

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET

DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MENTOURI-CONSTANTINE

FACULTE DES SCIENCES

DEPARTEMENT DE MATHÉMATIQUES



N° d'ordre :.....
Série :.....

Année : 2007

MEMOIRE

Présenté Pour Obtenir Le Diplôme De Magistère En Mathématiques

THEME

MODELES SEMI-PARAMETRIQUES EN ANALYSE DE SURVIE

OPTION

Mathématiques Appliquées

Présenté par

BOUOUDEN RABAH

Devant le Jury :

Président :	Z. MOHDEB	Prof	U. Mentouri Constantine
Rapporteur :	F. MESSACI	M. C	U. Mentouri Constantine
Examineur :	S. DJEZZAR	M. C	U. Mentouri Constantine
Examineur :	A. LANANI	M. C	U. Mentouri Constantine

النماذج النصف وسيطية في دراسة مدة الحياة

ملخص:

هذه المذكرة تقترح دراسة مفصلة للنموذج النصف وسيطي لـ Cox. الذي هو من أهم

النماذج النصف وسيطية التي تدرس تأثير العوامل الخارجية (Covariables) علي دالة البقاء على قيد الحياة.

في النهاية تطرقنا إلى بعض النماذج النصف وسيطية التي هي إما امتداد للنموذج النصف وسيطي لـ Cox أو تعميما له.

Modèles semi-paramétriques en analyse de survie

Résumé

Ce mémoire se propose de détailler l'étude du modèle semi-paramétrique de Cox, qui s'est avéré central dans les études de vie soumises à des influences extérieures.

En fin ce travail a été complété par l'étude d'autres modèles semi-paramétriques obtenus par extension ou généralisation du modèle de Cox.

Semi-parametric Models in Survival Analysis

Summary

This work suggests a detailed study of the semi-parametric Cox model which is one of the most important semi-parametric models that study the influence of the covariables (explanatory variables) on the survival function.

At the end, we presented a study of other semi-parametric models that are either extensions or generalisations of the Cox model.

Table des matières

1	Notions préliminaires	5
1.1	Les approches d'estimation	5
1.2	Données de survie	6
1.2.1	Censure à droite	7
1.2.2	Censure à gauche	7
1.2.3	Censure par intervalle	8
1.3	Fonctions de base en analyse de survie	8
1.4	Modélisation par un processus de comptage	11
1.5	Rappel sur les processus	13
2	Le modèle de Cox	19
2.1	Définition du modèle	19
2.2	L'estimation par la méthode de vraisemblance partielle	21
2.2.1	Construction de la vraisemblance partielle pour le modèle de Cox	22
2.3	L'estimation de β	27
2.4	Tests d'Hypothèses	31
2.4.1	Test global	31
2.4.2	Test local	31
2.5	Estimation de la fonction de hasard λ_0	33
2.6	Exemples	35

3	Propriétés asymptotiques	48
3.1	Rappels	48
3.2	consistance de $\hat{\beta}$	52
3.3	Normalité asymptotique	55
3.4	Cas particulier	68
4	Autres modèles semi-paramétriques	76
4.1	Modèle de Cox stratifié	76
4.2	Le modèle de fragilité	79
4.3	Le modèle de fragilité corrélé	81
4.4	Modèle de Cox avec covariable manquante	82
	4.4.1 Construction de la vraisemblance conjointe	83
	4.4.2 Propriétés asymptotiques des L'EMVNP	85
4.5	Modèle additif	87
	4.5.1 Définition de modèle	87
	4.5.2 Estimation de $B_j, 0 \leq j \leq p$	88
4.6	Le modèle de Cox-Aalen	91

Introduction

Etudier des durées de survie c'est s'intéresser à l'apparition d'un évènement au cours du temps, quelque soit son type (décès, apparition d'une maladie, récurrence d'une maladie,...). Il est nécessaire pour cela de disposer du temps de suivi de chaque sujet ainsi que du moment auquel l'évènement s'est produit, dans le cas où celui-ci a eu lieu. La particularité de ces études est l'existence de données incomplètes, appelées données censurées, pour les sujets n'ayant pas présenté l'évènement en question au moment du bilan de l'étude, ce qui nécessite une méthodologie adaptée pour leur analyse. L'analyse de durées de survie trouve donc des applications dans de nombreux domaines, biomédicaux ou non, tels que l'industrie (pour des études de fiabilité d'appareils...).

Les premières analyses statistiques d'une durée de survie remonteraient au XVII^{ème} siècle. D'autres contributions voient le jour par la suite :

- en 1951 un modèle paramétrique dans la fiabilité fut conçu par Weibull fournissant une nouvelle distribution de probabilité qui sera utilisée souvent en fiabilité (c'est la loi de Weibull),

- en 1958, furent présentés d'importants résultats par Kaplan et Meier concernant l'estimation non-paramétrique de la fonction de survie,

- en 1972, les travaux de Cox introduisent un modèle statistique semi-paramétrique comportant des variables exogènes.

Depuis sa parution, ce modèle s'est avéré central dans les études de durées de vie soumises à des influences extérieures (cofacteurs). Ainsi si on s'intéresse à l'estimation du taux de défaillance d'une pièce de voiture, à sa sortie d'usine, notée λ_0 , des observations recueillies dans différents pays ne pourront être considérées comme régies par le même taux de panne λ_0 . Ce taux dépend de l'état des routes, de l'entretien de la voiture, du climat etc. Le modèle de Cox peut alors servir à modéliser ces données en considérant que le taux de défaillance λ_i de la variable X_i s'obtient en multipliant λ_0 par un facteur dépendant du produit scalaire du vecteur des cofacteurs par le vecteur β de régression. Le succès de ce modèle tient au fait qu'il s'est avéré satisfaisant dans plusieurs applications.

Ce modèle fait intervenir deux inconnues, la fonction de hasard de base λ_0 et le vecteur de régression (de dimension finie β). On peut spécifier pour λ_0 une loi dont on estimera les paramètres (cf Kalbfleish pour les détails). Mais l'inconvénient est que la loi réelle peut être trop éloignée de celle choisie. Dans ce mémoire aucune loi n'est spécifiée pour λ_0 . Le fait que ce modèle comporte une composante non paramétrique (λ_0) et une composante paramétrique (β) le classe parmi les modèles dits semi-paramétriques. C'est le modèle semi-paramétrique le plus utilisé et il est à la base de l'introduction de plusieurs autres modèles. C'est pourquoi ce mémoire se propose d'abord de détailler l'étude du modèle de Cox en donnant les estimateurs des paramètres de ce modèle et leurs propriétés que nous retrouvons en utilisant l'approche par les processus de comptage et les martingales initiée par Aalen et bien développée par Andersen et Gill pour ce modèle. Le chapitre quatre se propose de passer en revue d'autres modèles qui sont obtenus par extension ou généralisation du modèle de Cox. Le premier chapitre regroupe les notions de base de l'analyse de survie ainsi que quelques notions et résultats sur les processus de comptage et les martingales utilisés dans la suite du mémoire.

Chapitre 1

Notions préliminaires

1.1 Les approches d'estimation

Commençons par les trois approches possibles d'estimation.

Définition 1

Un modèle statistique paramétrique consiste en l'observation d'une variable aléatoire X de loi $f(X/\theta)$ où seul le paramètre θ , appartenant à un sous ensemble Θ d'un espace vectoriel de dimension finie n'est pas connu.

Exemple 2

On suppose que f est la densité de la loi normale, les paramètres inconnus sont μ (moyenne) et σ^2 (variance), ils appartiennent à \mathbb{R}^2 .

L'avantage de ce modèle est la facilité attendue de la phase d'estimation des paramètres. L'inconvénient de la méthode paramétrique est l'inadéquation pouvant exister entre le phénomène étudié et le modèle retenu.

Définition 3

Un modèle statistique non-paramétrique consiste en l'observation d'une variable aléatoire X dont le paramètre appartient à un espace de dimension infinie.

Exemple 4

On s'intéresse à un phénomène de loi complètement inconnue et on cherche à estimer sa fonction de répartition ou sa densité (en supposant qu'elle existe) ou sa moyenne etc... L'inconvénient d'une telle approche est la nécessité de disposer d'un nombre important d'observations, le problème de l'estimation du paramètre étant délicat puisqu'il appartient à un espace de dimension infinie.

Définition 5

Un modèle statistique semi-paramétrique est une sorte de mélange entre les deux approches précédentes. La loi de probabilité est supposée appartenir à une classe de lois pour partie dépendant d'un paramètre appartenant à un espace de dimension finie et pour partie d'un paramètre appartenant à un espace de dimension infinie.

Cette approche est répandue en analyse de la survie, notamment à travers le modèle de régression de Cox (1972).

1.2 Données de survie

Les données de survie mesurent le temps d'occurrence d'un certain évènement qui peut être le développement d'une maladie, la guérison ou la mort suite à une traitement, la durée d'abonnement à un journal,...

Cependant les données sont le plus souvent incomplètes dans le sens où les observations recueillies ne sont pas toujours les véritables temps de survie. Ainsi, par exemple, pour une raison de coût et (ou) de temps il n'est pas toujours possible d'observer tous les éléments d'un échantillon jusqu'à ce qu'ils "tombent tous en panne". Dans la pratique l'expérimentateur est amené à fixer soit le temps d'étude, soit le nombre d'articles dont on attendra "la panne" pour terminer l'étude. Ce qui veut dire que pour certains articles les véritables temps de survie ne sont pas observés, les données correspondant à ces articles sont dites censurées. D'autres types de censures peuvent apparaître, nous résumons ci-dessous quelques types que l'on peut rencontrer.

1.2.1 Censure à droite

La censure à droite de type 1 correspond au cas où le véritable temps de survie n'est observé que s'il est inférieur à un temps spécifié. Si on note T le temps de survie et C le temps de censure, on observe seulement le couple :

$$X = T \wedge C \text{ et } \delta = 1_{\{T \leq C\}}$$

C'est-à-dire, on n'observe le véritable temps de survie que s'il est inférieur à C , dans ce cas la donnée n'est pas censurée et $\delta = 1$. Sinon on sait seulement que T dépasse C , la donnée est censurée et $\delta = 0$. δ est appelé indicateur de censure. La censure à droite de type 2 est le cas où l'étude continue jusqu'à ce qu'on observe r véritables temps de survie où r est déterminé d'avance. Ici c'est le temps de l'étude qui est aléatoire. Des sujets peuvent être retirés de l'étude aléatoirement à cause d'accidents, ou de migration d'individus etc.... Les données sont aléatoirement censurées.

Exemple 6 *On s'intéresse au temps de survie de personnes atteintes d'une maladie. On fixe le temps d'étude et à la fin de ce temps certaines personnes sont encore vivantes. Pour ces personnes tout ce que l'on sait est que leur temps de survie dépasse le temps observé, ce sont des données censurées à droite de type 1.*

1.2.2 Censure à gauche

Dans ce cas le temps de survie T n'est observé que s'il est supérieur à un temps de censure C , on observe seulement le couple :

$$X = T \vee C \quad \text{et} \quad \delta = 1_{\{T \geq C\}}.$$

Remarquons que les données peuvent être, en même temps censurées à gauche et à droite, c'est ce qu'on appelle censure mixte.

Exemple 7

Sur le même exemple (6) que précédemment, on ne peut pas toujours savoir le moment exact du déclenchement de la maladie. Ainsi si on s'intéresse à l'âge du début de la maladie, pour certaines personnes, on sait seulement que leur âge est inférieur à leur âge au moment de l'étude. Ces données sont censurées à gauche.

1.2.3 Censure par intervalle

Il arrive que le véritable temps de survie n'est pas connu mais on sait seulement qu'il appartient à un certain intervalle. Ceci est le cas lorsque des inspections des machines ne sont faites que périodiquement (par exemple chaque semaine). Si une panne survient, on sait seulement qu'elle est produite dans l'intervalle de temps séparant deux inspections. C'est une généralisation de la censure à droite et de la censure à gauche puisque pour la première on sait que le temps de survie appartient à un intervalle $[c, \infty[$ et pour la seconde il appartient à un intervalle $[0, c]$.

1.3 Fonctions de base en analyse de survie

Certaines fonctions de base jouent un rôle important en analyse de survie. Nous allons les introduire et donner les liens existant entre elles.

Définition 8

Notons T ($T \geq 0$), la variable aléatoire qui représente la durée de survie d'un sujet dans une expérience, de fonction de répartition

$$F(t) = P(T \leq t) \tag{1.1}$$

a) Fonction de survie

La fonction de survie, notée $S(t)$, est définie comme la probabilité qu'un individu survive

au-delà du temps t . i.e

$$S(t) = P(T > t) = 1 - F(t) \quad (1.2)$$

b) Taux de hasard

Si T admet une densité f alors le taux de hasard, noté λ , qui est la probabilité de défaillance dans un intervalle de temps très court, sachant que T dépasse t est défini par :

$$\lambda(t) = \begin{cases} \frac{f(t)}{S(t)} & \text{si } S(t) \neq 0 \\ 0 & \text{si } S(t) = 0 \end{cases} \quad (1.3)$$

En effet la terminologie taux se justifie par le fait que si f est continue, alors pour tout $t > 0$ tel que $P(T > t) > 0$, on a

$$\lambda(t) = \lim_{h \rightarrow 0} h^{-1} P \{t < T \leq t + h/T > t\} \quad (1.4)$$

En effet

$$\begin{aligned} \lim_{h \rightarrow 0} h^{-1} P \{t < T \leq t + h/T > t\} &= \lim_{h \rightarrow 0} h^{-1} \frac{P \{t < T \leq t + h, T > t\}}{P \{T > t\}} \\ &= \lim_{h \rightarrow 0} h^{-1} \frac{P \{t < T \leq t + h\}}{S(t)} \\ &= \frac{1}{S(t)} \lim_{h \rightarrow 0} h^{-1} (F(t + h) - F(t)) \\ &= \frac{1}{S(t)} F'(t) \end{aligned}$$

or f étant continue $F'(t) = f(t)$ donc

$$\frac{f(t)}{S(t)} = \lambda(t) = \lim_{h \rightarrow 0} h^{-1} P \{t < T \leq t + h/T > t\}$$

c) Fonction de hasard cumulé

Supposons que λ est intégrable sur tout compact (ce qui est le cas si f est continue),

on définit la fonction de hasard cumulé, notée Λ en posant

$$\Lambda(t) = \int_0^t \lambda(x) dx \quad (1.5)$$

La proposition suivante donne le lien entre ces trois notions.

Proposition 9

Supposons que T admette une densité f qui soit une fonction continue sur \mathbb{R}_+^* , et soit

$$A = \{t > 0 : S(t) \neq 0\}$$

alors les propriétés suivantes sont équivalentes.

- a) $\forall t \in A, \lambda(t) = \frac{f(t)}{S(t)}$
- b) $\forall t \in A, \lambda(t) = (-\log S(t))'$
- c) $\forall t \in A, S(t) = \exp(-\int_0^t \lambda(x) dx)$
- d) $\forall t \in A, f(t) = \lambda(t) \exp(-\int_0^t \lambda(x) dx)$

Preuve.

1) On a

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{F'(t)}{S(t)} = \frac{-S'(t)}{S(t)} = (-\log S(t))'$$

Ce qui montre l'équivalence entre a) et b)

2) La relation $\lambda(t) = (-\log S(t))'$ montre que $-\log S(t)$ est une primitive de $\lambda(t)$ (qui est continue puisque f et S le sont), donc

$$-\int_0^t \lambda(s) ds = \log S(t) - \log S(0) = \log S(t)$$

car $S(0) = p(T > 0) = p(T \geq 0) = 1$ (T est une v.a. positive et continue). d'où

$$S(t) = \exp\left(-\int_0^t \lambda(s) ds\right)$$

Ce qui montre l'équivalence entre $b)$ et $c)$

3)

$$\begin{aligned} S(t) &= \exp\left(-\int_0^t \lambda(s)ds\right) \implies \\ S'(t) &= -\lambda(t) \exp\left(-\int_0^t \lambda(s)ds\right) \end{aligned}$$

par continuité de λ . Par ailleurs on sait que

$$S'(t) = -f(t) \implies f(t) = \lambda(t) \exp\left(-\int_0^t \lambda(s)ds\right)$$

ce qui montre que $c) \implies d)$

4) Si

$$f(t) = \lambda(t) \exp\left(-\int_0^t \lambda(s)ds\right)$$

alors

$$\begin{aligned} S'(t) &= \left(\exp\left(-\int_0^t \lambda(s)ds\right)\right)' \implies \\ S(t) &= \exp\left(-\int_0^t \lambda(s)ds\right) + c \end{aligned}$$

or $S(0) = 1 = \exp(0) + c \implies c = 1$. D'où $d) \implies c)$. ■

1.4 Modélisation par un processus de comptage

L'étude des durées de survie dans le cas de censure à droite, c'est à dire que l'on observe, au lieu de T , la durée $X = T \wedge C$, ainsi que l'indicatrice de censure $\delta = 1_{\{T \leq C\}}$, peut être abordée d'une autre façon. Au lieu de considérer T , la durée étudiée, qui est une variable aléatoire réelle positive, généralement continue, de densité f , de fonction de répartition F et de fonction de survie $S = 1 - F$, on représente l'expérience par le processus de comptage associé $N(t)$, défini par

$$N(t) = 1_{\{X \leq t, \delta=1\}} \quad , \quad t \geq 0 \quad (1.6)$$

et l'indicateur de présence du sujet juste avant l'instant t défini par :

$$Y(t) = 1_{\{X \geq t\}} \quad , \quad t \geq 0$$

Considérons maintenant la durée de vie sous l'aspect de processus de comptage défini dans (1.6). Pour faire un modèle statistique, on doit préciser sur quelle information il est fondé. Pour un processus de comptage, cela est fait en spécifiant l'histoire, souvent appelée filtration, et notée $\{\mathcal{F}_t, t \geq 0\}$. Un choix naturel pour $\{\mathcal{F}_t, t \geq 0\}$ est l'histoire de l'expérience depuis le début (le temps 0) jusqu'à l'instant t inclus. Le processus N satisfait les conditions de Doob donc il existe un unique processus \mathcal{F}_t prévisible A tel que $N - A$ est un martingale. Donc

$$E(dN(t))/\mathcal{F}_{t-} = E(dA(t))/\mathcal{F}_{t-} = dA(t)$$

Si nous supposons que $dA(t) = l(s)ds$ où l est une fonction aléatoire le processus A est appelé processus d'intensité cumulé pour N et on appelle $l(s)$ processus d'intensité pour N . Dans l'approche du processus de comptage pour les données censurées le processus d'intensité est modelé par :

$$l(s) = Y(t)\lambda(t)$$

où $\lambda(t)$ est le taux de hasard de T (cf Dolivo 1974),(Aalen 1978) et (Gill 1980).

Cette approche par les processus de comptage à été initiée par Aalen et développée dans les ouvrages d'Andersen et autres (1993), Coccozza (1997) et Fleming Harrengton (1991). Elle sert à étudier les propriétés des estimateurs d'une manière très rigoureuse en se servant des propriétés des martingales. Nous allons donc rappeler certains résultats qui seront utilisés dans la suite de ce mémoire.

1.5 Rappel sur les processus

Soit (Ω, \mathcal{A}, P) un espace de probabilité.

Définition 10 Processus stochastique

Soit T un ensemble d'indices. Un processus stochastique est une famille de variables aléatoires $(X_t, t \in T)$ définies sur (Ω, \mathcal{A}, P) . En général, on prend $T = [0, \infty[$ ou $T = \mathbb{N}$.

Définition 11 Filtration

1) une filtration $(\mathcal{F}_t)_{t \geq 0}$ est une famille croissante de sous tribus de \mathcal{A} , c'est-à-dire : $\mathcal{F}_t \subset \mathcal{F}_s$ pour tout $t \leq s$.

2) quand $(\mathcal{F}_t)_{t \geq 0}$ est une filtration, la σ -algèbre $\bigcap_{h > 0} \mathcal{F}_{t+h}$ est noté par \mathcal{F}_{t+} .

Si $\mathcal{F}_t = \mathcal{F}_{t+}$ alors la filtration est dite continue à droite, on note cad.

3) \mathcal{F}_{t-} est la plus petite σ -algèbre qui contient toutes les σ -algèbres \mathcal{F}_{t-h} pour $h > 0$.

Définition 12

* Un processus stochastique $X = (X_t, t \geq 0)$ est dit adapté (par rapport à une filtration $(\mathcal{F}_t)_{t \geq 0}$ si X_t est \mathcal{F}_t mesurable pour tout t .

* Pour ω fixé, l'application $t \rightarrow X(t, \omega)$ est appelée trajectoire du processus.

* On dit que le processus $(X_t, t \geq 0)$ est à trajectoire continues (ou est continu), si pour presque tout ω l'application $t \rightarrow X(t, \omega)$ est continue.

* Un processus stochastique $X = (X_t, t \geq 0)$ est dit à variation bornée sur $[0, t]$ si $\exists K \geq 0$ telle que presque sûrement

$$\sup_{t_i} \sum_i |X_{t_{i+1}} - X_{t_i}| \leq K$$

le sup étant pris sur les subdivisions $0 \leq t_0 < \dots < t_i < t_{i+1} \leq t$.

