $\varepsilon > 0$,

$$\begin{aligned} P(\|U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot)\| \geq \varepsilon) &= P \left(\sup_{t \in [0,1]} |U_n^*(\beta_0, t) - U_n^{**}(\beta_0, t)| \geq \varepsilon \right) \\ &= P \left(\sup_{i=1, \dots, k_n} |U_n^*(\beta_0, t_i) - U_n^{**}(\beta_0, t_{i-1})| \geq \varepsilon \right) \\ &= P \left(\frac{1}{\sqrt{k_n}} \sup_{i=1, \dots, k_n} \left| \frac{Z(t_j) - E_{\beta_0}(Z(t_j) | \mathcal{F}_{t_j^-}^*)}{V_{\beta_0}(Z(t_j) | \mathcal{F}_{t_j^-}^*)^{1/2}} \right| \geq \varepsilon \right) \\ &\leq P \left(\frac{2L}{\sqrt{k_n} C_V} \geq \varepsilon \right), \end{aligned}$$

d'après les équations (3.18) et (3.40). Ainsi, $U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot) \xrightarrow[n \rightarrow \infty]{P} 0$. Le théorème de

Slutsky appliqué à la décomposition de l'équation (3.45) permet de conclure que

$$U_n^*(\beta_0, \cdot) - k_n^{-1/2} A_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} C_1(\beta, \beta_0) \mathcal{W},$$

dans $(C[0, 1], \mathbb{R})$ par rapport à la topologie de la convergence uniforme. □

3.5.3 Preuves de la Section 3.4

Nous rappelons que la norme sup matricielle n'est pas sous-multiplicative, mais on a les relations $\|Ax\| \leq p\|A\|\|x\|$ et $\|AB\| \leq p\|A\|\|B\|$ pour $A, B \in \mathcal{M}_{p \times p}(\mathbb{R})$ et $x \in \mathbb{R}^p$. On remarque que si A est diagonale et ses éléments sont positifs, on a

$$\|A^{1/2}\| = \max_{l=1, \dots, p} (A^{1/2})_{ll} \leq \left(\max_{l=1, \dots, p} (A)_{ll} \right)^{1/2} = \|A\|^{1/2}.$$

Preuve de la Proposition 3.4.1. D'après la condition **H3'**, Σ est une matrice symétrique à coefficients réels donc $\Sigma^{1/2}$ existe et $\|\Sigma^{1/2}\| < \infty$. Nous avons

$$\begin{aligned} & \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t) - \Sigma \right\| \\ &= \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t) - \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} \Sigma^{1/2} + \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} \Sigma^{1/2} - \Sigma \right\| \\ &\leq p\sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} - \Sigma^{1/2} \right\| \left(\sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} \right\| + \|\Sigma^{1/2}\| \right). \end{aligned}$$

Or, pour tout $t \in \{t_1, \dots, t_{k_n}\}$ et $\gamma \in \mathbb{B}'$, d'après la condition **H4'**,

$$\left\| \mathcal{V}_{\gamma(t)}(Z | t) \right\| \leq \left\| \mathcal{E}_{\gamma}(Z^{\otimes 2} | t) \right\| + \left\| \mathcal{E}_{\gamma}(Z | t)^{\otimes 2} \right\| \leq 2L'^2. \quad (3.46)$$

Ainsi, en choisissant une matrice de passage orthogonale $P_{\gamma(t)}(t)$ de norme 1 dans la diagonalisation de la matrice symétrique définie positive $\mathcal{V}_{\gamma(t)}(Z | t) = P_{\gamma(t)}(t) D_{\gamma(t)}(t) P_{\gamma(t)}(t)^T$ avec $D_{\gamma(t)}(t)$ diagonale, on a

$$\begin{aligned} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} \right\| &\leq p^2 \left\| P_{\gamma(t)}(t) \right\|^2 \left\| D_{\gamma(t)}(t)^{1/2} \right\| \\ &\leq p^2 \left\| D_{\gamma(t)}(t) \right\|^{1/2} \\ &= p^2 \left\| P_{\gamma(t)}(t)^T P_{\gamma(t)}(t) D_{\gamma(t)}(t) P_{\gamma(t)}(t)^T P_{\gamma(t)}(t) \right\|^{1/2} \\ &\leq p^2 \left(p^2 \left\| P_{\gamma(t)}(t) D_{\gamma(t)}(t) P_{\gamma(t)}(t)^T \right\| \left\| P_{\gamma(t)}(t) \right\|^2 \right)^{1/2} \\ &= p^3 \left\| \mathcal{V}_{\gamma(t)}(Z | t) \right\|^{1/2} \\ &\leq \sqrt{2} p^3 L'. \end{aligned}$$

Par conséquent,

$$\sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} \right\| \leq \sqrt{2} p^3 L', \quad (3.47)$$

et

$$\begin{aligned} & \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t) - \Sigma \right\| \\ & \leq p \left(\sqrt{2} p^3 L' + \left\| \Sigma^{1/2} \right\| \right) \sqrt{n} \sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} - \Sigma^{1/2} \right\|. \end{aligned}$$

L'équation (3.20) permet alors d'obtenir le résultat. \square

Preuve du Théorème 3.4.1. La preuve suit le même schéma que dans le cadre univarié. Rappelons que l'espace des fonctions càdlàg de $(D[0, 1], \mathbb{R}^p)$ est muni de la topologie produit de Skorokhod, introduite en Définition 2.1.3. De plus, la norme d'un vecteur de \mathbb{R}^p ou d'une matrice de $\mathcal{M}_{p \times p}(\mathbb{R})$ est sa norme sup. Le processus étant évalué au paramètre $\beta_0 \in \mathbb{B}'$, on a $k_n = k_n(\beta_0)$ et la majoration de l'équation (3.22) devient

$$\forall t \in \{t_1, \dots, t_{k_n}\}, \left\| \mathcal{V}_{\beta_0(t)}(Z | t)^{-1} \right\| \leq C_{\mathcal{V}}^{-1} \quad \text{p.s.} \quad (3.48)$$

On considère U_n^{**} le processus càdlàg ayant un saut à chaque temps d'évènement $t_i = i/k_n$, $i = 1, \dots, k_n$, tel que

$$U_n^{**}(\beta_0, t) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{\lfloor tk_n \rfloor} V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2} \left\{ \mathcal{Z}(t_i) - E_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right) \right\}, \quad 0 \leq t \leq 1.$$

Soit $\gamma \in \mathbb{B}'$. Rappelons les égalités de la Proposition 3.2.2 :

$$\mathcal{E}_{\gamma(t_i)}(Z | t_i) = E_{\gamma(t_i)} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right),$$

et

$$\mathcal{V}_{\gamma(t_i)}(Z | t_i) = V_{\gamma(t_i)} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right).$$

Remarquons que $U_n^{**}(\beta_0, t) = X_n(t) + \sqrt{k_n} B_n(t)$ pour $t \in [0, 1]$, où

$$\begin{aligned} X_n(t) &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{\lfloor tk_n \rfloor} \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \left\{ \mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right\} \\ B_n(t) &= \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \left\{ \mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\beta_0}(Z | t_i) \right\}. \end{aligned}$$

Nous avons

$$\left\| U_n^*(\beta_0, \cdot) - \sqrt{k_n} B_n - \mathcal{W}_p \right\| \leq \left\| U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot) \right\| + \left\| U_n^{**}(\beta_0, \cdot) - \sqrt{k_n} B_n - \mathcal{W}_p \right\|$$

$$= \|U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot)\| + \|X_n - \mathcal{W}_p\|. \quad (3.49)$$

Etudions les comportements limites des deux termes de droite de l'inégalité.

Premier terme. Soit $\varepsilon > 0$,

$$\begin{aligned} P\left(\sup_{t \in [0,1]} \|U_n^*(\beta_0, t) - U_n^{**}(\beta_0, t)\| \geq \varepsilon\right) &= P\left(\sup_{i=1, \dots, k_n} \left\|U_n^{**}\left(\beta_0, \frac{i}{k_n}\right) - U_n^{**}\left(\beta_0, \frac{i-1}{k_n}\right)\right\| \geq \varepsilon\right) \\ &= P\left(\frac{1}{\sqrt{k_n}} \sup_{i=1, \dots, k_n} \left\|V_{\beta_0}(Z | t_i)^{-1/2} \{Z(t_i) - \mathcal{E}_{\beta_0}(Z | t_i)\}\right\| \geq \varepsilon\right) \\ &\leq P\left(\frac{p}{\sqrt{k_n}} \sup_{i=1, \dots, k_n} \left\|V_{\beta_0}(Z | t_i)^{-1/2}\right\| \|Z(t_i) - \mathcal{E}_{\beta_0}(Z | t_i)\| \geq \varepsilon\right). \end{aligned} \quad (3.50)$$

Or, pour $i = 1, \dots, k_n$,

$$\begin{aligned} \left\|V_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)^{-1/2}\right\| &= \left\|V_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)^{-1} V_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)^{1/2}\right\| \\ &\leq p \left\|V_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)^{-1}\right\| \left\|V_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)^{1/2}\right\| \\ &\leq C_V^{-1} \sqrt{2} p^4 L', \end{aligned} \quad (3.51)$$

où la dernière inégalité est obtenue par les équations (3.47) et (3.48). De plus,

$$\left\|Z(t_i) - E_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)\right\| \leq \|Z(t_i)\| + \left\|E_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)\right\| \leq 2L'.$$

L'inégalité (3.50) implique alors

$$P\left(\sup_{t \in [0,1]} \|U_n^*(\beta_0, t) - U_n^{**}(\beta_0, t)\| \geq \varepsilon\right) \leq P\left(\frac{2\sqrt{2}}{\sqrt{k_n}} p^5 C_V^{-1} (L')^2 \geq \varepsilon\right), \quad (3.52)$$

et $\lim_{n \rightarrow \infty} P\left(\sup_{t \in [0,1]} \|U_n^*(\beta_0, t) - U_n^{**}(\beta_0, t)\| \geq \varepsilon\right) = 0$. Ceci implique la convergence en probabilité de $U_n^*(\beta_0, \cdot) - U_n^{**}(\beta_0, \cdot)$ vers 0 lorsque $n \rightarrow \infty$ dans l'espace $(D[0,1], \mathbb{R}^p)$ muni de la topologie produit de Skorhokod d'après les résultats de la Section 2.1.2.

Second terme. La convergence en loi de X_n vers \mathcal{W}_p est obtenue par l'application du théorème central limite fonctionnel multivarié 2.2.4 introduit par Helland (1982). Avant d'appliquer ce théorème, vérifions ses hypothèses. Soit $i \in \{1, 2, \dots, k_n\}$ et $t \in [0, 1]$. Notons

$$\xi_{i, k_n} = (\xi_{i, k_n}^1, \dots, \xi_{i, k_n}^p) = \frac{1}{\sqrt{k_n}} V_{\beta_0}(Z(t_i) | \mathcal{F}_{t_i}^*)^{-1/2} \left\{Z(t_i) - E_{\beta(t_i)}(Z(t_i) | \mathcal{F}_{t_i}^*)\right\} \quad (3.53)$$

le i ème incrément dans \mathbb{R}^p du processus X_n . On a ainsi $X_n(t) = \sum_{i=1}^{\lfloor tk_n \rfloor} \xi_{i, k_n}$. Remarquons que

ξ_{i,k_n} est $\mathcal{F}_{t_i}^*$ -mesurable et $E_{\beta(t_i)} \left(\xi_{i,k_n} \middle| \mathcal{F}_{t_i}^* \right) = 0$. Soit $l, m \in \{1, \dots, p\}$, $l \neq m$. Notons e_l le l ème vecteur de la base canonique de \mathbb{R}^p : tous ses éléments sont nuls sauf le l ème qui vaut 1. Remarquons que $e_l^T e_m = 0$, $e_l^T e_l = 1$ et $\|e_l\| = \max_{i=1, \dots, p} (e_l)_i = 1$. Pour $i = 1, \dots, k_n$, nous avons

$$\xi_{i,k_n}^l = e_l^T \xi_{i,k_n} = \xi_{i,k_n}^T e_l \in \mathbb{R}. \quad (3.54)$$

a.' (Tableau de différences de martingales.) L'inclusion des σ -algèbres $\mathcal{F}_{t_{i-1}}^* \subset \mathcal{F}_{t_i}^*$ et le centrage des incréments implique

$$E_{\beta(t_{i-1})} \left(\xi_{i,k_n}^l \middle| \mathcal{F}_{t_{i-1}}^* \right) = E_{\beta(t_{i-1})} \left(E_{\beta(t_i)} \left(\xi_{i,k_n}^l \middle| \mathcal{F}_{t_i}^* \right) \middle| \mathcal{F}_{t_{i-1}}^* \right) = 0.$$

b.' (Non-corrélation.) En utilisant l'équation (3.54), nous avons

$$\begin{aligned} E_{\beta(t_i)} \left(\xi_{i,k_n}^l \xi_{i,k_n}^m \middle| \mathcal{F}_{t_i}^* \right) &= e_l^T E_{\beta(t_i)} \left(\xi_{i,k_n} \xi_{i,k_n}^T \middle| \mathcal{F}_{t_i}^* \right) e_m \\ &= \frac{1}{k_n} e_l^T V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2} V_{\beta(t_i)} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right) V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2} e_m \\ &= \frac{1}{k_n} e_l^T V \left(t_i^-, \beta_0, \beta(t_i) \right) e_m, \end{aligned}$$

où

$$V \left(t_i^-, \beta_0, \beta(t_i) \right) = V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2} V_{\beta(t_i)} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right) V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2}.$$

Notons I_p est la matrice identité de $\mathcal{M}_{p \times p}(\mathbb{R})$. Remarquons que $e_l^T I_p e_m = 0$. Nous avons

$$\begin{aligned} &E \left(\left| \sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(\xi_{i,k_n}^l \xi_{i,k_n}^m \middle| \mathcal{F}_{t_{i-1}}^* \right) \right| \right) \\ &= E \left(\left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(e_l^T V \left(t_i^-, \beta_0, \beta(t_i) \right) e_m \middle| \mathcal{F}_{t_{i-1}}^* \right) \right| \right) \\ &\leq \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} E \left(\left| e_l^T V \left(t_i^-, \beta_0, \beta(t_i) \right) e_m \right| \right) \\ &= \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} E \left(\left| e_l^T \left(V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right) e_m \right| \right) \\ &\leq \frac{\lfloor tk_n \rfloor}{k_n} \sup_{i=1, \dots, \lfloor tk_n \rfloor} E \left(\left| e_l^T \left(V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right) e_m \right| \right) \\ &\leq \frac{\lfloor tk_n \rfloor}{k_n} E \left(\sup_{i=1, \dots, \lfloor tk_n \rfloor} \left| e_l^T \left(V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right) e_m \right| \right) \\ &\leq p^2 \frac{\lfloor tk_n \rfloor}{k_n} \|e_l^T\| \|e_m\| E \left(\sup_{i=1, \dots, \lfloor tk_n \rfloor} \left\| V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right\| \right) \end{aligned}$$

$$\begin{aligned}
&= p^2 \frac{\lfloor tk_n \rfloor}{k_n} E \left(\sup_{i=1, \dots, \lfloor tk_n \rfloor} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \right) \\
&\leq p^2 \frac{\lfloor tk_n \rfloor}{k_n} E \left(\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \right). \tag{3.55}
\end{aligned}$$

Montrons que

$$\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \xrightarrow[n \rightarrow \infty]{L^1} 0.$$

Nous avons

$$\begin{aligned}
&\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \\
&= \sup_{i=1, \dots, k_n} \left\| V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*)^{-1/2} \left(V_{\beta(t_i)}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) - V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) \right) V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*)^{-1/2} \right\| \\
&\leq p^2 \left(\sup_{i=1, \dots, k_n} \left\| V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*)^{-1/2} \right\| \right)^2 \sup_{i=1, \dots, k_n} \left\| V_{\beta(t_i)}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) - V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) \right\|.
\end{aligned}$$

L'équation (3.51) implique que

$$\begin{aligned}
&\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \\
&\leq 2p^{10} C_V^{-2} (L')^2 \sup_{i=1, \dots, k_n} \left\| V_{\beta(t_i)}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) - V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) \right\| \\
&\leq 2p^{10} C_V^{-2} (L')^2 \left(\sup_{i=1, \dots, k_n} \left\| V_{\beta(t_i)}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) - \Sigma \right\| + \sup_{i=1, \dots, k_n} \left\| \Sigma - V_{\beta_0}(\mathcal{Z}(t_i) \mid \mathcal{F}_{t_i}^*) \right\| \right).
\end{aligned}$$

Finalement, la convergence de l'équation (3.21) implique la convergence en probabilité

$$\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Enfin, les équations (3.46) et (3.51) impliquent que $\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\|$ est majoré par une constante finie. Donc la convergence est également une convergence en moyenne :

$$\sup_{i=1, \dots, k_n} \left\| V(t_i^-, \beta_0, \beta(t_i)) - I_p \right\| \xrightarrow[n \rightarrow \infty]{L^1} 0. \tag{3.56}$$

L'équation (3.55) permet de conclure que

$$\sum_{i=1}^{\lfloor k_n t \rfloor} E_{\beta(t_{i-1})}(\xi_{i, k_n}^l \xi_{i, k_n}^m \mid \mathcal{F}_{t_{i-1}}^*) \xrightarrow[n \rightarrow \infty]{L^1} 0,$$

et la convergence a également lieu en probabilité.

c.' (Variance.) Des arguments similaires aux points (a.) et (b.) impliquent l'égalité

$$\begin{aligned} & \sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(\left(\xi_{i,k_n}^l \right)^2 \middle| \mathcal{F}_{t_{i-1}}^* \right) - t \\ &= \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(e_l^T \left\{ V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right\} e_l \middle| \mathcal{F}_{t_{i-1}}^* \right) + \left(\frac{\lfloor k_n t \rfloor}{k_n} - t \right). \end{aligned}$$

Par conséquent,

$$\begin{aligned} & E \left(\left| \sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(\left(\xi_{i,k_n}^l \right)^2 \middle| \mathcal{F}_{t_{i-1}}^* \right) - t \right| \right) \\ & \leq \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} E \left(\left| e_l^T \left\{ V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right\} e_l \right| \right) + \left| \frac{\lfloor k_n t \rfloor}{k_n} - t \right| \\ & \leq \frac{p^2}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} E \left(\left\| V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right\| \right) + \left| \frac{\lfloor k_n t \rfloor}{k_n} - t \right| \\ & \leq p^2 \frac{\lfloor k_n t \rfloor}{k_n} \sup_{i=1, \dots, k_n} E \left(\left\| V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right\| \right) + \left| \frac{\lfloor k_n t \rfloor}{k_n} - t \right| \\ & \leq p^2 \frac{\lfloor k_n t \rfloor}{k_n} E \left(\sup_{i=1, \dots, k_n} \left\| V \left(t_i^-, \beta_0, \beta(t_i) \right) - I_p \right\| \right) + \left| \frac{\lfloor k_n t \rfloor}{k_n} - t \right|. \end{aligned}$$

L'équation (3.56) permet de conclure que $\sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(\left(\xi_{i,k_n}^l \right)^2 \middle| \mathcal{F}_{t_{i-1}}^* \right) \xrightarrow[n \rightarrow \infty]{L^1} t$, impliquant la convergence en probabilité.

d.' (Condition de Lyapunov.) Nous avons $\left| \xi_{i,k_n}^l \right| \leq \max_{l=1, \dots, p} \left| \xi_{i,k_n}^l \right| = \|\xi_{i,k_n}\|$. Ainsi,

$$\begin{aligned} E \left(\sum_{i=1}^{\lfloor k_n t \rfloor} E_{\beta(t_{i-1})} \left(\left| \xi_{i,k_n}^l \right|^3 \middle| \mathcal{F}_{t_{i-1}}^* \right) \right) & \leq \sum_{i=1}^{\lfloor k_n t \rfloor} E \left(\left| \xi_{i,k_n}^l \right|^3 \right) \\ & \leq \sum_{i=1}^{\lfloor k_n t \rfloor} E \left(\|\xi_{i,k_n}\|^3 \right) \\ & \leq \sum_{i=1}^{k_n} E \left(\|\xi_{i,k_n}\|^3 \right). \end{aligned}$$

D'après l'équation (3.51) et la condition **H4'**, nous avons la majoration

$$\begin{aligned} & E \left(\sum_{i=1}^{\lfloor k_n t \rfloor} E_{\beta(t_{i-1})} \left(\left| \xi_{i,k_n}^l \right|^3 \middle| \mathcal{F}_{t_{i-1}}^* \right) \right) \\ & \leq \frac{p^3}{k_n^{3/2}} \sum_{i=1}^{k_n} E \left(\left\| V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i^-}^* \right)^{-1/2} \right\|^3 \left\| \mathcal{Z}(t_i) - E_{\beta(t_i)} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i^-}^* \right) \right\|^3 \right) \\ & \leq \frac{p^3}{k_n^{1/2}} \left(C_V^{-1} \sqrt{2} p^4 L' \right)^3 (2L')^3. \end{aligned}$$

Par conséquent, $\sum_{i=1}^{\lfloor tk_n \rfloor} E_{\beta(t_{i-1})} \left(\left| \xi_{i,k_n}^l \right|^3 \middle| \mathcal{F}_{t_{i-1}}^* \right)$ converge vers 0 en moyenne et donc en probabilité lorsque $n \rightarrow \infty$.

En conclusion, toutes les hypothèses du Théorème 2.2.4 sont réunies, ce qui permet de conclure que X_n converge vers \mathcal{W}_p lorsque $n \rightarrow \infty$ dans l'espace $(D[0,1], \mathbb{R}^p)$ muni de la topologie produit de Skorokhod. Par suite, d'après l'équation (3.49) et le résultat de convergence de son premier terme de droite (équation (3.52)), l'équation (3.23) est démontrée. Pour terminer la démonstration du théorème, il reste à prouver l'équation (3.24).

Soit $t \in [0,1]$. L'inégalité multidimensionnelle de Taylor-Lagrange (rappelée au Théorème 2.3.1) et l'équation (3.19) de la condition **H3'** impliquent que pour $s \in \{t_1, \dots, t_{k_n}\}$,

$$\left\| \mathcal{E}_{\beta(s)}(Z | s) - \mathcal{E}_{\beta_0}(Z | s) - \mathcal{V}_{\beta_0}(Z | s)(\beta(s) - \beta_0) \right\| \leq \frac{1}{2} M_n \|\beta(s) - \beta_0\|^2. \quad (3.57)$$

Par conséquent,

$$\begin{aligned} & \left\| B_n(t) - \Sigma^{1/2} \int_0^t \{\beta(s) - \beta_0\} ds \right\| \\ &= \left\| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \left\{ \mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\beta_0}(Z | t_i) \right\} - \Sigma^{1/2} \int_0^t \{\beta(s) - \beta_0\} ds \right\| \\ &\leq \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \left\{ \mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\beta_0}(Z | t_i) - \mathcal{V}_{\beta_0}(Z | t_i) \{\beta(t_i) - \beta_0\} \right\} \right\| \\ &\quad + \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \left(\mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right) \{\beta(t_i) - \beta_0\} \right\| \\ &\quad + \left\| \Sigma^{1/2} \left(\frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta(t_i) - \int_0^t \beta(s) ds + \beta_0 \left(\frac{\lfloor tk_n \rfloor}{k_n} - t \right) \right) \right\| \\ &\leq \frac{pM_n}{2k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \right\| \|\beta(t_i) - \beta_0\|^2 + \frac{p}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right\| \|\beta(t_i) - \beta_0\| \\ &\quad + p \left\| \Sigma^{1/2} \right\| \left\| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta(t_i) - \int_0^t \beta(s) ds \right\| + \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right| \left\| \Sigma^{1/2} \beta_0 \right\| \\ &\leq \frac{\lfloor tk_n \rfloor}{k_n} \frac{pM_n}{2} \sup_{i=1, \dots, \lfloor tk_n \rfloor} \|\beta(t_i) - \beta_0\|^2 \sup_{i=1, \dots, \lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \right\| \\ &\quad + p \frac{\lfloor tk_n \rfloor}{k_n} \sup_{i=1, \dots, \lfloor tk_n \rfloor} \|\beta(t_i) - \beta_0\| \sup_{i=1, \dots, \lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right\| \\ &\quad + p \left\| \Sigma^{1/2} \right\| \left\| \sup_{l=1, \dots, p} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta_l(t_i) - \int_0^t \beta_l(s) ds \right| + p \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right| \left\| \Sigma^{1/2} \right\| \|\beta_0\|. \end{aligned}$$

Or, β_0 et $0 \in \mathbb{B}'$ donc $\|\beta_0\| \leq 2\delta_2$ et $\sup_{i=1, \dots, \lfloor tk_n \rfloor} \|\beta(t_i) - \beta_0\| \leq \sup_{t \in [0,1]} \|\beta(t) - \beta_0\| \leq \delta_2$.

De plus, l'équation (3.51) implique que

$$\sup_{i=1,\dots,k_n} \left\| V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2} \right\| \leq \sqrt{2} C_V^{-1} p^4 L'.$$

Ainsi,

$$\begin{aligned} & \left\| B_n(t) - \Sigma^{1/2} \int_0^t \{ \beta(s) - \beta_0 \} ds \right\| \\ & \leq M_n \frac{\lfloor tk_n \rfloor}{k_n} \frac{p^5}{\sqrt{2}} \delta_2^2 C_V^{-1} L' + p \frac{\lfloor tk_n \rfloor}{k_n} \delta_2 \sup_{i=1,\dots,k_n} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right\| \\ & \quad + p \left\| \Sigma^{1/2} \right\| \sup_{l=1,\dots,p} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta_l(t_i) - \int_0^t \beta_l(s) ds \right| + 2\delta_2 p \left| \frac{\lfloor tk_n \rfloor}{k_n} - t \right| \left\| \Sigma^{1/2} \right\|. \end{aligned}$$

Par passage au sup, nous avons

$$\begin{aligned} & \sup_{t \in [0,1]} \left\| B_n(t) - \Sigma^{1/2} \int_0^t \{ \beta(s) - \beta_0 \} ds \right\| \\ & \leq M_n \frac{p^5}{\sqrt{2}} \delta_2^2 C_V^{-1} L' + p \delta_2 \sup_{i=1,\dots,k_n} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right\| \\ & \quad + p \left\| \Sigma^{1/2} \right\| \sup_{t \in [0,1]} \sup_{l=1,\dots,p} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta_l(t_i) - \int_0^t \beta_l(s) ds \right| + \frac{2}{k_n} \delta_2 p \left\| \Sigma^{1/2} \right\|. \end{aligned} \quad (3.58)$$

D'après l'hypothèse **H3'**, $\lim_{n \rightarrow \infty} M_n = 0$, et

$$\sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}, \gamma \in \mathbb{B}'} \left\| \mathcal{V}_{\gamma(t)}(Z | t)^{1/2} - \Sigma^{1/2} \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Soit $l = 1, \dots, p$. La fonction réelle β_l est bornée, donc l'équation (3.27) de la preuve du Lemme 3.2.1 s'applique et

$$\lim_{n \rightarrow \infty} \sup_{t \in [0,1]} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta_l(t_i) - \int_0^t \beta_l(s) ds \right| = 0.$$

Par conséquent,

$$\sup_{t \in [0,1]} \sup_{l=1,\dots,p} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta_l(t_i) - \int_0^t \beta_l(s) ds \right| \leq \sum_{l=1}^p \sup_{t \in [0,1]} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta_l(t_i) - \int_0^t \beta_l(s) ds \right|,$$

et comme p est fini,

$$\lim_{n \rightarrow \infty} \sup_{t \in [0,1]} \sup_{l=1,\dots,p} \left| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \beta(t_i)_l - \int_0^t \beta(s)_l ds \right| = 0.$$

D'après l'équation (3.58), la convergence de l'équation (3.24) est ainsi démontrée. \square

Chapitre 4

Tests de comparaison de courbes de survie

Dans ce chapitre, nous dérivons un test de la valeur du paramètre de régression du modèle à risques proportionnels. Un cas particulier apparaît lorsque la valeur à tester du paramètre est nulle et la covariable du modèle est catégorielle. Il s'agit alors de comparer des courbes de survie. Ces tests sont construits à partir du processus du score standardisé, en utilisant les résultats de convergence du Chapitre 3. Ces travaux sont détaillés dans un article paru dans le journal *Biometrika* (Chauvel et O'Quigley, 2014a).

4.1 Introduction

Lorsque le critère principal d'un essai clinique est le temps de survie des patients, on cherche à comparer les durées de vie des patients dans les différents groupes de traitement pour évaluer si le nouveau traitement améliore la survie ou non. Dans un premier temps, par souci de simplification, considérons le cas de deux groupes. Les patients du premier groupe reçoivent le nouveau traitement, indicé par T , et les patients du second groupe reçoivent le placebo, indicé par P . La comparaison des deux courbes de survie s'effectue à l'aide de tests de l'hypothèse nulle H_0 contre l'alternative H_1 suivantes

$$\begin{cases} H_0 : \forall t, S_P(t) = S_T(t), \\ H_1 : \exists t, S_P(t) \neq S_T(t). \end{cases}$$

En se plaçant sous le modèle à risques non proportionnels de paramètre $\beta(t)$, cela revient à considérer H_0 et H_1 telles que

$$\begin{cases} H_0 : \forall t, \beta(t) = 0, \\ H_1 : \exists t, \beta(t) \neq 0. \end{cases} \quad (4.1)$$

Nous nous concentrons sur l'étude des tests bilatéraux, mais il est envisageable d'utiliser des tests unilatéraux. La comparaison de courbes de survie peut également s'appliquer aux études

épidémiologiques, par exemple pour comparer la survie de groupes de patients exposés à différents facteurs environnementaux.

4.1.1 Test du log-rank

Pour détecter la présence d'une différence entre les courbes de survie, le test du log-rank (Mantel, 1966) est le plus utilisé. Ce test a été détaillé dans le chapitre introductif (Section 1.3). Le test peut également s'exprimer comme un test du score sous le modèle à risques proportionnels dans lequel la covariable est discrète (Cox, 1972). Notons $U(0)$ (respectivement $-\mathcal{I}(0)$) la première (respectivement deuxième) dérivée de la log-vraisemblance partielle de Cox par rapport au paramètre β évaluée avec le paramètre $\beta = 0$. Sur l'échelle du temps standardisé entre 0 et 1, ces quantités s'écrivent

$$U(0) = \int_0^1 \{Z(s) - \mathcal{E}_0(Z | s)\} d\bar{N}^*(s) = \sum_{j=1}^{k_n} \{Z(t_j) - \mathcal{E}_0(Z | t_j)\},$$

et

$$\mathcal{I}(0) = \int_0^1 \mathcal{V}_0(Z | s) d\bar{N}^*(s) = \sum_{i=1}^{k_n} \mathcal{V}_0(Z | t_i).$$

Par définition, le test de statistique $U(0)^2/\mathcal{I}(0)$ est un test du score pour tester $H_0 : \beta = 0$ dans le modèle à risques proportionnels de paramètre β .

Définition 4.1.1. *Le test du log-rank rejette H_0 avec un risque α si $|L_n| \geq z^{\alpha/2}$, où $z^{\alpha/2}$ est le quantile supérieur d'ordre $\alpha/2$ de la loi normale centrée réduite et*

$$L_n = \frac{U(0)}{\sqrt{\mathcal{I}(0)}} = \sum_{j=1}^{k_n} \frac{Z(t_j) - \mathcal{E}_0(Z | t_j)}{\sqrt{\sum_{i=1}^{k_n} \mathcal{V}_0(Z | t_i)}}, \quad (4.2)$$

est la statistique du test.

Remarque 4.1.1. *On définit de manière équivalente le test du log-rank avec L_n ou L_n^2 . Lorsque $\beta \in \mathbb{R}^p$, la statistique L_n^2 est définie par $L_n^2 = U(0)^T \mathcal{I}(0)^{-1} U(0)$ et converge vers la loi du χ^2 à $p-1$ degré de liberté sous l'hypothèse nulle.*

Avec cette écriture, on peut étendre le test du log-rank à des variables continues, pour tester la présence d'un effet de ces variables (par exemple, l'âge du patient) sur la survie. La statistique du test est facile à mettre en oeuvre numériquement et la p-value exacte du test peut-être évaluée puisque la loi asymptotique est connue. La forte popularité du test s'explique notamment par le fait que ce test est le plus puissant (Peto et Peto, 1972) sous des hypothèses alternatives de type risques proportionnels, qui s'écrivent

$$H_{1,PH} : \exists \beta_0 \neq 0, \forall t, \beta(t) = \beta_0.$$

L'indice *PH* est l'acronyme du terme anglais "Proportional Hazards". Ce test est puissant pour détecter la présence d'un effet β constant dans le temps entre les groupes, sous le modèle

à risques proportionnels. En revanche, en présence de risques non proportionnels, c'est-à-dire d'un effet $\beta(t)$ qui varie dans le temps, le test du log-rank n'est plus optimal au sens où sa puissance peut fortement diminuer. C'est notamment le cas en présence de risques qui se croisent obtenus lorsque le coefficient change de signe (Leurgans, 1983, 1984). Formellement, l'alternative de type risques non proportionnels ("Non-Proportional Hazards" en anglais) s'écrit

$$H_{1,NPH} : \forall \beta_0 \neq 0, \exists t_0, \beta(t_0) \neq \beta_0.$$

Pour gagner de la puissance sous $H_{1,NPH}$, des tests ont été proposés, et notamment les tests du log-rank pondéré.

4.1.2 Tests du log-rank pondéré

Le test du log-rank pondéré a pour statistique LW_n telle que

$$LW_n = \sum_{j=1}^{k_n} W_n(t_j) \frac{Z(t_j) - \mathcal{E}_0(Z | t_j)}{\sqrt{\sum_{i=1}^{k_n} W_n(t_i)^2 \mathcal{V}_0(Z | t_i)}}, \quad (4.3)$$

où W_n est une fonction de poids positive, définie sur $[0, 1]$ et \mathcal{F}_t^* -prévisible. Le choix des poids permet de détecter différents types d'effet temporel $\beta(t)$. Nous rappelons que l'application ϕ_n^{-1} définie sur $[0, 1]$, permet de ramener le temps $t \in [0, 1]$ sur l'échelle des temps initiale (voir Section 3.2). On peut considérer, par exemple, les poids suivants :

- $W_n(t) = \sum_{i=1}^n Y_i(\phi_n^{-1}(t))$, le nombre d'individus à risque au temps t (Breslow, 1970, Gehan, 1965),
- $W_n(t) = \hat{S}(\phi_n^{-1}(t))$, l'estimateur de Kaplan-Meier de la survie au temps t (Peto et Peto, 1972),
- $W_n(t) = \tilde{S}(\phi_n^{-1}(t))$, où \tilde{S} est une modification de l'estimateur de Kaplan-Meier (Prentice, 1978),
- $W_n(t) = g\{\sum_{i=1}^n Y_i(\phi_n^{-1}(t))\}$, pour une fonction g donnée (par exemple, $g(x) = \sqrt{x}$, $\forall x$) (Tarone et Ware, 1977),
- $W_n(t) = \hat{S}(\phi_n^{-1}(t))^p \{1 - \hat{S}(\phi_n^{-1}(t))\}^q$, $p, q \geq 0$ des poids de la famille de Fleming et Harrington (1991, 2005).

Beaucoup d'autres poids ont été proposés (Lagakos *et al.*, 1990, Mantel et Stablein, 1988, Wu et Gilbert, 2002). Notons que le test avec la pondération de Gehan (1965) est une modification du test de Wilcoxon (1945) prenant en compte les données censurées. Ce test est équivalent au test de Mann-Whitney pour la comparaison de deux groupes et au test de Kruskal-Wallis en présence d'un nombre de groupes plus important (Breslow, 1970).

Aalen (1978) et Gill (1980) ont montré la convergence des statistiques du log-rank pondéré vers une loi normale centrée. Jones et Crowley (1989, 1990) ont défini une classe plus large de tests dont le log-rank, pondéré ou non, fait partie. Cette classe recouvre également des tests de comparaison de plus de deux courbes de survie, ainsi que des tests de tendance des risques instantanés.

