

Table des matières

Introduction	5
Notations	9
I Préliminaires	11
1 Analyse statistique des durées de vie	13
1.1 Les outils probabilistes	13
1.2 Modèles de censure	19
1.2.1 Mécanismes de censure	19
1.2.2 Estimation du maximum de vraisemblance dans un modèle de survie paramétrique avec censure à droite	20
2 Le modèle semiparamétrique de Cox	23
2.1 Présentation générale	23
2.2 Méthode du maximum de vraisemblance partielle	26
2.2.1 Introduction	26
2.2.2 La vraisemblance partielle de Cox	28
2.2.3 Loi de probabilité de la statistique de rang	31
2.2.4 Estimation de la fonction de risque	35
2.2.5 Prise en compte des ex aequo	36
2.3 Propriétés asymptotiques	37
2.4 Méthode du maximum de vraisemblance semiparamétrique	40
2.4.1 Concentration de la vraisemblance	40
2.4.2 Maximum de vraisemblance semiparamétrique dans le modèle de Cox	41

3	Modèle de Cox avec covariables dépendant du temps : travaux récents	47
3.1	Cas de l'erreur de mesure sur la covariable	48
3.2	Cas des valeurs manquantes de la covariable	51
3.2.1	La méthode proposée par Lin et Ying	51
3.2.2	La méthode proposée par Zhou et Pepe	54
3.2.3	La méthode proposée par Wang, Hsu, Feng et Prentice . . .	57
3.2.4	La méthode proposée par Paik et Tsai	58
3.3	Erreur de mesure et valeurs manquantes : modélisation conjointe pour une covariable fixe	62
II	Modèle de Cox avec valeurs manquantes d'une covariable dépendant du temps : proposition d'une approche par modélisation conjointe	65
4	Proposition d'un modèle conjoint et identifiabilité des paramètres	67
4.1	Notations et formulation du problème	67
4.2	Proposition d'une vraisemblance conjointe	70
4.3	Identifiabilité des paramètres dans le modèle conjoint	72
4.3.1	Définition de l'identifiabilité et information au sens de Kullback	72
4.3.2	Identifiabilité des paramètres du modèle conjoint	74
5	Estimation semiparamétrique dans le modèle de Cox avec valeurs manquantes d'une covariable dépendant du temps	79
5.1	Existence des estimateurs semiparamétriques	80
5.2	Une caractérisation des estimateurs semiparamétriques	83
5.3	Prise en compte des ex aequo	87
6	Etude des propriétés asymptotiques pour le modèle de Cox avec covariable dépendant du temps manquante	91
6.1	Preliminaires	91
6.2	Consistance des estimateurs semiparamétriques	96
6.3	Convergence en loi des estimateurs semiparamétriques	99
6.4	Estimation de la variance asymptotique des estimateurs	120
7	Mise en oeuvre de l'algorithme EM dans le modèle de Cox avec covariable manquante	125
7.1	Description de l'algorithme EM pour le modèle conjoint	126

7.2	Applications de l'approche par vraisemblance conjointe	133
7.2.1	Application aux abandons informatifs dans une étude longi- tudinale	133
7.2.2	Comparaison avec deux méthodes d'imputation de covariable dans le modèle de Cox	140
Conclusion et perspectives		143
Annexe A Approximation des intégrales intervenant dans la vraisemblance conjointe		145
Annexe B Listing Scilab		147
Bibliographie		153

Introduction

Le modèle semiparamétrique de Cox est l'un des plus utilisés dans le traitement statistique des durées de vie en présence de variables explicatives (covariables). Il exprime la relation entre le risque d'occurrence d'un événement (tel qu'un décès, la survenue d'une maladie, une panne, ...) et des covariables fixes ou qui dépendent du temps. Cette relation est définie par l'intermédiaire d'une fonction de risque instantané à l'instant t , qui dans le cas de covariables dépendant du temps, est conditionnée par l'observation du processus de covariables jusqu'à t . Cette fonction de risque s'écrit :

$$\lambda_T(t)e^{g(Z(s), 0 \leq s \leq t; \beta)},$$

où λ_T est une fonction positive déterministe inconnue, β est un paramètre de régression que l'on cherche à estimer, et g est une fonction du processus de covariable Z . L'estimation de β s'effectue par la méthode du maximum de vraisemblance, appliquée à une vraisemblance partielle qui ne contient pas la fonction λ_T considérée comme un paramètre de nuisance. Cette méthode nécessite l'observation, pour chaque sujet inclus dans l'étude de survie, de toutes les valeurs du processus de covariable Z jusqu'à la survenue de l'événement chez ce sujet.

Cependant, dans de nombreuses situations, la réalisation du processus Z au moment de l'événement ne peut être observée chez les sujets suivis. L'analyse par un modèle de Cox se heurte alors à un problème de données manquantes de Z , qui empêche le recours à l'inférence statistique basée sur la vraisemblance partielle.

Différentes méthodes ont été récemment proposées pour résoudre le problème de l'inférence dans un modèle de Cox avec covariables dépendant du temps manquantes : méthodes de la vraisemblance partielle approchée (Lin et Ying, [56]), de la vraisemblance partielle estimée (Zhou et Pepe, [96]), méthodes d'imputation de données manquantes (Paik et Tsai, [71]; Wang, Hsu, Feng et Prentice, [92]). Cependant, les conditions requises pour la validité de ces méthodes et des propriétés asymptotiques des estimateurs qui s'en déduisent sont contraignantes et ne sont pas toujours réunies dans la pratique de la modélisation des durées de vie.

L'objet de ce travail est de proposer une méthode d'estimation dans le modèle de Cox avec une covariable dépendant du temps manquante, qui soit applicable sous des conditions moins restrictives que les approches citées précédemment.

Dans la première partie de ce manuscrit, nous rappelons les définitions des outils de modélisation utilisés en analyse statistique des durées de vie, et la notion de censure. Puis nous présentons le modèle de Cox et nous rappelons les principaux résultats sur l'estimation de ses paramètres en cas de valeurs manquantes de covariables dépendant du temps.

Dans la deuxième partie, nous proposons une approche au problème de l'estimation dans le modèle de Cox avec une covariable dépendant du temps manquante. Elle consiste à modéliser conjointement les durées censurées et le processus de covariable pour en déduire une vraisemblance qui nous permette d'estimer le paramètre d'intérêt β (et aussi la fonction de risque cumulé $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$) au vu des données incomplètes. Dans le chapitre 4, nous décrivons le modèle conjoint proposé et nous établissons sa propriété d'identifiabilité.

Ce résultat nous permet ensuite de nous intéresser à l'estimation des paramètres du modèle conjoint, dont β et Λ_T . Dans le chapitre 5, nous adaptons à ce modèle la méthode du maximum de vraisemblance semiparamétrique qui, dans le cas du modèle de Cox où toutes les covariables sont observées, se ramène à estimer β par le maximum de vraisemblance partielle et Λ_T par l'estimateur de Breslow.

Dans le chapitre 5, nous montrons tout d'abord l'existence d'estimateurs semiparamétriques des paramètres du modèle conjoint. Nous obtenons ensuite une caractérisation de ces estimateurs en appliquant à la vraisemblance conjointe obtenue le principe de l'algorithme Espérance-Maximisation (EM).

Cette caractérisation est utile dans le chapitre 6, où nous montrons la consistance de ces estimateurs.

Nous nous intéressons ensuite au comportement en loi des estimateurs semiparamétriques dans le modèle conjoint. Nous montrons que la suite de ces estimateurs converge en loi vers un processus gaussien centré dont nous explicitons la fonction de covariance.

Nous proposons enfin un estimateur convergent de la variance asymptotique des estimateurs semiparamétriques.

Dans le chapitre 7, nous décrivons l'algorithme EM pour l'estimation des paramètres du modèle conjoint. Nous obtenons des formules explicites de calcul des valeurs itérées des estimations des paramètres de la loi de la covariable longitudinale et de Λ_T . Un algorithme numérique de type Newton est décrit pour estimer β .

Nous appliquons ensuite le modèle conjoint sur deux jeux de données réelles. Nous proposons tout d'abord de l'appliquer à l'analyse statistique des abandons de mesure informatifs dans les études longitudinales, et nous comparons ses résultats à ceux obtenus avec l'un des modèles les plus utilisés dans ce domaine, proposé par Diggle et Kenward [28]. Enfin, nous comparons les estimations du paramètre β du modèle de Cox, obtenues par l'approche conjointe et deux méthodes d'imputation.

Une partie des travaux présentés dans ce manuscrit a fait l'objet d'un article paru dans la revue *Lifetime Data Analysis* (2002).

Notations

Nous introduisons ici les notations générales communes aux différentes parties de ce manuscrit.

- Les variables aléatoires considérées seront définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.
- \mathbb{N} désigne l'ensemble des entiers naturels.
- \mathbb{R} l'ensemble des nombres réels.
- \mathbb{R}^+ l'ensemble des nombres réels positifs.
- \mathbb{R}^{+*} l'ensemble des nombres réels strictement positifs.
- \mathbb{R}^p l'espace vectoriel de dimension p sur \mathbb{R} .
- $|\cdot|$ la norme euclidienne sur \mathbb{R}^p .
- $a \wedge b$ le minimum de a et b .
- $a \vee b$ le maximum de a et b .
- $1_{\{A\}}$ l'indicatrice de A .
- $(x_1, \dots, x_n)_{\setminus x_i}$, ($i \in \{1, \dots, n\}$) l'ensemble des vecteurs (x_1, \dots, x_n) privés de l'élément x_i .
- $(x_1, \dots, x_n)_{\setminus x_i} \oplus \tilde{x}$, ($i \in \{1, \dots, n\}$) l'ensemble des vecteurs (x_1, \dots, x_n) où l'élément x_i est remplacé par la valeur fixée \tilde{x} .
- Soit T un ensemble :
 - $l^\infty(T)$ désigne l'espace des fonctions bornées de T dans \mathbb{R} .
 - $VB(T)$ l'espace des fonctions de T dans \mathbb{R} bornées et à variation bornée.
 - VB_p l'espace des fonctions de T dans \mathbb{R} bornées par p et à variation bornée par p .
 - si f est une fonction bornée de T dans \mathbb{R} , $\|f\|_\infty = \sup_{t \in T} |f(t)|$.
- Soit Ω un ensemble. $\mathcal{P}_b(\Omega)$ désigne l'ensemble des parties bornées de Ω .
- Soit A une matrice de dimension $n \times p$:

- A' désigne la transposée de A .
 - si $n = p$ et si A est inversible, A^{-1} désigne l'inverse de A .
-
- Si a est un vecteur de dimension p , $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ et $a^{\otimes 2} = aa'$.
-
- càdlàg : continue à droite et limitée à gauche.

Première partie

Préliminaires

Chapitre 1

Analyse statistique des durées de vie

L'analyse statistique des durées de vie a pour but l'étude des lois d'instantanéité d'occurrence d'événements, à partir d'observations de durées et éventuellement de variables explicatives, faites en continu ou de manière discrète dans le temps sur des individus. Dans ce premier chapitre, nous présentons les outils de modélisation utilisés en statistique des durées de vie, et nous rappelons le problème de la censure affectant les observations de durée de vies. Nous insistons plus particulièrement sur les censures dites aléatoires et de type I, qui interviendront dans ce travail. Enfin, nous expliquons comment la méthode du maximum de vraisemblance peut être appliquée au cas de données censurées.

1.1 Les outils probabilistes

Dans ce chapitre, nous désignons par T une variable aléatoire positive définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et représentant une durée jusqu'à un événement d'intérêt (cet événement peut être la mort, la guérison ou la rechute dans le domaine médical, la perte d'un emploi en économie, la panne en fiabilité... Par la suite, cette durée sera appelée *durée de vie*). Soit P_T la loi de T .

La fonction de répartition de T ou de sa loi P_T est la fonction sur \mathbb{R} définie par

$$F_T(t) = \mathbb{P}[T \leq t]$$

si $t \geq 0$ et $F_T(t) = 0$ si $t < 0$. La fonction de répartition F_T est croissante, continue à droite limitée à gauche, avec $\lim_{t \rightarrow +\infty} F_T(t) = 1$. Elle représente la probabilité

pour que l'événement d'intérêt survienne pendant l'intervalle de temps $[0, t]$.

Si le calcul des probabilités identifie en général la loi d'une variable aléatoire par sa densité ou sa fonction de répartition, l'interprétation d'une variable aléatoire en terme de durée de vie conduit à définir d'autres fonctions permettant de caractériser sa loi. Nous ne présentons pas toutes ces fonctions, seules les définitions des fonctions qui interviendront par la suite sont rappelées.

On appelle fonction de survie $S_T(t)$ la probabilité que la durée de vie T soit supérieure à une durée t :

$$S_T(t) = 1 - F_T(t) = \mathbb{P}[T > t].$$

Supposons que la loi de T admette une densité f_T par rapport à la mesure de Lebesgue sur \mathbb{R}^+ . On appelle fonction de risque instantané de la variable aléatoire T , la fonction définie presque partout en t tel que $S_T(t) > 0$ par :

$$\lambda_T(t) = -\frac{d}{dt} \log(S_T(t)) = \frac{f_T(t)}{S_T(t)}.$$

On adoptera la convention $c/0 = +\infty, \forall c \geq 0$, aboutissant à poser $\lambda_T(t) = +\infty$ pour les valeurs de t telles que $S_T(t) = 0$.

Supposons que la densité f_T de T soit continue en t , et que t soit tel que $S_T(t) > 0$. En remarquant que

$$\frac{1}{\Delta t} \mathbb{P}[t < T \leq t + \Delta t | T > t] = \frac{1}{S_T(t)} \cdot \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t}$$

et que $\frac{d}{dt} F_T(t) = f_T(t)$, nous obtenons la relation

$$\lambda_T(t) = \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}[t < T \leq t + \Delta t | T > t]}{\Delta t}.$$

La fonction de risque instantané est une fonction positive, et $\lambda_T(t) \cdot \Delta t$ s'interprète comme la probabilité de survenue de l'événement d'intérêt dans l'intervalle $]t, t + \Delta t]$ sachant que cet événement ne s'est pas encore produit à l'instant t . Cette fonction traduit donc l'évolution dans le temps du risque de survenue de l'événement.

On définit, pour tout $t \in \mathbb{R}^+$, la fonction de risque cumulé par

$$\Lambda_T(t) = \int_0^t \lambda_T(s) ds,$$

avec $\Lambda_T(t) = +\infty$ lorsque $S_T(t) = 0$. Pour presque tout $t \in \mathbb{R}^+$ tel que $S_T(t) > 0$, on a la relation suivante :

$$f_T(t) = \lambda_T(t) \exp \left[- \int_0^t \lambda_T(s) ds \right].$$

Il suffit donc de connaître l'une de ces cinq fonctions pour en déduire les autres. La fonction λ_T est généralement la plus intéressante. En effet, elle fournit une description probabiliste du futur immédiat d'un sujet encore observé et reflète des différences entre les modèles qui seraient plus difficiles à formuler en termes de fonction de survie, de répartition ou de densité. C'est l'interprétation de cette fonction qui permet le plus souvent de guider le choix d'un modèle pour des données de durée.

Nous illustrons les définitions précédentes avec les exemples des distributions IFR (*increasing failure rate*), DFR (*decreasing failure rate*), et de deux modèles paramétriques d'analyse des durées de vie parmi les plus utilisés.

Exemples

1. On appelle fonction de survie conditionnelle à la date t la probabilité que T dépasse $t + x$ sachant que T dépasse t :

$$\mathbb{P}[T > t + x | T > t] = \frac{S_T(t + x)}{S_T(t)}.$$

La variable aléatoire T , ou sa loi P_T , est dite IFR (respectivement DFR) si $\forall x > 0$, la fonction

$$t \longmapsto \frac{S_T(t + x)}{S_T(t)}$$

est décroissante (respectivement croissante).

Supposons que T admette une densité f_T continue sur \mathbb{R}^{+*} . En remarquant que λ_T peut s'écrire

$$\lambda_T(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \left[1 - \frac{S_T(t + \Delta t)}{S_T(t)} \right],$$

il revient au même de dire que T est IFR (respectivement DFR) si sa fonction de risque instantané λ_T est croissante (respectivement décroissante).

De plus amples développements sur les distributions IFR, DFR, sur d'autres

notions de vieillissement ou de rajeunissement (NBU : *New Better than Used*, NWU : *New Worse than Used*, ...), et sur les relations existant entre ces différentes notions, peuvent être obtenus dans les ouvrages de Aven et Jensen [2], Barlow et Proschan [10], Coccoza-Thivent [18], Gertsbakh [38].

2. Dans une étude de survie, si nous pouvons supposer qu'il n'y a pas d'effet d'usure, ou de vieillissement, la probabilité de survenue de l'événement d'intérêt à l'instant t , sachant que l'événement ne s'est pas encore produit, est la même pour tout t . La loi de la durée T jusqu'à cet événement est donc caractérisée par une fonction de risque instantané constante :

$$\lambda_T(t) = \lambda, t > 0, \lambda > 0.$$

Les fonctions f_T, F_T, S_T et Λ_T se déduisent des relations énoncées ci-dessus :

$$f_T(t) = \lambda e^{-\lambda t}, F_T(t) = 1 - e^{-\lambda t}, S_T(t) = e^{-\lambda t} \text{ et } \Lambda_T(t) = \lambda t, t > 0.$$

La loi de la variable aléatoire T est donc une loi exponentielle. Le modèle exponentiel est très utilisé dans le domaine de la fiabilité (analyse statistique des durées de vies dans le contexte industriel). Il est étudié dans ce contexte par Meeker et Escobar [63], Voinov et Nikulin [90]. Ce modèle fut historiquement l'un des premiers modèles d'analyse des durées de vie étudié (Johnson, Kotz et Balakrishnan [48], Klein et Moeschberger [50] et Lawless [54] fournissent des synthèses bibliographiques assez exhaustives de ces premiers travaux). Il repose néanmoins sur l'hypothèse parfois trop restrictive d'une fonction de risque instantané constante.

3. L'introduction d'un nouveau paramètre dans la fonction λ_T permet d'obtenir différentes formes de la fonction de risque.

La loi de Weibull à deux paramètres est définie par la fonction de risque suivante :

$$\lambda_T(t) = \alpha \lambda t^{\alpha-1}, t > 0, \alpha > 0, \lambda > 0.$$

Les paramètres λ et α sont respectivement appelés paramètre d'échelle et de forme. La position de α par rapport à 1 permet en effet de spécifier différentes formes de la fonction de risque instantané. Si $\alpha = 1$, la fonction λ_T est constante et le modèle de Weibull se ramène au modèle exponentiel. Si $\alpha < 1$ (respectivement $\alpha > 1$), λ_T est décroissante dans le temps (respectivement croissante). Par exemple, si $\alpha > 1$, la probabilité conditionnelle de survenue d'un événement à l'instant t , sachant qu'il ne s'est pas encore produit à t ,

croît avec t : il y a phénomène d'usure, ou de vieillissement.

Les fonctions f_T, F_T, S_T et Λ_T , d'expressions relativement simples, sont obtenues à partir de λ_T par les formules suivantes, pour tout $t > 0$:

$$f_T(t) = \alpha \lambda t^{\alpha-1} e^{-\lambda t^\alpha}, F_T(t) = 1 - e^{-\lambda t^\alpha}, S_T(t) = e^{-\lambda t^\alpha}, \Lambda_T(t) = \lambda t^\alpha.$$

La loi de Weibull est très largement utilisée dans les domaines de la fiabilité et de l'analyse des durées de vie en médecine [50, 54]. Deux raisons de son succès sont que cette loi a été obtenue au cours d'études expérimentales portant sur le risque instantané de défaillance de matériels, et qu'elle a pour cas particuliers des lois IFR, DFR et la loi exponentielle. Une synthèse bibliographique assez exhaustive des nombreux travaux portant sur la loi de Weibull est donnée par Johnson, Kotz et Balakrishnan [48].

Bagdonavičius et Nikulin [8], Carbon [15], Cox et Oakes [22], Johnson, Kotz et Balakrishnan [48], Kalbfleisch et Prentice [49], Klein et Moeschberger [50], Lawless [54], Meeker et Escobar [63], Voinov et Nikulin [90, 91] décrivent de nombreux autres modèles paramétriques d'analyse des durées de vie. Ces auteurs donnent aussi d'autres fonctions caractéristiques de la loi d'une variable aléatoire dont les interprétations peuvent être utiles dans l'analyse statistique des durées de vie.

Prise en compte de covariables. Modèle de régression paramétrique

Dans la plupart des domaines d'application de l'analyse des durées de vie, on dispose, en plus des observations de durées sur un ensemble d'individus, d'informations supplémentaires sous la forme de variables explicatives individuelles observées (covariables). Ces covariables peuvent être fixes dans le temps (sexe, catégorie socio-professionnelle, appartenance à une population à risque,...) ou au contraire dépendre du temps (mesures répétées d'une quantité biologique,...), ce qui est le cas lorsque l'on souhaite évaluer l'influence de l'évolution au cours du temps d'un phénomène individuel, sur la durée jusqu'à survenue d'un événement.

Nous considérons maintenant la paire (T, Z) , où T est une durée de vie aléatoire et $Z = (Z_1, \dots, Z_p)'$ est un vecteur de p variables explicatives réelles mesurées sur chaque sujet. Z_1, \dots, Z_p sont aussi appelées des covariables.

Ces covariables peuvent dépendre du temps. Dans ce cas, Z est considéré comme un processus p -dimensionnel $(Z(t))_{t \geq 0}$ indexé par le temps, et le vecteur $Z(t) = (Z_1(t), \dots, Z_p(t))'$ contient les valeurs des covariables à l'instant t . Le processus $(Z(t))_{t \geq 0}$ n'est observé que sur l'intervalle de temps $[0, T]$.

En analyse statistique des durées de vie, la famille des lois possibles pour (T, Z) est généralement déterminée en choisissant une loi conditionnelle pour T sachant Z . Notons $\mathcal{E}_t = \sigma(Z; T > t)$ la σ -algèbre engendrée par Z et l'événement $\{T > t\}$, et notons $\mathcal{F}_t = \sigma(Z(s) : 0 \leq s \leq t; T > t)$ la σ -algèbre engendrée par $\{Z(s) : 0 \leq s \leq t\}$ et $\{T > t\}$.

La loi conditionnelle de T sachant Z est caractérisée par la fonction de risque instantané conditionnelle $\lambda_{T|Z}$, définie, sous réserve de conditions générales de régularité, par

$$\lambda_{T|Z}(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}[t < T \leq t + \Delta t | \mathcal{E}_t]$$

si Z est fixe dans le temps, et par

$$\lambda_{T|Z}(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}[t < T \leq t + \Delta t | \mathcal{F}_t]$$

si Z dépend du temps.

Les deux familles principales de modèles ayant été proposés dans la littérature pour prendre en compte la présence de covariables dans l'analyse statistique des durées de vie sont les modèles *additifs* et les modèles *multiplicatifs*. Les modèles additifs ne seront pas abordés, le lecteur intéressé pourra se référer aux travaux de Bagdonavičius et Nikulin [3, 5, 8], Klein et Moeschberger [50], Lin et Ying [57] (on trouvera dans l'ouvrage très complet [8] une bibliographie détaillée de récents travaux sur les modèles additifs).

Les modèles multiplicatifs sont définis par une fonction de risque instantané conditionnelle $\lambda_{T|Z}$ de la forme

$$\lambda_{T|Z}(t) = \lambda_T(t)h(Z; \beta),$$

où λ_T est une fonction positive appelée fonction de risque de base, $h(Z; \beta)$ est une fonction de régression sur Z , et β est un vecteur de paramètres inconnus. La fonction $h(Z; \beta)$ doit être positive, on choisit généralement $h(Z; \beta) = e^{\beta'Z}$.

Les modèles paramétriques usuels d'analyse des durées de vie peuvent être adaptés pour prendre en compte des covariables fixes ou dépendant du temps. L'inférence statistique dans ces modèles, devenus des modèles de régression, l'analyse des résidus ainsi que d'autres méthodes de diagnostics, sont discutées, entre autres, par Bagdonavičius et Nikulin ([8], chapitre 5), Kalbfleisch et Prentice ([49], chapitre 3), Klein et Moeschberger ([50], chapitre 12), Lawless ([54], chapitre 6),

Meeker et Escobar ([63], chapitre 17).

Exemple

Le modèle multiplicatif permettant d'inclure des covariables dans le modèle de Weibull est défini par

$$\lambda_{T|Z}(t) = \alpha \lambda t^{\alpha-1} e^{\beta'Z}, \quad t > 0, \alpha > 0, \lambda > 0,$$

où $\beta, Z \in \mathbb{R}^p$, $\alpha \lambda t^{\alpha-1}$ s'interprète comme la fonction de risque de base, c'est-à-dire le risque instantané d'occurrence de l'événement pour un individu tel que $Z = 0$, et (α, λ, β) est le paramètre à estimer.

Les fonctions conditionnelles de densité, de répartition, de survie et de risque cumulé qui se déduisent de $\lambda_{T|Z}$ sont respectivement, pour $t > 0$:

$$\begin{aligned} f_{T|Z}(t) &= \alpha \lambda t^{\alpha-1} e^{\beta'Z} e^{-\lambda t^\alpha \exp(\beta'Z)}, & F_{T|Z}(t) &= 1 - e^{-\lambda t^\alpha \exp(\beta'Z)}, \\ S_{T|Z}(t) &= e^{-\lambda t^\alpha \exp(\beta'Z)}, & \Lambda_{T|Z}(t) &= \lambda t^\alpha e^{\beta'Z}. \end{aligned}$$

1.2 Modèles de censure

1.2.1 Mécanismes de censure

L'analyse des durées de vie pose des problèmes particuliers, dus au fait que les observations de durées de vie sont le plus souvent censurées, c'est-à-dire privées d'une partie de l'information qu'elle contiennent. Nous restreignons notre présentation des mécanismes de censure aux censures à droite, et parmi celles-ci, aux censures dites aléatoires et de type I, qui seront traitées dans ce travail. Des exposés détaillés d'autres types de censures sont proposés par Bagdonavičius et Nikulin ([8], chapitre 4), Huber [44, 45], Klein et Moeschberger ([50], chapitre 3) et Lawless ([54], chapitre 1). Dans la suite, T désigne une durée de vie aléatoire.

Les modèles de censure font intervenir, en sus de la variable T , une variable aléatoire auxiliaire C à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$ dite de *censure*. Dans le modèle de *censure aléatoire à droite*, on observe $X = T \wedge C$, ainsi que l'indicatrice $\Delta = 1_{\{T \leq C\}}$ qui vaut 1 si $T \leq C$ et 0 sinon. Le modèle de *censure aléatoire à gauche* est similaire : on observe $Z = T \vee C$ ainsi que l'indicatrice $1_{\{T < C\}} = 1 - 1_{\{T \geq C\}}$.

De manière générale, on parlera de *censure à droite* lorsqu'on observera conjointement $1_{\{T \leq C\}}$ et une variable $\psi(T, C)$ fonction de T et C , et de *censure à gauche* lorsqu'on

observera $1_{\{T < C\}}$ et $\psi(T, C)$.

Le cas particulier où C est une variable dégénérée (une constante) donne la *censure de type I*. Ce cas correspond à la situation où l'observation des durées de vie s'arrête au bout d'un temps fixé (par exemple, on observe la durée de fonctionnement de matériels sur une période de temps dont la durée est fixée à l'avance). Lorsque ce temps de censure, tout en étant fixé à l'avance, est différent selon les observations, on parle de *censure progressive à droite de type I*.

Dans ce qui suit, nous considérerons le modèle de censure aléatoire à droite, qui, lorsque la variable de censure C est déficiente et égale à $+\infty$ avec la probabilité 1, se résume à l'observation non censurée de la durée de vie T .

1.2.2 Estimation du maximum de vraisemblance dans un modèle de survie paramétrique avec censure à droite

Nous rappelons brièvement la méthode du maximum de vraisemblance pour estimer, au vu de données censurées, les paramètres réels d'un modèle d'analyse des durées de vie. Des exposés détaillés sont données par Bagdonavičius et Nikulin ([8], chapitre 4; [9]), Kalbfleisch et Prentice ([49], chapitre 3), Lawless ([54], chapitres 3-6), Meeker et Escobar ([63], chapitres 7, 8, 11), Pons et Huber [73], Voinov et Nikulin [90, 91].

Maximum de vraisemblance dans un modèle de survie paramétrique

Soit T une durée de vie aléatoire. On suppose que la loi $P_T = P_{\theta_0}$ de T appartient à une famille de lois de probabilité $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$, avec $\theta_0 \in \Theta$. Ici, on suppose que Θ est un ouvert de \mathbb{R}^p .

Les notations $f_T(t; \theta)$, $F_T(t; \theta)$, $S_T(t; \theta)$ et $\lambda_T(t; \theta)$ désignent respectivement la densité, la fonction de répartition, la fonction de survie et la fonction de risque instantané de T sous la loi P_θ .

Soit C un temps de censure, défini sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ que T . On caractérise la loi P_C de C par $F_C(t) = \mathbb{P}[C \leq t]$, $S_C(t) = 1 - F_C(t)$, et, lorsque cette expression a un sens $f_C(t) = \frac{d}{dt}F_C(t)$. On suppose que ces fonctions ne dépendent pas de θ : cette hypothèse est appelée hypothèse de *censure non-informative*. On suppose aussi que T et C sont indépendants.

Les observations sont des réalisations de la variable $X = T \wedge C$ et de l'indicatrice de censure $\Delta = 1_{\{T \leq C\}}$. Soit $(X_i, \Delta_i)_{1 \leq i \leq n}$ un échantillon indépendant des variables

(X, Δ) . Nous nous intéressons au problème de l'estimation de θ_0 par la méthode du maximum de vraisemblance, à partir de ces observations.