Définition 13

* Un processus stochastique $X = (X_t, t \geq 0)$ est un processus croissant s'il est à valeurs dans \mathbb{R}_+ , $X_0 = 0$ et si presque sûrement l'application $t \rightarrow X_t$ est une fonction

croissante, c'est-à-dire

$$\forall t \leq s \quad X_t \leq X_s \quad p.s$$

* Un processus $X = (X_t, t \geq 0)$ est dit borné si $\exists M$ (constante finie) telle que

$$P \left\{ \sup_{0 \leq t < \infty} |X_t| < M \right\} = 1$$

Définition 14

* Un temps d'arrêt est une variable aléatoire T à valeurs dans $\bar{\mathbb{R}}_+$ telle que

$$\forall t \in \mathbb{R}_+ \quad \{T \leq t\} \in \mathcal{F}_t.$$

* Une suite croissante de v.a. $(\tau_n)_{n \in \mathbb{N}^*}$ est une suite localisante (par rapport à la filtration $(\mathcal{F}_t)_{t \geq 0}$, si

1) $\forall n \in \mathbb{N}^*$, τ_n est un temps d'arrêt.

2) $\lim_{n \rightarrow \infty} \tau_n = +\infty$ p.s.

* Le processus $X = (X_t, t \geq 0)$ est dit localement borné s'il existe une suite localisante $(\tau_n)_{n \in \mathbb{N}^*}$ telle que

$$\forall n \in \mathbb{N}^*, X_n = \{X(t \wedge \tau_n), t \geq 0\}$$

est borné.

Définition 15

Soit (Ω, \mathcal{A}, P) un espace de probabilité et $(\mathcal{F}_t)_{t \geq 0}$ une filtration. La σ algèbre sur $[0, +\infty[\times \Omega$ engendrée par tous les ensembles de la forme

$$\{0\} \times A, A \in \mathcal{F}_0$$

ou

$$[a, b] \times A, 0 \leq a \leq b < \infty, A \in \mathcal{F}_a$$

est la σ algèbre des prévisibles pour la filtration $(\mathcal{F}_t)_{t \geq 0}$.

Proposition 16

1) Soit $X = (X_t, t \geq 0)$ un processus prévisible. Alors pour tout $t > 0$, X_t est \mathcal{F}_{t-} mesurable.

2) La tribu des prévisibles est la tribu engendrée par les processus adaptés et continus à gauche.

Définition 17

1) Un processus de comptage $\{N(t) : t \geq 0\}$ est un processus adapté à une filtration avec $N(0) = 0, N(t) < \infty$ p.s, dont les trajectoires sont continues à droite, constantes par morceaux et ont pour seules discontinuités des sauts de taille 1.

2) Soit le processus $N = (N_1, N_2, \dots, N_n)$, on dit que N est un processus de comptage multivarié si

i) pour tout $i = 1, 2, \dots, n$ N_i est un processus de comptage

ii) $\forall i \neq j, \forall t \geq 0,$

$$P \{ \Delta N_i(t) = \Delta N_j(t) = 1 \} = 0$$

Il n'y a presque sûrement pas de saut en même temps pour N_i et N_j .

Définition 18

Un processus $M = (M_t, t \geq 0)$ est une martingale à temps continu par rapport à la filtration $(\mathcal{F}_t)_{t \geq 0}$ si

1) M_t est \mathcal{F}_t adapté.

2) $\forall t, E(|M_t|) < \infty$

3) $E(X_t/\mathcal{F}_s) = X_s$ pour tout $0 < s < t$

Définition 19

1) Une famille de variables aléatoires $(M_t, t \geq 0)$ est une martingale locale par rapport à la filtration $(\mathcal{F}_t)_{t \geq 0}$ s'il existe une suite localisante $(\tau_n)_{n \geq 1}$ telle que

$$\forall n \in \mathbb{N}^*, M_n = \{M(t \wedge \tau_n) : 0 \leq t < \infty\}$$

est une martingale.

2) Si M_n est une martingale de carré intégrable, M est dite martingale locale de carré intégrable.

Définition 20 processus de variation quadratique

Soit M une martingale continue à droite telle que $E(M_t^2) < \infty \forall t \geq 0$. Il existe un seul processus croissant noté $\langle M, M \rangle$ prévisible, tel que :

* pour tout $t \geq 0$, $E(\langle M, M \rangle(t)) < \infty$

* $M^2 - \langle M, M \rangle$ est une martingale.

Proposition 21 (cf Fleming Harrington page 46)

Soit N un processus de comptage avec $E(N(t)) < \infty$ pour chaque t . Soit $(\mathcal{F}_t)_{t \geq 0}$ une filtration continue à droite, telle que :

1) $M = N - A$ est une \mathcal{F}_t martingale, où A est un processus croissant prévisible, avec $A(0) = 0$.

2) H est un processus borné et prévisible.

Alors le processus

$$L(t) = \int_0^t H(s) dM(s)$$

est une martingale.

Remarque 22

Soit N un processus de comptage. On peut montrer (cf Fleming) l'existence d'un unique processus A continu à droite croissant et prévisible tel que $A(0) = 0$ p.s et le processus $M = N - A$ est une \mathcal{F}_t martingale locale de carré intégrable, elle est dite associée à N

Proposition 23 (cf Fleming Harrington)

Soit $(\mathcal{F}_t)_{t \geq 0}$ une filtration continue à droite. H un processus \mathcal{F}_t prévisible et localement borné et N un processus de comptage. Soit $M = N - A$ la \mathcal{F}_t martingale associée à N de

carré intégrable. Alors

$$X(t) = \int_0^t H dM$$

est une martingale locale.

Proposition 24 (cf Fleming Harrington)

Soit $(\mathcal{F}_t)_{t \geq 0}$ une filtration continue à droite si

1) H_i est un processus \mathcal{F}_t prévisible et borné par $\Gamma_i (i = 1, 2)$

2) N_i est un processus de comptage borné par $K_i (i = 1, 2)$

3) la \mathcal{F}_t martingale $M_i = N_i - A_i$ vérifie $\forall t, E(M_i^2(t)) < \infty$, où A_i est le compensateur de $N_i (i = 1, 2)$ alors

$$\left\langle \int H_1 dM_1, \int H_2 dM_2 \right\rangle = \int H_1 H_2 d\langle M_1, M_2 \rangle$$

Théorème 25 théorème d'arrêt optionnel (cf Fleming Harrington)

Soit $(X_t, t \geq 0)$ une \mathcal{F}_t martingale (resp sous martingale) cad où $\{\mathcal{F}_t\}$ est cad. Soit τ un \mathcal{F}_t temps d'arrêt alors $(X, (t \wedge \tau), t \geq 0)$ est une martingale (resp sous martingale).

Théorème 26 Inégalité de Langlart (cf Fleming Harrington)

Soit X un processus continu à droite et adapté et Y un processus croissant prévisible avec $Y_0 = 0$. On suppose que pour tout temps d'arrêt T

$$E\{|X(T)|\} \leq E\{|Y(T)|\}$$

Alors pour tout temps d'arrêt T et pour tout $\epsilon, \eta > 0$, on a

$$P\left\{\sup_{t \leq T} |X_t| \geq \epsilon\right\} \leq \frac{\eta}{\epsilon} + P\{Y(t) \geq \eta\}$$

Corollaire 27 (cf Fleming Harrington)

Soit N un processus de comptage et soit $M = N - A$ la martingale locale de carré intégrable associée. On suppose que H est prévisible et localement borné. Alors pour tout

temps d'arrêt T tel que $P(T < \infty) = 1$ et pour tout $\epsilon, \eta > 0$

$$P \left\{ \sup_{t \leq T} \left| \int_0^t H(x) dM(x) \right| \geq \epsilon \right\} \leq \frac{\eta}{\epsilon} + P \left\{ \int_0^T H^2(x) d \langle M, M \rangle (x) \geq \eta \right\}$$

Corollaire 28 (cf Andersen et Gill (1982))

Soit W un martingale locale de carré intégrable, alors

$\forall \delta, \eta > 0$:

$$\Pr \left\{ \sup_{t \in [0,1]} W(t) > \eta \right\} \leq \frac{\delta}{\eta^2} + \Pr \{ \langle W, W \rangle(1) > \delta \}$$

Chapitre 2

Le modèle de Cox

Le modèle de Cox est employé lorsqu'on cherche à tenir compte et à évaluer l'effet de certains facteurs, appelés covariables, sur la durée de survie. Ces covariables peuvent être constantes (sexe, groupes de traitement, race, état initial de la maladie, ... , etc.). Comme ils peuvent aussi dépendre du temps (une covariable qui prend la valeur 0 jusqu'à ce que quelque évènement intermédiaire se produise et prend alors la valeur 1, est un exemple de covariables dépendant du temps). D'autres exemples de covariables dépendant du temps sont les mesures de quelques caractéristiques explicatives (tension, cholestérol, dimension de la tumeur,...etc.)). Dans le cas de covariables dépendantes du temps on suppose que : pour tout $i = 1, 2, \dots, n$ les valeurs de $Z_i(t)$ sont connues au temps où le malade sous observation (qui a pour vecteur de covariables $Z_i(t)$) est ni mort ni censuré.

2.1 Définition du modèle

Les $2n$ variables T_1, T_2, \dots, T_n et C_1, C_2, \dots, C_n que sont les durées de survie et les durées de censure des n individus sont considérés indépendantes. On observe la suite des n couples de variables (X_i, δ_i) où $X_i = T_i \wedge C_i$ et δ_i est l'indicatrice de censure ($\delta_i = 1$ c'est on observe T_i , 0 sinon) $\delta_i = 1_{\{T_i \leq C_i\}}$. Mais on a aussi observé sur chaque individu i , p covariables

$$Z_i(t) = (Z_{i1}(t), Z_{i2}(t), \dots, Z_{ip}(t))$$

dont dépend la durée de vie T_i . Le modèle de Cox suppose que le taux de hasard d'un individu qui a pour vecteur de covariables $Z_i(t)$ est donné par

$$\lambda(t/Z_i(t)) = \lambda_0(t) \exp \beta' Z_i(t) \quad (2.1)$$

où

$$\beta = (\beta_1, \beta_2, \dots, \beta_p)'$$

est le vecteur des coefficients de la régression. Il s'agit d'estimer ces coefficients pour évaluer l'impact de chacune des covariables sur la durée étudiée. λ_0 est le taux de hasard de base. C'est une fonction inconnue, qu'il faut estimer elle aussi. Nous sommes donc en présence d'un modèle semi-paramétrique. Comme la fonction de survie se déduit de la fonction de hasard, le modèle de Cox peut encore s'écrire

$$S(t/Z(t)) = [S_0(t)]^{\exp \beta' Z(t)} \quad (2.2)$$

Le modèle de Cox est souvent appelé le modèle de hasard proportionnel parce que le rapport des taux de hasard de deux individus i et j , qui ont les covariables fixes Z et Z^* , vaut

$$\frac{\lambda(t/Z)}{\lambda(t/Z^*)} = \frac{\lambda_0(t) \exp \beta' Z}{\lambda_0(t) \exp \beta' Z^*} = \exp \beta'(Z - Z^*)$$

L'estimation dans le modèle de Cox se base sur deux méthodes. La première fondée sur la vraisemblance partielle pour trouver l'estimateur de β et la deuxième basée sur la vraisemblance totale pour estimer le taux de hasard de base λ_0 .

2.2 L'estimation par la méthode de vraisemblance partielle

L'estimation basée sur la vraisemblance partielle a été proposée par Cox en 1975 pour les modèles qui contiennent des paramètres de nuisance. Ensuite il a utilisé cette idée dans la présentation du modèle de hasards proportionnels. Le taux de hasard de base λ_0 est traité comme un paramètre de nuisance quand on estime le vecteur de coefficients β . Supposons que X , un vecteur aléatoire de densité $f_X(x, \theta)$ où $\theta = (\varphi, \beta)$, est un vecteur des paramètres. La question d'estimation fondamentale se pose au sujet de β , avec φ un paramètre de nuisance. Dans quelques situations le vecteur X peut être transformé en deux composantes V et W et de même pour la densité de X . La vraisemblance pour θ peut s'écrire comme un produit d'une vraisemblance marginale et d'une vraisemblance conditionnelle.

$$f_X(x, \theta) = f_{W/V}(w/v, \theta) f_V(v, \theta)$$

où $X' = (W', V')$.

La vraisemblance partielle proposée par Cox, est fondée sur une idée tout à fait semblable à celle de la vraisemblance conditionnelle. Supposons que le vecteur des données X peut être transformé en une suite de paires

$$(V_1, W_1, V_2, W_2, \dots, V_L, W_L)$$

au lieu d'une seule paire. (le nombre L de paires dépend aléatoirement de X). La vraisemblance pour θ s'écrit

$$\begin{aligned} f_X(x, \theta) &= f_{V_1, W_1, V_2, W_2, \dots, V_L, W_L}(v_1, w_1, v_2, w_2, \dots, v_L, w_L, \theta) \\ f_X(x, \theta) &= \prod_{i=1}^L f_{W_i/V_i, W_1, \dots, V_i}(w_i/v_i, w_1, \dots, v_i, \theta) \\ &\quad \times f_{V_i/V_1, W_1, \dots, W_{i-1}}(v_i/v_1, w_1, \dots, w_{i-1}, \theta) \end{aligned}$$

D'où

$$f_X(x, \theta) = \prod_{i=1}^L f_{W_i / Q_i}(w_i / q_i, \theta) \times \prod_{i=1}^L f_{V_i / P_i}(v_i / p_i, \theta) \quad (2.3)$$

où $P_1 = \Phi$, $Q_1 = V_1$ et pour tout $i = 2, 3, \dots, L$

$$P_i = (V_1, W_1, V_2, W_2, \dots, V_{i-1}, W_{i-1}) \quad (2.4)$$

et

$$Q_i = (V_1, W_1, V_2, W_2, \dots, W_{i-1}, V_i) \quad (2.5)$$

Quand le premier produit dans l'équation (2.3) dépend seulement de β , Cox appelait ce produit "la vraisemblance partielle" pour β basée sur W dans la suite

$$(V_1, W_1, V_2, W_2, \dots, V_L, W_L)$$

2.2.1 Construction de la vraisemblance partielle pour le modèle de Cox

La vraisemblance partielle quand il n'y a pas d'ex-æquos

Supposons que les covariables ne dépendent pas du temps (constantes). Si les covariables dépendent du temps c'est une généralisation que nous donnons plus loin. Supposons aussi pour simplifier que la durée de vie a une distribution absolument continue et il n'y a pas d'ex-æquos parmi les données. Dans la suite nous appelons l'évènement d'intérêt une mort. Les données du modèle de Cox peuvent être transformées en une suite $(V_1, W_1, V_2, W_2, \dots, V_L, W_L)$ comme suit.

Supposons qu'il y ait L morts aux temps

$$T_{(1)} < T_{(2)} < \dots < T_{(L)}$$

avec $T_{(0)} = 0$ et $T_{(L+1)} = \infty$. Si (i) désigne l'individu mort à l'instant $T_{(i)}$ alors les covariables associées aux L morts sont respectivement $Z_{(1)}, Z_{(2)}, \dots, Z_{(L)}$. On pose $(L + 1) \equiv n + 1$. Supposons qu'il y ait m_i données censurées

$$T_{(i,1)} < T_{(i,2)} < \dots < T_{(i,m_i)}$$

dans l'intervalle $[T_{(i)}, T_{(i+1)}[$ et soit (i, j) désignant l'individu qui a pour temps de censure $T_{(i,j)}$ donc les covariables associées à ces données censurées sont $Z_{(i,1)}, Z_{(i,2)}, \dots, Z_{(i,m_i)}$. Pour $i = 0, 1, 2, \dots, L$ on pose

$$V_{i+1} = \{T_{(i+1)}, T_{(i,j)}, (i, j) : j = 1, 2, \dots, m_i\}$$

et

$$W_{i+1} = \{(i + 1)\}$$

Donc la paire (P_i, Q_i) définie par les formules (2.4) et (2.5) est : P_i contient toutes les données censurées qui sont avant $T_{(i-1)}$, tous les données complètes jusqu'à $T_{(i-1)}$ avec les indices correspondant à toutes ces données. Q_i contient tout ce qu'il y a dans P_i , de plus il contient $T_{(i)}$ et les données censurées de l'intervalle $[T_{(i-1)}, T_{(i)}[$ ainsi que les indices correspondant à ces données censurés. L'estimation de β est basée sur l'information dans P_{i+1} sachant Q_i , $i = 0, 1, 2, \dots, L$, donc basée sur la vraisemblance partielle

$$L(\beta) = \prod_{i=1}^L f_{W_i/Q_i}(w_i/q_i) \quad (2.6)$$

$$L(\beta) = \prod_{i=1}^L P\{W_i = (i)/Q_i\}. \quad (2.7)$$

Pour trouver l'expression de $L(\beta)$ notons que

$$P\{W_i = (i)/Q_i\}$$

est la probabilité que l'individu d'indice (i) meurt à l'instant $T_{(i)}$ sachant

$$R_{(i)} = \{j : X_j \geq T_{(i)}\}$$

et sachant qu'une mort se produit en $T_{(i)}$. Or La probabilité pour qu'une mort se produise dans un petit intervalle de temps $[T_{(i)}; T_{(i)} + dt)$ sachant $R_{(i)} = \{j : X_j \geq T_{(i)}\}$ vaut

$$\sum_{j \in R_{(i)}} \lambda(T_{(i)} / Z_j) dt$$

et la probabilité pour que cette mort soit (i) sachant $R_{(i)} = \{j : X_j \geq T_{(i)}\}$ et sachant qu'une mort a eu lieu vaut

$$\begin{aligned} P\{W_i = (i)/Q_i\} &= \frac{\lambda(T_{(i)} / Z_{(i)}) dt}{\sum_{j \in R_{(i)}} \lambda(T_{(i)} / Z_j) dt} \\ &= \frac{\exp \beta' Z_{(i)}}{\sum_{j \in R_{(i)}} \exp \beta' Z_j} \end{aligned}$$

Remplaçons $P\{W_i = (i)/Q_i\}$ dans (2.7) on obtient

$$L(\beta) = \prod_{i=1}^L \frac{\exp \beta' Z_{(i)}}{\sum_{j \in R_{(i)}} \exp \beta' Z_j} \quad (2.8)$$

La vraisemblance partielle quand il y a des ex-æquos

Eliminons la supposition selon laquelle le temps de mort a une distribution absolument continue. Supposons qu'il y ait L instants de mort aux temps

$$T_{(1)} < T_{(2)} < \dots < T_{(L)}$$

et supposons qu'il y ait m_i données censurées

$$T_{(i,1)} < T_{(i,2)} < \dots < T_{(i,m_i)}$$

dans l'intervalle $[T_{(i)}, T_{(i+1)}[$. Soit (i, j) désignant l'individu qui a pour temps de censure $T_{(i,j)}$. Supposons aussi qu'il y ait D_i morts à chaque instant $T_{(i)}$, $1 \leq i \leq L$ et soit

$$\{(i)^j : j = 1, 2, \dots, D_i\}$$

l'ensemble des indices des individus mourant en T_i . Pour chaque $i = 0, 1, 2, \dots, L$ on pose

$$V_{i+1} = \{T_{(i+1)}, D_i, T_{(i,j)}, (i, j) : j = 1, 2, \dots, m_i\}$$

et

$$W_{i+1} = \{(i+1)^1, (i+1)^2, \dots, (i+1)^{D_{i+1}}\}$$

La vraisemblance partielle basée sur P_{i+1} sachant Q_i pour $i = 1, 2, \dots, L$ est

$$L(\beta) = \prod_{i=1}^L P \{W_i = \{(i)^1, (i)^2, \dots, (i)^{D_i}\} / Q_i\} \quad (2.9)$$

où

$$P \{W_i = \{(i)^1, (i)^2, \dots, (i)^{D_i}\} / Q_i\}$$

est la probabilité selon laquelle les individus d'indices $(i)^1, (i)^2, \dots, (i)^{D_i}$ meurent à l'instant $T_{(i)}$, sachant

$$R_{(i)} = \{j : X_j \geq T_{(i)}\}$$

et sachant qu'il y a D_i morts à l'instant $T_{(i)}$. Soit $\mathcal{G}_{(i)}$ l'ensemble de tous les sous ensembles de D_i élément que l'on peut choisir de $R_{(i)}$. si $g \in \mathcal{G}_{(i)}$ alors $g = (g_1, g_2, \dots, g_{D_i})$. On pose $S_g^* = \sum_{j=1}^{D_i} Z_{g_j}$ et soit $S_{(i)} = \sum_{j=1}^{D_i} Z_{(i)^j}$ alors

$$\begin{aligned}
P \{W_i = \{(i)^1, (i)^2, \dots, (i)^{D_i}\} / Q_i\} &= \frac{\prod_{j=1}^{D_i} \lambda(T_{(i)} / Z_{(i)^j}) dt}{\sum_{g \in \mathcal{G}_{(i)}} \prod_{j=1}^{D_i} \lambda(T_{(i)} / Z_{g_j}) dt} \\
&= \frac{\exp \beta' S_{(i)}}{\sum_{g \in \mathcal{G}_{(i)}} \exp \beta' S_g^*}
\end{aligned}$$

remplaçons $P \{W_i = \{(i)^1, (i)^2, \dots, (i)^{D_i}\} / Q_i\}$ dans 2.9 on obtient

$$L(\beta) = \prod_{i=1}^L \frac{\exp \beta' S_{(i)}}{\sum_{g \in \mathcal{G}_{(i)}} \exp \beta' S_g^*} \quad (2.10)$$

Remarque 29

Deux autres formes de La vraisemblance partielle quand il y a des ex-æquos, sont proposées. La première dûe à Breslow (1974) est donnée par la formule

$$L_1(\beta) = \prod_{i=1}^L \frac{\exp \beta' S_{(i)}}{\sum_{j \in R_{(i)}} \exp \beta' Z_j}$$

La deuxième dûe à Efron (1977) est donnée par la formule

$$L_2(\beta) = \prod_{i=1}^L \frac{\exp \beta' S_{(i)}}{\prod_{j=1}^{D_i} \sum_{k \in R_{(i)}} \exp \beta' Z_k - \frac{j-1}{D_i} \sum_{k=1}^{D_i} \exp \beta' Z_k}$$

S'il n'y a pas d'ex-æquos, les 3 méthodes d'estimations donnent exactement les mêmes résultats et s'il y a peu d'ex-æquos les 3 méthodes ont des résultats assez proches.