Tous ces tests du log-rank pondéré ont été développés sur l'échelle du temps initiale. Cependant, s'agissant de tests de rangs, il est équivalent de les définir sur l'échelle du temps transformée dans $[0, 1]$ par la transformation ϕ_n de l'équation (3.6). Les poids sont souvent plus grands pour les temps initiaux, permettant ainsi de détecter des effets précoces, c'est-à-dire un coefficient $\beta(t)$ qui diminue dans le temps. En effet, ils pondèrent plus fortement les contributions précoces à la statistique LW_n et détectent plus facilement la présence d'un effet au début de l'étude. Cependant, les pondérations de Tarone et Ware (1977) et Fleming et Harrington (1991) définissent des classes de poids pour lesquelles un bon choix de fonction g ou de couple (p, q) permettent de détecter des effets tardifs, c'est-à-dire un coefficient $\beta(t)$ qui augmente avec le temps. Le choix optimal de fonction g ou du couple (p, q) dépend des données et peut-être délicat. Peckova et Fleming (2003) ont proposé un test du log-rank pondéré dont le poids, choisi de manière adaptative en fonction des données, appartient à une classe prédéfinie et est tel que l'efficacité du test log-rank pondéré est maximale.

Alors que ces tests sont puissants sous des alternatives de type risques non proportionnels $H_{1,NPH}$, ils perdent de la puissance par rapport au log-rank non pondéré sous des alternatives de type risques proportionnels $H_{1,PH}$. De plus, pour appliquer des poids convenables, si l'alternative est de type risques non proportionnels, il faut connaître le sens de variation de l'effet (effets précoces, effets tardifs, effets plus importants au milieu de l'étude, etc.). Yang et Prentice (2010) ont proposé un test adaptatif qui se réduit au test du log-rank lorsque les risques sont proportionnels. Ce test est basé sur la classe de modèles de Yang et Prentice (2005) défini à l'équation (1.20), dont le modèle à risques proportionnels et le modèle à odds proportionnels font partie (voir Section 1.2.3). La statistique du test est alors le maximum entre les valeurs absolues des deux statistiques du log-rank pondéré $LW_{1,n}$ et $LW_{2,n}$ de poids respectifs $W_n^1(t) = \hat{\lambda}_T(t)/\hat{\lambda}_P(t)$ et $W_n^2(t) = \hat{\lambda}_P(t)/\hat{\lambda}_T(t)$, où les estimateurs des risques instantanés dans les groupes qui reçoivent le placebo, $\hat{\lambda}_P$ ou le traitement, $\hat{\lambda}_T$ sont basés sur le modèle (1.20).

De nombreux ouvrages mentionnent les tests du log-rank et du log-rank pondéré, nous renvoyons à l'un d'entre eux pour plus de renseignements (Andersen *et al.*, 1993, Balakrishnan et Rao, 2003, Klein et Moeschberger, 2003).

4.1.3 Combinaison de tests

La réglementation des essais cliniques impose la rédaction du plan d'analyse statistique et, par conséquent, le choix du test à appliquer avant de collecter les données. Il est souvent difficile d'anticiper si l'effet du groupe de traitement sur le temps de survie est constant ou s'il change avec le temps, et, dans ce dernier cas, de connaître son comportement temporel. Le test du

log-rank est le plus couramment utilisé, de par son bon comportement sous une alternative de type risques proportionnels. Cependant, face à une situation de risques non proportionnels, la faible puissance du test peut empêcher la détection de l'effet du groupe sur la survie. Pour ces raisons, développer un test qui se comporte comme le log-rank sous $H_{1,PH}$ et qui a une bonne puissance sous $H_{1,NPH}$ semble judicieux. Plusieurs possibilités ont été envisagées.

Combinaison de tests. La première possibilité consiste à combiner plusieurs statistiques du log-rank pondéré. Il peut s'agir d'une combinaison linéaire. Par exemple, pour détecter la présence d'un effet qui apparaît après un décalage dans le temps, [Zucker et Lakatos \(1990\)](#) ont étudié un test dont la statistique est la somme de la statistique du log-rank et de celle d'un test du log-rank pondéré puissant pour détecter les effets tardifs. [Lee \(1996\)](#) a considéré une combinaison linéaire de 4 statistiques du log-rank pondéré avec des poids dans la famille de [Fleming et Harrington \(1991\)](#). En notant $G^{p,q}$ la statistique du log-rank pondéré de poids $\hat{S}(t)^p(1 - \hat{S}(t))^q$, $p, q \geq 0$, les 4 statistiques considérées par Lee sont

- $G^{0,0} = L_n$: pour détecter des effets constants (risques proportionnels),
- $G^{2,0}$: pour détecter des effets précoces,
- $G^{0,2}$: pour détecter des effets tardifs,
- $G^{2,2}$: pour détecter des effets au milieu de l'étude.

Dans la même optique, [Wu et Gilbert \(2002\)](#) ont proposé une combinaison linéaire de plusieurs statistiques du log-rank pondéré avec des poids de la forme

$$W_n^a(t) = \left\{ \hat{S}(t) - (a\hat{S}(\tau_0) + 1 - a) \right\}^2, \quad (4.4)$$

où \hat{S} est l'estimateur de Kaplan-Meier de la survie, τ_0 est la durée maximale de suivi et a est un paramètre compris entre 0 et 1 permettant de détecter différents types d'effets dépendants du temps. L'inconvénient de toutes ces procédures linéaires est de choisir les poids optimaux à donner aux différentes statistiques. Un mauvais choix de poids pourrait entraîner une perte de puissance importante, même dans une situation normalement couverte par une des statistiques de la combinaison ([Lee, 1996](#)). En revanche, ces tests sont intéressants si l'investigateur a une idée a priori de la forme de l'effet.

Une autre possibilité pour combiner plusieurs statistiques est de retenir la statistique maximale en valeur absolue. [Tarone \(1981\)](#) a étudié le test dont la statistique est le maximum entre le log-rank et le log-rank pondéré avec les poids introduits par [Gehan \(1965\)](#). [Self \(1991\)](#) décrit le test du supremum sur la classe des statistiques du log-rank pondéré avec des poids de la forme $t^{\theta_1}(1 - t)^{\theta_2}$, $\theta_1, \theta_2 \geq 0$. [Fleming et Harrington \(1991\)](#) étudient les propriétés du test dont la statistique est le maximum entre plusieurs statistiques $G^{p,q}$. [Lee \(1996\)](#) a étudié ce test par simulation en considérant la classe des 4 statistiques $G^{0,0}$, $G^{2,0}$, $G^{0,2}$ et $G^{2,2}$. [Wu et Gilbert \(2002\)](#) ont également étudié les propriétés théoriques d'un test ayant pour statistique

le maximum entre plusieurs statistiques du log-rank pondéré de poids $W_n^a(t)$ (équation (4.4)). [Kosorok et Lin \(1999\)](#) considèrent le supremum sur une classe de statistiques indexées par des fonctions, étendant les résultats de [Jones et Crowley \(1990\)](#) au cas de plusieurs covariables. Enfin, le dernier exemple que nous citons est le travail de [Breslow et al. \(1984\)](#) qui considèrent le maximum entre la statistique du log-rank et la statistique du score associée à une covariable pour tester l'hypothèse nulle d'absence de l'effet groupe et de l'effet de la covariable.

Supremum sur le temps. Plutôt que de considérer les statistiques précédentes au temps final, une idée est de travailler avec leur supremum sur le temps. La statistique du log-rank est alors considérée comme un processus, qui retrace l'historique de la force de l'effet. Si l'effet dépend du temps, un écart à l'hypothèse nulle pourra être plus facilement détecté. L'introduction de poids, dont le choix est toujours délicat, est possible, mais non nécessaire pour détecter un effet qui dépend du temps. Ces tests sont de type Rényi, c'est-à-dire une adaptation du test de Kolmogorov-Smirnov pour comparer deux distributions de données censurées. Leur comportement asymptotique ainsi que sur des petits échantillons a été étudié par plusieurs auteurs, théoriquement ou numériquement ([Fleming et Harrington, 1991, 1984](#), [Fleming et al., 1987, 1980](#), [Gill, 1980](#), [Schumacher, 1984](#)). [Eng et Kosorok \(2005\)](#) donnent la formule du nombre de sujets nécessaires pour le test du supremum sur le temps. Notons que [Schumacher \(1984\)](#) s'est également intéressé aux statistiques de type Cramér-von Mises. [Kosorok et Lin \(1999\)](#) ont contribué à l'étude des tests du supremum sur le temps. Ils se sont également intéressés à la statistique du supremum plus de l'infimum sur le temps, produisant un test puissant sous des alternatives ordonnées, par exemple $H_1 : S_P(t) < S_T(t), \forall t$.

Statistiques de Kaplan-Meier pondérées. D'autres tests que le test du log-rank et ses dérivés ont été introduits pour comparer des courbes de survie. [Pepe et Fleming \(1989\)](#) ont proposé une classe de tests basés sur les différences entre les estimateurs de Kaplan-Meier des courbes de survie. Ces différences peuvent être pondérées ou non. [Shen et Cai \(2001\)](#) ont proposé une version adaptative de ces tests, dont la statistique est le maximum entre plusieurs statistiques de la classe, avec des poids différents. [Lee \(2011\)](#) a étudié le comportement asymptotique de ces tests par des simulations. Cependant, ces tests ne sont pas des tests de rangs, ils sont sensibles à une transformation monotone croissante sur les temps. De plus, même s'ils sont puissants sous des hypothèses alternatives avec des fonctions de survie ordonnées, par exemple, $H_1 : S_T(t) > S_P(t), \forall t$, leur puissance est faible lorsque les risques se croisent. Pour ces deux raisons, ce type de tests n'est pas retenu dans la suite de notre travail.

Dans la section suivante, nous introduisons une classe de tests construite à partir du processus du score standardisé U_n^* . Plusieurs tests sont étudiés, chacun basé sur des caractéristiques différentes du processus. Nous montrons en particulier qu'un des tests, le test de la distance depuis l'origine, est asymptotiquement équivalent au log-rank. Ces tests sont comparés par un travail de simulation en Section 4.3 avant d'être appliqués sur des données réelles (Section 4.4). Une discussion et des perspectives terminent ce chapitre.

4.2 Tests basés sur le processus du score standardisé

Le comportement limite du processus du score standardisé U_n^* , donné par les Théorèmes 3.3.1 et 3.3.2, permet de construire des tests d'hypothèse sur la valeur du paramètre de régression $\beta(t)$. Les hypothèses nulle H_0 et alternative H_1 que nous voulons tester sont

$$\begin{cases} H_0 : \forall t, \beta(t) = \beta_0, \\ H_1 : \exists t, \beta(t) \neq \beta_0, \end{cases}$$

avec β_0 fixé. Sous H_0 , la dérive du processus du score standardisé $U_n^*(\beta_0, \cdot)$ évalué au paramètre β_0 s'annule et le processus converge vers un mouvement Brownien standard. On peut alors baser des tests sur les propriétés du mouvement Brownien. Lorsque $\beta_0 = 0$, cela revient à tester l'absence d'effet entre les différentes valeurs de la covariable Z . En particulier, si la covariable est catégorielle et représente des groupes de traitement différents, on testera alors la présence d'un effet du traitement sur la survie. Le cas $\beta_0 = 0$ correspond ainsi aux hypothèses nulle et alternative (4.1) du test du log-rank. Plusieurs possibilités sont mentionnées dans l'ouvrage d'O'Quigley (2008), par exemple les tests de la distance depuis l'origine au temps t , de la distance maximale parcourue pendant un intervalle de temps, du mouvement Brownien intégré, du supremum d'un pont Brownien, du mouvement Brownien reflété ou encore de l'arcsine. Nous détaillons d'abord les deux premiers tests cités, avant d'introduire une nouvelle proposition. Dans les sections 4.2.1 jusqu'à 4.2.3, nous travaillons avec le cas univarié d'une covariable et d'un seul effet à tester ($p = 1$), avant d'étendre ces tests pour plusieurs coefficients ($p > 1$) en Section 4.2.4.

4.2.1 Distance depuis l'origine

On se place sous le modèle à risques proportionnels de paramètre β_0 fixé, c'est-à-dire sous l'hypothèse nulle $H_0 : \beta(t) = \beta_0$. Dans cette section et jusqu'à la Section 4.2.3, on considère les hypothèses **H1.**, **H2.**, **H3.** et **H4.** vérifiées et $\beta, \beta_0 \in \mathbb{B}$. Le Théorème 3.3.1 implique le corollaire suivant.

Corollaire 4.2.1. *Pour tout $t \in [0, 1]$, sous l'hypothèse H_0 ,*

$$\frac{1}{\sqrt{t}} U_n^*(\beta_0, t) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (4.5)$$

Ce résultat permet de définir le test de la distance depuis l'origine au temps t .

Définition 4.2.1. *Soit $t \in [0, 1]$. La statistique du test de la distance depuis l'origine au temps t est $U_n^*(\beta_0, t)$. Le test de la distance depuis l'origine au temps t rejette H_0 au niveau asymptotique α si*

$$\frac{1}{\sqrt{t}} |U_n^*(\beta_0, t)| \geq z^{\alpha/2},$$

où $z^{\alpha/2}$ est le quantile supérieur d'ordre $\alpha/2$ de la loi normale centrée réduite. La p -value

asymptotique de ce test est

$$2 \left[1 - \Phi \left(\frac{1}{\sqrt{t}} |U_n^*(\beta_0, t)| \right) \right],$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

La proposition suivante indique que la statistique du test de la distance est une statistique convergente. Ainsi, plus l'échantillon est grand, meilleure est la détection de l'hypothèse alternative.

Proposition 4.2.1. *Sous l'hypothèse alternative $H_1 : \beta(t) \neq \beta_0$, c'est-à-dire sous le modèle à risques non proportionnels de paramètre $\beta(t) \neq \beta_0$, on a*

$$\lim_{n \rightarrow \infty} P \left(\frac{1}{\sqrt{t}} |U_n^*(\beta_0, t)| \geq z^{\alpha/2} \right) = 1.$$

Preuve de la Proposition 4.2.1. Voir la preuve de la Proposition 4.2.4. □

Remarque 4.2.1. *Le test peut être appliqué à n'importe quel temps ce qui le rend flexible. Par exemple, si l'on sait a priori que l'effet disparaît à partir d'un certain temps τ , on peut appliquer le test de la distance depuis l'origine au temps τ . Lorsque l'on ne dispose pas d'une telle information, on considère $t = 1$ de manière à maximiser la puissance. Dans la suite, on considèrera toujours $t = 1$ et on parlera du "test de la distance".*

La statistique du test de la distance est la dernière valeur observée du processus du score standardisé. Cette valeur peut donc être visualisée sur le graphique du processus $U_n^*(\beta_0, \cdot)$ en fonction du temps. Ce test, qui contrôle asymptotiquement le niveau à α , est puissant sous des hypothèses alternatives de type risques proportionnels d'après le résultat suivant.

Proposition 4.2.2. *En notant L_n la statistique du log-rank et $U_n^*(0, 1)$ la statistique de la distance $U_n^*(\beta_0, 1)$ évaluée au paramètre $\beta_0 = 0$, on a*

$$|U_n^*(0, 1) - L_n| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Preuve de la Proposition 4.2.2. La preuve suit celle du Théorème 3.3.2. La statistique L_n est décomposée asymptotiquement en un mouvement Brownien plus une dérive. L'hypothèse **H3**. d'homoscédasticité implique que la décomposition est la même que celle du processus du score standardisé $U_n^*(0, 1)$ évalué au paramètre 0 et au temps $t = 1$, impliquant le résultat. □

Sous les conditions que nous avons introduites, la statistique du test de la distance converge vers celle du log-rank, garantissant une puissance maximale sous des hypothèses alternatives de type risques proportionnels. Ce résultat n'est pas surprenant, car l'écriture de la statistique du log-rank L_n (équation (4.2)) et celle du test de la distance (équation (3.11)) ne diffèrent que par la standardisation adoptée. Pour la première, la standardisation s'effectue globalement pour tous les temps à la fin de l'expérience, alors que pour la seconde, la standardisation a lieu à chaque temps de décès. Or, sous la condition **H3**. d'homoscédasticité, la standardisation ne

change pas avec le temps. Rappelons que cette condition, que nous avons justifiée en Section 3.3.2, a été utilisée par d'autres auteurs (Grambsch et Therneau, 1994, Xu, 1996). Le résultat de la Proposition 4.2.2 implique également que le test de la distance n'est pas optimal sous des alternatives $H_{1,NPH}$, héritant des propriétés de puissance du test du log-rank. Il n'existe pas d'équivalent du log-rank pondéré (équation (4.3)) avec la statistique de la distance puisque les poids introduits au numérateur et au dénominateur du processus du score standardisé s'annulent (voir Remarque 3.3.2).

Dans la section suivante, nous étudions un deuxième test basé sur le processus du score standardisé U_n^* et sur les propriétés du mouvement Brownien.

4.2.2 Aire sous la courbe

Définissons l'aire sous la courbe du processus du score standardisé au temps t .

Définition 4.2.2. *Pour tout $\gamma \in \mathbb{B}$, le processus de l'aire sous la courbe du score standardisé au temps t , noté $\{J_n(\gamma, t), 0 \leq t \leq 1\}$ est défini par*

$$J_n(\gamma, t) = \int_0^t U_n^*(\gamma, u) du, \quad 0 \leq t \leq 1.$$

Le Théorème 3.3.1 et les propriétés du mouvement Brownien (Bhattacharya et Waymire, 1990) permettent d'établir le résultat suivant :

Proposition 4.2.3. *Sous le modèle à risques proportionnels de paramètre β_0 , le processus de l'aire sous la courbe du score standardisé converge en distribution vers un mouvement Brownien intégré :*

$$J_n(\beta_0, t) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \int_0^t \mathcal{W}(s) ds, \quad t > 0.$$

De plus, la covariance du processus est telle que

$$\lim_{n \rightarrow \infty} \text{Cov} \{J_n(\beta_0, s), J_n(\beta_0, t)\} = s^2 \left(\frac{t}{2} - \frac{s}{6} \right), \quad 0 \leq s \leq t \leq 1. \quad (4.6)$$

Nous rappelons que la variable aléatoire $\int_0^t \mathcal{W}(s) ds$ est distribuée selon une loi normale centrée de variance $t^3/3$. Cette proposition permet de définir le test de l'aire sous la courbe du processus du score au temps t .

Définition 4.2.3. *Soit $t \in [0, 1]$. La statistique de l'aire sous la courbe du processus du score standardisé au temps t est $J_n(\beta_0, t)$. Le test de l'aire sous la courbe au temps t rejette H_0 au niveau asymptotique α si*

$$\left(3t^{-3}\right)^{1/2} |J_n(\beta_0, t)| \geq z^{\alpha/2}.$$

La p -value asymptotique du test est

$$2 \left[1 - \Phi \left(\left(3t^{-3}\right)^{1/2} |J_n(\beta_0, t)| \right) \right].$$

Remarque 4.2.2. *Les mêmes observations qu'à la Remarque 4.2.1 peuvent être étendues au test de l'aire sous la courbe du processus du score standardisé. Nous considérons dorénavant $t = 1$ et, pour ne pas alourdir la rédaction, appelons ce test le "test de l'aire sous la courbe".*

La proposition suivante indique que la statistique du test de l'aire sous la courbe est une statistique convergente.

Proposition 4.2.4. *Sous l'hypothèse alternative $H_1 : \beta(t) \neq \beta_0$, c'est-à-dire sous le modèle à risques non proportionnels de paramètre $\beta(t) \neq \beta_0$, on a*

$$\lim_{n \rightarrow \infty} P \left(\left((3t^{-3})^{1/2} |J_n(\beta_0, t)| \geq z^{\alpha/2} \right) \right) = 1.$$

Preuves des Propositions 4.2.1 et 4.2.4. Rappelons le résultat du Théorème 3.3.2 :

$$U_n^*(\beta_0, t) - \sqrt{k_n} A_n(t) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} C_1(\beta, \beta_0) \mathcal{W}(t), \quad 0 \leq t \leq 1,$$

où

$$A_n(t) \xrightarrow[n \rightarrow \infty]{P} C_2 \int_0^t \{\beta(s) - \beta_0\} ds, \quad 0 \leq t \leq 1.$$

Soit $t \in [0, 1]$. Les fonctions $U_n^*(\beta_0, \cdot)$ et A_n sont bornées presque sûrement, donc

$$J_n(\beta_0, t) - \sqrt{k_n} \int_0^t A_n(s) ds = \int_0^t U_n^*(\beta_0, s) ds - \sqrt{k_n} \int_0^t A_n(s) ds \xrightarrow[n \rightarrow \infty]{\mathcal{L}} C_1(\beta, \beta_0) \int_0^t \mathcal{W}(s) ds.$$

Par conséquent, nous avons les convergences en probabilité

$$\frac{1}{\sqrt{k_n}} U_n^*(\beta_0, t) \xrightarrow[n \rightarrow \infty]{P} C_2 \int_0^t \{\beta(s) - \beta_0\} ds,$$

et

$$\frac{1}{\sqrt{k_n}} J_n(\beta_0, t) \xrightarrow[n \rightarrow \infty]{P} C_2 \int_0^t \int_0^s (\beta(u) - \beta_0) du ds.$$

On rappelle que $\lim_{n \rightarrow \infty} \sqrt{k_n} = \infty$, et pour tout $s \in [0, 1]$, $\int_0^s \{\beta(u) - \beta_0\} du \neq 0$, ce qui entraîne les convergences

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left(\frac{1}{\sqrt{k_n}} |U_n^*(\beta_0, t)| \geq z^{\alpha/2} \right) &= 1, \\ \lim_{n \rightarrow \infty} P \left(\left((3t^{-3})^{1/2} |J_n(\beta_0, t)| \geq z^{\alpha/2} \right) \right) &= 1. \end{aligned}$$

□

Intuitivement, sous des alternatives de type risques proportionnels, le test de la distance a une puissance légèrement plus grande que le test de l'aire sous la courbe puisqu'il est asymptotiquement équivalent au test du log-rank. En revanche, lorsque l'effet $\beta(t)$ diminue dans le temps, le test de l'aire sous la courbe devient le plus puissant des deux. Cela peut être vu sur un exemple en considérant la situation d'un modèle de rupture dans lequel l'effet est

constant jusqu'à un certain temps τ et s'annule après (voir Figure 3.2(a)). Jusqu'au temps τ la dérive du processus $U_n^*(\beta_0, \cdot)$ est linéaire de pente strictement positive, mais après τ , la dérive est parallèle à l'axe du temps. Un test basé sur la distance depuis l'origine ne prendra en compte que la dernière valeur du processus, sans tenir compte de la force de l'effet avant τ . Ceci amoindrira la puissance du test de la distance par rapport à celle du test de l'aire sous la courbe dans lequel toute l'information collectée dans l'historique du processus est conservée.

Remarque 4.2.3. *D'après les Propositions 4.2.1 et 4.2.4, les statistiques de la distance et de l'aire sous la courbe sont toutes les deux convergentes. Ainsi, il est impossible de comparer les fonctions puissance des tests correspondants, qui tendent toutes les deux vers 1 lorsque la taille de l'échantillon tend vers l'infini. Une comparaison numérique des puissances des tests pour des échantillons de taille finie sera effectuée par un travail de simulation en Section 4.3.*

Les deux statistiques de test sont corrélées, d'après le lemme suivant (Ross, 1996) :

Lemme 4.2.1. *Soit $t \in [0, 1]$. Sous le modèle à risques proportionnels de paramètre β_0 , la fonction de covariance de $J_n(\beta_0, t)$ et $U_n^*(\beta_0, t)$ est telle que*

$$\text{Cov}\{J_n(\beta_0, t), U_n^*(\beta_0, t)\} \xrightarrow[n \rightarrow \infty]{P} t^2/2.$$

Un test basé sur une combinaison des deux tests précédents aura une bonne puissance sous les deux types d'alternatives (risques proportionnels ou non). Dans la section suivante, nous définissons une classe de tests adaptatifs combinant les statistiques de la distance $U_n^*(\beta_0, t)$ et de l'aire sous la courbe $J_n(\beta_0, t)$. La statistique résultante prend la valeur de l'une ou l'autre statistique. La classe est adaptative au sens où l'utilisateur ne choisit pas directement quelle statistique appliquer, cette dernière est choisie selon les données et un paramètre fixé par l'utilisateur. Nous proposons également un test adaptatif restreint qui ne dépend d'aucun paramètre.

4.2.3 Tests adaptatifs restreints

L'idée d'un test adaptatif restreint est de limiter la classe des possibilités considérées, en favorisant ou non un membre particulier de la classe. L'utilisateur peut préférer utiliser le test du log-rank, à moins que des indices forts de la non-proportionnalité des risques soient présents. Dans ce cas, un autre test peut être préféré. On peut introduire un paramètre γ permettant de favoriser un test ou l'autre. Dans notre cas, une large valeur de γ favorise le test du log-rank, résultant en une perte de puissance négligeable sous l'hypothèse des risques proportionnels comparée à l'application directe du test du log-rank. La contrepartie est la faible puissance de détection des effets dépendants du temps. De plus faibles valeurs du paramètre γ permettent de mieux détecter des effets lorsque les risques ne sont pas proportionnels.

Sous des alternatives de type risques proportionnels $H_{1,PH}$, le processus $U_n^*(\beta_0, \cdot)$ peut être approché par un mouvement Brownien avec une dérive linéaire d'après le Théorème 3.3.2. Ainsi, au temps t , deux fois le processus intégré $J_n(\beta_0, t)$ devrait être proche de la distance

depuis l'origine $U_n^*(\beta_0, t)$. Notons $\Delta_n(\beta_0, t)$ la différence entre ces deux processus

$$\Delta_n(\beta_0, t) = 2J_n(\beta_0, t) - U_n^*(\beta_0, t).$$

Si la valeur absolue de $\Delta_n(\beta_0, t)$ est assez petite, on peut s'attendre à des risques proportionnels et le test de la distance ou, de manière équivalente, le test du log-rank (Proposition 4.2.2), sera puissant. Sinon, on appliquera le test de l'aire sous la courbe. On peut ainsi définir une classe de tests adaptatifs restreints dépendants du seuil $\gamma \geq 0$.

Définition 4.2.4. La classe de statistiques des tests adaptatifs restreints au temps t est notée $\{M_n^\gamma(\beta_0, t), \gamma \geq 0\}$, où, pour $\gamma \geq 0$,

$$M_n^\gamma(\beta_0, t) = \left\{ \sqrt{3} |J_n(\beta_0, t)| - |U_n^*(\beta_0, t)| \right\} \mathbf{1}_{|\Delta_n(\beta_0, t)| \geq \gamma + |U_n^*(\beta_0, t)|}.$$

Remarque 4.2.4. Le temps t est choisi de la même manière qu'aux Remarques 4.2.1 et 4.2.2. On posera dans la suite $t = 1$.

Remarque 4.2.5. La p -value du test peut être évaluée numériquement, en simulant N couples indépendants $(X_1, Y_1), \dots, (X_N, Y_N)$ de loi normale dans \mathbb{R}^2 centrée et de matrice de variance-covariance

$$\begin{pmatrix} t & t^2/2 \\ t^2/2 & t^3/3 \end{pmatrix}.$$

La p -value peut alors être estimée par

$$\frac{1}{N} \sum_{i=1}^N \mathbf{1}_{f(X_i, Y_i) \geq M_n^\gamma(\beta_0, t)},$$

où, pour $i = 1, \dots, N$,

$$f(X_i, Y_i) = \left\{ \sqrt{3} |X_i| - |Y_i| \right\} \mathbf{1}_{|2X_i - Y_i| \geq \gamma + |Y_i|}.$$

Dans les simulations de la Section 4.3, nous avons considéré $N = 10^6$.

Quand $\gamma = 0$, $M_n^\gamma(\beta_0, t)$ est la valeur absolue de l'aire sous la courbe. A l'autre extrême, quand $\gamma \rightarrow \infty$, $M_n^\gamma(\beta_0, t)$ vaut la valeur absolue de la distance depuis l'origine. Entre ces deux situations, $M_n^\gamma(\beta_0, t)$ prend la valeur absolue de l'une ou l'autre statistique. Le choix optimal de γ dépend de la forme inconnue de l'alternative, $H_{1,PH}$ ou $H_{1,NPH}$. L'utilisateur peut vouloir travailler avec une grande valeur de γ , ce qui revient essentiellement à appliquer le test du log-rank en gardant un peu de puissance en réserve pour des alternatives non anticipées dans lesquelles l'effet change fortement avec le temps (par exemple, s'il change de signe). Cependant, ce test sera moins puissant que le log-rank pour détecter la présence d'un effet constant non nul. Si l'utilisateur n'a pas d'idée a priori sur le type d'alternative, plutôt que d'estimer le paramètre γ optimal et de perdre de la puissance pour détecter l'hypothèse alternative, nous suggérons d'utiliser le test suivant.

Définition 4.2.5. *La statistique du test adaptatif restreint est définie par*

$$M_n(\beta_0, t) = \max \left\{ |U_n^*(\beta_0, t)|, \sqrt{3} |J_n(\beta_0, t)| \right\}.$$

Proposition 4.2.5. *Notons $\phi\{\cdot; 0, \Sigma(t)\}$ la densité de la loi normale centrée dans \mathbb{R}^2 de matrice de variance-covariance*

$$\Sigma(t) = \begin{pmatrix} t & \sqrt{3}t^2/2 \\ \sqrt{3}t^2/2 & t^3 \end{pmatrix}.$$

Le test adaptatif restreint rejette H_0 au niveau asymptotique α si

$$M_n(\beta_0, t) \geq q^\alpha,$$

où q^α est le quantile supérieur de la loi $\mathcal{N}(0, \Sigma(t))$ tel que

$$1 - 2 \int_0^{q^\alpha} \int_0^{q^\alpha} \phi\{u, v; 0, \Sigma(t)\} dudv = \alpha.$$

La p -value de ce test est

$$1 - 2 \int_0^{M_n(\beta_0, t)} \int_0^{M_n(\beta_0, t)} \phi\{u, v; 0, \Sigma(t)\} dudv.$$

Remarque 4.2.6. *La p -value peut être évaluée numériquement par simulation, en adaptant la méthode proposée à la Remarque 4.2.5 à ce nouveau test.*

Ce test est adaptatif au sens où les données seules déterminent laquelle des deux statistiques sera utilisée. Les possibilités sont très restreintes, seuls deux choix sont possibles. Les performances du test sont comprises entre les performances des tests de la distance et de l'aire sous la courbe. Ainsi, l'hypothèse alternative $H_1 : \beta(t) \neq \beta_0$ est bien détectée en général, que l'effet dépende du temps ou non. Cependant, on observe une légère perte de puissance sous $H_{1,PH}$ par rapport au test du log-rank et une légère perte de puissance sous $H_{1,NPH}$ comparée au test de l'aire sous la courbe.

Avant d'étudier les comportements de ces différents tests avec des simulations, nous développons ces tests dans le cadre multivarié de plusieurs covariables.

4.2.4 Tests multivariés

On considère un vecteur de covariables de dimension p , noté $Z(t) = (Z^{(1)}(t), Z^{(2)}(t), \dots, Z^{(p)}(t))$. Dans ce contexte multivarié, on rappelle que le processus du score standardisé a été étudié en Section 3.4. On se place sous les conditions **H1'**, **H1'**, **H3'** et **H4'**. On suppose que $\beta, \beta_0 \in \mathbb{B}'$. Le processus du score standardisé est une fonction aléatoire de $[0, 1]$ dans \mathbb{R}^p et on note

$$U_n^*(\beta_0, t) = (U_n^*(\beta_0, t)_1, U_n^*(\beta_0, t)_2, \dots, U_n^*(\beta_0, t)_p), \quad 0 \leq t \leq 1.$$

En intégrant chaque composante de ce processus en fonction du temps, on peut également définir le processus de l'aire sous la courbe du processus du score standardisé multivarié.

Définition 4.2.6. *Le processus de l'aire sous la courbe du processus du score standardisé multivarié est une fonction aléatoire de $[0, 1]$ dans \mathbb{R}^p telle que*

$$J_n(\beta_0, t) = \left(\int_0^t U_n^*(\beta_0, s)_1 ds, \int_0^t U_n^*(\beta_0, s)_2 ds, \dots, \int_0^t U_n^*(\beta_0, s)_p ds \right), \quad 0 \leq t \leq 1.$$

Sous $H_0 : \beta(t) = \beta_0$, le Théorème 3.4.1 implique les convergences

$$\begin{aligned} U_n^*(\beta_0, \cdot) &\xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{W}_p, \\ J_n(\beta_0, \cdot) &\xrightarrow[n \rightarrow \infty]{\mathcal{L}} \int \mathcal{W}_p(s) ds. \end{aligned}$$

Dans la suite du chapitre, pour simplifier les notations et sans perdre de généralité, on considère $p = 2$. Les équations (4.5), (4.6) et le Lemme 4.2.1 qui indiquent le comportement limite des fonctions de covariance entre les différentes quantités aléatoires impliquent le résultat suivant.

Proposition 4.2.6. *Soit $t \in [0, 1]$. Sous l'hypothèse nulle $H_0 : \beta = \beta_0$, c'est-à-dire sous le modèle à risques proportionnels de paramètre β_0 , le vecteur*

$$\left(U_n^*(\beta_0, t)_1, U_n^*(\beta_0, t)_2, \sqrt{3}J_n(\beta_0, t)_1, \sqrt{3}J_n(\beta_0, t)_2 \right)^T$$

converge en loi vers un vecteur gaussien centré de matrice de variance covariance $\tilde{\Sigma}_t$ telle que

$$\tilde{\Sigma}_t = \begin{pmatrix} t & 0 & \sqrt{3}t^2/2 & 0 \\ 0 & t & 0 & \sqrt{3}t^2/2 \\ \sqrt{3}t^2/2 & 0 & t^3 & 0 \\ 0 & \sqrt{3}t^2/2 & 0 & t^3 \end{pmatrix}.$$

Plusieurs corollaires à cette proposition étendent le test de la distance depuis l'origine, de l'aire sous la courbe et le test restreint adaptatif au cas multivarié pour tester l'hypothèse nulle $H_0 : \beta(t) = \beta_0$ contre $H_1 : \beta(t) \neq \beta_0$. Le premier corollaire étend le test de la distance depuis l'origine.

Corollaire 4.2.2. *La statistique de la distance depuis l'origine est notée $\mathbb{U}_n(\beta_0, t)$ telle que*

$$\mathbb{U}_n(\beta_0, t) = t^{-1} \left\{ U_n^*(\beta_0, t)_1^2 + U_n^*(\beta_0, t)_2^2 \right\}.$$

Sous l'hypothèse $H_0 : \beta(t) = \beta_0$, $\mathbb{U}_n(\beta_0, t)$ converge en loi vers une loi du chi deux à 2 degrés de liberté χ_2^2 lorsque $n \rightarrow \infty$. Le test de la distance depuis l'origine rejette H_0 au niveau asymptotique α si $\mathbb{U}_n(\beta_0, t) \geq Q_2(\alpha)$, où $P(\chi_2^2 \geq Q_2(\alpha)) = \alpha$.

Le corollaire suivant étudie le test de l'aire sous la courbe multivarié.

Corollaire 4.2.3. *La statistique de l'aire sous la courbe est*

$$\mathbb{J}_n(\beta_0, t) = 3t^{-3} \left\{ J_n(\beta_0, t)_1^2 + J_n(\beta_0, t)_2^2 \right\}.$$

Sous $H_0 : \beta(t) = \beta_0$, $\mathbb{J}_n(\beta_0, t)$ converge en loi vers une loi χ_2^2 du chi deux à 2 degrés de liberté lorsque $n \rightarrow \infty$. Le test de l'aire sous la courbe rejette H_0 au niveau asymptotique α si $\mathbb{J}_n(\beta_0, t) \geq Q_2(\alpha)$, où $P(\chi_2^2 \geq Q_2(\alpha)) = \alpha$.