Sous l'hypothèse que $S_C(t) = S_C(t; \theta)$ et $f_C(t) = f_C(t; \theta)$ dépendent de $\theta \in \Theta$, la vraisemblance de θ au vu du n -échantillon $(X_i, \Delta_i)_{1 \leq i \leq n}$ est égale à :

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n [f_T(X_i; \theta) S_C(X_i; \theta)]^{\Delta_i} [S_T(X_i; \theta) f_C(X_i; \theta)]^{1-\Delta_i} \\ &= \prod_{i=1}^n \lambda_T(X_i; \theta)^{\Delta_i} S_T(X_i; \theta) S_C(X_i; \theta)^{\Delta_i} f_C(X_i; \theta)^{1-\Delta_i}. \end{aligned}$$

Sous l'hypothèse de censure non-informative (c'est-à-dire si S_C et f_C ne dépendent pas de θ), il est équivalent de chercher l'estimateur du maximum de vraisemblance de θ_0 à partir du facteur suivant de la fonction de vraisemblance :

$$\prod_{i=1}^n \lambda_T(X_i; \theta)^{\Delta_i} S_T(X_i; \theta),$$

obtenu en retirant de $L_n(\theta)$ les termes relatifs à la loi de la censure.

Maximum de vraisemblance dans un modèle de survie paramétrique avec covariables

Supposons que la loi conditionnelle $P_{T|Z} = P_{\theta_0, Z}$ de T sachant Z appartienne à une famille de lois de probabilité $\mathcal{P}_Z = \{P_{\theta, Z} : \theta \in \Theta \subseteq \mathbb{R}^p\}$, avec $\theta_0 \in \Theta$. On suppose que Θ est un ouvert de \mathbb{R}^p .

Les notations $f_{T|Z}(t; \theta)$, $F_{T|Z}(t; \theta)$, $S_{T|Z}(t; \theta)$ et $\lambda_{T|Z}(t; \theta)$ désignent respectivement la densité, la fonction de répartition, la fonction de survie et la fonction de risque instantané de T , conditionnelles à Z , sous la loi $P_{\theta, Z}$. On suppose que Z admet une densité $f_Z(Z; \phi)$ qui dépend d'un paramètre ϕ .

On dispose d'un n -échantillon $(X_i, \Delta_i, Z_i)_{1 \leq i \leq n}$ du triplet (X, Δ, Z) .

Les hypothèses usuelles de censure indépendante et non-informative sont reprises : on suppose que T et C sont indépendants conditionnellement à Z , et que C est non-informatif pour les paramètres de la loi conditionnelle de T sachant Z . L'estimateur du maximum de vraisemblance de θ_0 est alors obtenu à partir de la vraisemblance

$$\prod_{i=1}^n \lambda_{T|Z}(X_i; \theta)^{\Delta_i} S_{T|Z}(X_i; \theta) f_Z(Z_i; \phi).$$

Si la loi de T sachant Z et la loi de Z n'ont pas de paramètre en commun, il est équivalent, pour faire de l'inférence sur θ , d'utiliser l'expression

$$\prod_{i=1}^n \lambda_{T|Z}(X_i; \theta)^{\Delta_i} S_{T|Z}(X_i; \theta).$$

Ceci peut s'étendre au cas de covariables dépendant du temps en remplaçant Z par $\{Z(s), 0 \leq s \leq t\}$.

Chapitre 2

Le modèle semiparamétrique de Cox

Le modèle de régression de Cox définit une famille de lois conditionnelles d'instants d'occurrence d'événements, sachant un ensemble de covariables fixes ou qui dépendent du temps. C'est un modèle semiparamétrique : les lois sont paramétrées par un vecteur de paramètres de régression et par une fonction positive inconnue. Ce modèle est donc moins contraignant que les modèles paramétriques usuels d'analyse des durées de vie.

Dans ce chapitre, nous présentons tout d'abord le modèle de Cox avec covariables fixes et covariables dépendant du temps. Puis nous présentons le principe des méthodes du maximum de vraisemblance partielle et du maximum de vraisemblance semiparamétrique, et nous rappelons les propriétés asymptotiques des estimateurs obtenus. Nous présentons enfin l'estimateur de Breslow de la fonction de risque cumulé, et le traitement des *ex aequo* parmi les instants d'événements.

2.1 Présentation générale

Nous définissons tout d'abord les notations utilisées dans ce chapitre et nous formulons le modèle de Cox. Dans ce chapitre, certains développements seront valables à la fois pour des covariables fixes et des covariables dépendant du temps, certains autres ne le seront que pour des covariables fixes. Nous les signalerons le cas échéant.

Les modèles paramétriques de régression décrits dans le chapitre 1 permettent de modéliser des données de durée qui peuvent être censurées, en présence de cova-

riables. Le choix d'un modèle paramétrique particulier n'est cependant pas toujours aisé en pratique. Il peut ne pas être utile non plus : si le paramètre d'intérêt est le paramètre de régression, les autres paramètres du modèle apparaissent comme des paramètres de nuisance. Il peut donc être intéressant de ne définir le modèle que partiellement, en ne spécifiant précisément que la partie relative aux paramètres de régression : c'est l'approche semiparamétrique.

Le modèle semiparamétrique le plus utilisé pour analyser des durées de vie avec des covariables est le modèle de Cox [20]. Il définit la fonction de risque instantané d'un événement, conditionnelle à un ensemble de covariables fixes ou dépendant du temps, comme le produit d'une fonction positive déterministe inconnue $\lambda_T(t)$ ($t \geq 0$) et d'une fonction de régression sur l'ensemble de ces covariables.

Pour que $\lambda_{T|Z}$ soit positive, la fonction de régression choisie est la fonction exponentielle. Cox a formulé son modèle original dans le cas de covariables fixes, en définissant la fonction de risque instantanée suivante :

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta'Z}, \quad (2.1)$$

où $\beta = (\beta_1, \dots, \beta_p)$ est un vecteur de p coefficients inconnus. La fonction λ_T est appelée fonction de risque instantané de base (ou taux de base) et est considérée comme un paramètre de nuisance, l'intérêt résidant en général dans l'estimation du paramètre de régression β .

Si Z dépend du temps, la fonction de risque s'écrit sous la forme suivante :

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{g(Z(s), 0 \leq s \leq t; \beta)}, \quad (2.2)$$

où $g(Z(s), 0 \leq s \leq t; \beta)$ est une fonction de β et $\{Z(s), 0 \leq s \leq t\}$.

Le lecteur intéressé par le modèle de Cox avec des covariables dépendant du temps pourra se référer aux exposés détaillés de Bagdonavičius et Nikulin [8], Huber [45], Kalbfleisch et Prentice [49], Klein et Moeschberger [50], Lawless [54], Marubini et Valsecchi [61], Pons et Huber [73], Therneau et Grambsch [82]. Il ressort de leur lecture que la fonction $g(Z(s), 0 \leq s \leq t; \beta)$ la plus souvent utilisée dans les applications est $\beta'Z(t)$: le risque instantané d'occurrence de l'événement, à l'instant t , dépend du processus Z au travers de sa valeur en t .

Dans ce chapitre, pour simplifier les notations, les développements relatifs au modèle de Cox avec des covariables dépendant du temps seront donc obtenus à partir de la formulation suivante de la fonction $\lambda_{T|Z}$:

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta'Z(t)}. \quad (2.3)$$

Marubini et Valsecchi [61] notent néanmoins qu'une autre fonction intéressante est

$$g(Z(s), 0 \leq s \leq t; \beta) = \beta'(Z(t) - Z(t - h)), \quad h > 0.$$

Ce choix de g permet d'évaluer l'influence de l'évolution du processus de covariable Z entre les instants $(t - h)$ et t sur le risque de survenue de l'événement d'intérêt en t .

Bagdonavičius et Nikulin [4, 6] ont proposé un modèle où le risque d'occurrence d'un événement à l'instant t dépend de la valeur de Z à cet instant et d'un processus de dégradation de Z , qui est une fonction de la quantité $\int_0^t e^{\beta'Z(s)} ds$, appelée accumulation additive des dommages.

Lorsque Z est fixe dans le temps, le modèle de Cox appartient à la famille des modèles à risques proportionnels : si Z_1 et Z_2 désignent deux vecteurs de covariables, le rapport des risques instantanés $\lambda_T(t)e^{\beta'Z_1} / \lambda_T(t)e^{\beta'Z_2}$ est constant pour tout t , égal à $e^{\beta'(Z_1 - Z_2)}$. Par exemple, si la i -ème composante du vecteur Z prend ses valeurs dans $\{0, 1\}$, e^{β_i} peut s'interpréter comme le rapport des risques instantanés de deux individus (numérotés respectivement 1 et 2) pour lesquels $Z_{1,i}$ et $Z_{2,i}$ valent respectivement 1 et 0 et tels que $Z_{1,j} = Z_{2,j}$ ($j \neq i$).

Notons que le taux de base λ_T peut s'interpréter comme la fonction de risque instantané en l'absence d'influence de la covariable, et que le taux de base cumulé $\Lambda_T = \int_0^t \lambda_T(s) ds$ est une fonction positive, croissante, continue, et telle que $\Lambda_T(0) = 0$. Le modèle de Cox comprend comme cas particuliers la plupart des modèles paramétriques usuels, obtenus pour des spécifications particulières du taux de base λ_T . En particulier, si $\lambda_T(t) = \lambda$ (respectivement $\lambda_T(t) = \alpha \lambda t^{\alpha-1}$), le modèle de Cox devient le modèle exponentiel (respectivement de Weibull).

Le problème statistique consiste à estimer le paramètre vectoriel β à partir d'un n -échantillon $(X_i, \Delta_i, Z_i)_{1 \leq i \leq n}$ du triplet (X, Δ, Z) . Il pourra également être intéressant d'estimer le paramètre fonctionnel $\Lambda_T = \int_0^t \lambda_T(s) ds$. L'estimation est abordée dans la section suivante.

2.2 Méthode du maximum de vraisemblance partielle

2.2.1 Introduction

Cox [20] a proposé de baser l'inférence statistique pour les modèles (2.1) et (2.2), qui contiennent un paramètre de nuisance fonctionnel λ_T , sur une vraisemblance partielle ne contenant pas ce paramètre. Plus généralement, Cox [21] a proposé l'utilisation d'une vraisemblance partielle pour les modèles contenant des paramètres de nuisance fonctionnels ou de grande dimension.

Nous rappelons tout d'abord le principe de cette méthode et nous l'appliquons ensuite pour construire la vraisemblance partielle dite de Cox, pour un échantillon de durées et de covariables. Pour un exposé plus complet, le lecteur pourra se référer à Bagdonavičius et Nikulin ([8], chapitre 7), Cox [21], Fleming et Harrington ([37], chapitre 4), Kalbfleisch et Prentice ([49], chapitres 4 et 5) et Pons et Huber [73].

Soit X un vecteur aléatoire de densité $f_X(x; \theta)$, où θ est un vecteur de paramètres décomposable en $\theta = (\phi, \beta)$. Supposons que l'on s'intéresse à β , paramètre d'intérêt, et non à ϕ , paramètre de nuisance. Par exemple dans le modèle de Cox, qui définit une famille de lois paramétrées par un vecteur β et une fonction λ_T , β est le paramètre d'intérêt et λ_T est le paramètre de nuisance.

Une solution consiste à estimer β et ϕ mais à n'utiliser de la matrice de covariance asymptotique de l'estimateur de θ que la sous-matrice relative à β , pour faire de l'inférence statistique sur β . Cependant, lorsque le paramètre de nuisance est une fonction, l'emploi de cette méthode est rendu difficile (problème de maximisation de la vraisemblance).

Supposons que X puisse se décomposer en (V, W) et que la densité de X puisse être réécrite comme le produit

$$f_X(x; \theta) = f_{W|V}(w|v; \theta) \cdot f_V(v; \theta). \quad (2.4)$$

Un exemple de décomposition d'un vecteur d'observations X est obtenu pour le vecteur V des éléments de X ordonnés par ordre croissant et le vecteur W des rangs de ces éléments.

Si l'un des facteurs du produit (2.4) ne dépend pas de ϕ , on peut penser baser l'inférence statistique pour β sur ce seul facteur. Ceci entraîne une perte d'information si le terme ignoré dépend de β . Néanmoins, s'il ne contient que peu d'informa-

tion sur β , cette perte peut apparaître comme un inconvénient mineur au regard de la simplicité de l'inférence développée sur le terme retenu. Ceci est particulièrement vrai lorsque le paramètre de nuisance est une fonction.

Le principe de l'estimation par maximum de vraisemblance partielle proposé par Cox [21] est basé sur cette idée.

Supposons plus généralement que le vecteur X puisse se décomposer sous la forme

$$(V_1, W_1, V_2, W_2, \dots, V_N, W_N)$$

d'une suite de paires (V_i, W_i) , $i = 1, \dots, N$. La vraisemblance de θ peut alors s'écrire :

$$\begin{aligned} f_X(x; \theta) &= f_{V_1, W_1, \dots, V_N, W_N}(v_1, w_1, \dots, v_n, w_n; \theta) \\ &= \prod_{i=1}^N f_{W_i|V_1, W_1, \dots, V_i}(w_i|v_1, w_1, \dots, v_{i-1}, w_{i-1}, v_i; \theta) \\ &\quad f_{V_i|V_1, W_1, \dots, V_{i-1}, W_{i-1}}(v_i|v_1, w_1, \dots, v_{i-1}, w_{i-1}; \theta) \\ &= \left[\prod_{i=1}^N f_{W_i|Q_i}(w_i|q_i; \theta) \right] \cdot \left[\prod_{i=1}^N f_{V_i|P_i}(v_i|p_i; \theta) \right], \end{aligned} \quad (2.5)$$

où $P_1 = \{\emptyset\}$, $Q_1 = V_1$ et pour $i = 2, \dots, N$,

$$P_i = (V_1, W_1, \dots, V_{i-1}, W_{i-1}) \quad (2.6)$$

et

$$Q_i = (V_1, W_1, \dots, W_{i-1}, V_i). \quad (2.7)$$

Si le premier facteur de (2.5) dépend uniquement de β , Cox [21] l'appelle vraisemblance partielle de β basée sur W , et souligne son intérêt pratique pour l'estimation de β lorsque son expression est plus simple que celle de $f_X(x; \theta)$.

Cependant, l'emploi de la méthode du maximum de vraisemblance partielle pose la question, par exemple, de la consistance, ou de la normalité asymptotique des estimateurs obtenus. Pour le modèle de Cox, Tsiatis [84] et Andersen et Gill [1] ont montré, respectivement dans le cas de covariables fixes et de covariables dépendant du temps, que l'estimateur du maximum de vraisemblance partielle du paramètre

de régression possède les propriétés usuelles d'un estimateur du maximum de vraisemblance. Plus généralement, Wong [93] a étudié les propriétés asymptotiques des estimateurs obtenus à partir d'une vraisemblance partielle.

La méthode du maximum de vraisemblance partielle est particulièrement adaptée lorsqu'il existe un ordre naturel pour les observations, tel qu'un ordre dans le temps. Ceci permet en effet de décomposer un vecteur d'observations X en une suite $(V_1, W_1, \dots, V_N, W_N)$ d'instantanés d'événements ordonnés.

Nous appliquons cette méthode pour construire la vraisemblance partielle proposée par Cox [20] pour des instantanés d'événements censurés et des covariables.

2.2.2 La vraisemblance partielle de Cox

Nous considérons le triplet (X, Δ, Z) , où $X = T \wedge C$, $\Delta = 1_{\{T \leq C\}}$ et Z est un p -vecteur ou un processus p -dimensionnel de covariables. Nous notons S_C la fonction de survie de C , f_C la densité de C et f_Z la densité de Z .

Nous supposons que T et C sont indépendants conditionnellement à Z . Nous supposons que la fonction de risque instantané est $\lambda_T(t)e^{\beta'Z}$ si Z est fixé, et $\lambda_T(t)e^{\beta'Z(t)}$ si Z dépend du temps.

La densité de (X, Δ, Z) s'écrit

$$\left[\lambda_T(X) e^{\beta'Z} e^{-\Lambda_T(X) \exp(\beta'Z)} S_{C|Z}(X) \right]^\Delta \left[e^{-\Lambda_T(X) \exp(\beta'Z)} f_{C|Z}(X) \right]^{1-\Delta} f_Z(Z)$$

si Z est fixé, et

$$\left[\lambda_T(X) e^{\beta'Z(X)} e^{-\int_0^t \exp(\beta'Z(s)) d\Lambda_T(s)} S_{C|Z}(X) \right]^\Delta \left[e^{-\int_0^t \exp(\beta'Z(s)) d\Lambda_T(s)} f_{C|Z}(X) \right]^{1-\Delta} f_Z(Z)$$

si Z dépend du temps.

Sous l'hypothèse supplémentaire de censure non-informative et si la distribution de Z ne dépend pas des paramètres β et Λ_T , les termes impliquant les distributions de C et Z peuvent être ignorés pour estimer β et Λ_T par la méthode du maximum de vraisemblance. La vraisemblance obtenue pour estimer β et Λ_T au vu d'un n -échantillon $(X_i, \Delta_i, Z_i)_{1 \leq i \leq n}$ s'écrit donc

$$L_n(\beta, \Lambda_T) = \prod_{i=1}^n \left[\lambda_T(X_i) e^{\beta'Z_i} e^{-\Lambda_T(X_i) \exp(\beta'Z_i)} \right]^{\Delta_i} \left[e^{-\Lambda_T(X_i) \exp(\beta'Z_i)} \right]^{1-\Delta_i} \quad (2.8)$$

ou

$$L_n(\beta, \Lambda_T) = \prod_{i=1}^n \left[\lambda_T(X_i) e^{\beta' Z_i(X_i)} e^{-\int_0^t \exp(\beta' Z_i(s)) d\Lambda_T(s)} \right]^{\Delta_i} \left[e^{-\int_0^t \exp(\beta' Z_i(s)) d\Lambda_T(s)} \right]^{1-\Delta_i}. \quad (2.9)$$

selon que Z est fixe (2.8) ou dépend du temps (2.9).

Estimer le paramètre $\Lambda_T = \int_0^t \lambda_T(s) ds$ relève d'un problème d'estimation fonctionnelle. Il n'existe pas de maximum de (2.8) ou (2.9) lorsque λ_T appartient à l'ensemble des fonctions positives sur \mathbb{R}^+ (Van der Vaart, [87]). La méthode d'estimation du maximum de vraisemblance ne peut pas être appliquée à (2.8) et (2.9).

Cox a donc proposé d'estimer β à partir d'une vraisemblance partielle obtenue par le principe général décrit en introduction de cette section. Nous décrivons la construction de cette vraisemblance partielle et, pour ne pas alourdir les notations, nous traitons le cas de covariables fixes. L'argument se généralise aux covariables dépendant du temps (on se référera aux ouvrages de Fleming et Harrington [37] et Kalbfleisch et Prentice [49]).

Nous disposons d'un n -échantillon $(X_i, \Delta_i, Z_i)_{1 \leq i \leq n}$ du triplet (X, Δ, Z) . Supposons que l'on observe L instants d'occurrence de l'événement d'intérêt (ces instants sont supposés distincts). Nous les réordonnons par ordre croissant : $T_{(1)} < \dots < T_{(L)}$, où (l) ($l = 1, \dots, L$) désigne l'indice de l'observation i ($i = 1, \dots, n$) telle que $T_i = T_{(l)}$. Notons $Z_{(l)}$ le vecteur de covariables associé à cette observation. Supposons que m_i censures se produisent dans l'intervalle $[T_{(i)}, T_{(i+1)})$ et notons $X_{(i,1)}, \dots, X_{(i,m_i)}$ ces instants, auxquels sont associés les vecteurs de covariables $Z_{(i,1)}, \dots, Z_{(i,m_i)}$. Dans un souci de clarté des formules, nous serons amenés, dans la suite de ce chapitre, à utiliser parfois les notations $f_{T|Z}(t|Z)$, $S_{T|Z}(t|Z)$, $\lambda_{T|Z}(t|Z)$ au lieu des notations $f_{T|Z}(t)$, $S_{T|Z}(t)$, $\lambda_{T|Z}(t)$ adoptées jusqu'ici.

Reprenant les notations de l'introduction de cette section, nous définissons, pour $i = 1, \dots, L$:

$$V_i = \{Z_l : l = 1, \dots, n; T_{(i)}, X_{(i-1,j)} : j = 1, \dots, m_{i-1}\}$$

et

$$W_i = \{(i)\} = \{j : T_j = T_{(i)}\}.$$

En utilisant la décomposition (2.6), nous pouvons définir une suite P_i , qui contient tous les instants de censure jusqu'à $T_{(i-1)}$, tous les instants d'événement jusqu'à $T_{(i-1)}$ inclus, et les indices des observations correspondant à ces censures et événements. P_i contient également l'information sur les covariables de tous les sujets. La suite Q_i obtenue à partir de la décomposition (2.7) contient en plus l'événement $T_{(i)}$, ainsi que les instants de censure précédant $T_{(i)}$ et les indices des observations correspondant à ces censures.

Ignorer les termes $f_{V_i|P_i}(v_i|p_i; \beta, \lambda_T)$ dans (2.5) pour faire de l'inférence sur β revient à ignorer, pour chaque i , la probabilité qu'aucun événement d'intérêt ne survienne dans l'intervalle $(T_{(i-1)}, T_{(i)})$, qu'un événement survienne en $T_{(i)}$, et que m_{i-1} censures se produisent en $X_{(i-1,j)}$ ($j = 1, \dots, m_{i-1}$), sachant tous les événements survenus jusqu'à $T_{(i-1)}$ inclus, toutes les censures jusqu'à $T_{(i-1)}$, et connaissant $Z_l : l = 1, \dots, n$.

Sous certaines conditions, on peut considérer que l'information perdue sur β reste cependant limitée. En effet, le fait qu'un événement se produise en $T_{(i)}$ mais qu'aucun ne survienne entre $T_{(i-1)}$ et $T_{(i)}$ pourrait être expliqué seulement par une fonction de risque λ_T voisine de 0 sur $(T_{(i-1)}, T_{(i)})$ et prenant une valeur élevée en $T_{(i)}$. Il n'y a donc pas d'information perdue sur β . De plus, on peut souvent considérer que les instants de censure et les covariables qui leur sont associées ne contiennent pas d'information sur β : c'est l'hypothèse de censure non-informative. Sous cette hypothèse, Cox a proposé de baser l'inférence sur β sur la vraisemblance partielle

$$\prod_{i=1}^L P[W_i = (i)|q_i]. \quad (2.10)$$

L'ensemble des sujets pour lesquels ni l'événement d'intérêt ni une censure ne se sont produits à l'instant t est appelé ensemble des sujets à risque à t et est noté $R(t)$. Le terme $P[W_i = (i)|q_i]$ s'interprète comme la probabilité que ce soit pour le sujet (i) que se produise l'événement d'intérêt à l'instant $T_{(i)} = t_i$, sachant qu'un événement se produit en t_i et connaissant l'ensemble des sujets à risque à cet instant.

$P[W_i = (i)|q_i]$ se calcule comme suit :

$$\begin{aligned}
& P[W_i = (i)|q_i] \\
&= \frac{P[T_{(i)} \in [t_i, t_i + \Delta t)|q_i] \cdot \prod_{l \in R(t_i) \setminus (i)} P[T_l \notin [t_i, t_i + \Delta t)|q_i]}{\sum_{j \in R(t_i)} \left[P[T_j \in [t_i, t_i + \Delta t)|q_i] \cdot \prod_{l \in R(t_i) \setminus j} P[T_l \notin [t_i, t_i + \Delta t)|q_i] \right]} \\
&= \frac{\lambda_{T|Z}(t_i|Z_{(i)})\Delta t \cdot \prod_{l \in R(t_i) \setminus (i)} [1 - \lambda_{T|Z}(t_i|Z_l)\Delta t]}{\sum_{j \in R(t_i)} \left[\lambda_{T|Z}(t_i|Z_j)\Delta t \cdot \prod_{l \in R(t_i) \setminus j} [1 - \lambda_{T|Z}(t_i|Z_l)\Delta t] \right]} \\
&\xrightarrow{\Delta t \downarrow 0} \frac{\lambda_{T|Z}(t_i|Z_{(i)})}{\sum_{j \in R(t_i)} \lambda_{T|Z}(t_i|Z_j)} \\
&= \frac{\lambda_T(t_i)e^{\beta'Z_{(i)}}}{\sum_{j \in R(t_i)} \lambda_T(t_i)e^{\beta'Z_j}} \\
&= \frac{e^{\beta'Z_{(i)}}}{\sum_{j \in R(t_i)} e^{\beta'Z_j}}.
\end{aligned}$$

D'après (2.10), la vraisemblance partielle de Cox est donnée par

$$\prod_{i=1}^L \frac{e^{\beta'Z_{(i)}}}{\sum_{j \in R(t_i)} e^{\beta'Z_j}} = \prod_{i=1}^n \left[\frac{e^{\beta'Z_i}}{\sum_{j=1}^n e^{\beta'Z_j} 1_{\{X_i \leq X_j\}}} \right]^{\Delta_i}, \quad (2.11)$$

Cette vraisemblance sera notée $L_n^{Cox}(\beta)$ par la suite.

Dans le cas de covariables dépendant du temps, V_i contient l'information sur les covariables sur l'intervalle $[T_{(i-1)}, T_{(i)})$. La vraisemblance partielle de Cox s'écrit :

$$L_n^{Cox}(\beta) = \prod_{i=1}^n \left[\frac{e^{\beta'Z_i(X_i)}}{\sum_{j=1}^n e^{\beta'Z_j(X_i)} 1_{\{X_i \leq X_j\}}} \right]^{\Delta_i}. \quad (2.12)$$

Dans le cas de covariables fixes, la vraisemblance partielle de Cox peut aussi être obtenue en calculant la loi de probabilité de la statistique des rangs associée à l'échantillon des instants d'événements observés. L'argument ci-dessous est rappelé notamment par Fleming et Harrington [37] et Kalbfleisch et Prentice [49].

2.2.3 Loi de probabilité de la statistique de rang

Nous nous intéressons tout d'abord au cas où aucun événement n'est censuré.

Loi de probabilité de la statistique de rang. Cas non censuré

Nous disposons d'un n -échantillon $(X_i, \Delta_i, Z_i)_{1 \leq i \leq n}$ du triplet (X, Δ, Z) tel que

$$X = T \wedge C = T \text{ et } \Delta = 1.$$

Z est fixe dans le temps. Soit (T_1, \dots, T_n) les instants d'occurrence de l'événement d'intérêt, supposés tous distincts.

Rappelons que la statistique d'ordre de (T_1, \dots, T_n) est l'application qui au vecteur (T_1, \dots, T_n) associe le vecteur $(T_{(1)}, \dots, T_{(n)})$, obtenu en ordonnant les composantes $(T_i)_{1 \leq i \leq n}$ par ordre croissant.

Soit $R_i = \sum_{j=1}^n 1_{\{T_i - T_j \in \mathbb{R}^+\}}$ le rang de T_i ($i = 1, \dots, n$) dans l'échantillon (T_1, \dots, T_n) , c'est-à-dire le nombre d'observations T_j ($j = 1, \dots, n$) qui sont inférieures ou égales à T_i . La statistique de rang de (T_1, \dots, T_n) est le vecteur des rangs $R = (R_1, \dots, R_n)$. Par exemple, si on observe la réalisation $(28, 15, 17, 6)$ de (T_1, T_2, T_3, T_4) , la statistique d'ordre de (T_1, T_2, T_3, T_4) est le vecteur $(6, 15, 17, 28)$ et le vecteur des rangs est $(4, 2, 3, 1)$.

Nous notons $Z_{(i)}$ ($1 \leq i \leq n$) le vecteur de covariables associé au sujet pour lequel l'événement survient en $T_{(i)}$ (on prendra garde au fait que $Z_{(i)}$ ne désigne pas la statistique d'ordre des Z_i).

Proposition 2.1 *La loi de probabilité de la statistique de rang R associée à l'échantillon (T_1, \dots, T_n) est donnée par*

$$Pr(R = (r_1, \dots, r_n)) = \prod_{i=1}^n \frac{e^{\beta' Z_{(i)}}}{\sum_{j=1}^n e^{\beta' Z_{(j)}} 1_{\{T_{(i)} \leq T_{(j)}\}}}.$$

Démonstration. Notons tout d'abord que la loi de probabilité de la statistique de rang $R = (r_1, \dots, r_n)$ est donnée par :

$$\begin{aligned} Pr(R = (r_1, \dots, r_n)) &= P_T(T_{(1)} < \dots < T_{(n)}) \\ &= \int_0^\infty \int_{t_{(1)}}^\infty \dots \int_{t_{(n-1)}}^\infty \prod_{i=1}^n f_{T|Z}(t_{(i)} | Z_{(i)}) dt_{(n)} \dots dt_{(1)}. \end{aligned} \quad (2.13)$$

L'intégrale $\int_{t_{(n-1)}}^\infty f_{T|Z}(t_{(n)} | Z_{(n)}) dt_{(n)}$ se calcule comme suit :

$$\begin{aligned} \int_{t_{(n-1)}}^\infty f_{T|Z}(t_{(n)} | Z_{(n)}) dt_{(n)} &= S_{T|Z}(t_{(n-1)} | Z_{(n)}) \\ &= \frac{e^{\beta' Z_{(n)}}}{e^{\beta' Z_{(n)}}} e^{-\Lambda_T(t_{(n-1)}) \exp(\beta' Z_{(n)})}. \end{aligned}$$

Une fois ce résultat obtenu, nous pouvons calculer l'intégrale suivante, issue de l'intégrale multiple figurant dans l'expression (2.13) :

$$\begin{aligned}
& \int_{t_{(n-2)}}^{\infty} f_{T|Z}(t_{(n-1)}|Z_{(n-1)}) S_{T|Z}(t_{(n-1)}|Z_{(n)}) dt_{(n-1)} \\
&= \int_{t_{(n-2)}}^{\infty} \lambda_{T|Z}(t_{(n-1)}) e^{\beta' Z_{(n-1)}} e^{-\Lambda_{T|Z}(t_{(n-1)})[\exp(\beta' Z_{(n-1)}) + \exp(\beta' Z_{(n)})]} dt_{(n-1)} \\
&= \frac{e^{\beta' Z_{(n)}}}{e^{\beta' Z_{(n)}} e^{\beta' Z_{(n-1)}} + e^{\beta' Z_{(n)}}} e^{-\Lambda_T(t_{(n-2)})[\exp(\beta' Z_{(n-1)}) + \exp(\beta' Z_{(n)})]}.
\end{aligned}$$

En effectuant ce calcul de manière récursive, l'équation (2.13) se ramène finalement :

$$Pr(R = (r_1, \dots, r_n)) = \prod_{i=1}^n \frac{e^{\beta' Z_{(i)}}}{\sum_{j=1}^n e^{\beta' Z_{(j)}} 1_{\{T_{(i)} \leq T_{(j)}\}}},$$

qui s'écrit encore $\prod_{i=1}^n \frac{e^{\beta' Z_i}}{\sum_{j=1}^n e^{\beta' Z_j} 1_{\{T_i \leq T_j\}}}$ et qui est la vraisemblance partielle de Cox. ■

Fleming et Harrington [37] et Kalbfleisch et Prentice [49] généralisent l'argument développé dans cette partie aux données censurées. Nous expliquons cette généralisation.