Remarque 30

Dans le cas des covariables dépendant du temps la vraisemblance partielle dans (2.8) est donnée par la formule

$$L(\beta) = \prod_{i=1}^L \frac{\exp \beta' Z_{(i)}(T_{(i)})}{\sum_{j \in R_{(i)}} \exp \beta' Z_j(T_{(i)})}$$

qui s'écrit aussi

$$L(\beta) = \prod_{i=1}^L \left[\frac{\exp \beta' Z_i(T_i)}{\sum_{j \in R_i} \exp \beta' Z_j(T_i)} \right]^{\delta_i} \quad (2.11)$$

et la vraisemblance partielle dans (2.10) est donnée par la formule

$$L(\beta) = \prod_{i=1}^L \frac{\exp \beta' S_{(i)}(T_{(i)})}{\sum_{g \in \mathcal{G}_{(i)}} \exp \beta' S_g^*(T_{(i)})} \quad (2.12)$$

2.3 L'estimation de β

L'estimateur de Cox s'obtient en maximisant $L(\beta)$ ou $\mathcal{L}(\beta)$ le logarithme de $L(\beta)$. L'observation de $(X_i, \delta_i, Z_i)_{1 \leq i \leq n}$ nous permet d'observer $(N_i, Y_i, Z_i)_{1 \leq i \leq n}$ où

$$N_i(t) = 1_{\{X_i \leq t, \delta_i = 1\}}$$

est un processus de comptage,

$$Y_i(t) = 1_{\{X_i \geq t\}}$$

est l'indicateur de présence du sujet juste avant l'instant t et $Z_i(t)$ est le vecteur des covariables du i^{eme} individu. Nous supposons que Z_i est un processus prévisible et localement borné. Supposons que

- 1) $N = (N_1, N_2, \dots, N_n)$ est un processus de comptage multivarié.
- 2) le processus $N = (N_1, N_2, \dots, N_n)$ à pour processus d'intensité aléatoire $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ où

$$\lambda_i(t) = Y_i(t) \exp \beta_0' Z_i(t) \lambda_0(t) \quad (2.13)$$

i.e.

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(x) dx \quad t \geq 0. \quad (2.14)$$

est un martingale locale par rapport à la filtration

$$\mathcal{F}_{t,i} = \sigma \{N_i(u_+), Y_i(u_+), Z_i(u_+) , 0 \leq u \leq t\}$$

pour tout $1 \leq i \leq n$ où $X(u_+)$ est la limite à droite de X au point u .

En réalité M_i est un martingale locale de carré intégrable (c.f. Fleming p 61). De plus

1) pour $i = 1, 2, \dots, n$.

$$\langle M_i, M_i \rangle(t) = \int_0^t \lambda_i(x) dx \quad (2.15)$$

2) si $i \neq j$,

$$\langle M_i, M_j \rangle(t) = 0 \quad \text{ps} \quad (2.16)$$

(c.f. Fleming pg 75)

En effet cette forme donnée au compensateur peut découler, sous certaines conditions de régularité du fait que

$$\lambda_i(t) = \lim_{h \searrow 0} \Pr \{(N_i(t+h) - N_i(t) = 1 / \mathcal{F}_t)\}$$

(c.f Aalen 1978).

Pour β et t donnés , introduisons les processus

$$S^{(0)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n Y_i(t) \exp \beta' Z_i(t), \quad (2.17)$$

$S^{(1)}(\beta, t)$ et $E(\beta, t)$ les vecteurs colonnes de dimension p donnés par

$$S^{(1)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n Z_i(t) Y_i(t) \exp \beta' Z_i(t) \quad (2.18)$$

$$E(\beta, t) = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \quad (2.19)$$

$S^{(2)}(\beta, t)$ et $V(\beta, t)$ les matrices de dimension $p \times p$ données par

$$S^{(2)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n \{Z_i(t)\}^{\otimes 2} Y_i(t) \exp \beta' Z_i(t) \quad (2.20)$$

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \{E(\beta, t)\}^{\otimes 2}$$

(ou pour tout vecteur colonne X , $X^{\otimes 2} = XX'$).

Les quantités $E(\beta, t)$ et $V(\beta, t)$ peuvent être interprétées comme suit. Si à l'instant t nous sélectionnons un individu i parmi tous les individus sous observation (i.e. $Y_i(t) = 1$) alors $E(\beta, t)$ (resp $V(\beta, t)$) est la moyenne (resp covariance) du vecteur de covariables de l'individu sélectionné pondéré par les poids

$$\frac{Y_i(t) \exp \beta' Z_i(t)}{\sum_{i=1}^n Y_j(t) \exp \beta' Z_j(t)}$$

$V(\beta, t)$ s'écrit aussi

$$V(\beta, t) = \frac{n^{-1} \sum_{i=1}^n \{Z_i(t) - E(\beta, t)\}^{\otimes 2} Y_i(t) \exp \beta' Z_i(t)}{S^{(0)}(\beta, t)} \quad (2.21)$$

Le logarithme de vraisemblance partielle donnée par la formule(2.11) est

$$\mathcal{L}(\beta) = \sum_{i=1}^n \int_0^{\infty} \beta' Z_i(x) dN_i(x) - \int_0^{\infty} \log \left\{ \sum_{i=1}^n Y_i(x) \exp \beta' Z_i(x) \right\} d\bar{N}(x). \quad (2.22)$$

où $\bar{N} = \sum_{i=1}^n N_i$.

Le vecteur des dérivées partielles de $\mathcal{L}(\beta)$ par rapport aux composantes de β , ou vecteur des scores, noté $U(\beta)$ vaut

$$U(\beta) = \sum_{i=1}^n \int_0^{\infty} \{Z_i(x) - E(\beta, x)\} dN_i(x) \quad (2.23)$$

qui s'écrit aussi

$$U(\beta) = \sum_{i=1}^n \int_0^{\infty} \{Z_i(x) - E(\beta, x)\} dM_i(x) \quad (2.24)$$

la matrice des dérivées secondes est

$$\frac{\partial}{\partial \beta} U(\beta) = - \int_0^{\infty} \sum_{i=1}^n V(\beta, x) dN_i(x)$$

La matrice d'information est égale à moins l'espérance de la matrice des dérivées secondes.

On pose

$$\mathcal{I}(\beta) = - \frac{\partial}{\partial \beta} U(\beta) = \int_0^{\infty} \sum_{i=1}^n V(\beta, x) dN_i(x) \quad (2.25)$$

L'estimateur du maximum de vraisemblance partielle est obtenu en résolvant le système d'équations :

$$U(\beta) = 0$$

Il y a en tout p équations, une pour chacune des p variables $j = 1, 2, \dots, p$. En général, les solutions ne peuvent être obtenues que par itération (Newton Raphson). Nous verrons au chapitre 3 que la solution $\hat{\beta}$ est consistante et asymptotiquement normale avec pour moyenne β et pour matrice de variance-covariance $[E(\mathcal{I}(\beta))]^{-1}$.

La matrice de variance-covariance $[E(\mathcal{I}(\beta))]^{-1}$ elle aussi estimée par $(\mathcal{I}(\hat{\beta}))^{-1}$

2.4 Tests d'Hypothèses

2.4.1 Test global

$$H_0 : \beta = \beta_0$$

1) Test de Wald (maximum de vraisemblance)

Il est basé sur la statistique

$$\chi_{WV}^2 = (\hat{\beta} - \beta_0)' \mathcal{I}(\hat{\beta})(\hat{\beta} - \beta_0)$$

qui suit un χ^2 avec p degrés de liberté sous H_0 pour les grands échantillons.

2) Test de rapport de vraisemblance

Il utilise la statistique

$$\chi_{RV}^2 = 2 \left\{ \mathcal{L}(\hat{\beta}) - \mathcal{L}(\beta_0) \right\}$$

qui suit un χ^2 avec p degrés de liberté sous H_0 pour les grands échantillons.

3) Test score :

Il est fondé sur la statistique

$$\chi_{SC}^2 = U'(\beta_0) [\mathcal{I}(\beta_0)]^{-1} U(\beta_0)$$

qui suit un χ^2 avec p degrés de liberté sous H_0 pour les grands échantillons.

2.4.2 Test local

Souvent on trouve intéressant de tester une hypothèse au sujet d'une sous ensemble de β . L'hypothèse est :

$$H_0 : \beta_1 = \beta_{10}$$

où $\beta = (\beta_1', \beta_2')'$. Ici β_1 est un vecteur $q \times 1$ de β et β_2 est un vecteur $p - q \times 1$ de β .

1) Test de Wald : (maximum de vraisemblance)

Soit $\hat{\beta} = (\hat{\beta}'_1, \hat{\beta}'_2)'$ l'estimateur de maximum de vraisemblance partielle de β . Nous divisons la matrice d'information $\mathcal{I}(\beta_0)$ comme suit :

$$\mathcal{I}(\beta_0) = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}$$

où \mathcal{I}_{11} (resp \mathcal{I}_{22}) est la $q \times q$ (resp $(p - q) \times (p - q)$) sous matrice de $\mathcal{I}(\beta_0)$ associée à β_1 (resp β_2). Ce test se base sur la statistique

$$\chi_W^2 = (\hat{\beta}_1 - \beta_{10})' \left[\mathcal{I}^{11}(\hat{\beta}) \right]^{-1} (\hat{\beta}_1 - \beta_{10})$$

où $\mathcal{I}^{11}(\hat{\beta})$ est la matrice $q \times q$ de $\left[\mathcal{I}(\hat{\beta}) \right]^{-1}$. Pour les grands échantillons χ_W^2 suit un χ^2 avec q degrés de liberté sous H_0 .

2) Test de rapport de vraisemblance :

Soit $\hat{\beta}_2(\beta_{10})$ l'estimateur du maximum de vraisemblance partielle de β_2 avec β_1 fixé à la valeur β_{10} . La statistique de test est

$$\chi_{RV}^2 = 2 \left\{ \mathcal{L}(\hat{\beta}) - \mathcal{L} \left[\beta_{10}, \hat{\beta}_2(\beta_{10}) \right] \right\}$$

Pour les grands échantillons χ_{RV}^2 suit un χ^2 avec q degrés de liberté sous H_0 .

3) Test score :

Soit $U_1 \left[\beta_{10}, \hat{\beta}_2(\beta_{10}) \right]$ le $q \times 1$ vecteur de score de β_1 , évalué à la valeur supposée de β_{10} et à l'estimateur de maximum de vraisemblance partielle $\hat{\beta}_2$ de β_2 . Ce test se base sur la statistique

$$\chi_{SC}^2 = U_1' \left[\beta_{10}, \hat{\beta}_2(\beta_{10}) \right] \left[\mathcal{I}^{11}(\beta_{10}, \hat{\beta}_2(\beta_{10})) \right]^{-1} U_1 \left[\beta_{10}, \hat{\beta}_2(\beta_{10}) \right].$$

Pour les grands échantillons χ_{SC}^2 suit un χ^2 avec q degrés de liberté sous H_0 .

2.5 Estimation de la fonction de hasard λ_0

Pour déduire l'estimateur du taux de hasard de base λ_0 il est nécessaire d'estimer β , puis on maximise la vraisemblance totale par rapport à λ_0 . Pour construire la vraisemblance dans le cas des données censurées nous avons besoin de considérer quelle information nous donne chaque observation.

1) une observation complète donne l'information sur la probabilité que l'évènement a eu lieu à ce temps, qui est approximativement égal à la fonction de la densité prise en ce temps.

2) Pour une observation censurée à droite, tout ce que nous savons c'est que le temps de l'évènement est plus grand que ce temps. Donc l'information est la fonction de survie évaluée au temps de l'étude.

Construisons maintenant la vraisemblance totale pour le modèle de hasard proportionnel dans le cas où il n'y a pas d'ex-æquos et les covariables $(Z_i)_{i=1,2,\dots,n}$ sont constantes. Soit $X = (X_i, \delta_i, Z_i)_{1 \leq i \leq n}$ comme dans la sous section (2.2.1) supposons qu'il y ait L morts aux temps

$$t_{(1)} < t_{(2)} < \dots < t_{(L)}$$

donc la vraisemblance totale dans notre modèle est donnée par

$$\begin{aligned} L(\beta, \lambda_0) &= \prod_{i=1}^n (f(t_i/Z_i))^{\delta_i} S(t_i/Z_i)^{1-\delta_i} \\ &= \prod_{i=1}^n (\lambda(t_i/Z_i))^{\delta_i} \exp \{-\Lambda(t_i/Z_i)\} \\ &= \prod_{i=1}^n (\lambda_0(t_i))^{\delta_i} (\exp \beta' Z_i)^{\delta_i} \exp \{-\Lambda_0(t_i)(\exp \beta' Z_i)\} \\ &= \left[\prod_{i=1}^L \lambda_0(t_{(i)}) \exp \{\beta' Z_{(i)}\} \right] \exp \left\{ - \sum_{j=1}^n \Lambda_0(t_j) \exp \{\beta' Z_j\} \right\} \end{aligned}$$

Posons

$$\lambda_0(t_{(i)}) = \lambda_{0i}, \quad i = 1, 2, \dots, L.$$

On obtient

$$\Lambda_0(t_j) = \sum_{t_{(i)} \leq t_j} \lambda_{0i}.$$

Maximer $L(\beta, \lambda_0)$ revient à maximiser

$$L^*(\beta, \lambda_0) = \prod_{i=1}^L \lambda_{0i} \exp \left\{ -\lambda_{0i} \sum_{j \in R_{(i)}} \exp \{ \beta' Z_j \} \right\}$$

cette fonction est maximale quand

$$\lambda_{0i} = \frac{1}{\sum_{j \in R_{(i)}} \exp \{ \hat{\beta}' Z_j \}} \quad (2.26)$$

Ceci suggère d'estimer la fonction de hasard cumulée Λ_0 par $\hat{\Lambda}_0$, appelé l'estimateur de Breslow, défini par

$$\hat{\Lambda}_0(t) = \sum_{t_{(i)} \leq t} \frac{1}{\sum_{j \in R_{(i)}} \exp \{ \hat{\beta}' Z_j \}} \quad (2.27)$$

Cet estimateur se réduit à l'estimateur de Nelson Aalen quand il n'y a aucune covariable (cf Fleming).

On peut estimer S_0 par

$$\hat{S}_0(t) = \exp \hat{\Lambda}_0(t) \quad (2.28)$$

et l'estimateur de la fonction de survie pour un individu qui a pour vecteur de covariables Z_i est donné par

$$\hat{S}(t/Z_i) = \left[\hat{S}_0(t) \right]^{\exp \{ \hat{\beta}' Z_i \}} \quad (2.29)$$

Remarque 31

S'il y a des *ex-æquo* parmi les données, l'estimateur de λ_0 est donné par

$$\hat{\lambda}_0(t_{(i)}) = \frac{D_i}{\sum_{j \in R_{(i)}} \exp \left\{ \hat{\beta}' Z_j \right\}}$$

et l'estimateur de $\hat{\Lambda}_0$ est donné par

$$\hat{\Lambda}_0(t) = \sum_{t_{(i)} \leq t} \frac{D_i}{\sum_{j \in R_{(i)}} \exp \left\{ \hat{\beta}' Z_j \right\}} \quad (2.30)$$

Remarque 32

L'estimateur de taux de hasard cumulé Λ_0 dans le cas où les covariables dépendent du temps est donné par

$$\hat{\Lambda}_0(t) = \int_0^t \left\{ \sum_{i=1}^n Y_i(x) \exp \hat{\beta}' Z_i(x) \right\}^{-1} d\bar{N}(x) \quad (2.31)$$

où $\bar{N}(x) = \sum_{i=1}^n N_i(x)$ (cf Fleming Harrington pg 151).

2.6 Exemples

Exemple 33

Afin de mieux comprendre les différentes composantes de ce modèle, nous allons étudier un exemple tiré de la recherche clinique. Il s'agit de l'étude de la durée de survie de 90 patients atteints de la maladie du cancer du larynx. On possède un tableau avec les informations suivantes

- 1) temps : le nombre de mois de vie jusqu'à la mort ou la censure.
- 2) ind : indicateur de censure

3) L'étape de la maladie au moment du diagnostic (il existe 4 étapes).

On pose : L'étape 1 est l'étape de référence.

$Z_1 = 1$ si la maladie est à l'étape 2 et $Z_1 = 0$ sinon.

$Z_2 = 1$ si la maladie est à l'étape 3 et $Z_2 = 0$ sinon.

$Z_3 = 1$ si la maladie est à l'étape 4 et $Z_3 = 0$ sinon.

4) âge : est l'âge au moment du diagnostic, noté Z_4 .

Le tableau est le suivant

	<i>temps</i>	<i>ind</i>	<i>étap2</i>	<i>étap3</i>	<i>étap4</i>	<i>âge</i>
1	0.6	1	0	0	0	77
2	1.3	1	0	0	0	53
3	2.4	1	0	0	0	45
4	3.2	1	0	0	0	58
5	3.3	1	0	0	0	76
6	3.5	1	0	0	0	43
7	3.5	1	0	0	0	60
8	4	1	0	0	0	52
9	4	1	0	0	0	63
10	4.3	1	0	0	0	86
11	5.3	1	0	0	0	81
12	6	1	0	0	0	75
13	6.4	1	0	0	0	77
14	6.5	1	0	0	0	77
15	7.4	1	0	0	0	67
16	2.5	0	0	0	0	57
17	3.2	0	0	0	0	51
18	3.3	0	0	0	0	63

	<i>temps</i>	<i>ind</i>	<i>étap2</i>	<i>étap3</i>	<i>étap4</i>	<i>âge</i>
19	4.5	0	0	0	0	48
20	4.5	0	0	0	0	68
21	5.5	0	0	0	0	70
22	5.9	0	0	0	0	47
23	5.9	0	0	0	0	58
24	6.1	0	0	0	0	77
25	6.2	0	0	0	0	64
26	6.5	0	0	0	0	79
27	6.7	0	0	0	0	61
28	7	0	0	0	0	66
29	7.4	0	0	0	0	73
30	8.1	0	0	0	0	56
31	8.1	0	0	0	0	73
32	9.6	0	0	0	0	58
33	10.7	0	0	0	0	68
34	0.2	1	1	0	0	86
35	1.8	1	1	0	0	64
36	2	1	1	0	0	63
37	3.6	1	1	0	0	70
38	4	1	1	0	0	81
39	6.2	1	1	0	0	74
40	7	1	1	0	0	62
41	2.2	0	1	0	0	71
42	2.6	0	1	0	0	67
43	3.3	0	1	0	0	51
44	3.6	0	1	0	0	72

	<i>temps</i>	<i>ind</i>	<i>étap2</i>	<i>étap3</i>	<i>étap4</i>	<i>âge</i>
45	4.3	0	1	0	0	47
46	4.3	0	1	0	0	64
47	5	0	1	0	0	66
48	7.5	0	1	0	0	50
49	7.6	0	1	0	0	53
50	9.3	0	1	0	0	61
51	0.3	1	0	1	0	49
52	0.3	1	0	1	0	71
53	0.5	1	0	1	0	57
54	0.7	1	0	1	0	79
55	0.8	1	0	1	0	82
56	1	1	0	1	0	49
57	1.3	1	0	1	0	60
58	1.6	1	0	1	0	64
59	1.8	1	0	1	0	74
60	1.9	1	0	1	0	53
61	1.9	1	0	1	0	72
62	3.2	1	0	1	0	54
63	3.5	1	0	1	0	81
64	5	1	0	1	0	59
65	6.3	1	0	1	0	70
66	6.4	1	0	1	0	65
67	7.8	1	0	1	0	68
68	3.7	0	0	1	0	52
69	4.5	0	0	1	0	66
70	4.8	0	0	1	0	45

	<i>temps</i>	<i>ind</i>	<i>étap2</i>	<i>étap3</i>	<i>étap4</i>	<i>âge</i>
71	4.8	0	0	1	0	63
72	5	0	0	1	0	49
73	5.1	0	0	1	0	69
74	6.5	0	0	1	0	65
75	8	0	0	1	0	78
76	9.3	0	0	1	0	69
77	10.1	0	0	1	0	51
78	0.1	1	0	0	1	65
79	0.3	1	0	0	1	71
80	0.4	1	0	0	1	76
81	0.8	1	0	0	1	65
82	0.8	1	0	0	1	78
83	1	1	0	0	1	41
84	1.5	1	0	0	1	68
85	2	1	0	0	1	69
86	2.3	1	0	0	1	62
87	3.6	1	0	0	1	71
88	3.8	1	0	0	1	84
89	2.9	0	0	0	1	74
90	4.3	0	0	0	1	48

Nous avons donc le modèle suivant

$$\lambda(t) = \lambda_0(t) \exp \{ \beta_1 \times \text{étap2} + \beta_2 \times \text{étap3} + \beta_3 \times \text{étap4} + \beta_4 \times \text{âge} \}$$

ou

$$S(t) = (S_0(t))^{\exp \{ \beta_1 \times \text{étap2} + \beta_2 \times \text{étap3} + \beta_3 \times \text{étap4} + \beta_4 \times \text{âge} \}} .$$

1) L'estimation des paramètres

La fonction "Coxph" dans le logiciel S-Plus nous permet d'estimer le vecteur du paramètre $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ et nous obtenons les résultats suivants

```
Working data will be in C:\Program Files\sp2000\users\dounia\_Data
> cancl<-coxph(Surv(temps, ind) ~ etap2 + etap3 + etap4 + age, data = cancer)
> summary(cancl)
Call:
coxph(formula = Surv(temps, ind) ~ etap2 + etap3 + etap4 + age, data = cancer)

n= 90

      coef exp(coef) se(coef)      z      p
etap2 0.140      1.15  0.4625 0.302 0.760000
etap3 0.642      1.90  0.3561 1.803 0.071000
etap4 1.706      5.51  0.4219 4.044 0.000052
age 0.019      1.02  0.0143 1.335 0.180000

      exp(coef) exp(-coef) lower .95 upper .95
etap2      1.15      0.869      0.465      2.85
etap3      1.90      0.526      0.946      3.82
etap4      5.51      0.182      2.409     12.59
age      1.02      0.981      0.991      1.05
```

D'où

$$\hat{\beta} = (0.140, 0.642, 1.706, 0.019).$$

2) Tests d'Hypothèses.

a) Test global

$H_0 : \beta = 0$ contre $H_1 : \beta \neq 0$ et on prend le seuil $\alpha = 0.05$. On a les résultats suivants

```
Likelihood ratio test= 18.3 on 4 df, p=0.00107
Wald test              = 21.2 on 4 df, p=0.000295
Score (logrank) test = 24.8 on 4 df, p=0.0000557
```

Le vecteur des coefficients de régression est

$$\hat{\beta} = (0.140, 0.642, 1.706, 0.019),$$

avec un niveau de significativité (test de rapport de vraisemblance) de 0,00107. Le vecteur des coefficients est donc significativement différent de 0. Le résultat est le même si nous utilisons le test de Wald ou le test de Score.

b) Test local

Nous testons, par exemple, l'hypothèse

$$H_0 : (\beta_1, \beta_2, \beta_3) = 0,$$

contre

$$H_1 : (\beta_1, \beta_2, \beta_3) \neq 0$$

et nous prenons le seuil $\alpha = 0.05$.