Enfin, on peut étendre le test restreint adaptatif au cas multivarié comme suit.

Corollaire 4.2.4. *Soit $\mathbb{M}_n(\beta_0, t)$ la statistique du test restreint adaptatif multivarié telle que*

$$\mathbb{M}_n(\beta_0, t) = \max \{ \mathbb{U}_n(\beta_0, t), \mathbb{J}_n(\beta_0, t) \}.$$

Le test restreint adaptatif multivarié rejette H_0 au niveau asymptotique α si $\mathbb{M}_n(\beta_0, t) \geq Q(\alpha)$, où

$$1 - \int_{\mathbb{R}^4} \mathbf{1}_{p^2+q^2 \leq Q(\alpha), r^2+s^2 \leq Q(\alpha)} \phi(p, q, r, s; 0, \tilde{\Sigma}_1) dpdqdrds = \alpha,$$

et $\phi(\cdot; 0, \tilde{\Sigma}_1)$ est la densité de la distribution normale $\mathcal{N}_4(0, \tilde{\Sigma}_1)$. La p -value du test est

$$1 - \int_{\mathbb{R}^4} \mathbf{1}_{p^2+q^2 \leq \mathbb{M}_n(\beta_0, t), r^2+s^2 \leq \mathbb{M}_n(\beta_0, t)} \phi(p, q, r, s; 0, \tilde{\Sigma}_1) dpdqdrds.$$

Plusieurs possibilités pour l'extension des tests au cadre multivarié ont été considérées. Celle que nous avons retenue présente l'avantage d'être invariante selon le codage utilisé et de prendre en compte les covariances entre les p covariables.

4.3 Simulations

Nous étudions le comportement des tests présentés à taille d'échantillon finie et nous comparons ces tests à plusieurs tests de référence.

Scénarii de simulation

Les temps de décès T sont simulés selon une loi exponentielle de paramètre 1. Les temps de censure C , indépendants de T , suivent également une distribution exponentielle, dont le paramètre change. Le paramètre de la loi de la censure est déterminé de manière à calibrer les pourcentages de censure, comme mentionné Table 4.1. Le taux de censure dans l'échantillon

λ	0.5	1	2
$P(C \leq T)$	0.3	0.5	0.7

TABLE 4.1 – Lien entre le pourcentage de censure $P(C \leq T)$ et le paramètre de la loi $\mathcal{E}(\lambda)$ de la censure.

est donc 30%, 50% ou 70%. Les données sont simulées avec une covariable ($p = 1$). La

covariable Z peut suivre une distribution de Bernoulli de paramètre $1/2$, une distribution normale de moyenne $1/2$ et de variance $1/4$, une loi uniforme continue sur $\left[\frac{1-\sqrt{3}}{2}, \frac{1+\sqrt{3}}{2}\right]$ ou une loi exponentielle de paramètre $1/2$. Les paramètres de ces distributions ont été choisis afin d'égaliser leurs deux premiers moments. L'échantillon est de taille 60, 100 ou 200. Les tests sont calibrés pour avoir un risque de première espèce $\alpha = 0.05$.

Plusieurs coefficients de régression sont considérés. Le cas $\beta(t) = 0$ pour tout $t \in [0, 1]$ permet de vérifier le niveau des tests. Les cas $\beta(t) = 0.5$ ou $\beta(t) = 0.8$ pour tout $t \in [0, 1]$ sont les deux situations de risques proportionnels considérées. Les autres cas correspondent au modèle à risques non proportionnels, avec un effet continu et décroissant ($\beta(t) = 1.5(1-t)$ ou $\beta(t) = 2(1-t)^2$, pour $t \in [0, 1]$), ou des risques qui se croisent ($\beta(t) = -1.5\mathbf{1}_{t \leq 0.5} + 1.5\mathbf{1}_{t > 0.5}$ pour $t \in [0, 1]$, ou son opposé). On considère également 3 modèles de rupture, avec des temps de rupture différents ($\beta(t) = \mathbf{1}_{t \leq 0.3}$, $\beta(t) = \mathbf{1}_{t \leq 0.5}$ ou $\beta(t) = \mathbf{1}_{t \leq 0.7}$, $t \in [0, 1]$). Pour chaque coefficient de régression et chaque pourcentage de censure considérés, 3000 échantillons ont été générés selon la méthode détaillée en annexe pour évaluer le niveau et les puissances empiriques de chacun des tests.

Remarque 4.3.1. *Nous avons prouvé les théorèmes de la Section 3 pour des covariables uniformément bornées. Ces théorèmes sont à la base des tests présentés dans ce chapitre. Cependant, nous effectuons des simulations avec des covariables normales et exponentielles puisque ces cas sont souvent considérés dans les applications, bien qu'il s'agisse de variables non bornées uniformément.*

Choix du paramètre γ

Un travail de simulation a été effectué pour calibrer le paramètre γ du test restreint adaptatif de statistique $M_n^\gamma(0, 1)$. Rappelons que l'intérêt de ce test réside dans la possibilité de perdre très peu de puissance par rapport au test du log-rank sous l'hypothèse de proportionnalité des risques, tout en gagnant de la puissance sous des alternatives de type risques non proportionnels. La taille d'un échantillon est fixée à $n = 100$, avec une covariable Z simulée selon une loi de Bernoulli de paramètre $1/2$ et un taux de censure égal à 30%. Pour chaque coefficient de régression $\beta(t)$, les tests du log-rank, de l'aire sous la courbe de statistique $J_n(0, 1)$ et le test restreint adaptatif $M_n^\gamma(0, 1)$ ont été évalués, avec $\gamma = 0.5, 1, 1.5$ ou 2. Les résultats sont donnés Table 4.2.

Lorsque $\gamma = 2$, la puissance du test $M_n^\gamma(0, 1)$ est très proche de celle du test du log-rank, sauf lorsque les risques se croisent, c'est-à-dire lorsque le coefficient de régression change de signe dans le temps. Dans ces situations, le test de statistique $M_n^2(0, 1)$ est nettement plus puissant que le test du log-rank. Hormis dans ces situations, le test est peu puissant pour détecter la présence d'un effet qui change avec le temps, héritant du comportement du log-rank.

Lorsque $\gamma = 0.5$, la puissance du test de statistique $M_n^\gamma(0, 1)$ est, à l'opposé, très proche de celle du test de l'aire sous la courbe. Ainsi, on observe une forte perte de puissance (en-

$\beta(t)$	Log-rank	$J_n(0,1)$	$M_n^\gamma(0,1)$			
			$\gamma = 0.5$	$\gamma = 1$	$\gamma = 1.5$	$\gamma = 2$
0	4.9	5.2	5.0	5.1	4.8	4.6
0.5	50.9	40.9	42.7	46.3	49.5	50.3
0.8	88.1	76.8	79.2	83.5	86.6	87.5
$\mathbf{1}_{t \leq 0.3}$	21.8	38.5	35.4	32.0	25.3	21.2
$\mathbf{1}_{t \leq 0.5}$	54.1	72.6	70.1	66.7	58.3	52.8
$\mathbf{1}_{t \leq 0.7}$	83.4	89.0	88.3	85.5	82.3	80.9
$-1.5 \mathbf{1}_{t \leq 0.5}$ $+1.5 \mathbf{1}_{t \geq 0.5}$	16.8	85.3	82.7	83.4	84.4	80.0
$1.5 \mathbf{1}_{t \leq 0.5}$ $-1.5 \mathbf{1}_{t \geq 0.5}$	15.5	85.1	82.3	83.2	84.4	79.5
$1.5(1-t)$	84.7	91.6	91.0	88.8	85.2	83.0
$2(1-t)^2$	73.8	90.6	89.2	86.0	78.7	72.8

TABLE 4.2 – Niveaux de signification et puissances empiriques des tests (en %) basés sur 3000 jeux de données pour chaque $\beta(t)$. Chaque jeu de données a une taille 100 (50 sujets par groupe). Le taux de censure est fixé à 30%.

viron 10%) par rapport au test du log-rank sous les alternatives considérées de type risques proportionnels.

Enfin, les cas $\gamma = 1$ et $\gamma = 1.5$ sont intermédiaires à ces deux extrêmes. Le test de statistique $M_n^1(0,1)$ a une puissance plus proche du test de l'aire sous la courbe, alors que le comportement sous l'alternative du test de statistique $M_n^{1.5}(0,1)$ se rapproche plus de celui du log-rank. A l'inverse des cas précédents, les tests ont des puissances convenables sous toutes les alternatives considérées. Nous choisissons de retenir la valeur $\gamma = 1.5$ de manière à minimiser la perte de puissance comparée au test du log-rank sous des alternatives de type risques proportionnels.

Description des tests comparés

Les tests de la distance depuis l'origine, de l'aire sous la courbe du processus du score standardisé et les tests restreints adaptatifs au temps $t = 1$ de statistiques respectives $U_n^*(0,1)$, $J_n(0,1)$, $M_n(0,1)$ et $M_n^{1.5}(0,1)$ sont étudiés. De plus, le test du log-rank (Mantel, 1966) et deux autres tests adaptatifs sont évalués. Ces deux tests ont été décrits dans la Section 4.1, nous les rappelons brièvement ici.

Le premier test est le test de Lee (1996), dont la statistique est le maximum entre quatre statistiques du log-rank pondéré avec des poids dans la famille $G^{\rho,\gamma}$ de Fleming et Harrington (1991). Les poids permettent de détecter respectivement un effet précoce, un effet au milieu de l'étude ou un effet tardif entre les courbes de survie des deux groupes ainsi qu'une situation de risques proportionnels. Cette procédure est similaire au test proposé par Self (1991). La classe de tests de Fleming et Harrington (1991) est connue pour avoir une bonne puissance de détection d'un effet dépendant du temps.

La seconde statistique de test, proposée par Yang et Prentice (2010), est le maximum

entre deux statistiques du log-rank pondéré, avec des poids basés sur l'inférence des quantités définies par le modèle (1.20) de Yang et Prentice (2005). Asymptotiquement, cette statistique équivaut à celle du log-rank lorsque les risques sont proportionnels. Yang et Prentice (2010) ont recommandé l'utilisation de ce test après avoir comparé les comportements de nombreux tests sous différentes alternatives. Ce test est implémenté sous R dans la librairie `Ypmodel`.

Résultats

On considère le cas d'une covariable Z de loi de Bernoulli de paramètre $1/2$. Les résultats sont donnés à la Table 4.3 pour le cas $n = 100$, Table 4.4 pour une taille d'échantillon $n = 60$ et Table 4.5 pour $n = 200$. Les conclusions sont les mêmes pour les 3 tailles de population considérées.

A l'exception des tests de Lee et de Yang et Prentice, les niveaux des tests sont bien contrôlés à 5%. L'inflation du niveau du test de Yang et Prentice confirme leurs propres résultats de simulation pour des tailles d'échantillon et des niveaux de censure comparables. Comme attendu (Proposition 4.2.2), le test de la distance a une puissance très proche du test du log-rank sous des alternatives de type risques proportionnels et non proportionnels.

Lorsque les risques sont proportionnels, le test de l'aire sous la courbe est moins puissant que le log-rank, alors que la puissance du test adaptatif restreint M_n se situe entre les deux, avec une perte moyenne comparée au log-rank de seulement 3%. Les puissances du test restreint adaptatif M_n^γ avec $\gamma = 1.5$ et le test de Lee sont comparables à celles du log-rank. Le test le plus puissant semble être celui de Yang et Prentice. Ce résultat est surprenant puisque le test du log-rank est le plus puissant dans cette situation (Peto et Peto, 1972). Cette apparente contradiction s'explique par le contrôle insuffisant de l'erreur de type I du test.

La première situation de risques non proportionnels considère un effet constant par morceaux, correspondant à un modèle de rupture. Nous faisons varier les temps de rupture. Le test de l'aire sous la courbe est plus puissant que le log-rank ou le test de la distance. La puissance du test restreint adaptatif M_n est plus proche de celle du test de l'aire sous la courbe. D'autre part, la puissance du test de statistique $M_n^{1.5}$ est faible et proche du log-rank. Les quatre nouveaux tests introduits dans cette thèse ont de meilleures puissances que le test de Lee. Le test de Yang et Prentice est le plus puissant de tous les tests étudiés, mais ce résultat doit être à nouveau tempéré par le contrôle insuffisant de l'erreur de première espèce. Les puissances du test M_n sont relativement proches de celles du test de Yang et Prentice.

Dans les situations de risques qui se croisent, les puissances des tests du log-rank, de la distance et de Lee sont faibles. Le test de Yang et Prentice subit une sévère diminution de puissance, jusqu'à 30%, entre un effet ayant une valeur négative puis positive et un effet positif puis négatif. Nous avons étudié plus en détail ce résultat surprenant. Sur un même jeu de données, nous avons interverti la valeur de covariable : le groupe qui était codé 0 a été codé 1 et inversement. La statistique du test a été modifiée par le nouveau codage. L'expérience a été répétée plusieurs fois, avec la même conclusion. L'utilisation de ce test en pratique est donc délicate puisqu'il dépend d'un codage arbitraire. Les tests restreints adaptatifs et de l'aire

sous la courbe montrent de bonnes puissances en situation de risques qui se croisent.

Dans les derniers cas étudiés, l'effet $\beta(t)$ diminue de manière continue avec le temps. Toutes les puissances sont comparables, le test le plus puissant est celui de l'aire sous la courbe.

En conclusion, dans les situations générales pour lesquelles aucune hypothèse n'est faite sur la forme de l'effet, nous recommandons l'utilisation du test restreint adaptatif de statistique M_n . Les résultats de simulation montrent de bonnes puissances sous des alternatives de type risques proportionnels et non proportionnels. L'utilisation du test de Yang et Prentice est délicate à cause du manque de contrôle du risque de première espèce et à la sensibilité au codage. Si l'utilisateur souhaite perdre le moins de puissance possible comparée au test du log-rank, nous conseillons d'utiliser un test dans la classe des tests adaptatifs restreints de statistique M_n^γ . Cependant, ce test est moins puissant que le test de statistique M_n en cas de risques non proportionnels.

Influence de la distribution de la covariable

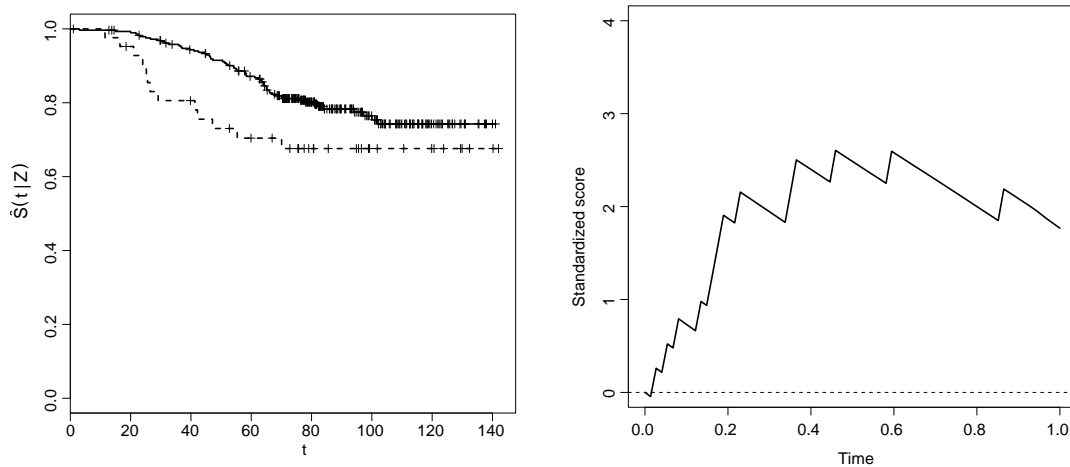
Les tests de Lee et de Yang et Prentice ont été développés pour comparer les survies de 2 groupes de patients. Ainsi, ils ne peuvent être appliqués qu'avec des covariables binaires. Nous étudions le comportement des tests proposés dans cette thèse avec des covariables continues. Les paramètres de ces lois ont été choisis pour égaliser les deux premiers moments. La taille d'un échantillon est fixée à $n = 100$, avec un taux de censure de 30%. Les résultats sont donnés Table 4.6. Pour un test et un coefficient de régression donnés, les puissances respectives de chaque test sont similaires pour les trois distributions de covariable.

4.4 Applications

Deux applications sont proposées. Nous disposons de données collectées par l'Institut Curie pour l'étude du cancer du sein. Nous n'étudierons qu'un effet pronostique, la taille de la tumeur sur la survie afin d'illustrer nos méthodes sur un cas univarié. Dans un second temps, nous étudions des données d'un essai clinique avec trois groupes de traitement afin d'illustrer le cas multivarié.

4.4.1 Données de l'Institut Curie

La base de données, collectée par l'Institut Curie, est composée de 1504 patientes atteintes de cancer du sein. La base de données est complexe avec plusieurs sous-groupes et, pour l'illustration, nous considérons un sous-groupe particulier de 332 patientes. Parmi les facteurs pronostiques, on s'intéresse à l'influence de la taille de la tumeur en millimètres (mm) sur la durée de vie des patientes. Les patientes sont séparées en deux groupes selon que la taille de leur tumeur soit inférieure ou supérieure à 60 mm. Les patientes ont été suivies pendant 12 ans. Le taux de censure de la base de données est de 22%. Les estimateurs de Kaplan-Meier de



(a) Estimateurs de Kaplan-Meier de la survie des patients avec une taille de tumeur inférieure (ligne solide) ou supérieure (ligne pointillée) à 60 mm. (b) Processus du score standardisé pour la taille de la tumeur.

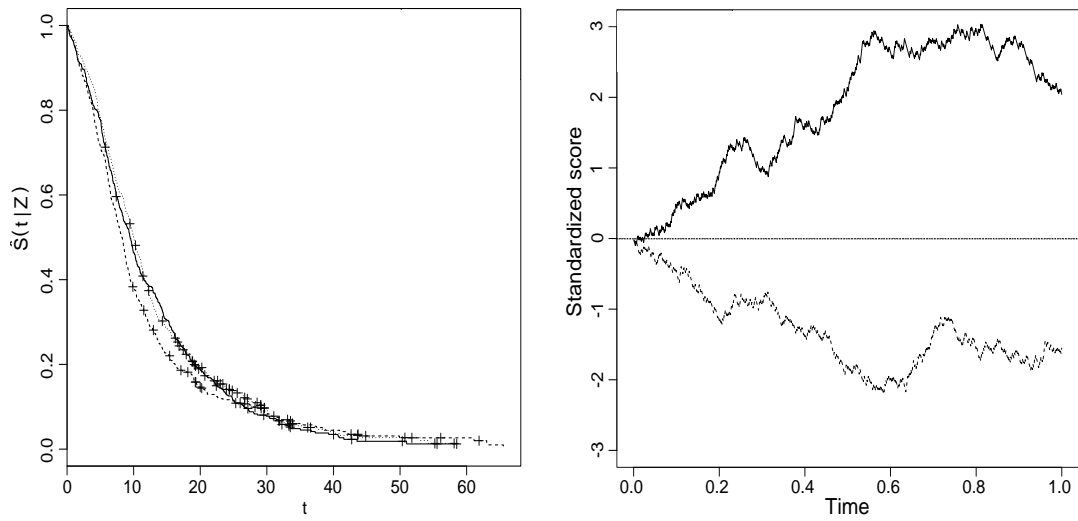
FIGURE 4.1 – Estimateurs de Kaplan-Meier et processus du score standardisé pour la base de données de l'Institut Curie.

la survie sont présentés Figure 4.1(a). Pendant les 60 premiers mois, les survies estimées sont très différentes, correspondant à un effet fort entre les groupes de patientes. Après 60 mois, les courbes se rapprochent, reflétant une diminution de l'effet. De plus, on remarque que 80% des patientes dont la taille de la tumeur est inférieure à 60 mm survit au moins 75 mois. Cette durée est divisée par 3 et diminuée à 28 mois pour les patientes dont la taille de la tumeur est supérieure à 60 mm.

Le processus du score standardisé correspondant est tracé en Figure 4.1(b). Ce processus n'est pas centré autour de 0, la présence d'une dérive, et donc d'un effet, est claire. La dérive n'étant pas linéaire, l'effet diminue dans le temps. Cette diminution a un impact sur les tests du log-rank ($p = 0.06$) et de la distance depuis l'origine ($p = 0.08$) qui ne sont pas significatifs. Cependant, le test de l'aire sous la courbe et le test restreint adaptatif sont très significatifs, avec des p -values respectives de 0.001 et 0.002. Les conclusions de ces deux derniers tests suivent notre analyse des estimateurs de Kaplan-Meier.

4.4.2 Essai clinique

Durant un essai clinique récent, 1636 sujets ont été randomisés dans 3 groupes de traitement et suivis pendant 65 mois. Le but de cet essai est d'étudier l'influence des traitements sur la survie de patients atteints d'une même maladie. Au moment de la rédaction de cette thèse, l'étude et ses résultats n'ont pas été publiés dans un journal clinique et sont confidentiels, seuls les éléments nécessaires à l'illustration sont décrits ici. Les estimateurs de Kaplan-Meier (1.1) des courbes de survie sont présentés en Figure 4.2(a). Les survies des patients randomisés dans les groupes de traitement 1 et 3 semblent très proches, à l'inverse de la survie des patients dans



(a) Estimateurs de Kaplan-Meier de la survie dans le groupe de traitement 1 (ligne solide), groupe de traitement 2 (tirets) ou groupe de traitement 3 (ligne pointillée). (b) Processus du score standardisé du groupe de traitement 2 contre 1 (ligne solide) ou du groupe de traitement 3 contre 1 (tirets).

FIGURE 4.2 – Estimateurs de Kaplan-Meier et processus du score standardisés pour les données de l'essai clinique avec 3 groupes de traitement.

le groupe de traitement 2. Ainsi, les courbes de Kaplan-Meier nous encouragent à trouver un test assez puissant pour rejeter l'hypothèse nulle $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ où β_i est le coefficient de régression associé au groupe de traitement $i = 1, 2$ ou 3 . Deux covariables binaires ont été créées à partir de la covariable traitement, avec la modalité "groupe de traitement 1" en référence. Le processus du score standardisé multivarié $U_n^*(0, \cdot)$ est représenté en fonction du temps en Figure 4.2(b).

Malgré le bruit, les dérives des processus sont nettement visibles, correspondant à l'hypothèse alternative de la présence d'un effet. L'effet apparaît être non linéaire. Sa valeur diminue dans le temps de telle sorte que le test du log-rank multivarié est non significatif, avec une p-value égale à 0.10. Le test de la distance multivarié a une p-value égale à 0.09. En revanche, le test de l'aire sous la courbe multivarié et le test restreint adaptatif multivarié sont très significatifs, avec des p-values de 0.005 et 0.008 respectivement. Ces tests sont moins impactés par la diminution des effets et, pour cette raison, ils détectent leur présence. Les résultats de ces deux derniers tests sont plus proches de notre intuition lors de l'étude des courbes de Kaplan-Meier.

4.5 Discussion

Conclusion du chapitre

Dans ce chapitre, nous étudions des tests de la présence d'effet des covariables sur la survie. En considérant le cas de variables discrètes, il s'agit de comparer les survies de groupes d'individus. Ces tests peuvent également être appliqués sur des variables continues. Sous le modèle à risques non proportionnels, le test revient à considérer l'hypothèse nulle $H_0 : \beta(t) = 0$ contre son alternative $H_0 : \beta(t) \neq 0$. Nous avons développé ces tests dans le cadre multivarié de plusieurs covariables, en prenant en compte la structure de covariance entre les covariables.

Pour comparer les survies de plusieurs groupes, le test du log-rank est couramment utilisé. Ce test est le plus puissant pour détecter un effet fixe dans le temps. En revanche, plusieurs auteurs ont montré ses mauvaises propriétés en présence d'un effet qui varie dans le temps. Notre but était alors de proposer un test ayant de bonnes puissances sous ces deux types d'alternatives. Les propriétés asymptotiques du processus du score standardisé étudiées au Chapitre 3 ont permis de construire plusieurs tests. Nous avons d'abord introduit le test de la distance depuis l'origine, asymptotiquement équivalent au test du log-rank. Puis, le test de l'aire sous la courbe a été étudié. Ce test, un peu moins puissant que le test du log-rank en présence d'un effet constant, permet de gagner de la puissance en présence d'effet qui change avec le temps. Enfin, nous avons introduit une classe de tests adaptatifs, dont les statistiques dépendent d'un paramètre. Ces statistiques prennent la valeur des statistiques de la distance ou de l'aire sous la courbe. Le paramètre permet de calibrer la proximité de ce test avec celui du test du log-rank. Cependant, le choix du paramètre pouvant être délicat, ou si l'utilisateur n'a pas besoin de se rapprocher du test du log-rank, nous proposons l'utilisation du test restreint adaptatif, dont la statistique est la valeur absolue maximale des statistiques de la distance depuis l'origine et de l'aire sous la courbe. Une étude de simulation a permis de conclure à la bonne performance de ces tests adaptatifs, en les comparant avec d'autres tests de la littérature. En plus du contrôle du niveau des tests et de leurs bonnes propriétés de puissance, leur avantage est leur interprétation simple et leur représentation graphique. Dans les applications, en même temps que les p-values, les estimations des paramètres et leurs intervalles de confiance, il serait certainement utile de considérer des graphiques des processus pour obtenir une impression visuelle de l'écart à la significativité des tests de la distance ou de l'aire sous la courbe.

En évaluant le processus du score standardisé U_n^* au paramètre β_0 , les résultats peuvent s'étendre immédiatement à l'hypothèse nulle $H_0 : \beta(t) = \beta_0, \forall t$ contre son alternative $H_1 : \exists t_0, \beta(t_0) \neq \beta_0$. Si β_0 n'est pas nul, il ne s'agit plus alors de tester la présence d'une différence entre les survies des groupes, mais de tester si le risque relatif entre les groupes est égal à β_0 ou non.

Perspectives

Nous nous sommes concentrés sur des tests d'une valeur constante du paramètre de régression. Il est possible de considérer le test de l'hypothèse nulle $H_0 : \beta(t) = \beta_0(t)$ contre son alternative $H_1 : \beta(t) \neq \beta_0(t)$, où le coefficient de régression $\beta_0(t)$ change avec le temps.

D'autres combinaisons des tests de la distance depuis l'origine et de l'aire sous la courbe peuvent être étudiées. Nous nous sommes intéressés à leur combinaison linéaire

$$\theta U_n^*(\beta_0, t) + (1 - \theta) J_n(\beta_0, t), \quad 0 \leq \theta \leq 1.$$

Cependant, selon la proportionnalité ou non des risques, des simulations ont montré que les puissances maximales étaient toujours atteintes pour $\theta = 1$ ou $\theta = 0$ respectivement, se réduisant au test restreint adaptatif. Aussi, il serait intéressant d'étudier d'autres tests basés sur les propriétés du mouvement Brownien, qui est le processus limite de $U_n^*(\beta_0, \cdot)$.

Nous avons appliqué ces tests pour $t = 1$. Il serait possible d'appliquer les tests à n'importe quel temps τ si l'on sait, par exemple, que l'effet s'annule après le temps τ , afin d'augmenter la puissance des tests. Il serait également intéressant d'étudier les tests dont les statistiques sont les supremum sur le temps des statistiques de la distance ou de l'aire sous la courbe.

Dans le cadre multivarié, les tests introduits dans ce chapitre permettent de tester simultanément la nullité globale des paramètres. Une extension serait d'étudier comment utiliser le processus pour tester la nullité de certains paramètres seulement. Supposons deux variables Z_1 et Z_2 dans le modèle à risques non proportionnels de paramètres respectifs $\beta_1(t)$ et $\beta_2(t)$. Pour tester $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$, une idée serait de stratifier sur la covariable Z_2 en estimant le paramètre $\beta_2(t)$ et en considérant le processus dont la construction est basée sur les probabilités $\{\pi_i(\hat{B}(t), t), i = 1, \dots, n\}$, avec $\hat{B}(t) = (0, \hat{\beta}_2(t))$.

$\beta(t)$	cens. (%)	Log- rank	$U_n^*(0,1)$	$J_n(0,1)$	$M_n(0,1)$	$M_n^{1.5}(0,1)$	Lee	Yang Prentice
0	30	4.8	4.5	4.9	5.0	4.4	6.1	6.4
	50	5.2	5.0	5.4	5.0	5.0	6.3	6.8
	70	5.1	4.7	4.8	4.8	4.4	5.8	6.2
0.5	30	50.9	49.9	39.6	47.3	48.8	50.4	54.2
	50	51.0	50.2	39.0	47.4	48.9	51.0	53.8
	70	51.7	50.8	40.9	48.7	49.7	50.8	55.5
0.8	30	87.1	86.6	74.6	84.1	85.3	86.3	89.0
	50	87.2	86.7	75.8	85.3	85.7	86.8	88.4
	70	88.0	87.1	76.8	85.8	85.9	87.3	89.0
$\mathbf{1}_{t \leq 0.3}$	30	23.7	22.2	40.1	34.6	26.2	19.8	40.4
	50	18.4	17.6	30.8	26.9	21.0	16.1	34.4
	70	13.5	12.7	21.7	18.5	14.4	11.1	26.6
$\mathbf{1}_{t \leq 0.5}$	30	55.5	52.4	73.2	69.4	59.0	49.5	69.1
	50	44.3	41.9	61.0	55.9	46.7	37.3	57.7
	70	32.8	30.3	44.4	40.9	32.6	27.2	43.6
$\mathbf{1}_{t \leq 0.7}$	30	83.1	80.6	88.2	86.9	82.1	79.0	87.3
	50	72.4	69.4	77.3	76.4	71.1	66.6	78.1
	70	56.0	51.9	60.2	59.0	53.3	49.2	61.2
$-1.5\mathbf{1}_{t \leq 0.5}$ $+1.5\mathbf{1}_{t \geq 0.5}$	30	17.1	15.1	85.7	81.3	84.8	27.5	61.5
	50	13.5	11.7	72.7	67.7	69.6	39.5	42.3
	70	12.0	11.4	56.0	49.5	49.0	37.0	25.8
$1.5\mathbf{1}_{t \leq 0.5}$ $-1.5\mathbf{1}_{t \geq 0.5}$	30	15.9	13.8	85.3	81.4	84.3	26.0	87.8
	50	14.0	12.4	74.0	68.0	71.0	40.6	77.1
	70	11.0	10.2	55.2	49.1	48.7	36.7	61.6
$1.5(1-t)$	30	85.0	82.9	90.7	90.0	84.7	81.3	90.1
	50	74.5	71.0	81.5	79.9	73.2	68.6	81.2
	70	58.0	54.9	64.7	62.8	56.3	52.0	65.8
$2(1-t)^2$	30	73.5	70.8	89.7	87.0	78.2	67.8	88.1
	50	60.7	57.4	78.7	74.7	64.6	54.0	77.9
	70	46.7	43.0	62.1	58.1	47.2	39.4	62.8

TABLE 4.3 – Niveaux de signification et puissances empiriques des tests (en %) basés sur 3000 jeux de données pour chaque $\beta(t)$ et chaque niveau de censure. Chaque jeu de données a une taille 100 (50 sujets par groupe). Les tests ont un niveau nominal de 5%.

$\beta(t)$	cens. (%)	Log- rank	$U_n^*(0,1)$	$J_n(0,1)$	$M_n(0,1)$	$M_n^{1.5}(0,1)$	Lee	Yang Prentice
0	30	4.9	4.7	4.7	4.3	4.7	5.7	6.6
	50	5.4	5.5	4.9	5.1	5.3	6.5	7.1
	70	4.9	4.7	4.3	4.4	4.7	5	4.9
0.5	30	35.6	34.6	26.9	32.5	33.5	36.4	38.1
	50	27.6	26.7	19.5	25.4	26.0	27.7	29.7
	70	19.9	19.9	14.8	18.6	19.3	21.1	21.1
0.8	30	68.2	67.6	53.7	64.8	66.0	68.9	70.4
	50	53.7	52.8	40.8	50.6	51.3	54.6	56.2
	70	38.9	38.2	28.7	36.6	37.1	40.0	39.4
$\mathbf{1}_{t \leq 0.3}$	30	16.5	15.2	25.4	22.4	17.6	14.9	29.3
	50	13.5	12.7	19.9	17.6	13.7	12.5	23.6
	70	10.5	10.2	15.0	13.2	10.5	9.2	16.6
$\mathbf{1}_{t \leq 0.5}$	30	36.8	34.1	50.5	46.3	38.2	33.0	49.2
	50	29.3	26.7	40.2	36.6	28.9	25.0	40.1
	70	23.1	21.0	28.7	26.5	21.6	18.8	29.0
$\mathbf{1}_{t \leq 0.7}$	30	61.9	58.3	66.4	64.7	59.1	57.0	67.4
	50	51.6	47.9	53.8	54.4	48.8	47.2	57.5
	70	38.4	34.9	37.9	38.4	34.8	33.4	40.2
$-1.5\mathbf{1}_{t \leq 0.5}$ $+1.5\mathbf{1}_{t \geq 0.5}$	30	11.9	10.2	65.8	59.4	61.0	15.6	42.7
	50	10.7	9.3	51.8	45.8	43.3	22.4	29.6
	70	9.4	8.4	36.6	30.8	25.5	20.2	18.9
$1.5\mathbf{1}_{t \leq 0.5}$ $-1.5\mathbf{1}_{t \geq 0.5}$	30	13	11.2	66.6	59.9	61.7	16.8	67.6
	50	11.2	9.7	51.3	44.8	43.4	22.6	51.3
	70	8.7	8.1	36.7	31.2	26.0	19.1	38.6
$1.5(1-t)$	30	65.8	62.4	71.6	70.7	63.8	61.4	73.8
	50	53.1	48.6	57.7	57.1	50.3	46.9	60.9
	70	40.8	37.6	42.4	42.1	37.8	36.3	44.7
$2(1-t)^2$	30	53.3	49.4	69.8	65.6	55.5	48.4	71.0
	50	44.9	41.3	57.4	53.4	44.2	39.3	58.3
	70	29.7	27.6	37.5	34.7	29.2	25.2	37.5

TABLE 4.4 – Niveaux de signification et puissances empiriques des tests (en %) basés sur 3000 jeux de données pour chaque $\beta(t)$ et chaque niveau de censure. Chaque jeu de données a une taille 60 (30 sujets par groupe). Les tests ont un niveau nominal de 5%.

$\beta(t)$	cens. (%)	Log- rank	$U_n^*(0,1)$	$J_n(0,1)$	$M_n(0,1)$	$M_n^{1.5}(0,1)$	Lee	Yang Prentice
0	30	5.5	5.4	4.6	5.2	5.1	6.7	6.3
	50	5.2	5.3	4.9	5	5.2	6.2	6.6
	70	5.2	5.2	5.5	5.3	5.1	5.8	6.4
0.5	30	79.8	79.3	68.4	76.7	78.8	78.4	81.7
	50	67	66.3	54.8	64.2	65.4	64.2	69
	70	50.3	48.9	41.1	47.9	48.2	47.2	52.2
0.8	30	99.2	99.1	97.1	98.9	98.8	99.1	99.3
	50	96.9	96.5	91.7	96.3	95.7	95.9	97.1
	70	86.7	86.3	76.8	84.8	85	84.6	87.8
$\mathbf{1}_{t \leq 0.3}$	30	40.3	38.9	69.6	63.2	52.4	33.7	62.7
	50	29.9	28.8	55.6	49	38.4	24.3	49.5
	70	22.3	21.5	39.8	35.2	26.4	17.6	39.9
$\mathbf{1}_{t \leq 0.5}$	30	82.9	81.6	96.4	95	89.5	78	93.3
	50	69.8	68.2	88.7	85.4	76.2	63.4	83.6
	70	53.7	51	72.2	67.9	57.4	45.1	67.5
$\mathbf{1}_{t \leq 0.7}$	30	98.2	97.9	99.3	99.1	98.3	97.2	99
	50	94.2	93	96.7	96.1	94	91.3	96.2
	70	83.8	80.9	88.1	87.3	82.7	77.8	87.8
$-1.5\mathbf{1}_{t \leq 0.5}$ $+1.5\mathbf{1}_{t \geq 0.5}$	30	23.9	19.9	98.9	98.4	98.8	56.4	94
	50	20.7	17.2	95.3	93.5	95.1	73.3	70
	70	15.6	13	84.2	80.4	83.1	63.3	41.8
$1.5\mathbf{1}_{t \leq 0.5}$ $-1.5\mathbf{1}_{t \geq 0.5}$	30	23.7	19.7	98.7	98.3	98.7	53.9	98.8
	50	20.6	17	94.8	92.9	94.3	74	96.9
	70	15.7	13.7	83.1	79.3	81.6	61	88.5
$1.5(1-t)$	30	98.7	98.4	99.8	99.7	98.8	98.3	99.6
	50	95.5	94.2	98.4	97.9	95.8	93	97.8
	70	86.5	84.3	92.1	91.1	86.6	81.7	91.4
$2(1-t)^2$	30	95.5	94.8	99.7	99.5	98	93.3	99.5
	50	89	87.1	98.1	97.3	93.7	84.1	97.3
	70	73.1	70.1	89.7	86.9	78.3	65.7	87.8

TABLE 4.5 – Niveaux de signification et puissances empiriques des tests (en %) basés sur 3000 jeux de données pour chaque $\beta(t)$ et chaque niveau de censure. Chaque jeu de données a une taille 200 (100 sujets par groupe). Les tests ont un niveau nominal de 5%.