Loi de probabilité de la statistique de rang. Cas censuré

Nous supposons maintenant que les durées T_i ($i = 1, \dots, n$) peuvent être censurées à droite (i.e. nous ne supposons plus que $\Delta_i = 1, i = 1, \dots, n$). Un échantillon de durées censurées ne fournit plus qu'une information partielle sur le vecteur des rangs. Nous illustrons ceci par un exemple extrait de l'ouvrage de Kalbfleisch et Prentice [49].

Supposons que l'on ait observé les durées de vie suivantes : 114, 90*, 63, 108*, où un astérisque indique une durée de vie censurée. Il est maintenant possible de construire six vecteurs de rangs différents, cohérents avec les durées observées :

$$\begin{aligned}
r_1 &= (3, 2, 4, 1), & r_2 &= (3, 4, 2, 1), & r_3 &= (3, 2, 1, 4), \\
r_4 &= (3, 4, 1, 2), & r_5 &= (3, 1, 2, 4), & r_6 &= (3, 1, 4, 2).
\end{aligned}$$

Notons par (X_1, \dots, X_n) l'échantillon des durées de vie éventuellement censurées à droite et par $(\Delta_1, \dots, \Delta_n)$ les indicateurs de censure correspondant. $(X_{(1)}, \dots, X_{(n)})$

désigne la statistique d'ordre de (X_1, \dots, X_n) . Notons $\Delta_{(i)}$ l'indicateur de censure associé à l'instant $X_{(i)}$ ($i = 1, \dots, n$). Comme précédemment, R désigne la statistique de rang associée à (X_1, \dots, X_n) .

Une première démarche consiste à calculer la loi de probabilité de (r_1, \dots, r_n) et $(\Delta_{(1)}, \dots, \Delta_{(n)})$. Mais cette démarche, faisant intervenir la loi de la censure serait complexe, aussi une autre approche est privilégiée. Elle consiste à construire une vraisemblance pour l'ensemble des rangs possibles de (T_1, \dots, T_n) , cohérents avec les vecteurs $(X_{(1)}, \dots, X_{(n)})$ et $(\Delta_{(1)}, \dots, \Delta_{(n)})$. Ceci revient à ignorer l'ordre exact des censures entre deux événements d'intérêt.

Nous notons donc (X_1, \dots, X_L) ($L \leq n$) le vecteur des instants d'occurrence de l'événement d'intérêt, supposés tous distincts, et $(X_{(1)}, \dots, X_{(L)})$ sa statistique d'ordre. Supposons que m_i censures se produisent dans l'intervalle $[X_{(i)}, X_{(i+1)})$ et notons $(X_{(i,j)} : j = 1 \dots, m_i)$ ces instants de censure. Les vecteurs $Z_{(i,1)}, \dots, Z_{(i,m_i)}$ désignent les vecteurs de covariables des individus censurés en $X_{(i,1)}, \dots, X_{(i,m_i)}$. Un vecteur de rang $R^* = (r_1^*, \dots, r_n^*)$ est obtenu de la manière suivante :

- si X_i est une observation non censurée, r_i^* est le rang de X_i dans la suite $(X_{(1)}, \dots, X_{(L)})$,
- si X_i est une observation censurée, r_i^* est le rang de la plus grande observation non censurée précédent X_i .

La vraisemblance pour le paramètre β est obtenue en calculant la probabilité d'observer le vecteur $(r_1^*, \dots, r_n^*, \Delta_{(1)}, \dots, \Delta_{(n)})$, ce qui revient à calculer

$$Pr(X_{(1)} < \dots < X_{(L)}, X_{(i)} \leq X_{(i,j)} : i = 1, \dots, L; j = \dots, m_i)$$

qui est donnée par la proposition suivante.

Proposition 2.2 *La loi de probabilité de $(R^*, \Delta_{(1)}, \dots, \Delta_{(n)})$ est donnée par :*

$$\begin{aligned} & Pr(X_{(1)} < \dots < X_{(L)}, X_{(i)} \leq X_{(i,j)} : i = 1, \dots, L; j = \dots, m_i) \\ &= \prod_{i=1}^L \frac{e^{\beta' Z_{(i)}}}{\sum_{j=1}^n e^{\beta' Z_j} 1_{\{X_{(i)} \leq X_j\}}} \end{aligned}$$

Démonstration. Remarquons tout d'abord que

$$Pr(X_{(1)} < \dots < X_{(L)}, X_{(i)} \leq X_{(i,j)} : i = 1, \dots, L; j = \dots, m_i),$$

est égale à

$$\int_0^\infty \int_{x_{(1)}}^\infty \cdots \int_{x_{(L-1)}}^\infty \prod_{i=1}^L \left[f_{T|Z}(x_{(i)}|Z_{(i)}) \prod_{j=1}^{m_i} S_{T|Z}(x_{(i)}|Z_{(i,j)}) \right] dx_{(L)} \cdots dx_{(1)}. \quad (2.14)$$

Notons que

$$\begin{aligned} \prod_{j=1}^{m_i} S_{T|Z}(x_{(i)}|Z_{(i,j)}) &= \prod_{j=1}^{m_i} e^{-\Lambda_T(x_{(i)}) \exp(\beta' Z_{(i,j)})} \\ &= e^{-\Lambda_T(x_{(i)}) \sum_{j=1}^{m_i} \exp(\beta' Z_{(i,j)})}. \end{aligned}$$

Nous calculons d'abord l'intégrale $\int_{x_{(L-1)}}^\infty f_{T|Z}(x_{(L)}|Z_{(L)}) \prod_{j=1}^{m_i} S_{T|Z}(x_{(L)}|Z_{(L,j)}) dx_{(L)}$:

$$\begin{aligned} &\int_{x_{(L-1)}}^\infty f_{T|Z}(x_{(L)}|Z_{(L)}) e^{-\Lambda_T(x_{(L)}) \sum_{j=1}^{m_i} \exp(\beta' Z_{(L,j)})} dx_{(L)} \\ &= \frac{e^{\beta' Z_{(L)}}}{e^{\beta' Z_{(L)}} + \sum_{j=1}^{m_i} e^{\beta' Z_{(L,j)}}} e^{-\Lambda(t_{L-1})[\exp(\beta' Z_{(L)}) + \sum_{j=1}^{m_i} \exp(\beta' Z_{(L,j)})]}. \end{aligned}$$

Puis, en effectuant ce calcul de manière récursive, l'équation (2.14) se ramène à :

$$\prod_{i=1}^L \frac{e^{\beta' Z_{(i)}}}{\sum_{j=1}^n e^{\beta' Z_j} 1_{\{X_{(i)} \leq X_j\}}},$$

qui est égale à la vraisemblance partielle de Cox (2.11). ■

2.2.4 Estimation de la fonction de risque

Dans cette partie, nous supposons toujours que la censure est non-informative. Nous supposons de plus l'absence d'*ex aequo* parmi les instants d'événements. La méthode du maximum de vraisemblance partielle ne permet pas d'estimer directement la fonction de risque cumulé $\Lambda_T = \int_0^t \lambda_T(s) ds$, puisque λ_T ne figure pas dans les vraisemblances partielles (2.11) et (2.12). L'estimateur de Λ_T le plus souvent retenu est l'estimateur proposé par Breslow [13], dont nous donnons l'expression pour le modèle (2.3). Si le coefficient de régression β a été estimé par $\hat{\beta}$, Breslow propose d'estimer Λ_T par

$$\hat{\Lambda}_n(\hat{\beta}, t) = \sum_{i=1}^n \frac{\Delta_i 1_{\{X_i \leq t\}}}{\sum_{j=1}^n e^{\hat{\beta}' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}. \quad (2.15)$$

Il s'appuie sur un raisonnement que nous expliquerons dans la section consacrée à l'estimation semiparamétrique dans le modèle de Cox.

Dans le cas particulier où il n'y a pas de covariables dans le modèle, l'estimateur de Breslow devient l'estimateur de Nelson-Aalen de Λ_T , égal à

$$\hat{\Lambda}_{n,NA}(\hat{\beta}, t) = \sum_{i=1}^n \frac{\Delta_i 1_{\{X_i \leq t\}}}{\sum_{j=1}^n 1_{\{X_i \leq X_j\}}}.$$

Pour de plus amples détails sur les estimateurs des fonctions de risque cumulé et de survie, on se référera à Bagdonavičius et Nikulin [8], Fleming et Harrington [37], Huber et Lecoutre [46], Klein et Moeschberger [50] et Pons et Huber [73].

2.2.5 Prise en compte des *ex aequo*

Nous avons supposé jusqu'à présent qu'il n'y avait pas d'*ex aequo* parmi les instants d'événements d'intérêt observés. Cependant, dans la pratique de l'analyse statistique des durées de vie, cette hypothèse n'est pas toujours vérifiée. En effet, l'observation des sujets se fait souvent à des instants successifs et non en continu. Diverses modifications de la vraisemblance partielle de Cox ont été proposées pour tenir compte des *ex aequo*.

Nous donnons ici l'expression de la vraisemblance partielle proposée par Breslow [14]. D'autres modifications de la vraisemblance de Cox sont rappelées par Klein et Moeschberger dans [50]. La version de Breslow nous intéresse car sa construction est liée à la méthode du maximum de vraisemblance semiparamétrique que nous décrirons dans la dernière section de ce chapitre.

Supposons que l'on ait observé L ($L \leq n$) instants d'événements d'intérêt parmi n sujets. Supposons que seulement D ($D \leq L$) d'entre eux soient distincts. Notons ces instants distincts par t_1, \dots, t_D . Notons d_i le nombre de sujets pour lesquels l'événement d'intérêt se produit en t_i et D_i l'ensemble de ces sujets. Soit $S_i = \sum_{j=1}^n Z_j 1_{\{j \in D_i\}}$ la somme des vecteurs de covariables associés à ces individus. Breslow [14] a proposé la vraisemblance suivante :

$$L_n(\beta) = \prod_{i=1}^D \frac{e^{\beta' S_i}}{\left[\sum_{j=1}^n e^{\beta' Z_j} 1_{\{t_i \leq X_j\}} \right]^{d_i}},$$

ce qui revient à calculer la contribution à la vraisemblance partielle de Cox de chaque sujet pris séparément, et à multiplier ces contributions. Cette expression se

généralise à des covariables dépendant du temps :

$$L_n(\beta) = \prod_{i=1}^D \frac{e^{\beta' S_i(t_i)}}{\left[\sum_{j=1}^n e^{\beta' Z_j(t_i)} 1_{\{t_i \leq X_j\}} \right]^{d_i}}.$$

Dans le cas où il n'y a pas d'*ex aequo*, ces vraisemblances se ramènent aux vraisemblances (2.11) et (2.12).

L'estimateur (2.15) de la fonction de risque Λ_T proposé par Breslow peut être modifié pour prendre en compte des *ex aequo* parmi les instants d'occurrence de l'événement d'intérêt.

Supposons que l'on observe un échantillon de durées de vie (X_1, \dots, X_n) . Notons $t_1 < \dots < t_D$ les D instants d'événement distincts observés. L'estimateur de Breslow s'écrit maintenant

$$\hat{\Lambda}_n(\beta, t) = \sum_{i=1}^D \frac{1_{\{t_i \leq t\}} \left[\sum_{j=1}^n 1_{\{X_j = t_i\}} \Delta_j \right]}{\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{t_i \leq X_j\}}}, \quad (2.16)$$

où $\sum_{j=1}^n 1_{\{X_j = t_i\}} \Delta_j$ représente le nombre d'événements d'intérêt survenant à l'instant t_i . L'expression (2.16) se ramène à (2.15) dans le cas où tous les instants d'événement sont distincts.

2.3 Propriétés asymptotiques

Nous nous intéressons aux propriétés asymptotiques de l'estimateur du maximum de vraisemblance partielle de β et de l'estimateur de Breslow de $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$ dans le modèle de Cox défini par la fonction de risque

$$\lambda_{T|Z}(t) = \lambda_T(t) e^{\beta' Z(t)}.$$

L'étude asymptotique de ces estimateurs, lorsque n tend vers l'infini, a été faite par Andersen et Gill [1]. Des exposés détaillés sont également donnés par Bagdonavičius et Nikulin [8], Gouriéroux, Lecoutre et Tassi [39] et Pons et Huber [73].

Soit un n -échantillon $(X_i, \Delta_i, Z_i)_{1 \leq i \leq n}$ du triplet (X, Δ, Z) , où $X = T \wedge C$, $\Delta = 1_{\{T \leq C\}}$ et $Z = (Z(t))_{t \geq 0}$ est un processus p -dimensionnel de covariables. Nous notons τ la durée maximale d'observation : chaque sujet est observé pendant une durée au plus égale à τ , mais il n'est pas nécessairement observé pendant tout l'intervalle de temps $[0, \tau]$.

Nous désignons par $l_n^{Cox}(\beta)$ le logarithme de la vraisemblance partielle de Cox $L_n^{Cox}(\beta)$. Cette log-vraisemblance partielle est égale à :

$$l_n^{Cox}(\beta) = \sum_{i=1}^n \Delta_i \left[\beta' Z_i(X_i) - \ln \left[\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}} \right] \right].$$

Le vecteur du score $\frac{\partial}{\partial \beta} l_n^{Cox}(\beta) = (\frac{\partial}{\partial \beta_r} l_n^{Cox}(\beta))_{1 \leq r \leq p}$ est :

$$\frac{\partial}{\partial \beta} l_n^{Cox}(\beta) = \sum_{i=1}^n \Delta_i \left[Z_i(X_i) - \frac{\sum_{j=1}^n Z_j(X_i) e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}} \right]. \quad (2.17)$$

Pratiquement, les p équations de vraisemblance $\frac{\partial}{\partial \beta_r} l_n^{Cox}(\beta) = 0$ se résolvent par un algorithme numérique (algorithme de Newton-Raphson).

Soit $\hat{\beta}_n$ l'estimateur du maximum de vraisemblance partielle de β et $\hat{\Lambda}_n(\hat{\beta}_n, \cdot)$ l'estimateur de Breslow de $\Lambda_T(\cdot)$ calculé pour $\beta = \hat{\beta}_n$ (Nous le noterons par la suite $\hat{\Lambda}_n(\cdot)$ pour alléger les notations).

Pour β et t donnés, notons, pour un i quelconque ($i = 1, \dots, n$) :

$$s^{(r)}(\beta, t) = E \left[\{Z_i(t)\}^{\otimes r} e^{\beta' Z_i(t)} 1_{\{t \leq X_i\}} \right], \quad r = 0, 1, 2.$$

Remarquons que $s^{(1)}(\beta, t) = \partial s^{(0)}(\beta, t) / \partial \beta$ et $s^{(2)}(\beta, t) = \partial s^{(1)}(\beta, t) / \partial \beta$. Notons également par $\Sigma(\beta)$ la matrice de dimension $(p \times p)$

$$\Sigma(\beta) = \int_0^\tau \left[s^{(2)}(\beta, t) - \frac{\{s^{(1)}(\beta, t)\}^{\otimes 2}}{s^{(0)}(\beta, t)} \right] d\Lambda_T(t)$$

et par $j(t)$ le vecteur de dimension p

$$j(t) = \int_0^t \frac{s^{(1)}(\beta, s)}{s^{(0)}(\beta, s)} d\Lambda_T(s).$$

Andersen et Gill [1] supposent les conditions suivantes pour l'étude du comportement asymptotique des estimateurs $\hat{\beta}_n$ et $\hat{\Lambda}_n$.

C1 Le paramètre β appartient à un sous-ensemble borné B de \mathbb{R}^p .

C2 La fonction Λ_T est telle que $\Lambda_T(\tau) < \infty$.

C3 $\mathbb{P}[t \leq X] > 0$ pour tout $t \in [0, \tau]$.

C4 Z est un processus à trajectoires continues à gauche avec limite à droite tel que

$$E \left[\sup_{t \in [0, \tau], \beta \in B} |Z(t)|^2 e^{\beta' Z(t)} \right] < \infty.$$

C5 La matrice $\Sigma(\beta)$ est définie positive.

Sous ces conditions, Andersen et Gill ont montré le résultat suivant (où les limites sont prises lorsque n tend vers l'infini) :

Théorème 2.1 (Andersen et Gill, 1982) *L'estimateur du maximum de vraisemblance partielle $\hat{\beta}_n$ et l'estimateur de Breslow $\hat{\Lambda}_n$ de Λ_T possèdent les propriétés suivantes :*

1. $\hat{\beta}_n$ est un estimateur consistant de β .
2. $\sqrt{n}(\hat{\beta}_n - \beta)$ converge en loi vers un vecteur aléatoire de dimension p , de loi normale centrée et de matrice de covariance $\Sigma^{-1}(\beta)$.
3. $\sqrt{n}(\hat{\Lambda}_n - \Lambda_T)$ converge en loi vers un processus gaussien centré G , de fonction de covariance

$$\text{cov}[G(s), G(t)] = \int_0^{s \wedge t} \frac{d\Lambda_T(u)}{s^{(0)}(\beta, u)} + j'(t)\Sigma^{-1}(\beta)j(s), \quad s, t \leq \tau.$$

Remarques

1. En particulier, d'après le point 3., $\sqrt{n}(\hat{\Lambda}_n(t) - \Lambda_T(t))$ converge en loi vers une variable aléatoire normale centrée et de variance

$$\int_0^t \frac{d\Lambda_T(u)}{s^{(0)}(\beta, u)} + j'(t)\Sigma^{-1}(\beta)j(t), \quad t \leq \tau.$$

2. Plus généralement, on peut exprimer les points 2. et 3. sous la forme unique du résultat énoncé ci-dessous :

Soit t_1, \dots, t_m m instants dans $[0, \tau]$. Le vecteur

$$\sqrt{n}(\hat{\beta}_n - \beta, \hat{\Lambda}_n(t_1) - \Lambda_T(t_1), \dots, \hat{\Lambda}_n(t_m) - \Lambda_T(t_m))$$

converge en loi vers un vecteur aléatoire de dimension $(p+m)$, de loi normale centrée et de matrice de covariance

$$\begin{pmatrix} \Sigma^{-1}(\beta) & -\Sigma^{-1}(\beta)J \\ -J'\Sigma^{-1}(\beta) & W \end{pmatrix},$$

où $W(i, j) = \int_0^{t_i \wedge t_j} \frac{d\Lambda_T(u)}{s^{(0)}(\beta, u)} + j'(t_j)\Sigma^{-1}(\beta)j(t_i)$ et $J = (j(t_1), \dots, j(t_m))$.

2.4 Méthode du maximum de vraisemblance semiparamétrique

La méthode du maximum de vraisemblance semiparamétrique permet de considérer la vraisemblance partielle comme une vraisemblance “légitime”, au sens où elle est obtenue à partir de la densité de (X, Δ, Z) , mais sur un espace des paramètres modifié. Mais au-delà de la justification de la vraisemblance partielle qu’elle fournit, cette méthode jouera un rôle prépondérant dans la deuxième partie de ce travail aussi la présentons-nous maintenant.

Nous rappelons tout d’abord la notion de vraisemblance concentrée.

2.4.1 Concentration de la vraisemblance

Considérons un modèle statistique $(\mathcal{Y}, P_\theta : \theta \in \Theta \subset \mathbb{R}^p)$ dans lequel la loi du vecteur aléatoire Y à valeurs dans \mathcal{Y} admet une densité $f_Y(y; \theta)$ par rapport à une certaine mesure. Supposons que le vecteur θ des paramètres puisse se décomposer en deux sous-vecteurs θ_1 et $\theta_2 : \theta = (\theta'_1, \theta'_2)'$.

La recherche des estimations de θ par la méthode du maximum de vraisemblance peut être facilitée en effectuant des maximisations successives de la vraisemblance par rapport à chacun des sous-vecteurs θ_1 et θ_2 , pour une valeur fixe de l’autre sous-vecteur. Cette démarche conduit à la définition de la vraisemblance concentrée [40, 68, 86, 87].

Définition 2.1 *L’application $\theta_2 \mapsto \max_{\theta_1} f_Y(y; \theta_1, \theta_2)$ est appelée vraisemblance concentrée en θ_1 . On appelle log-vraisemblance concentrée en θ_1 , notée $\ln f_Y^c(y; \theta_2)$, l’application $\theta_2 \mapsto \max_{\theta_1} \ln f_Y(y; \theta_1, \theta_2)$.*

Le calcul des estimations du maximum de vraisemblance de θ_1 et θ_2 à partir de la vraisemblance concentrée procède de la manière suivante.

Proposition 2.3 *Les solutions $\hat{\theta}_n = (\hat{\theta}'_{1,n}, \hat{\theta}'_{2,n})'$ du problème de maximisation $\max_{\theta_1, \theta_2} f_Y(y; \theta_1, \theta_2)$ peuvent être obtenues en utilisant la démarche suivante :*

1. *On maximise la log-vraisemblance $\ln f_Y(y; \theta_1, \theta_2)$ par rapport à θ_1 , pour une valeur fixée de θ_2 . Soit $\hat{\Theta}_{1,n}(\theta_2)$ l’ensemble des valeurs de θ_1 maximisant $\ln f_Y(y; \theta_1, \theta_2)$.*

2. On maximise $\ln f_Y(y; \theta_1, \theta_2)$ par rapport à θ_2 , pour $\theta_1 \in \hat{\Theta}_{1,n}(\theta_2)$, c'est-à-dire $\ln f_Y^c(y; \theta_2)$. Soit $\hat{\Theta}_2$ l'ensemble des θ_2 rendant $\ln f_Y(y; \theta_1, \theta_2)$ maximum. Les solutions $\hat{\theta}_n = (\hat{\theta}'_{1,n}, \hat{\theta}'_{2,n})'$ sont alors les éléments de $\cup_{\theta_2 \in \hat{\Theta}_2} \{\hat{\Theta}_{1,n}(\theta_2) \times \theta_2\}$.

Cette démarche peut être généralisée à une décomposition du vecteur θ en un plus grand nombre de sous-vecteurs : $\theta = (\theta'_1, \dots, \theta'_N)'$. Elle est surtout intéressante lorsque la correspondance $\theta_2 \mapsto \hat{\Theta}_{1,n}(\theta_2)$ peut être explicitée, en particulier lorsque la concentration conduit à une valeur unique (c'est-à-dire $\hat{\Theta}_{1,n}(\theta_2) = \{\hat{\theta}_{1,n}(\theta_2)\}$).

Breslow [13] et Johansen [47] ont expliqué comment cette démarche peut être appliquée à la vraisemblance totale (2.9) pour estimer le paramètre de régression β et la fonction de risque Λ_T dans le modèle de Cox. Nous expliquons cette démarche.

2.4.2 Maximum de vraisemblance semiparamétrique dans le modèle de Cox

Considérons le modèle de Cox

$$\lambda_{T|Z}(t) = \lambda_T(t) e^{\beta' Z(t)}.$$

La vraisemblance des paramètres β et $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$ au vu d'un n -échantillon $(X_i, \Delta_i, Z_i(s), 0 \leq s \leq X_i)_{1 \leq i \leq n}$ peut s'écrire

$$L_n(\beta, \Lambda_T) = \prod_{i=1}^n \left(\lambda_T(X_i) e^{\beta' Z_i(X_i)} \right)^{\Delta_i} e^{-\int_0^{X_i} \lambda_T(s) \exp(\beta' Z_i(s)) ds}.$$

La méthode du maximum de vraisemblance n'étant pas applicable à cette fonction, Breslow [13] et Johansen [47] ont proposé de définir une vraisemblance en considérant que la fonction de risque cumulé Λ_T est une fonction en escalier (notons-la $\Lambda_{T,n}$), dont les sauts se produisent aux L ($L \leq n$) instants X_i tels que $\Delta_i = 1$, $i = 1, \dots, n$ (notons par $X_{(1)} < \dots < X_{(L)}$ ces instants ordonnés par ordre croissant, et par $X_{(L+1)} \leq \dots \leq X_{(n)}$ les instants de censure).

Ceci ramène le problème de maximisation de la vraisemblance à un problème de "dimension finie" : les paramètres du modèle sont le paramètre de régression β et les valeurs des sauts de la fonction $\Lambda_{T,n}$ aux instants $X_{(i)}$ tels que $\Delta_{(i)} = 1$. Notons $\Delta \Lambda_{T,n}(X_{(l)})$ ($l = 1, \dots, L$) les valeurs de ces sauts et $L_n(\beta, \Lambda_{T,n})$ la vraisemblance à maximiser.

Plutôt que d'estimer simultanément $(\beta, \Delta\Lambda_{T,n}(X_{(1)}), \dots, \Delta\Lambda_{T,n}(X_{(L)}))$, nous pouvons utiliser la méthode d'estimation par vraisemblance concentrée pour estimer successivement ces paramètres. Ceci consiste alors à estimer les $\Delta\Lambda_{T,n}(X_{(l)})$ ($l = 1, \dots, L$), à les remplacer par leurs estimateurs dans la vraisemblance $L_n(\beta, \Lambda_{T,n})$, puis à estimer β . Nous vérifions, après deux remarques, que ceci revient à estimer β par la méthode du maximum de vraisemblance partielle et Λ_T par l'estimateur de Breslow (ce résultat est rappelé par Bagdonavičius et Nikulin [8] et Klein et Moeschberger [50]).

Remarques

1. Murphy, Rossini et Van der Vaart [67] et Nielsen, Gill, Andersen et Sørensen [70] appliquent cette méthode respectivement à l'estimation des paramètres du modèle

$$-\text{logit}(S_{T|Z}(t)) = \alpha(t) + \beta'Z$$

(où $\alpha(t)$ est une fonction croissante du temps) et du modèle de fragilité.

Bagdonavičius et Nikulin [8] l'utilisent pour estimer les paramètres du modèle GPH (Generalized Proportional Hazards), qu'il définissent par la fonction de risque instantané

$$\lambda_T(t)e^{\beta'Z(t)}q\left(\int_0^t e^{\beta'Z(u)} d\Lambda_T(u); \gamma\right),$$

où γ est un paramètre inconnu. Pour des choix appropriés de la fonction q , le modèle GPH comprend comme cas particuliers les modèles à risques proportionnels et les modèles de la vie accélérée. Un traitement complet de l'estimation dans les modèles GPH est donné dans [8].

2. Méthode non-paramétrique vs méthode semiparamétrique.

Cette méthode d'estimation de β et Λ_T est parfois appelée méthode non-paramétrique dans la littérature consacrée à l'estimation dans les modèles de durée de vie (on la désigne par le sigle NPML pour Non-Parametric Maximum Likelihood dans la littérature anglo-saxonne).

Notons néanmoins que le terme non-paramétrique n'est pas le plus approprié : seul le traitement de Λ_T est réellement non-paramétrique dans cette méthode. Par la suite, nous utiliserons le nom de méthode d'estimation du maximum de vraisemblance semiparamétrique, utilisé par Bagdonavičius et Nikulin [8].

Nous constatons qu'il existe une relation, donnée dans la proposition suivante, entre la vraisemblance partielle de Cox $L_n^{Cox}(\beta)$ et la concentration en $\Lambda_{T,n}$ de $L_n(\beta, \Lambda_{T,n})$. Cette relation montre que l'estimateur du maximum de vraisemblance partielle de β peut être vu comme un estimateur du maximum de vraisemblance au sens habituel, lorsque la fonction de risque Λ_T est considérée comme une fonction en escalier.

Proposition 2.4 $\max_{\Lambda_{T,n}} L_n(\beta, \Lambda_{T,n}) = L_n^{Cox}(\beta) \cdot e^{-\sum_{i=1}^n \Delta_i}$.