Test de rapport de vraisemblance

```
> canc1r<-coxph(Surv(temps, ind) ~ age, data = cancer)
> canc1$loglik
[1] -196.8635 -187.7074
> canc1r$loglik
[1] -196.8635 -195.5478
> 2*(-187.707+195.548)
[1] 15.682
> pchisq(15.682,df=3)
[1] 0.9986824
> 1-0.9987
[1] 0.0013
```

Donc

$$\chi_{RV}^2 = 2 \left\{ \mathcal{L}(\hat{\beta}) - \mathcal{L}(\beta_0) \right\} = 15.682$$

avec un niveau de significativité de 0,0013, d'où $(\beta_1, \beta_2, \beta_3) \neq 0$. le résultat est le même si nous utilisons le test de Wald ou le test de Score.

Estimation de la fonction de covariance

```
> canc1$var
      [,1]      [,2]      [,3]      [,4]
[1,] 0.2139048730 0.0683665643 0.068916396 0.0008240989
[2,] 0.0683665643 0.1268209256 0.068075539 0.0003271962
[3,] 0.0689163963 0.0680755394 0.177987411 -0.0003914110
[4,] 0.0008240989 0.0003271962 -0.000391411 0.0002032944
```

3) Estimation de la fonction de survie de base S_0 .

La fonction (survfit) dans le logiciel S-Plus nous permet d'estimer la fonction de survie de base et nous avons les résultats suivants

```

> summary(survfit(canc1))
Call: survfit.coxph(object = canc1)

   time  n.risk  n.event  survival  std.err  lower 95% CI  upper 95% CI
0.1      90       1      0.991 0.00875      0.974      1.000
0.2      89       1      0.982 0.01252      0.958      1.000
0.3      88       3      0.955 0.02009      0.917      0.995
0.4      85       1      0.946 0.02212      0.904      0.990
0.5      84       1      0.937 0.02406      0.891      0.985
0.6      83       1      0.927 0.02585      0.878      0.979
0.7      82       1      0.918 0.02751      0.865      0.973
0.8      81       3      0.888 0.03213      0.827      0.954
1.0      78       2      0.868 0.03499      0.802      0.939
1.3      76       2      0.847 0.03758      0.776      0.924
1.5      74       1      0.837 0.03877      0.764      0.916
1.6      73       1      0.826 0.03995      0.751      0.908
1.8      72       2      0.804 0.04216      0.726      0.891
1.9      70       2      0.783 0.04418      0.701      0.874
2.0      68       2      0.761 0.04604      0.676      0.857
2.3      65       1      0.750 0.04696      0.663      0.848
2.4      64       1      0.738 0.04786      0.650      0.838
3.2      60       2      0.713 0.04975      0.622      0.817
3.3      57       1      0.700 0.05063      0.607      0.807
3.5      54       3      0.661 0.05308      0.565      0.774
3.6      51       2      0.635 0.05444      0.537      0.751
3.8      47       1      0.621 0.05516      0.522      0.739
4.0      46       3      0.575 0.05699      0.473      0.698

4.3      43       1      0.560 0.05743      0.458      0.684
5.0      34       1      0.540 0.05859      0.437      0.668
5.3      30       1      0.518 0.06012      0.413      0.651
6.0      26       1      0.495 0.06182      0.388      0.632
6.2      24       1      0.471 0.06339      0.362      0.613
6.3      22       1      0.446 0.06481      0.336      0.593
6.4      21       2      0.396 0.06676      0.285      0.551
6.5      19       1      0.371 0.06721      0.260      0.529
7.0      15       1      0.342 0.06818      0.231      0.505
7.4      13       1      0.311 0.06880      0.202      0.480
7.8       9       1      0.276 0.07016      0.167      0.454

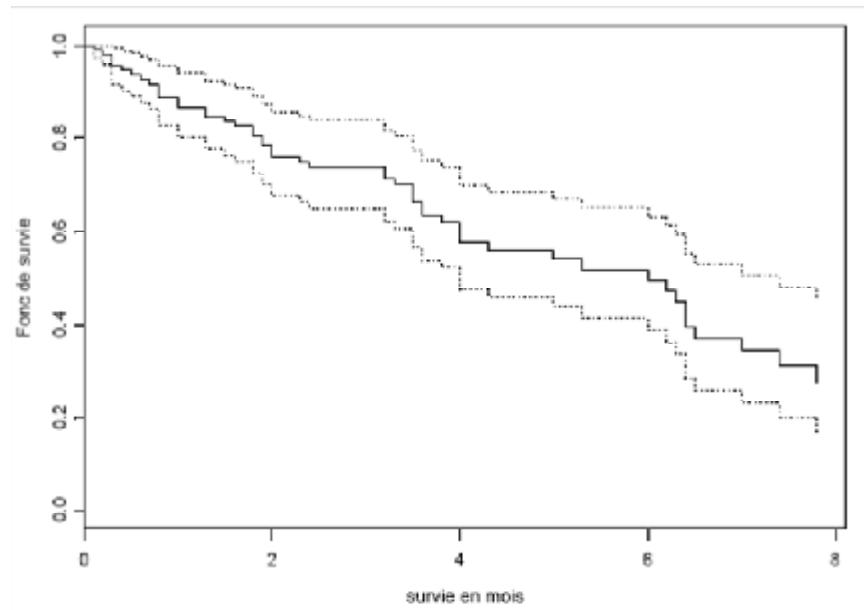
```

La figure suivante donne la courbe de l'estimateur de la fonction de survie de base S_0

```

> plot(survfit(canc1),xlab="survie en mois",ylab="Fonc de survie")

```



Maintenant, si nous considérons l'équation (2.2), la fonction de survie de base $S_0(t)$, et les coefficients de régression, nous pouvons estimer la fonction de survie pour un individu quelconque.

Exemple 34

Nous allons étudier un second exemple. Il s'agit de l'étude de la durée de survie de 26 patients atteints de la maladie du cancer. On a les données suivantes

	Temps	Ind	âge	Tr
1	59	1	72.33	1
2	115	1	74.49	1
3	156	1	66.47	1
4	421	0	53.36	2
5	431	1	50.34	1
6	448	0	56.43	1
7	464	1	56.94	2

	<i>Temps</i>	<i>Ind</i>	<i>âge</i>	<i>Tr</i>
8	475	1	59.85	2
9	477	0	64.18	1
10	563	1	55.18	2
11	638	1	56.76	1
12	744	0	50.11	2
13	769	0	59.63	2
14	770	0	57.05	2
15	803	0	39.27	1
16	855	0	43.12	1
17	1040	0	38.89	1
18	1106	0	44.60	1
19	1129	0	53.91	2
20	1206	0	44.21	2
21	1227	0	59.59	2
22	268	1	74.50	1
23	329	1	43.14	1
24	353	1	63.22	2
25	365	1	64.42	2
26	377	0	58.31	2

où

1) *Temps* est le nombre de jours du début du traitement jusqu'à la mort ou la censure.

2) *Ind* est l'indicateur de censure

3) *âge* est l'âge au moment du diagnostic.

4) *Tr* est un indicateur du traitement donné au malade.

Nous avons donc le modèle suivant

$$\lambda(t) = \lambda_0(t) \exp \{ \beta_1 \times \text{age} + \beta_2 \text{Tr} \}$$

ou

$$S(t) = (S_0(t))^{\exp\{\beta_1 \times \text{age} + \beta_2 \text{Tr}\}}$$

1) L'estimation des paramètres

$\beta = (\beta_1, \beta_2)$ et nous obtenons les résultats suivants

```
Working data will be in C:\Program Files\sp2000\users\d
ounia\_Data
> car<-coxph(Surv(Temps, Ind) ~ age + Tr,cancer5)
> summary(car)
Call:
coxph(formula = Surv(Temps, Ind) ~ age + Tr, data = cancer5)

n= 26

      coef exp(coef) se(coef)      z      p
age  0.147    1.159   0.0461   3.19 0.0014
Tr -0.804    0.448   0.6320  -1.27 0.2000

      exp(coef) exp(-coef) lower .95 upper .95
age    1.159    0.863    1.06    1.27
Tr    0.448    2.234    0.13    1.54
```

D'où $\beta = (0.147, -0.448)$

2) Test d'hypothèses

a) Test global

$H_0 : \beta = 0$ contre $H_1 : \beta \neq 0$ et en prend le seuil $\alpha = 0.05$.

On a les résultats suivants

```
Likelihood ratio test= 15.9 on 2 df, p=0.000355
Wald test              = 13.5 on 2 df, p=0.00119
Score (logrank) test = 18.6 on 2 df, p=0.0000934
```

Le coefficient de régression est $\hat{\beta} = (0.147, -0.448)$, avec un niveau de significativité (test de rapport de vraisemblance) de 0,000355. Le vecteur des coefficients est donc significativement différent de 0. le résultat est le même si nous utilisons le test de Wald ou le test de Scor.

b) Test local.

Si nous nous intéressons à tester l'effet du traitement sur la fonction de hasard de base λ_0 nous obtenons les résultats suivants (nous utilisons par exemple le test de rapport de

vraisemblance).

```
> car1<-coxph(Surv(Temps, Ind) ~ age,cancer5)
> car$loglik
[1] -34.98494 -27.04190
> car1$loglik
[1] -34.98494 -27.83815
> 2*(-27.042+27.838)
[1] 1.592
> pchisq(1.592,df=1)
[1] 0.7929594
> 1-0.793
[1] 0.207
```

Donc

$$\chi_{RV}^2 = 2 \left\{ \mathcal{L}(\hat{\beta}) - \mathcal{L}(\beta_0) \right\} = 1.592$$

avec un niveau de significativité de 0.207, alors on accepte H_0 donc le traitement n'a pas d'effet multiplicatif sur la durée de survie.

3) Estimation de la fonction de covariance

```
> car$var
      [,1] [,2]
[1,] 0.002129550 0.004696922
[2,] 0.004696922 0.399486402
```

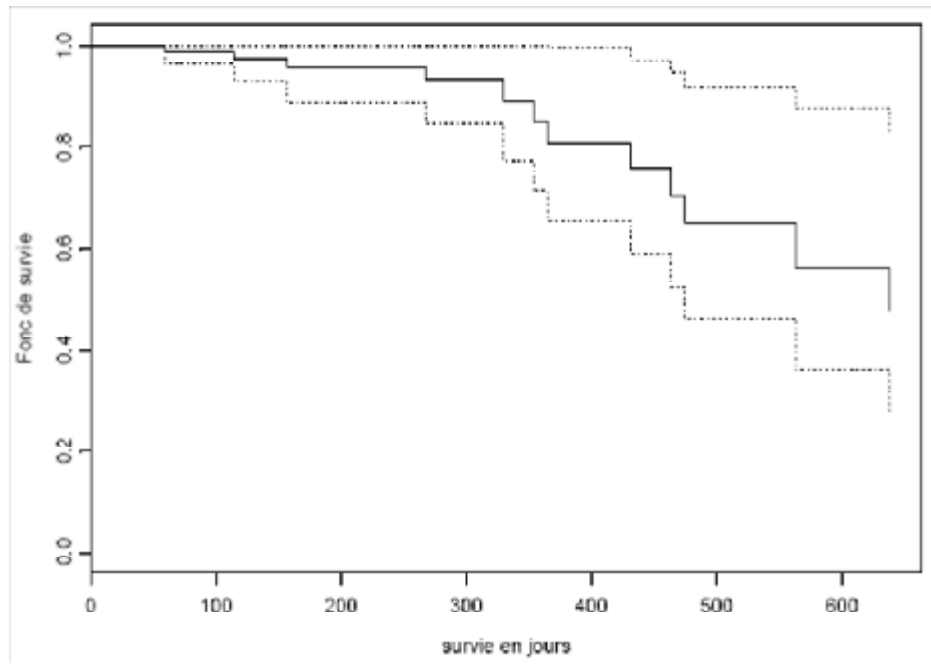
4) Estimation de la fonction de survie de base S_0 .

```
> summary(survfit(car))
Call: survfit.coxph(object = car)

   time  n.risk  n.event  survival  std.err  lower 95% CI  upper 95% CI
   59      26       1      0.989  0.0132    0.963    1.000
  115      25       1      0.975  0.0229    0.931    1.000
  156      24       1      0.956  0.0351    0.890    1.000
  268      23       1      0.935  0.0474    0.846    1.000
  329      22       1      0.892  0.0644    0.774    1.000
  353      21       1      0.851  0.0763    0.714    1.000
  365      20       1      0.808  0.0883    0.656    0.996
  431      17       1      0.759  0.0965    0.591    0.973
  464      15       1      0.705  0.1066    0.524    0.948
  475      14       1      0.652  0.1140    0.463    0.918
  563      12       1      0.563  0.1277    0.361    0.878
  638      11       1      0.479  0.1328    0.278    0.825
```

La figure suivante donne la courbe de l'estimateur de la fonction de survie de base S_0

```
> plot(survfit(car),xlab="survie en jours",ylab="Fonc de survie")
```



Chapitre 3

Propriétés asymptotiques

Nous allons nous intéresser aux propriétés asymptotiques de l'EMVP $\hat{\beta}$ de β et $\hat{\Lambda}_0$ de Λ_0 . Andersen et Gill (1982) ont montré que $\hat{\beta}$ est un estimateur consistant et asymptotiquement normal et que $\hat{\Lambda}$ est asymptotiquement normal en utilisant l'approche par les martingales.

3.1 Rappels

Notation 35

Pour chaque scalaire y soit $\|y\| = |y|$ et pour chaque matrice M soit

$$\|M\| = \max_{i,j} |M_{(i,j)}|.$$

Pour $Y = (y_1, \dots, y_p)$ de \mathbb{R}^p ,

$$|Y| = \left(\sum_{i=1}^p y_i^2 \right)^{\frac{1}{2}}$$

(la norme euclidienne).

Si $a = (a_1, \dots, a_p)$ et $b = (b_1, \dots, b_p)$ on pose

$$[a, b] = \left\{ x = (x_1, \dots, x_p) \in \mathbb{R}^p : a_i \leq x_i \leq b_i, i = 1, 2, \dots, p \right\}$$

Définition 36

Soit $\{f_\alpha : \alpha \in A\}$ une famille de fonctions réelles sur un intervalle $[a, b]$ de \mathbb{R}^p . la famille $(f_\alpha)_{\alpha \in A}$ est dite équi continue au point $x \in [a, b]$ si

$$\forall \epsilon > 0, \exists \delta > 0, |x - y| < \delta \implies |f_\alpha(x) - f_\alpha(y)| < \epsilon$$

pour tout $\alpha \in A$. La famille est dite équi continue sur $[a, b]$ si elle est équi continue en chaque point x de $[a, b]$.

Théorème 37 (Application du théorème central limite de Rebolledo sur les martingales locales de carré intégrable)

Pour chaque $n \in \mathbb{N}^*$ soit $N^{(n)}$ un processus de comptage multivarié à n composantes. i.e.

$$\forall i \neq j, \forall t \geq 0, P \{ \Delta N_i(t) = \Delta N_j(t) = 1 \} = 0$$

Soit $H^{(n)}$ une matrice $n \times p$ (p fixé) telle que chaque élément $H_{i,l}^{(n)}$ est un processus prévisible localement borné. Supposons que $N^{(n)}$ a un processus d'intensité $\lambda^{(n)}$. Définissons la martingale locale de carré intégrable

$$W^{(n)} = (W_1^{(n)}, \dots, W_p^{(n)})$$

par

$$W_i^{(n)}(t) = \int_0^t \sum_{l=1}^n H_{i,l}^{(n)}(x) \left\{ dN_l^{(n)}(x) - \lambda_l^{(n)}(x) dx \right\}, i = 1, 2, \dots, p.$$

Soit A une matrice de dimension $p \times p$ de fonctions continues sur $[0, \tau]$ ayant la forme de la fonction de covariance d'une martingale gaussienne $W^{(\infty)}$ avec $W^{(\infty)}(0) = 0$. i.e.

$$\text{Cov}(W_i^{(\infty)}(t), W_j^{(\infty)}(x)) = A_{ij}(t \wedge x)$$

pour tout i, j, t, x . Supposons que

1)

$$\langle W_i^{(n)}, W_j^{(n)} \rangle (t) = \int_0^t \sum_{l=1}^n H_{i,l}^{(n)}(x) H_{j,l}^{(n)}(x) \lambda_l^{(n)}(x) dx \xrightarrow[n \rightarrow \infty]{P} A_{i,j}(t) \quad (\text{C1})$$

2) $\forall i, 1 \leq i \leq p$ et $\forall \epsilon > 0$

$$\int_0^t \sum_{l=1}^n \left(H_{i,l}^{(n)}(x) \right)^2 \lambda_l^{(n)}(x) 1_{\{|H_{i,l}^{(n)}(x)| > \epsilon\}} dx \xrightarrow[n \rightarrow \infty]{P} 0. \quad (\text{C2})$$

alors

$$W^{(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} W^{(\infty)}$$

Dans $D([0, 1]^p)$.

Lemme 38 (Inégalité de Langlart)

soit N un processus de comptage univarié de compensateur continu A , et soit

$$M = N - A$$

et H un processus prévisible localement borné.

alors $\forall \delta > 0, \forall \rho > 0$ et $\forall t \geq 0$ on a

1)

$$P \{ N(t) \geq \rho \} \leq \frac{\delta}{\rho} + P \{ A(t) \geq \delta \}.$$

2)

$$P \left\{ \sup_{0 \leq y \leq t} \left| \int_0^y H(x) dM(x) \right| \geq \rho \right\} \leq \frac{\delta}{\rho^2} + P \left\{ \int_0^t H^2(x) dA(x) \geq \delta \right\}.$$

Preuve.

2) Dans le corollaire (27), on prend $\eta = \delta$ et $\rho = \sqrt{\epsilon}$ et le temps d'arrêt $T = t$ on trouve le résultat (2).

1) soit (τ_n) une suite localisante telle que

$$\forall n, N(\cdot \wedge \tau_n), A(\cdot \wedge \tau_n) \text{ et } H(\cdot \wedge \tau_n)$$

soient bornées et tel que $\forall n$, $M(\cdot \wedge \tau_n)$ est une martingale locale de carré intégrable.

D'après le théorème d'arrêt optionnel pour tout temps d'arrêt fini T on a

$$E(N(T \wedge \tau_n)) = E(A(T \wedge \tau_n))$$

on pose

$$X(\cdot) = N(\cdot \wedge \tau_n)$$

et

$$Y(\cdot) = A(\cdot \wedge \tau_n)$$

et en appliquant le théorème (26) on trouve pour tout n

$$P \left\{ \sup_{s \leq T} N(s \wedge \tau_n) \geq \rho \right\} \leq \frac{\delta}{\rho} + P \{A(T) \geq \delta\}$$

Or $\tau_n \nearrow \infty$ p.s et $\Pr \left\{ \sup_{s \leq T} N(s \wedge \tau_n) \geq \rho \right\}$ est croissant en n donc

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{s \leq T} N(s \wedge \tau_n) \geq \rho \right\} = P \left\{ \sup_{s \leq T} N(s) \geq \rho \right\}$$

d'où

$$P \left\{ \sup_{s \leq T} N(s) \geq \rho \right\} \leq \frac{\delta}{\rho} + P \{A(t) \geq \delta\}.$$

En prenant $T = t$ on trouve

$$P \{N(t) \geq \rho\} \leq \frac{\delta}{\rho} + \Pr \{A(t) \geq \delta\}$$

■

3.2 constance de $\hat{\beta}$

Nous commençons par donner des hypothèses imposées dans toute la suite de ce chapitre.