$\beta(t)$	Covariable	$U_n^*(0,1)$	$J_n(0,1)$	$M_n(0,1)$	$M_n^{1.5}(0,1)$
0	Normale	4.9	4.5	4.9	4.9
	Uniforme	5.4	4.8	5.2	5.5
	Exponentielle	5.0	5.0	5.0	5.3
0.5	Normale	46.9	40.6	45.9	46.4
	Uniforme	48.7	41.0	46.5	48.2
	Exponentielle	48.3	38.5	45.4	47.9
0.8	Normale	84.6	76.4	83.5	84.1
	Uniforme	84.5	76.3	82.1	84.0
	Exponentielle	84.7	76.1	83.0	83.9
$\mathbf{1}_{t \leq 0.3}$	Normale	22.4	41.7	37.2	28.2
	Uniforme	21.2	41.1	35.7	27.4
	Exponentielle	21.6	42.1	36.1	27.7
$\mathbf{1}_{t \leq 0.5}$	Normale	49.1	71.2	66.7	56.4
	Uniforme	50.6	71.5	66.7	56.9
	Exponentielle	49.0	71.5	67.5	56.9
$\mathbf{1}_{t \leq 0.7}$	Normale	77.0	87.3	84.9	79.4
	Uniforme	75.9	85.6	83.6	77.5
	Exponentielle	75.7	85.9	83.7	78.1
$-1.5\mathbf{1}_{t \leq 0.5}$ $+1.5\mathbf{1}_{t \geq 0.5}$	Normale	9.8	71.9	66.2	70.1
	Uniforme	9.1	73.3	67.9	71.4
	Exponentielle	9.6	73.1	67.9	71.2
$1.5\mathbf{1}_{t \leq 0.5}$ $-1.5\mathbf{1}_{t \geq 0.5}$	Normale	9.4	74.8	69.6	73.0
	Uniforme	9.5	73.8	68.2	71.9
	Exponentielle	9.6	72.5	67.6	70.5
$1.5(1-t)$	Normale	81.5	91.4	89.4	84.0
	Uniforme	79.9	91.3	89.9	83.2
	Exponentielle	79.7	90.5	88.7	83.
$2(1-t)^2$	Normale	70.9	90.5	87.9	79.7
	Uniforme	68.7	89.6	86.4	78.6
	Exponentielle	70.2	90.2	87.1	79.1

TABLE 4.6 – Niveaux de signification et puissances empiriques des tests (en %) basés sur 3000 jeux de données pour chaque $\beta(t)$ et chaque distribution de covariable. Le taux de censure est fixé à 30%. Chaque jeu de données a une taille 100.

Chapitre 5

Construction et évaluation du modèle à risques non proportionnels

En utilisant les propriétés de convergence du processus du score standardisé établies au Chapitre 3, nous développons une méthode de construction du modèle à risques non proportionnels. Le modèle résultant aura une bonne adéquation et une capacité prédictive optimale, mesurée par le coefficient R^2 . Nous dérivons un test d'adéquation du modèle à risques proportionnels ainsi que des bandes de confiance pour le processus correspondant à un effet constant. Comme nous l'avons illustré dans le chapitre introductif, l'adéquation et la capacité prédictive sont deux aspects différents d'un modèle statistique. Le lien entre ces deux notions n'est pas toujours clair. Dans ce chapitre, nous étudions comment l'amélioration de l'ajustement du modèle à risques non proportionnels peut permettre d'augmenter sa puissance prédictive. Le plan du chapitre est le suivant.

Dans la Section 5.1, nous présentons une revue de littérature sur les méthodes graphiques et les tests d'adéquation du modèle à risques proportionnels. Dans la Section 5.2, nous introduisons le coefficient R^2 du modèle à risques non proportionnels, qui est la mesure de capacité prédictive utilisée dans la suite du travail. La Section 5.3 présente une méthode graphique et un test d'adéquation du modèle à risques proportionnels basés sur le processus du score standardisé. Nous exposons également une méthode de construction du modèle à risques non proportionnels à partir du processus et du coefficient R^2 . Enfin, un travail de simulation est présenté en Section 5.4 pour comparer les performances du test d'adéquation que nous avons développé avec des tests standards. La méthode de construction du modèle à risques non proportionnels est illustrée sur des données simulées. Des applications sur des données réelles sont présentées en Section 5.5.2. Une discussion termine ce chapitre.

Ce chapitre est en partie constitué des résultats présentés dans [Chauvel et O'Quigley \(2014b\)](#).

5.1 Adéquation du modèle à risques proportionnels

L'utilisation d'un modèle statistique permet de simplifier la représentation des observations. Une question importante soulevée par l'utilisation d'un modèle particulier sur des observations est celle de l'adéquation : est-il plausible de considérer que les données sont issues du modèle ? Les procédures d'évaluation de l'adéquation peuvent être utilisées pour étudier la qualité de l'ajustement du modèle dans sa globalité ou seulement de certaines parties. Par exemple, dans le contexte d'une étude clinique, on peut se demander si les hypothèses de travail concernant les différences de traitement sont satisfaites en présence d'autres covariables, sans s'intéresser à la qualité de la modélisation de ces covariables. L'adéquation d'un modèle peut être évaluée avec des tests d'hypothèses ou en utilisant des méthodes graphiques. L'interprétation d'un test d'adéquation est délicate. La plupart des tests d'adéquation sont puissants pour détecter des alternatives particulières, par exemple un effet qui augmente ou qui diminue dans le temps. L'utilisateur doit alors avoir une idée de l'alternative à considérer quand il traite un jeu de données, ce qui n'est pas toujours le cas. De plus, le résultat d'un test est souvent limité à la lecture de la p -value. Cette dernière peut être diminuée par une augmentation du nombre d'observations et pas nécessairement par un ajustement moins bon. Pourtant, dans les deux cas, le test rejette l'hypothèse nulle d'une bonne adéquation du modèle aux données. Les méthodes graphiques, quant à elles, présentent souvent l'avantage de visualiser le type d'alternative, c'est-à-dire le modèle à considérer. Cependant, les résultats peuvent prêter à confusion ou être difficiles à interpréter si l'utilisateur n'est pas habitué à lire de tels graphiques. Les tests d'adéquation et les méthodes graphiques sont donc deux procédures complémentaires pour évaluer la qualité de l'ajustement d'un modèle particulier à un jeu de données. De nombreuses procédures d'adéquation du modèle à risques proportionnels ont été proposées.

5.1.1 Méthodes graphiques

La première méthode graphique pour vérifier l'hypothèse de proportionnalité des risques pour des variables catégorielles a été proposée par [Kay \(1977\)](#). Cette méthode est basée sur l'écriture suivante du modèle à risques proportionnels :

$$\log\{-\log(S(t|Z))\} = \beta^T Z + \log\left(\int_0^t \lambda_0(s) ds\right).$$

Ainsi, si la covariable Z prend les modalités z_1, \dots, z_M , les courbes $\log(-\log(S(t | Z = z_m)))$ en fonction du temps t pour chaque modalité $m = 1, \dots, M$, sont parallèles. En pratique, la fonction de survie conditionnelle $S(\cdot | Z)$ est remplacée par son estimateur de Kaplan-Meier $\hat{S}(\cdot | Z)$ (équation (1.1.1)). Cette méthode présente quelques inconvénients. Les covariables continues doivent être discrétisées et le choix de la coupure pour la discrétisation n'est pas toujours évident ([Andersen, 1982](#)). Le second inconvénient réside dans le fait que le parallélisme entre les courbes estimées, qui sont bruitées, n'est pas toujours détectable. On peut alors envisager de tracer les différences entre les courbes et de regarder si cette différence est

constante au cours du temps. Aussi, en présence de 2 modalités, on peut tracer une courbe en fonction de l'autre pour les différentes valeurs du temps, ce qui devrait produire une ligne droite (Andersen, 1982). Cependant, les estimateurs de Kaplan-Meier sont évalués avec peu de sujets vers la fin de l'étude, produisant de plus grandes variances. Ainsi, on peut observer un écart à la linéarité pour les grandes valeurs du temps et rejeter à tort l'hypothèse de proportionnalité des risques.

Les autres procédures sont essentiellement basées sur l'étude des résidus. Les premiers résidus utilisés dans les méthodes graphiques d'adéquation sont les résidus de Cox et Snell (1968). Leur distribution sous le modèle de Cox a été étudiée par Kay (1977). Lagakos (1981), O'Quigley (1982) et Crowley et Storer (1983) ont souligné leur faible puissance de détection d'une mauvaise adéquation. De plus, ces résidus reflètent aussi la variabilité de la covariable (Baltazar-Aban et Peña, 1995). Les méthodes graphiques pour vérifier l'adéquation du modèle à risques proportionnels basées sur les résidus peuvent être classées en deux catégories, selon que les résidus soient cumulés ou non. Parmi les méthodes non cumulatives, une grande classe de résidus de martingale décrite par Barlow et Prentice (1988) peut être considérée en traçant ses membres en fonction des temps de décès observés ou de leur rang. Le résidu de martingale de l'individu $j = 1, \dots, n$, au temps t est défini par

$$M_j(t) = N_j(t) - \hat{\Lambda}_j(t), \quad \text{avec} \quad \hat{\Lambda}_j(t) = \int_0^t \pi_j(\hat{\beta}, s) d\bar{N}(s),$$

où $\hat{\beta}$ est l'estimateur du maximum de vraisemblance partielle de β . Si le modèle est valide, $M_1(t), \dots, M_n(t)$ sont des réalisations de martingales indépendantes (Gill, 1980). La classe de résidus de martingale de Barlow et Prentice (1988) est alors définie par

$$\int_0^t \phi(s) M_j(s) ds, \quad j = 1, \dots, n, \quad 0 \leq t \leq \mathcal{T},$$

où ϕ est une fonction du temps déterministe ou aléatoire et prévisible. L'intégrale d'un processus prévisible par rapport aux résidus de martingale est une martingale et présente de bonnes propriétés comme le centrage et la non-corrélation. Ces propriétés sont observées sur le graphique des résidus si le modèle est valide. Les résidus de Schoenfeld (1982), les résidus de Schoenfeld pondérés (Lin, 1991) et les résidus de Kay (1977) font partie de cette classe de résidus. Les résidus de Schoenfeld ont une place importante dans les méthodes d'adéquation. Nous rappelons que ces résidus sont définis à chaque temps de décès t par

$$r_\beta(t) = Z_j(t) - \mathcal{E}_\beta(Z | t), \quad (5.1)$$

où j est l'individu qui décède en t . Grambsch et Therneau (1994) ont suggéré de tracer les résidus de Schoenfeld standardisés au cours du temps pour détecter la validité de l'hypothèse des risques proportionnels et, en cas de rejet, avoir une indication sur la forme de l'effet

dépendant du temps. En effet, ils ont montré que, pour chaque temps de décès t ,

$$E\left(\left[\mathcal{V}_{\hat{\beta}}(Z | t)^{-1}r_{\hat{\beta}}(t)\right]_j\right) + \hat{\beta}_j \approx \beta_j(t), \quad j = 1, \dots, p,$$

où $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ est l'estimateur du maximum de vraisemblance partielle du modèle à risques proportionnels et $\beta(t) = (\beta_1(t), \dots, \beta_p(t))$ est le vrai paramètre dépendant du temps du modèle à risques non proportionnels. Cette approche est très utilisée et est implémentée dans le langage de programmation R dans le package `survival`. De plus, une bande de confiance a été proposée pour l'effet de régression, mais cette bande de confiance n'est pas fiable si l'hypothèse des risques proportionnels n'est pas vérifiée. Plus récemment, [Sasieni et Winnett \(2003\)](#) ont proposé d'utiliser des résidus lissés de différences de martingales. Ces résidus peuvent être tracés en fonction du temps pour une valeur fixe de la covariable, permettant d'analyser la dépendance temporelle. Ils peuvent également être tracés en fonction de la covariable à une valeur fixe du temps, pour étudier sous quelle forme introduire la covariable dans le modèle. Dans le premier cas, plusieurs figures doivent être produites pour différentes valeurs des covariables et, dans le second cas, pour différents temps. En pratique, l'interprétation de tous ces graphiques peut être compliquée. Toutes ces méthodes basées sur des résidus non cumulés font appel à des fonctions de lissage pour relier les résidus. Comme l'ont souligné [Lin et al. \(1993\)](#), les résultats sont souvent sensibles au choix des techniques de lissage utilisées. Des lissages différents pouvant mener à différentes conclusions, la vérification du modèle à risques proportionnels avec ces méthodes est délicate.

Pour remédier à ce problème, plusieurs auteurs ont proposé l'utilisation de résidus de martingales cumulés. [Arjas \(1988\)](#) a suggéré de tracer le nombre attendu de décès sous le modèle à risques proportionnels en fonction des rangs des temps de décès. Le modèle s'ajuste bien aux données si la courbe résultante est proche de la première bissectrice des axes. Le processus du score, inspiré du test d'adéquation de [Wei \(1984\)](#), a été introduit par [Therneau et al. \(1990\)](#) et est noté $\{U(\beta, t), 0 \leq t \leq \mathcal{T}\}$, où

$$U(\beta, t) = \sum_{i=1}^n \int_0^t \{Z_i(s) - \mathcal{E}_{\beta}(Z | s)\} dN_i(s), \quad 0 \leq t \leq \mathcal{T}. \quad (5.2)$$

Sous le modèle à risques proportionnels, une version standardisée de $\{U(\hat{\beta}, t), 0 \leq t \leq \mathcal{T}\}$ converge vers un pont Brownien. Un test dont la statistique est le supremum de la valeur absolue du processus standardisé peut alors être appliqué. La loi limite de la statistique est la distribution de Kolmogorov. [Wei \(1984\)](#) a proposé ce test pour une covariable binaire et [Therneau et al. \(1990\)](#) l'ont étendu à une covariable dépendant du temps. En présence de covariables corrélées, la loi limite de la statistique n'est plus le supremum d'un pont Brownien. [Lin et al. \(1993\)](#) ont alors montré que le processus du score (5.2) non standardisé converge vers un processus gaussien centré de matrice de variance-covariance basé sur les observations. Ainsi, la comparaison entre ce processus observé au cours du temps et un grand nombre de processus simulés sous la loi limite gaussienne peut donner une indication de la distribution

asymptotique du supremum du processus. En pratique, l'interprétation de tels graphiques n'est pas toujours évidente, notamment parce que l'enveloppe du processus n'est pas déterministe, mais est bruitée.

La régression linéaire avec un bruit gaussien et des covariables gaussiennes est un modèle de régression particulier. Chaque sous-modèle est encore un modèle linéaire et le bruit résiduel est encore une variable gaussienne. Cette structure de modèle est très particulière et ne se retrouve pas souvent dans les modèles de régression. C'est le cas notamment du modèle à risques proportionnels. En introduisant plusieurs covariables, si le modèle à risques proportionnels est valide, les sous-modèles ne le seront plus. Ainsi, l'évaluation de l'adéquation du modèle à risques proportionnels multivarié par l'évaluation des modèles univariés ne suffit pas. Cependant, en l'absence d'outils pour vérifier la validité globale du modèle, la plupart des méthodes pour l'évaluation de l'hypothèse de proportionnalité des risques pour une covariable supposent l'hypothèse vérifiée pour les autres covariables, ce qui est très souvent une hypothèse fautive (Scheike et Martinussen, 2004). De plus, beaucoup de méthodes dépendent des covariances entre les covariables, comme celles proposées par Grambsch et Therneau (1994) et Lin *et al.* (1993). Pour répondre à ce problème, Scheike et Martinussen (2004) ont considéré un modèle à risques non proportionnels et ont développé des procédures d'estimation et des tests d'adéquation du modèle à risques proportionnels pour chaque covariable, en laissant la possibilité aux autres covariables d'avoir un effet qui dépend du temps. Leurs simulations indiquent une bonne performance de leur méthode comparée aux méthodes couramment utilisées de Grambsch et Therneau (1994) et de Lin *et al.* (1993) lorsque l'hypothèse des risques proportionnels n'est pas valide et/ou en présence de covariables corrélées. Les statistiques de leurs tests dépendent de l'estimation du paramètre de régression, qui fait appel à un algorithme difficile à appréhender et basé sur un lissage par noyau. De plus, la forme de l'estimateur du paramètre de régression n'est pas une fonction explicite et lisse du temps. Les tests d'adéquation sont de type Kolmogorov-Smirnov et Cramér-von Mises et les p-values sont calculées par simulations numériques, étant donné que les lois limites des statistiques ne sont pas connues.

Nous renvoyons aux ouvrages de Therneau et Grambsch (2000), Klein et Moeschberger (2003) ainsi que Martinussen et Scheike (2006) pour plus de détails sur les méthodes d'adéquation du modèle à risques proportionnels.

La procédure d'évaluation de l'adéquation du modèle à risques proportionnels développée dans cette thèse est basée sur une méthode graphique qui permet de détecter la forme temporelle de l'effet en cas de mauvaise adéquation, ainsi que sur un test d'adéquation pour l'effet de chaque covariable. La procédure ne requiert pas d'estimation du paramètre et peut s'appliquer avec plusieurs covariables corrélées.

5.1.2 Tests d'adéquation

Dans son article introduisant le modèle à risques proportionnels, Cox (1972) a suggéré de remplacer le coefficient de régression fixe au cours du temps par un coefficient non fixe puis de tester sa dépendance au temps. Formellement, cela revient à considérer la classe de tests

basée sur le modèle à risques non proportionnels (1.11) de paramètre

$$\beta(t) = \beta_0 + \beta_1 g(t),$$

où β_0 et β_1 sont deux paramètres fixes et g est une fonction du temps. Il s'agit alors de tester l'hypothèse nulle de la validité du modèle à risques proportionnels en considérant l'hypothèse nulle $H_0 : \beta_1 = 0$ et son alternative $H_1 : \beta_1 \neq 0$ (Grambsch et Therneau, 1994, Therneau et Grambsch, 2000). Un large choix de g est possible. Si g est une fonction connue du temps, cela revient à effectuer un test du score comme proposé par Cox (1972). La fonction g peut également être un processus prévisible. Par exemple, $g(t) = \bar{N}(t^-)$ revient à considérer le test de Breslow *et al.* (1984) qui est basé sur les rangs des temps d'évènements. Enfin, le dernier exemple que nous citons est le cas d'une fonction g constante par morceaux, avec des temps de rupture prédéfinis (Moreau *et al.*, 1985, O'Quigley et Pessione, 1989, Schoenfeld, 1980).

Wei (1984) a développé un test d'adéquation pour une variable binaire, dont la statistique est $\sup_t |U(\hat{\beta}, t)|$, où $\hat{\beta}$ est l'estimateur du maximum de vraisemblance partielle de Cox. Sous le modèle, asymptotiquement, cela revient à appliquer le test du supremum d'un pont Brownien et la loi asymptotique de la statistique est la distribution de Kolmogorov. Ce test a motivé plusieurs méthodes graphiques que nous avons mentionnées dans la section précédente (Lin *et al.*, 1996, 1993, Therneau *et al.*, 1990). Schoenfeld (1980) a développé un test du χ^2 en partitionnant l'espace des covariables et du temps et en comparant les valeurs observées dans les éléments de cette partition aux valeurs attendues sous le modèle. Lin (1991) a proposé un test du score basé sur un estimateur du maximum de la log-vraisemblance pondérée du coefficient de régression β , en remarquant que sous le modèle, la loi limite de la différence entre cet estimateur et l'estimateur usuel du maximum de vraisemblance partielle est une loi normale centrée.

Murphy et Sen (1991) et Gray (1992) ont également proposé des tests d'adéquation du modèle à risques proportionnels, en modélisant le coefficient de régression $\beta(t)$ dans un modèle à risques non proportionnels par projection sur des bases de fonctions puis en testant si les projetés dépendent du temps.

La méthode d'adéquation que nous proposons est basée sur les résultats des Théorèmes 3.3.2 et 3.4.1, qui indiquent que la forme du coefficient de régression se retrouve dans la dérive du processus du score standardisé. Vérifier l'adéquation du modèle à risques proportionnels revient à s'intéresser à la linéarité de la dérive du processus. Avant de présenter la méthode, introduisons la notion de variance expliquée dans le modèle à risques non proportionnels, ainsi que son estimateur, le coefficient R^2 , qui permet de mesurer la capacité prédictive du modèle.

5.2 Mesure de la capacité de prédiction du modèle à risques non proportionnels

5.2.1 Variance expliquée

Soit X une variable aléatoire admettant un moment d'ordre 2. L'inégalité de Bienaymé-Chebyshev

$$\forall \varepsilon > 0, \quad P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2},$$

indique que la dispersion de X autour de son espérance peut être contrôlée par sa variance. Plus la variance est petite, plus la probabilité pour que X soit plus proche de son espérance qu'une distance fixée est grande. En utilisant un modèle statistique, l'idée d'évaluer la proximité entre une variable réponse et sa prédiction à travers la variance semble alors naturelle. Plus la variance est petite, meilleure sera la qualité de la prédiction.

Pour deux variables aléatoires X et Y ayant des moments d'ordre 2, la formule de décomposition de la variance

$$\text{Var}(Y) = \text{Var}(E(Y|X)) + E(\text{Var}(Y|X)), \quad (5.3)$$

permet de définir le paramètre de variance expliquée. Supposons que l'on cherche à modéliser la variable réponse réelle Y à l'aide du vecteur de variables explicatives X . Si le modèle et le vecteur X permettent de bien représenter la variable Y , alors les valeurs de l'espérance de Y sachant X sont très dispersées. Leur variance est forte et correspond à un signal. De plus, en moyenne, la variance résiduelle de Y conditionnellement à X est faible et peut être assimilée à un bruit. La formule (5.3) devient alors

$$\text{Var}(Y) = \text{signal} + \text{bruit}.$$

Définition 5.2.1. La proportion de variance de Y expliquée par X et le modèle, notée $\Omega_{Y|X}^2$, est

$$\Omega_{Y|X}^2 = \frac{\text{Var}(E(Y|X))}{\text{Var}(Y)} = \frac{\text{signal}}{\text{signal} + \text{bruit}}. \quad (5.4)$$

Le coefficient $\Omega_{Y|X}^2$ est couramment appelé variance expliquée. Il s'agit du rapport de la variance des valeurs attendues de la variable réponse sous le modèle sachant les variables explicatives sur la variance marginale de la variable réponse. Dans le modèle à risques non proportionnels, deux coefficients peuvent être définis, selon que T soit la variable réponse et Z la variable explicative ou l'inverse.

Variance expliquée de T sachant Z .

En remplaçant Y par T et X par Z dans la formule (5.4), on définit la variance expliquée de T sachant Z , $\Omega_{T|Z}^2$, de la façon suivante :

$$\Omega_{T|Z}^2 = \frac{\text{Var}(E(T | Z))}{\text{Var}(T)}. \quad (5.5)$$

Intuitivement, cette quantité semble être celle que l'on doit utiliser pour représenter la capacité prédictive du modèle à risques proportionnels. En effet, lorsque l'on construit un modèle de survie à partir de données, un de nos objectifs est de prédire la survie d'un individu, en prenant en compte la valeur de certaines de ses caractéristiques regroupées dans le vecteur Z . On aimerait donc calculer le pouvoir prédictif de Z sur T . Le résultat suivant indique que ce coefficient n'est pas un bon indicateur de la qualité de la prédiction en utilisant le modèle à risques proportionnels.

Proposition 5.2.1. *Sous le modèle à risques proportionnels avec une covariable $Z \in \mathbb{R}$ et un risque de base constant $\lambda_0(t) = a$, $a > 0$, on a*

$$\Omega_{T|Z}^2 \leq \frac{1}{2}.$$

Preuve de la Proposition 5.2.1. Sous les hypothèses de la Proposition 5.2.1, la loi de T conditionnellement à Z est une loi exponentielle $\mathcal{E}(\alpha \exp(\beta Z))$. Ainsi, $E(T | Z) = \exp(-\beta Z)/\alpha$ et $E(T^2 | Z) = 2 \exp(-2\beta Z)/\alpha^2$. On a alors

$$\begin{aligned} \text{Var}(T) &= E(T^2) - E(T)^2 \\ &= E(E(T^2 | Z)) - E(E(T | Z))^2 \\ &= \frac{1}{\alpha^2} \left(2E(\exp(-2\beta Z)) - E(\exp(-\beta Z))^2 \right) \\ &= \frac{1}{\alpha^2} \left(2\phi_Z(-2\beta) - \phi_Z(-\beta)^2 \right), \end{aligned} \quad (5.6)$$

où ϕ_Z est la fonction génératrice des moments de Z . De plus,

$$\text{Var}(T | Z) = \frac{1}{\alpha^2} \exp(-2\beta Z),$$

et

$$E(\text{Var}(T | Z)) = \frac{1}{\alpha^2} E(\exp(-2\beta Z)) = \frac{1}{\alpha^2} \phi_Z(-2\beta). \quad (5.7)$$

En utilisant les équations (5.6) et (5.7) et d'après la définition de $\Omega_{T|Z}^2$ donnée équation (5.5), on obtient

$$\Omega_{T|Z}^2 = 1 - \frac{\phi_Z(-2\beta)}{2\phi_Z(-2\beta) - \phi_Z(-\beta)^2} = 1 - \frac{1}{2 - \frac{\phi_Z(-\beta)^2}{\phi_Z(-2\beta)}}. \quad (5.8)$$

La positivité de la fonction génératrice des moments permet de conclure. \square

Si le coefficient de régression β dans le modèle à risques proportionnels est élevé, cela signifie que l'association entre T et Z est forte et la probabilité que l'individu décède tôt ou tard est fortement influencée par la valeur de Z . La valeur de Z influençant plus la survie lorsque le coefficient de régression est élevé en valeur absolue, la valeur prédictive de Z sur T devrait augmenter lorsque $|\beta|$ augmente et ne devrait pas être bornée par une valeur strictement inférieure à 1. Le résultat de la Proposition 5.2.1 empêche la réalisation de cette propriété.

De plus, d'après la formule (5.8), on peut montrer que si Z suit une loi de Bernoulli de paramètre $1/2$, alors $\Omega_{T|Z}^2$ est majoré par $1/3$ et le majorant change avec le paramètre de la loi de Bernoulli. Ainsi, la comparaison de deux valeurs de $\Omega_{T|Z}^2$ avec des distributions marginales de Z différentes est délicate puisque les coefficients ne sont pas définis sur la même échelle. Ceci est illustré par la Figure 5.1 qui représente des courbes de survie conditionnelles sous

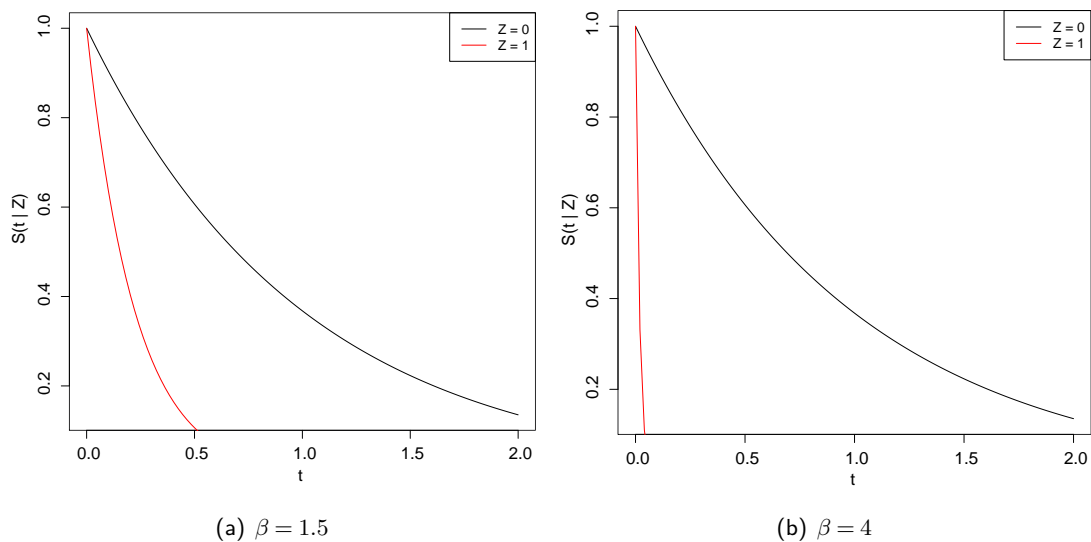


FIGURE 5.1 – Courbes de survie conditionnelles $S(\cdot | Z)$ en fonction du temps sous le modèle à risques proportionnels avec $\lambda_0(t) = 1$, $\beta = 1.5$ ou 4 et des covariables binaires.

le modèle à risques proportionnels avec $\beta = 1.5$ ou $\beta = 4$, $\lambda_0(t) = 1$ pour tout t et Z une covariable binaire. On remarque que les courbes de survie conditionnelles sont plus éloignées lorsque la valeur de β est la plus grande. Ceci devrait améliorer la qualité de la prédiction puisque la covariable est plus fortement liée à la survie. Cependant, en utilisant l'équation (5.8), on peut montrer que pour $\beta = 1.5$ et $Z \sim \mathcal{B}(0.5)$ et pour $\beta = 4$ et $Z \sim \mathcal{B}(0.3)$, la valeur de la variance expliquée est la même, $\Omega_{T|Z}^2 = 0.22$.

Aussi, O'Quigley et Flandre (1994) ont montré qu'une transformation non linéaire qui change l'axe du temps tout en conservant l'ordre d'arrivée des temps de décès modifie la variance expliquée de T sachant Z . Avec le modèle à risques proportionnels, modifier de cette manière le temps ne change pourtant pas l'estimation du coefficient de régression β . La capacité prédictive du modèle ne devrait pas être différente. La variance expliquée de T sachant Z semble donc être un mauvais indicateur du pouvoir prédictif des variables dans le modèle à risques proportionnels et, à plus forte raison, dans le modèle à risques non proportionnels.

Étudions maintenant la variance expliquée de T sachant Z .

Variance expliquée de Z sachant T

On rappelle que $Z = Z(t)$ est un vecteur de p covariables pouvant dépendre du temps. La variance expliquée de Z sachant T sous le modèle à risques non proportionnels n'est pas définie de la même manière dans les cas univarié ($p = 1$) et multivarié ($p > 1$). Dans la suite du document, nous considérons uniquement la variance expliquée basée sur la loi de Z conditionnellement à T et, pour alléger l'écriture, nous ne précisons plus cette loi conditionnelle.

Définition 5.2.2. Dans le modèle à risques non proportionnels univarié, la variance expliquée, exprimée comme une fonction du coefficient de régression dépendant du temps $\beta(t)$, est définie par

$$\Omega^2(\beta(t)) = \frac{\text{Var}(E_{\beta(t)}(Z|T))}{\text{Var}(Z)} = 1 - \frac{E(\text{Var}_{\beta(t)}(Z|T))}{\text{Var}(Z)}. \quad (5.9)$$

Lorsque $Z(t)$ est un vecteur de p covariables ($p > 1$), au temps t l'individu i est caractérisé par son indice pronostique réel $\eta_i(t) = \beta(t)^T Z_i(t)$ qui est une réalisation de la variable aléatoire $\eta(t) = \beta(t)^T Z(t)$. Ainsi, la qualité de la prédiction du modèle peut être évaluée de manière équivalente par Z ou η (Altman et Andersen, 1986, Andersen et al., 1983). Nous adoptons la dernière possibilité afin de nous ramener à une quantité unidimensionnelle.

Définition 5.2.3. La variance expliquée du modèle à risques non proportionnels avec plusieurs covariables peut être définie comme une fonction de $\beta(t)$ par

$$\Omega^2(\beta(t)) = \frac{\text{Var}(E(\eta|T))}{\text{Var}(\eta)} = 1 - \frac{E(\text{Var}(\eta|T))}{\text{Var}(\eta)}, \quad \eta(t) = \beta(t)^T Z(t). \quad (5.10)$$

On peut définir l'espérance par rapport à la distribution de T de la variance de la covariable ou de l'indice pronostique évalué avec le coefficient $\alpha_2(t)$ sous le modèle à risques non proportionnels de paramètre $\alpha_1(t)$ de dimension p par

$$Q(F, \alpha_1(t), \alpha_2(t)) \quad (5.11)$$

$$= \begin{cases} \int_0^1 E_{\alpha_1(t)} \left(\left\{ Z(t) - E_{\alpha_1(t)}(Z | T = t) \right\}^2 \middle| T = t \right) dF(t) & \text{si } p = 1 \\ \int_0^1 E_{\alpha_1(t)} \left(\left\{ \alpha_2(t)^T Z(t) - E_{\alpha_1(t)}(\alpha_2(t)^T Z | T = t) \right\}^2 \middle| T = t \right) dF(t) & \text{si } p > 1, \end{cases}$$

où F est la fonction de répartition de T . Ainsi, on a l'égalité

$$\Omega^2(\beta(t)) = 1 - \frac{Q(F, \beta(t), \beta(t))}{Q(F, 0, \beta(t))}.$$

Cette écriture de la variance expliquée sera utilisée pour construire un estimateur. Les difficultés rencontrées avec la variance expliquée basée sur la loi de T sachant Z définie au paragraphe

précédent disparaissent en considérant la variance expliquée de Z conditionnellement à T d'après le résultat suivant :

Proposition 5.2.2. *Sous le modèle à risques non proportionnels de paramètre $\beta(t)$ et en considérant un vecteur de p covariables ($p \in \mathbb{N}^*$), les trois assertions suivantes sont vérifiées*

1. $\Omega^2(0) = 0$ et $0 \leq \Omega_{Z|T}^2(\beta(t)) \leq 1$,
2. Ω^2 est invariant à des transformations strictement croissantes de T et linéaires en Z ,
3. Si le paramètre $\beta(t) = \beta_0$ est constant et $p = 1$, Ω^2 est une fonction croissante de $|\beta_0|$ et $\lim_{\beta \rightarrow \infty} \Omega^2(\beta) = 1$.

Preuve de la Proposition 5.2.2. Nous renvoyons au Chapitre 13 de l'ouvrage d'O'Quigley (2008) et aux travaux d'O'Quigley et Xu (2001). Dans ces références, Ω^2 est une fonction d'un paramètre de régression constant, correspondant au modèle à risques proportionnels. L'extension à un modèle avec un paramètre dépendant du temps est immédiate. Pour le cas multivarié, nous renvoyons à Xu (1996). \square

La transposition de la propriété 3. au modèle à risques non proportionnels est délicate. Le paramètre de régression étant une fonction, la notion de croissance en le paramètre n'est pas claire. Il faudrait définir une mesure de distance convenable entre les paramètres de régression afin de pouvoir les ordonner.