Démonstration. La vraisemblance $L_n(\beta, \Lambda_{T,n})$ du paramètre $(\beta, \Lambda_{T,n})$ au vu du n -échantillon $(X_i, \Delta_i, Z_i(s), 0 \leq s \leq X_i)$ se calcule comme suit :

$$\begin{aligned} L_n(\beta, \Lambda_{T,n}) &= \prod_{i=1}^L \Delta \Lambda_{T,n}(X_{(i)}) e^{\beta' Z_{(i)}(X_{(i)})} e^{-\sum_{j=1}^L \Delta \Lambda_{T,n}(X_{(j)}) \exp(\beta' Z_{(i)}(X_{(j)})) 1_{\{X_{(j)} \leq X_{(i)}\}}} \\ &\quad \times \prod_{i=L+1}^n e^{-\sum_{j=1}^L \Delta \Lambda_{T,n}(X_{(j)}) \exp(\beta' Z_{(i)}(X_{(j)})) 1_{\{X_{(j)} \leq X_{(i)}\}}} \\ &= \prod_{i=1}^L \Delta \Lambda_{T,n}(X_{(i)}) e^{\beta' Z_{(i)}(X_{(i)})} \prod_{i=1}^n e^{-\sum_{j=1}^L \Delta \Lambda_{T,n}(X_{(j)}) \exp(\beta' Z_{(i)}(X_{(j)})) 1_{\{X_{(j)} \leq X_{(i)}\}}}, \end{aligned}$$

d'où l'on déduit la log-vraisemblance

$$\begin{aligned} \ln L_n(\beta, \Lambda_{T,n}) &= \sum_{i=1}^n [\Delta_{(i)} \beta' Z_{(i)}(X_{(i)}) + \Delta_{(i)} \ln \Delta \Lambda_{T,n}(X_{(i)}) \\ &\quad - \sum_{j=1}^L \Delta \Lambda_{T,n}(X_{(j)}) e^{\beta' Z_{(i)}(X_{(j)})} 1_{\{X_{(j)} \leq X_{(i)}\}}] . \end{aligned}$$

Il est immédiat de dériver cette log-vraisemblance par rapport à $\Delta \Lambda_{T,n}(X_{(l)})$:

$$\frac{\partial \ln L_n(\beta, \Lambda_{T,n})}{\partial \Delta \Lambda_{T,n}(X_{(l)})} = \frac{1}{\Delta \Lambda_{T,n}(X_{(l)})} - \sum_{i=1}^n e^{\beta' Z_{(i)}(X_{(l)})} 1_{\{X_{(l)} \leq X_{(i)}\}}, \quad l = 1, \dots, L,$$

et d'obtenir un estimateur du maximum de vraisemblance de $\Delta \Lambda_{T,n}(X_{(l)})$, en résolvant l'équation $\frac{\partial \ln L_n(\beta, \Lambda_{T,n})}{\partial \Delta \Lambda_{T,n}(X_{(l)})} = 0$, ce qui donne

$$\Delta \hat{\Lambda}_{T,n}(X_{(l)}) = \frac{1}{\sum_{j=1}^n e^{\beta' Z_{(j)}(X_{(l)})} 1_{\{X_{(l)} \leq X_{(j)}\}}}, \quad l = 1, \dots, L.$$

Pour une valeur fixée de β , la concentration de $L_n(\beta, \Lambda_{T,n})$ en $\Lambda_{T,n}$ conduit à un unique $\hat{\Lambda}_{T,n}$. Ceci se révèle intéressant pour estimer β , en calculant $L_n(\beta, \hat{\Lambda}_{T,n})$ de la manière suivante :

$$\begin{aligned} & \prod_{i=1}^L \frac{e^{\beta' Z_{(i)}(X_{(i)})}}{\sum_{j=1}^n e^{\beta' Z_{(j)}(X_{(i)})} 1_{\{X_{(i)} \leq X_{(j)}\}}} \prod_{i=1}^n \exp \left[- \sum_{j=1}^L \frac{e^{\beta' Z_{(i)}(X_{(j)})} 1_{\{X_{(j)} \leq X_{(i)}\}}}{\sum_{l=1}^n e^{\beta' Z_{(l)}(X_{(j)})} 1_{\{X_{(j)} \leq X_{(l)}\}}} \right] \\ &= \prod_{i=1}^L \frac{e^{\beta' Z_{(i)}(X_{(i)})}}{\sum_{j=1}^n e^{\beta' Z_{(j)}(X_{(i)})} 1_{\{X_{(i)} \leq X_{(j)}\}}} \exp \left[- \sum_{j=1}^L \frac{\sum_{i=1}^n e^{\beta' Z_{(i)}(X_{(j)})} 1_{\{X_{(j)} \leq X_{(i)}\}}}{\sum_{l=1}^n e^{\beta' Z_{(l)}(X_{(j)})} 1_{\{X_{(j)} \leq X_{(l)}\}}} \right] \\ &= L_n^{Cox}(\beta) \cdot e^{-\sum_{i=1}^n \Delta_i}. \end{aligned}$$

La vraisemblance totale $L_n(\beta, \Lambda_{T,n})$, concentrée en $\Lambda_{T,n}$, est donc proportionnelle à la vraisemblance partielle $L_n^{Cox}(\beta)$. ■

Remarques

1. Un estimateur du maximum de vraisemblance de Λ_T peut être obtenu par

$$\begin{aligned} \hat{\Lambda}_{T,n}(t) &= \sum_{i=1}^L \Delta \hat{\Lambda}_{T,n}(X_{(i)}) 1_{\{X_{(i)} \leq t\}} \\ &= \sum_{i=1}^L \frac{1_{\{X_{(i)} \leq t\}}}{\sum_{j=1}^n e^{\beta' Z_{(j)}(X_{(i)})} 1_{\{X_{(i)} \leq X_{(j)}\}}}. \end{aligned}$$

Nous pouvons réécrire cette expression sous la forme

$$\hat{\Lambda}_{T,n}(t) = \sum_{i=1}^n \frac{\Delta_i 1_{\{X_i \leq t\}}}{\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}$$

qui est l'estimateur de Breslow (2.15) de Λ_T .

Dans le cas d'*ex aequo*, où nous observons D ($D \leq n$) événements distincts aux instants t_1, \dots, t_D , l'estimateur de Breslow (2.16) découle naturellement de la méthode d'estimation semiparamétrique lorsque $\prod_{i=1}^L \Delta \Lambda_{T,n}(X_{(i)})$ est remplacé par

$$\prod_{i=1}^n \prod_{l=1}^D \Delta \Lambda_{T,n}(t_l)^{1_{\{X_i = t_l\}} \Delta_i}$$

dans $L_n(\beta, \Lambda_{T,n})$.

2. Dans le cas du modèle de Cox “usuel”, la méthode du maximum de vraisemblance semiparamétrique se ramène à la méthode du maximum de vraisemblance partielle. En revanche, elle lui fournit une alternative utile quand l’utilisation de celle-ci n’est plus possible. En particulier, la méthode semiparamétrique a été récemment utilisée pour estimer les paramètres du modèle de Cox avec erreur de mesure sur une covariable fixe (Hu, Tsiatis et Davidian [43]) ou dépendant du temps (Wulfsohn et Tsiatis [95]), et du modèle de Cox avec valeurs manquantes d’une covariable fixe (Chen et Little [16]; Martinussen [60]) ou dépendant du temps (Dupuy et Mesbah [35, 36]).

Chapitre 3

Modèle de Cox avec covariables dépendant du temps : travaux récents

Deux problèmes relatifs aux covariables dans un modèle de Cox ont récemment suscité une importante littérature : les valeurs manquantes et l'erreur de mesure. Dans une étude de survie, le problème des valeurs manquantes d'une covariable survient le plus souvent pour des raisons pratiques, liées à la difficulté d'obtenir régulièrement des mesures de cette covariable sur tous les sujets suivis.

En général, une covariable est mesurée avec erreur lorsqu'elle représente la mesure d'une quantité physique ou biologique qu'il est difficile d'obtenir précisément. Par exemple, la mesure de la quantité de lymphocytes CD4, marqueurs de l'évolution du SIDA chez les patients séropositifs, est souvent entachée d'erreur. Une évaluation de la relation entre la survie de ces patients et des mesures répétées du marqueur doit prendre en compte cette erreur. De nombreux travaux ont été récemment proposés en ce sens par Dafni et Tsiatis [24], Hu, Tsiatis et Davidian [43], Tsiatis, De Gruttola et Wulfsohn [85] et Wulfsohn et Tsiatis [95] lorsque cette relation est décrite par un modèle de Cox.

Les problèmes de biais induits par ces deux problèmes dans l'estimation du paramètre d'intérêt β sont connus depuis longtemps (Collett [19] ; Prentice [74]). Des solutions élaborées n'ont en revanche été développées que récemment. Nous proposons dans ce chapitre une synthèse de travaux récents portant sur ces problèmes. Nous développons plus particulièrement la question des valeurs manquantes d'une covariable dépendant du temps, sujet du travail réalisé dans la deuxième partie de ce manuscrit.

Nous décrivons d'abord brièvement deux solutions récentes au problème de l'estimation dans un modèle de Cox avec erreur de mesure sur une covariable dépendant du temps (Dafni et Tsiatis [24] ; Tsiatis, De Gruttola et Wulfsohn [85] ; Wulfsohn et Tsiatis [95]). Ces solutions n'ont pas encore été étayées par des résultats de convergence ou de distribution asymptotique des estimateurs obtenus.

Nous décrivons ensuite quatre méthodes d'estimation des paramètres d'un modèle de Cox en présence de données manquantes d'une covariable dépendant du temps (Lin et Ying [56] ; Paik et Tsai [71] ; Wang, Hsu, Feng et Prentice [92] ; Zhou et Pepe [96]). Une cinquième méthode, relative à des covariables fixes sera évoquée (Chen et Little [16] ; Martinussen [60]).

Nous remarquerons que si des résultats de consistance et de normalité asymptotique des estimateurs obtenus grâce à ces méthodes ont été démontrés, ces méthodes reposent sur des hypothèses contraignantes qui ne sont pas toujours satisfaites en pratique.

3.1 Cas de l'erreur de mesure sur la covariable

Considérons le triplet (X, Δ, Z) où $X = T \wedge C$, $\Delta = 1_{\{T \leq C\}}$ et $Z = (Z(t))_{t \geq 0}$ est un processus de covariables. Supposons que la relation entre T et Z soit définie par le modèle de Cox

$$\lambda_{T|Z}(t) = \lambda_T(t) e^{g(Z(s), 0 \leq s \leq t; \beta)}.$$

Supposons que l'on n'observe pas exactement $Z(t)$, mais une valeur de $Z(t)$ entachée d'erreur. Nous notons $W(t)$ cette valeur erronée. Prentice a montré [74] que l'estimateur du maximum de vraisemblance partielle du paramètre β (qui quantifie l'intensité de la relation entre le processus de covariables Z et la durée jusqu'à survenue d'un événement) obtenu à partir de n répliques indépendantes $(X_i, \Delta_i, W_i(s), 0 \leq s \leq X_i)_{1 \leq i \leq n}$ est un estimateur biaisé de l'association entre T et Z .

Approche en deux étapes

Tsiatis, DeGruttola et Wulfsohn [85] ont proposé une méthode d'estimation de β qui procède en deux étapes.

La première consiste à modéliser conjointement $Z(t)$ et $\{W(s), 0 \leq s \leq t\}$ et à estimer l'espérance (que nous noterons par la suite $m_{Z|W}(Z(t)|W(s), 0 \leq s \leq t)$) de la loi conditionnelle de $Z(t)$ sachant $\{W(s), 0 \leq s \leq t\}$.

Dans une deuxième étape, le paramètre β est estimé en utilisant une estimation $\hat{L}_n^{Cox}(\beta)$ de la vraisemblance partielle de Cox dans laquelle les valeurs $Z(t)$ sont remplacées par les estimations $\hat{m}_{Z|W}(Z(t)|W(s), 0 \leq s \leq t)$:

$$\hat{L}_n^{Cox}(\beta) = \prod_{i=1}^n \left[\frac{e^{\beta' \hat{m}_{Z|W}(Z_i(X_i)|W_i(s), 0 \leq s \leq X_i)}}{\sum_{j=1}^n e^{\beta' \hat{m}_{Z|W}(Z_j(X_i)|W_j(s), 0 \leq s \leq X_i)} 1_{\{X_i \leq X_j\}}} \right]^{\Delta_i}.$$

Plus précisément, Tsiatis, DeGruttola et Wulfsohn supposent que pour chaque sujet i ($i = 1, \dots, n$), la covariable Z est observée avec erreur aux instants t_j ($j \geq 1$), jusqu'au décès du sujet ou jusqu'à une censure. Les variables observées sont donc, pour chaque i , $(X_i, \Delta_i, W_i(t_1), \dots, W_i(t_{m_i}))$, où t_{m_i} est le dernier instant de mesure de Z avant l'instant X_i . Z est supposé unidimensionnel.

Les auteurs supposent que pour chaque i , $Z(t)$ peut être modélisé par

$$Z_i(t) = \alpha_{0i} + \alpha_{1i}t,$$

où $(\alpha_{0i}, \alpha_{1i})'$ est un vecteur aléatoire de loi normale bivariée, d'espérance θ et de matrice de covariance V . Les auteurs supposent ensuite un modèle d'erreur de mesure de la forme

$$W_i(t) = Z_i(t) + \epsilon_i,$$

où ϵ_i a une distribution normale centrée et de variance σ^2 , où les ϵ_i sont indépendants, et où ϵ_i et $(\alpha_{0i}, \alpha_{1i})'$ sont indépendants.

La loi conjointe de $Z_i(t)$ et $(W_i(t_1), \dots, W_i(t_{m_i}))$ est une loi normale multivariée et l'espérance $m_{Z|W}(Z_i(t)|W_i(t_1), \dots, W_i(t_{m_i}))$ de la loi conditionnelle de $Z_i(t)$ sachant $(W_i(t_1), \dots, W_i(t_{m_i}))$ peut être calculée explicitement. Elle dépend des paramètres θ , V et σ^2 que Tsiatis, DeGruttola et Wulfsohn estiment par la méthode des moindres carrés, ce qui permet ensuite de calculer les estimations $\hat{m}_{Z|W}(Z_i(t)|W_i(t_1), \dots, W_i(t_{m_i}))$.

Des simulations ont été réalisées par Dafni et Tsiatis [24] pour comparer les estimations du paramètre de régression du modèle de Cox obtenues en utilisant cette méthode et l'approche consistant à utiliser les observations de W dans la vraisemblance partielle. Ces simulations montrent que l'approche en deux étapes permet de diminuer le biais de l'estimateur.

Wulfsohn et Tsiatis [95] soulignent néanmoins plusieurs inconvénients de cette approche.

D'une part, la nécessité d'estimer, pour chaque instant d'événement t , les paramètres θ , V et σ^2 pour le calcul des espérances $\hat{m}_{Z|W}(Z_i(t)|W_i(t_1), \dots, W_i(t_{m_i}))$ pose un problème de parcimonie et d'interprétation de ces paramètres. D'autre part, leur estimation ne tient pas compte de l'information relative aux durées d'observation des sujets. Wulfsohn et Tsiatis émettent l'hypothèse que prendre en compte cette information peut permettre de diminuer le biais résiduel de l'estimateur de β , observé par Dafni et Tsiatis.

Approche par modélisation conjointe

Wulfsohn et Tsiatis [95] proposent de modéliser conjointement les durées de vie censurées et le processus de covariable.

Pour chaque sujet i , on dispose des observations $(X_i, \Delta_i, W_i(t_1), \dots, W_i(t_{m_i}))$. Un modèle pour (X, Δ, Z, W) est obtenu en spécifiant un modèle de Cox pour la loi conditionnelle de T sachant (Z, W) et une loi marginale pour (Z, W) . Wulfsohn et Tsiatis supposent que

$$W_i(t_j) = Z_i(t_j) + \epsilon_i(t_j)$$

et

$$Z_i(t_j) = \alpha_{0i} + \alpha_{1i}t_j,$$

où $(\alpha_{0i}, \alpha_{1i})'$ est un vecteur aléatoire de loi normale bivariée, d'espérance θ et de matrice de covariance V , et $\epsilon_i(t_j)$ est une erreur aléatoire de loi normale centrée et de variance σ^2 .

La fonction de risque $\lambda_{T|W,Z}(t)$ est supposée égale à

$$\lambda_{T|Z,W}(t) = \lambda_T(t)e^{\beta(\alpha_{0i} + \alpha_{1i}t)}.$$

Une vraisemblance est obtenue pour estimer les paramètres $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$, β , θ , V et σ^2 sous les hypothèses que la censure est indépendante des effets aléatoires α_{0i} et α_{1i} et est non-informative pour β , Λ_T , θ , V et σ^2 . Wulfsohn et Tsiatis utilisent la méthode d'estimation semiparamétrique et l'algorithme Espérance-Maximisation [26] pour estimer les paramètres du modèle conjoint proposé.

En revanche, les auteurs ne fournissent pas de résultats concernant l'identifiabilité du modèle proposé, l'existence d'estimateurs semiparamétriques ou les propriétés asymptotiques de tels estimateurs.

3.2 Cas des valeurs manquantes de la covariable

L'estimation des paramètres d'un modèle de Cox avec covariables dépendant du temps nécessite, pour chaque sujet i ($i = 1, \dots, n$) inclus dans l'étude, l'observation de toutes les valeurs $\{Z_i(s), 0 \leq s \leq X_i\}$ de la covariable jusqu'à ce que le sujet quitte l'étude. Cette condition n'est le plus souvent pas vérifiée dans la pratique de la modélisation des durées de vie.

Plusieurs méthodes ont été proposées pour estimer le paramètre de régression dans des situations de valeurs manquantes de covariables dépendant du temps. Nous expliquons dans cette partie le principe de quatre méthodes récemment développées.

3.2.1 La méthode proposée par Lin et Ying

Lin et Ying [56] proposent une méthode dite méthode Approximate Partial Likelihood Estimation (ou méthode d'estimation par vraisemblance partielle approchée, notée APLE par la suite).

Hypothèses et notations

Les auteurs supposent que la fonction de risque instantané associée à la survenue d'un événement d'intérêt est de la forme $\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta'Z(t)}$.

Le problème statistique est d'estimer les paramètres β et $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$ au vu de n observations indépendantes $(X_i, \Delta_i, Z_i(s), 0 \leq s \leq X_i)_{1 \leq i \leq n}$, lorsque $\{Z_i(s), 0 \leq s \leq X_i\}$ n'est pas entièrement observé pour certains sujets i ($i = 1, \dots, n$).

Les données disponibles pour chaque sujet i ($i = 1, \dots, n$) sont constituées du quintuplet $(X_i, \Delta_i, Z_i(s), H_i(s), H_{0i}(s), 0 \leq s \leq X_i)$, où l'ensemble d'observations $\{Z_i(s), 0 \leq s \leq X_i\}$ peut être incomplet. $H_i(t)$ est une matrice diagonale de dimension p dont les éléments diagonaux $H_{ir}(t)$ ($r = 1, \dots, p$) sont égaux à 1 si $Z_{ir}(t)$ est observé et 0 sinon, et $H_{0i}(t) = \prod_{r=1}^p H_{ir}(t)$. Avec ces notations, à l'instant t , $H_{0i}(t)$ indique si le vecteur de covariable $Z(t)$ est entièrement observé ou pas pour le sujet i . Si le vecteur $Z_i(t)$ n'est pas entièrement observé, la matrice $H_i(t)$ indique quelles sont ses composantes manquantes et observées. Les auteurs définissent $S^{(r)}(\beta, t)$ par :

$$S^{(r)}(\beta, t) = \frac{1}{n} \sum_{i=1}^n H_{0i}(t) \{Z_i(t)\}^{\otimes r} e^{\beta'Z_i(t)} 1_{\{t \leq X_i\}}, \quad r = 0, 1, 2,$$

qui sera un estimateur de

$$s^{(r)}(\beta, t) = E \left[H_{01}(t) \{Z_1(t)\}^{\otimes r} e^{\beta' Z_1(t)} 1_{\{t \leq X_1\}} \right], \quad r = 0, 1, 2,$$

où $E[U]$ désigne l'espérance de U .

Vraisemblance partielle approchée

La méthode APLE consiste à estimer les termes

$$\frac{\sum_{j=1}^n Z_j(X_i) e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}$$

figurant dans le vecteur du score (2.17) en utilisant seulement les sujets j ($j = 1, \dots, n$) à risque à l'instant X_i , pour lesquels la réalisation $Z_j(X_i)$ du processus p -dimensionnel à l'instant X_i est entièrement observée, c'est-à-dire les sujets j à risque tels que $H_{0j}(X_i) = 1$.

De plus, si pour un sujet i ($i = 1, \dots, n$) tel que $\Delta_i = 1$, une composante r ($r = 1, \dots, p$) du vecteur $Z_i(X_i)$ n'est pas observée, c'est-à-dire si $H_{ir}(X_i) = 0$, le terme

$$Z_{ir}(X_i) - \frac{\sum_{j=1}^n Z_{jr}(X_i) e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}.$$

est retiré de la somme (2.17).

Une approximation du score (2.17) est ainsi obtenue à partir des sujets pour lesquels Z est observé aux instants considérés, justifiant le nom de “vraisemblance partielle approchée” donné à la méthode. En utilisant les notations introduites précédemment, cette approximation s'écrit :

$$\sum_{i=1}^n \Delta_i H_i(X_i) \left[Z_i(X_i) - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right].$$

Ce vecteur sera appelé score approché et nous le noterons

$$\frac{\partial}{\partial \beta} l_n^{Cox}(\beta)_{appr}.$$

L'estimateur $\hat{\beta}_{n,appr}$ du maximum de vraisemblance partielle approchée de β est défini comme la solution de l'équation $\frac{\partial}{\partial \beta} l_n^{Cox}(\beta)_{appr} = 0$.

Sous des conditions de régularité énoncées dans [56], Lin et Ying établissent les propriétés asymptotiques de $\hat{\beta}_{n,appr}$.

Théorème 3.1 (Lin et Ying, 1993) *L'estimateur $\hat{\beta}_{n,appr}$ vérifie les propriétés suivantes :*

1. $\hat{\beta}_{n,appr}$ est un estimateur consistant de β .

2. Soit $A(\beta) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Delta_i H_i(X_i) \left[\frac{S^{(2)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} - \left\{ \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\}^{\otimes 2} \right]$ et

$$B(\beta) = E \left[\left\{ \Delta_1 H_1(X_1) \left[Z_1(X_1) - \frac{s^{(1)}(\beta, X_1)}{s^{(0)}(\beta, X_1)} \right] - \int_0^{X_1} \frac{E[H_1(t)]}{E[H_{01}(t)]} H_{01}(t) e^{\beta' Z_1(t)} \left[Z_1(t) - \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right] d\Lambda_T(t) \right\}^{\otimes 2} \right].$$

Alors $\sqrt{n}(\hat{\beta}_{n,appr} - \beta)$ converge en loi vers une variable aléatoire de loi normale centrée et de matrice de covariance $\Sigma(\beta)$ où $\Sigma(\beta) = A^{-1}(\beta)B(\beta)A^{-1}(\beta)'$.

Lin et Ying proposent ensuite d'estimer la fonction de risque cumulé $\Lambda_T(t)$ par :

$$\hat{\Lambda}_{n,appr}(\hat{\beta}_{n,appr}, t) = \sum_{i=1}^n \frac{\Delta_i H_{0i}(X_i) 1_{\{X_i \leq t\}}}{\sum_{j=1}^n H_{0j}(X_i) e^{\hat{\beta}_{n,appr}' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}.$$

Cet estimateur se ramène à l'estimateur de Breslow (2.15) dans le cas particulier où toutes les covariables sont entièrement observées. L'estimateur $\hat{\Lambda}_{n,appr}$ dépend de β , mais pour ne pas alourdir les notations, nous omettons cette dépendance. Dans ce qui suit, l'estimateur $\hat{\Lambda}_{n,appr}$ est supposé calculé pour $\beta = \hat{\beta}_{n,appr}$.

Lin et Ying démontrent la convergence en loi du processus $\sqrt{n}(\hat{\Lambda}_{n,appr} - \Lambda_T)$:

Théorème 3.2 (Lin et Ying, 1993) $\sqrt{n}(\hat{\Lambda}_{n,appr} - \Lambda_T)$ converge en loi vers un processus gaussien centré G , de fonction de covariance

$$\begin{aligned} cov[G(s), G(t)] &= \int_0^{s \wedge t} \frac{d\Lambda_T(u)}{s^{(0)}(\beta, u)} + J'(t)A^{-1}(\beta)B(\beta)A^{-1}(\beta)'J(s) \\ &\quad - J'(s)A^{-1}(\beta)G(t) - J'(t)A^{-1}(\beta)G(s), \end{aligned}$$

où $J(t)$ et $G(t)$ sont définis par

$$\begin{aligned} J(t) &= \int_0^t \frac{s^{(1)}(\beta, u) d\Lambda_T(u)}{s^{(0)}(\beta, u)}, \\ G(t) &= E \left[\int_0^\infty \int_0^{t \wedge v} \frac{H_{01}(u) e^{\beta' Z_1(u)} d\Lambda_T(u)}{s^{(0)}(\beta, u)} \left[H_1(v) - \frac{H_{01}(v)}{E[H_{01}(v)]} E[H_1(v)] \right] \right. \\ &\quad \times \left. \left[Z_1(v) - \frac{s^{(1)}(\beta, v)}{s^{(0)}(\beta, v)} \right] 1_{\{v \leq X_1\}} e^{\beta' Z_1(v)} d\Lambda_T(v) \right]. \end{aligned}$$

Remarques

1. Lin et Ying soulignent que la validité de la méthode APLE repose sur l'hypothèse suivante : sachant qu'un individu i est à risque à l'instant t , les variables aléatoires $H_{ir}(t)$, ($r = 1, \dots, p$) sont indépendantes de toutes les autres variables (X_i, Δ_i, Z_i) . En particulier, la probabilité de ne pas observer $Z_i(t)$ ne dépend pas de la valeur de $Z_i(t)$.
2. L'hypothèse précédente est appelée une hypothèse de valeurs manquantes complètement au hasard (ou MCAR pour Missing Completely At Random). Diggle, Liang et Zeger décrivent [29] trois types de données manquantes. Reprenant les notations précédentes, Z est dite manquante complètement au hasard si la variable $H_{ir}(t)$ ($t \geq 0, r = 1, \dots, p$) est indépendante de toutes les autres variables du modèle.

La covariable Z est dite manquante au hasard (ou MAR pour Missing At Random) si la variable $H_{ir}(t)$ ne dépend pas de la valeur de $Z_i(t)$ ($t \geq 0, r = 1, \dots, p$), c'est-à-dire si la probabilité de ne pas observer la réalisation de $Z_i(t)$ ne dépend pas de cette réalisation. Néanmoins, $H_{ir}(t)$ peut dépendre des autres variables, observées, du modèle.

Enfin, le mécanisme de données manquantes de Z est dit informatif si $H_{ir}(t)$ dépend de la valeur de $Z_i(t)$.

La possibilité d'estimer les termes

$$\frac{\sum_{j=1}^n Z_j(X_i) e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta' Z_j(X_i)} 1_{\{X_i \leq X_j\}}}$$

en utilisant seulement les sujets pour lesquels le processus Z est entièrement observé suppose que ces sujets soient représentatifs de l'échantillon des n sujets, soit, que le mécanisme de données manquantes soit MAR ou MCAR. Cette hypothèse ne serait pas satisfaite si les sujets à valeurs manquantes étaient ceux pour lesquels, par exemple, les valeurs de Z sont en moyenne inférieures à un certain seuil.

3.2.2 La méthode proposée par Zhou et Pepe

Zhou et Pepe [96] proposent une méthode d'imputation pour estimer le paramètre de régression d'un modèle de Cox avec valeurs manquantes de covariables dépendant du temps. Cette méthode consiste à remplacer la valeur manquante d'une covariable à un instant t par une moyenne calculée sur un sous-échantillon de sujets

pour lesquels le processus de covariable est entièrement observé.

Hypothèses

Zhou et Pepe considèrent le modèle de Cox défini par la fonction de risque instantané $\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta'Z(t)}$, et supposent que l'on observe les variables (X, Δ) et le processus Z sur n sujets indépendants. $Z(t)$ est supposé observé pour tout t sur un sous-échantillon V des n sujets. Son ensemble complémentaire \bar{V} contient les sujets pour lesquels Z n'est jamais observé.

Zhou et Pepe supposent que V est un sous-échantillon tiré aléatoirement de la population des n sujets. Le mécanisme de données manquantes de Z est donc MAR. Il serait informatif si par exemple, les sujets de l'ensemble \bar{V} étaient les sujets pour lesquels les valeurs de Z sont les plus basses.

Pour chaque sujet i ($i = 1, \dots, n$), Zhou et Pepe supposent aussi que l'on observe complètement un processus $(W(t))_{t \geq 0}$, où pour tout t , $W(t)$ est une variable catégorielle. W ne participe pas au modèle mais est supposée corrélée à Z : elle est appelée variable auxiliaire. Son rôle sera défini ci-dessous.

Estimation

Zhou et Pepe proposent d'estimer la vraisemblance partielle de Cox en remplaçant chaque valeur manquante $Z_i(t)$ par une moyenne des $Z_j(t)$ calculée sur les sujets j pour lesquels Z est entièrement observé (c'est-à-dire : $j \in V$) et tels que $W_j(t) = W_i(t)$. Cette condition sur W est motivée par l'argument suivant : si deux sujets partagent la même valeur d'une variable auxiliaire W fortement corrélée à Z , on peut penser que leurs valeurs de Z sont proches. La condition $W_j(t) = W_i(t)$ permet donc de sélectionner un sous-échantillon de sujets j dont le sujet i puisse être considéré comme un élément représentatif.