Hypothèses **R**

R₁ : Soit τ un temps tel que

$$\int_0^\tau \lambda_0(x) dx < \infty$$

R₂ : Il existe un voisinage B de β_0 , un réel $s^{(0)}(\beta, t)$, un vecteur $s^{(1)}(\beta, t)$ et une matrice $s^{(2)}(\beta, t)$ définis sur $B \times [0, \tau]$ tels que : pour $j = 0, 1, 2$ on ait

$$\sup_{\substack{x \in [0, \tau] \\ \beta \in B}} \|S^{(j)}(\beta, x) - s^{(j)}(\beta, x)\| \xrightarrow[n \rightarrow \infty]{P} 0$$

R₃ : Il existe $\delta > 0$ tel que

$$n^{-\frac{1}{2}} \sup_{\substack{x \in [0, \tau] \\ 1 \leq i \leq n}} |Z_i(x)| Y_i(x) \mathbf{1}_{\{\beta' Z_i(x) > -\delta |Z_i(x)|\}} \xrightarrow[n \rightarrow \infty]{P} 0$$

R₄ : Pour les fonctions $s^{(j)}$, $j = 0, 1, 2$ définies dans **R₂** nous définissons :

$e = \frac{s^{(1)}}{s^{(0)}}$ et $v = \frac{s^{(2)}}{s^{(0)}} - e^{\otimes 2}$ alors : $\forall \beta \in B$ et $\forall t \in [0, \tau]$ on a

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, t)$$

et

$$s^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta^2} s^{(0)}(\beta, t).$$

R₅ : Les fonctions $s^{(j)}$, $j = 0, 1, 2$ sont bornées sur $\beta \times [0, \tau]$.

Il existe $a > 0$ tel que la fonction $t \rightarrow s^{(0)}(\beta, t)$ est bornée sur $\beta \times [a, \tau]$.

Pour $j = 0, 1, 2$ la famille des fonctions $s^{(j)}(\cdot, t)$, $0 \leq t \leq \tau$ est équi continue en β_0 .

R₆ : La matrice

$$\Sigma(\beta_0; \tau) = \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt$$

est définie positive.

Comme nous ne connaissons pas l'expression explicite de $\hat{\beta}$, c'est le lemme suivant donné en appendice de l'article de Andersen et Gill (1982), qui nous servira à montrer la consistance de $\hat{\beta}$.

Lemme 39

Soit E un ensemble convexe et ouvert de \mathbb{R}^p , et soit $(F_n)_{n \geq 1}$ une suite de fonctions aléatoires concaves sur E et f une fonction réelle sur E telle que

$$\forall x \in E, F_n(x) \xrightarrow[n \rightarrow \infty]{P} f(x)$$

alors

- 1) la fonction f est concave.
- 2) pour tout compact A de E

$$\sup_{x \in A} |F_n(x) - f(x)| \xrightarrow[n \rightarrow \infty]{P} 0$$

- 3) Si F_n a un seul maximum en X_n et f a un seul maximum en X , alors

$$X_n \xrightarrow[n \rightarrow \infty]{P} X$$

Théorème 40

Soit $\hat{\beta}$ l'EMVP de β et β_0 la vraie valeur de β dans le modèle de Cox. Alors

$$\hat{\beta} \xrightarrow[n \rightarrow \infty]{P} \beta_0$$

Preuve.

Soit le processus $X_n(\beta, \cdot)$ défini par

$$\begin{aligned} & X_n(\beta, t) \\ &= n^{-1} \{ \mathcal{L}(\beta, t) - \mathcal{L}(\beta_0, t) \} \\ &= n^{-1} \left[\sum_{i=1}^n \int_0^t (\beta - \beta_0)' Z_i(x) dN_i(x) - \int_0^t \log \left\{ \frac{\sum_{i=1}^n Y_i(x) \exp(\beta' Z_i(x))}{\sum_{i=1}^n Y_i(x) \exp(\beta_0' Z_i(x))} \right\} d\bar{N}(x) \right]. \end{aligned}$$

On pose

$$A_n(\beta, t) = n^{-1} \left[\sum_{i=1}^n \int_0^t (\beta - \beta_0)' Z_i(x) \lambda_i(x) dx - \int_0^t \log \left\{ \frac{S^{(0)}(\beta, x)}{S^{(0)}(\beta_0, x)} \right\} \bar{\lambda}(x) dx \right],$$

où $\bar{\lambda} = \sum_{i=1}^n \lambda_i$. Alors

$$\begin{aligned} & X_n(\beta, t) - A_n(\beta, t) \\ &= n^{-1} \left[\sum_{i=1}^n \int_0^t \left\{ (\beta - \beta_0)' Z_i(x) - \log \left\{ \frac{S^{(0)}(\beta, t)}{S^{(0)}(\beta_0, t)} \right\} \right\} dM_i(x) \right]. \end{aligned}$$

Le processus $X_n(\beta, \cdot) - A_n(\beta, \cdot)$ est une martingale locale de carré intégrable de processus de variation quadratique

$$\begin{aligned} & \langle X_n(\beta, \cdot) - A_n(\beta, \cdot) - X_n(\beta, \cdot) - A_n(\beta, \cdot) \rangle (t) \\ &= n^{-2} \sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)' Z_i(x) - \log \left\{ \frac{S^{(0)}(\beta, t)}{S^{(0)}(\beta_0, t)} \right\} \right]^2 \lambda_i(x) dx \\ &= n^{-1} \int_0^t [(\beta - \beta_0)' S^{(2)}(\beta_0, x) (\beta - \beta_0) - 2(\beta - \beta_0)' S^{(1)}(\beta_0, x) \log \left\{ \frac{S(\beta, x)}{S^{(0)}(\beta_0, x)} \right\} \\ & \quad + \left[\log \left\{ \frac{S^{(0)}(\beta, x)}{S^{(0)}(\beta_0, x)} \right\} \right]^2 S^{(0)}(\beta_0, x)] \lambda_0(x) dx. \end{aligned}$$

D'après les hypothèses R_1, R_2 et R_5 on a

$$n \langle X_n(\beta, \cdot) - A_n(\beta, \cdot), X_n(\beta, \cdot) - A_n(\beta, \cdot) \rangle (\tau)$$

converge en probabilité vers une limite finie et d'après le corollaire (28) on obtient :

$$X_n(\beta, \tau) - A_n(\beta, \tau) \xrightarrow[n \rightarrow \infty]{P} 0$$

mais $A_n(\beta, \tau)$ converge en probabilité vers $A(\beta, \tau)$ donnée par la formule

$$A(\beta, \tau) = \int_0^\tau \left[(\beta - \beta_0)' s^{(1)}(\beta_0, x) - \log \left\{ \frac{s^{(0)}(\beta, x)}{s^{(0)}(\beta_0, x)} \right\} s^{(0)}(\beta_0, x) \right] \lambda_0(x) dx.$$

D'où $\forall \beta \in B$, $X_n(\beta, \tau)$ converge en probabilité vers la même limite .

Or $X_n(\beta, \tau)$ est une fonction concave avec un seul maximum en $\hat{\beta}$ et $A(\beta, \tau)$ a un seul maximum en β_0 . Le lemme (39) donne

$$\hat{\beta} \xrightarrow[n \rightarrow \infty]{P} \beta_0$$

■

3.3 Normalité asymptotique

Théorème 41

Sous les hypothèses R_1, \dots, R_6 on a

1) *Le processus*

$$n^{-\frac{1}{2}}U(\beta_0, \cdot) = \left\{ n^{-\frac{1}{2}}U(\beta_0, t) : 0 \leq t \leq \tau \right\}$$

où

$$n^{-\frac{1}{2}}U(\beta_0, t) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t \{Z_i(x) - E(\beta_0, x)\} dN_i(x)$$

converge en loi dans $D([0, \tau]^p)$ vers $\mathcal{N}(0, \Sigma(\beta_0, t))$ avec

$$\{\Sigma(\beta_0, t)\}_{il} = \int_0^t \{v(\beta_0, x)\}_{il} s^{(0)}(\beta_0, x) \lambda_0(x) dx.$$

2) Soit $\hat{\beta}$ L'EMVP. Alors

$$\sup_{0 \leq t \leq \tau} \left\| \frac{1}{n} \int_0^t \sum_{i=1}^n V(\hat{\beta}, x) dN_i(x) - \Sigma(\beta_0, t) \right\| \xrightarrow[n \rightarrow \infty]{P} 0$$

Preuve.

1)

$$U(\beta_0, t) = \sum_{i=1}^n \int_0^t \{Z_i(x) - E(\beta_0, x)\} dM_i(x).$$

La composante l de $U(\beta_0, t)$ est donnée par

$$U_l(\beta_0, t) = \sum_{i=1}^n \int_0^t \{Z_{il}(x) - E_l(\beta_0, t)\} dM_i(x),$$

où

$$E_l(\beta_0, t) = \frac{\sum_{i=1}^n Z_{il}(t) Y_i(t) \exp\{\beta'_0 Z_i(t)\}}{\sum_{i=1}^n Y_i(t) \exp\{\beta'_0 Z_i(t)\}}.$$

On pose

$$H_{il}(x) = Z_{il}(x) - E_l(\beta_0, x), \quad l = 1, 2, \dots, p.$$

On obtient

$$U_l(\beta_0, t) = \sum_{i=1}^n \int_0^t H_{il}(x) dM_i(x)$$

et

$$U(\beta_0, t) = \sum_{i=1}^n \int_0^t H_i(x) dM_i(x).$$

alors les fonctions H_{il} vérifient les conditions de théorème (37).

Il reste à vérifier les conditions (C1) et (C2) du théorème (37). Soit

$$U^{(n)}(.,.) = n^{-\frac{1}{2}}U(.,.).$$

Pour vérifier la condition (C1) notons que

$$\begin{aligned} \left\langle U_l^{(n)}(\beta_0, .), U_l^{(n)}(\beta_0, .) \right\rangle(t) &= n^{-1} \int_0^t \sum_{i=1}^n H_{il}(x) H_{il}(x) \lambda_i(x) dx \\ &= \left\{ \int_0^t V(\beta_0, x) S^{(0)}(\beta_0, x) \lambda_0(x) dx \right\}_W \end{aligned}$$

qui converge en probabilité vers

$$\left\{ \int_0^t v(\beta_0, x) s^{(0)}(\beta_0, x) \lambda_0(x) dx \right\}_W$$

d'après les hypothèses R_1, R_2 et R_5 , où $M_{ll'}$ est l'élément (l, l') de la matrice M .

Vérifions maintenant la condition (C2). Posons $H_i^{(n)} = n^{-\frac{1}{2}} H_i$ alors

$$U_l^{(n)}(\beta_0, t) = \sum_{i=1}^n \int_0^t H_{il}^{(n)}(x) dM_i(x)$$

et

$$U_{l\epsilon}^{(n)}(\beta_0, t) = \sum_{i=1}^n \int_0^t H_{il}^{(n)}(x) \mathbf{1}_{\{|H_{il}^{(n)}(x)| \geq \epsilon\}} dM_i(x).$$

D'où

$$\begin{aligned} &\left\langle U_{l\epsilon}^{(n)}(\beta_0, .), U_{l\epsilon}^{(n)}(\beta_0, .) \right\rangle(t) \\ &= \sum_{i=1}^n \int_0^t \left\{ H_{il}^{(n)}(x) \right\}^2 \mathbf{1}_{\{|H_{il}^{(n)}(x)| \geq \epsilon\}} \lambda_i(x) dx \\ &= n^{-1} \sum_{i=1}^n \int_0^t \left\{ H_{il}(x) \right\}^2 \mathbf{1}_{\{|H_{il}^{(n)}(x)| \geq \epsilon\}} Y_i(x) \exp \beta_0' Z_i(x) \lambda_0(x) dx. \end{aligned}$$

Utilisons l'inégalité

$$\forall (a, b) \in \mathbb{R}^2 : |a - b|^2 1_{\{|a-b|>\epsilon\}} \leq 4|a|^2 1_{\{|a|>\frac{\epsilon}{2}\}} + 4|b|^2 1_{\{|b|>\frac{\epsilon}{2}\}}$$

Nous déduisons

$$\begin{aligned} & \left\langle U_{i\epsilon}^{(n)}(\beta_0, \cdot), U_{i\epsilon}^{(n)}(\beta_0, \cdot) \right\rangle (t) \\ & \leq 4n^{-1} \sum_{i=1}^n \int_0^\tau |Z_{il}(x)|^2 1_{\{n^{-\frac{1}{2}}|Z_{il}(x)|>\frac{\epsilon}{2}\}} Y_i(x) \exp \beta'_0 Z_i(x) \lambda_0(x) dx \\ & \quad + 4n^{-1} \sum_{i=1}^n \int_0^\tau |E_i(\beta_0, x)|^2 1_{\{n^{-\frac{1}{2}}|E_i(\beta_0, x)|>\frac{\epsilon}{2}\}} Y_i(x) \exp \beta'_0 Z_i(x) \lambda_0(x) dx. \end{aligned} \quad (3.1)$$

Le 2^{ème} terme dans l'équation (3.1) peut s'écrire

$$4 \int_0^\tau |E_i(\beta_0, x)|^2 1_{\{n^{-\frac{1}{2}}|E_i(\beta_0, x)|>\frac{\epsilon}{2}\}} S^{(0)}(\beta_0, x) \lambda_0(x) dx.$$

Soit $\epsilon' > 0$ (fixé) les hypothèses R_2 et R_5 impliquent que pour n assez grand il existe un ensemble A_1 vérifiant

$$P(A_1) \geq 1 - \epsilon'.$$

Et sur A_1 , on a

$$1_{\{n^{-\frac{1}{2}}|E_i(\beta_0, x)|>\frac{\epsilon}{2}\}} = 0,$$

pour tout $t \in [0, \tau]$. Donc le 2^{ème} terme dans l'équation (3.1) converge vers 0 en probabilité.

Montrons la convergence du 1^{er} terme

Ecrivons

$$\left\{ n^{-\frac{1}{2}} |Z_{il}(x)| > \frac{\epsilon}{2} \right\} = B_{1i}^l(x) \cup B_{2i}^l(x),$$

où

$$B_{1i}^l(x) = \left\{ n^{-\frac{1}{2}} |Z_{il}(x)| > \frac{\epsilon}{2}, \beta'_0 Z_i(x) > -\delta |Z_i(x)| \right\}$$

et

$$B_{2i}^l(x) = \left\{ n^{-\frac{1}{2}} |Z_{il}(x)| > \frac{\epsilon}{2}, \beta_0' Z_i(x) \leq -\delta |Z_i(x)| \right\}.$$

On a

$$B_{1i}^l(x) \cap B_{2i}^l(x) = \emptyset$$

Par l'hypothèse R_3 et R_5 et pour $\epsilon^1 > 0$ et n suffisamment grand, il existe un ensemble A_2 vérifiant

$$P(A_2) \geq 1 - \epsilon^1$$

tel que sur A_2 on ait

$$1_{\{B_{1i}^l(x)\}} Y_i(x) = 0$$

pour tout $x \in [0, \tau]$ et $1 \leq i \leq n$. Alors

$$\int_0^\tau n^{-1} \sum_{i=1}^n |Z_{il}(x)|^2 1_{\{B_{1i}^l(x)\}} Y_i(x) \exp\{\beta_0' Z_i(x)\} \lambda_0(x) dx \xrightarrow[n \rightarrow \infty]{P} 0.$$

Pour prouver que

$$\int_0^\tau n^{-1} \sum_{i=1}^n |Z_{il}(x)|^2 1_{\{B_{2i}^l(x)\}} Y_i(x) \exp\{\beta_0' Z_i(x)\} \lambda_0(x) dx \xrightarrow[n \rightarrow \infty]{P} 0 \quad (3.2)$$

remarquons que cette intégrale est bornée par

$$\int_0^\tau n^{-1} \sum_{i=1}^n |Z_{il}(x)|^2 1_{\left\{ |Z_{il}(x)| > \frac{n^{\frac{1}{2}} \epsilon}{2} \right\}} \exp\{-\delta |Z_i(x)|\} \lambda_0(x) dx.$$

Or

$$\lim_{x \rightarrow \infty} x^2 \exp(-\delta x) = 0, \text{ si } \delta > 0.$$

et pour $\eta > 0$ fixé et n suffisamment grand on a

$$|Z_{il}(x)|^2 1_{\left\{ |Z_{il}(x)| > \frac{n^{\frac{1}{2}} \epsilon}{2} \right\}} \exp\{-\delta |Z_i(x)|\} < \eta$$

pour tout i . Donc l'intégrale dans (3.2) est bornée par

$$\eta \int_0^t \lambda_0(x) dx$$

d'ou le résultat. Toutes les conditions du théorème (37) étant remplies, nous avons

$$n^{-\frac{1}{2}}U(\beta_0, t) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma(\beta_0, t)).$$

2) Le processus de comptage $\bar{N} = \sum_{i=1}^n N_i$ a pour compensateur $\sum_{i=1}^n A_i$.

Posons $\bar{M} = \bar{N} - \sum_{i=1}^n A_i$, le lemme (38) partie (1) entraîne
 $\forall c > 0$ et $\forall \delta > 0$

$$\begin{aligned} & P \{ (n^{-1}\bar{N}(\tau) > c) \\ & \leq \frac{\delta}{c} + P \left\{ \int_0^\tau n^{-1} \sum_{i=1}^n Y_i(x) \exp \{ \beta_0' Z_i(x) \} \lambda_0(x) dx > \delta \right\} \\ & = \frac{\delta}{c} + P \left\{ \int_0^\tau S^{(0)}(\beta_0, x) \lambda_0(x) dx > \delta \right\}. \end{aligned}$$

Par la loi forte des grands nombres $n^{-1}\bar{N}(\tau)$ converge *p.s* et

$$\lim_{n \rightarrow \infty} P \{ (n^{-1}\bar{N}(\tau) > c) \}$$

existe $\forall c < \infty$. Par l'hypothèse R_2 on a

$$\frac{\delta}{c} + P \left\{ \int_0^\tau S^{(0)}(\beta_0, x) \lambda_0(x) dx > \delta \right\}$$

converge vers

$$\frac{\delta}{c} + P \left\{ \int_0^\tau s^{(0)}(\beta_0, x) \lambda_0(x) dx > \delta \right\}.$$

Choisissons $\delta > \int_0^\tau s^{(0)}(\beta_0, x) \lambda_0(x) dx$, nous obtenons

$$\lim_{c \rightarrow \infty} \lim_{n \rightarrow \infty} P \{ (n^{-1} \bar{N}(\tau) > c) \} = 0. \quad (3.3)$$

Rappelons que

$$\mathcal{I}(\beta, t) = -\frac{\partial}{\partial \beta} U(\beta, t) = \int_0^t \sum_{i=1}^n V(\beta, x) dN_i(x).$$

Donc

$$\begin{aligned} & \left\| n^{-1} \mathcal{I}(\hat{\beta}, \tau) - \Sigma(\beta_0, \tau) \right\| \quad (3.4) \\ \leq & \left\| \int_0^\tau \{ V(\hat{\beta}, x) - v(\hat{\beta}, x) \} n^{-1} d\bar{N}(x) \right\| \\ & + \left\| \int_0^\tau \{ v(\hat{\beta}, x) - v(\beta_0, x) \} n^{-1} d\bar{N}(x) \right\| \\ & + \left\| \int_0^\tau v(\beta_0, x) n^{-1} \left\{ d\bar{N}(x) - \sum_{i=1}^n Y_i(x) \exp \{ \beta_0^\lambda Z_i(x) \} \lambda_0(x) dx \right\} \right\| \\ & + \left\| \int_0^\tau v(\beta_0, x) \{ S^{(0)}(\beta_0, x) - s^{(0)}(\beta_0, x) \} \lambda_0(x) dx \right\|. \end{aligned}$$

Les hypothèses R_2 et R_5 entraînent que

$$\sup_{0 \leq x \leq \tau} \left\| V(\hat{\beta}, x) - v(\hat{\beta}, x) \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

Ce résultat et la formule (3.3) montrent que le 1^{er} terme dans l'inégalité (3.4) converge en probabilité vers 0.

La formule (3.3) et la continuité uniforme de $s^{(j)}$ $j = 0, 1, 2$ impliquent que le 2^{ème} terme dans l'inégalité (3.4) converge vers 0 en probabilité.

La négligeabilité asymptotique du 4^{ème} terme est directement déduite de l'application des hypothèses R_1 , R_2 et R_5 .