Le coefficient de variance expliquée ne change pas en appliquant une transformation strictement croissante sur les temps. Ainsi, nous travaillons avec l'échelle du temps standardisée décrite au Chapitre 3. La variance expliquée est un paramètre dans la population qui doit être estimé.

5.2.2 Coefficient R^2

De nombreuses mesures de la qualité de la prédiction du modèle à risques proportionnels ont été proposées dans la littérature (Choodari-Oskooei et al., 2012). Certaines sont appelées "coefficient R^2 ", qui désigne classiquement un estimateur de Ω^2 , ou "variance expliquée", sans pour autant être des estimateurs de la variance expliquée définie à la section 5.2.1. Il s'agit plus exactement de mesures ad hoc, construites généralement sans justification théorique. De plus, ces mesures ont été construites sous le modèle à risques proportionnels, mais, pour la plupart des mesures, leur comportement en cas de mauvaise spécification du modèle n'a pas été étudié. Notre but n'est pas de présenter une revue exhaustive de ces estimateurs. Nous nous concentrons sur un estimateur de Ω^2 , le coefficient R^2 construit avec les mêmes résidus que le processus du score standardisé et introduit par O'Quigley et Flandre (1994). Nous rappelons sa définition en l'étendant au modèle à risques non proportionnels.

Les résidus de Schoenfeld, définis équation (1.22), peuvent être étendus au modèle à risques non proportionnels de paramètre $\beta(t)$ et sur l'échelle transformée par

$$r_{\beta(t)}(t) = \mathcal{Z}(t) - \mathcal{E}_{\beta(t)}(\mathcal{Z} | t), \quad 0 \leq t \leq 1,$$

où $Z(t)$ est la valeur de la covariable de l'individu qui décède en t et $\mathcal{E}_{\beta(t)}(Z | t)$ est son espérance sous le modèle. Notons \tilde{F} l'estimateur de la fonction de répartition empirique de T sur l'échelle transformée :

$$\tilde{F}(t) = \frac{1}{k_n} \bar{N}^*(t) = \frac{1}{k_n} \sum_{i=1}^n \mathbf{1}_{\phi_n(X_i) \leq t, \delta_i = 1}, \quad 0 \leq t \leq 1.$$

En présence de données non censurées, \tilde{F} correspond à l'estimateur usuel de la fonction de répartition empirique de T sur l'échelle transformée. La valeur $Q(F, \alpha_1(t), \alpha_2(t))$ de l'équation (5.11) peut être estimée par

$$\tilde{Q}(\tilde{F}, \alpha_1(t), \alpha_2(t)) = \begin{cases} \int_0^1 \{r_{\alpha_1(s)}(s)\}^2 d\tilde{F}(s) = \frac{1}{k_n} \sum_{i=1}^{k_n} \{r_{\alpha_1(t_i)}(t_i)\}^2 & \text{si } p = 1 \\ \int_0^1 \{\alpha_2(s)^T r_{\alpha_1(s)}(s)\}^2 d\tilde{F}(s) = \frac{1}{k_n} \sum_{i=1}^{k_n} \{\alpha_2(t_i)^T r_{\alpha_1(t_i)}(t_i)\}^2 & \text{si } p > 1. \end{cases}$$

Définition 5.2.4. Le coefficient R^2 est défini par $R^2 = R^2(\hat{\beta}(t))$, où $\hat{\beta}(t)$ est un estimateur convergent de $\beta(t)$ et, pour toute fonction α de $[0, 1]$ à valeurs dans \mathbb{R}^p ,

$$R^2(\alpha(t)) = 1 - \frac{\tilde{Q}(\tilde{F}, \alpha(t), \alpha(t))}{\tilde{Q}(\tilde{F}, 0, \alpha(t))} = \begin{cases} 1 - \frac{\sum_{i=1}^{k_n} r_{\alpha(t_i)}(t_i)^2}{\sum_{i=1}^{k_n} r_0(t_i)^2} & \text{si } p = 1 \\ 1 - \frac{\sum_{i=1}^{k_n} \{\alpha(t_i)^T r_{\alpha(t_i)}(t_i)\}^2}{\sum_{i=1}^{k_n} \{\alpha(t_i)^T r_0(t_i)\}^2} & \text{si } p > 1. \end{cases} \quad (5.12)$$

Avec une covariable dans le modèle, le coefficient R^2 est le rapport de la somme des résidus de Schoenfeld au carré évalués sous le modèle de paramètre $\hat{\beta}(t)$ sur la somme des carrés des résidus du modèle de paramètre nul. Lorsque plusieurs covariables sont introduites dans le modèle, le coefficient est toujours le rapport de sommes des résidus au carré, mais les résidus sont évalués avec l'indice pronostique $\hat{\beta}(t)^T Z(t)$ et non plus directement avec la covariable $Z(t)$.

En présence de censure, l'estimateur \tilde{F} ne converge pas vers la fonction de répartition empirique de T , comme nous l'avons rappelé dans le Chapitre 1. Ainsi, le coefficient R^2 ne converge pas vers la variance expliquée Ω^2 . Le coefficient R^2 étant influencé par des mécanismes de censure, un coefficient alternatif, noté R_S^2 a été proposé par O'Quigley et Xu (2001). Il est défini par

$$R_S^2(\alpha(t)) = 1 - \frac{\tilde{Q}(\hat{F}, \alpha(t), \alpha(t))}{\tilde{Q}(\hat{F}, 0, \alpha(t))}$$

$$= \begin{cases} 1 - \frac{\sum_{i=1}^{k_n} d\hat{S}(t_i) \{\alpha(t_i)^T r_{\alpha(t_i)}(t_i)\}^2}{\sum_{i=1}^{k_n} d\hat{S}(t_i) \{\alpha(t_i)^T r_0(t_i)\}^2} & \text{si } p > 1 \\ 1 - \frac{\sum_{i=1}^{k_n} d\hat{S}(t_i) r_{\alpha(t_i)}(t_i)^2}{\sum_{i=1}^{k_n} d\hat{S}(t_i) r_0(t_i)^2} & \text{si } p = 1, \end{cases}$$

où nous rappelons que $\hat{F} = 1 - \hat{S}$ est l'estimateur de [Kaplan et Meier \(1958\)](#) de la fonction de répartition de T (Définition 1.1.1) et

$$d\hat{S}(t_i) = \hat{S}(\phi_n^{-1}(t_i)) - \hat{S}(\phi_n^{-1}(t_i)^-)$$

est le saut de l'estimateur au temps de décès t_i , pour $i = 1, \dots, k_n$ (équation (1.2)). Cette modification du coefficient R^2 par pondération avec les sauts de l'estimateur de Kaplan-Meier permet d'obtenir la convergence de $R_S^2(\hat{\beta}(t))$ vers la proportion de variance expliquée Ω^2 lorsque $\hat{\beta}(t)$ est un estimateur convergent de $\beta(t)$ ([O'Quigley et Xu, 2001](#), [Xu, 1996](#)). Cependant, dans les applications pratiques, il a été observé que R^2 ne dépend pas ou que très peu de la censure, même pour des taux de censure très élevés (jusqu'à 90%). Un travail de simulation effectué par [Choodari-Oskooei et al. \(2012\)](#) a confirmé l'indépendance du coefficient R^2 sans les pondérations à la censure et a révélé une variance plus élevée pour R_S^2 , en contrepartie de son absence de biais. Pour ces raisons, dans notre travail, nous choisissons d'utiliser le coefficient R^2 non pondéré.

Le coefficient de variance expliquée $\Omega^2(\beta(t))$ peut-être estimé par $R^2(\hat{\beta}(t))$, où $\hat{\beta}(t)$ est un estimateur consistant du vrai paramètre de régression $\beta(t)$. Le théorème suivant justifie l'emploi du coefficient R^2 pour évaluer l'adéquation du modèle à risques non proportionnels et donne son comportement limite sous un modèle mal spécifié.

Théorème 5.2.1. *Sous le modèle à risques non proportionnels de paramètre $\beta(t)$, si $p = 1$ et sous les conditions **H1.**, **H2.**, **H3.** et **H4.** de la Section 3.3, il existe une constante $C(\beta) > 0$ telle que pour tout $\alpha \in \mathbb{B}$,*

$$\lim_{n \rightarrow \infty} R^2(\alpha(t)) = 1 - \frac{C(\beta) + \int_0^1 (e(\alpha(t), t) - e(\beta(t), t))^2 dt}{C(\beta) + \int_0^1 (e(0, t) - e(\beta(t), t))^2 dt}, \quad (5.13)$$

et

$$\arg \max_{b \in \mathbb{B}} \lim_{n \rightarrow +\infty} R^2(b(t)) = \beta \quad p.s.$$

Si $p > 1$ et sous les conditions **H1'**, **H2'**, **H3'** et **H4'** de la Section 3.4, pour tout $\alpha \in \mathbb{B}'$, $\alpha \neq 0$, nous avons

$$\lim_{n \rightarrow \infty} R^2(\alpha(t)) = 1 - \frac{\int_0^1 \alpha(t)^T \Sigma \alpha(t) dt + \int_0^1 (\alpha(t)^T \{e(\beta(t), t) - e(\alpha(t), t)\})^2 dt}{\int_0^1 \alpha(t)^T \Sigma \alpha(t) dt + \int_0^1 (\alpha(t)^T \{e(\beta(t), t) - e(0, t)\})^2 dt}. \quad (5.14)$$

Pour $t \in [0, 1]$ et γ fonction de $[0, 1]$ à valeurs dans \mathbb{R}^p , nous rappelons la définition

$$e(\gamma(t), t) = s^{(1)}(\gamma(t), t) / s^{(0)}(0, t),$$

Le Théorème 5.2.1 indique que, dans le cas univarié, si $\beta(t)$ est le vrai coefficient de régression et si la taille de l'échantillon est assez grande, le maximum de la fonction R^2 est atteint pour le vrai paramètre $\beta(t)$. La généralisation au cas multivarié n'est pas immédiate. Considérons le cas de deux covariables $Z^{(1)}$ et $Z^{(2)}$ dont les effets sont respectivement $\beta_1(t)$ et $\beta_2(t)$. Si on suppose l'effet $\beta_2(t)$ connu, cela revient à estimer $\beta_1(t)$ dans un modèle univarié. Le Théorème 5.2.1 s'applique alors et $\beta_1(t)$ réalise le maximum de la limite du coefficient R^2 . En pratique, l'effet $\beta_2(t)$ ne sera pas connu, mais on peut l'approcher par un estimateur convergent. En conditionnant sur cet estimateur, c'est-à-dire en traitant encore $\beta_2(t)$ comme fixe et connu, le résultat s'applique à nouveau sur $\beta_1(t)$. On peut donc considérer que le résultat du Théorème 5.2.1 s'applique pour chaque variable prise séparément, après avoir conditionné sur des estimateurs convergents des autres coefficients de régression.

Preuve du Théorème 5.2.1. Nous traitons dans un premier temps le cas d'une covariable dans le modèle, pour lequel $p = 1$. Etudions le numérateur du coefficient R^2 défini à l'équation (5.12). Nous avons

$$\begin{aligned} & \frac{1}{k_n} \sum_{i=1}^{k_n} \left(Z(t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right)^2 \\ &= \frac{1}{k_n} \sum_{i=1}^{k_n} \left(Z(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right)^2 + \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right)^2 \\ &+ \frac{2}{k_n} \sum_{i=1}^{k_n} \left(Z(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right). \end{aligned} \quad (5.15)$$

Etudions les termes de droite de l'équation (5.15). Pour le premier terme, nous avons

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left(Z(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right)^2 = \frac{1}{k_n} \sum_{i=1}^{k_n} V_i + \frac{1}{k_n} \sum_{i=1}^{k_n} \mathcal{V}_{\beta(t_i)}(Z | t_i),$$

où $V_i = \left(Z(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right)^2 - \mathcal{V}_{\beta(t_i)}(Z | t_i)$. On rappelle que la variable V_i est $\mathcal{F}_{t_i}^*$ -mesurable et $E(V_i | \mathcal{F}_{t_i}^*) = 0$ pour $i = 1, \dots, k_n$. Montrons que les variables V_1, V_2, \dots, V_{k_n} vérifient les hypothèses de la loi des grands nombres pour des variables non corrélées (Théorème 2.2.1). Soit $i, j \in \{1, \dots, k_n\}$ tels que $i < j$. La variable V_i est centrée :

$$E(V_i) = E \left(E \left(V_i \middle| \mathcal{F}_{t_i}^* \right) \right) = 0.$$

Les variables V_i et V_j ne sont pas corrélées. En effet, V_i est $\mathcal{F}_{t_i}^*$ -mesurable, $\mathcal{F}_{t_i}^* \subset \mathcal{F}_{t_j}^*$ et

$E\left(V_j \middle| \mathcal{F}_{t_j}^*\right) = 0$. Donc

$$E(V_i V_j) = E\left(E\left(V_i V_j \middle| \mathcal{F}_{t_j}^*\right)\right) = E\left(V_i E\left(V_j \middle| \mathcal{F}_{t_j}^*\right)\right) = 0.$$

Il reste à montrer que la variance de la variable V_i est bornée. On a

$$\begin{aligned} \text{Var}(V_i) &= E(V_i^2) = E\left[\left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{V}_{\beta(t_i)}(Z | t_i)\right)^2\right] \\ &\leq E\left[\left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i)\right)^4\right] + E\left[\mathcal{V}_{\beta(t_i)}(Z | t_i)^2\right]. \end{aligned}$$

Or, nous avons les majorations $|\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i)| \leq 2L$ et $\mathcal{V}_{\beta(t_i)}(Z | t_i) \leq 4L^2$. Ainsi,

$$\text{Var}(V_i) \leq 32L^4 < \infty.$$

La loi faible des grands nombres pour des variables non corrélées, rappelée au Théorème 2.2.1, s'applique alors et nous avons la convergence en probabilité

$$\frac{1}{k_n} \sum_{i=1}^{k_n} V_i = \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i)\right)^2 - \frac{1}{k_n} \sum_{i=1}^{k_n} \mathcal{V}_{\beta(t_i)}(Z | t_i) \xrightarrow[n \rightarrow \infty]{P} 0. \quad (5.16)$$

D'après le Lemme 3.5.2, il existe $C(\beta) > 0$ telle que

$$\sup_{t \in \{t_1, t_2, \dots, t_{k_n}\}} \left| \mathcal{V}_{\beta(t)}(Z | t) - C(\beta) \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Le Lemme 3.2.1 s'applique et la convergence uniforme résultante entraîne la convergence ponctuelle au temps $t = 1$:

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \mathcal{V}_{\beta(t_i)}(Z | t_i) \xrightarrow[n \rightarrow \infty]{P} C(\beta). \quad (5.17)$$

Les équations (5.16) et (5.17) entraînent la convergence en probabilité

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i)\right)^2 \xrightarrow[n \rightarrow \infty]{P} C(\beta).$$

La convergence du second terme de (5.15) est également une conséquence du Lemme 3.2.1 :

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i)\right)^2 \xrightarrow[n \rightarrow \infty]{P} \int_0^1 (e(\beta(t), t) - e(\alpha(t), t))^2 dt,$$

où la convergence uniforme

$$\sup_{t \in \{t_1, \dots, t_{k_n}\}} \left| \left(\mathcal{E}_{\beta(t)}(Z | t) - \mathcal{E}_{\alpha(t)}(Z | t)\right)^2 - (e(\beta(t), t) - e(\alpha(t), t))^2 \right| \xrightarrow[n \rightarrow \infty]{P} 0,$$

peut être démontrée par des arguments similaires que ceux utilisés dans la preuve du Lemme 3.5.2. Finalement, le dernier terme de (5.15) converge en probabilité vers 0 lorsque $n \rightarrow \infty$ par la loi des grands nombres du Théorème 2.2.1. L'équation (5.13) est alors obtenue en remplaçant la fonction α par $0 \in \mathbb{B}$ dans les calculs précédents, et le maximum de $\lim_{n \rightarrow \infty} R^2$ est atteint en $\beta(t)$.

Considérons le cas multivarié ($p > 1$). On peut décomposer le numérateur du coefficient $R^2(\alpha(t))$ défini équation (5.12) de la façon suivante :

$$\begin{aligned} & \frac{1}{k_n} \sum_{i=1}^{k_n} \left(\alpha(t_i)^T \mathcal{Z}(t_i) - \alpha(t_i)^T \mathcal{E}_{\alpha(t)}(Z|t_i) \right)^2 \\ &= \frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z|t_i) \right) \right\}^2 \\ &+ \frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z|t_i) - \mathcal{E}_{\alpha(t_i)}(Z|t_i) \right) \right\}^2 \\ &+ \frac{2}{k_n} \sum_{i=1}^{k_n} \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z|t_i) \right) \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z|t_i) - \mathcal{E}_{\alpha(t_i)}(Z|t_i) \right). \end{aligned} \quad (5.18)$$

Comme dans le cas univarié, étudions séparément les comportements limites des 3 termes de droite. Le premier terme se décompose de la manière suivante

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left(\alpha(t_i)^T \mathcal{Z}(t_i) - \alpha(t_i)^T \mathcal{E}_{\beta(t_i)}(Z|t_i) \right)^2 = \frac{1}{k_n} \sum_{i=1}^{k_n} V_i' + \frac{1}{k_n} \sum_{i=1}^{k_n} \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z|t_i) \alpha(t_i),$$

où

$$V_i' = \left(\alpha(t_i)^T \mathcal{Z}(t_i) - \alpha(t_i)^T \mathcal{E}_{\beta(t_i)}(Z|t_i) \right)^2 - \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z|t_i) \alpha(t_i).$$

Montrons que les variables V_1', \dots, V_{k_n}' vérifient les hypothèses de la loi faible des grands nombres (Théorème 2.2.1). Soit $i, j \in \{1, \dots, k_n\}$ tels que $i < j$. La variable V_i' est $\mathcal{F}_{t_i}^*$ -mesurable et $E(V_i' | \mathcal{F}_{t_i}^*) = 0$. Donc

$$E(V_i') = E \left(E \left(V_i' \middle| \mathcal{F}_{t_i}^* \right) \right) = 0.$$

Les variables V_i' et V_j' ne sont pas corrélées :

$$E \left(E \left(V_i' V_j' \middle| \mathcal{F}_{t_j}^* \right) \right) = E \left(V_i' E \left(V_j' \middle| \mathcal{F}_{t_j}^* \right) \right) = 0,$$

d'après l'inclusion $\mathcal{F}_{t_i}^* \subset \mathcal{F}_{t_j}^*$. Enfin, montrons que la variance de V_i' est bornée. On rappelle que $\mathcal{V}_{\beta(t_i)}(Z|t_i)$ est définie positive et $\alpha \neq 0$. Ainsi,

$$\text{Var}(V_i') = E(V_i'^2) = E \left[\left(\left\{ \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z|t_i) \right) \right\}^2 - \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z|t_i) \alpha(t_i) \right)^2 \right]$$

$$\leq E \left[\left\{ \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \right\}^4 \right] + E \left[\left(\alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z | t_i) \alpha(t_i) \right)^2 \right].$$

Puisque $0, \alpha, \beta \in \mathbb{B}'$ qui est un voisinage de centre β et de rayon δ_2 , nous avons

$$\sup_{i=1, \dots, k_n} \|\alpha(t_i)\| \leq \sup_{t \in [0,1]} \|\alpha(t) - \beta(t)\| + \sup_{t \in [0,1]} \|\beta(t)\| \leq 2\delta_2.$$

La condition **H4'** implique alors que

$$\begin{aligned} \left| \alpha(t_i)^T \mathcal{Z}(t_i) \right| &\leq p \|\alpha(t_i)\| \sup_{j=1, \dots, n} \|Z_j(X_j)\| \\ &\leq 2p\delta_2 L', \end{aligned}$$

et, par conséquent,

$$\left\{ \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \right\}^2 \leq (4p\delta_2 L')^2.$$

De plus,

$$\begin{aligned} \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z | t_i) \alpha(t_i) &= E \left[\alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right)^T \alpha(t_i) \middle| \mathcal{F}_{t_i}^* \right] \\ &= E \left[\left\{ \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \right\}^2 \middle| \mathcal{F}_{t_i}^* \right] \\ &\leq (4p\delta_2 L')^2. \end{aligned}$$

Ainsi, $Var(V'_i) \leq 2(4p\delta_2 L')^4$. La loi des grands nombres 2.2.1 appliquée aux variables V'_1, \dots, V'_{k_n} entraîne

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \alpha(t_i)^T \left(\mathcal{Z}(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \right\}^2 - \frac{1}{k_n} \sum_{i=1}^{k_n} \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z | t_i) \alpha(t_i) \xrightarrow[n \rightarrow \infty]{P} 0. \quad (5.19)$$

De plus,

$$\begin{aligned} \sup_{i=1, \dots, k_n} \left| \alpha(t_i)^T \left(\mathcal{V}_{\beta(t_i)}(Z | t_i) - \Sigma \right) \alpha(t_i) \right| &\leq p^2 \left(\sup_{i=1, \dots, k_n} \|\alpha(t_i)\| \right)^2 \sup_{i=1, \dots, k_n} \left\| \mathcal{V}_{\beta(t_i)}(Z | t_i) - \Sigma \right\| \\ &\leq 4p^2 \delta_2^2 \sup_{i=1, \dots, k_n} \left\| \mathcal{V}_{\beta(t_i)}(Z | t_i) - \Sigma \right\|. \end{aligned}$$

D'après l'équation (3.21), on conclut que

$$\sup_{i=1, \dots, k_n} \left| \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z | t_i) \alpha(t_i) - \alpha(t_i)^T \Sigma \alpha(t_i) \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Le Lemme 3.2.1 s'applique alors et nous avons la convergence

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \alpha(t_i)^T \mathcal{V}_{\beta(t_i)}(Z | t_i) \alpha(t_i) \xrightarrow[n \rightarrow \infty]{P} \int_0^1 \alpha(t)^T \Sigma \alpha(t) dt. \quad (5.20)$$

Ainsi, la convergence du premier terme de droite de l'équation (5.18) est démontrée :

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \alpha(t_i)^T \left(Z(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \right\}^2 \xrightarrow[n \rightarrow \infty]{P} \int_0^1 \alpha(t)^T \Sigma \alpha(t) dt.$$

Pour obtenir la convergence du second terme, il suffit de remarquer que pour $\gamma \in \mathbb{B}'$,

$$\begin{aligned} \sup_{i=1, \dots, k_n} \left\| \mathcal{E}_{\gamma(t_i)}(Z | t_i) - e(\gamma(t_i), t_i) \right\| &= \sup_{i=1, \dots, k_n} \left\| \frac{S^{(1)}(\gamma(t_i), t_i)}{S^{(0)}(\gamma(t_i), t_i)} - \frac{s^{(1)}(\gamma(t_i), t_i)}{s^{(0)}(\gamma(t_i), t_i)} \right\| \\ &\leq \sup_{i=1, \dots, k_n} \left\| S^{(1)}(\gamma(t_i), t_i) \right\| \sup_{i=1, \dots, k_n} \left| \frac{1}{S^{(0)}(\gamma(t_i), t_i)} - \frac{1}{s^{(0)}(\gamma(t_i), t_i)} \right| \\ &\quad + \sup_{i=1, \dots, k_n} \left| \frac{1}{s^{(0)}(\gamma(t_i), t_i)} \right| \sup_{i=1, \dots, k_n} \left\| S^{(1)}(\gamma(t_i), t_i) - s^{(1)}(\gamma(t_i), t_i) \right\|. \end{aligned}$$

Sous les conditions **H1'** et **H2'**, en utilisant les mêmes arguments que dans la preuve du Lemme 3.5.2, on peut conclure que

$$\sup_{i=1, \dots, k_n} \left\| \mathcal{E}_{\gamma(t_i)}(Z | t_i) - e(\gamma(t_i), t_i) \right\| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (5.21)$$

On rappelle qu'il existe $M'_1, m'_0 \in \mathbb{R}^{+*}$ telles que pour tout $t \in [0, 1]$ et $\gamma \in \mathbb{B}'$, $s^{(0)}(\gamma(t), t) \geq m'_0$ et $\|s^{(1)}(\gamma(t), t)\| \leq M'_1$. Ainsi,

$$\sup_{t \in [0, 1], \gamma \in \mathbb{B}'} \|e(\gamma(t), t)\| = \sup_{t \in [0, 1], \gamma \in \mathbb{B}'} \left\| \frac{s^{(1)}(\gamma(t), t)}{s^{(0)}(\gamma(t), t)} \right\| \leq M'_1 m'_0{}^{-1}.$$

De plus, la condition **H4'** implique que

$$\sup_{i=1, \dots, k_n, \gamma \in \mathbb{B}'} \left\| \mathcal{E}_{\gamma(t_i)}(Z | t_i) \right\| \leq L'.$$

Par conséquent, nous avons l'inégalité

$$\begin{aligned} &\sup_{i=1, \dots, k_n} \left| \left\{ \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right) \right\}^2 - \left\{ \alpha(t_i)^T \left(e(\beta(t_i), t_i) - e(\alpha(t_i), t_i) \right) \right\}^2 \right| \\ &= \sup_{i=1, \dots, k_n} \left(\left| \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) - e(\beta(t_i), t_i) + e(\alpha(t_i), t_i) \right) \right| \right. \\ &\quad \left. \times \left| \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) + e(\beta(t_i), t_i) - e(\alpha(t_i), t_i) \right) \right| \right) \\ &\leq p^2 \left(\sup_{i=1, \dots, k_n} \|\alpha(t_i)\| \right)^2 \left(2 \sup_{i=1, \dots, k_n, \gamma \in \mathbb{B}'} \left\| \mathcal{E}_{\gamma(t_i)}(Z | t_i) \right\| + 2 \sup_{i=1, \dots, k_n, \gamma \in \mathbb{B}'} \|e(\gamma(t), t)\| \right) \\ &\quad \times \left(\sup_{i=1, \dots, k_n} \left\| \mathcal{E}_{\beta(t_i)}(Z | t_i) - e(\beta(t_i), t_i) \right\| + \sup_{i=1, \dots, k_n} \left\| \mathcal{E}_{\alpha(t_i)}(Z | t_i) - e(\alpha(t_i), t_i) \right\| \right) \\ &\leq 8p^2 \delta_2^2 (L' + M'_1 m'_0{}^{-1}) \\ &\quad \times \left(\sup_{i=1, \dots, k_n} \left\| \mathcal{E}_{\beta(t_i)}(Z | t_i) - e(\beta(t_i), t_i) \right\| + \sup_{i=1, \dots, k_n} \left\| \mathcal{E}_{\alpha(t_i)}(Z | t_i) - e(\alpha(t_i), t_i) \right\| \right). \end{aligned}$$

Le résultat (5.21) implique que

$$\sup_{i=1, \dots, k_n} \left| \left\{ \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right) \right\}^2 - \left\{ \alpha(t_i)^T \left(e(\beta(t_i), t_i) - e(\alpha(t_i), t_i) \right) \right\}^2 \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

L'application du Lemme 3.2.1 entraîne alors la convergence en probabilité du second terme de l'équation (5.18) :

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right) \right\}^2 \xrightarrow[n \rightarrow \infty]{P} \int_0^1 \left\{ \alpha(t)^T \left(e(\beta(t), t) - e(\alpha(t), t) \right) \right\}^2 dt.$$

Enfin, pour le dernier terme, la loi des grands nombres (Théorème 2.2.1) donne

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \alpha(t_i)^T \left(Z(t_i) - \mathcal{E}_{\beta(t_i)}(Z | t_i) \right) \alpha(t_i)^T \left(\mathcal{E}_{\beta(t_i)}(Z | t_i) - \mathcal{E}_{\alpha(t_i)}(Z | t_i) \right) \xrightarrow[n \rightarrow \infty]{P} 0,$$

ce qui permet de déduire le résultat de l'équation (5.14). \square

Le coefficient R^2 et le processus du score standardisé sont construits avec les mêmes résidus. Le processus du score standardisé permet de vérifier l'adéquation du modèle à risques non proportionnels alors que le coefficient R^2 est une mesure de la capacité prédictive du modèle. Bien que ces aspects de la modélisation soient différents (voir la Section 1.4), leur construction avec les mêmes quantités semble naturelle.

5.3 Adéquation et construction de modèles

5.3.1 Adéquation du modèle à risques proportionnels

Le processus du score standardisé U_n^* introduit au Chapitre 3 est un outil utile pour vérifier l'adéquation du modèle à risques proportionnels. Dans un premier temps, nous introduisons des bandes de confiance pour le processus, qui permettent d'évaluer si le paramètre du modèle est constant ou non. Dans un second temps, nous présentons un test d'adéquation du modèle.

Dans cette section, on se place sous le modèle à risques non proportionnels de coefficient de régression $\beta(t) = (\beta_1(t), \dots, \beta_p(t))$ et sous les conditions **H1'**, **H2'**, **H3'** et **H4'**. On suppose de plus que la suite $(M_n)_n$ de l'hypothèse **H3'** est telle que $\sqrt{n}M_n \rightarrow 0$ lorsque $n \rightarrow \infty$. Pour $i = 1, \dots, p$, on considère les hypothèses nulles et alternatives suivantes :

$$\begin{cases} H_{0,i} : \exists b_i, \forall t, \beta_i(t) = b_i \\ H_{1,i} : \nexists b_i, \forall t, \beta_i(t) = b_i. \end{cases}$$

Sous l'hypothèse $H_{0,i}$, la covariable $Z^{(i)}$ respecte l'hypothèse des risques proportionnels alors que, sous $H_{1,i}$, son coefficient change avec le temps. La proposition suivante permet de définir le comportement limite du i ème composant du processus du score standardisé sous $H_{0,i}$.

Proposition 5.3.1. *Soit $i \in \{1, \dots, p\}$ et $\beta_0 \in \mathbb{B}'$. Sous $H_{0,i}$, c'est-à-dire sous le modèle à risques proportionnels de paramètre constant b non nécessairement égal à β_0 , on a, pour $a \geq 0$,*

$$\lim_{n \rightarrow +\infty} P \left(\left\| \hat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2^{-1} \sup_{t \in [0,1]} \left| \left(\hat{\Sigma}^{-1/2} \{U_n^*(\beta_0, t) - tU_n^*(\beta_0, 1)\} \right)_i \right| \leq a \right) = P \left(\sup_{t \in [0,1]} |\mathcal{B}(t)| \leq a \right),$$

où \mathcal{B} est un pont Brownien et $\left\| \hat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2 = \left(\sum_{j=1}^p \left(\hat{\Sigma}_{j,i}^{-1/2} \right)^2 \right)^{1/2}$ est la norme L^2 de la i ème colonne de $\hat{\Sigma}^{-1/2}$.

Preuve de la Proposition 5.3.1. Montrons que

$$\sqrt{n} \sup_{t \in [0,1]} \left\| B_n(t) - t\Sigma^{1/2}(b - \beta_0) \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Nous avons

$$\begin{aligned} & \sqrt{n} \sup_{t \in [0,1]} \left\| B_n(t) - t\Sigma^{1/2}(b - \beta_0) \right\| \\ & \leq \sqrt{n} \sup_{t \in [0,1]} \left\| B_n(t) - \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \Sigma^{1/2}(b - \beta_0) \right\| + \sqrt{n} \sup_{t \in [0,1]} \left(t - \frac{\lfloor tk_n \rfloor}{k_n} \right) \left\| \Sigma^{1/2}(b - \beta_0) \right\| \\ & \leq \sqrt{n} \sup_{t \in [0,1]} \left\| B_n(t) - \frac{\lfloor tk_n \rfloor}{k_n} \Sigma^{1/2}(b - \beta_0) \right\| + \frac{\sqrt{n}}{k_n} p \left\| \Sigma^{1/2} \right\| \|b - \beta_0\|. \end{aligned}$$

Or, nous avons $\|\Sigma^{1/2}\| < \infty$, $\|b - \beta_0\| \leq \delta_2$ car β_0 est dans le voisinage \mathbb{B}' de $\beta(t) = b$ de rayon δ_2 et $\sqrt{n}/k_n \rightarrow 0$ quand $n \rightarrow \infty$. Il reste donc à montrer que

$$\sqrt{n} \sup_{t \in [0,1]} \left\| B_n(t) - \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \Sigma^{1/2}(b - \beta_0) \right\| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (5.22)$$

En utilisant l'inégalité de l'équation (3.57) avec $\beta(s) = b$:

$$\|\mathcal{E}_b(Z | s) - \mathcal{E}_{\beta_0}(Z | s) - \mathcal{V}_{\beta_0}(Z | s)(b - \beta_0)\| \leq \frac{1}{2} M_n \|b - \beta_0\|^2, \quad s \in \{t_1, \dots, t_{k_n}\},$$

et l'équation (3.51) :

$$\left\| V_{\beta_0} \left(\mathcal{Z}(t_i) \middle| \mathcal{F}_{t_i}^* \right)^{-1/2} \right\| \leq C_V^{-1} \sqrt{2} p^4 L', \quad i = 1, \dots, k_n,$$

nous avons

$$\begin{aligned} & \sqrt{n} \left\| B_n(t) - \frac{\lfloor tk_n \rfloor}{k_n} \Sigma^{1/2}(b - \beta_0) \right\| \\ & = \sqrt{n} \left\| \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \{ \mathcal{E}_b(Z | t_i) - \mathcal{E}_{\beta_0}(Z | t_i) \} - \frac{1}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \Sigma^{1/2}(b - \beta_0) \right\| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{\sqrt{n}}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \{ \mathcal{E}_b(Z | t_i) - \mathcal{E}_{\beta_0}(Z | t_i) - \mathcal{V}_{\beta_0}(Z | t_i)(b - \beta_0) \} \right\| \\
&\quad + \frac{\sqrt{n}}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \left(\mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right) (b - \beta_0) \right\| \\
&\leq p \frac{\sqrt{n}}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{-1/2} \right\| \frac{1}{2} M_n \|b - \beta_0\|^2 + p \|b - \beta_0\| \frac{\sqrt{n}}{k_n} \sum_{i=1}^{\lfloor tk_n \rfloor} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right\| \\
&\leq \frac{1}{\sqrt{2}} p^5 C_{\mathcal{V}}^{-1} L' \delta_2^2 \frac{\lfloor tk_n \rfloor}{k_n} \sqrt{n} M_n + p \frac{\lfloor tk_n \rfloor}{k_n} \delta_2 \sqrt{n} \sup_{i=1, \dots, k_n} \left\| \mathcal{V}_{\beta_0}(Z | t_i)^{1/2} - \Sigma^{1/2} \right\|.
\end{aligned}$$

La condition **H3'** et la convergence $\sqrt{n}M_n \rightarrow 0$ lorsque $n \rightarrow \infty$ terminent la conclusion de l'équation (5.22). Le Corollaire 3.4.1 indique la convergence au sens de la topologie produit de Skorokhod

$$\widehat{\Sigma}^{-1/2} U_n^*(\beta_0, \cdot) - \sqrt{k_n} \widehat{\Sigma}^{-1/2} B_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Sigma^{-1/2} \mathcal{W}_p.$$

Par le lemme de Slutsky, nous avons alors

$$\widehat{\Sigma}^{-1/2} U_n^*(\beta_0, \cdot) - \sqrt{k_n} \widehat{\Sigma}^{-1/2} B_{\beta_0, b} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Sigma^{-1/2} \mathcal{W}_p,$$

où $B_{\beta_0, b}(t) = t(b - \beta_0)$ pour $t \in [0, 1]$. Ainsi, en posant $\widetilde{U}_n^*(\beta_0, t) = U_n^*(\beta_0, t) - tU_n^*(\beta_0, 1)$ pour $t \in [0, 1]$, on a

$$\widehat{\Sigma}^{-1/2} \widetilde{U}_n^*(\beta_0, \cdot) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \Sigma^{-1/2} \mathcal{B}_p,$$

où $\mathcal{B}_p = (\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(p)})$ est un vecteur de p ponts Browniens $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(p)}$ indépendants. Or, la combinaison linéaire

$$\frac{1}{\sqrt{\sum_{i=1}^p c_i^2}} \sum_{i=1}^p c_i \mathcal{B}^{(i)}$$

est un pont Brownien unidimensionnel \mathcal{B} pour $c_1, \dots, c_p \in \mathbb{R}$. Par conséquent, en standardisant chaque composante $\left(\widehat{\Sigma}^{-1/2} \widetilde{U}_n^*(\beta_0, \cdot) \right)_i$ par $\left\| \widehat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2^{-1}$ pour $i = 1, \dots, p$, on obtient la convergence

$$\left\| \widehat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2^{-1} \left(\widehat{\Sigma}^{-1/2} \widetilde{U}_n^*(\beta_0, \cdot) \right)_i \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{B}, \quad i = 1, \dots, p,$$

et la convergence a lieu uniformément sur les temps car le processus limite \mathcal{B} est continu. \square

Ce résultat permet de construire des bandes de confiance de la constante d'un paramètre autour de la composante correspondante du processus, ainsi qu'un test d'adéquation du modèle à risques proportionnels. Rappelons que la loi limite du supremum de la valeur absolue d'un pont Brownien est la loi de Kolmogorov. Pour $\alpha = 5\%$, le quantile supérieur d'ordre α de la loi de Kolmogorov vaut $a(\alpha) = 1.358$.