Plus précisément, Zhou et Pepe proposent d'estimer les termes $\{Z_i(t)\}^{\otimes r} e^{\beta'Z_i(t)}$ ($i \in \bar{V}$) ($r = 0, 1, 2$) par $\tilde{E} [\{Z_i(t)\}^{\otimes r} e^{\beta'Z_i(t)}]$, défini par :

$$\tilde{E} [\{Z_i(t)\}^{\otimes r} e^{\beta'Z_i(t)}] = \frac{\sum_{j=1}^n 1_{\{j \in V\}} 1_{\{W_j(t)=W_i(t)\}} 1_{\{t \leq X_j\}} \{Z_j(t)\}^{\otimes r} e^{\beta'Z_j(t)}}{\sum_{j=1}^n 1_{\{j \in V\}} 1_{\{W_j(t)=W_i(t)\}} 1_{\{t \leq X_j\}}}.$$

La notation suivante sera utile par la suite :

$$r_i^{(r)}(\beta, t) = \{Z_i(t)\}^{\otimes r} e^{\beta'Z_i(t)} 1_{\{i \in V\}} + \tilde{E} [\{Z_i(t)\}^{\otimes r} e^{\beta'Z_i(t)}] 1_{\{i \in \bar{V}\}}, \quad r = 0, 1, 2.$$

Les auteurs proposent ensuite de remplacer la vraisemblance partielle de Cox par une vraisemblance partielle estimée (cette terminologie justifie le nom donné à la méthode : Estimated Partial Likelihood) égale à

$$L_n^{Cox}(\beta)_{est} = \prod_{i=1}^n \left[\frac{r_i^{(0)}(\beta, X_i)}{\sum_{j=1}^n r_j^{(0)}(\beta, X_i) 1_{\{X_i \leq X_j\}}} \right]^{\Delta_i}.$$

L'estimateur $\hat{\beta}_{n,est}$ du maximum de vraisemblance partielle estimée de β est défini comme la solution de l'équation $\frac{\partial}{\partial \beta} \ln L_n^{Cox}(\beta)_{est} = 0$, où

$$\frac{\partial}{\partial \beta} \ln L_n^{Cox}(\beta)_{est} = \sum_{i=1}^n \left[\frac{r_i^{(1)}(\beta, X_i)}{r_i^{(0)}(\beta, X_i)} - \frac{\sum_{j=1}^n r_j^{(1)}(\beta, X_i) 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n r_j^{(0)}(\beta, X_i) 1_{\{X_i \leq X_j\}}} \right].$$

Les notations suivantes sont définies dans [96] :

$$\begin{aligned} s^{(r)}(\beta, t) &= E[r^{(r)}(\beta, t)], \\ \Sigma(\beta) &= \int_0^\tau \left[\frac{s^{(2)}(\beta, s)}{s^{(0)}(\beta, s)} - \left\{ \frac{s^{(1)}(\beta, s)}{s^{(0)}(\beta, s)} \right\}^{\otimes 2} \right] s^{(0)}(\beta, s) d\Lambda_T(s), \\ \Sigma_1(\beta) &= \int_0^\tau \left[E \left[\frac{\{E[Z_i(s)e^{\beta' Z_i(s)}]\}^{\otimes 2}}{E[e^{\beta' Z_i(s)}]} 1_{\{s \leq X_i\}} \right] - \frac{\{s^{(1)}(\beta, s)\}^{\otimes 2}}{s^{(0)}(\beta, s)} \right] d\Lambda_T(s), \\ \Sigma_2(\beta) &= E \left[\left\{ \int_0^\tau \left[\frac{Z_i(s)e^{\beta' Z_i(s)}}{E[e^{\beta' Z_i(s)}]} - \frac{s^{(1)}(\beta, s)}{s^{(0)}(\beta, s)} \right] dM_i(s) - \frac{1-\gamma}{\gamma} \right. \right. \\ &\quad \left. \left. \times \int_0^\tau \left[\frac{E[Z_i(s)e^{\beta' Z_i(s)}]}{E[e^{\beta' Z_i(s)}]} - \frac{s^{(1)}(\beta, s)}{s^{(0)}(\beta, s)} \right] \left[e^{\beta' Z_i(s)} - E[e^{\beta' Z_i(s)}] \right] 1_{\{s \leq X_i\}} d\Lambda_T(s) \right\}^{\otimes 2} \right], \end{aligned}$$

où $\gamma = \lim_{n \rightarrow \infty} \frac{\text{card } V}{n}$, $N_i(s) = 1_{\{\Delta_i=1, X_i \leq s\}}$ et $M_i(s) = N_i(s) - \int_0^s r_i^{(0)}(s) 1_{\{s \leq X_i\}} d\Lambda_T(s)$.

Sous des conditions de régularité énoncées dans [96], Zhou et Pepe démontrent le résultat suivant :

Théorème 3.3 (Zhou et Pepe, 1995) *L'estimateur $\hat{\beta}_{n,est}$ vérifie les propriétés suivantes :*

1. $\hat{\beta}_{n,est}$ est un estimateur consistant de β .
2. $\sqrt{n}(\hat{\beta}_{n,est} - \beta)$ converge en loi vers une variable aléatoire de loi normale centrée et de matrice de covariance $\Sigma(\beta)$, où

$$\Sigma(\beta) = \Sigma^{-1}(\beta) [(1-\gamma)\Sigma_1(\beta) + \gamma\Sigma_2(\beta)] \Sigma^{-1}(\beta)'$$

Remarques

1. Zhou et Pepe signalent que lorsque γ tend vers 0, c'est-à-dire lorsque la proportion de sujets utilisables pour l'estimation des données manquantes devient voisine de 0, les termes de la matrice de covariance asymptotique augmentent, entraînant l'imprécision des estimations obtenues.
2. Un inconvénient de la méthode EPL réside dans la nécessité de disposer pour tous les sujets d'une covariable auxiliaire complètement observée et qui soit suffisamment corrélée avec la covariable entrant dans le modèle de Cox.

3.2.3 La méthode proposée par Wang, Hsu, Feng et Prentice

Le principe de la méthode proposée par Wang, Hsu, Feng et Prentice [92] est analogue, dans le contexte des données manquantes d'une covariable, à l'approche en deux étapes proposée par Tsiatis, DeGruttola et Wulfsohn [85] dans le contexte de l'erreur de mesure. Un modèle de régression de la variable manquante sur des variables explicatives toujours observées est d'abord proposé. Ses paramètres sont estimés à partir d'un sous-échantillon de la population pour lequel toutes les variables sont observées. Les valeurs manquantes de la covariable du modèle de Cox sont ensuite estimées à l'aide du modèle de régression, et utilisées dans la vraisemblance partielle pour estimer le paramètre β du modèle de Cox.

Cette méthode est motivée par des difficultés survenant dans l'implémentation de la méthode EPL lorsque la variable auxiliaire W est continue, ou que la covariable manquante est de grande dimension.

Pour simplifier les notations, nous présentons cette méthode dans le cas particulier où la covariable manquante Z est unidimensionnelle.

Wang, Hsu, Feng et Prentice supposent que la fonction de risque d'occurrence d'un événement d'intérêt est

$$\lambda_{T|Z,Y}(t) = \lambda_T(t) e^{\beta_0 Z(t) + \beta_1' Y(t)},$$

où Y est une covariable dépendant du temps toujours observée et Z peut être manquant. On observe aussi sur les n sujets suivis les réalisations d'une variable W (qui peut aussi dépendre du temps) explicative de Z . Soit V un sous-échantillon tiré aléatoirement parmi les n sujets.

On suppose que l'on observe complètement $(X, \Delta, W, Y(s), Z(s), 0 \leq s \leq X)$ sur les

sujets de V , tandis que l'on observe complètement que $(X, \Delta, W, Y(s), 0 \leq s \leq X)$ sur les autres sujets.

Estimation du paramètre de régression

Wang, Hsu, Feng et Prentice supposent un modèle de régression de Z sur les variables W et Y :

$$E(Z(t)|W, Y(s), 0 \leq s \leq t) = g(t, W, Y; \alpha)$$

où α est un vecteur de paramètres inconnus.

Le paramètre α est d'abord estimé à partir du sous-échantillon V puis les valeurs manquantes $Z(t)$ sont estimées par $g(t, W, Y; \hat{\alpha})$, où $\hat{\alpha}$ désigne une estimation de α . Soit $\eta_i(t)$ la fonction indicatrice valant 1 si $Z_i(t)$ est observé et 0 sinon. Notons $Z_i^*(t) = Z_i(t)^{\eta_i(t)} \cdot g(t, W, Y; \alpha)^{1-\eta_i(t)}$ et $\hat{Z}_i^*(t) = Z_i(t)^{\eta_i(t)} \cdot g(t, W, Y; \hat{\alpha})^{1-\eta_i(t)}$.

Le vecteur $\beta = (\beta_0, \beta_1)'$ est estimé en utilisant la vraisemblance partielle de Cox dans laquelle les $Z(t)$ manquants sont remplacés par $g(t, W, Y; \hat{\alpha})$, ce qui revient à résoudre le système d'équations de vraisemblance

$$U_n(\beta, \hat{\alpha}) = \sqrt{n} \sum_{i=1}^n \Delta_i \begin{pmatrix} \hat{Z}_i^*(X_i) - \frac{\sum_{j=1}^n \hat{Z}_j^*(X_i) e^{\beta_0 \hat{Z}_j^*(X_i) + \beta_1' Y_j(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta_0 \hat{Z}_j^*(X_i) + \beta_1' Y_j(X_i)} 1_{\{X_i \leq X_j\}}} \\ Y_i(X_i) - \frac{\sum_{j=1}^n Y_j(X_i) e^{\beta_0 \hat{Z}_j^*(X_i) + \beta_1' Y_j(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta_0 \hat{Z}_j^*(X_i) + \beta_1' Y_j(X_i)} 1_{\{X_i \leq X_j\}}} \end{pmatrix} = 0$$

Un inconvénient de cette méthode est que la suite des estimateurs $\hat{\beta}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n})'$ obtenus à partir de ce système d'équations ne converge pas vers β . Wang, Hsu, Feng et Prentice montrent que $\hat{\beta}_n$ converge vers la solution $\beta^* = (\beta_0^*, \beta_1^{*T})'$ de $E[U_1(\beta, \alpha)] = 0$. Les auteurs soulignent que l'estimation du paramètre β est d'autant plus proche de β que $g(t, W, Y; \alpha)$ est proche de la vraie relation entre $Z(t)$ et (W, Y) .

De même que la méthode EPL, cette méthode présente l'inconvénient de nécessiter l'existence et l'observation de variables explicatives de la variable manquante ainsi que l'existence d'un sous-échantillon de validation. Elle requiert aussi la détermination d'un modèle $g(t, W, Y; \alpha)$.

3.2.4 La méthode proposée par Paik et Tsai

La méthode proposée par Paik et Tsai [71] pour estimer le paramètre de régression d'un modèle de Cox en cas de valeurs manquantes de covariables dépendant du

temps est une méthode d'imputation. Elle consiste à remplacer les valeurs manquantes à un instant t pour un sujet i ($i = 1, \dots, n$) par une estimation calculée à partir des individus à risque et observés à t .

Dans cette méthode, pour qu'un sujet j ($j = 1, \dots, n$) participe à l'estimation d'un $Z_i(t)$ manquant, il n'est pas nécessaire que $Z_j(t)$ soit observé pour tout t , contrairement à ce qui est proposé par Zhou et Pepe [96].

Soit $t < t' < t''$ trois instants distincts. Un sujet j peut être utilisé pour estimer $Z_i(t)$ et $Z_i(t'')$ si $Z_j(t)$ et $Z_j(t'')$ sont observés, et ne pas être utilisé pour estimer $Z_i(t')$ si $Z_j(t')$ est manquant. $Z_j(t')$ sera dans ce cas lui-même estimé. Le sous-échantillon utilisé pour l'estimation des valeurs manquantes est dynamique.

Cette méthode est donc applicable même lorsque la proportion de sujets pour lesquels Z est observé pour tout t est faible, ce qui n'était pas le cas de la méthode EPL [96].

Hypothèses et notations

Paik et Tsai supposent que la fonction de risque instantané est de la forme $\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta'Z(t)}$ et s'intéressent à l'estimation du paramètre β au vu d'un n -échantillon $(X_i, \Delta_i, Z_i(s), 0 \leq s \leq X_i)_{1 \leq i \leq n}$.

Ils supposent que $Z(t) = (Z'_1(t), Z'_2(t))'$ est un $(p + q)$ -vecteur de covariables dont la première composante est le q -vecteur $Z_1(t)$, toujours observé. Les valeurs du p -vecteur $Z_2(t)$ peuvent manquer. Le vecteur de paramètres β se décompose de même en $(\beta'_1, \beta'_2)'$ où β_1 et β_2 sont respectivement un q et un p -vecteur.

Paik et Tsai adoptent dans [71] des notations proches de celles de Lin et Ying [56]. En particulier, les observations sont, pour chaque sujet i ($i = 1, \dots, n$), le quintuplet $(X_i, \Delta_i, Z_i(s), H_i(s), H_{0i}(s), 0 \leq s \leq X_i)$. Soit $H_{ir}(t)$ la fonction indicatrice valant 1 si la r -ème coordonnée de $Z_i(t)$ est observée et 0 sinon (notons que $H_{ir}(t) = 1$ si $r \in \{1, \dots, q\}$). $H_i(t)$ est une matrice diagonale de dimension p et d'éléments diagonaux $H_{ir}(t)$ ($r = q + 1, \dots, q + p$). Enfin $H_{0i}(t) = \prod_{r=q+1}^{q+p} H_{ir}(t)$ indique si le sous-vecteur $Z_{2i}(t)$ est complètement observé ou pas.

Paik et Tsai supposent que $H_{ir}(t)$ et $Z'_{2i}(t)$ sont indépendants. Dans la terminologie de Diggle, Liang et Zeger [29], Paik et Tsai supposent que les valeurs sont manquantes au hasard ou complètement au hasard.

En tenant compte de la décomposition de Z , nous pouvons réécrire le vecteur

du score $\frac{\partial}{\partial \beta} l_n^{Cox}(\beta)$ sous la forme suivante :

$$\sum_{i=1}^n \Delta_i \left[\begin{pmatrix} Z_{1i}(X_i) \\ Z_{2i}(X_i) \end{pmatrix} - \frac{\sum_{j=1}^n \begin{pmatrix} Z_{1j}(X_i) \\ Z_{2j}(X_i) \end{pmatrix} e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}} \right]. \quad (3.1)$$

Principe de la méthode

Si pour un sujet i ($i = 1, \dots, n$) non censuré, la coordonnée r ($r = 1, \dots, p$) de $Z_{2i}(X_i)$ (notée $Z_{2ir}(X_i)$, qui est aussi la coordonnée $(q + r)$ de $Z_i(X_i)$) n'est pas observée, le terme

$$Z_{2ir}(X_i) - \frac{\sum_{j=1}^n Z_{2jr}(X_i) e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}}$$

est retiré du vecteur du score (3.1).

Le principe d'imputation à partir duquel Paik et Tsai obtiennent une fonction score pour le paramètre β intervient dans le calcul des termes

$$\frac{\sum_{j=1}^n \begin{pmatrix} Z_{1j}(X_i) \\ Z_{2j}(X_i) \end{pmatrix} e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}}{\sum_{j=1}^n e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}}$$

figurant dans l'expression (3.1).

Nous décrivons la démarche proposée, pour le terme $\sum_{j=1}^n e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}$.

Si $H_{j(q+r)}(t) = 0$ pour certains sujets j ($j = 1, \dots, n$), c'est-à-dire si $Z_{2jr}(t)$ n'est pas observé (pour $r \in \{1, \dots, p\}$), Paik et Tsai proposent d'estimer les termes

$$\frac{\begin{pmatrix} Z_{1j}(t) \\ Z_{2j}(t) \end{pmatrix} e^{\beta'_1 Z_{1j}(t) + \beta'_2 Z_{2j}(t)} 1_{\{t \leq X_j\}}}{e^{\beta'_1 Z_{1j}(t) + \beta'_2 Z_{2j}(t)} 1_{\{t \leq X_j\}}}$$

en remplaçant chacun des $e^{\beta_{2r} Z_{2jr}(t)}$ manquants par une moyenne des termes $e^{\beta_{2r} Z_{2kr}(t)}$ obtenus sur les individus k ($k = 1, \dots, n$) à risque à l'instant t , pour lesquels $Z_{2kr}(t)$ est observé et tels que $Z_{1k}(t) = Z_{1j}(t)$. Cette moyenne est notée $\tilde{E} [e^{\beta_{2r} Z_{2jr}(t)}]$ et se calcule comme suit :

$$\tilde{E} [e^{\beta_{2r} Z_{2jr}(t)}] = \frac{\sum_{k=1}^n H_{k(q+r)}(t) e^{\beta_{2r} Z_{2kr}(t)} 1_{\{Z_{1k}(t) = Z_{1j}(t)\}} 1_{\{t \leq X_k\}}}{\sum_{k=1}^n H_{k(q+r)}(t) 1_{\{Z_{1k}(t) = Z_{1j}(t)\}} 1_{\{t \leq X_k\}}},$$

où la condition $\{Z_{1k}(t) = Z_{1j}(t)\}$ permet de sélectionner un sous-ensemble de sujets k dont le sujet j puisse être considéré comme un élément représentatif (Z_1 joue ici le même rôle que la variable auxiliaire requise par Zhou et Pepe pour appliquer leur méthode).

Le terme $e^{\beta'_2 Z_{2j}(t)}$ peut ensuite être estimé par

$$\tilde{E} \left[e^{\beta'_2 Z_{2j}(t)} \right] = \prod_{r=1}^p e^{H_{j(q+r)}(t) \beta_{2r} Z_{2jr}(t)} \left[\tilde{E} \left[e^{\beta_{2r} Z_{2jr}(t)} \right] \right]^{1-H_{j(q+r)}(t)}$$

Finalement, $\sum_{j=1}^n e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}$ peut être estimé par

$$\begin{aligned} \tilde{S}^{(0)}(\beta, X_i) = \\ \sum_{j=1}^n e^{\beta'_1 Z_{1j}(X_i)} \left[H_{0j}(X_i) e^{\beta'_2 Z_{2j}(X_i)} + (1 - H_{0j}(X_i)) \tilde{E} \left[e^{\beta'_2 Z_{2j}(X_i)} \right] \right] 1_{\{X_i \leq X_j\}}. \end{aligned}$$

Le terme $\begin{pmatrix} Z_{1j}(X_i) \\ Z_{2j}(X_i) \end{pmatrix} e^{\beta'_1 Z_{1j}(X_i) + \beta'_2 Z_{2j}(X_i)} 1_{\{X_i \leq X_j\}}$ peut être estimé de même par $\tilde{S}_r^{(1)}(\beta, X_i) = \frac{\partial}{\partial \beta_r} \tilde{S}^{(0)}(\beta, X_i)$ $r = 1, 2$.

Paik et Tsai définissent alors un estimateur $\hat{\beta}_{n,imp}$ de β comme la solution du système d'équations

$$\frac{\partial}{\partial \beta} l_n^{Cox}(\beta)_{imp} = \begin{pmatrix} \sum_{i=1}^n \Delta_i \left[Z_{1i}(X_i) - \frac{\tilde{S}_1^{(1)}(\beta, X_i)}{\tilde{S}^{(0)}(\beta, X_i)} \right] \\ \sum_{i=1}^n \Delta_i H_i(X_i) \left[Z_{2i}(X_i) - \frac{\tilde{S}_2^{(1)}(\beta, X_i)}{\tilde{S}^{(0)}(\beta, X_i)} \right] \end{pmatrix} = 0,$$

et énoncent, sous des conditions de régularité données dans [71] le résultat de convergence suivant :

Théorème 3.4 (Paik et Tsai, 1997) $\hat{\beta}_{n,imp}$ est un estimateur consistant de β .

Un résultat de normalité asymptotique de $\hat{\beta}_{n,imp}$ est obtenu par Paik et Tsai lorsque

Z est binaire. Notons

$$\begin{aligned} a_i(\beta, t) &= [Z_i(t) - \tilde{e}(\beta, t)] e^{\beta' Z_i(t)} H_{0i}(t) [E[H_{0i}(t) | Z_{1i}(t)]]^{-1} 1_{\{t \leq X_i\}}, \\ \Psi_0(\beta, Z_{1i}(t), t) &= E \left[\frac{a_1(\beta, t) 1_{\{Z_{11}(t) = Z_{1i}(t)\}}}{Pr[X_i \geq t, Z_{11}(t) = Z_{1i}(t)]} \right], \\ \Psi_1(\beta, Z_{1i}(t), t) &= E \left[\frac{a_1(\beta, t) 1_{\{Z_{11}(t) = Z_{1i}(t)\}}}{Pr[X_i \geq t, Z_{11}(t) = Z_{1i}(t)] E[H_{0i}(t) | Z_{11}(t) = Z_{1i}(t)]} \right], \\ \tilde{e}(\beta, t) &= \begin{pmatrix} E[\tilde{S}_1^{(1)}(\beta, t)] / E[\tilde{S}^{(0)}(\beta, t)] \\ E[\tilde{S}_2^{(1)}(\beta, t)] / E[\tilde{S}^{(0)}(\beta, t)] \end{pmatrix}. \end{aligned}$$

Notons enfin

$$\begin{aligned} A(\beta) &= \lim_{n \rightarrow \infty} \frac{\partial^2}{\partial \beta \partial \beta'} l_n^{Cox}(\beta)_{imp}, \\ B(\beta) &= E \left[\frac{1}{n} \sum_{i=1}^n \{ \Delta_i H_i(X_i) [Z_i(X_i) - \tilde{e}(\beta, X_i)] \right. \\ &\quad - \int_0^{X_i} \frac{E[H_i(t)]}{E[H_{0i}(t) | Z_{1i}(t)]} H_{0i}(t) e^{\beta' Z_i(t)} [Z_i(t) - \tilde{e}(\beta, t)] d\Lambda_T(t) \\ &\quad \left. - \int_0^{X_i} E[H_i(t)] [\Psi_0(\beta, Z_{1i}(t), t) - H_{0i}(t) \Psi_1(\beta, Z_{1i}(t), t)] d\Lambda_T(t) \}^{\otimes 2} \right]. \end{aligned}$$

Paik et Tsai montrent le résultat suivant :

Théorème 3.5 (Paik et Tsai, 1997) $\sqrt{n}(\hat{\beta}_{n,imp} - \beta)$ converge en loi vers une variable aléatoire de loi normale centrée et de matrice de covariance $\Sigma(\beta)$ où $\Sigma(\beta) = A^{-1}(\beta)B(\beta)A^{-1}(\beta)'$.

3.3 Erreur de mesure et valeurs manquantes : modélisation conjointe pour une covariable fixe

Pour estimer les paramètres d'un modèle de Cox avec erreur de mesure sur une covariable dépendant du temps, Wulfsohn et Tsiatis ont proposé de modéliser conjointement les durées de vie censurées et le processus de covariables. Des approches basées sur le même principe ont été récemment proposées pour estimer les paramètres dans le modèle de Cox avec des covariables fixes, dans des situations d'erreur de mesure et de données manquantes des covariables.

Ainsi, Hu, Tsiatis et Davidian [43] ont réalisé des simulations pour comparer les estimations du paramètre de régression du modèle de Cox, obtenues à partir de la modélisation conjointe et de l'approche consistant à utiliser les mesures erronées dans la vraisemblance partielle. Ces simulations ont indiqué une diminution du biais de l'estimateur de β lorsque l'on utilise la modélisation conjointe.

Li et Lin [55] ont récemment proposé une modélisation conjointe pour estimer les paramètres d'un modèle de fragilité avec erreur de mesure sur une covariable fixe dans le temps. Les auteurs adoptent la méthode d'estimation semiparamétrique dans ce modèle. Leurs simulations fournissent des résultats encourageants, relativement à la convergence et à la normalité asymptotique de l'estimateur semiparamétrique du paramètre de régression β .

Martinussen [60] a proposé une approche par modélisation conjointe pour estimer les paramètres d'un modèle de Cox en présence de données manquantes d'une covariable fixe dans le temps, lorsque les données sont manquantes au hasard. Martinussen estime les paramètres du modèle par la méthode semiparamétrique et un algorithme Espérance-Maximisation. Dans [16], Chen et Little proposent une approche similaire sous la même hypothèse d'un mécanisme MAR de données manquantes. Chen et Little obtiennent des résultats de convergence et de normalité asymptotique pour l'estimateur du paramètre de régression β .

Trois méthodes d'estimation dans le modèle de Cox avec valeurs manquantes d'une covariable dépendant du temps présentées ci-dessus, et la méthode proposée par Martinussen et Chen et Little dans le cas de covariables fixes permettent d'obtenir des estimateurs convergents et asymptotiquement normaux du paramètre de régression β . Des résultats de convergence et de distribution asymptotique sont également obtenus pour l'estimateur de la fonction de risque cumulé.

Néanmoins, ces méthodes reposent sur des hypothèses contraignantes. Les valeurs des covariables doivent être manquantes au hasard ou complètement au hasard. De plus, pour utiliser la méthode EPL [96] et la méthode proposée par Wang, Hsu, Feng et Prentice [92], il faut disposer d'un sous-échantillon de sujets pour lesquels l'ensemble des valeurs des covariables est observé (notons que Kulich et Lin [53] supposent une condition analogue pour estimer les paramètres d'un modèle additif semiparamétrique dont une covariable est mesurée avec erreur). Ces hypothèses sont restrictives et souvent, ne sont pas satisfaites dans la pratique des études de survie.

Dans la deuxième partie de ce travail, nous nous intéressons au problème de

l'estimation des paramètres du modèle de Cox lorsque des valeurs d'une covariable dépendant du temps sont manquantes, que le mécanisme de données manquantes n'est ni MCAR, ni MAR, mais informatif, et que l'on ne dispose pas d'un sous-échantillon de sujets pour lesquels la covariable est entièrement observée.

Deuxième partie

Modèle de Cox avec valeurs
manquantes d'une covariable
dépendant du temps : proposition
d'une approche par modélisation
conjointe

Chapitre 4

Proposition d'un modèle conjoint et identifiabilité des paramètres

Dans ce chapitre, nous proposons une méthode pour estimer les paramètres d'un modèle de Cox lorsqu'une covariable dépendant du temps n'est pas complètement observée. Elle consiste à modéliser conjointement les durées de vie censurées et le processus de covariable.

Nous formalisons tout d'abord le problème, puis nous décrivons le modèle conjoint proposé. Nous établissons ensuite une vraisemblance pour ses paramètres. Le paramètre de régression du modèle (marginal) de Cox est le paramètre d'intérêt de ce modèle conjoint, tandis que la fonction de risque cumulé et les paramètres de la loi de la covariable sont des paramètres de nuisance. Nous démontrons enfin que le modèle conjoint proposé est identifiable sous certaines conditions.

4.1 Notations et formulation du problème

Dans cette section, nous définissons les notations utilisées dans la suite de ce travail et nous formulons précisément le problème statistique étudié.

Soit $(\Omega, \mathcal{A}, \mathbb{P}, (Z(t))_{t \geq 0})$ un processus stochastique à valeurs réelles indexé par le temps. Supposons que l'on suive n sujets au cours du temps, et que sur chacun de ces sujets i ($i = 1, \dots, n$), on observe une réplique $(Z_i(t))_{t \geq 0}$ du processus.

Soit T une variable aléatoire positive représentant une durée jusqu'à un événement d'intérêt. Nous supposons que la durée T peut être censurée à droite à un instant aléatoire C . Nous notons $\Delta = 1_{\{T \leq C\}}$ et $X = T \wedge C$.

Nous nous intéressons à l'évaluation de l'influence du processus de covariable $(Z(t))_{t \geq 0}$ sur la survenue de l'événement d'intérêt. Le modèle semiparamétrique de Cox incluant $(Z(t))_{t \geq 0}$ comme une covariable dépendant du temps, et défini par la fonction de risque instantané

$$\lambda_{T|Z}(t) = \lambda_T(t) e^{g(Z(s), 0 \leq s \leq t; \beta)}$$

permet cette évaluation.

Mais nous allons nous intéresser au problème de cette évaluation dans un contexte de données manquantes de la covariable $(Z(t))_{t \geq 0}$ où le recours à la méthode du maximum de vraisemblance partielle et aux méthodes décrites dans le chapitre 3 n'est pas possible.

Soit $t_0 < t_1 < \dots < t_{j-1} < t_j < \dots < \infty$ un ensemble d'instants fixés, connus, et vérifiant $t_0 = 0$ et $0 < \epsilon_0 \leq t_j - t_{j-1} \leq \epsilon_1 < \infty$, $j = 1, 2, \dots$. Les instants t_j sont les instants d'observation du processus $(Z(t))_{t \geq 0}$. Nous supposons que $(Z(t))_{t \geq 0}$ est un processus en escalier, constant sur les intervalles $[t_{j-1}, t_j]$. Pour $j \geq 0$, posons $Z_j = Z(t_j)$, et notons z_j une réalisation de la variable Z_j . De même, posons $Z_{ij} = Z_i(t_j)$ et notons z_{ij} une réalisation de la variable Z_{ij} , pour un sujet i ($i = 1, \dots, n$). Nous supposons que pour des raisons liées au plan de l'étude mesurant le processus de covariable, la réalisation z_j de Z_j sur l'intervalle $[t_{j-1}, t_j]$ n'est connue qu'à la fin de l'intervalle, soit à l'instant t_j (deux exemples pratiques de telles situations sont décrits dans le chapitre 7 de ce manuscrit).

Nous introduisons une notation qui permettra de faire référence à la dernière réalisation z_j observée avant un instant t ($t > 0$) quelconque. Pour un instant t ($t > 0$) quelconque, soit $a_t = \max(k : t_k < t)$. Notons que si l'instant t appartient à l'intervalle $[t_j, t_{j+1}]$, alors $a_t = j$, et la dernière valeur observée de la covariable avant t est $z_{a_t} = z_j$ (si $t \in [t_j, t_{j+1}]$, la dernière valeur de la covariable observée avant t sur le sujet i ($i = 1, \dots, n$) est $z_{ia_t} = z_{ij}$).

Pour un instant t quelconque ($t > 0$), nous notons par $W(t)$ le vecteur aléatoire

$$W(t) = (Z_{a_t}, Z(t))',$$

ayant pour composantes la dernière valeur de la covariable observée avant t et sa valeur à l'instant t . Nous notons par $W_i(t) = (Z_{ia_t}, Z_i(t))'$ ce vecteur pour un sujet i ($i = 1, \dots, n$).