La partie (2) du lemme (38) implique que pour chaque paire (j, k) on a

$$\begin{aligned}
& P \left\{ \left| \int_0^\tau \{v(\beta_0, x)\}_{jk} n^{-1} dM(x) \right| > \rho \right\} \\
& \leq \frac{\delta}{\rho^2} + P \left\{ n^{-1} \int_0^\tau [\{v(\beta_0, x)\}_{jk}]^2 S^{(0)}(\beta_0, t) \lambda_0(x) dx > \delta \right\}
\end{aligned}$$

Les hypothèses R_1 , R_2 et R_5 impliquent

$$P \left\{ \left| \int_0^\tau \{v(\beta_0, x)\}_{jk} n^{-1} dM(x) \right| > \rho \right\} \xrightarrow[n \rightarrow \infty]{P} 0$$

Donc le 3^{ème} terme dans l'inégalité (3.4) converge vers 0 en probabilité.

D'où le résultat. ■

Nous sommes maintenant en mesure de montrer le théorème suivant.

Théorème 42

Soit $\Sigma(\beta_0, t)$ la matrice définie dans R_6 alors

$$n^{\frac{1}{2}} (\hat{\beta} - \beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1}(\beta_0, \tau))$$

Preuve.

Le développement de Taylor de $U(\beta, \tau)$ au voisinage de β_0 donne

$$U(\hat{\beta}, \tau) = U(\beta_0, \tau) - \mathcal{I}(\beta^*, \tau) (\hat{\beta} - \beta_0)$$

où β^* est un point entre $\hat{\beta}$ et β_0 . $U(\hat{\beta}, \tau)$ étant nul on a

$$n^{-1} \mathcal{I}(\beta^*, \tau) n^{\frac{1}{2}} (\hat{\beta} - \beta_0) = n^{-\frac{1}{2}} U(\beta_0, \tau).$$

Or

$$n^{-\frac{1}{2}} U(\beta_0, t) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma(\beta_0, t))$$

d'après le théorème (41). De plus $\hat{\beta}$ est consistant alors

$$\beta^* \xrightarrow[n \rightarrow \infty]{P} \beta_0,$$

d'où

$$\mathcal{I}(\beta^*, \tau) \xrightarrow[n \rightarrow \infty]{P} \Sigma(\beta_0, \tau).$$

Finalement

$$n^{\frac{1}{2}} (\hat{\beta} - \beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1}(\beta_0, \tau)).$$

■

Le résultat suivant nous donne la normalité asymptotique de $\hat{\Lambda}_0$.

Théorème 43

Soit $\hat{\Lambda}_0$ l'estimateur de la fonction de hasard cumulé Λ_0 dans le modèle de Cox. i.e.

$$\hat{\Lambda}_0(t) = \int_0^t \left\{ \sum_{i=1}^n Y_i(x) \exp \hat{\beta}' Z_i(x) \right\}^{-1} d\bar{N}(x).$$

Alors

$$n^{\frac{1}{2}} (\hat{\Lambda}_0 - \Lambda_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Psi)$$

où

$$\Psi(t) = \int_0^t \frac{\lambda_0(x)}{s^{(0)}(\beta_0, x)} dx + \varphi'(\beta_0, t) \Sigma^{-1}(\beta_0, \tau) \varphi(\beta_0, t)$$

et

$$\varphi(\beta_0, t) = \int_0^t e(\beta_0, x) \lambda_0(x) dx.$$

Preuve.

$$\begin{aligned}
& n^{\frac{1}{2}}(\hat{\Lambda}_0(t) - \Lambda_0(t)) \tag{3.5} \\
&= n^{\frac{1}{2}} \int_0^t \left[\left\{ \sum_{i=1}^n Y_i(x) \exp \{ \hat{\beta}' Z_i(x) \} \right\}^{-1} d\bar{N}(x) - \lambda_0(x) dx \right] \\
&= n^{\frac{1}{2}} \int_0^t \left[\left\{ \sum_{i=1}^n Y_i(x) \exp \{ \hat{\beta}' Z_i(x) \} \right\}^{-1} - \left\{ \sum_{i=1}^n Y_i(x) \exp \{ \beta_0' Z_i(x) \} \right\}^{-1} \right] d\bar{N}(x) \\
&\quad + n^{\frac{1}{2}} \left[\int_0^t \left\{ \sum_{i=1}^n Y_i(x) \exp \{ \beta_0' Z_i(x) \} \right\}^{-1} d\bar{N}(x) - \Lambda_0^*(t) \right] + n^{\frac{1}{2}} \{ \Lambda_0^*(t) - \Lambda_0(t) \},
\end{aligned}$$

où

$$\Lambda_0^*(t) = \int_0^t 1_{\{\bar{Y}(x) > 0\}} \lambda_0(x) dx.$$

On va montrer que

- 1) Le 3^{ème} terme est asymptotiquement négligeable,
- 2) le 2^{ème} terme converge en loi vers $\mathcal{N}(\cdot, \cdot)$,
- 3) le 1^{er} terme converge en loi vers $\mathcal{N}(\cdot, \cdot)$,
- 4) l'indépendance asymptotique entre le 1^{er} et le 2^{ème} terme.

1) La négligeabilité asymptotique de 3^{ème} terme dans (3.5).

Par les hypothèses R₂ et R₅ on a $n^{\frac{1}{2}} \{ \Lambda_0^*(t) - \Lambda_0(t) \}$ est asymptotiquement négligeable.

2) La convergence asymptotique de 2^{ème} terme dans (3.5)

$$\begin{aligned}
& n^{\frac{1}{2}} \left[\int_0^t \left\{ \sum_{i=1}^n Y_i(x) \exp \beta_0' Z_i(x) \right\}^{-1} d\bar{N}(x) - \Lambda_0^*(x) \right] \\
&= \int_0^t \frac{n^{\frac{1}{2}} d\bar{M}(x)}{\sum_{i=1}^n Y_i(x) \exp \beta_0' Z_i(x)} \\
&= \int_0^t \frac{n^{-\frac{1}{2}}}{S^{(0)}(\beta_0, t)} d\bar{M}(x)
\end{aligned}$$

a la forme $\int_0^t H(x)dM(x)$ qui converge d'après le théorème (37) vers un processus gaussien de variance

$$\int_0^t \frac{\lambda_0(x)}{s^{(0)}(\beta_0, x)} dx.$$

3) La convergence asymptotique de 1^{er} terme dans (3.5).

Par le développement de Taylor au voisinage de β_0 de

$$\left\{ \sum_{i=1}^n Y_i(t) \exp \beta' Z_i(t) \right\}^{-1}$$

le 1^{er} terme dans (3.5) peut être écrit

$$\Gamma(\beta^*, t) n^{\frac{1}{2}} (\hat{\beta} - \beta_0)$$

où β^* est un point entre $\hat{\beta}$ et β_0 et $\Gamma(\beta, t)$ est donnée par

$$\Gamma(\beta, t) = - \int_0^t n^{-1} \frac{S^{(1)}(\beta, x)}{\{S^{(0)}(\beta, x)\}^2} d\bar{N}(x).$$

Montrons que $\Gamma(\beta^*, \cdot)$ converge en probabilité vers

$$\int_0^\cdot s^{(1)}(\beta, x) \{s^{(0)}(\beta, x)\}^{-1} \lambda_0(x) dx.$$

Ecrivons

$$\Gamma(\beta_0, t) = - \int_0^t n^{-1} \frac{S^{(1)}(\beta_0, x)}{\{S^{(0)}(\beta_0, x)\}^2} d\bar{N}(x) \quad (3.6)$$

$$\begin{aligned} &= - \int_0^t n^{-1} \frac{S^{(1)}(\beta_0, x)}{\{S^{(0)}(\beta_0, x)\}^2} d\bar{M}(x) \\ &\quad - \int_0^t n^{-1} \frac{S^{(1)}(\beta_0, x)}{\{S^{(0)}(\beta_0, x)\}^2} \sum_{i=1}^n Y_i(x) \exp \{\beta_0' Z_i(x)\} \lambda_0(x) dx. \end{aligned} \quad (3.7)$$

le 1^{er} terme dans (3.6) est une martingale locale de carré intégrable de processus de

variation quadratique

$$\begin{aligned} & \int_0^t n^{-2} \left[\frac{S^{(1)}(\beta_0, x)}{\{S^{(0)}(\beta_0, x)\}^2} \right]^2 \sum_{i=1}^n Y_i(x) \exp\{\beta_0' Z_i(x)\} \lambda_0(x) dx \\ &= \int_0^t n^{-1} \left[\frac{S^{(1)}(\beta_0, x)}{\{S^{(0)}(\beta_0, x)\}^2} \right]^2 \{S^{(0)}(\beta_0, x)\} \lambda_0(x) dx \end{aligned}$$

qui converge en probabilité vers 0, $\forall t \in [0, \tau]$.

D'après le théorème (37) le 1^{ème} terme converge en loi vers $\mathcal{N}(0, 0)$ donc converge en probabilité vers 0.

Le 2^{ème} terme dans (3.6) est

$$\int_0^t \frac{S^{(1)}(\beta_0, x)}{\{S^{(0)}(\beta_0, x)\}} \lambda_0(x) dx \xrightarrow[n \rightarrow \infty]{P} - \int_0^t \frac{s^{(1)}(\beta_0, x)}{\{s^{(0)}(\beta_0, x)\}} \lambda_0(x) dx,$$

donc $\Gamma(\beta_0, \mathbf{t})$ converge en probabilité vers

$$- \int_0^t \frac{s^{(1)}(\beta_0, x)}{\{s^{(0)}(\beta_0, x)\}} \lambda_0(x) dx.$$

De plus d'après le théorème (40) on a

$$n^{\frac{1}{2}} (\hat{\beta} - \beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1}(\beta_0, \tau)).$$

Donc la fonction de covariance de 1^{er} terme dans (3.5) est

$$\varphi'(\beta_0, t) \Sigma^{-1}(\beta_0, t) \varphi(\beta_0, t).$$

4) L'indépendance asymptotique entre le 1^{er} terme et le 2^{ème} terme.

Puisque $\hat{\beta}$ est une fonction de vecteur score $U(\beta_0, \tau)$ alors il suffit de montrer L'indépendance asymptotique entre le vecteur score et le 2^{ème} terme.

$$\begin{aligned}
& \left\langle U_l(\beta_0, \cdot), \int_0^\cdot \left\{ \sum_{i=1}^n Y_i(x) \exp \beta'_0 Z_i(x) \right\}^{-1} d\bar{M}(x) \right\rangle (t) \\
&= \sum_{i=1}^n \int_0^t \left[\frac{\{Z_{il}(x) - E_l(\beta_0, x)\}}{\sum_{i=1}^n Y_i(x) \exp \beta'_0 Z_i(x)} \right] Y_i(x) \exp \{\beta'_0 Z_i(x)\} \lambda_0(x) dx \\
&= \int_0^t \left\{ \frac{S_l^{(1)}(\beta_0, x)}{S^{(0)}(\beta_0, x)} - E_l(\beta_0, x) \right\} \lambda_0(x) dx \\
&= 0.
\end{aligned}$$

La fonction de variance de $\hat{\Lambda}_0$ contient des paramètres inconnus que nous devons estimer. Dans la représentation de la fonction de variance nous remplaçons chaque terme inconnu par son estimateur empirique il vient

$$\widehat{Var} \left[n^{\frac{1}{2}} (\hat{\Lambda}_0(t) - \Lambda_0(t)) \right] = \int_0^t \frac{d\hat{\Lambda}_0(x)}{S^{(0)}(\hat{\beta}, x)} + \Gamma'(\hat{\beta}, \mathbf{t}) \hat{\Sigma}^{-1}(\hat{\beta}, \tau) \Gamma(\hat{\beta}, \mathbf{t}),$$

où

$$\hat{\Sigma}^{-1}(\hat{\beta}, \tau) = \frac{1}{n} \int_0^\tau \sum_{i=1}^n V(\hat{\beta}, x) dN_i(x).$$

■

Théorème 44

Sous les hypothèses R_1, \dots, R_6 on a

1) *sous l'hypothèse $H_0 : \beta = \beta_0$,*

$$2 \left\{ \mathcal{L}(\hat{\beta}, t) - \mathcal{L}(\beta_0, t) \right\}$$

converge en loi vers χ_p^2 .

2) *Posons $\beta = (\beta'_1, \beta'_2)'$ où β'_1 (resp β'_2) est un vecteur de dimension $q \times 1$ (resp $(p-q) \times 1$).*

Soit $\hat{\beta}_2(\beta_{10})$ l'estimateur du maximum de vraisemblance partielle de β_2 avec β_1 fixé à la

valeur β_{10} . Alors

$$\left\{ \mathcal{L}(\hat{\beta}) - \mathcal{L} \left[\beta_{10}, \hat{\beta}_2(\beta_{10}) \right] \right\}$$

converge en loi vers χ_p^2 .

les résultats de ce chapitre sont la base des tests donnés au chapitre 2.

3.4 Cas particulier

Les hypothèses R_1, \dots, R_6 sont plus faciles à vérifier dans le cas des données indépendantes et identiquement distribuées (iid), où les covariables sont constantes et bornées. Dans ce cas les fonction $s^{(j)}, j = 0, 1, 2$ définies dans R_2 prennent des formes simples. Chaque $S^{(j)}, j = 0, 1, 2$ est une somme de termes indépendants et puisque les covariables sont bornés alors $S^{(j)}$ converge presque sûrement vers son espérance. Par exemple $S^{(1)}(\beta, t)$ converge *p.s* vers

$$\begin{aligned} E \{ S^{(1)}(\beta, t) \} &= E \left\{ Z_i Y_i(t) e^{\beta' Z_i} \right\} \\ &= E \left\{ Z_i e^{\beta' Z_i} P \{ Y_i(t) = 1/Z_i \} \right\} \\ &= E \left\{ Z_i e^{\beta' Z_i} \Pi(t/Z_i) \right\} \end{aligned}$$

où

$$\Pi(t/Z_i) = P \{ X_i(t) \geq t/Z_i \}.$$

On peut montrer aussi que

$$\begin{aligned} S^{(0)}(\beta, t) &\text{ converge } p.s \text{ vers } E \left\{ e^{\beta' Z_i} \Pi(t/Z_i) \right\} \\ S^{(2)}(\beta, t) &\text{ converge } p.s \text{ vers } E \left\{ Z_i^{\otimes 2} e^{\beta' Z_i} \Pi(t/Z_i) \right\}. \end{aligned}$$

Théorème 45

Supposons que les observations (T, C, Z) et $(T_i, C_i, Z_i)_{i=1,2,\dots,n}$ dans le modèle de Cox sont (iid) et T_i et C_i sont indépendantes conditionnellement à Z_i . Supposons aussi que

les covariables $(Z_i)_{1 \leq i \leq n}$ sont constantes et bornées. Alors les hypothèses R_1, \dots, R_6 sont satisfaites si

$$1) P \{Y_i(\tau) > 0\} > 0$$

et

2) la matrice $\Sigma(\beta_0, \tau)$ est définie positive.

Preuve.

Vérifions l'hypothèse R_1

$$\begin{aligned} 0 &< P \{Y_i(\tau) > 0\} = P \{T_i \geq \tau, C_i \geq \tau\} \\ &= E \{P \{T_i \geq \tau, C_i \geq \tau/Z_i\}\} \\ &= E \{P \{T_i \geq \tau/Z_i\} P \{C_i \geq \tau/Z_i\}\} \\ &= E \left\{ [S_0(\tau)]^{\exp\{\beta'_0 Z_i\}} P \{C_i \geq \tau/Z_i\} \right\}, \end{aligned}$$

donc

$$P \left\{ [S_0(\tau)]^{\exp\{\beta'_0 Z_i\}} > 0 \right\} > 0$$

d'où $S_0(\tau) > 0$ et $\int_0^\tau \lambda_0(x) dx < \infty$.

Pour montrer l'hypothèse R_1 notons que $\forall t \in [0, \tau]$ et $\forall \beta \in \mathbb{R}^p$

$$S^{(j)}(\beta, t) \xrightarrow[n \rightarrow \infty]{p.s.} s^{(j)}(\beta, t)$$

par la loi forte des grands nombres. Nous devons montrer aussi qu'il existe un voisinage B de β_0 tel que

$$\sup_{\substack{x \in [0, \tau] \\ \beta \in B}} \|S^{(j)}(\beta, x) - s^{(j)}(\beta, x)\| \xrightarrow[n \rightarrow \infty]{p.s.} 0.$$

Nous montrons ce dernier résultat pour $S^{(0)}$ et la démonstration est la même pour $S^{(1)}$ et $S^{(2)}$.

1) Supposons que $P \{C_i \geq t/Z_i\}$ est continue en t pour chaque valeur de Z_i alors

$$\Pi(t/Z_i) = [S_0(\tau)]^{\exp\{\beta'_0 Z_i\}} P \{C_i \geq t/Z_i\}$$

est aussi continu en t pour chaque valeur de Z_i , le théorème de convergence dominée implique que

$$s^{(0)}(\beta, t) = E \left\{ e^{\beta' Z_i} \Pi(t/Z_i) \right\}$$

est continu en t . Pour chaque $t \in [0, \tau]$ la loi forte des grands nombres implique

$$S^{(0)}(\beta, t) \xrightarrow[n \rightarrow \infty]{p.s.} s^{(0)}(\beta, t),$$

et comme $S^{(0)}(\beta, \cdot)$, $n = 1, 2, \dots$ est une suite de fonctions monotones et bornées qui converge en chaque point vers une fonction monotone et bornée $s^{(0)}(\beta, \cdot)$ donc la convergence est uniforme sur $t \in [0, \tau]$.

2) Si $P \{C_i \geq t/Z_i\}$ est discontinu alors $\Pi(t/Z_i)$ et $s^{(0)}(\beta, t) = E \left\{ e^{\beta' Z_i} \Pi(t/Z_i) \right\}$ peuvent être discontinus. Or Z_i est borné donc $s^{(0)}(\beta, t)$ est bornée, continue à gauche et non croissante sur $[0, \infty)$, donc $s^{(0)}(\beta, \cdot)$ a au plus un ensemble dénombrable des sauts sur $[0, \infty)$ noté $J = \{t_1, t_2, \dots\}$. Soit \mathbb{Q} l'ensemble des nombres rationnels. Une preuve identique à celle du théorème de Glivenko-Cantelli (c.f. Laha) montre que

$$S^{(0)}(\beta, t) \xrightarrow[n \rightarrow \infty]{p.s.} s^{(0)}(\beta, t),$$

pour tout $t \in [0, \tau] \cap \mathbb{Q}$ et

$$\frac{1}{n} \sum_{i=1}^n e^{\beta' Z_i} \{Y_i(t^+) - Y_i(t)\} \xrightarrow[n \rightarrow \infty]{p.s.} E \left\{ e^{\beta' Z_i} P \{X_i > t/Z_i\} P \{X_i \geq t/Z_i\} \right\}$$

pour tout $t \in J \cap [0, \tau]$.

La convergence sur un ensemble dense dans $[0, \tau]$ et en tout point de discontinuité de $s^{(0)}$ sont utilisés pour montrer que la convergence sur $[0, \tau]$ est uniforme sur le même

ensemble comme dans la preuve de théorème de Glivenko-Cantelli.

Si B est un voisinage compact de β_0 , la bornitude de Z_i est les résultats ci-dessus impliquent

$$\sup_{\substack{x \in [0, \tau] \\ \beta \in B}} \|S^{(j)}(\beta, x) - s^{(j)}(\beta, x)\| \xrightarrow[n \rightarrow \infty]{p.s.} 0$$

Comme les Z_i sont bornés donc on n'a pas besoin de vérifier l'hypothèse R_3 .

La bornitude de Z_i et le théorème de convergence dominée sont suffisants pour montrer l'hypothèse R_4 .

Il est clair que $s^{(j)}$ est borné sur $B \times [0, \tau]$, pour voir que $s^{(0)}$ est borné au delà de 0 sur $B \times [0, \tau]$, notons que la compacité de B et la bornitude de Z_i impliquent qu'il existe un constant k tel que $|\beta' Z_i| \leq k$, $\forall \beta \in B$. Alors pour $t \in [0, \tau]$

$$\begin{aligned} s^{(0)}(\beta, t) &= E \left\{ e^{\beta' Z_i} P \{X_i \geq t/Z_i\} \right\} \\ &\leq e^k E \left\{ [S_0(\tau)]^{\beta' Z_i} P \{C_i \geq \tau/Z_i\} \right\} \\ &\leq e^k [S_0(\tau)]^{-k} P \{C_i \geq \tau\} \\ &\leq e^k [S_0(\tau)]^{-k}. \end{aligned}$$

Finalement nous montrons que la continuité de $s^{(0)}(\beta, t)$ en β_0 est uniforme pour $t \in [0, \tau]$ (même preuve pour $s^{(1)}(\beta, t)$, $s^{(2)}(\beta, t)$).

Soit $\{\beta_m\}_{m \geq 1}$ telle que $\beta_m \xrightarrow[n \rightarrow \infty]{} \beta_0$ (i.e. $\beta_{ml} \xrightarrow[n \rightarrow \infty]{} \beta_{0l}$ pour tout $l = 1, 2, \dots, p$). Alors

$$\begin{aligned} \sup_{0 \leq t \leq \tau} |s^{(0)}(\beta_m, t) - s^{(0)}(\beta_0, t)| &\leq \sup_{0 \leq t \leq \tau} E \left\{ \left| e^{\beta'_m Z_i} - e^{\beta'_0 Z_i} \right| \Pi(t/Z_i) \right\} \\ &\leq E \left\{ \left| e^{\beta'_m Z_i} - e^{\beta'_0 Z_i} \right| \right\} \xrightarrow[n \rightarrow \infty]{} 0 \end{aligned}$$

par la bornitude de Z_i . ■

Théorème 46

Sous les conditions de théorème (45), soit u un temps tel que

$$\forall \tau < u, P \{Y_i(\tau) > 0\} > 0$$

et $\Sigma(\beta_0, \tau)$ est matrice définie positive. Alors

$$U^{(n)}(\beta_0, u) = n^{-\frac{1}{2}}U(\beta_0, u) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma(\beta_0, u)).$$

où

$$\Sigma(\beta_0, u) = \int_0^u v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt.$$

Preuve.