Corollaire 5.3.1 (Bande de confiance). *Soit $i \in \{1, \dots, p\}$ et $\beta_0 \in \mathbb{B}^l$. Sous $H_{0,i}$, c'est-à-dire sous le modèle à risques proportionnels de paramètre constant b non nécessairement égal à β_0 ,*

en notant $a(\alpha)$ le quantile supérieur d'ordre α de la loi de Kolmogorov, on a

$$\lim_{n \rightarrow +\infty} P\left(\forall t \in [0, 1], \left[\hat{\Sigma}^{-1/2} U_n^*(\beta_0, t)\right]_i \in BC_i(\alpha)\right) = 1 - \alpha,$$

où $BC_i(\alpha)$ est une bande de confiance de niveau $1 - \alpha$ telle que

$$BC_i(\alpha) = \left[t \left[\hat{\Sigma}^{-1/2} U_n^*(\beta_0, 1) \right]_i - \left\| \hat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2 a(\alpha); t \left[\hat{\Sigma}^{-1/2} U_n^*(\beta_0, 1) \right]_i + \left\| \hat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2 a(\alpha) \right].$$

Corollaire 5.3.2 (Test d'adéquation). *Soit $i \in \{1, \dots, p\}$ et $\beta_0 \in \mathbb{B}'$. Sous le modèle à risques non proportionnels de paramètre $\beta(t)$ non nécessairement égal à β_0 , l'hypothèse $H_{0,i}$ est rejetée au niveau asymptotique α si*

$$\left\| \hat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2^{-1} \sup_{t \in [0, 1]} \left| \left(\hat{\Sigma}^{-1/2} \{U_n^*(\beta_0, t) - tU_n^*(\beta_0, 1)\} \right)_i \right| \geq a(\alpha),$$

où $a(\alpha)$ est le quantile supérieur d'ordre α de la loi de Kolmogorov. En notant K une variable aléatoire de distribution de Kolmogorov, la p -value asymptotique du test est

$$P\left(K \geq \left\| \hat{\Sigma}_{\cdot, i}^{-1/2} \right\|_2^{-1} \sup_{t \in [0, 1]} \left| \left(\hat{\Sigma}^{-1/2} \{U_n^*(\beta_0, t) - tU_n^*(\beta_0, 1)\} \right)_i \right| \right).$$

Si le i ème composant du processus $\hat{\Sigma}^{-1/2} U_n^*(\beta_0, \cdot)$ sort de la bande de confiance $IC(\alpha)_i$ ou si la p -value est inférieure à α , l'hypothèse $H_{0,i}$ que l'effet $\beta_i(t)$ soit constant est rejetée au niveau α . Cependant, en testant simultanément plusieurs hypothèses $H_{0,i}$ pour différentes covariables, on observe une inflation du niveau global du test. Cela signifie qu'un processus pourrait sortir de la bande de confiance alors que l'effet correspondant est constant au cours du temps, avec un niveau global supérieur à α . Ceci pourrait être un problème si l'on basait une conclusion définitive sur l'adéquation du modèle à risques proportionnels sur ces bandes de confiance ou ce test d'adéquation. Ce n'est pas le cas, les bandes de confiance ne sont qu'un outil dans la construction du modèle. La non-détection d'un effet constant sera corrigée par la suite avec les autres étapes de la méthode de sélection de variables respectant l'hypothèse de proportionnalité des risques que nous proposons dans la section suivante. Notons que l'hypothèse de proportionnalité des risques est évaluée pour chaque covariable en tenant compte de la corrélation entre les différentes covariables et en laissant la possibilité aux autres effets de dépendre du temps. Des simulations comparant les performances de ce test d'adéquation du modèle à risques proportionnels avec d'autres tests de référence seront présentées en Section 5.4.1.

5.3.2 Construction du modèle à risques non proportionnels

Avec les résultats du Théorème 3.4.1 et de son corollaire, le processus du score standardisé U_n^* peut être utilisé pour déterminer la forme du coefficient de régression multivarié $\beta(t)$. La dérive du processus reflète la forme de $\beta(t)$. Aucun autre intermédiaire tel que le lissage, la projection

sur une base de fonctions ou l'estimation à noyau n'est nécessaire (Cai et Sun, 2003, Hastie et Tibshirani, 1990, Scheike et Martinussen, 2004). Par exemple, comme l'illustre la Figure 3.2(a), un effet constant jusqu'à un temps τ suivi d'un effet nul est facilement détectable, même pour des tailles d'échantillons modérées.

Supposons que les p composantes du paramètre de régression dépendant du temps

$$\beta(t) = (\beta_1(t), \dots, \beta_p(t))$$

peuvent être écrites $\beta_j(t) = \beta_{0,j} h_j(t)$ pour $j = 1, \dots, p$, avec

- $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,p}) \in \mathbb{R}^p$ un paramètre de régression fixe dans le temps et inconnu,
- $h = (h_1, \dots, h_p)$ une fonction de $[0, 1]$ à valeurs dans \mathbb{R}^p connue.

Un estimateur convergent de $\beta(t)$ est alors $\hat{\beta}(t) = (\hat{\beta}_1(t), \dots, \hat{\beta}_p(t))$ avec $\hat{\beta}_j(t) = \hat{\beta}_{0,j} h_j(t)$ ($j = 1, \dots, p$) où $\hat{\beta}_0 = (\hat{\beta}_{0,1}, \dots, \hat{\beta}_{0,p})$ est l'estimateur du maximum de la vraisemblance partielle de β_0 . Cette estimation peut s'obtenir en transférant la partie temporelle h du coefficient de régression dans la covariable (Cox, 1972). La fonction h peut être déterminée graphiquement en utilisant le processus du score standardisé (voir les exemples de la Section 5.4). Par exemple, si la j ème composante de la dérive est une fonction linéaire, la fonction h_j est constante. Si cette dérive est concave, la fonction h_j est décroissante et une dérive convexe correspond à une fonction h_j croissante. De plus, les bandes de confiance définies en Section 5.3.1 peuvent aider à évaluer la plausibilité d'un effet constant au cours du temps pour $\beta_j(t)$, résultant en une fonction h_j constante, pour $j = 1, \dots, p$.

En présence de risques non proportionnels, l'investigateur a besoin d'un second instrument qui ne soit pas basé uniquement sur la qualité de l'adéquation du modèle. Nous proposons d'utiliser le coefficient R^2 qui, en plus d'indiquer la capacité prédictive du modèle, converge vers une fonction dont le maximum est atteint lorsque la vraie fonction h est choisie (Théorème 5.2.1). Quand plusieurs modèles fournissent des formes plausibles pour h , les estimateurs $\hat{\beta}(t)$ et les coefficients R^2 correspondants sont évalués. La fonction temporelle h retenue sera celle qui maximise le coefficient R^2 . Avec cette procédure, nous obtenons un modèle à risques non proportionnels avec une bonne adéquation et une capacité prédictive maximale. La mesure de capacité prédictive est maximisée sur l'ensemble \mathbf{B} des m effets de régression temporels sélectionnés par l'investigateur. Plus formellement, notons $\mathbf{B} = \{\beta_1(t), \dots, \beta_m(t)\}$ l'ensemble des m fonctions de $[0, 1]$ dans \mathbb{R}^p . Le coefficient de régression sélectionné $\beta^*(t)$ est tel que

$$\beta^*(t) = \arg \max_{\alpha(t) \in \mathbf{B}} R^2(\alpha(t)).$$

Le théorème suivant donne une équivalence entre ce problème de maximisation lorsque $n \rightarrow \infty$ et un problème de minimisation de norme L^2 .

Théorème 5.3.1. *Soit $p = 1$. On suppose les conditions **H1.**, **H2.**, **H3.** et **H4.** vérifiées, $\beta(t) \in \mathbb{B}$ et $\mathbf{B} \subset \mathbb{B}$. Sous le modèle à risques non proportionnels de paramètre de régression*

$\beta(t)$, avec $\beta(t)$ non nécessairement dans \mathbf{B} , le coefficient

$$\beta^*(t) = \arg \max_{\alpha(t) \in \mathbf{B}} \lim_{n \rightarrow \infty} R^2(\alpha(t))$$

est solution de

$$\beta^*(t) = \arg \min_{\alpha(t) \in \mathbf{B}} \|\beta(t) - \alpha(t)\|_2,$$

où pour $\alpha(t) \in \mathbb{B}$, $\|\beta(t) - \alpha(t)\|_2^2 = \int_0^1 (\beta(t) - \alpha(t))^2 dt$.

Preuve du Théorème 5.3.1. Soit $t \in [0, 1]$. Sous les conditions du théorème, un développement de Taylor de $e(\alpha(t), t)$ donne

$$e(\alpha(t), t) = e(\beta(t), t) - (\alpha(t) - \beta(t))v(\gamma(t), t),$$

où γ se situe dans la boule de centre α et de rayon $\sup_{t \in [0, 1]} |\alpha(t) - \beta(t)|$. Sous **H3.**, il existe une constante $C(\gamma)$ telle que $v(\gamma(t), t) = C(\gamma)$.

En utilisant ce développement dans l'équation (5.13), on a

$$\lim_{n \rightarrow \infty} R^2(\alpha(t)) = 1 - \frac{C(\beta) + C(\gamma)^2 \int_0^1 (\alpha(t) - \beta(t))^2 dt}{C(\beta) + \int_0^1 (e(0, t) - e(\beta(t), t))^2 dt}.$$

Par conséquent, maximiser la fonction $\alpha \mapsto \lim_{n \rightarrow \infty} R^2(\alpha)$ revient à minimiser la fonction $\alpha \mapsto \int_0^1 (\alpha(t) - \beta(t))^2 dt$.

□

En d'autres termes, pour un échantillon de taille assez grande, choisir le coefficient de régression qui maximise le coefficient R^2 revient à choisir le coefficient le plus proche de la vraie fonction de régression au sens de la distance L^2 . Il est possible que le vrai coefficient de régression, s'il existe, ne soit pas inclus dans l'ensemble \mathbf{B} . Un modèle est choisi pour représenter des données en raison soit d'une bonne adéquation, soit d'une bonne capacité prédictive. Plusieurs modèles peuvent présenter un de ces aspects ou les deux, pas seulement le "vrai" modèle. Ici, la priorité est donnée à l'adéquation, avec la sélection des coefficients de régression candidats, et, dans un second temps, la capacité prédictive du modèle est considérée. Nous avons choisi de travailler avec le coefficient R^2 , mais toute mesure de capacité prédictive vérifiant le Théorème 5.2.1 pourrait être considérée.

Lorsque la tendance de la dérive du processus est une fonction concave, l'effet diminue au cours du temps alors qu'en présence d'une fonction convexe, l'effet augmente. Afin d'obtenir le R^2 le plus grand possible, il serait possible de créer un effet temporel de plus en plus proche de la dérive observée du processus, par exemple un effet constant par morceaux avec de nombreuses ruptures. En général, cela résulterait en un modèle surajusté. Dans ce cas, l'interprétation du coefficient ne serait pas claire. Un compromis doit être établi entre une forte capacité prédictive et la simplicité du coefficient, notamment en terme d'interprétation. Cette propriété est comparable à la situation du modèle linéaire dans laquelle la variance expliquée estimée, le

coefficient R^2 , est biaisée positivement et le biais augmente avec la dimension du modèle. Un équilibre doit être établi entre le but d'améliorer la qualité de la prédiction et les dangers de mauvaises prédictions dues à un surajustement du modèle.

5.4 Simulations

Deux types de simulations sont présentées. Les premières simulations portent sur l'adéquation du modèle à risques proportionnels, alors que les secondes sont des illustrations de la méthode de construction de modèles à risques non proportionnels à l'aide du processus du score standardisé et du coefficient R^2 . Les échantillons ont été générés sous le modèle à risques non proportionnels selon la méthode détaillée en annexe.

5.4.1 Test d'adéquation

Le comportement du test d'adéquation du modèle à risques proportionnels présenté en Section 5.3.1 est étudié par un travail de simulation. Nous l'avons comparé à des tests de référence qui présentent également l'avantage d'être basés sur des méthodes graphiques permettant de visualiser la forme de l'effet dépendant du temps en cas de mauvaise adéquation du modèle.

Description des tests comparés

Le premier test implémenté a été proposé par [Lin et al. \(1993\)](#). Il est basé sur le processus du score défini à l'équation (5.2). Ce processus diffère du processus du score standardisé défini et étudié dans cette thèse selon deux aspects : les incréments du processus de [Lin et al. \(1993\)](#) sont les résidus de Schoenfeld non standardisés sommés selon les temps de décès de l'échelle de temps initiale et une standardisation globale est appliquée sur tous les incréments. La statistique du test est le supremum sur le temps de la valeur absolue du processus standardisé globalement. Si les covariables sont non corrélées, la loi limite de la statistique de test est la distribution de Kolmogorov ([Therneau et al., 1990](#)). Si les covariables sont corrélées, cette loi limite n'est pas connue, mais elle peut être évaluée numériquement par des simulations de Monte Carlo. Ainsi, l'enveloppe de confiance du processus et le test d'adéquation sont obtenus par simulation de N processus selon la loi limite gaussienne du processus. Nous avons choisi $N = 10^3$. Les codes R pour l'évaluation du processus et de l'enveloppe simulée et le test d'adéquation sont disponibles dans le package `timereg` ([Martinussen et Scheike, 2006](#)). Dans la suite, ce test est abrégé par le sigle LWY.

Le second test d'adéquation du modèle à risques proportionnels a été proposé par [Grambsch et Therneau \(1994\)](#). Ils considèrent un effet de la forme $\beta(t) = \beta_0 + \beta_1 \log(t)$ et proposent un test du score de l'hypothèse nulle $H_0 : \beta_1 = 0$ contre son alternative $H_1 : \beta_1 \neq 0$. Le test est implémenté dans le langage R dans le package `survival`. Nous référons à ce test par le sigle GT.

Enfin, le dernier test d'adéquation étudié a été proposé par [Scheike et Martinussen \(2004\)](#). En se plaçant sous le modèle à risques non proportionnels, ils proposent un test de type

Cramér-von Mises, basé sur l'estimateur du paramètre $\beta(t)$ cumulé. L'estimation est faite par un algorithme faisant appel à un lissage par noyau. Le code du test est programmé dans le package `timereg` de R. Ce test est abrégé par le sigle SM. Les auteurs ont également proposé un test de Kolmogorov-Smirnov, mais qui semblait moins puissant au vu de leurs simulations, c'est pourquoi nous ne l'avons pas retenu.

Le test basé sur le processus du score standardisé introduit en Section 5.3.1 est noté U_n^* dans les tableaux de résultats.

Scénarii de simulation

Les données sont simulées sous le modèle à risques proportionnels avec deux covariables $Z^{(1)}$ et $Z^{(2)}$ normales centrées de variance 2 et de corrélation $\rho(Z^{(1)}, Z^{(2)})$ valant 0, 0.3, 0.5 ou 0.7. La taille de l'échantillon vaut $n = 50$, $n = 100$ ou $n = 200$. Le taux de censure moyen est de 25% et la censure est administrative : les données sont censurées au temps 1.5 sur l'échelle initiale. Le modèle à risques non proportionnels est

$$\lambda(t | Z^{(1)}, Z^{(2)}) = \exp\left(0.5Z^{(1)} + \beta_2(t)Z^{(2)}\right),$$

où $\beta_2(t)$ prend les valeurs 0, 0.5, $\mathbf{1}_{t \leq 0.5}$, $1.5\mathbf{1}_{t \leq 0.5}$ ou $0.5\mathbf{1}_{t \leq 0.5} - 0.5\mathbf{1}_{t \geq 0.5}$ sur l'échelle du temps transformée. La covariable $Z^{(1)}$ respecte l'hypothèse de proportionnalité des risques. Dans les deux premiers cas, $Z^{(2)}$ respecte également l'hypothèse de proportionnalité des risques. Dans les deux cas suivants, cette hypothèse n'est pas vérifiée puisque l'effet est constant au départ et s'annule lorsque la moitié des décès a été observée. Dans le dernier cas, l'hypothèse n'est pas vérifiée et la nature de l'effet change à la moitié de l'étude (l'effet de la variable sur la survie est positif au début et négatif à la fin). Cette dernière situation correspond à des risques qui se croisent. Pour chaque configuration, 3000 échantillons ont été simulés afin d'évaluer le niveau ou la puissance empiriques des tests.

Résultats

Les Tables 5.1 et 5.2 présentent les résultats pour les deux modèles à risques proportionnels. On remarque que le niveau est bien estimé à 5% pour chaque configuration et chaque test, à l'exception du test de Scheike et Martinussen qui a un niveau supérieur à 5% pour les petits échantillons ($n = 50$ et $n = 100$).

Dans les situations de risques non proportionnels (Tables 5.3 à 5.5), l'augmentation de la corrélation entre les covariables se traduit par une augmentation du niveau estimé du test de Lin, Wei et Ying. Ce résultat avait déjà été souligné par [Scheike et Martinussen \(2004\)](#). Ceci rend l'utilisation du test en présence de covariables corrélées inadéquate, ce qui est pourtant très souvent le cas en pratique.

Les autres tests étudiés ne présentent pas ce phénomène. Dans leur simulation, [Scheike et Martinussen \(2004\)](#) retrouvent pourtant cette inflation du niveau du test de Grambsch et Therneau en présence de corrélation entre les covariables. Ceci est sûrement dû au fait que les

n	$\rho(Z^{(1)}, Z^{(2)})$	LWY		GT		SM		U_n^*	
		$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$
50	0.0	6.3	5.3	5.5	4.4	7.4	7.0	3.3	2.9
50	0.3	5.4	4.8	5.2	4.6	8.1	7.0	2.8	3.1
50	0.5	5.0	4.7	5.2	4.5	6.8	6.7	3.0	2.6
50	0.7	5.7	6.5	5.0	5.1	7.2	8.8	3.1	3.2
100	0.0	5.4	5.7	4.8	5.0	5.7	6.5	3.4	3.7
100	0.3	5.5	5.2	4.9	4.7	5.9	5.9	3.4	3.6
100	0.5	5.7	5.9	4.5	4.8	5.4	6.6	3.4	3.5
100	0.7	5.4	5.4	5.7	6.1	6.5	6.7	3.9	3.7
200	0.0	4.6	5.6	4.3	5.1	5.0	5.7	3.9	3.9
200	0.3	5.5	5.7	4.9	4.8	5.2	5.4	4.5	4.2
200	0.5	6.2	5.7	4.4	4.9	4.9	5.1	3.8	3.9
200	0.7	5.5	5.2	5.7	5.0	5.3	5.0	4.6	3.8
400	0.0	5.4	5.7	4.7	4.7	4.6	4.7	4.5	4.6
400	0.3	5.6	4.5	5.1	4.4	5.0	5.2	4.4	3.6
400	0.5	4.8	5.6	4.8	5.2	5.3	5.2	4.4	5.2
400	0.7	6.0	5.2	5.1	5.3	5.7	5.9	5.1	5.0

TABLE 5.1 – Niveaux empiriques des différents tests (en %) pour $\beta_2(t) = 0$.

effets non constants $\beta_2(t)$ que nous considérons sont assez réguliers. Le test de Grambsch et Therneau suppose la constance du coefficient $\beta_2(t)$ lorsque le test est effectué pour $\beta_1(t)$. Ce test doit donc être utilisé avec précaution. Notons que la puissance du test de Grambsch et Therneau est équivalente ou inférieure à celle du test d'adéquation basé sur le processus du score.

Enfin, le test de Scheike et Martinussen est moins puissant que le test d'adéquation basé sur le processus du score standardisé pour les situations que nous avons considérées. De plus, pour les plus faibles tailles d'échantillon ($n = 50$), le test de Scheike et Martinussen ne respecte pas le niveau de 5%.

En conclusion, les simulations montrent que le test d'adéquation basé sur le processus du score est le plus puissant des tests étudiés pour les situations que nous avons considérées. Ce test n'est pas influencé par la corrélation entre les covariables et son niveau est bien calibré, même pour les petits échantillons.

n	$\rho(Z^{(1)}, Z^{(2)})$	LWY		GT		SM		U_n^*	
		$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$
50	0.0	5.3	5.9	4.8	4.9	7.3	6.7	3.0	2.8
50	0.3	6.1	5.4	5.0	4.5	7.8	7.0	3.2	3.0
50	0.5	5.1	5.1	5.0	4.5	7.0	7.2	2.9	2.6
50	0.7	5.4	5.5	5.1	5.0	8.1	7.7	3.1	3.0
100	0.0	5.9	5.3	4.3	4.3	5.6	5.3	3.6	3.3
100	0.3	5.0	5.7	4.9	4.3	5.3	5.0	3.5	3.8
100	0.5	5.4	5.5	4.8	4.9	6.1	6.7	3.5	3.4
100	0.7	5.4	5.8	5.6	5.4	5.4	6.0	3.6	3.6
200	0.0	5.2	5.5	3.9	4.6	5.3	5.7	3.5	3.4
200	0.3	6.0	5.6	4.8	4.1	5.7	4.9	4.1	3.6
200	0.5	5.7	5.8	4.9	4.5	5.3	5.0	4.0	4.1
200	0.7	4.6	4.9	5.1	4.4	5.2	4.7	3.4	4.2
400	0.0	4.7	5.6	4.7	5.6	4.6	5.4	3.9	4.6
400	0.3	4.2	5.0	4.5	4.4	5.2	4.6	4.2	4.5
400	0.5	6.6	5.3	5.8	4.9	5.1	5.4	4.6	5.0
400	0.7	5.7	6.3	5.2	4.9	5.1	5.3	4.9	5.1

TABLE 5.2 – Niveaux empiriques des différents tests (en %) pour $\beta_2(t) = 0.5$.

n	$\rho(Z^{(1)}, Z^{(2)})$	LWY		GT		SM		U_n^*	
		$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$
50	0.0	5.3	14.9	4.4	13.6	7.2	7.8	2.6	11.3
50	0.3	6.3	14.7	5.0	12.8	7.4	7.6	3.3	9.6
50	0.5	6.1	15.1	4.6	11.9	6.8	7.6	2.7	8.3
50	0.7	9.3	16.8	5.4	11.0	6.3	7.4	2.8	7.9
100	0.0	4.6	28.3	4.8	21.2	5.2	9.0	2.4	22.4
100	0.3	6.4	27.9	5.2	20.7	5.8	8.3	4.0	21.7
100	0.5	9.4	30.6	5.0	20.3	5.5	9.1	4.1	19.8
100	0.7	14.6	30.0	5.5	15.1	5.8	7.2	3.4	13.5
200	0.0	5.5	55.5	4.6	37.9	5.4	14.7	3.7	50.4
200	0.3	8.2	57.5	4.7	38.4	4.9	13.2	4.1	50.7
200	0.5	13.9	59.2	5.0	34.0	4.4	12.8	4.6	42.7
200	0.7	29.3	59.9	4.9	23.8	5.3	9.2	3.8	28.3
400	0.0	5.3	89.6	4.7	60.8	4.4	23.4	4.3	87.6
400	0.3	11.5	90.4	5.1	62.5	5.4	25.9	4.2	86.0
400	0.5	27.7	91.2	5.6	56.0	5.1	21.5	4.6	79.6
400	0.7	56.3	92.5	5.5	46.2	4.9	16.8	4.7	63.3

TABLE 5.3 – Niveaux (colonne $\beta_1(t)$) et puissances (colonne $\beta_2(t)$) empiriques des différents tests (en %) pour $\beta_2(t) = \mathbf{1}_{t \leq 0.5}$.

n	$\rho(Z^{(1)}, Z^{(2)})$	LWY		GT		SM		U_n^*	
		$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$
50	0.0	7.0	51.6	5.6	57.4	8.9	53.4	2.9	56.3
50	0.3	5.8	44.9	5.4	53.1	7.4	47.5	3.1	49.5
50	0.5	10.0	41.9	5.9	49.2	8.0	43.1	3.4	41.9
50	0.7	14.6	37.3	5.5	37.2	7.9	31.4	2.3	27.4
100	0.0	7.5	82.2	4.9	85.4	6.0	81.9	3.4	88.0
100	0.3	7.0	80.9	5.5	83.9	6.2	80.0	4.3	86.6
100	0.5	13.9	76.5	5.6	78.6	6.4	72.4	3.8	77.9
100	0.7	26.1	70.4	5.7	63.6	5.4	56.5	2.8	58.8
200	0.0	8.7	99.0	5.3	98.9	5.1	98.6	4.6	99.7
200	0.3	8.4	98.5	6.0	98.9	5.4	97.9	4.4	99.5
200	0.5	22.3	97.5	5.9	97.4	5.1	95.8	4.5	98.6
200	0.7	53.4	96.6	5.7	92.4	5.1	88.3	3.6	92.9
400	0.0	13.7	100.0	5.0	100.0	4.9	100.0	4.6	100.0
400	0.3	11.4	100.0	5.4	100.0	4.7	100.0	4.3	100.0
400	0.5	43.5	100.0	6.1	100.0	4.6	99.9	5.0	100.0
400	0.7	84.5	100.0	5.7	99.6	4.9	99.4	5.0	100.0

TABLE 5.4 – Niveaux (colonne $\beta_1(t)$) et puissances (colonne $\beta_2(t)$) empiriques des différents tests (en %) pour $\beta_2(t) = 1.5\mathbf{1}_{t \leq 0.5}$.

n	$\rho(Z^{(1)}, Z^{(2)})$	LWY		GT		SM		U_n^*	
		$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$	$\beta_1(t)$	$\beta_2(t)$
50	0.0	5.4	14.4	4.6	12.6	6.8	7.8	2.8	9.8
50	0.3	5.8	14.9	4.9	13.0	7.0	8.1	2.7	9.8
50	0.5	8.0	15.2	5.4	12.6	7.1	8.4	3.2	9.4
50	0.7	8.4	15.2	5.3	9.8	6.6	6.7	3.0	6.7
100	0.0	6.2	26.8	5.4	21.6	5.9	9.1	3.4	21.3
100	0.3	6.8	29.0	5.0	21.3	5.5	8.7	3.6	21.6
100	0.5	9.8	29.6	5.2	20.1	5.8	8.2	4.3	19.9
100	0.7	15.2	31.1	5.3	16.0	5.4	7.5	3.1	14.6
200	0.0	6.0	52.8	5.2	35.8	5.5	12.4	4.2	47.7
200	0.3	8.3	56.7	5.3	37.1	5.1	13.0	4.3	48.6
200	0.5	14.9	59.7	4.7	34.0	5.1	11.6	3.6	43.0
200	0.7	29.9	61.8	6.2	26.0	5.2	10.2	4.3	29.7
400	0.0	5.4	87.3	4.5	62.4	5.6	25.0	4.3	86.2
400	0.3	11.1	90.0	5.3	61.6	5.3	22.4	4.3	85.2
400	0.5	27.0	91.7	4.5	55.4	4.1	18.7	4.1	79.6
400	0.7	56.6	93.2	4.9	46.6	4.6	15.4	4.0	64.5

TABLE 5.5 – Niveaux (colonne $\beta_1(t)$) et puissances (colonne $\beta_2(t)$) empiriques des différents tests (en %) pour $\beta_2(t) = 0.5\mathbf{1}_{t \leq 0.5} - 0.5\mathbf{1}_{t \geq 0.5}$.

5.4.2 Construction du modèle à risques non proportionnels

Dans cette section, nous illustrons la méthode de construction du modèle à risques non proportionnels à partir du processus du score standardisé U_n^* et du coefficient R^2 . Dans un premier temps, nous illustrons la méthode avec une seule covariable, avant de considérer des cas multivariés. Nous traçons la bande de confiance de chaque processus comme décrit en Section 5.3.1, afin d'avoir une indication sur la validité de l'hypothèse de proportionnalité des risques. Le package R nommé `PHeval`¹ que nous avons développé permet d'évaluer et de tracer le processus du score standardisé ainsi que de calculer le coefficient R^2 .

Cas univariés

Les échantillons simulés sont de taille $n = 200$ avec une covariable Z distribuée selon une loi de Bernoulli de paramètre 0.5. Les temps de censure sont simulés indépendamment des temps d'évènement selon une loi uniforme sur $[0, 3]$.

Risques proportionnels. On simule d'abord un échantillon sous le modèle à risques proportionnels de paramètre 1 :

$$\lambda(t | Z) = \exp(Z). \quad (5.23)$$

Le taux de censure dans l'échantillon est de 18%. Le processus du score standardisé $U_n^*(0, \cdot)$ et

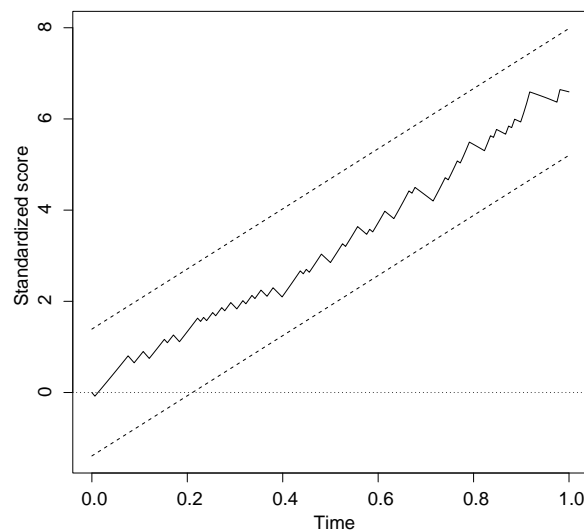


FIGURE 5.2 – Processus du score standardisé $U_n^*(0, \cdot)$ (ligne pleine) et sa bande de confiance (lignes pointillées) pour le jeu de données simulé avec $\beta(t) = 1$.

sa bande de confiance sous l'hypothèse des risques proportionnels sont tracés en fonction du temps en Figure 5.2. On peut observer une dérive, ce qui implique que l'effet est non nul. La dérive paraît être linéaire et le processus ne sort pas des bandes de confiance à 95% : l'hypothèse de proportionnalité des risques semble raisonnable. La dérive du processus est croissante, ce

1. Disponible à l'adresse : <http://cran.r-project.org/web/packages/PHeval/index.html>

qui signifie que le coefficient est positif. L'estimateur du maximum de vraisemblance partielle du modèle à risques proportionnels est estimé à 1.08 et le coefficient R^2 vaut 0.21.

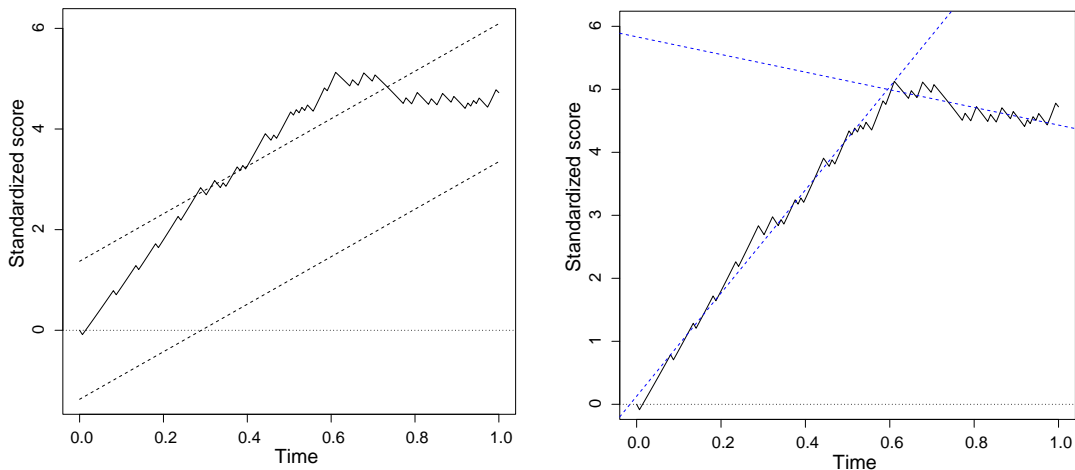
Modèle de rupture. On simule un jeu de données sous le modèle de rupture

$$\lambda(t | Z) = \exp(\beta(t)Z),$$

avec $\beta(t) = \mathbf{1}_{t \leq 0.4}$ sur l'échelle de temps initiale. Le coefficient vaut 1 pour la première partie de l'étude et s'annule dans un second temps. L'échantillon simulé comporte 26% de censure. La figure 5.3(a) représente le processus du score standardisé $U_n^*(0, \cdot)$ en fonction du temps transformé entre 0 et 1. Une dérive est observée, indiquant la présence d'un effet non nul. La dérive ne semble pas linéaire et le processus quitte la bande de confiance à 95%, ce qui permet de conclure à la présence d'un effet qui change dans le temps. La dérive paraît linéaire jusqu'au temps 0.6 sur l'échelle des temps transformés et l'effet semble faible par la suite. Plusieurs coefficients de régression sont envisagés. Le premier est le coefficient avec un effet constant jusqu'au temps 0.6 sur l'échelle transformée puis nul par la suite, noté $\beta_{0.6a}(t) = \beta_0 \mathbf{1}_{t \leq 0.6}$. Dans ce cas, la dérive observée après le temps 0.6 correspond à du bruit. L'effet après le temps 0.6 peut également être non nul, c'est pourquoi nous considérons l'effet

$$\beta_{0.6b}(t) = \beta_0 \mathbf{1}_{\{ \mathbf{1}_{t \leq 0.6} + C_{0.6} \mathbf{1}_{t > 0.6} \}},$$

avec β_0 et $C_{0.6}$ inconnus. La valeur $C_{0.6}$ est la valeur qui multiplie le coefficient de régression dans la deuxième partie de l'étude. Dans la Figure 5.3(b), deux droites ont été ajustées au processus par régression linéaire, avant et après $t = 0.6$. Le rapport de la deuxième pente sur la première donne $C_{0.6} = -0.17$.



(a) Processus $U_n^*(0, \cdot)$ (ligne pleine) et sa bande de confiance (lignes pointillées) (b) $U_n^*(0, \cdot)$ et ajustement d'un coefficient constant par morceaux avec une rupture à $t = 0.6$

FIGURE 5.3 – Processus du score standardisé $U_n^*(0, \cdot)$ pour le jeu de données simulé selon un modèle de rupture.

Nous considérons également d'autres temps de rupture à $t = 0.4$, $t = 0.5$ et $t = 0.7$, et plus particulièrement les coefficients $\beta_{0.4a}(t) = \beta_0 \mathbf{1}_{t \leq 0.4}$, $\beta_{0.5a}(t) = \beta_0 \mathbf{1}_{t \leq 0.5}$ et $\beta_{0.7a}(t) = \beta_0 \mathbf{1}_{t \leq 0.7}$ qui s'annulent dans la seconde partie de l'étude. Les coefficients constants par morceaux ne s'annulant pas en deuxième partie d'étude, $\beta_{0.4b}$, $\beta_{0.5b}$ et $\beta_{0.7b}$ ont été considérés et tracés sur le processus en Figure 5.4. L'ajustement a été effectué par régression linéaire.