La méthode du maximum de vraisemblance partielle requiert l'observation, pour chaque sujet i ($i = 1, \dots, n$), de l'ensemble des valeurs $\{Z_i(s) : 0 \leq s \leq X_i\}$ du processus $(Z(t))_{t \geq 0}$ jusqu'à l'instant X_i où ce sujet quitte l'étude. L'ensemble $\{Z_i(s) : 0 \leq s \leq X_i\}$ est ici l'ensemble $\{Z_{i0}, \dots, Z_{ia_{X_i}}, Z_i(X_i)\}$. Or dans le cadre de l'étude décrite, la valeur $Z_i(X_i)$ n'est pas observée.

Les méthodes d'estimation dans le modèle de Cox avec valeurs manquantes d'une covariable dépendant du temps rappelées dans le chapitre 3 supposent la condition de données manquantes au hasard ou complètement au hasard. Or cette hypothèse n'est pas valable dans le contexte que nous décrivons. En effet, pour le modèle $\lambda_{T|Z}(t) = \lambda_T(t)e^{g(Z(s), 0 \leq s \leq t; \beta)}$, la probabilité de survenue de l'événement d'intérêt à l'instant t dépend de la valeur inobservée $Z(t)$. La probabilité de ne pas observer $Z(t)$ dépend donc de la valeur inobservée $Z(t)$. Ceci place notre problème dans le contexte des données manquantes dites informatives (c'est le troisième type de données manquantes décrit par Diggle, Liang et Zeger [29]).

La réalisation de $Z_i(X_i)$ n'est observée pour aucun sujet i , nous ne disposons donc pas non plus d'un échantillon de sujets pour lesquels la covariable est complètement observée. Les méthodes proposées par Wang, Hsu, Feng et Prentice [92] et Zhou et Pepe [96] ne sont donc pas applicables.

La loi conditionnelle de T sachant $(Z(t))_{t \geq 0}$ est caractérisée par la fonction de risque instantané (que nous noterons $\lambda_{T|Z}$), que nous supposons de la forme suivante dans la suite de ce travail :

$$\begin{aligned} \lambda_{T|Z}(t) &= \lambda_T(t)e^{\beta'W(t)} \\ &= \lambda_T(t)e^{\beta_1 Z_{a_t} + \beta_2 Z(t)}, \quad t > 0, \end{aligned}$$

où $\beta = (\beta_1, \beta_2)'$ est un vecteur de paramètres inconnus et λ_T est une fonction de risque de base inconnue.

Ce choix de la fonction de régression nous permet d'inclure dans une même formulation les modèles

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta Z(t)} \tag{4.1}$$

et

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta(Z(t) - Z_{a_t})} \tag{4.2}$$

rappelés dans le chapitre 2, qui sont les plus utilisés dans les applications du modèle de Cox.

Supposons que l'on dispose d'un échantillon indépendant $(X_i, \Delta_i, (Z_i(t))_{t \geq 0})_{1 \leq i \leq n}$ du triplet $(X, \Delta, (Z(t))_{t \geq 0})$, où $X = T \wedge C$ et $\Delta = 1_{\{T \leq C\}}$. Nous notons par

$$Y_i = (X_i, \Delta_i, Z_{i0}, \dots, Z_{ia_{X_i}})$$

le vecteur des observations incomplètes pour le sujet i ($i = 1, \dots, n$) et par $y_i = (x_i, \delta_i, z_{i0}, \dots, z_{ia_{x_i}})$ une réalisation de ce vecteur.

Le problème statistique est d'estimer les paramètres β et $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$ au vu des observations incomplètes fournies par les réalisations de n vecteurs $(Y_i)_{1 \leq i \leq n}$.

4.2 Proposition d'une vraisemblance conjointe

Nous proposons de modéliser conjointement les durées de vie censurées et le processus de covariable, pour en déduire une vraisemblance qui nous permette d'estimer β et Λ_T au vu des données incomplètes.

Nous supposons les conditions suivantes sur le modèle conjoint proposé :

- C1** L'expérience est interrompue à un instant τ tel que $\mathbb{P}[X > \tau] > 0$.
- C2** La covariable est uniformément bornée en variation totale, c'est-à-dire que $\int_0^\infty |dZ_i(t)| + |Z_i(0)| < K$ ($0 < K < \infty$) pour tout i ($i = 1, \dots, n$). Cette condition assure que la covariable ne connaisse pas trop de fluctuations. Notons c le réel positif fini tel que $|Z_i(t)| \leq c$ pour tout i ($i = 1, \dots, n$) et $t \in [0, \tau]$.
- C3** Pour chaque $j \geq 1$, le vecteur aléatoire $(Z_0 = Z(t_0), \dots, Z_j = Z(t_j))$ à valeurs dans \mathbb{R}^{j+1} admet une densité $f(z_0, \dots, z_j; \alpha)$ sur \mathbb{R}^{j+1} , paramétrée par $\alpha \in \mathbb{R}^p$ ($p \geq 1$). Nous supposons que f est continue en α , uniformément bornée, deux fois dérivable par rapport aux composantes de α et que ses dérivées secondes sont continues.

Nous supposons que la densité de la loi de $Z_j = Z(t_j)$ sachant $Z_{j-1} = z_{j-1}, \dots, Z_0 = z_0$ (nous noterons $f(z_j | z_{j-1}, \dots, z_0; \alpha)$ cette densité) ne dépend pas de z_{j-2}, \dots, z_0 , et est de la forme

$$f(z_j | z_{j-1}, \dots, z_0; \alpha) = f(z_j | z_{j-1}; \alpha), \alpha \in \mathbb{R}^p.$$

Cette hypothèse correspond à l'idée de pouvoir résumer l'information contenue dans les valeurs passées de la covariable par un nombre fini (ici égal à 1) de ses valeurs retardées. Elle est une hypothèse raisonnable pour les exemples

d'applications traités dans le chapitre 7.

Soit $f(z_0; \alpha)$ la densité de Z_0 pour le paramètre α . Nous supposons la condition d'identifiabilité suivante : $f(z_0; \alpha) = f(z_0; \tilde{\alpha})$ presque partout implique $\alpha = \tilde{\alpha}$.

C4 L'espace des paramètres α est un compact de \mathbb{R}^p noté $A = [-a, a]^p$. L'espace des paramètres β est un compact de \mathbb{R}^2 noté $B = [-b, b]^2$. Nous désignerons par α_0 et β_0 les valeurs "vraies" des paramètres α et β .

C5 L'espace des fonctions Λ_T est l'espace L des fonctions définies sur $[0, \infty[$, continues, croissantes, positives et telles que $\Lambda_T(0) = 0$. Nous supposons que $\Lambda_T(\tau) < \infty$. Soit $t^* = \inf\{t : \Lambda_T(t) > 0\}$. Nous supposons que $t^* \neq t_j$ ($j = 0, 1, \dots$). Nous noterons par $\Lambda_{T,0}$ la "vraie" fonction de risque cumulé.

Par la suite, nous noterons $\theta = (\alpha, \beta, \Lambda_T)$ le paramètre du modèle conjoint. Nous noterons $\theta_0 = (\alpha_0, \beta_0, \Lambda_{T,0})$ sa vraie valeur.

C6 Nous supposons que les matrices

$$-E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z; \alpha_0) \right] \text{ et } E_{\theta_0} \left[\int_0^X W(u) W(u)' e^{\beta_0' W(u)} d\Lambda_{T,0}(u) \right]$$

sont définies positives. Nous supposons qu'il existe un q fini ($q > 0$) tel que

$$E_{\theta_0} \left[e^{\beta_0' W(u)} 1_{\{u \leq X\}} \right] > q, \forall u \in [0, \tau].$$

Nous étendons l'hypothèse de censure non-informative à la condition suivante : la distribution de la censure C ne dépend pas des paramètres β, Λ_T et α . Nous supposons que T et C sont indépendants conditionnellement à la covariable et que C ne dépend pas des valeurs inobservées de cette covariable.

Nous obtenons une vraisemblance pour les paramètres α, β et Λ_T au vu d'une réalisation $y = (x, \delta, z_0, \dots, z_{a_x})$ de $Y = (X, \Delta, Z_0, \dots, Z_{a_x})$ en calculant la densité $f_Y(y; \theta)$ de y . Nous obtenons cette densité en intégrant la densité (que nous noterons $f_{Y,Z}(y, z; \theta)$) du vecteur d'observations complètes $(X, \Delta, Z_0, \dots, Z_{a_x}, Z(X))$ sur l'ensemble des valeurs de la variable inobservée $Z(X)$.

Sous les hypothèses de censure indépendante et non-informative formulées ci-dessus, nous pouvons retirer de cette densité les termes relatifs à la distribution de C pour estimer θ par la méthode du maximum de vraisemblance.

Nous notons $L^{(i)}(\theta)$ la contribution à la vraisemblance d'une observation y_i , une fois retirés de $f_Y(y_i; \theta)$ les termes relatifs à la censure. Cette contribution s'écrit :

$$L^{(i)}(\theta) = \int l(y_i, z; \theta) dz, \quad (4.3)$$

où

$$l(y_i, z; \theta) = \lambda_T(x_i)^{\delta_i} \exp \left[\delta_i \beta' w_i(x_i) - \int_0^{x_i} e^{\beta' w_i(u)} d\Lambda_T(u) \right] f(z_{i0}, \dots, z_{ia_{x_i}}, z; \alpha),$$

et $w_i(u) = (z_{ia_{x_i}}, z)'$ si $u \in]t_{a_{x_i}}, x_i]$. Par convention, nous posons $0^0 = 1$.

Avant de chercher à estimer θ , nous devons nous intéresser au problème de l'identifiabilité de ce paramètre. Si on suppose que la loi P_θ dont sont issues les observations est parfaitement connue, peut-on retrouver la valeur de θ ? La réponse à ce problème pour le modèle conjoint proposé est développée dans la section suivante. Nous rappelons la définition de la notion d'identifiabilité et nous démontrons, à partir de la vraisemblance (4.3), que le paramètre θ est identifiable sous les conditions C1 à C6 énoncées ci-dessus.

4.3 Identifiabilité des paramètres dans le modèle conjoint

Dans cette partie, nous étudions la question de l'identifiabilité des paramètres du modèle conjoint proposé. Nous rappelons tout d'abord les définitions de l'identifiabilité et de l'information au sens de Kullback, ainsi que quelques propriétés reliant ces deux notions (ces propriétés nous seront utiles quand nous établirons les propriétés asymptotiques des estimateurs proposés). Puis nous démontrons que le modèle conjoint proposé est identifiable.

4.3.1 Définition de l'identifiabilité et information au sens de Kullback

La définition générale de l'identifiabilité des paramètres d'un modèle est la suivante. Considérons la famille de lois de probabilités paramétrée $\mathcal{P} = \{P_\phi : \phi \in \Phi\}$.

Définition 4.1 *Un modèle est dit identifiable si l'application de Φ dans l'espace des lois de probabilité \mathcal{P} , qui à ϕ associe P_ϕ est injective.*

On dit aussi dans ce cas que le paramètre ϕ est identifiable. La condition d'identifiabilité permet donc de passer sans ambiguïté d'une loi P_{ϕ_1} à la valeur ϕ_1 du paramètre.

Si la famille de lois de probabilités \mathcal{P} peut être définie par l'intermédiaire de la famille \mathcal{F} de densités : $\mathcal{P} = \{P_\phi = (f(y; \phi) \cdot \mu) : f \in \mathcal{F}\}$, pour une mesure μ , la condition d'identifiabilité précédente peut s'exprimer de la manière suivante :

$$\forall \phi_1, \phi_2 \in \Phi, f(y; \phi_1) = f(y; \phi_2) \mu - pp \implies \phi_1 = \phi_2. \quad (4.4)$$

Nous supposons dans la suite de ce paragraphe que les lois considérées admettent des densités strictement positives par rapport à une même mesure dominante μ .

Définition 4.2 Soit deux distributions $P_{\phi_1} = f(y; \phi_1) \cdot \mu$ et $P_{\phi_2} = f(y; \phi_2) \cdot \mu$. On appelle *information au sens de Kullback* de P_{ϕ_2} sur P_{ϕ_1} la quantité

$$K(P_{\phi_1}, P_{\phi_2}) = \int_{\mathcal{Y}} \ln \frac{f(y; \phi_2)}{f(y; \phi_1)} \cdot f(y; \phi_2) \mu(dy).$$

L'information au sens de Kullback n'est pas une distance au sens classique du terme car les conditions de symétrie et l'inégalité triangulaire ne sont pas satisfaites. Elle traduit néanmoins une idée de proximité entre les lois P_{ϕ_1} et P_{ϕ_2} , due aux propriétés énoncées dans la proposition suivante :

Proposition 4.1 Soit deux lois de probabilité P_{ϕ_1} et P_{ϕ_2} de \mathcal{P} . L'information au sens de Kullback vérifie les propriétés suivantes :

1. $K(P_{\phi_1}, P_{\phi_2}) \geq 0$,
2. $K(P_{\phi_1}, P_{\phi_2}) = 0 \iff P_{\phi_1} = P_{\phi_2}$.

Comme l'information au sens de Kullback s'annule si et seulement si les deux lois sont confondues, on peut caractériser la propriété d'identifiabilité d'un modèle à l'aide de la fonction K .

Proposition 4.2 Le paramètre ϕ est identifiable si et seulement si :

$$\forall \phi_1, \phi_2 \in \Phi, K(P_{\phi_1}, P_{\phi_2}) = 0 \implies \phi_1 = \phi_2.$$

On note qu'une condition nécessaire à la convergence d'un estimateur du maximum de vraisemblance est l'identifiabilité du paramètre à estimer.

Supposons en effet que Y_1, \dots, Y_n soit un échantillon indépendant extrait d'une loi P^Y de densité $f_Y(y; \phi_0)$, $\phi_0 \in \Phi$. Désignons par $E_{\phi_0}[g(Y)]$ l'espérance de g sous la loi P^Y . Sous des conditions de régularité rappelées par Van der Vaart et Wellner [88], l'estimateur du maximum de vraisemblance $\hat{\phi}_n$ de ϕ_0 , tel que $\hat{\phi}_n = \operatorname{argmax}_{\phi \in \Phi} n^{-1} \sum_{i=1}^n \ln f(y_i; \phi)$, converge vers la solution, qui doit alors être unique, du problème limite $\max_{\phi \in \Phi} E_{\phi_0}[\ln f(Y; \phi)]$. La vraie valeur ϕ_0 est une solution de ce problème limite. La proposition suivante assure que ϕ_0 est l'unique solution si et seulement si ϕ_0 est identifiable.

Proposition 4.3 *La valeur ϕ_0 est identifiable si et seulement si la fonction $\phi \mapsto E_{\phi_0}[\ln f(Y; \phi)]$ admet un maximum unique $\phi = \phi_0$.*

4.3.2 Identifiabilité des paramètres du modèle conjoint

Dans cette partie, nous démontrons que le paramètre $\theta = (\alpha, \beta, \Lambda_T)$ du modèle conjoint proposé est identifiable. Nous obtenons ce résultat en montrant que la condition (4.4) est vérifiée. Dans cette condition, la famille de densité \mathcal{F} est la famille de fonctions $\{f_Y(\cdot, \theta) : \theta \in \Theta\}$ (où $\Theta = A \times B \times L$), mais sous l'hypothèse de censure non-informative et si la distribution de la censure ne dépend pas de la réalisation du processus Z à l'instant de censure, il est équivalent de démontrer la condition (4.4) à partir de la famille $\{L(\theta) : \theta \in \Theta\}$, où $L(\theta)$ est donnée par (4.3). Nous démontrons le résultat suivant :

Proposition 4.4 *Le paramètre $\theta = (\alpha, \beta, \Lambda_T)$ est identifiable, c'est-à-dire, si $\theta = (\alpha, \beta, \Lambda_T)$ et $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{\Lambda}_T)$ sont deux éléments de Θ et si $L(\theta) = L(\tilde{\theta})$ presque partout, alors $\theta = \tilde{\theta}$.*

Démonstration. La démonstration que nous proposons se décompose en trois lemmes. Nous montrons tout d'abord que pour $\theta = (\alpha, \beta, \Lambda_T)$, $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{\Lambda}_T) \in \Theta$, l'égalité $L(\theta) = L(\tilde{\theta})$ implique que $\alpha = \tilde{\alpha}$. Nous montrons ensuite que si $\theta = (\alpha, \beta, \Lambda_T)$, $\tilde{\theta} = (\alpha, \tilde{\beta}, \tilde{\Lambda}_T)$ et $L(\theta) = L(\tilde{\theta})$, alors $\beta = \tilde{\beta}$ et $\Lambda_T(t_{a_{t^*}+1}) = \tilde{\Lambda}_T(t_{a_{t^*}+1})$. Nous montrons finalement, en tenant compte de ces résultats, que $\Lambda_T(x) = \tilde{\Lambda}_T(x)$ pour $x \in [0, \tau]$.

Lemme 4.1 *Soit $\theta, \tilde{\theta} \in \Theta$. Si $L(\theta) = L(\tilde{\theta})$ presque partout, alors $\alpha = \tilde{\alpha}$.*

Démonstration. Désignons par N l'ensemble

$$\{t : t \in]t_{a_t^*}, t^*[/ (z_0, \dots, z_{a_t}, t) \text{ ne vérifie pas } L(\theta) = L(\tilde{\theta})\}.$$

Pour presque tous les vecteurs (z_0, \dots, z_{a_t}) (c'est-à-dire hors d'un ensemble Ω_Z négligeable), N est un ensemble négligeable.

Soit (z_0, \dots, z_{a_t}) un vecteur tel que $(z_0, \dots, z_{a_t}) \notin \Omega_Z$. Alors (z_0, \dots, z_{a_t}, t) vérifie $L(\theta) = L(\tilde{\theta})$ pour presque tous les $t \in]t_{a_t^*}, t^*[$. Soit $\nu \in]t_{a_t^*}, t^*[$ une valeur de t tel que $(z_0, \dots, z_{a_\nu}, \nu)$ vérifie $L(\theta) = L(\tilde{\theta})$ (il en existe forcément puisque $t^* > t_{a_t^*}$). Considérons alors $L(\theta) = L(\tilde{\theta})$ pour le vecteur $(z_0, \dots, z_{a_\nu}, \nu)$ et $\delta = 0$.

Il vient $f(z_0, \dots, z_{a_\nu}; \alpha) = f(z_0, \dots, z_{a_\nu}; \tilde{\alpha})$ presque partout. D'après l'hypothèse (C3), $\alpha = \tilde{\alpha}$.

■

Le principe de la suite de la démonstration du résultat d'identifiabilité est le suivant : nous montrons que l'égalité $L(\theta) = L(\tilde{\theta})$ peut se réécrire sous la forme d'une égalité

$$g_1(z_0, \dots, z_{a_x-1}, z_{a_x}, x, \delta; \theta, \tilde{\theta}) = g_2(z_{a_x}, x, \delta; \theta, \tilde{\theta}).$$

A (z_{a_x}, x, δ) fixé, g_1 doit être constante pour presque tous les (z_0, \dots, z_{a_x-1}) . Nous choisissons deux valeurs appropriées $(z_0^{(i)}, \dots, z_{a_x-1}^{(i)})$ ($i = 1, 2$) de (z_0, \dots, z_{a_x-1}) , dont nous justifions l'existence, et à partir de l'égalité

$$g_1(z_0^{(1)}, \dots, z_{a_x-1}^{(1)}, z_{a_x}, x, \delta; \theta, \tilde{\theta}) = g_1(z_0^{(2)}, \dots, z_{a_x-1}^{(2)}, z_{a_x}, x, \delta; \theta, \tilde{\theta}),$$

nous obtenons une condition sur θ et $\tilde{\theta}$ qui ne peut être vérifiée que si $\beta = \tilde{\beta}$ (lemme 4.2) et si $\Lambda_T = \tilde{\Lambda}_T$ (lemme 4.3).

Lemme 4.2 Soit $\theta = (\alpha, \beta, \Lambda_T)$ et $\tilde{\theta} = (\alpha, \tilde{\beta}, \tilde{\Lambda}_T)$. Si $L(\theta) = L(\tilde{\theta})$ presque partout, alors $\beta = \tilde{\beta}$ et $\Lambda_T(t_{a_t^*+1}) = \tilde{\Lambda}_T(t_{a_t^*+1})$.

Démonstration. Si $\theta = (\alpha, \beta, \Lambda_T)$ et $\tilde{\theta} = (\alpha, \tilde{\beta}, \tilde{\Lambda}_T)$, l'égalité $L(\theta) = L(\tilde{\theta})$ peut s'écrire sous la forme :

$$\begin{aligned} & \delta \ln \left(\frac{\lambda_T}{\tilde{\lambda}_T}(x) \right) + \ln \int \exp \left[\delta \beta' w(x) - \int_{t_{a_x}}^x e^{\beta' w(u)} d\Lambda_T(u) \right] f(z|z_{a_x}; \alpha) dz \\ & - \ln \int \exp \left[\delta \tilde{\beta}' w(x) - \int_{t_{a_x}}^x e^{\tilde{\beta}' w(u)} d\tilde{\Lambda}_T(u) \right] f(z|z_{a_x}; \alpha) dz = \\ & \int_0^{t_{a_x}} \left[e^{\beta' w(u)} d\Lambda_T(u) - e^{\tilde{\beta}' w(u)} d\tilde{\Lambda}_T(u) \right] \text{ p.p..} \end{aligned} \quad (4.5)$$

Le membre droit de (4.5) dépend d'une réalisation du processus Z au travers de z_0, \dots, z_{a_x} . Le membre gauche de (4.5) ne dépend de Z qu'au travers de z_{a_x} presque partout. Donc pour une réalisation (x, δ, z_{a_x}) fixée, le membre droit de (4.5) doit être constant pour presque tous les (z_0, \dots, z_{a_x-1}) . Notons \mathcal{P} cette propriété.

\mathcal{P} est vérifiée pour presque tous les (x, δ, z_{a_x}) . Soit (x, δ, z_{a_x}) une valeur pour laquelle \mathcal{P} est vérifiée, et telle que $x > t_{a_t^*} + 2$ (nous n'introduisons pas de notation supplémentaire pour ce vecteur, pour ne pas surcharger la démonstration).

Pour une valeur fixée $z_{a_t^*}$, notons par $\Omega_{(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*}}$ l'ensemble des vecteurs $(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*}$ tels que (z_0, \dots, z_{a_x-1}) ne vérifie pas \mathcal{P} . Notons que l'ensemble $\Omega_{(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*}}$ est bien défini. En effet, si $x > t_{a_t^*} + 2$, alors $a_x - 1 > a_t^*$. L'ensemble $\Omega_{(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*}}$ est négligeable pour presque tous les $z_{a_t^*}$ (c'est-à-dire pour $z_{a_t^*} \notin \Omega_{Z_{a_t^*}}$ où $\Omega_{Z_{a_t^*}}$ est un ensemble négligeable). Soit z_a et z_b deux valeurs de $z_{a_t^*}$ telles que $z_a \neq z_b$ et $z_a, z_b \notin \Omega_{Z_{a_t^*}}$. Alors $(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*} \oplus z_a$ et $(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*} \oplus z_b$ vérifient \mathcal{P} pour presque tous les $(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*}$. Ceci peut s'exprimer, en utilisant (4.5), par l'égalité suivante :

$$\begin{aligned} & \int_0^{t_{a_t^*}+1} \left[e^{\beta' w_a} d\Lambda_T(u) - e^{\tilde{\beta}' w_a} d\tilde{\Lambda}_T(u) \right] + \int_{t_{a_t^*}+1}^{t_{a_x}} \left[e^{\beta' w(u)} d\Lambda_T(u) - e^{\tilde{\beta}' w(u)} d\tilde{\Lambda}_T(u) \right] \\ &= \int_0^{t_{a_t^*}+1} \left[e^{\beta' w_b} d\Lambda_T(u) - e^{\tilde{\beta}' w_b} d\tilde{\Lambda}_T(u) \right] + \int_{t_{a_t^*}+1}^{t_{a_x}} \left[e^{\beta' w(u)} d\Lambda_T(u) - e^{\tilde{\beta}' w(u)} d\tilde{\Lambda}_T(u) \right], \end{aligned}$$

où $w_a = (z_a, z_{a_t^*+1})'$ et $w_b = (z_b, z_{a_t^*+1})'$. Les seconds termes des membres gauche et droit de cette égalité sont égaux. Cette expression se simplifie donc sous la forme suivante :

$$e^{\beta' w_a} \Lambda_T(t_{a_t^*+1}) - e^{\tilde{\beta}' w_a} \tilde{\Lambda}_T(t_{a_t^*+1}) = e^{\beta' w_b} \Lambda_T(t_{a_t^*+1}) - e^{\tilde{\beta}' w_b} \tilde{\Lambda}_T(t_{a_t^*+1}).$$

Puis, il vient

$$\frac{\Lambda_T(t_{a_t^*+1})}{\tilde{\Lambda}_T(t_{a_t^*+1})} = \frac{e^{\tilde{\beta}' w_a} - e^{\tilde{\beta}' w_b}}{e^{\beta' w_a} - e^{\beta' w_b}}. \quad (4.6)$$

Notons que le rapport $\Lambda_T(t_{a_t^*+1})/\tilde{\Lambda}_T(t_{a_t^*+1})$ est constant. Notons γ ($\gamma > 0$) ce rapport.

Pour $z_a, z_b \notin \Omega_{Z_{a_t^*}}$, $[e^{\tilde{\beta}' w_a} - e^{\tilde{\beta}' w_b}]/[e^{\beta' w_a} - e^{\beta' w_b}]$ est égal à γ pour presque tous les $(z_0, \dots, z_{a_x-1}) \setminus z_{a_t^*}$. La fonction f de z_a et z_b définie par $f(z_a, z_b) = e^{\tilde{\beta}' w_a} - e^{\tilde{\beta}' w_b} -$

$\gamma(e^{\beta'w_a} - e^{\beta'w_b})$ est égale à 0 pour presque tous les $(z_0, \dots, z_{a_x-1})_{\setminus z_{a_t^*}}$. Dérivons f par rapport à z_a , pour obtenir $\gamma = \frac{\tilde{\beta}_1}{\beta_1} e^{[\tilde{\beta}-\beta]'w_a}$, pour $z_a, z_b \notin \Omega_{z_{a_t^*}}$. Alors il vient : $\beta = \tilde{\beta}$.

En utilisant ce résultat dans l'égalité (4.6), nous pouvons conclure que $\Lambda_T(t_{a_t^*+1}) = \tilde{\Lambda}_T(t_{a_t^*+1})$.

■

Lemme 4.3 Soit $\theta = (\alpha, \beta, \Lambda_T)$ et $\tilde{\theta} = (\alpha, \beta, \tilde{\Lambda}_T)$. Si $L(\theta) = L(\tilde{\theta})$ presque partout, alors $\Lambda_T(x) = \tilde{\Lambda}_T(x)$, $x \in [0, \tau]$.

Démonstration. La démonstration que nous proposons se décompose en deux étapes. Nous montrons tout d'abord que pour tout $x \geq t^*$, $\Lambda_T(t_{a_x}) = \tilde{\Lambda}_T(t_{a_x})$.

Soit $x \in [t^*, t_{a_t^*+1}]$. Alors $a_x = a_{t^*}$. Puisque $\Lambda_T(t_{a_t^*}) = \tilde{\Lambda}_T(t_{a_t^*}) = 0$, il vient $\Lambda_T(t_{a_x}) = \tilde{\Lambda}_T(t_{a_x})$.

Soit $x \in]t_{a_t^*+1}, t_{a_t^*+2}]$. Alors $a_x = a_{t^*} + 1$. D'après le lemme 3.2, $\Lambda_T(t_{a_t^*+1}) = \tilde{\Lambda}_T(t_{a_t^*+1})$. Il vient $\Lambda_T(t_{a_x}) = \tilde{\Lambda}_T(t_{a_x})$.

Soit finalement $x > t_{a_t^*+2}$. La propriété \mathcal{P} nous permet de dire que pour (x, δ, z_{a_x}) fixé hors d'un ensemble négligeable, le membre droit de (4.5) doit être constant pour presque tous les (z_0, \dots, z_{a_x-1}) . Soit (x, δ, z_{a_x}) une valeur pour laquelle \mathcal{P} est vérifiée.

Soit z_j ($j = a_{t^*} + 1, \dots, a_x - 1$) une valeur fixée de Z_j . Notons que l'ensemble des indices $\{a_{t^*} + 1, \dots, a_x - 1\}$ est bien défini. En effet, si $x > t_{a_t^*+2}$, alors $a_x - 1 \geq a_{t^*} + 1$. Nous notons par $\Omega_{(z_0, \dots, z_{a_x-1}) \setminus z_j}$ ($j = a_{t^*} + 1, \dots, a_x - 1$) l'ensemble des vecteurs $(z_0, \dots, z_{a_x-1}) \setminus z_j$ tels que (z_0, \dots, z_{a_x-1}) ne vérifie pas \mathcal{P} . Pour presque tous les z_j (c'est-à-dire pour $z_j \notin \Omega_{Z_j}$, où Ω_{Z_j} est un ensemble négligeable), $\Omega_{(z_0, \dots, z_{a_x-1}) \setminus z_j}$ est négligeable.

Soit z_a et z_b deux valeurs de Z_j telles que $z_a \neq z_b$ et $z_a, z_b \notin \Omega_{Z_j}$. Alors \mathcal{P} est vérifiée par $(z_0, \dots, z_{a_x-1})_{\setminus z_j} \oplus z_a$ et $(z_0, \dots, z_{a_x-1})_{\setminus z_j} \oplus z_b$ pour presque tous les $(z_0, \dots, z_{a_x-1})_{\setminus z_j}$.