La normalité asymptotique de $U^{(n)}(\beta_0, \tau)$ pour $\tau < u$ est évidente d'après le théorème (45). Supposons que

$$U^{(n)}(\beta_0, \tau) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U^\infty(\beta_0, \tau)$$

et soit $U^\infty(\beta_0, u)$ un processus gaussien de moyenne nulle et de matrice de covariance $\Sigma(\beta_0, u)$. Il suffit de montrer que

1)

$$U^\infty(\beta_0, \tau) \xrightarrow[\tau \nearrow u]{\mathcal{L}} U^\infty(\beta_0, u).$$

et

2)

$$\lim_{\tau \nearrow u} \limsup_{n \rightarrow \infty} P \left\{ \sup_{\tau \leq s \leq u} \|U^{(n)}(\beta_0, s) - U^{(n)}(\beta_0, \tau)\| > \epsilon \right\} = 0.$$

1) $U^\infty(\beta_0, \tau)$ et $U^\infty(\beta_0, u)$ sont deux processus gaussiens de moyenne nulle, de plus $\Sigma(\beta_0, \tau)$ converge vers $\Sigma(\beta_0, u)$ quand $\tau \nearrow u$ donc

$$U^\infty(\beta_0, \tau) \xrightarrow[\tau \nearrow u]{\mathcal{L}} U^\infty(\beta_0, u)$$

2) La preuve est très difficile et nous la donnons pour $p = 1$ (sans perte de généralité).

$$U^{(n)}(\beta_0, s) - U^{(n)}(\beta_0, \tau) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_{\tau}^s \{Z_i - E(\beta_0, t)\} dM_i(x)$$

par le corollaire (28) on a $\forall \varepsilon, \delta > 0$

$$\begin{aligned} & P \left\{ \sup_{\tau \leq s \leq u} n^{-\frac{1}{2}} \left| \sum_{i=1}^n \int_{\tau}^s \{Z_i - E(\beta_0, x)\} dM_i(x) \right| > \varepsilon \right\} \\ & \leq \frac{\delta}{\varepsilon^2} + \Pr \left\{ \frac{1}{n} \sum_{i=1}^n \int_{\tau}^u \{Z_i - E(\beta_0, s)\}^2 d\langle M_i, M_i \rangle(s) > \delta \right\} \\ & \leq \frac{\delta}{\varepsilon^2} + \frac{1}{\delta} E \left\{ \frac{1}{n} \sum_{i=1}^n \int_{\tau}^u \{Z_i - E(\beta_0, s)\}^2 Y_i(s) \exp\{\beta_0 Z_i\} \lambda_0(s) ds \right\} \\ & \leq \frac{\delta}{\varepsilon^2} + \frac{1}{\delta} \int_{\tau}^u E \left[\{Z_i - E(\beta_0, s)\}^2 Y_i(s) \exp\{\beta_0 Z_i\} \right] \lambda_0(s) ds. \end{aligned}$$

Il suffit de montrer que

$$\lim_{\tau \nearrow u} \int_{\tau}^u E \left[\{Z_i - E(\beta_0, s)\}^2 Y_i(s) \exp\{\beta_0 Z_i\} \right] \lambda_0(s) ds = 0.$$

Or $\forall s > 0$:

$$E(\beta_0, s) = \frac{\sum_{i=1}^n Z_i Y_i(s) \exp \beta_0 Z_i}{\sum_{i=1}^n Y_i(s) \exp \beta_0 Z_i}$$

est l'espérance de la covariable à risque à l'instant s , comptée par rapport à la loi discrète définie par

$$h(Z_i) = \frac{Y_i(t) \exp \beta_0 Z_i}{\sum_{i=1}^n Y_i(t) \exp \beta_0 Z_i}.$$

Or pour chaque loi de probabilité d'un variable aléatoire Z de moyenne μ on a

$$E(Z - \mu)^2 < EZ^2,$$

d'où

$$\frac{\sum_{i=1}^n \{Z_i - E(\beta_0, s)\}^2 Y_i(s) \exp \beta_0 Z_i}{\sum_{i=1}^n Y_i(s) \exp \beta_0 Z_i} \leq \frac{\sum_{i=1}^n Z_i^2 Y_i(s) \exp \beta_0 Z_i}{\sum_{i=1}^n Y_i(s) \exp \beta_0 Z_i}.$$

Donc

$$\sum_{i=1}^n \{Z_i - E(\beta_0, s)\}^2 Y_i(s) \exp \beta_0 Z_i \leq \sum_{i=1}^n Z_i^2 Y_i(s) \exp \beta_0 Z_i.$$

D'où

$$E [\{Z_i - E(\beta_0, s)\}^2 Y_i(s) \exp \beta_0 Z_i] \leq E [Z_i^2 Y_i(s) \exp \beta_0 Z_i].$$

Il suffit donc de prouver que

$$\lim_{\tau \nearrow u} \int_{\tau}^u E [Z_i^2 Y_i(s) \exp \{\beta_0 Z_i\}] \lambda_0(s) ds = 0,$$

i.e. il suffit de prouver que

$$\int_0^u E [Z_i^2 Y_i(t) \exp \{\beta_0 Z_i\}] \lambda_0(s) ds < \infty.$$

Or si $|Z_i| \leq c$, on a

$$\int_0^u E [Z_i^2 Y_i(s) \exp \{\beta_0 Z_i\}] \lambda_0(s) ds \leq c^2 E \int_0^u Y_i(s) \exp \{\beta_0 Z_i\} \lambda_0(s) ds$$

$$\begin{aligned} \int_0^u E [Z_i^2 Y_i(s) \exp \{\beta_0 Z_i\}] \lambda_0(s) ds &\leq c^2 EN(u) \\ &\leq c^2, \end{aligned}$$

d'où le résultat. ■

Chapitre 4

Autres modèles semi-paramétriques

4.1 Modèle de Cox stratifié

Dans cette section nous nous intéressons au modèle de régression de Cox stratifié, défini par les fonctions de hasard

$$\lambda_j(t | Z(t)) = \lambda_{j0}(t) \exp \beta'_0 Z(t) \quad , \quad j = 1, \dots, s,$$

où β est un paramètre de régression, et les $\{\lambda_{j0}, j = 1, \dots, s\}$ sont s fonctions de hasard de base inconnues considérées comme des paramètres de nuisance. Ce modèle généralise le modèle semi-paramétrique de Cox usuel en permettant que différents groupes d'individus (les strates) soient gouvernées par des fonctions de hasard de base $\lambda_{j0}, j = 1, \dots, s$ distinctes.

L'inférence statistique dans ce modèle repose sur la méthode du maximum de vraisemblance partielle (MVP), qui fournit un estimateur consistant et asymptotiquement normal. L'estimation et les tests d'hypothèses se font comme auparavant où le log de vraisemblance partielle est donné par

$$\mathcal{L}(\beta) = \mathcal{L}_1(\beta) + \mathcal{L}_2(\beta) + \dots + \mathcal{L}_s(\beta) \tag{4.1}$$

où $\mathcal{L}_j(\beta)$ est le log de vraisemblance qui utilise seulement les données des sujet dans la $j^{\text{ème}}$ strate. Les dérivées de $\mathcal{L}(\beta)$ donnée par la formule(4.1) sont trouvées en additionnant les dérivées à travers chaque strate.

La fonction de hasard cumulé pour la $j^{\text{ème}}$ strate peut être estimée comme décrit dans la section (2.5).

$$\hat{\Lambda}_{j0}(t) = \int_0^t \left\{ \sum_{i=1}^{n_j} Y_{ji}(x) \exp \hat{\beta}' Z_{ji}(x) \right\}^{-1} \left(\sum_{i=1}^{n_j} dN_{ji}(x) \right), \quad j = 1, \dots, s,$$

où Y_{ji}, N_{ji} et Z_{ji} sont les processus correspondant au $i^{\text{ème}}$ individu dans le $j^{\text{ème}}$ strate.

Une supposition clé dans l'utilisation du modèle de Cox stratifié, est que les covariables agissent de la même façon sur la fonction de hasard de base dans toutes les strates.

Exemple 47

Nous avons vu dans l'exemple (34) que le traitement n'a pas un effet multiplicatif sur le taux de hazard de base λ_0 . Nous considérons donc les deux groupes de traitements comme deux strates au lieu de considérer comme des covariables. Nous obtenons les résultats suivants

1) Estimation des paramètres

```
> rar10<-coxph(Surv(Temps,Ind)~age+strata(Tr),cancer5)
> summary(rar10)
Call:
coxph(formula = Surv(Temps, Ind) ~ age + strata(Tr), data = cancer5)

n= 26

      coef exp(coef) se(coef)      z      p
age 0.137      1.15  0.0474 2.9 0.0038

      exp(coef) exp(-coef) lower .95 upper .95
age      1.15      0.872  1.05  1.26
```

Alors

$$\hat{\beta} = 0.137$$

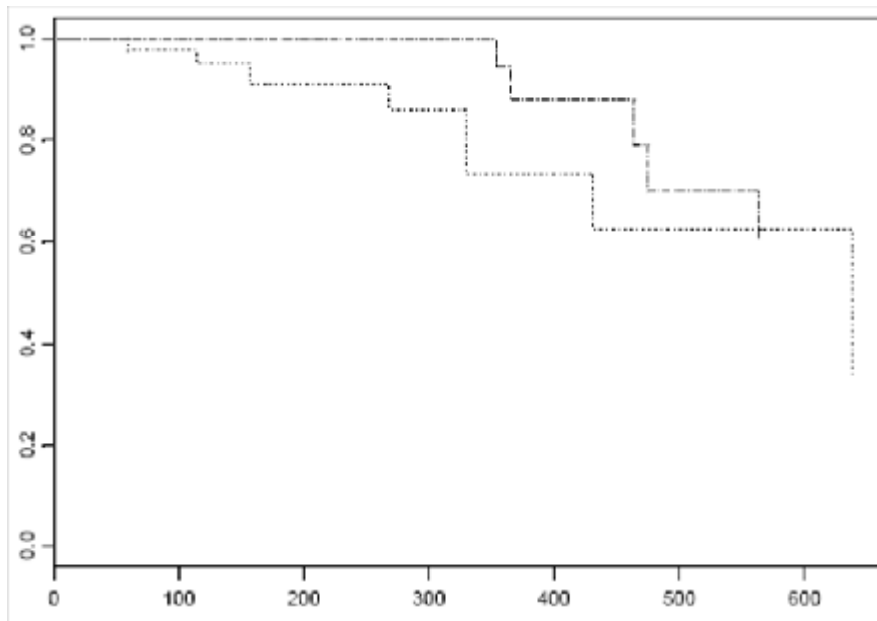
2) Estimation des fonctions de survie de base

```
> summary(survfit(rar10))
Call: survfit.coxph(object = rar10)
```

Tr=1								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
59	13	1	0.978	0.0269		0.9264		1
115	12	1	0.950	0.0481		0.8607		1
156	11	1	0.910	0.0758		0.7725		1
268	10	1	0.862	0.1050		0.6793		1
329	9	1	0.736	0.1525		0.4902		1
431	8	1	0.625	0.1698		0.3671		1
638	5	1	0.341	0.2225		0.0947		1

Tr=2								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
353	13	1	0.943	0.0558		0.840		1.000
365	12	1	0.880	0.0812		0.735		1.000
464	9	1	0.791	0.1125		0.599		1.000
475	8	1	0.701	0.1318		0.485		1.000
563	7	1	0.602	0.1460		0.374		0.968

```
> plot(Survfit(rar10), lty=2:3)
```



La courbe du haut donne l'estimation de la fonction de base dans la strate (1).

La courbe du bas donne l'estimation de la fonction de base dans la strate (2).

4.2 Le modèle de fragilité

Les modèles de fragilité sont une généralisation du modèle de Cox.

Quelle est la raison pour laquelle nous devons généraliser le modèle de Cox ? Les modèles et méthodes standards supposent que la population est homogène. Or dans certaines situations, cette hypothèse n'est pas réaliste. Les gens sont différents, par exemple, ils peuvent avoir une prédisposition génétique à certaines maladies. On peut essayer de modéliser cette hétérogénéité en l'introduisant dans le modèle, Aussi introduisons nous dans le modèle une nouvelle covariable, non observée, Z_0

$$\lambda(t/Z, Z_0) = \lambda_0(t) \exp\{\beta_0 Z_0\} \exp\{\beta' Z\}.$$

On note

$$\eta = \exp\{\beta_0 Z_0\},$$

où η est une variable aléatoire réelle positive de fonction de répartition $F_\eta(\eta)$ appelée la fragilité "frailty". L'estimation est basée sur la fonction de vraisemblance non-paramétrique (cf Nielsen(1991)).

Murphy (1995) a montré la consistance et la normalité asymptotique de l'estimateur de maximum de vraisemblance non-paramétrique. La fonction de survie s'écrit donc

$$\begin{aligned} S(t/Z, \eta) &= \exp\left\{-\int_0^t \lambda_0(s) \eta \exp\{\beta' Z\}\right\} \\ &= \exp\{-\eta \exp\{\beta' Z\} \Lambda_0(t)\}. \end{aligned}$$

Comme η n'est pas observée, la survie doit être moyennée sur η

$$S(t/Z) = \int_0^\infty \exp\{-\eta \exp\{\beta' Z\} \Lambda_0(t)\} F_\eta(\eta)$$

Exemple 48

Soit le modèle de Cox

$$S(t/Z) = \exp \{ - \exp \{ \beta' Z \} \Lambda_0(t) \}$$

où Λ_0 est le taux de hasard cumulé de base, alors

$$S(t/Z) = \exp \{ - \exp \{ \beta' Z \} \lambda_0(t) \}$$

Le choix le plus habituel pour la loi F_η de la fragilité est la loi gamma de densité

$$f(a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}.$$

Alors on a

$$\begin{aligned} E(\eta) &= ab \\ \text{var}(\eta) &= ab^2. \end{aligned}$$

On suppose en général que la moyenne de η est égale à 1 et on prend alors comme unique paramètre de la loi sa variance, notée c ce qui donne

$$\begin{aligned} E(\eta) &= ab = 1 \\ \text{var}(\eta) &= ab^2 = c \\ \eta &\sim g\left(\frac{1}{c}, c\right) \end{aligned}$$

Le paramètre c , qui caractérise la variabilité de la fragilité peut être supposé connu ou inconnu. Regardons que devient la fonction de survie dans ce cas.

$$\begin{aligned} S(t/Z) &= \int_0^\infty e^{\{-\eta e^{\{\beta' Z\} \Lambda_0(t)}\}} f_\eta(\eta) d\eta \\ &= \int_0^\infty e^{\{-\eta e^{\{\beta' Z\} \Lambda_0(t)}\}} \frac{1}{c^{\frac{1}{c}} \Gamma(\frac{1}{c})} \eta^{\frac{1}{c}-1} e^{-\frac{\eta}{c}} d\eta \end{aligned}$$

$$\begin{aligned}
S(t/Z) &= \frac{1}{c^{\frac{1}{c}}\Gamma(\frac{1}{c})} \int_0^\infty \eta^{\frac{1}{c}-1} e^{-\{\frac{1}{c}+e^{\{\beta'Z\}\Lambda_0(t)}\}\eta} d\eta \\
&= \frac{1}{c^{\frac{1}{c}}\Gamma(\frac{1}{c})} \Gamma(\frac{1}{c}) \frac{(\frac{1}{c} + e^{\{\beta'Z\}\Lambda_0(t)})^{-1}}{(\frac{1}{c} + e^{\{\beta'Z\}\Lambda_0(t)})^{\frac{1}{c}-1}} \\
&= \frac{(\frac{1}{c} + e^{\{\beta'Z\}\Lambda_0(t)})^{-\frac{1}{c}}}{c^{\frac{1}{c}}} \\
&= \frac{(\frac{1}{c})^{-\frac{1}{c}} (1 + ce^{\{\beta'Z\}\Lambda_0(t)})^{-\frac{1}{c}}}{c^{\frac{1}{c}}} \\
&= \left(1 + ce^{\{\beta'Z\}\Lambda_0(t)}\right)^{-\frac{1}{c}} \\
&= e^{\frac{-1}{c} \log(1+ce^{\{\beta'Z\}\Lambda_0(t)})} \\
&= e^{-\left(\log(1+ce^{\{\beta'Z\}\Lambda_0(t)})\right)^{\frac{1}{c}}} \\
&= e^{-G(e^{\{\beta'Z\}\Lambda_0(t)})}.
\end{aligned}$$

On voit donc par ce calcul qu'un modèle de fragilité gamma généralise le modèle de Cox de la manière suivante. Pour la fragilité gamma on a le modèle

$$S(t/Z) = e^{-G(e^{\{\beta'Z\}\Lambda_0(t)})}$$

où

$$G(u) = \log\left(\left(1 + cu\right)^{\frac{1}{c}}\right)$$

alors que pour le modèle de Cox G est simplement l'identité $G(u) = u$.

4.3 Le modèle de fragilité corrélé

Ce modèle est une généralisation du modèle de fragilité. Le taux de hasard sous le modèle de fragilité corrélé d'un individu i est donné par

$$\lambda(t / Z_i, Q^i) = \lambda_0(t) Q^i \exp \{ \beta' Z_i(t) \},$$

où

$$Z_i(t) = (Z_{i1}(t), Z_{i2}(t), \dots, Z_{ip}(t))'$$

est le vecteur de covariables, $Q^i = Q_0 + Q_i$ tel que

$$Q_0, Q_1, Q_2, \dots, Q_n$$

sont des variables aléatoires indépendantes de loi gamma de paramètres

$(v, \eta), (v^*, \eta), (v^*, \eta), \dots, (v^*, \eta)$ respectivement, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ est le vecteur de paramètre et $\lambda_0(\cdot)$ est une fonction de hasard de base. Dans ce cas Q^i est une variable aléatoire de loi gamma de paramètre $(v - v^*, \eta)$.

4.4 Modèle de Cox avec covariable manquante

Ce modèle a été introduit par Dupuy et autres (cf Dupuy (2006)). Le modèle de Cox définit la loi d'un instant d'évènement T conditionnelle à une variable explicative (covariable) dépendant du temps $(Z(t))$, $t \geq 0$ par la fonction de hasard

$$\lambda(t / Z(t)) = \lambda_0(t) \exp \beta' Z(t)$$

où β est un paramètre de régression et λ_0 une fonction de hasard de base considérée comme un paramètre de nuisance. Cox a proposé d'estimer β au vu d'un échantillon de données censurées

$$X = (X_i, \delta_i, \{Z_i(s), 0 \leq s \leq X_i\})_{1 \leq i \leq n}$$

où $X_i = T_i \wedge C_i$, T_i est le temps de survie du $i^{\text{ème}}$ malade et C_i est le temps de de censure du $i^{\text{ème}}$ malade, $\delta_i = 1_{\{T_i > C_i\}}$ est l'indicateur de censure. Si les valeurs

$$\{Z_i(X_i), i = 1, 2, \dots, n\}$$

aux instants d'évènement ne sont pas observées, nous modélisons conjointement (X, δ) et $Z(\cdot)$ pour en déduire une vraisemblance permettant d'estimer β et Λ_0 au vu des données incomplètes.

4.4.1 Construction de la vraisemblance conjointe

Étudions le cas où Z est un scalaire.

1. Supposons que l'on observe une covariable $Z(\cdot)$ dans l'intervalle de temps $[0, \tau]$. Supposons que $Z(\cdot)$ soit une fonction en escalier de valeur constante $Z(t) = Z_j$ sur les intervalles $(t_{j-1}, t_j]$, où t_j est l'instant d'observation de Z_j , ($j = 1, \dots, K, t_0 = 0, t_K = \tau$). Z_0 est une valeur initiale mesurée en t_0 . La valeur Z_j n'est pas mesurée si $X \leq t_j$. Soit

$$a_t = \max(k : t_k < t), (t > 0)$$

l'indice de la dernière observation de $Z(\cdot)$ avant t . Le problème est le suivant. Soit $(X_i, \delta_i, Z_i(\cdot))_{1 \leq i \leq n}$, $(X, \delta, Z(\cdot))$, $(n + 1)$ triplets iid. Pour tout i les observations portent sur vecteur aléatoire

$$Y_i = \left\{ X_i, \delta_i, Z_{i;0}, \dots, Z_{i;a_{X_i}} \right\}.$$

L'observation de $Z_i(\cdot)$ à l'instant X_i est manquante et la loi de X_i en dépend. On dit que $Z_i(X_i)$ est une valeur manquante.

Le problème statistique est d'estimer β et le taux de hasard cumulé Λ_0 au vu des observations incomplètes fournies par $(Y_i)_{1 \leq i \leq n}$.

2. Supposons que pour chaque $j \geq 1$, le vecteur aléatoire (Z_0, \dots, Z_j) à valeurs dans \mathbb{R}^{j+1} admette une densité $f(Z_0, \dots, Z_j, \alpha)$ sur \mathbb{R}^{j+1} , paramétrée par $\alpha \in \mathbb{R}^p$ ($p \geq 1$). Construisons un modèle conjoint pour $(X, \delta, Z(\cdot))$ à partir du modèle de Cox. Ce modèle conjoint est paramétré par $\theta = (\beta, \Lambda_0, \alpha)$, où β et Λ_0 sont les paramètres d'intérêt du modèle.