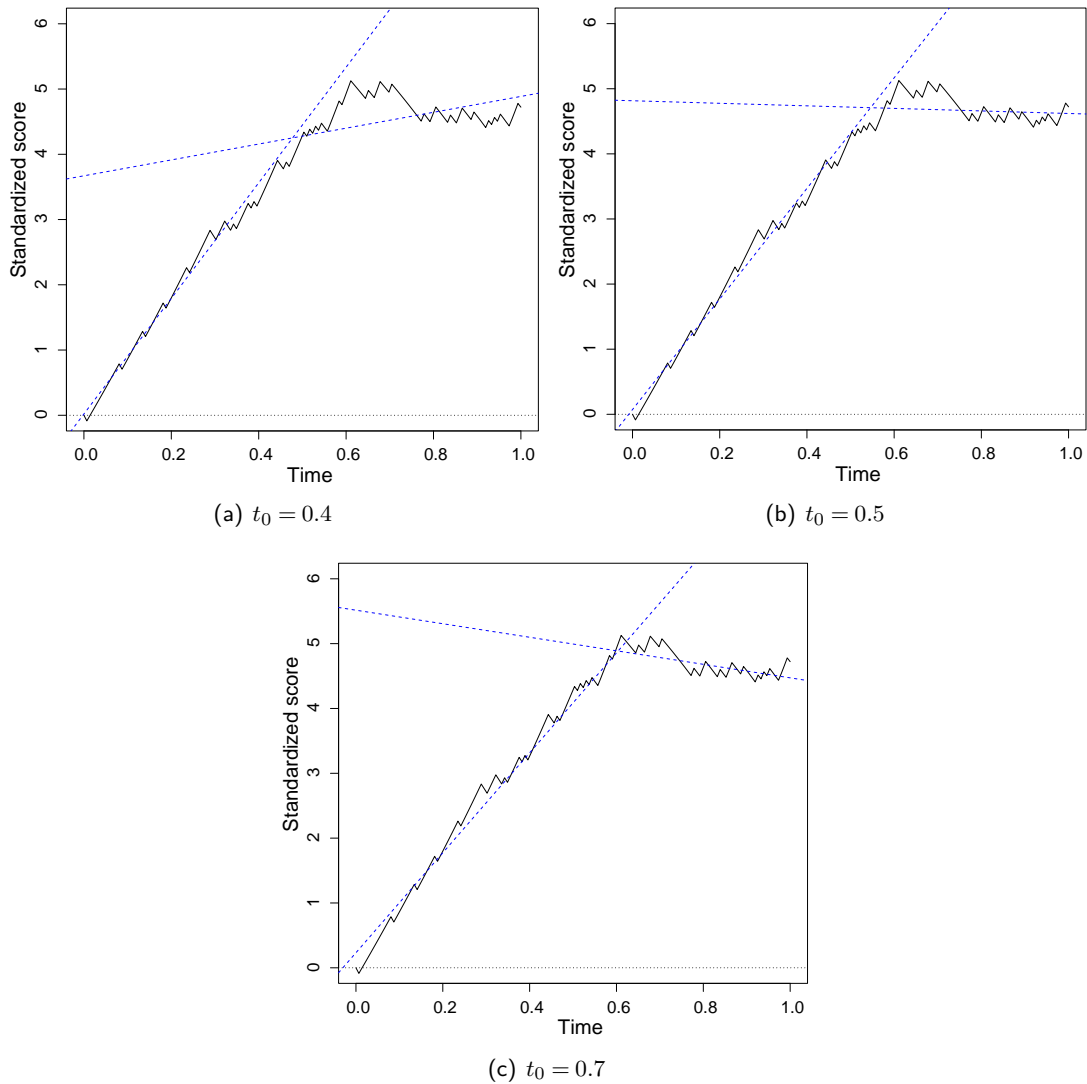


FIGURE 5.4 – Processus du score standardisé $U_n^*(0, \cdot)$ (ligne solide) et ajustement d'un coefficient constant par morceaux (lignes pointillées) avec une rupture au temps t_0 pour le jeu de données simulé selon un modèle de rupture.

La forme du processus du score standardisé indique que l'effet $\beta(t)$ diminue avec le temps. Plusieurs modèles avec des effets décroissants continus $\beta(t) = \beta_0 h(t)$ ont été sélectionnés, en plus des coefficients déjà mentionnés. Pour chaque modèle, l'estimateur du maximum de la vraisemblance partielle $\hat{\beta}_0$ et le coefficient R^2 ont été évalués. Les résultats sont donnés en Table 5.6. Les deux coefficients $\beta_{0.6a}(t)$ et $\beta_{0.6b}(t)$ donnent la plus grande valeur de R^2 égale

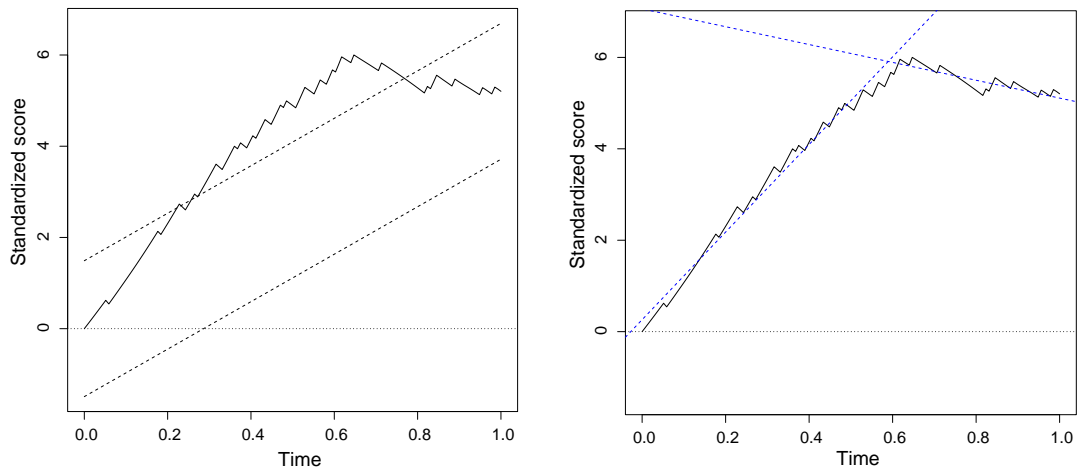
à 0.25, ce qui correspond à une augmentation de 80% de la capacité prédictive par rapport au modèle à risques proportionnels, pour lequel $R^2 = 0.14$. Entre les deux modèles à grande valeur prédictive, nous retenons le modèle de paramètre de régression $\beta_{0.6a}(t) = \beta_0 \mathbf{1}_{t \leq 0.6}$. En effet, puisque les deux modèles s'ajustent bien aux données et ont une bonne capacité prédictive, nous retenons le modèle avec le coefficient le plus simple. Le temps $t = 0.6$ sur l'échelle transformée correspond au temps 0.39 sur l'échelle initiale. Rappelons que, sur l'échelle initiale, le temps de rupture utilisé pour simuler les données est 0.4. Le coefficient de régression sélectionné est donc proche du coefficient utilisé pour simuler les données.

$\beta(t)$	$\hat{\beta}_0$	R^2
$\beta_{0.6a}(t)$	1.46	0.25
$\beta_{0.6b}(t)$	1.42	0.25
$\beta_0(1-t)^2$	2.78	0.24
$\beta_{0.5a}(t)$	1.58	0.23
$\beta_{0.5b}(t)$	1.57	0.23
$\beta_0(1-t)$	1.90	0.23
$\beta_{0.7b}(t)$	1.20	0.22
$\beta_{0.7a}(t)$	1.21	0.22
$\beta_0(1-t^2)$	1.38	0.21
$\beta_{0.4b}(t)$	1.57	0.19
$\beta_{0.4a}(t)$	1.51	0.17
β_0	0.74	0.14

TABLE 5.6 – Coefficients R^2 et estimateurs du maximum de vraisemblance partielle $\hat{\beta}_0$ pour le jeu de données simulé selon un modèle de rupture.

Effet décroissant continu. Considérons maintenant le cas d'un effet décroissant continu. Nous simulons un jeu de données avec $\beta(t) = 3(1-t)^2$ sur l'échelle transformée dans $[0, 1]$, $n = 200$ individus et une censure uniforme sur $[0, 3]$.

Le processus du score standardisé $U_n^*(0, \cdot)$ est tracé en fonction du temps en Figure 5.5(a) avec sa bande de confiance sous l'hypothèse de risques proportionnels. Le processus quitte la bande de confiance à 95%, ce qui indique que l'hypothèse de proportionnalité des risques n'est pas vérifiée. La tendance de la dérive est concave, l'effet paraît décroître dans le temps. Parmi d'autres possibilités, l'effet pourrait être linéaire, de forme quadratique, ou encore constant par morceaux. Dans le dernier cas, la tendance apparaît linéaire jusqu'au temps $t = 0.6$, correspondant à un coefficient constant. Ensuite, la dérive change pour une valeur plus faible, ce qui nous amène à considérer le coefficient $\beta_{06a}(t) = \beta_0 (\mathbf{1}_{t \leq 0.6} + C_{06} \mathbf{1}_{t > 0.6})$, où β_0 et C_{06} sont inconnus. C_{06} est la valeur par laquelle le coefficient est multiplié dans la seconde partie de l'étude. En utilisant la régression linéaire, deux droites ont été ajustées au processus, avant et après la rupture au temps $t = 0.6$ (Figure 5.5(b)). Le rapport de la deuxième pente sur la première donne la valeur $C_{06} = -0.2$. Nous avons également considéré le coefficient $\beta_{06b}(t) = \beta_0 \mathbf{1}_{t \leq 0.6}$, qui s'annule après $t = 0.6$. En utilisant le même procédé, nous avons considéré les coefficients constants par morceaux avec des temps de rupture à $t = 0.4, 0.5$ ou 0.7



(a) Processus $U_n^*(0, \cdot)$ (ligne pleine) et sa bande de confiance (lignes pointillées) (b) $U_n^*(0, \cdot)$ et ajustement d'un coefficient constant par morceaux avec une rupture à $t = 0.6$

FIGURE 5.5 – Processus du score standardisé $U_n^*(0, \cdot)$ pour le jeu de données simulé avec un effet décroissant continu.

(voir les définitions des coefficients dans l'exemple précédent). Plusieurs modèles avec des effets décroissants continûment dans le temps ont été considérés.

$\beta(t)$	$\hat{\beta}_0$	R^2
$\beta_0(1-t)^2$	3.54	0.36
$\beta_{0.6b}(t)$	1.64	0.34
$\beta_{0.6a}(t)$	1.65	0.34
$\beta_0(1-t)$	2.34	0.34
$\beta_{0.5a}(t)$	1.78	0.31
$\beta_0(1-t^2)$	1.69	0.31
$\beta_{0.5b}(t)$	1.77	0.31
$\beta_{0.7b}(t)$	1.44	0.31
$\beta_{0.7a}(t)$	1.43	0.31
$\beta_{0.4b}(t)$	2.01	0.30
$\beta_0 \exp(-t)$	1.79	0.29
$\beta_{0.4a}(t)$	1.98	0.28
β_0	0.95	0.21

TABLE 5.7 – Coefficients R^2 et estimateurs du maximum de vraisemblance partielle $\hat{\beta}_0$ pour le jeu de données simulé selon le modèle avec effet décroissant continu.

Les coefficients R^2 et les estimateurs du maximum de vraisemblance partielle $\hat{\beta}_0$ de β_0 sont présentés Table 5.7. Le coefficient R^2 le plus faible correspond au modèle à risques proportionnels. Le modèle sélectionné est celui avec le plus grand coefficient R^2 . Il s'agit du modèle associé au coefficient $\beta(t) = \beta_0(1-t)^2$, avec une estimation de β_0 égale à 3.54. Le coefficient R^2 de ce modèle vaut 0.36, correspondant à une augmentation de 70% de la valeur prédictive du modèle. Encore une fois, notons que les données ont été simulées sous le modèle

sélectionné.

Cas multivarié

Deux covariables $Z^{(1)}$ et $Z^{(2)}$ normales, centrées, de variance 1 et de corrélation 0.5 ont été simulées selon le modèle

$$\lambda(t | Z) = \exp\left(\beta_1(t)Z^{(1)} + \beta_2 Z^{(2)}\right),$$

avec $\beta_1(t) = \mathbf{1}_{t \leq 0.7}$ et $\beta_2 = -1$. Pour cet exemple, nous avons choisi de simuler les données sans censure, avec $n = 200$ individus. Chaque composante du processus bivarie $\hat{\Sigma}^{-1/2}U_n^*(0, \cdot)$ et sa bande de confiance à 95% sont tracées en fonction du temps transformé entre 0 et 1 sur les Figures 5.6(a) et 5.6(b). L'hypothèse de proportionnalité des risques pour la covariable

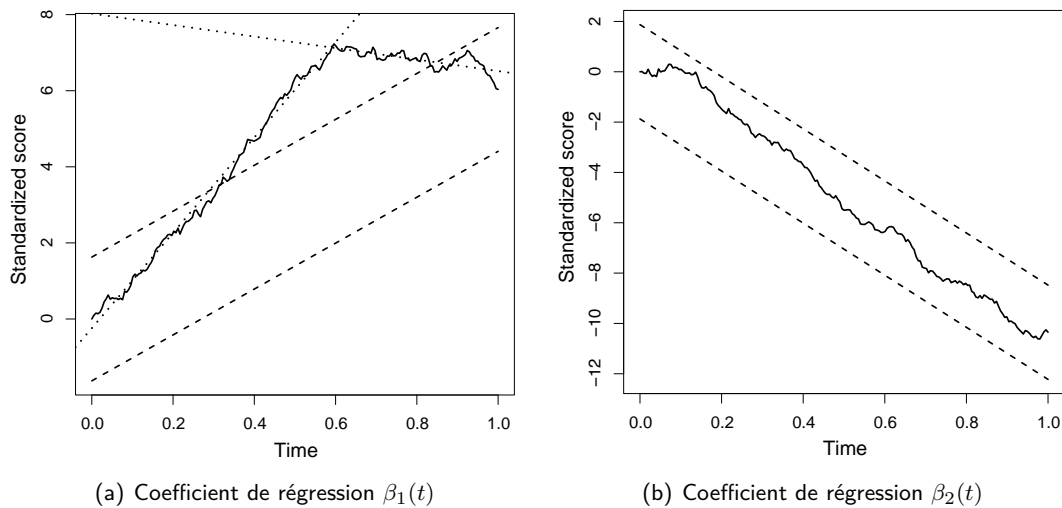


FIGURE 5.6 – Processus du score standardisé $\hat{\Sigma}^{-1/2}U_n^*(0, \cdot)$ (lignes pleines), bandes de confiance (tirets) et ajustement d'un coefficient constant par morceaux (lignes pointillées) pour les données simulées multivariées.

$Z^{(1)}$ est rejetée au niveau 5% puisque le processus quitte la bande de confiance. La forme de la dérive du processus indique un effet constant par morceaux avec une rupture au temps $t_0 = 0.6$ sur l'échelle transformée. Comme dans le cas univarié, deux droites ont été ajustées par régression linéaire au processus, une avant $t_0 = 0.6$ et une après. Le rapport des pentes est -0.12 , c'est pourquoi nous considérons l'effet de régression $\beta_1(t) = \beta_1 B_{0.6}(t)$ avec $B_{0.6}(t) = \mathbf{1}_{t \leq 0.6} - 0.12 \mathbf{1}_{t > 0.6}$. D'autres coefficients de régression constants par morceaux $\beta(t) = \beta_1 B_{t_0}(t)$ ont été étudiés, avec des ruptures aux temps $t_0 \in \{0.45, 0.5, \dots, 0.7\}$ sur l'échelle transformée. Pour chaque temps t_0 , le rapport des pentes a été évalué afin de déterminer la valeur qui multiplie le coefficient dans la deuxième partie de l'étude. La deuxième covariable, $Z^{(2)}$, semble avoir un effet de régression constant puisque le processus reste dans les bandes de confiance à 5% et a une dérive linéaire (Figure 5.6(b)). Par conséquent, nous considérons seulement le coefficient de régression $\beta_2(t) = \beta_2$. Les résultats d'estimation sont donnés Table 5.8. Le

$\beta_1(t)$	β_1	$\beta_1 B_{0.45}(t)$	$\beta_1 B_{0.5}(t)$	$\beta_1 B_{0.55}(t)$	$\beta_1 B_{0.6}(t)$	$\beta_1 B_{0.65}(t)$	$\beta_1 B_{0.7}(t)$
$\hat{\beta}_1$	0.45	0.93	0.96	0.89	0.95	0.86	0.72
$\hat{\beta}_2$	-0.73	-0.72	-0.73	-0.74	-0.79	-0.80	-0.77
R^2	0.24	0.35	0.37	0.35	0.39	0.37	0.32

TABLE 5.8 – Estimateurs du maximum de vraisemblance partielle $\hat{\beta}(t)$ et coefficients R^2 pour les données simulées multivariées.

modèle à risques proportionnels donne un coefficient R^2 de 0.24. Le R^2 maximal est obtenu avec $\beta_1(t) = \beta_1 B_{0.6}(t)$, augmentant la capacité prédictive de 60% par rapport au modèle à risques proportionnels. En conclusion, nous retenons le modèle avec $\beta_1(t) = \beta_1 B_{0.6}(t)$ et $\beta_2(t) = \beta_2$. Notons que le temps $t_0 = 0.6$ sur l'échelle transformée correspond au temps 0.68 sur l'échelle du temps initiale, ce qui est proche de la valeur 0.7 utilisée pour simuler les données.

5.5 Applications

5.5.1 Données de Freireich

Les données de Freireich sont souvent utilisées en analyse de survie ([Acute Leukemia Group B et al., 1963](#), [Cox, 1972](#)). Ces données sont connues pour respecter l'hypothèse de proportionnalité des risques. La moitié des 42 patients de l'étude atteints de leucémie reçoit le nouveau traitement étudié, la 6-Mercaptopurine, et l'autre moitié reçoit un traitement contrôle. Le processus du score standardisé $U_n^*(0, \cdot)$ et sa bande de confiance sont tracés en fonction du temps transformé dans l'échelle $[0, 1]$ en Figure 5.7. Comme nous l'attendions, le processus ne sort pas de la bande de confiance à 95%, il y a une bonne adéquation entre le modèle à risques proportionnels et les données.

5.5.2 Etude clinique

Lors d'une étude clinique, les données de 1504 patientes atteintes de cancer du sein traitées à l'Institut Curie ont été collectées. A la suite d'une première guérison, les patientes ont été suivies pendant une période de 15 ans. Le taux de censure dans la base de données est de 24%. Un des objectifs de l'étude de cette cohorte est de construire un modèle de survie descriptif qui puisse permettre de mieux connaître l'influence de certains facteurs pronostiques sur la survie des patientes après le traitement. Les effets de plusieurs de ces facteurs pronostiques peuvent changer avec le temps. Les indices pronostiques résultants, c'est-à-dire les combinaisons linéaires des effets et des facteurs pronostiques, devraient refléter ces dépendances au temps.

Les facteurs pronostiques de la base de données sont la présence d'un récepteur à progestérone, la présence de la taille de la tumeur supérieure à 60 mm et la présence du grade

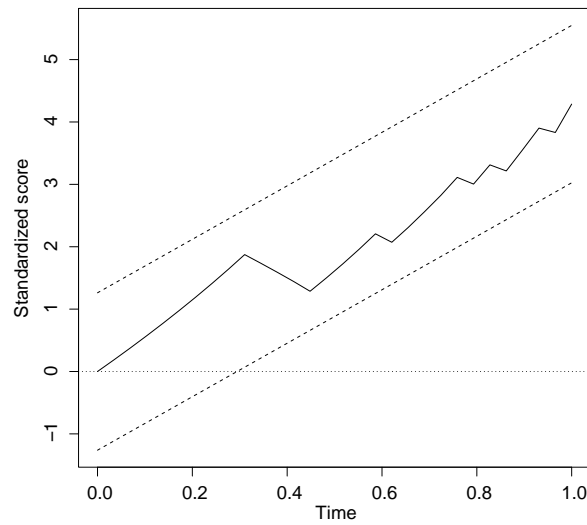


FIGURE 5.7 – Processus du score standardisé $U_n^*(0, \cdot)$ (ligne pleine) et sa bande de confiance (lignes pointillées) en fonction du temps pour les données de Freireich.

histologique supérieur à 2. Le processus du score multivarié standardisé $\hat{\Sigma}^{-1/2}U_n^*(0, \cdot)$ et ses bandes de confiance sont tracés en fonction du temps sur l'échelle transformée dans $[0, 1]$ en Figure 5.8. La Figure 5.8(a) illustre le processus correspondant à l'effet de la taille de la tumeur. Le processus quitte la bande de confiance à 95% et l'effet semble non constant avec une pente diminuant graduellement au cours du temps. Ainsi, un modèle s'ajustant mieux aux données que le modèle à risques proportionnels est un modèle avec un effet constant par morceaux pour la variable taille de la tumeur, présentant une rupture au temps $t = 0.2$ sur l'échelle du temps transformée, et des effets constants pour les deux autres facteurs pronostiques. Comme dans les simulations, deux droites ont été ajustées au processus avant et après $t = 0.2$, ce qui mène à considérer l'effet de régression pour la variable taille de la tumeur $\beta_{Taille}(t) = \beta_0(\mathbf{1}_{t \leq 0.2} + 0.24\mathbf{1}_{t > 0.2})$. La capacité prédictive de ce modèle par rapport au modèle à risques proportionnels augmente de plus de 30% du R^2 , passant de 0.29 à 0.39 (Table 5.9).

La Figure 5.8(b) représente le processus pour l'effet de la présence du récepteur à progestérone en fonction du temps sur l'échelle transformée. La dérive ne paraît pas tout à fait linéaire, mais le processus reste dans la bande de confiance correspondant à un effet constant dans le temps. Nous avons considéré plusieurs effets de régressions potentiels : un modèle de rupture avec une rupture à $t = 0.5$ sur l'échelle transformée, $\beta_{rec0}(t) = \beta_0(\mathbf{1}_{t \leq 0.5} + 0.39\mathbf{1}_{t > 0.5})$ et plusieurs effets continus $\beta_{rec1}(t) = \beta_0(1 - t)$, $\beta_{rec2}(t) = \beta_0(1 - t)^2$, $\beta_{rec3}(t) = \beta_0(1 - t^2)$ et $\beta_{rec4}(t) = \beta_0 \log(t)$.

La Figure 5.8(c) représente le processus pour l'effet du grade histologique. Le processus atteint la borne inférieure de sa bande de confiance pour un effet constant, mais ne la dépasse pas. La dérive ne paraît pas linéaire, mais donne plutôt l'impression d'un effet négatif qui s'atténue dans le temps. L'effet le plus simple, c'est-à-dire constant dans le temps semble

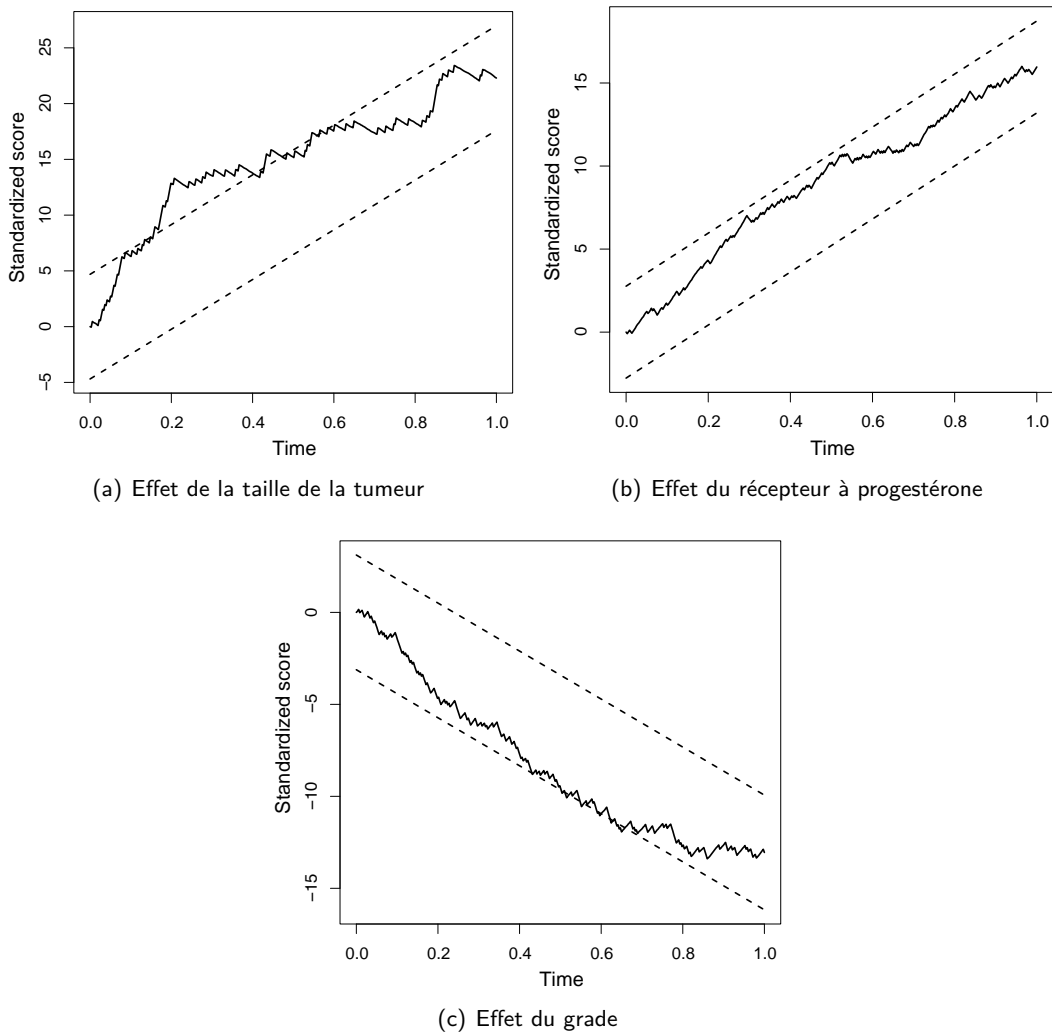


FIGURE 5.8 – Processus du score standardisé $\hat{\Sigma}^{-1/2}U_n^*(0, \cdot)$ (lignes pleines) et ses bandes de confiance (lignes pointillées) construits avec les données de l'étude clinique pour les facteurs pronostiques taille de la tumeur, récepteur à progestérone et grade histologique.

être un bon candidat, mais, dans un contexte de construction de modèle, nous considérons également des effets qui dépendent du temps. En particulier, nous sélectionnons l'effet constant par morceaux $\beta_{Grade}(t) = \beta_0(t)(\mathbf{1}_{t \leq 0.4} + 0.69\mathbf{1}_{t > 0.4})$.

Toutes les combinaisons de ces effets ainsi que le modèle à risques proportionnels ont été étudiés. Pour chaque combinaison, les effets de régression ont été estimés en maximisant la vraisemblance partielle et le coefficient R^2 a été évalué.

Une partie des résultats est présentée à la Table 5.9. Le modèle à risques proportionnels induit un coefficient R^2 de 0.29. Comme nous l'avons déjà mentionné, un modèle permettant à l'effet de la taille de la tumeur de varier de manière simple au cours du temps implique une grande amélioration de la capacité prédictive du modèle, de l'ordre de 30%. Le coefficient R^2 le plus élevé est obtenu avec des effets constants par morceaux pour la taille de la tumeur et le

Taille de la tumeur	Récepteur à progestérone	Grade	R^2
0.84	1.03	-0.68	0.29
$1.77(\mathbf{1}_{t \leq 0.2} + 0.24\mathbf{1}_{t > 0.2})$	1.03	-0.66	0.39
0.85	$-1.02 \log(t)$	-0.67	0.39
$1.74(\mathbf{1}_{t \leq 0.2} + 0.24\mathbf{1}_{t > 0.2})$	$-1.02 \log(t)$	-0.66	0.51
$1.72(\mathbf{1}_{t \leq 0.2} + 0.24\mathbf{1}_{t > 0.2})$	$-1.02 \log(t)$	$-0.82(\mathbf{1}_{t \leq 0.4} + 0.69\mathbf{1}_{t > 0.4})$	0.52

TABLE 5.9 – Estimateurs de la vraisemblance partielle et coefficients R^2 pour les données de l'étude clinique.

grade et un effet proportionnel à $\log(t)$ pour le récepteur à progestérone. La capacité prédictive de ce modèle a augmenté de 80% par rapport au modèle à risques proportionnels pour les trois facteurs pronostiques. Cependant, introduire un effet qui dépend du temps pour le grade, en ayant déjà permis aux effets de la taille de la tumeur et du récepteur à progestérone de varier dans le temps entraîne une augmentation du R^2 de 0.51 à 0.52. Une augmentation si faible ne justifie pas la complexité supplémentaire du modèle, c'est pourquoi nous retenons le modèle avec l'effet du grade constant dans le temps et des effets non constants pour la taille de la tumeur et le récepteur à progestérone.

5.6 Discussion

Conclusion du chapitre

Le processus du score standardisé est un outil permettant la construction d'un modèle à risques non proportionnels. L'idée de ce processus est d'utiliser les données brutes, sans estimer aucun paramètre et d'évaluer graphiquement la forme temporelle de l'effet de régression, comme c'est le cas pour un nuage de points en deux dimensions pour des données non censurées. Notons que les processus développés jusqu'à présent pour vérifier l'adéquation du modèle à risques proportionnels aux données utilisent une estimation du paramètre de régression.

Une seconde différence entre le processus du score standardisé et les processus introduits dans la littérature réside dans les standardisations. Deux types de standardisation sont appliqués : la première sur les temps et la seconde sur les incréments du processus. Nous avons choisi de standardiser chaque incrément plutôt que d'appliquer une standardisation globale. Ceci permet de connaître analytiquement la loi limite du processus et de décorrélérer les covariables. La connaissance de la distribution asymptotique du processus permet de dériver un test d'adéquation et une bande de confiance d'un effet constant du modèle à risques proportionnels autour de chaque composante du processus. Ces tests ne sont pas influencés par les corrélations entre les covariables puisqu'elles ont été prises en compte dans la construction du processus.

Les techniques liées et complémentaires d'adéquation et de capacité prédictive des modèles statistiques fournissent une procédure cohérente de construction de modèles. Intuitivement,

les modèles construits de cette manière devraient avoir de meilleures performances prédictives, ce qui est justifié théoriquement dans ce chapitre de thèse. Nous avons choisi d'utiliser des techniques basées sur les résidus de Schoenfeld pour les modèles à risques proportionnels et non proportionnels, à la fois pour évaluer l'adéquation du modèle par le processus du score standardisé et sa capacité prédictive par le coefficient R^2 . Un grand nombre d'approches concurrentes pourrait être utilisé, sous réserve de vérifier les mêmes résultats théoriques.

Perspectives

Un modèle de rupture, c'est-à-dire un modèle à risques non proportionnels avec un effet constant par morceaux, est un modèle plus raffiné que le modèle à risques proportionnels. La méthode de construction de modèle présentée dans cette thèse pourrait permettre d'estimer les temps de rupture, ainsi que les différentes valeurs du paramètre de régression. Dans les illustrations présentées, nous avons considéré un unique temps de rupture et le rapport entre les deux valeurs du coefficient de régression a été estimé par régression linéaire à partir de la pente du processus du score standardisé. Il serait envisageable de ne pas faire cette estimation par régression linéaire, mais d'estimer les deux parties du coefficient par maximisation de la vraisemblance partielle.

Dans la méthode que nous proposons, l'utilisateur joue un rôle important en déterminant les coefficients de régression pouvant permettre de bien représenter les données. Nous avons choisi de ne pas estimer le paramètre à partir de bases de fonctions, comme des splines ou des bases d'histogramme, afin de conserver une bonne interprétation du coefficient de régression. Le modèle retenu approchera bien les données et aura une bonne capacité prédictive, même si un autre modèle aurait pu convenir. Evidemment, il est possible d'estimer le coefficient de régression non paramétriquement et de manière plus automatique à partir de la dérive du processus.

Dans le Chapitre 4, nous avons utilisé la distance entre le processus du score standardisé et deux fois l'aire sous sa courbe pour évaluer l'écart à l'hypothèse de proportionnalité des risques. En effet, si le modèle à risques proportionnels est valide, cette distance est faible et permet de définir une statistique de test. En présence d'une covariable, la connaissance de la loi limite de cette statistique implique la construction d'un test d'adéquation. L'extension au cas multivarié n'est pas immédiate et devrait être étudiée plus en détail.

Sous le modèle à risques non proportionnels de paramètre $\beta(t)$, le processus du score standardisé $\{U_n^*(\beta(t), t), 0 \leq t \leq 1\}$ converge vers un mouvement Brownien standard. Il serait intéressant d'étudier le comportement limite du processus $\{U_n^*(\widehat{\beta}(t), t), 0 \leq t \leq 1\}$ où $\widehat{\beta}(t)$ est un estimateur convergent de $\beta(t)$. La connaissance de la distribution asymptotique de ce processus permettrait de vérifier si l'estimateur de $\beta(t)$ sélectionné par la méthode de construction de modèles présentée dans cette thèse est proche du vrai paramètre du modèle. Si le processus est évalué en $\widehat{\beta}(t)$ où $\widehat{\beta}(t)$ est l'estimateur de maximum de vraisemblance partielle de $\beta(t)$ alors le processus n'est plus (\mathcal{F}_t^*) -prévisible. Nos résultats asymptotiques étant basés sur cette propriété, la dérivation de la loi limite n'est pas immédiate. Notons toutefois que

l'évaluation de l'adéquation du modèle à risques non proportionnels de paramètre $\beta_0(t)$ fixé peut être effectuée par le test de l'hypothèse nulle $H_0 : \beta(t) = \beta_0(t)$ étudié au Chapitre 4.

Annexe - Simulations de données sous le modèle à risques non proportionnels

Dans cette annexe, nous présentons les deux méthodes développées durant la thèse pour simuler un échantillon $\{T_i, Z_i, i = 1, \dots, n\}$ sous le modèle à risques non proportionnels (1.11). La première permet de simuler des données avec un coefficient $\beta(t)$ constant par morceaux correspondant à un modèle de rupture, en utilisant la loi de T conditionnelle à Z . La deuxième méthode est basée sur la loi de Z sachant T et permet de générer des données pour n'importe quel $\beta(t)$ non constant au cours du temps. La censure est indépendante, nous simulons donc les temps $\{C_1, \dots, C_n\}$ indépendamment de cet échantillon.

.1 Méthode 1 - Modèle de rupture

Nous considérons un modèle à risques non proportionnels (1.11) avec un coefficient de régression constant par morceaux qui s'écrit, pour tout $t \in [0, \mathcal{T}]$,

$$\beta(t) = \sum_{i=1}^k \beta_i \mathbf{1}_{t_i \leq t < t_{i+1}}, \quad t_1 = 0, \quad t_{k+1} = +\infty,$$

où les valeurs β_1, \dots, β_k sont fixées. Dans un premier temps, nous considérons que les $k-1$ temps de rupture t_2, t_3, \dots, t_k sont connus et nous montrons comment simuler des données sous le modèle de paramètre $\beta(t)$. Dans un second temps, nous expliquons comment définir les points de rupture correspondant aux changements de valeur de $\beta(t)$.