Si nous réécrivons le membre droit de (4.5) avec $\beta = \tilde{\beta}$ de la manière suivante :

$$\begin{aligned} & \int_0^{t_{a_x}} \left[e^{\beta'w(u)} d(\Lambda_T(u) - \tilde{\Lambda}_T(u)) \right] \\ &= \sum_{\kappa=0}^{a_x-1} e^{\beta_1 z_\kappa + \beta_2 z_{\kappa+1}} [(\Lambda_T - \tilde{\Lambda}_T)(t_{\kappa+1}) - (\Lambda_T - \tilde{\Lambda}_T)(t_\kappa)], \end{aligned}$$

le précédent argument peut s'exprimer de la manière suivante :

$$\begin{aligned} & (e^{\beta_1 z_{j-1} + \beta_2 z_a} - e^{\beta_1 z_{j-1} + \beta_2 z_b})[(\Lambda_T - \tilde{\Lambda}_T)(t_j) - (\Lambda_T - \tilde{\Lambda}_T)(t_{j-1})] \\ & = (e^{\beta_1 z_b + \beta_2 z_{j+1}} - e^{\beta_1 z_a + \beta_2 z_{j+1}})[(\Lambda_T - \tilde{\Lambda}_T)(t_{j+1}) - (\Lambda_T - \tilde{\Lambda}_T)(t_j)], \end{aligned}$$

pour $j = a_{t^*} + 1, \dots, a_x - 1$.

Nous remarquons que si $\Lambda_T(t_{j-1}) = \tilde{\Lambda}_T(t_{j-1})$ et $\Lambda_T(t_j) = \tilde{\Lambda}_T(t_j)$, alors cette expression se réduit à $(e^{\beta_1 z_b + \beta_2 z_{j+1}} - e^{\beta_1 z_a + \beta_2 z_{j+1}})[(\Lambda_T - \tilde{\Lambda}_T)(t_{j+1})] = 0$, d'où il vient $\Lambda_T(t_{j+1}) = \tilde{\Lambda}_T(t_{j+1})$.

Or, ceci est vrai pour $j = a_{t^*} + 1$, car $\Lambda_T(t_{a_{t^*}}) = \tilde{\Lambda}_T(t_{a_{t^*}})$ et d'après le lemme 4.2, $\Lambda_T(t_{a_{t^*}+1}) = \tilde{\Lambda}_T(t_{a_{t^*}+1})$. Il vient $\Lambda_T(t_{a_{t^*}+2}) = \tilde{\Lambda}_T(t_{a_{t^*}+2})$.

En répétant ce raisonnement de manière récursive, pour $j = a_{t^*} + 2, \dots, a_x - 1$, il vient $\Lambda_T(t_{a_x}) = \tilde{\Lambda}_T(t_{a_x})$.

La deuxième étape de la démonstration de ce lemme consiste à montrer que $\Lambda_T(x) = \tilde{\Lambda}_T(x)$ pour $x \in [0, \tau]$. Supposons directement que $x \geq t^*$. En utilisant le résultat précédent et $\beta = \beta'$, nous pouvons réécrire (4.5) avec $\delta = 0$ sous la forme :

$$\ln \frac{\int \exp \left[- \int_{t_{a_x}}^x e^{\beta' w(u)} d\Lambda_T(u) \right] f(z|z_{a_x}; \alpha) dz}{\int \exp \left[- \int_{t_{a_x}}^x e^{\beta' w(u)} d\tilde{\Lambda}_T(u) \right] f(z|z_{a_x}; \alpha) dz} = 0.$$

D'où l'intégrale I définie par

$$I = \int \left\{ \exp \left[- \int_{t_{a_x}}^x e^{\beta' w(u)} d\Lambda_T(u) \right] - \exp \left[- \int_{t_{a_x}}^x e^{\beta' w(u)} d\tilde{\Lambda}_T(u) \right] \right\} f(z|z_{a_x}; \alpha) dz$$

est nulle. Réécrivons I de la manière suivante :

$$\int \left\{ \left(\exp \left[-e^{\beta' w(x)} \right] \right)^{\Lambda_T(x) - \Lambda_T(t_{a_x})} - \left(\exp \left[-e^{\beta' w(x)} \right] \right)^{\tilde{\Lambda}_T(x) - \Lambda_T(t_{a_x})} \right\} f(z|z_{a_x}; \alpha) dz.$$

Si $\Lambda_T(x) > \tilde{\Lambda}_T(x)$, alors $I < 0$, qui est une contradiction. Si $\Lambda_T(x) < \tilde{\Lambda}_T(x)$, on aboutit à la contradiction $I > 0$. Il vient $\Lambda_T(x) = \tilde{\Lambda}_T(x)$ pour $x \in [0, \tau]$.

■

Chapitre 5

Estimation semiparamétrique dans le modèle de Cox avec valeurs manquantes d'une covariable dépendant du temps

Lorsque la condition d'identifiabilité d'un modèle est satisfaite, nous pouvons espérer développer une méthode d'estimation convergente de ses paramètres. Notre problème consiste à estimer l'inconnue paramétrique β du modèle de Cox

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta'W(t)}$$

lorsque la composante $Z(t)$ du vecteur $W(t)$ n'est pas observée. Nous avons proposé de modéliser conjointement les durées de vie censurées et le processus de covariable, et nous avons obtenu la vraisemblance conjointe (4.3) pour les paramètres β , λ_T et le paramètre α de la loi du processus. Dans ce modèle, les paramètres λ_T et α peuvent être considérés comme des paramètres de nuisance. La méthode du maximum de vraisemblance n'est pas adaptée pour estimer ces paramètres à partir de la vraisemblance (4.3) car il n'existe pas de maximum de cette vraisemblance lorsque $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$ appartient à l'ensemble L .

Dans ce chapitre, nous proposons d'estimer le paramètre de régression β d'un modèle de Cox avec covariable dépendant du temps manquante, en adaptant la méthode semiparamétrique à l'estimation des paramètres α, β, Λ_T dans le modèle conjoint proposé.

Nous démontrons tout d'abord l'existence d'estimateurs semiparamétriques des pa-

ramètres de ce modèle. Puis, en utilisant le principe de l'algorithme Espérance-Maximisation (EM), nous obtenons une caractérisation de ces estimateurs qui sera cruciale pour démontrer leurs propriétés asymptotiques.

5.1 Existence des estimateurs semiparamétriques

Pour chaque sujet i ($i = 1, \dots, n$), nous observons les données incomplètes $Y_i = (X_i, \Delta_i, Z_{i0}, \dots, Z_{ia_{X_i}})$.

Supposons que l'on ait observé $p(n)$ événements d'intérêt ($p(n) \leq n$) aux instants réordonnés par ordre croissant $x_{(1)} < \dots < x_{(p(n))}$ (nous supposons dans un premier temps qu'il n'y a pas d'*ex aequo* parmi ces instants). Les instants $x_{(p(n)+1)} \leq \dots \leq x_{(n)}$ sont les instants de censure. La méthode semiparamétrique consiste à modifier le paramètre Λ_T : nous considérons que Λ_T appartient à l'ensemble des fonctions en escalier $\Lambda_{T,n}$ définies sur $[0, \infty[$, positives, croissantes, et dont les sauts se produisent aux instants d'événement $x_{(k)}$ ($k = 1, \dots, p(n)$).

Par la suite, les valeurs $\Lambda_{T,n}(x_{(k)})$ de ces fonctions aux instants $x_{(k)}$ seront notées $\Lambda_{n,k}$. Les sauts $\Delta\Lambda_{T,n}(x_{(k)}) = \Lambda_{n,k} - \Lambda_{n,k-1}$ ($k = 2, \dots, p(n)$) des fonctions $\Lambda_{T,n}$ aux instants $x_{(k)}$ seront notés $\Delta\Lambda_{n,k}$ (avec $\Delta\Lambda_{n,1} = \Lambda_{n,1}$).

Notons Θ_n l'espace modifié des paramètres, défini par :

$$\Theta_n = \{\theta = (\alpha, \beta, \Delta\Lambda_{n,k}) : \alpha \in A, \beta \in B, \Delta\Lambda_{n,k} \in \mathbb{R}, k = 1, \dots, p(n)\}.$$

La vraisemblance du paramètre modifié $(\alpha, \beta, \Lambda_{T,n})$ au vu d'une observation incomplète y_i est obtenue à partir de la vraisemblance conjointe (4.3). Elle s'écrit

$$L^{(i)}(\theta) = \int \Delta\Lambda_{T,n}(x_i) \exp \left[\beta' w_i(x_i) - \sum_{k=1}^{p(n)} \Delta\Lambda_{n,k} e^{\beta' w_i(x_{(k)})} 1_{\{x_{(k)} \leq x_i\}} \right] \\ \times f(z_{i0}, \dots, z_{ia_{x_i}}, z; \alpha) dz$$

si l'observation x_i correspond à un événement d'intérêt et

$$L^{(i)}(\theta) = \int \exp \left[- \sum_{k=1}^{p(n)} \Delta\Lambda_{n,k} e^{\beta' w_i(x_{(k)})} 1_{\{x_{(k)} \leq x_i\}} \right] f(z_{i0}, \dots, z_{ia_{x_i}}, z; \alpha) dz$$

si x_i correspond à une censure.

La vraisemblance de θ au vu des observations incomplètes y_i issues de n répliques indépendantes de $(X, \Delta, (Z(t))_{t \geq 0})$ est obtenue par le produit

$$L_n(\theta) = \prod_{i=1}^n L^{(i)}(\theta),$$

soit

$$L_n(\theta) = \prod_{i=1}^{p(n)} \Delta \Lambda_{n,i} \prod_{i=1}^n \int \exp \left[\delta_i \beta' w_i(x_i) - \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} e^{\beta' w_i(x_{(k)})} 1_{\{x_{(k)} \leq x_i\}} \right] \times f(z_{i0}, \dots, z_{ia_{x_i}}, z; \alpha) dz.$$

Nous montrons le résultat suivant :

Proposition 5.1 *La fonction $\theta \mapsto L_n(\theta)$ est bornée sur Θ_n et atteint son maximum en $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \Delta \hat{\Lambda}_{n,1}, \dots, \Delta \hat{\Lambda}_{n,p(n)})$.*

Démonstration. Supposons d'abord que $\Delta \Lambda_{n,i} \leq U < \infty$ ($i = 1, \dots, p(n)$) pour un réel U positif fini. La vraisemblance L_n est une fonction continue en α, β et $\Delta \Lambda_{n,i}$ ($i = 1, \dots, p(n)$) sur l'espace compact $A \times B \times [0, U]^{p(n)}$. La fonction L_n est donc bornée sur cet espace et y atteint son maximum.

Pour montrer qu'il existe un maximum de L_n sur l'espace $A \times B \times [0, \infty]^{p(n)}$, nous montrons qu'il existe un réel positif U fini tel que pour tout élément $\theta_U = (\alpha_U, \beta_U, \Delta \Lambda_{n,1,U}, \dots, \Delta \Lambda_{n,p(n),U})$ de $\{A \times B \times [0, \infty]^{p(n)}\} \setminus \{A \times B \times [0, U]^{p(n)}\}$, il existe un élément $\theta = (\alpha, \beta, \Delta \Lambda_{n,1}, \dots, \Delta \Lambda_{n,p(n)})$ de $A \times B \times [0, U]^{p(n)}$ tel que $L_n(\theta) > L_n(\theta_U)$.

Pour démontrer cette affirmation, nous supposons que sa contre-apposée est vraie, puis nous montrons que celle-ci engendre une contradiction.

Nous supposons que pour tout réel positif fini U , il existe un élément $\theta_U = (\alpha_U, \beta_U, \Delta \Lambda_{n,1,U}, \dots, \Delta \Lambda_{n,p(n),U})$ de $\{A \times B \times [0, \infty]^{p(n)}\} \setminus \{A \times B \times [0, U]^{p(n)}\}$ tel que pour tout $\theta = (\alpha, \beta, \Delta \Lambda_{n,1}, \dots, \Delta \Lambda_{n,p(n)})$ de $A \times B \times [0, U]^{p(n)}$, on ait $L_n(\theta) \leq L_n(\theta_U)$. Notons \mathcal{A} cette assertion.

Notons $m = \min_{B,i,k} e^{\beta' w_i(x_k)}$ et $M = \max_{B,i,k} e^{\beta' w_i(x_k)}$. Nous pouvons minorer les termes

$$\sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} e^{\beta' w_i(x_{(k)})} 1_{\{x_{(k)} \leq x_i\}}$$

par $m \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} 1_{\{x_{(k)} \leq x_i\}}$ ($i = 1, \dots, n$).

Nous pouvons donc majorer la fonction $L_n(\theta)$ par

$$\prod_{i=1}^n \left[(M \Delta \Lambda_{T,n}(x_i))^{\delta_i} \exp(-m \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} 1_{\{x_{(k)} \leq x_i\}}) f(z_{i0}, \dots, z_{ia_{x_i}}; \alpha) \right]. \quad (5.1)$$

Si $\theta_U = (\alpha_U, \beta_U, \Delta \Lambda_{n,1,U}, \dots, \Delta \Lambda_{n,p(n),U})$ appartient à l'espace $\{A \times B \times [0, \infty^{p(n)}] \setminus \{A \times B \times [0, U]^{p(n)}\}$, alors il existe au moins un sujet j ($j \in \{1, \dots, p(n)\}$) pour lequel $\Delta \Lambda_{n,j,U} = \Delta \Lambda_{T,n,U}(x_{(j)}) > U$.

Donc il existe au moins un sujet i_U , $i_U \in \{1, \dots, n\}$ tel que

$$\sum_{k=1}^{p(n)} \Delta \Lambda_{n,k,U} 1_{\{x_{(k)} \leq x_{i_U}\}} > U.$$

Il s'ensuit que $\sum_{i=1}^n \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k,U} 1_{\{x_{(k)} \leq x_i\}}$ tend vers l'infini lorsque U tend vers l'infini. Donc d'après (5.1), la borne supérieure de $L_n(\theta_U)$, et donc $L_n(\theta_U)$ peuvent être rendus arbitrairement petits en faisant tendre U vers l'infini. Ceci contredit l'assertion \mathcal{A} .

Ainsi, si la fonction L_n atteint son maximum, c'est sur l'ensemble $A \times B \times [0, U]^{p(n)}$ pour U réel positif fini. Nous avons montré au début de cette démonstration qu'un maximum de $L_n(\theta)$ existe sur un tel ensemble. Par la suite, nous noterons $\hat{\theta}_n$ l'estimateur semiparamétrique de θ .

■

Remarque : La fonction $\theta \mapsto \prod_{i=1}^n f_Y(y_i; \theta)$ atteint aussi son maximum en $\hat{\theta}_n$.

Il n'existe pas d'expression explicite pour l'estimateur du maximum de vraisemblance partielle du paramètre β du modèle de Cox. De même, il n'existe pas d'expression explicite pour l'estimateur semiparamétrique de β dans le modèle conjoint proposé.

Nous obtenons en revanche une caractérisation de l'estimateur semiparamétrique du paramètre $\theta = (\alpha, \beta, \Delta \Lambda_{n,1}, \dots, \Delta \Lambda_{n,p(n)})$ en utilisant le principe de l'algorithme EM. Cette caractérisation sera centrale dans la démonstration de la consistance et de la normalité asymptotique des estimateurs.

5.2 Une caractérisation des estimateurs semiparamétriques

L'algorithme Espérance-Maximisation (EM) est un algorithme de recherche d'estimation par le maximum de vraisemblance, applicable dans des situations de données manquantes. Il a été décrit par Dempster, Laird et Rubin [26]. Le lecteur intéressé par de plus amples développements pourra consulter l'ouvrage très complet de McLachlan et Krishnan [62].

Si les variables $Z_1(X_1), \dots, Z_n(X_n)$ étaient observables, l'estimation du maximum de vraisemblance de $\theta = (\alpha, \beta, \Lambda_{T,n})$ serait déterminée en maximisant la log-vraisemblance $\sum_{i=1}^n \ln f_{Y,Z}(y_i, z_i(x_i); \theta)$. Comme ces variables ne sont pas disponibles, on peut penser remplacer cette fonction par sa meilleure approximation calculable à partir des variables observables Y_1, \dots, Y_n . Ceci revient à maximiser l'espérance conditionnelle de la log-vraisemblance $\sum_{i=1}^n \ln f_{Y,Z}(y_i, z_i(x_i); \theta)$ sachant les observations $y_i, i = 1, \dots, n$.

Dans la suite de ce travail, si X_1 et X_2 sont deux variables aléatoires, nous noterons $f_{X_1|X_2}(x_1; \theta)$ la densité conditionnelle de X_1 sachant $X_2 = x_2$, paramétrée par θ . L'espérance conditionnelle $E_\theta[g(X_1)|X_2 = x_2]$ de $g(X_1)$ sachant $X_2 = x_2$ sera notée $E_\theta[g(X_1)|x_2]$.

Dans ce chapitre, nous utilisons le principe de l'algorithme EM pour obtenir une caractérisation de $\hat{\theta}_n$ à partir de la densité $f_{Y,Z}(y, z; \theta)$. L'implémentation de l'algorithme EM et les formules nécessaires au calcul des estimations semiparamétriques des paramètres du modèle conjoint seront détaillées dans le chapitre 7 de ce manuscrit.

Le résultat intermédiaire suivant, qui est une adaptation aux notations de notre problème de la démarche de l'EM, est d'abord obtenu :

Proposition 5.2 *L'estimateur semiparamétrique $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{T,n})$ vérifie la condition*

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} [E_{\hat{\theta}_n} [\ln f_{Y,Z}(Y, Z; \theta) | y_i]] \Big|_{\theta=\hat{\theta}_n} = 0.$$

Démonstration. Soit $\theta = (\alpha, \beta, \Lambda_{T,n})$. La densité conditionnelle de Z sachant

$Y = y$ peut s'écrire sous la forme suivante :

$$f_{Z|Y}(z; \theta) = \frac{f_{Y,Z}(y, z; \theta)}{f_Y(y; \theta)}.$$

Il vient

$$\ln f_Y(y; \theta) = \ln f_{Y,Z}(y, z; \theta) - \ln f_{Z|Y}(z; \theta).$$

Soit $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{\Lambda}_{T,n})$ une valeur quelconque de θ . Si nous multiplions les deux membres de l'égalité précédente par la densité conditionnelle $f_{Z|Y}(z; \tilde{\theta})$ et que nous intégrons sur z , nous obtenons le résultat suivant :

$$E_{\tilde{\theta}}[\ln f_Y(Y; \theta)|y] = E_{\tilde{\theta}}[\ln f_{Y,Z}(Y, Z; \theta) - \ln f_{Z|Y}(Z; \theta)|y],$$

qui peut être réécrit comme

$$\begin{aligned} E_{\tilde{\theta}}[\ln f_Y(Y; \theta)|y] &= \ln f_Y(y; \theta) \\ &= E_{\tilde{\theta}}[\ln f_{Y,Z}(Y, Z; \theta)|y] - E_{\tilde{\theta}}[\ln f_{Z|Y}(Z; \theta)|y]. \end{aligned} \quad (5.2)$$

Nous voulons montrer que

$$\left. \frac{\partial \ln f_Y(y; \theta)}{\partial \theta} \right|_{\theta=\tilde{\theta}} = \left. \frac{\partial}{\partial \theta} [E_{\tilde{\theta}}[\ln f_{Y,Z}(Y, Z; \theta)|y]] \right|_{\theta=\tilde{\theta}}.$$

Nous allons donc montrer que

$$\left. \frac{\partial}{\partial \theta} [E_{\tilde{\theta}}[\ln f_{Z|Y}(Z; \theta)|y]] \right|_{\theta=\tilde{\theta}} = 0.$$

Notons pour cela que

$$\begin{aligned} E_{\tilde{\theta}}[\ln f_{Z|Y}(Z; \theta)|y] - E_{\tilde{\theta}}[\ln f_{Z|Y}(Z; \tilde{\theta})|y] &= E_{\tilde{\theta}} \left[\ln \frac{f_{Z|Y}(Z; \theta)}{f_{Z|Y}(Z; \tilde{\theta})} | y \right] \\ &= -E_{\tilde{\theta}} \left[\ln \frac{f_{Z|Y}(Z; \tilde{\theta})}{f_{Z|Y}(Z; \theta)} | y \right] \\ &\leq 0, \end{aligned}$$

d'après la proposition 4.1 (remarquons en effet que $E_{\tilde{\theta}} \left[\ln \frac{f_{Z|Y}(Z; \tilde{\theta})}{f_{Z|Y}(Z; \theta)} | y \right]$ est l'information de Kullback de la loi de densité $f_{Z|Y}(Z; \tilde{\theta})$ sur la loi de densité $f_{Z|Y}(Z; \theta)$). Ainsi, nous avons montré que

$$E_{\tilde{\theta}}[\ln f_{Z|Y}(Z; \theta)|y] \leq E_{\tilde{\theta}}[\ln f_{Z|Y}(Z; \tilde{\theta})|y],$$

d'où $\tilde{\theta}$ réalise le maximum de $\theta \mapsto E_{\tilde{\theta}} [\ln f_{Z|Y}(Z; \theta) | y]$ et on a donc

$$\left. \frac{\partial}{\partial \theta} [E_{\tilde{\theta}} [\ln f_{Z|Y}(Z; \theta) | y]] \right|_{\theta=\tilde{\theta}} = 0.$$

Si nous dérivons (5.2) par rapport à θ , ce dernier résultat nous permet d'écrire

$$\left. \frac{\partial \ln f_Y(y; \theta)}{\partial \theta} \right|_{\theta=\tilde{\theta}} = \left. \frac{\partial}{\partial \theta} [E_{\tilde{\theta}} [\ln f_{Y,Z}(Y, Z; \theta) | y]] \right|_{\theta=\tilde{\theta}}.$$

Supposons que $\tilde{\theta} = \hat{\theta}_n$. Puisque

$$\sum_{i=1}^n \left. \frac{\partial \ln f_Y(y_i; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0,$$

nous concluons que

$$\sum_{i=1}^n \left. \frac{\partial}{\partial \theta} [E_{\hat{\theta}_n} [\ln f_{Y,Z}(Y, Z; \theta) | y_i]] \right|_{\theta=\hat{\theta}_n} = 0.$$

■

Ce résultat intermédiaire nous permet de démontrer la caractérisation suivante de l'estimateur semiparamétrique $\hat{\theta}_n$:

Proposition 5.3 *L'estimateur semiparamétrique $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{T,n})$ satisfait l'équation*

$$\hat{\Lambda}_{T,n}(t) = \int_0^t \frac{dH_n(u)}{W_n(u; \hat{\theta}_n)},$$

où $H_n(u) = n^{-1} \sum_{i=1}^n \Delta_i 1_{\{X_i \leq u\}}$ et $W_n(u; \theta) = n^{-1} \sum_{i=1}^n E_{\theta} [e^{\beta' W(u)} 1_{\{u \leq X\}} | y_i]$.

Démonstration. D'après la proposition 5.2, $\hat{\theta}_n$ vérifie

$$\sum_{j=1}^n \left. \frac{\partial}{\partial \theta} [E_{\hat{\theta}_n} [\ln f_{Y,Z}(Y, Z; \theta) | y_j]] \right|_{\theta=\hat{\theta}_n} = 0.$$

Sous les hypothèses de censure indépendante et non-informative, nous pouvons ignorer les termes relatifs à la distribution de la censure pour estimer θ par la méthode

semiparamétrique. Le résultat précédent peut donc être écrit en utilisant la fonction $l(y, z; \theta)$ au lieu de $f_{Y,Z}(y, z; \theta)$:

$$\sum_{j=1}^n \frac{\partial}{\partial \theta} [E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_j]] \Big|_{\theta=\hat{\theta}_n} = 0.$$

En utilisant l'expression de $l(y, z; \theta)$ donnée par (4.3), nous pouvons calculer explicitement les termes $E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}]$, selon que l'observation $y_{(j)}$ corresponde à un événement d'intérêt ($j = 1, \dots, p(n)$) :

$$\begin{aligned} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}] &= E_{\hat{\theta}_n} \left[\beta' W(X) - \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} e^{\beta' W(x_{(k)})} 1_{\{x_{(k)} \leq X\}} | y_{(j)} \right] \\ &\quad + \ln \Delta \Lambda_{n,j} + E_{\hat{\theta}_n} [\ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y_{(j)}], \end{aligned}$$

ou à une censure ($j = p(n) + 1, \dots, n$) :

$$\begin{aligned} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}] &= E_{\hat{\theta}_n} \left[- \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} e^{\beta' W(x_{(k)})} 1_{\{x_{(k)} \leq X\}} + \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y_{(j)} \right]. \end{aligned}$$

Nous dérivons $\sum_{j=1}^{p(n)} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}]$ et $\sum_{j=p(n)+1}^n E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}]$ par rapport à l'inconnue $\Delta \Lambda_{n,i}$ ($i = 1, \dots, p(n)$) et nous obtenons respectivement :

$$\sum_{j=1}^{p(n)} \frac{\partial}{\partial \Delta \Lambda_{n,i}} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}] = \frac{1}{\Delta \Lambda_{n,i}} - \sum_{j=1}^{p(n)} E_{\hat{\theta}_n} [e^{\beta' W(x_{(i)})} 1_{\{x_{(i)} \leq X\}} | y_{(j)}]$$

et

$$\sum_{j=p(n)+1}^n \frac{\partial}{\partial \Delta \Lambda_{n,i}} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_{(j)}] = - \sum_{j=p(n)+1}^n E_{\hat{\theta}_n} [e^{\beta' W(x_{(i)})} 1_{\{x_{(i)} \leq X\}} | y_{(j)}].$$

Nous obtenons donc le résultat suivant :

$$\sum_{j=1}^n \frac{\partial}{\partial \Delta \Lambda_{n,i}} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_j] = \frac{1}{\Delta \Lambda_{n,i}} - \sum_{j=1}^n E_{\hat{\theta}_n} [e^{\beta' W(x_{(i)})} 1_{\{x_{(i)} \leq X\}} | y_j].$$

D'après la proposition 5.2, l'estimateur semiparamétrique $\Delta \hat{\Lambda}_{n,i}$ de $\Delta \Lambda_{n,i}$ satisfait l'équation

$$\sum_{j=1}^n \frac{\partial}{\partial \Delta \Lambda_{n,i}} E_{\hat{\theta}_n} [\ln l(Y, Z; \hat{\theta}_n) | y_j] = 0,$$

d'où l'on déduit

$$\Delta \hat{\Lambda}_{n,i} = \frac{1}{\sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\hat{\beta}'_n W(x_{(i)})} 1_{\{x_{(i)} \leq X\}} | y_j \right]}.$$

Un estimateur de la fonction $\Lambda_{T,n}$ satisfait l'équation

$$\begin{aligned} \hat{\Lambda}_{T,n}(t) &= \sum_{i=1}^{p(n)} \Delta \hat{\Lambda}_{n,i} 1_{\{x_{(i)} \leq t\}} \\ &= \sum_{i=1}^n \frac{\Delta_i 1_{\{x_i \leq t\}}}{\sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\hat{\beta}'_n W(x_i)} 1_{\{x_i \leq X\}} | y_j \right]}. \end{aligned} \quad (5.3)$$

Cette équation ne permet pas le calcul numérique direct de $\hat{\Lambda}_{T,n}(t)$ car le terme $E_{\hat{\theta}_n} \left[e^{\hat{\beta}'_n W(x_i)} 1_{\{x_i \leq X\}} | y_j \right]$ dépend lui-même de $\hat{\Lambda}_{T,n}$. Ce calcul, détaillé dans le chapitre 7, nécessitera une procédure itérative. En revanche, (5.3) est une caractérisation intéressante, pour la démonstration de sa consistance, de l'estimateur $\hat{\theta}_n$. Nous remarquons que si tous les $Z_i(X_i)$ étaient observés, (5.3) se ramènerait à l'estimateur de Breslow (2.15).

En utilisant les notations des processus ponctuels, nous pouvons réécrire (5.3) sous la forme suivante, qui sera utile par la suite :

$$\hat{\Lambda}_{T,n}(t) = \int_0^t \frac{dH_n(u)}{W_n(u; \hat{\theta}_n)},$$

où $H_n(u) = n^{-1} \sum_{i=1}^n \Delta_i 1_{\{X_i \leq u\}}$ et $W_n(u; \theta) = n^{-1} \sum_{i=1}^n E_{\theta} \left[e^{\beta' W(u)} 1_{\{u \leq X\}} | y_i \right]$.

■

Remarque : Par la suite, l'expression $E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_j]$ introduite dans cette démonstration sera notée $L_{\hat{\theta}_n}^{(j)}(\theta)$. Plus généralement, nous noterons

$$L_{\hat{\theta}}^{(j)}(\theta) = E_{\hat{\theta}} [\ln l(Y, Z; \theta) | y_j]$$

et $L_{n,\hat{\theta}}(\theta) = \sum_{j=1}^n L_{\hat{\theta}}^{(j)}(\theta)$ sera appelée log-vraisemblance EM.

5.3 Prise en compte des ex aequo

Nous avons supposé depuis le début de ce chapitre que les $p(n)$ événements d'intérêt observés étaient distincts. Cependant, dans la pratique de l'analyse des

durées de vie, et pour des raisons liées au recueil des données, nous disposons très souvent d'*ex aequo*. La méthode d'estimation semiparamétrique pour le modèle conjoint proposé peut être modifiée pour en tenir compte.