Nous obtenons la vraisemblance de θ au vu d'une observation incomplète Y en intégrant la densité du vecteur complet $(Y, Z(X))$ sur la valeur inobservée $Z(X)$.

Nous supposons que la distribution de C ne dépend pas de θ (censure non informative), que T et C sont indépendants conditionnellement à $Z(\cdot)$ et que C ne dépend pas des valeurs inobservées de $Z(\cdot)$. On obtient la vraisemblance

$$\int \{\lambda(x)\}^\delta \exp \left[\delta \beta Z(X) - \int_0^x \lambda(x) \exp \{\beta Z(x)\} dx \right] \times f(Z_0, \dots, Z_{a_{X_i}}, Z(X), \alpha) dZ(X)$$

3. Au contraire de la vraisemblance partielle, la vraisemblance (notée L_n par la suite) dépend du paramètre Λ_0 . Pour trouver un maximum de cette vraisemblance, l'espace des fonctions $\Lambda_0(\cdot)$ est restreint à l'ensemble des fonctions en escalier $\Lambda_{0,n}(\cdot)$ positives, croissantes sur $[0, \infty)$, et dont les sauts $\Delta\Lambda_{0,n}(X_k)$ se produisent aux $p(n)$ ($p(n) \leq n$) instants d'événements non censurés X_k ($k = 1, \dots, p(n)$).

L'estimateur de θ obtenu sera dit estimateur semiparamétrique. Notons

$\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n(\cdot), \hat{\alpha}_n)$ l'estimateur de θ obtenu en maximisant sur l'espace des paramètres modifié

$$\left\{ \begin{array}{l} (\beta, \Delta\Lambda_{0,n}(X_1), \dots, \Delta\Lambda_{0,n}(X_{p(n)}), \alpha) : \beta \in \mathbb{R}, \Delta\Lambda_{0,n}(X_k) \in \mathbb{R}, \\ k = 1, 2, \dots, p(n), \alpha \in \mathbb{R}^p \end{array} \right\}$$

la vraisemblance :

$$L_n(\theta) = \prod_{i=1}^n \int \Delta\Lambda_{0,n}^{\delta_i}(X_i) \exp \left[\delta_i \beta Z(X) - \sum_{k=1}^{p(n)} \Delta\Lambda_{0,n}(X_k) \exp \{\beta Z_i(X_k)\} \right] \times 1_{\{X_k \leq X_i\}} f(Z_{i0}, \dots, Z_{ia_{X_i}}, Z(X); \alpha) dZ(X)$$

L'estimateur $\hat{\theta}_n$ obtenu on maximisant $L_n(\theta)$ est appelé estimateur de maximum de vraisemblance non paramétrique (L'EMVNP).

4. Pour maximiser le logarithme de $L_n(\theta)$ nous utilisons l'algorithme itératif EM. Notons que pour estimer les paramètres d'intérêt β et Λ_0 , il est nécessaire d'estimer α qui inter-

vient dans le calcul des espérances conditionnelles E_θ . Notons que si X_1 et X_2 sont deux variables aléatoires, nous notons $E_\theta [X_1/X_2]$ l'espérance conditionnelle de X_1 sachant X_2 , paramétrée par θ .

La proposition suivante fournit une caractérisation de $\hat{\theta}_n$ (L'EMVNP).

Sous certaines conditions (cf dupuy 2006) on a

Proposition 49

(L'EMVNP) $\hat{\theta}_n$ satisfait l'équation suivante,

$$\hat{\Lambda}_n(t) = \int_0^t \frac{dH_n(x)}{W_n(u, \hat{\theta}_n)}$$

où

$$H_n(x) = n^{-1} \sum_{i=1}^n \Delta_i 1_{\{X_i \leq x\}}$$

et

$$W_n(u, \theta) = n^{-1} \sum_{i=1}^n E_\theta [e^{\beta Z(x)} 1_{\{x \leq X\}} / y_i]$$

4.4.2 Propriétés asymptotiques des L'EMVNP

Théorème 50

L'ENPMV $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n(\cdot), \hat{\alpha}_n)$ est consistant. i.e.

$$\begin{aligned} \|\hat{\alpha}_n - \alpha\| &\xrightarrow[n \rightarrow \infty]{p.s.} 0 \\ \left| \hat{\beta}_n - \beta \right| &\xrightarrow[n \rightarrow \infty]{p.s.} 0 \\ \left\| \hat{\Lambda}_n - \Lambda_0 \right\|_\infty &\xrightarrow[n \rightarrow \infty]{p.s.} 0 \end{aligned}$$

où $\|\cdot\|$ est la norme euclidienne et $\|\cdot\|_\infty$ est la norme sup sur $[0, \tau]$.

Soit $VB[0, \tau]$ l'espace des applications à variation bornée de $[0, \tau]$ dans \mathbb{R} , et

$$H = \{h = (h_1, h_2, h_3) : h_1 \in \mathbb{R}^p, h_2 \in \mathbb{R}, h_3 \in VB[0, \tau]\}.$$

Nous définissons sur H la norme suivante, si $h \in H$,

$$\|h\|_H = \|h_1\| + |h_2| + \|h_3\|_v$$

où $\|h_3\|_v$, est la valeur absolue de $h_3(0)$ plus le total de variation de h_3 sur l'intervalle $[0, \tau]$. Soit

$$H_p = \{h \in H, \|h\|_H \leq p\},$$

et

$$H_\infty = \{h \in H, \|h\|_H < \infty\}$$

et $VB_p[0, \tau]$ l'espace des fonctions réelles sur $[0, \tau]$ bornées par p et de variation bornée par p . Soit

$$\theta(h) = (\beta, \Lambda_0, \alpha) = h'_1\alpha + h_2\beta + \int_0^\tau h_0(u)d\Lambda_0(u).$$

Alors on peut considérer θ comme une fonctionnelle sur H_p et l'espace des paramètres Θ est un sous ensemble de $l^\infty(H_p)$ l'espace des fonctions réelles bornées sur H_p .

Théorème 51

La suite $(\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_0 - \Lambda_0))$ converge en loi dans $l^\infty(H_p)$ vers un processus gaussien centré G , de fonction de covariance

$$\begin{aligned} \text{Cov}(G(g), G(g^*)) &= \int_0^\tau g_3(u)\sigma_{3,\theta_0}^{-1}(g^*)(u)d\Lambda_0(u) + \\ &\quad \sigma_{2,\theta_0}^{-1}(g^*)g_2 + \sigma_{1,\theta_0}^{-1}(g^*)'g_1(u) \end{aligned}$$

où

$$\sigma_{\theta_0}^{-1} = (\sigma_{1,\theta_0}^{-1}, \sigma_{2,\theta_0}^{-1}, \sigma_{3,\theta_0}^{-1})$$

est l'inverse de l'opérateur continu

$$\sigma_{\theta_0} = (\sigma_{1,\theta_0}, \sigma_{2,\theta_0}, \sigma_{3,\theta_0})$$

de H_∞ dans H_∞ défini par :

$$\begin{aligned}\sigma_{1,\theta_0}(h) &= E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_{X_i}}, Z, \alpha_0) h_1 \right] \\ \sigma_{2,\theta_0}(h) &= E_{\theta_0} \left[\int_0^X Z(u) e^{\beta_0 Z(u)} (Z(u) h_2 + h_3(u)) d\Lambda_0(u) \right] \\ \sigma_{3,\theta_0}(h)(u) &= E_{\theta_0} [[Z(u) h_2 + h_3(u)] e^{\beta_0 Z(u)} 1_{\{u \leq X\}}].\end{aligned}$$

4.5 Modèle additif

Dans le modèle de Cox nous avons supposé que les covariables ont un effet multiplicatif sur le taux de hasard de base λ_0 . Dans cette section nous présentons un modèle alternatif en supposant que les covariables agissent d'une manière additive sur le taux de hasard de base inconnue. Ce modèle proposé par Odd Aalen (1989) est connu sous le nom du modèle de hasard additif. Les coefficients de risque inconnus dans ce modèle peuvent dépendre du temps.

4.5.1 Définition de modèle

Soit $(X_i, \delta_i, Z_i)_{1 \leq i \leq n}$, où $X_i = T_i \wedge C_i$ (C_i est la variable aléatoire de censure et T_i est la variable d'intérêt) et δ_i est l'indicateur de censure $\delta_i = 1_{\{T_i \leq C_i\}}$.

$$Z_i(t) = (Z_{i1}(t), Z_{i2}(t), \dots, Z_{ip}(t))'$$

est le vecteur de covariables dont dépend la durée de survie T_i . Le taux de hasard d'un sujet qui à pour vecteur de covariables $Z_i(t)$, sous le modèle additif; est

$$\lambda(t/Z_i(t)) = \beta_0(t) + \sum_{j=1}^p \beta_j(t) Z_{ij}(t) \quad (4.2)$$

où β_0 est la fonction de taux de hasard de base et les β_j , $1 \leq j \leq p$ sont des fonctions de risque inconnues. L'estimation directe des fonctions β_j , $1 \leq j \leq p$ étant très difficile

nous estimons les fonctions de risque cumulé B_j , $1 \leq j \leq p$ définies par

$$B_j(t) = \int_0^t \beta_j(x) dx \quad , \quad 1 \leq j \leq p.$$

4.5.2 Estimation de B_j , $0 \leq j \leq p$

Nous définissons les processus de comptage

$$N_i(t) = 1_{\{X_i \leq t, \delta_i = 1\}},$$

et

$$Y_i(t) = 1_{\{X_i \geq t\}}.$$

le processus de présence du $i^{\text{ème}}$ sujet juste avant l'instant t et le processus

$$N(t) = (N_1(t), N_2(t), \dots, N_n(t))'.$$

Soit \mathbb{Y} la matrice de dimension $n \times (p+1)$, où la $j^{\text{ème}}$ ligne de cette matrice est donnée par

$$[Y_j(t), Y_j(t) Z_{j1}(t), Y_j(t) Z_{j2}(t), \dots, Y_j(t) Z_{jp}(t)].$$

Le processus d'intensité du processus $N_j(t)$ sous le modèle additif est :

$$\lambda_j(t) = [(Y_j(t), Y_j(t) Z_{j1}(t), Y_j(t) Z_{j2}(t), \dots, Y_j(t) Z_{jp}(t))] \beta(t) \quad (4.3)$$

où

$$\beta(t) = (\beta_0(t), \beta(t)_1, \dots, \beta_p(t))'$$

donc

$$M(t) = N(t) - \int_0^t \mathbb{Y}(x) \beta(x) dx$$

est une martingale de dimensions $n \times 1$. D'où

$$dN(t) = dM(t) + \mathbb{Y}(t)\beta(t)dt \quad (4.4)$$

On peut supposer que la martingale M est un bruit et on pose $dM(t) = 0$ dans (4.4) on trouve

$$\mathbb{Y}(t)\beta(t)dt = dN(t) \quad (4.5)$$

et nous résolvons (4.5) pour obtenir un estimateur de $B(t) = \int_0^t \beta(x)dx$. la solution existe si la matrice $\mathbb{Y}(t)$ est de rang maximal, dans ce cas soit $\mathbb{Y}^-(t)$ la matrice inverse généralisée de $\mathbb{Y}(t)$ (i.e. $\mathbb{Y}^-(t) \mathbb{Y}(t) = I_{(p+1)(p+1)}$ matrice unité), l'estimateur de $B(t)$ est

$$\begin{aligned} \hat{B}(t) &= \int_0^t \mathbb{Y}^-(x)dN(x) \\ \hat{B}(t) &= \sum_{T_j \leq t} \delta_j \mathbb{Y}^-(T_j)(\Delta N_1(T_j), \dots, \Delta N_1(T_j))', \text{ pour } t \leq \tau \end{aligned}$$

où $\tau = \max \{T_j : \mathbb{Y}'(T_j)\mathbb{Y}(T_j) \text{ est une matrice nonsingulière}\}$ et la matrice $\mathbb{Y}^-(t)$ donnée par

$$\mathbb{Y}^-(t) = [\mathbb{Y}'(t)\mathbb{Y}(t)]^{-1} \mathbb{Y}'(t).$$

On pose $A(t) = \mathbb{Y}'(t)\mathbb{Y}(t)$ d'élément (g, k) donné par

$$a_{g,k}(t) = \sum_{j=1}^n Y_j(t)Z_{j(g-1)}(t)Z_{j(k-1)}(t), \quad g, k = 1, \dots, p+1 \quad (4.6)$$

où $Z_{j0}(t) = 1$. Pour $t \leq \tau$ soit $R(t) = A^{-1}(t)$ et $R_{g,k}(t)$ l'élément (g, k) de la matrice $R(t)$. Alors l'estimateur de $B_g(t)$ est

$$\hat{B}_g(t) = \sum_{T_j \leq t} \delta_j \sum_{k=1}^{p+1} R_{g+1,k}(T_j)Y_j(T_j)Z_{j(k-1)}(T_j), \quad 0 \leq g \leq p, \quad t \leq \tau \quad (4.7)$$

Exemple 52

Supposons qu'on a un échantillon construit à partir de deux groupes (G_1 et G_2). Et supposons que les Z_j sont des scalaires tels que $Z_j = 1$ si le malade dans le groupe 1 et $Z_j = 0$ si non. Alors la matrice $A(t)$ est donnée par

$$A(t) = \begin{bmatrix} \bar{Y}(t) & \bar{Y}_1(t) \\ \bar{Y}_1(t) & \bar{Y}_1(t) \end{bmatrix},$$

où $\bar{Y}(t)$ est le nombre des individus à risque à l'instant t et $\bar{Y}_1(t)$ est le nombre des individus du 1^{er} groupe (G_1) à risque à l'instant t . D'où

$$R(t) = \begin{bmatrix} \frac{1}{\bar{Y}_2(t)} & \frac{-1}{\bar{Y}_2(t)} \\ \frac{-1}{\bar{Y}_2(t)} & \frac{1}{\bar{Y}_1(t)} + \frac{1}{\bar{Y}_2(t)} \end{bmatrix}.$$

D'après la formule (4.7) on a

$$\begin{aligned} \hat{B}_0(t) &= \sum_{\substack{T_j \leq t \\ j \in G_2}} \frac{\delta_j}{\bar{Y}_2(T_j)} - \sum_{\substack{T_j \leq t \\ j \in G_1}} \frac{\delta_j}{\bar{Y}_2(T_j)} \\ &= \sum_{\substack{T_j \leq t \\ j \in G_2}} \frac{\delta_j}{\bar{Y}_2(T_j)}. \end{aligned}$$

et

$$\begin{aligned} \hat{B}_1(t) &= - \sum_{\substack{T_j \leq t \\ j \in G_1}} \frac{\delta_j}{\bar{Y}_2(T_j)} + \sum_{\substack{T_j \leq t \\ j \in G_1}} \delta_j \left[\frac{1}{\bar{Y}_1(t)} + \frac{1}{\bar{Y}_2(t)} \right] \\ &= \sum_{\substack{T_j \leq t \\ j \in G_1}} \frac{\delta_j}{\bar{Y}_1(T_j)} - \sum_{\substack{T_j \leq t \\ j \in G_2}} \frac{\delta_j}{\bar{Y}_2(T_j)}. \end{aligned}$$

Remarque 53

Huffer et Mckeague (1991) proposent d'utiliser la matrice inverse généralisée des

moindres carrés donnée par

$$\mathbb{Y}^-(t) = [\mathbb{Y}'(t)W(t)\mathbb{Y}(t)]^{-1} \mathbb{Y}'(t)W(t)$$

où $W(t)$ est une matrice diagonale de dimensions $n \times n$ avec l'élément (j, j) proportionnel avec l'inverse de la variance de $dM_j(t)$.

4.6 Le modèle de Cox-Aalen

Ce modèle est proposé par Scheike et Zhang en 2002. Le processus d'intensité sous le modèle de Cox-Aalen est donné par

$$\lambda_i(t) = Y_i(t)Q_i'(t)\alpha(t) \exp \{ \beta' Z_i(t) \},$$

où

$$Q_i(t) = (Q_{i1}(t), Q_{i2}(t), \dots, Q_{ip}(t))',$$

et

$$Z_i(t) = (Z_{i1}(t), Z_{i2}(t), \dots, Z_{iq}(t))',$$

sont deux vecteurs de covariables et

$$\alpha(t) = (\alpha_1(t), \dots, \alpha_p(t))',$$

est le vecteur de fonctions de risque qu'on doit estimer. C'est-à-dire que quelques covariables ont un effet additif et les autres ont un effet multiplicatif.

Si β est connu nous utilisons l'estimateur de Huffer et Mckeague (1991) de fonction de risque cumulé $B(t) = \int_0^t \alpha(x)dx$ en résolvant l'équation

$$\mathbb{Y}'(t, \beta)W(t) \{dN(t) - \mathbb{Y}'(t, \beta)d\Lambda(t)\} = 0,$$

où $N(t)$ est un processus de comptage multivarié et $\mathbb{Y}(t, \beta)$ est donné par

$$\mathbb{Y}(t, \beta) = (Y_1(t) \exp \{\beta' Z_1(t)\} Q_1(t), \dots, Y_n(t) \exp \{\beta' Z_n(t)\} Q_n(t))',$$

et $\mathbb{Y}^-(t, \beta)$ est la matrice inverse généralisée de $\mathbb{Y}(t, \beta)$ tel que

$$\mathbb{Y}^-(t, \beta) = [\mathbb{Y}'(t, \beta)W(t)\mathbb{Y}(t, \beta)]^{-1} \mathbb{Y}'(t, \beta)W(t),$$

où

$$W(t) = \text{diag} [w_1(t, \beta), \dots, w_n(t, \beta)],$$

est une matrice diagonale telle que

$$w_i(t, \beta) = \frac{Y_i(t) \exp \{-\beta' Z_i(t)\}}{h_i(t)},$$

où $h_i(t)$ est une fonction connue ne dépendant pas de β , un choix simple de la fonction $h_i(t)$ est $h_i(t) = 1$ pour tout i . L'estimateur de B est donné par

$$\widehat{B}(t, \beta) = \int_0^t \mathbb{Y}^-(s, \beta) dN(s).$$

Nous résolvons l'équation

$$U(\beta) = \int_0^\tau [Z'(s) - Z'(s)\mathbb{Y}(s, \beta)\mathbb{Y}^-(s, \beta)] dN(s) = 0,$$

pour obtenir un estimateur $\widehat{\beta}$ de β . Où

$$Z = (Z_1, Z_2, \dots, Z_n)'$$

Bibliographie

- [1] Aalen, O. O. (1978). Nonparametric regression for a family of counting processes. *Z. Ann. Statist.* **6**, 701-726.
- [2] Aalen, O. O. (1989). A Linear Regression Model for the Analysis of Life Times, *Statistics in Medicine*, **8**, 907-925
- [3] Andersen, P. K. Gill. R. D. (1982). Cox's regression model for counting processes : a large sample study, *Ann. Statist.* **10**, 1100-1120
- [4] Bousseboua, M. (2004). *Statistique Mathématique*. Les éditions de l'université Mentouri Constantine.
- [5] Breslow, N. E. (1972). Contribution to the discussion on the paper by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34**, 216-217.
- [6] Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.
- [7] Coccozza, C-Thivent. (1997). *Processus stochastiques et fiabilité des systèmes*. Springer-verlag Berlin. Heidelberg.
- [8] Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society B*. **74**, 187-220.
- [9] Cox, D. R. (1975). "Partial Likelihood", *Biometrika*. May/Aug. **62(2)**, pp. 269-76.
- [10] Dolivo, F.G. (1974) *Counting processus and integrated conditional rates* : a martingale approach with application to theory. Ph.D. thesis, University of Michigan.

- [11] Dupuy J.-F. Grama, I. Mesbah, M. (2006). Asymptotic theory for the Cox model with missing time-dependent covariate. *The Annals of Statistics*. **34**. 903-924
- [12] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72**, 557-565.
- [13] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- [14] Gill. R. D.(1980). Censoring and stochastic integrals. *Mathematical Centre Tracts* 124.Mathematisch Centrum Amsterdam.
- [15] Huffer, F. W. and Mckeague, I. W. (1991). Weighted Least Squares Estimation for Aalen's Additive Risk Model. *Journal of the American Statistical Association*. **86**. 114-129.
- [16] Kalbfleish, J. D. and Prentice, R. L. (1980). *The statistical Analysis of Failure Time Data*. Wiley and Sons.
- [17] Klein, John P. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. Springer.
- [18] Laha, R. G. (1979). *Probability theory*. John Wiley.
- [19] Messaci, F. *Notions fondamentales dela théotie de probabilités*. Les éditions de l'université Mentouri Constantine, 2001.
- [20] Murphy, S. A.(1995). Asymptotic theory for the frailty model. *Ann Statist*. **23**, 182-198
- [21] Nielsen, G. G, Gill.R.D, Andersen. P. K and Sørensen, T. I. A.(1991). A counting process approach to maximum likelihood estimation in frailty models. *Scand.J. Statis*. **19**, 25-44.
- [22] Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete* **51**, 269-286.
- [23] Scheike, T. H. and Zhang, M. J. (2002). An additive-multiplicative Cox-Aalen regression model. *Scand. J. Statist*. **29**, 75-88