.1.1 Simulation des données

Pour faciliter l'exposition de la méthode, nous nous plaçons dans le cadre d'une covariable $Z \in \mathbb{R}$ unidimensionnelle fixe dans le temps et nous considérons que le risque de base est constant : $\lambda_0(t) = 1$, pour tout $t \in [0, \mathcal{T}]$. La généralisation au cas de plusieurs covariables et/ou à des covariables qui dépendent du temps s'obtient facilement. Les données sont alors

simulées sous le modèle

$$\lambda(t | Z) = \lambda_0(t) \exp\{\beta(t)Z\} = \exp\left(Z \sum_{i=1}^k \beta_i \mathbf{1}_{t_i \leq t < t_{i+1}}\right), \quad t \in [0, T]. \quad (24)$$

Calcul de $S(\cdot | Z)$

Soit $t \in [0, T]$. Calculons la valeur de la survie conditionnelle $S(t | Z)$. Si t est compris entre $t_1 = 0$ et t_2 , la survie est

$$\begin{aligned} S(t | Z) &= \exp\left(-\int_0^t \lambda(u | Z) du\right) \\ &= \exp\left(-\int_0^t \exp(\beta(u)Z) du\right) \\ &= \exp\left(-\int_{t_1=0}^t \exp(\beta_1 Z) du\right) \\ &= \exp(-t \exp(\beta_1 Z)). \end{aligned}$$

Si t compris entre t_j et t_{j+1} , $j \in \{2, \dots, k\}$, la survie de T conditionnelle à Z et calculée en t est

$$\begin{aligned} S(t | Z) &= \exp\left(-\int_0^t \lambda(u | Z) du\right) \\ &= \exp\left(-\int_0^t \exp(\beta(u)Z) du\right) \\ &= \exp\left(-\sum_{i=1}^{j-1} \int_{t_i}^{t_{i+1}} \exp(\beta_i Z) du - \int_{t_j}^t \exp(\beta_j Z) du\right) \\ &= \exp\left(-\sum_{i=1}^{j-1} (t_{i+1} - t_i) \exp(\beta_i Z) - (t - t_j) \exp(\beta_j Z)\right). \end{aligned}$$

Pour simuler une variable aléatoire dont la fonction de survie conditionnelle aux covariables Z est $S(\cdot | Z)$, on utilise la propriété suivante

Proposition .1.1. *Soit F une fonction de répartition. Si $U \sim \mathcal{U}[0, 1]$, alors $F^{-1}(U)$ a pour fonction de répartition F .*

Il faut donc inverser la fonction de répartition $F(\cdot | Z) = 1 - S(\cdot | Z)$.

Inverse de $F(\cdot | Z)$

Soit $\gamma \in [0, 1]$ tel que

$$F(t_j | Z) \leq \gamma < F(t_{j+1} | Z), \quad j \in \{1, \dots, k\}.$$

On cherche t tel que $\gamma = F(t | Z)$. Si $0 \leq \gamma < F(t_2 | Z)$,

$$\gamma = F(t | Z) = 1 - \exp(-t \exp(\beta_1 Z)),$$

ce qui équivaut à

$$t = -\log(1 - \gamma) \exp(-\beta_1 Z).$$

Si γ est tel que $F(t_j | Z) \leq \gamma < F(t_{j+1} | Z)$, pour $j \in \{2, \dots, k\}$ alors

$$\gamma = F(t | Z) = 1 - \exp\left(-\sum_{i=1}^{j-1} (t_{i+1} - t_i) \exp(\beta_i Z) - (t - t_j) \exp(\beta_j Z)\right),$$

qui équivaut à

$$\begin{aligned} -\log(1 - \gamma) &= \sum_{i=1}^{j-1} (t_{i+1} - t_i) \exp(\beta_i Z) + (t - t_j) \exp(\beta_j Z) \\ &= \sum_{i=2}^j t_i \{\exp(\beta_{i-1} Z) - \exp(\beta_i Z)\} + t \exp(\beta_j Z). \end{aligned}$$

En conclusion,

$$t = \exp(-\beta_j Z) \left(-\log(1 - \gamma) - \sum_{i=2}^j t_i \{\exp(\beta_{i-1} Z) - \exp(\beta_i Z)\} \right).$$

En résumé, pour tout $\gamma \in [0, 1]$, on a

$$\begin{aligned} F^{-1}(\gamma | Z) &= \sum_{j=2}^k \left(e^{-\beta_j Z} \left(-\log(1 - \gamma) - \sum_{i=2}^j t_i (e^{\beta_{i-1} Z} - e^{\beta_i Z}) \right) \right) \mathbf{1}_{F(t_j | Z) \leq \gamma < F(t_{j+1} | Z)} \\ &\quad - \log(1 - \gamma) e^{-\beta_1 Z} \mathbf{1}_{\gamma \leq F(t_2 | Z)}. \end{aligned}$$

A partir d'une variable Z , d'un vecteur de coefficients $(\beta_1, \dots, \beta_k)$ et d'un vecteur de temps de rupture (t_2, \dots, t_k) , en simulant U selon une loi $\mathcal{U}[0, 1]$, alors $F^{-1}(U | Z)$ est la réalisation d'une variable aléatoire qui a pour fonction de répartition $F(\cdot | Z)$. Cette variable est le temps de survie T , qui respecte le modèle à risques non proportionnels.

Les temps de rupture doivent être judicieusement choisis afin de ne pas rencontrer de problèmes dus à l'épuisement des individus. En effet, les temps doivent être calibrés de telle sorte que des individus soient présents dans chaque intervalle $[t_j, t_{j+1}[$ et en nombre suffisant. Le choix de ces temps s'effectue en fonction des coefficients β_1, \dots, β_k choisis.

.1.2 Choix des points de rupture

Les temps de rupture t_2, \dots, t_{k-1} sont choisis numériquement, de manière à ce qu'il y ait le même effectif dans chaque intervalle de temps. Considérons l'exemple suivant : on veut simuler des données de survie selon le modèle (24) avec 3 valeurs successives de β :

$$\beta(t) = \sum_{i=1}^3 \beta_i \mathbf{1}_{t_i \leq t < t_{i+1}}, \quad t_1 = 0, \quad t_4 = +\infty,$$

et $\beta_1 = 2$, $\beta_2 = 1$ et $\beta_3 = 0$. Avant de pouvoir simuler les données selon la méthode décrite en Section 5.6, il faut choisir des valeurs pour t_2 et t_3 .

Etape 1 : On simule des données de survie avec 2000 individus selon le modèle (24) avec $\beta(t) = \beta_1 = 2$ pour tout t . On calcule l'estimateur de Kaplan-Meier correspondant. On choisit alors le temps tel que $1/3$ des individus soient décédés à ce temps. Cette valeur correspond à t_2 . Dans notre exemple, il s'agit de $t_2 = 0.1$.

Etape 2 : On simule un nouveau jeu de données avec 2000 individus selon le modèle (24) avec $\beta(t) = \beta_1 \mathbf{1}_{t \leq t_2} + \beta_2 \mathbf{1}_{t > t_2} = 2 \mathbf{1}_{t \leq t_2} + \mathbf{1}_{t > t_2}$. On calcule alors l'estimateur de Kaplan-Meier correspondant. Le temps t_3 est le temps correspondant au décès de $2/3$ des individus. Dans notre exemple, $t_3 = 0.4$.

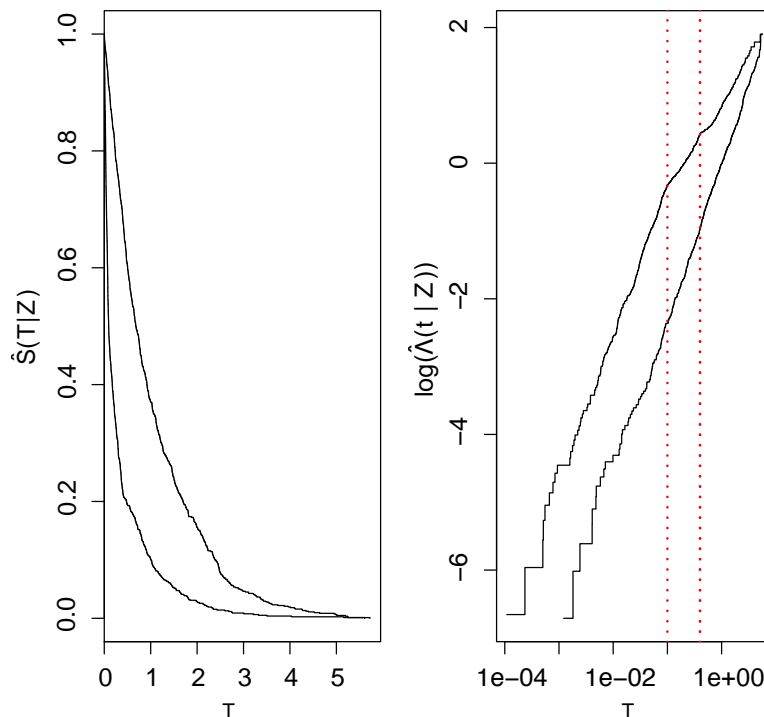


FIGURE 9 – Estimateur de Kaplan-Meier conditionnel à la covariable $\hat{S}(\cdot | Z)$ et sa transformation $\log(-\log(\hat{S}(\cdot | Z)))$ en fonction du temps.

Un jeu de données a été simulé avec les temps de ruptures déterminés ci-dessus. En Figure 9, l'estimateur de Kaplan–Meier conditionnel aux covariables $\hat{S}(\cdot | Z)$ est représenté au cours du temps, ainsi que sa transformation $\log(-\log(\hat{S}(\cdot | Z)))$. Avec cette transformation, chaque courbe représente le logarithme du risque instantané pour un groupe. On remarque une différence dans l'écart des risques pour chacune des trois parties du temps (entre 0 et $t_2 = 0.1$, entre $t_2 = 0.1$ et $t_3 = 0.4$ et pour des temps supérieurs à t_3). Cela signifie que les risques ne sont pas proportionnels.

Nous avons choisi des intervalles de temps pour lesquels le même nombre d'individus décèdent. Il est possible de déséquilibrer ces nombres de décès, tout en faisant attention au problème de l'épuisement : si aucun individu ne décède dans une tranche, l'effet correspondant ne sera pas représenté dans le jeu de données.

.2 Méthode 2 - Modèle à risques non proportionnels

Afin de générer des données sous un modèle à risques non proportionnels plus général, nous utilisons la loi conditionnelle de Z sachant T . Pour cette méthode, on considère une covariable Z qui ne dépend pas du temps. L'extension à de multiples covariables est aisée et passe par l'utilisation de l'indice pronostique.

Nous simulons les vecteurs de temps de décès (T_1, \dots, T_n) , de censure (C_1, \dots, C_n) et les vecteurs covariables (Z_1, \dots, Z_n) de manière indépendante, avec leur loi marginale. A l'aide des probabilités $\{\pi_i(\beta(t), t), i = 1, \dots, n\}$ définies équation (1.5), nous relierons les covariables et les rangs des temps de décès de la manière suivante :

1. On ordonne les temps de décès : $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$.
2. On pose $\mathcal{R} = \{1, \dots, n\}$, qui correspondra par la suite à l'ensemble courant des sujets à risques.
3. Pour i allant de 1 à n :
 - On calcule $\pi_j(\beta(T_{(i)}), T_{(i)})$ pour chaque individu $j \in \mathcal{R}$, en fixant $Y_l(t) = 1$ si $l \in \mathcal{R}$ et $Y_l(t) = 0$ sinon.
 - On échantillonne un individu j^* dans \mathcal{R} où chaque individu j de \mathcal{R} peut être choisi avec la probabilité $\pi_j(\beta(T_{(i)}), T_{(i)})$.
 - On associe Z_{j^*} à $T_{(i)}$.
 - On enlève j^* de \mathcal{R} .
4. On pose $X_i = \min(T_i, C_i)$ et $\delta_i = \mathbf{1}_{T_i \leq C_i}$ pour $i = 1, \dots, n$.

Ainsi, dans notre jeu de données $\{X_i, \delta_i, Z_i; i = 1, \dots, n\}$, les temps de décès ont été simulés sous le modèle à risques non proportionnels de paramètre $\beta(t)$ et la censure est indépendante du temps de décès. [Abrahamowicz et al. \(1996\)](#) ont proposé une méthode similaire dans laquelle les probabilités d'échantillonnage de Z sont différentes en présence de censure ou non.

Bibliographie

- AALLEN, O. (1980). A model for nonparametric regression analysis of counting processes. Dans *Mathematical statistics and probability theory. Lecture Notes in Statistics 2*. Springer-Verlag, New York, pages 1–25.
- AALLEN, O., BORGAN, O. et GJESSING, H. (2008). *Survival and Event History Analysis : A Process Point of View*. Springer. New-York.
- AALLEN, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6 : 701–726.
- AALLEN, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8 : 907–925.
- ABRAHAMOWICZ, M., MACKENZIE, T. et ESDAILE, J. (1996). Time-dependent hazard ratio : Modeling and hypothesis testing with application in lupus nephritis. *Journal of the American Statistical Association*, 91 : 1432–1439.
- ACUTE LEUKEMIA GROUP B, FREIREICH, E. J., GEHAN, E., FREI, E., SCHROEDER, L. R., WOLMAN, I. J., ANBARI, R., BURGERT, E. O., MILLS, S. D., PINKEL, D., SELAWRY, O. S., MOON, J. H., GENDEL, B. R., SPURR, C. L., STORRS, R., HAURANI, F., HOOGSTATEN, B. et LEE, S. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia : A model for evaluation of other potentially useful therapy. *Blood*, 21(6) : 699–716.
- ALLISON, P. D. (1995). *Survival Analysis Using the SAS System*. SAS Institute Inc. : Cary, North Carolina, U.S.A.
- ALTMAN, D. G. et ANDERSEN, P. (1986). A note on the uncertainty of a survival probability estimated from Cox's regression model. *Biometrika*, 73(3) : 722–724.
- ANDERSEN, P. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics*, 38 : 67–77.
- ANDERSEN, P. et GILL, R. (1982). Cox's regression model for counting processes : A large sample study. *The Annals of Statistics*, 10 : 1100–1120.

- ANDERSEN, P. K., BORGAN, O., GILL, R. et KEIDING, N. (1993). *Statistical models based on counting Processes*. Springer-Verlag. New York.
- ANDERSEN, P. K., CHRISTENSEN, E., FAUERHOLDT, L. et SCHLICHTING, P. (1983). Evaluating prognoses based on the proportional hazards model. *Scandinavian Journal of Statistics*, 10(2) : 141–144.
- ANDERSON, J. et SENTHILSELVAN, A. (1982). A two-step regression model for hazard function. *Journal of the Royal Statistical Society. Series C*, 31 : 44–51.
- ARJAS, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association*, 83 : 204–212.
- ARJAS, E. et GASBARRA, D. (1994). Nonparametric bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, 4 : 505–524.
- BALAKRISHNAN, N. et RAO, C. R., rédacteurs (2003). *Advances in Survival Analysis*. Handbook of Statistics. Elsevier.
- BALTAZAR-ABAN, I. et PEÑA, E. (1995). Properties of hazard-based residuals and implications in model diagnostics. *Journal of the American Statistical Association*, 90 : 185–219.
- BARLOW, W. E. et PRENTICE, R. L. (1988). Residuals for relative risk regression. *Biometrika*, 75(1) : 65–74.
- BENNETT, S. (1983a). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2 : 273–277.
- BENNETT, S. (1983b). Log-logistic regression models for survival data. *Applied Statistics*, 32 : 165–171.
- BERAN, R. (1981). Nonparametric regression with randomly censored survival data. *Technical Report, University of California, Berkeley*.
- BHATTACHARYA, R. et WAYMIRE, E. (1990). *Stochastic Processes with Applications*. Wiley. New York.
- BILLINGSLEY, P. (1999). *Convergence of probability measures, Second Edition*. John Wiley & Sons Inc. New York.
- BLANCHE, P. (2013). *Contribution à l'évaluation de capacités pronostiques en présence de données censurées, de risques concurrents et de marqueurs longitudinaux. Inférence et applications à la prédiction de la démence*. Thèse de doctorat, Université Bordeaux Segalen.
- BLANCHE, P., LATOUCHE, A. et VIALON, V. (2013). Time-dependent AUC with right-censored data : A survey study. Dans *Risk Assessment and Evaluation of Predictions*, édité par LEE, M., GAIL, G., CAI, T., PFEIFFER, R. et GANDY, A. Springer.

- BRESLOW, N. (1970). A generalized Kruskal–Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, 57 : 579–594.
- BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30 : 89–99.
- BRESLOW, N., ELDER, L. et BERGER, L. (1984). A two sample censored–data rank test for acceleration. *Biometrics*, 40 : 1042–1069.
- BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78 : 1–3.
- BROWN, C. (1975). On the use of indicator variables for studying the time–dependence of parameters in a response–time model. *Biometrics*, 31 : 863–872.
- CAI, Z. et SUN, Y. (2003). Local linear estimation for time–dependent coefficients in Cox’s regression models. *Scandinavian Journal of Statistics*, 30 : 93–111.
- CHAUVEL, C. et O’QUIGLEY, J. (2014a). Tests for comparing estimated survival functions. *Biometrika*, In press.
- CHAUVEL, C. et O’QUIGLEY, J. (2014b). Survival model construction guided by fit and predictive strength. *Submitted*.
- CHENG, S., WEI, L. et YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82 : 835–845.
- CHOODARI-OSKOOEI, B., ROYSTON, P. et PARMAR, M. K. B. (2012). A simulation study of predictive ability measures in a survival model I : Explained variation measures. *Statistics in Medicine*, 31(23) : 2627–2643.
- COX, D. R. (1972). Regression models and life–tables (with discussion). *Journal of the Royal Statistical Society. Series B*, 34(2) : 187–220.
- COX, D. R. (1975). Partial likelihood. *Biometrika*, 63 : 269–276.
- COX, D. R. et SNELL, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B*, 30(2) : 248–275.
- CROWLEY, J. J. et STORER, B. E. (1983). Comment on 'A reanalysis of the Stanford Heart Transplant Data', by M. Aitkin, N. Laird and B. Francis. *Journal of the American Statistical Association*, 78 : 277–281.
- DABROWSKA, D. M. et DOKSUM, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, 18 : 1–23.
- DONSKER, M. (1951). An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6 : 1–10.

- ENG, K. H. et KOSOROK, M. R. (2005). A sample size formula for the supremum log-rank statistic. *Biometrics*, 61 : 86–91.
- FLEMING, T. et HARRINGTON, D. (1991). *Counting Processes and Survival Analysis*. Wiley. New York.
- FLEMING, T. et HARRINGTON, D. (2005). *Counting Processes and Survival Analysis. Second edition*. Wiley. New York.
- FLEMING, T. R. et HARRINGTON, D. P. (1984). Evaluation of censored survival data test procedures based on single and multiple statistics. Dans *Topics in Applied Statistics*. Marcel Dekker : New York, page 97–123.
- FLEMING, T. R., HARRINGTON, D. P. et O’SULLIVAN, M. (1987). Supremum versions of the log-rank and generalized Wilcoxon statistics. *Journal of the American Statistical Association*, 82 : 312–320.
- FLEMING, T. R., O’FALLON, J. R., O’BRIEN, P. C. et HARRINGTON, D. P. (1980). Modified Kolmogorov–Smirnov test procedures with application to arbitrarily right-censored data. *Biometrics*, 36 : 607–625.
- GAMERMAN, D. (1991). Dynamic bayesian models for survival data. *Journal of the Royal Statistical Society. Series C*, 40 : 63–79.
- GEHAN, E. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52 : 203–223.
- GERDS, T., CAI, T. et SCHUMACHER, M. (2008). The performance of risk prediction models. *Biometrical Journal*, 50 : 457–479.
- GERDS, T. A. et SCHUMACHER, M. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48 : 1029–1040.
- GILL, R. (1980). Censoring and stochastic integrals. *Mathematical Center Tract 124*.
- GRAF, E., SCHMOOR, C., SAUERBREI, W. et SCHUMACHER, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18 : 2529–2545.
- GRAMBSCH, P. M. et THERNEAU, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3) : 515–526.
- GRAY, R. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87 : 942–951.
- GU, W. et PEPE, M. (2009). Measures to summarize and compare the predictive capacity of markers. *International Journal of Biostatistics*, 5(1) : Article 27.

- GUSTAFSON, P. (1998). Flexible bayesian modelling for survival data. *Lifetime Data Analysis*, 4 : 281–299.
- GÖNEN, M. et HELLER, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92 : 965–970.
- HAARA, P. (1987). A note on the asymptotic behaviour of the empirical score in Cox's regression model for counting processes. Dans *Proceedings of the 1st World Congress of the Bernoulli Society*. VNU Science Press, Tashkent, Soviet Union, pages 139–142.
- HANLEY, J. A. et MCNEIL, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143 : 29–36.
- HARRELL, F. E., CALIFF, R. M., PRYOR, D. B., LEE, K. L. et ROSATI, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247 : 2543–2546.
- HARRELL, F. E., LEE, K. L., CALIFF, R. M., PRYOR, D. B. et ROSATI, R. A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, 3 : 143–152.
- HASTIE, T. et TIBSHIRANI, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, 46 : 1005–1016.
- HASTIE, T. et TIBSHIRANI, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B.*, 55 : 757–796.
- HEAGERTY, P. J., LUMLEY, T. et PEPE, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2) : 337–344.
- HEAGERTY, P. J. et ZHENG, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 61(1) : 92–105.
- HELLAND, I. (1982). Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics*, 9 : 79–94.
- HELLER, G. (2012). A measure of explained risk in the proportional hazards model. *Biostatistics*, 13 : 315–325.
- HESS, K. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine*, 13 : 1045–1062.
- HIELSCHER, T., ZUCKNICK, M., WERFT, W. et BENNER, A. (2010). On the prognostic value of survival models with application to gene expression signatures. *Statistics in Medicine*, 29 : 818–829.

- HILDEN, J. et GERDS, T. (2013). A note on the evaluation of novel biomarkers : do not rely on integrated discrimination improvement and net reclassification index. *Statistics in Medicine*. *In press*.
- IBRAHIM, J. G., CHEN, M.-H. et SINHA, D. (2001). *Bayesian Survival Analysis*. Springer. New-York.
- JONES, M. P. et CROWLEY, J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics*, 45 : 157–170.
- JONES, M. P. et CROWLEY, J. (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *The Annals of Statistics*, 18 : 1203–1220.
- KALBFLEISCH, J. et PRENTICE, R. (1973). Marginal likelihood based on Cox's regression and life model. *Biometrika*, 60 : 267–278.
- KALBFLEISCH, J. D. et PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, Inc.
- KAPLAN, E. L. et MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282) : 457–481.
- KAY, R. (1977). Proportional hazard regression models and the analysis of censored survival data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(3) : 227–237.
- KENT, J. et O'QUIGLEY, J. (1988). Measures of dependence for censored survival data. *Biometrika*, 75(3) : 525–534.
- KENT, J. T. (1983). Information gain and a general measure of correlation. *Biometrika*, 70 : 163–173.
- KLEIN, J. et MOESCHBERGER, M. (2003). *Survival Analysis : Techniques for Censored and Truncated Data*. Springer.
- KLEIN, J. P., VAN HOUWELINGEN, H. C., IBRAHIM, J. G. et SCHEIKE, T. H., rédacteurs (2013). *Handook of Survival Analysis*. Chapman & Hall/CRC.
- KORN, E. L. et SIMON, R. (1990). Measures of explained variation for survival data. *Statistics in Medicine*, 9 : 487–503.
- KOSOROK, M. R. et LIN, C. Y. (1999). The versatility of function-indexed weighted log-rank statistics. *Journal of the American Statistical Association*, 94 : 320–332.
- LAGAKOS, S. (1981). The graphical evaluation of explanatory variables in proportional hazards regression models. *Biometrika*, 68 : 93–98.
- LAGAKOS, S. W., KIM, L. L.-Y. et ROBINS, J. M. (1990). Adjusting for early treatment termination in comparative clinical trials. *Statistics in Medicine*, 9 : 1417–1424.

- LEE, J. (1996). Some versatile tests based on the simultaneous use of weighted log-rank statistics. *Biometrics*, 52 : 721–725.
- LEE, S.-H. (2011). Maximum of the weighted Kaplan–Meier tests for the two-sample censored data. *Journal of Statistical Computation and Simulation*, 81 : 1017–1026.
- LENGLART, E. (1977). Relation de domination entre deux processus. *Annales de l'Institut Henri Poincaré*, 13 : 171–179.
- LEURGANS, S. (1983). Three classes of censored data rank tests : Strengths and weaknesses under censoring. *Biometrika*, 70 : 651–658.
- LEURGANS, S. (1984). Asymptotic behavior of two—sample rank tests in the presence of random censoring. *The Annals of Statistics*, 12 : 572–589.
- LIN, D. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, 86 : 153–180.
- LIN, D., ROBINS, J. et WEI, L. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*, 83 : 381–393.
- LIN, D., WEI, L. et YING, Z. (1993). Checking the Cox model with cumulative sums of martingale based residuals. *Biometrika*, 80 : 557–572.
- LIN, D. Y. (2013). Transformation models. Dans *Handbook of Survival Analysis*. Chapman & Hall/CRC.
- LIN, D. Y. et YING, Z. (1995). Semiparametric analysis of general additive–multiplicative intensity for counting processes. *The Annals of Statistics*, 23 : 1712–1734.
- MADDALA, G. S. (1983). *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge U.K.
- MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50 : 163–70.
- MANTEL, N. et STABLEIN, D. (1988). The crossing hazard function problem. *The Statistician*, 37 : 59–64.
- MARTINUSSEN, T. et PENG, L. (2013). Alternatives to the Cox Model. Dans *Handbook of Survival Analysis*, édité par KLEIN, J. P., VAN HOUWELINGEN, H. C., IBRAHIM, J. G. et SCHEIKE, T. H. Chapman & Hall/CRC.
- MARTINUSSEN, T. et SCHEIKE, T. (2006). *Dynamic Regression Models for Survival Data*. Springer.
- MARZEC, L. et MARZEC, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika*, 84 : 901–908.

- MCKEAGUE, I. et SASIENI, P. (1994). A partly parametric additive risk model. *Biometrika*, 81 : 501–514.
- MCKEAGUE, I. et UTIKAL, K. J. (1990). Inference for a nonlinear counting process regression model. *The Annals of Statistics*, 18 : 1172–1187.
- MOREAU, T., O’QUIGLEY, J. et MESBAH, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. *Journal of the Royal Statistical Society. Series C*, 34(3) : 212–218.
- MÜLLER, M., DÖRING, A., KÜCHENHOFF, H., LAMINA, C., MALZAHN, D., BICKEBÖLLER, H., VOLLMERT, C., KLOPP, N., MEISINGER, C., HEINRICH, J., KRONENBERG, F., ERICH WICHMANN, H. et HEID, I. (2008). Quantifying the contribution of genetic variants for survival phenotypes. *Genetic Epidemiology*, 32(6) : 574–585.
- MURPHY, S. (1993). Testing for a time-dependent coefficient in Cox’s regression model. *Scandinavian journal of Statistics*, 20 : 35–50.
- MURPHY, S. et SEN, P. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications*, 39 : 153–180.
- NAGELKERKE, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78 : 691–692.
- NELSON, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14 : 945–965.
- O’QUIGLEY, J. (1982). Regression models and survival prediction. *Statistician*, 31 : 106–116.
- O’QUIGLEY, J. (2003). Khmaladze-type graphical evaluation of the proportional hazards assumption. *Biometrika*, 90 : 577–584.
- O’QUIGLEY, J. (2008). *Proportional Hazards Regression*. Springer. New York.
- O’QUIGLEY, J. et FLANDRE, P. (1994). Predictive capability of proportional hazards regression. *Proceedings of the National Academy of Sciences*, 91(6) : 2310–2314.
- O’QUIGLEY, J., FLANDRE, P. et REINER, E. (1999). Large sample theory for Schemper’s measure of explained variation in the Cox model. *Journal of the Royal Statistical Society. Series D.*, 48 : 53–62.
- O’QUIGLEY, J. et PESSIONE, F. (1989). Score tests for homogeneity of regression effect in the proportional hazards model. *Biometrics*, 45 : 135–144.
- O’QUIGLEY, J. et PESSIONE, F. (1991). The problem of a covariate-time qualitative interaction in a survival study. *Biometrics*, 47 : 101–115.

- O'QUIGLEY, J. et XU, R. (2001). Explained variation in proportional hazards regression. Dans *Handbook of Statistics in Clinical Oncology*, édité par CROWLEY, J. et HOERING, A. Marcel Dekker, New York, pages 397–410.
- PECKOVA, M. et FLEMING, T. R. (2003). Adaptive test for testing the difference in survival distributions. *Lifetime Data Analysis*, 9 : 223–238.
- PENCINA, M. J., D'AGOSTINO, R. B. et VASAN, R. S. (2008). Evaluating the added predictive ability of a new marker : From area under the ROC curve to reclassification and beyond. *Statistics in medicine*, 27 : 157–172.
- PEPE, M. et FLEMING, T. (1989). Weighted Kaplan–Meier statistics : A class of distance tests for censored survival data. *Biometrics*, 45 : 497–507.
- PEPE, M. et JANES, H. (2013). Methods for evaluating prediction performance of biomarkers and tests. Dans *Risk Assessment and Evaluation of Predictions*, édité par LEE, M.-L., GAIL, G., CAI, T., PFEIFFER, R. et GANDY, A. Springer.
- PEPE, M., JANES, H., LONGTON, G., LEISENRING, W., NEWCOMB, P. et FLEMING, T. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology*, 159 : 882–890.
- PEPE, M., ZHENG, Y., JIN, Y., HUANG, Y., PARIKH, C. et LEVY, W. (2008). Evaluating the ROC performance of markers for future events. *Lifetime Data Analysis*, 14 : 86–113.
- PETO, R. et PETO, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A*, 135 : 185–207.
- PRENTICE, R. et SELF, S. (1983). Asymptotic distribution theory for Cox–type regression models with general relative risk form. *The Annals of Statistics*, 11 : 804–813.
- PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, 65 : 167–179.
- REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete*, 51 : 269–286.
- ROSS, S. (1996). *Stochastic Processes. Second edition*. Wiley. New York.
- ROUAM, S. (2011). *Développement d'un indice de séparabilité adapté aux données de génomique en analyse de survie*. Thèse de doctorat, Université Paris Sud.
- ROYSTON, P. (2006). Explained variation for survival models. *Stata Journal*, 6(1) : 83–96.
- SASIENI, P. (1993). Some new estimators for Cox regression. *Annals of Statistics*, 21 : 1721–1759.

- SASIENI, P. et WINNETT, A. (2003). Martingale difference residuals as a diagnostic tool for the Cox model. *Biometrika*, 90 : 899–912.
- SATTEN, G. A. et DATTA, S. (2001). The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician*, 55(3) : 207–210.
- SCHEIKE, T. et MARTINUSSEN, T. (2004). On estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics*, 31 : 51–62.
- SCHEIKE, T. et ZHANG, M. (2002). An additive–multiplicative Cox–Aalen regression model. *Scandinavian Journal of Statistics*, 29 : 79–92.
- SCHEMPER, M. (1990). The explained variation in proportional hazards regression. *Biometrika*, 77(1) : 216–218.
- SCHEMPER, M. et HENDERSON, R. (2000). Predictive accuracy and explained variation in Cox regression. *Biometrics*, 56 : 249–255.
- SCHEMPER, M. et STARE, J. (1996). Explained variation in survival analysis. *Statistics in Medicine*, 15 : 1999–2012.
- SCHMID, M., HIELSCHER, T., AUGUSTIN, T. et GEFELLER, O. (2010). A robust alternative to the Schemper–Henderson estimator of prediction error. *Biometrics* : 524–545.
- SCHOENFELD, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67 : 145–153.
- SCHOENFELD, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69 : 239–241.
- SCHOOP, R., GRAF, E. et SCHUMACHER, M. (2008). Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*, 64 : 603–610.
- SCHUMACHER, M. (1984). Two-sample tests of Cramer–von Mises and Kolmogorov–Smirnov–type for randomly censored data. *International Statistical Review*, 62 : 263–281.
- SELF, S. (1991). An adaptive weighted log–rank test with application to cancer prevention and screening trials. *Biometrics*, 47 : 975–86.
- SHEN, Y. et CAI, J. (2001). Maximum of the weighted Kaplan–Meier tests with application to cancer prevention and screening trials. *Biometrics*, 57 : 837–843.
- SINHA, D., CHEN, M.-H. et GHOSH, S. (1999). Bayesian analysis and model selection for interval-censored survival data. *Biometrics*, 55 : 585–590.
- SONG, X. et ZHOU, X. H. (2008). A semiparametric approach for the covariate specific ROC curve with survival outcome. *Statistica Sinica*, (18) : 947–965.

- STABLEIN, D., CARTER, W. et NOVAX, J. (1981). Analysis of survival data with non—proportional hazard functions. *Controlled Clinical Trials*, 2 : 149–159.
- STARE, J., PERME, M. P. et HENDERSON, R. (2011). A measure of explained variation for event history data. *Biometrics*, 67(3) : 750–759.
- STEYERBERG, E. W., VICKERS, A. J., COOK, N. R., GERDS, T., GONEN, M., OBUCHOWSKI, N., PENCINA, M. J. et KATTAN, M. W. (2010). Assessing the performance of prediction models : A framework for some traditional and novel measures. *Epidemiology*, 21.
- STRUTHERS, C. et KALBFLEISCH, J. (1986). Misspecified proportional hazards model. *Biometrika*, 73 : 363–369.
- STUTE, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, 45(1) : 89–103.
- STUTE, W. (1995). The central limit theorem under random censorship. *The Annals of Statistics*, 23 : 422–439.
- TARONE, R. E. (1981). On the distribution of the maximum of the log–rank statistic and the modified Wilcoxon statistic. *Biometrics*, 37 : 79–85.
- TARONE, R. E. et WARE, J. H. (1977). On distribution–free tests for equality for survival distributions. *Biometrika*, 64 : 156–160.
- TAULBEE, J. (1979). A general model for the hazard rate with covariables. *Biometrics*, 35 : 439–450.
- THERNEAU, T. et GRAMBSCH, P. (2000). *Modeling Survival Data : Extending the Cox Model*. Springer. New York.
- THERNEAU, T., GRAMBSCH, P. et FLEMING, T. (1990). Martingale–based residuals for survival models. *Biometrika*, 77(1) : 147–160.
- VALSECCHI, M. et SILVESTRI, D. (1996). Evaluation of long–term survival : Use of diagnostics and robust estimators with Cox’s proportional hazards model. *Statistics in Medicine*, 15 : 2763–2780.
- VERWEIJ, P. et VAN HOUWELINGEN, H. (1995). Time–dependent effects of fixed covariates in Cox regression. *Biometrics*, 51 : 1550–1556.
- WARE, J. (2006). The limitations of risk factors as prognostic tools. *New England Journal of Medicine*, 355 : 2615–2617.
- WEI, L. (1984). Testing goodness–of–fit for proportional hazards model with censored observations. *Journal of the American Statistical Association*, 79 : 649–652.

- WEI, L. (1992). The accelerated failure time model : A useful alternative to the Cox regression model in survival analysis. *Statistics In Medicine*, 11 : 1871–1879.
- WHITT, W. (2002). *Stochastic-process limits : An introduction to stochastic-process limits and their application to queues*. Springer. New York.
- WILCOXON, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1 : 80–83.
- WINNETT, A. et SASIENI, P. (2003). Iterated residuals and time-varying covariate effects in Cox regression. *Journal of the Royal Statistical Society. Series B*, 65 : 473–488.
- WU, L. et GILBERT, P. B. (2002). Flexible weighted log-rank tests optimal for detecting early and/or late survival differences. *Biometrics*, 58 : 997–1004.
- XU, R. (1996). *Inference for the proportional hazards model*. Thèse de doctorat, University of California, San Diego.
- XU, R. et O'QUIGLEY, J. (1999). A R2 type measure of dependence for proportional hazards models. *Journal of Nonparametric Statistics*, 12 : 83–107.
- XU, R. et O'QUIGLEY, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1(4) : 423–439.
- YANG, S. et PRENTICE, R. (2005). Semiparametric analysis of short term and long term relative risks with two sample survival data. *Biometrika*, 92 : 1–17.
- YANG, S. et PRENTICE, R. (2010). Improved logrank-type tests for survival data using adaptive weights. *Biometrics*, 66 : 30–38.
- ZUCKER, D. M. et KARR, A. F. (1990). Nonparametric survival analysis with time-dependent effects : A penalized partial likelihood approach. *The Annals of Applied Statistics*, 18 : 329–353.
- ZUCKER, D. M. et LAKATOS, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika*, 77 : 853–864.