Supposons que l'on ait observé $p(n)$ instants d'événements dont D seulement soient distincts. Nous notons u_1, \dots, u_D ces instants distincts. La méthode semiparamétrique consistera à chercher un estimateur de la fonction $\Lambda_T(t) = \int_0^t \lambda_T(s) ds$ dans l'ensemble des fonctions en escalier définies sur $[0, \infty[$, positives, croissantes, et dont les sauts $\Delta\Lambda_{T,n}(u_i)$ (notés $\Delta\Lambda_{n,i}$ par la suite) se produisent aux instants u_1, \dots, u_D . L'espace modifié des paramètres est maintenant :

$$\tilde{\Theta}_n = \{\theta = (\alpha, \beta, \Lambda_{T,n}) : \alpha \in A, \beta \in B, \Delta\Lambda_{n,i} \in \mathbb{R}, i = 1, \dots, D\}.$$

La vraisemblance du paramètre θ au vu d'une observation $(x_i, \delta_i, z_{i0}, \dots, z_{ia_{x_i}})$ s'écrit maintenant :

$$\prod_{l=1}^D \Delta\Lambda_{n,l}^{\delta_i 1_{\{x_i=u_l\}}} \int \exp \left[\delta_i \beta' w_i(x_i) - \sum_{l=1}^D \Delta\Lambda_{n,l} e^{\beta' w_i(u_l)} 1_{\{u_l \leq x_i\}} \right] \\ \times f(z_{i0}, \dots, z_{ia_{x_i}}; \alpha).$$

La proposition 5.3 peut être modifiée pour prendre en compte les *ex aequo*. En utilisant la même démarche que dans sa démonstration, nous obtenons une caractérisation de l'estimateur semiparamétrique $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \Delta\hat{\Lambda}_{n,1}, \dots, \Delta\hat{\Lambda}_{n,D})$ de θ .

Nous notons que

$$E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_j] = E_{\hat{\theta}_n} \left[\sum_{l=1}^D \Delta 1_{\{X=u_l\}} \ln \Delta\Lambda_{n,l} + \Delta\beta' W(X) \right. \\ \left. - \sum_{l=1}^D \Delta\Lambda_{n,l} e^{\beta' W(u_l)} 1_{\{u_l \leq X\}} + \ln f(Z_0, \dots, Z_{a_X}; \alpha) \right] | y_j.$$

Nous en déduisons, pour $i = 1, \dots, D$:

$$\sum_{j=1}^n \frac{\partial}{\partial \Delta\Lambda_{n,i}} E_{\hat{\theta}_n} [\ln l(Y, Z; \theta) | y_j] = \sum_{j=1}^n \delta_j 1_{\{x_j=u_i\}} \frac{1}{\Delta\Lambda_{n,i}} - \sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\beta' W(u_i)} 1_{\{u_i \leq X\}} | y_j \right].$$

Enfin, nous obtenons

$$\Delta\hat{\Lambda}_{n,i} = \frac{\sum_{j=1}^n \delta_j 1_{\{x_j=u_i\}}}{\sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\hat{\beta}' W(u_i)} 1_{\{u_i \leq X\}} | y_j \right]}, \quad i = 1, \dots, D.$$

Un estimateur de la fonction $\Lambda_{T,n}$ satisfait l'équation

$$\hat{\Lambda}_{T,n}(t) = \sum_{i=1}^D \frac{1_{\{u_i \leq t\}} \left[\sum_{j=1}^n \delta_j 1_{\{x_j = u_i\}} \right]}{\sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\hat{\beta}'_n W(u_i)} 1_{\{u_i \leq X\}} | y_j \right]}.$$

De nouveau, une démarche itérative sera nécessaire pour calculer numériquement les estimations $\hat{\Lambda}_{T,n}(t)$.

Chapitre 6

Etude des propriétés asymptotiques pour le modèle de Cox avec covariable dépendant du temps manquante

Nous rappelons d'abord quelques définitions et résultats de calcul différentiel et de théorie des processus empiriques. Puis nous démontrons que les estimateurs semiparamétriques obtenus dans le modèle de Cox avec covariable dépendant du temps manquante sont consistants et asymptotiquement normalement distribués.

6.1 Préliminaires

Nous rappelons ci-dessous quelques notions de calcul différentiel (Bickel, Klaassen, Ritov et Wellner [11] ; Christol, Cot et Marle [17]) et de théorie des processus empiriques (Huber et Lecoutre [46] ; Van der Vaart [86] ; Van der Vaart et Wellner [88]).

Rappels sur la différentiabilité

Soit E et F deux espaces vectoriels sur \mathbb{R} . E et F sont normés et munis respectivement des normes $\|\cdot\|_E$ et $\|\cdot\|_F$. Soit $a \in E$.

Une application f de E dans F est différentiable au sens de Gâteaux au point a s'il existe une application linéaire continue $Df(a)$ de E dans F telle que pour tout

$h \in E$,

$$D_h f(a) = \lim_{t \rightarrow 0} \frac{f(a + th) - f(a)}{t}.$$

L'application $Df(a)$ est appelée différentielle de f au sens de Gâteaux au point a . On dit encore que f admet une dérivée directionnelle en a suivant la direction h .

Une application f est différentiable au sens de Fréchet au point a s'il existe une application linéaire continue $\dot{f}(a)$ de E dans F telle que

$$\|f(a + h) - f(a) - \dot{f}(a)(h)\|_F = R(h) \quad (6.1)$$

où le reste $R(h)$ est un $o(\|h\|_E)$ lorsque h tend vers 0 dans E . L'application $\dot{f}(a)$ est appelée différentielle de f au sens de Fréchet au point a .

Bickel, Klaassen, Ritov et Wellner [11] donnent la définition équivalente suivante de la différentiabilité au sens de Fréchet, que nous utiliserons de préférence à (6.1) dans notre démonstration. Si E est un ensemble, nous désignons par $\mathcal{P}_b(E)$ l'ensemble des parties bornées de E .

Définition 6.1 *Une application f est différentiable au sens de Fréchet au point a et sa différentielle est $\dot{f}(a)$ si pour tout $H \in \mathcal{P}_b(E)$,*

$$\lim_{\epsilon \rightarrow 0} \left[\frac{f(a + \epsilon h) - f(a) - \dot{f}(a)(\epsilon h)}{\epsilon} \right] = 0$$

uniformément en $h \in H$.

Rappelons que si une application est différentiable en a , sa différentielle en ce point est unique.

Lorsqu'une application admet, en un point a , une différentielle au sens de Fréchet, elle admet aussi en ce point une différentielle au sens de Gâteaux, égale à sa différentielle au sens de Fréchet. Ceci permet de calculer la différentielle de Fréchet en commençant par la différentielle de Gâteaux, ce qui est parfois plus simple. Pour calculer $\dot{f}(a)$, on calcule d'abord $D_h f(a)$; puis on vérifie que $h \mapsto D_h f(a)$ est linéaire continue; enfin on vérifie que $D_h f(a)$ satisfait à la définition de la différentielle de Fréchet.

Rappels de théorie des processus empiriques

Nous donnons ici quelques éléments de théorie des processus empiriques dont nous aurons besoin. Pour des exposés détaillés de cette théorie, on se référera aux ouvrages de Huber et Lecoutre [46], Van der Vaart [86] et Van der Vaart et Wellner [88].

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes et de même loi P_X sur \mathbb{R} muni des boréliens. On note F_X la fonction de répartition de P_X .

Définition 6.2 *Pour chaque $\omega \in \Omega$ et chaque entier $n \geq 1$, on appelle mesure empirique, et on note $\mathbb{P}_n(\omega)$, la mesure de probabilité sur \mathbb{R} définie par*

$$\mathbb{P}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)},$$

où δ_a désigne la mesure de Dirac au point a .

On écrit généralement $\mathbb{P}_n(A)$ la valeur de la mesure empirique sur le borélien A (au lieu de $\mathbb{P}_n(A)(\omega)$). La mesure empirique correspond à la loi d'une variable aléatoire uniforme sur $\{X_1, \dots, X_n\}$. Elle vérifie $\mathbb{P}_n(A) = n^{-1} \text{card}\{i : X_i \in A\}$. On désigne par \mathbb{F}_n sa fonction de répartition (aléatoire) définie, pour tout $x \in \mathbb{R}$ et $\omega \in \Omega$, par

$$\mathbb{F}_n(x) = \mathbb{F}_n(x)(\omega) = \mathbb{P}_n([-\infty, x])(\omega).$$

Notons que

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}, \quad x \in \mathbb{R}.$$

La fonction \mathbb{F}_n est appelée fonction de répartition empirique, et $\sqrt{n}(\mathbb{F}_n - F_X)$ est le processus empirique réel.

La fonction de répartition empirique est l'estimateur le plus naturel de la fonction de répartition F_X lorsque celle-ci est totalement inconnue. Pour chaque $x \in \mathbb{R}$, $\mathbb{F}_n(x)$ est un estimateur sans biais de $F_X(x)$.

Pour chaque $x \in \mathbb{R}$, la loi des grands nombres permet de montrer que $\mathbb{F}_n(x)$ converge presque sûrement (p.s.) vers $F_X(x)$, et le théorème central limite permet de montrer que $\sqrt{n}(\mathbb{F}_n(x) - F_X(x))$ converge en loi vers une variable de loi

$\mathcal{N}(0, F_X(x)(1-F_X(x)))$. Le théorème de Glivenko-Cantelli, énoncé ci-dessous, étend le premier résultat à une convergence uniforme.

Théorème 6.1 (Glivenko-Cantelli) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes et de même loi, et soit F_X leur fonction de répartition commune. La suite des fonctions de répartition empiriques \mathbb{F}_n converge vers F_X uniformément en x presque sûrement :*

$$\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F_X(x)| = 0 \text{ p.s..}$$

Le théorème de Donsker donne la convergence en loi de la suite des processus empiriques $\sqrt{n}(\mathbb{F}_n - F_X)$ vers un processus gaussien.

Théorème 6.2 (Donsker) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes et de même loi, et soit F_X leur fonction de répartition commune. La suite des processus empiriques $\sqrt{n}(\mathbb{F}_n - F_X)$ converge en loi dans l'espace $D[-\infty, +\infty]$ des fonctions càdlàg sur $\mathbb{R} \cup \{-\infty, \infty\}$ vers un processus gaussien centré \mathbb{G}_{F_X} dont la fonction de covariance au point (s, t) est $F_X(s \wedge t) - F_X(s)F_X(t)$.*

Le processus \mathbb{G}_{F_X} est appelé un F_X -pont brownien. La notion de convergence en loi du processus $\sqrt{n}(\mathbb{F}_n - F_X)$ vers \mathbb{G}_{F_X} dépend de la topologie de l'espace où vivent ces processus. Le lecteur intéressé par l'étude de la convergence en loi du processus empirique réel pourra se reporter aux articles de Donsker [30, 31] et Doob [32], ainsi qu'au livre de Billingsley [12], qui étudie en détail la convergence en loi sur les espaces métriques. Une importante littérature est consacrée à différentes approches de la convergence en loi des processus à trajectoires càdlàg. Van der Vaart et Wellner [88] en donnent une synthèse exhaustive.

Les théorèmes de Glivenko-Cantelli et Donsker concernent respectivement la convergence uniforme de la fonction de répartition empirique et la convergence en loi du processus empirique $\sqrt{n}(\mathbb{F}_n - F_X)$. Mais des versions plus générales de ces théorèmes existent pour des ensembles de fonctions mesurables appelés respectivement classe de Glivenko-Cantelli et classe de Donsker. Ces notions interviennent dans l'étude des processus indexés par des ensembles de fonctions. Nous énonçons succinctement les notions dont nous aurons besoin dans ce chapitre, le

lecteur intéressé par un exposé détaillé pourra se reporter à l'ouvrage de Van der Vaart et Wellner [88] (chapitre 2).

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même loi P_X à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{E})$. Soit \mathcal{F} un ensemble de fonctions mesurables sur $(\mathcal{X}, \mathcal{E})$ à valeurs réelles, intégrables pour P_X . Notons

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i) \quad \text{et} \quad P_X f = \int f dP_X.$$

$\{\mathbb{G}_n f : f \in \mathcal{F}\}$ désigne le processus empirique indexé par \mathcal{F} , où

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P_X)f = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) - P_X f \right).$$

Soit $f \in \mathcal{F}$. $\mathbb{P}_n f$ converge presque sûrement vers $P_X f$. Si ce résultat peut être étendu à une convergence uniforme sur l'ensemble \mathcal{F} des fonctions f , c'est-à-dire si

$$\lim_{n \rightarrow +\infty} \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - P_X f| = 0, \text{ p.s..}$$

alors \mathcal{F} est appelée classe de Glivenko-Cantelli.

Supposons que $P_X f^2 < \infty$. $\mathbb{G}_n f$ converge en loi vers $\mathcal{N}(0, P_X(f - P_X f)^2)$. Les définitions et résultats suivants étendent ce résultat et considèrent la convergence en loi de la suite des processus $\{\mathbb{G}_n f : f \in \mathcal{F}\}$. Si on considère l'hypothèse supplémentaire $\sup_{f \in \mathcal{F}} |f(x) - P_X f| < \infty$ pour tout x , \mathbb{G}_n peut être vu comme un élément de l'espace $l^\infty(\mathcal{F})$ des fonctions de \mathcal{F} dans \mathbb{R} bornées.

Un ensemble \mathcal{F} de fonctions est une classe de Donsker si le processus empirique \mathbb{G}_n converge en loi vers un processus G dans l'espace $l^\infty(\mathcal{F})$. G est un processus gaussien centré de fonction de covariance $\text{cov}(Gf, Gg) = P_X fg - P_X f P_X g$ (pont Brownien sur \mathcal{F}).

Van der Vaart [86] et Van der Vaart et Wellner [88] donnent les exemples suivants de classes de Donsker.

Si $\mathcal{X} = \mathbb{R}$ et si \mathcal{F} est l'ensemble $\{f_x = 1_{[-\infty, x]} : x \in \mathbb{R}\}$ des indicatrices d'intervalles $]-\infty, x]$, alors \mathcal{F} est une classe de Donsker. $\mathbb{P}_n f_x$ se ramène à $\mathbb{F}_n(x)$ et le processus empirique \mathbb{G}_n indexé par \mathcal{F} se ramène au processus empirique réel $\sqrt{n}(\mathbb{F}_n - F_X)$. L'ensemble des fonctions uniformément bornées et de variation uniformément bornée est Donsker (exemple 19.11, [86]).

Un ensemble de fonctions $x \mapsto f_t(x)$ ($t \in T \subset \mathbb{R}^d, d \geq 1, T$ borné) indexée par t ,

lipschitzienne en t est Donsker [86].

Soit x une fonction continue à gauche ou à droite, à variation bornée sur $[0, 1]$. Parner montre [72] (lemme 2) que l'ensemble $\{f_u : u \in [0, 1]\}$ des fonctions $x(\cdot) \xrightarrow{f_u} x(u)$ ($u \in [0, 1]$) est Donsker.

D'autre part il est possible de construire des classes de Donsker à partir d'autres classes de Donsker : en particulier si \mathcal{F}_1 et \mathcal{F}_2 sont Donsker et $\sup_{\mathcal{F}_1 \cup \mathcal{F}_2} |P_X f| < \infty$, alors $\{f_1 + f_2 : f_j \in \mathcal{F}_j\}$ est Donsker. Si $\sup_{\mathcal{F}_1 \cup \mathcal{F}_2} |f| < \infty$, alors $\{f_1 f_2 : f_j \in \mathcal{F}_j\}$ est Donsker. Si \mathcal{F} est Donsker et $\mathcal{G} \subset \mathcal{F}$, alors \mathcal{G} est Donsker.

Si $\mathcal{F} = \{f : \mathcal{X} \rightarrow \mathbb{R}\}$ est une classe de Donsker telle que $\sup_{f \in \mathcal{F}} |P_X f| < \infty$, et si $\phi : \mathbb{R} \rightarrow \mathbb{R}$ est lipschitzienne sur $\mathcal{F}(\mathcal{X})$, alors l'ensemble $\{\phi \circ f : f \in \mathcal{F}\}$ est une classe de Donsker si il existe une fonction $f \in \mathcal{F}$ tel que $\phi \circ f$ soit de carré intégrable. Notons enfin que toute classe de Donsker est aussi Glivenko-Cantelli.

6.2 Consistance des estimateurs semiparamétriques

Dans cette section, nous établissons, sous les conditions C1 à C6, le résultat suivant de consistance des estimateurs semiparamétriques des paramètres du modèle conjoint. Nous désignons par $\theta_0 = (\alpha_0, \beta_0, \Lambda_{T,0})$ la valeur “vraie” du paramètre θ .

Théorème 6.3 *L'estimateur semiparamétrique $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{T,n})$ vérifie, pour tout $0 < \tau < \infty$,*

$$\lim_{n \rightarrow +\infty} \sup_{t \in [0, \tau]} |\hat{\Lambda}_{T,n}(t) - \Lambda_{T,0}(t)| = 0 \text{ p.s.}, \quad \lim_{n \rightarrow +\infty} |\hat{\beta}_n - \beta_0| = 0 \text{ p.s.},$$

$$\text{et } \lim_{n \rightarrow +\infty} |\hat{\alpha}_n - \alpha_0| = 0 \text{ p.s.}$$

Démonstration. La démarche adoptée pour montrer la convergence des estimateurs est une démarche générale, lorsque l'on veut montrer la convergence d'une suite $(a_n)_{n \in \mathbb{N}}$ dans un espace topologique. Elle procède en deux étapes.

On commence par montrer que la suite $(a_n)_{n \in \mathbb{N}}$ est relativement compacte, c'est-à-dire que l'ensemble $\{a_n : n \in \mathbb{N}\}$ est d'adhérence compacte. Puis on montre que l'ensemble des valeurs d'adhérence contient un unique élément, que l'on identifie éventuellement [12, 73].

En particulier, cette démarche a été adoptée par Murphy [65] et Parner [72] pour montrer la consistance d'estimateurs respectivement dans le modèle de fragilité sans

covariables et avec covariables. Elle a aussi été récemment utilisée pour montrer la consistance des estimateurs dans les modèles d'analyse des durées de vie définis par la relation $-\text{logit}(S_{T|Z}(t)) = \alpha(t) + \beta'Z$ (où $\alpha(t)$ est une fonction croissante du temps) (Murphy, Rossini et Van der Vaart [67]; Scharfstein, Tsiatis et Gilbert [78]). Nous adoptons ici cette démarche.

Nous montrons tout d'abord que la suite $(\hat{\theta}_n)_{n \in \mathbb{N}}$ est relativement compacte dans l'espace $(\mathbb{R}^p \times \mathbb{R}^2 \times L, |\cdot| \times |\cdot| \times \|\cdot\|_\infty)$, où $|\cdot|$ désigne la norme euclidienne et $\|\cdot\|_\infty$ désigne la norme sup sur $[0, \tau]$. Pour cela, nous montrons tout d'abord que la suite $(\hat{\Lambda}_{T,n})_{n \in \mathbb{N}}$ reste bornée quand $n \rightarrow \infty$. Nous nous plaçons sur un ensemble de réalisations des données de probabilité 1.

D'après (5.3), l'estimateur semiparamétrique $\hat{\Lambda}_{T,n}$ est donné par :

$$\hat{\Lambda}_{T,n}(t) = \sum_{i=1}^n \frac{\Delta_i 1_{\{X_i \leq t\}}}{\sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\hat{\beta}'_n W(X_i)} 1_{\{X_i \leq X\}} | y_j \right]}.$$

Remarquons que $\hat{\Lambda}_{T,n}(t) \leq \hat{\Lambda}_{T,n}(\tau)$ et que

$$\begin{aligned} \hat{\Lambda}_{T,n}(\tau) &= \sum_{i=1}^n \frac{\Delta_i 1_{\{X_i \leq \tau\}}}{\sum_{j=1}^n E_{\hat{\theta}_n} \left[e^{\hat{\beta}'_n W(X_i)} 1_{\{X_i \leq X\}} | y_j \right]} \\ &\leq \frac{1}{m} \sum_{i=1}^n \frac{\Delta_i 1_{\{X_i \leq \tau\}}}{\sum_{j=1}^n 1_{\{X_i \leq X_j\}}} \\ &\leq \frac{1}{m} \frac{\sum_{i=1}^n \Delta_i 1_{\{X_i \leq \tau\}}}{\sum_{j=1}^n 1_{\{\tau \leq X_j\}}}. \end{aligned}$$

Remarquons ensuite que $\frac{1}{n} \sum_{j=1}^n 1_{\{\tau \leq X_j\}}$ converge vers $\mathbb{P}[\tau \leq X]$, qui est strictement positif par hypothèse, donc $\left[\frac{1}{n} \sum_{j=1}^n 1_{\{\tau \leq X_j\}} \right]^{-1}$ converge vers $\frac{1}{\mathbb{P}[\tau \leq X]}$. Donc il existe c ($0 < c < \infty$) et $N \in \mathbb{N}$ tel que $\left[\frac{1}{n} \sum_{j=1}^n 1_{\{\tau \leq X_j\}} \right]^{-1} \leq \frac{1}{\mathbb{P}[\tau \leq X]} + c$ lorsque $n \geq N$. De plus, $\frac{1}{n} \sum_{j=1}^n \Delta_i 1_{\{X_i \leq \tau\}} \leq 1$. Finalement, $\hat{\Lambda}_{T,n}(t)$ est borné par $\frac{1}{m} \left[\frac{1}{\mathbb{P}[\tau \leq X]} + c \right]$ quand n devient suffisamment grand (c'est-à-dire $n \geq N$).

Soit $\phi(n)$ une sous-suite quelconque de (n) . D'après le théorème de Bolzano-Weierstrass, il existe une sous-suite $(\hat{\alpha}_{\eta(\phi(n))}, \hat{\beta}_{\eta(\phi(n))})_{n \in \mathbb{N}}$ de $(\hat{\alpha}_{\phi(n)}, \hat{\beta}_{\phi(n)})_{n \in \mathbb{N}}$ qui converge vers une limite (α^*, β^*) . Le théorème de Helly [41] assure que si $(f_n)_{n \in \mathbb{N}}$ est une suite de fonctions croissantes et uniformément bornées sur \mathbb{R} , il existe une

sous-suite de $(f_n)_{n \in \mathbb{N}}$ qui converge simplement vers une limite f . $\hat{\Lambda}_{T,n}$ reste borné quand $n \rightarrow \infty$; il existe une sous-suite de $(\hat{\Lambda}_{T,\eta(\phi(n))})_{n \in \mathbb{N}}$ (nous la notons $\hat{\Lambda}_{T,g(n)}$ pour simplifier les notations) qui converge simplement vers une limite Λ^* . Enfin, la sous-suite $(\hat{\alpha}_{g(n)}, \hat{\beta}_{g(n)})_{n \in \mathbb{N}}$ converge vers (α^*, β^*) .

Nous utiliserons par la suite le lemme de Helly-Bray rappelé par Korsholm [51]. Soit $(f_n)_{n \in \mathbb{N}}$ et $(g_n)_{n \in \mathbb{N}}$ deux suites de fonctions réelles, définies sur $[0, \tau]$, uniformément convergentes vers f et g bornées. Le lemme de Helly-Bray assure que $\int_0^\cdot g_n df_n$ converge uniformément vers $\int_0^\cdot g df$ sur $[0, \tau]$.

Il sera utile pour la suite de définir la quantité $\bar{\Lambda}_{T,n}$, intermédiaire entre $\hat{\Lambda}_{T,n}$ et $\Lambda_{T,0}$, par $\bar{\Lambda}_{T,n}(t) = \int_0^t \frac{dH_n(u)}{W_n(u; \theta_0)}$, où $H_n(u) = n^{-1} \sum_{i=1}^n \Delta_i 1_{\{X_i \leq u\}}$ et $W_n(u; \theta_0) = n^{-1} \sum_{i=1}^n E_{\theta_0} [e^{\beta_0' W(u)} 1_{\{u \leq X\}} | y_i]$. Notons que $\Lambda_{T,0}(t) = \int_0^t \frac{dH(u)}{W(u; \theta_0)}$ où $W(u; \theta_0) = E_{\theta_0} [e^{\beta_0' W(u)} 1_{\{u \leq X\}}]$ et $H(u) = E_{\theta_0} [\Delta 1_{\{X \leq u\}}]$. H_n converge uniformément vers H sur $[0, \tau]$ (les fonctions $\delta 1_{\{x \leq u\}}$ forment une classe de Donsker). Les fonctions $u \rightarrow E_{\theta_0} [e^{\beta_0' W(u)} 1_{\{u \leq X\}} | y]$ sont uniformément bornées et de variation bornée, d'où $W_n(u; \theta_0)$ converge uniformément vers $W(u; \theta_0)$. Par hypothèse, $E_{\theta_0} [e^{\beta_0' W(u)} 1_{\{u \leq X\}}]$ est strictement positif pour tout $u \in [0, \tau]$, d'où $\frac{1}{W_n(u; \theta_0)}$ converge uniformément vers $\frac{1}{W(u; \theta_0)}$ sur $[0, \tau]$. D'après le lemme de Helly-Bray,

$$\sup_{t \in [0, \tau]} \left| \int_0^t \frac{dH_n(u)}{W_n(u; \theta_0)} - \int_0^t \frac{dH(u)}{W(u; \theta_0)} \right| \rightarrow 0 \text{ quand } n \rightarrow \infty,$$

soit $\|\bar{\Lambda}_{T,n} - \Lambda_{T,0}\|_\infty \rightarrow 0$. Par le même raisonnement, on montre que $\|\hat{\Lambda}_{T,g(n)} - \Lambda^*\|_\infty \rightarrow 0$.

Ainsi, pour une sous-suite quelconque $(\hat{\theta}_{\phi(n)})_{n \in \mathbb{N}}$ de $(\hat{\theta}_n)_{n \in \mathbb{N}}$, il existe une sous-suite $(\hat{\theta}_{g(n)})_{n \in \mathbb{N}}$ qui converge vers une limite $\theta^* = (\alpha^*, \beta^*, \Lambda^*)$, donc $(\hat{\theta}_n)_{n \in \mathbb{N}}$ est relativement compacte.

Notons que

$$\frac{1}{g(n)} \sum_{i=1}^{g(n)} \left[\ln L^{(i)}(\hat{\theta}_{g(n)}) - \ln L^{(i)}(\alpha_0, \beta_0, \bar{\Lambda}_{T,g(n)}) \right] \geq 0,$$

car $\hat{\theta}_{g(n)}$ maximise $\sum_{i=1}^{g(n)} \ln L^{(i)}(\theta)$, et que pour tout $g(n)$,

$$\frac{1}{m} \sum_{i=1}^m \left[\ln L^{(i)}(\hat{\theta}_{g(n)}) - \ln L^{(i)}(\alpha_0, \beta_0, \bar{\Lambda}_{T,g(n)}) \right]$$

converge vers $E_{\theta_0} \left[\ln L(\hat{\theta}_{g(n)}) - \ln L(\alpha_0, \beta_0, \bar{\Lambda}_{T,g(n)}) \right]$ quand $m \rightarrow \infty$. Finalement, $E_{\theta_0} \left[\ln L(\hat{\theta}_{g(n)}) - \ln L(\alpha_0, \beta_0, \bar{\Lambda}_{T,g(n)}) \right] \geq 0$ quand $n \rightarrow \infty$.

Notons que $\ln L(\hat{\theta}_{g(n)}) - \ln L(\alpha_0, \beta_0, \bar{\Lambda}_{T,g(n)})$ converge vers $\ln L(\theta^*) - \ln L(\theta_0)$.

D'après le théorème de Lebesgue, $E_{\theta_0} \left[\ln L(\hat{\theta}_{g(n)}) - \ln L(\alpha_0, \beta_0, \bar{\Lambda}_{T,g(n)}) \right]$ converge vers $E_{\theta_0} [\ln L(\theta^*) - \ln L(\theta_0)]$. D'où $E_{\theta_0} \left[\ln \frac{L(\theta^*)}{L(\theta_0)} \right] \geq 0$. Or cette quantité est toujours négative ou nulle (propriété de la distance de Kullback-Leibler).

D'où $E_{\theta_0} \left[\ln \frac{L(\theta^*)}{L(\theta_0)} \right] = 0$.

D'après la proposition 4.4 (identifiabilité du paramètre θ), $\theta^* = \theta_0$. Finalement, $(\hat{\theta}_n)_{n \in \mathbb{N}}$ converge vers θ_0 .

■

6.3 Convergence en loi des estimateurs semiparamétriques

Préliminaires

Considérons un modèle paramétrique ou semiparamétrique $(\mathcal{Y}, \mathcal{P} = \{P_\psi : \psi \in \Psi\})$ dans lequel intervient un paramètre inconnu ψ , et soit Y un vecteur aléatoire à valeurs dans \mathcal{Y} et de loi $P^Y \in \mathcal{P}$. Soit Y_1, \dots, Y_n un n -échantillon de Y . Van der Vaart et Wellner [88] étudient les propriétés asymptotiques des estimateurs de ψ obtenus comme solution du problème

$$S_n(\psi) = 0,$$

où S_n est une fonction donnée.

Dans le cas particulier de l'estimation par maximum de vraisemblance dans un modèle d'échantillonnage paramétrique défini par la densité $f_Y(Y; \psi)$,

$$S_n(\psi) = \sum_{i=1}^n \frac{\partial}{\partial \psi} \ln f_Y(Y_i; \psi).$$

Lorsque ψ est un paramètre de dimension finie, le nombre d'équations d'estimation $\sum_{i=1}^n \partial \ln f_Y(Y_i; \psi) / \partial \psi_j$ est égal à la dimension de ψ . Un problème se pose lorsque ψ contient un paramètre fonctionnel (modèles semiparamétriques).

Van der Vaart et Wellner [88] se placent dans le cadre général d'une fonction S_n définie sur un espace de Banach Ψ et à valeurs dans un espace de Banach (ce qui permet de considérer le cas où S_n est à valeurs dans un espace fonctionnel par exemple) et énoncent le théorème suivant, relatif à la normalité asymptotique des estimateurs obtenus à partir d'une équation $S_n(\psi) = 0$.

Théorème 6.4 (Van der Vaart et Wellner, 1996) *Soit S_n et S deux applications définies sur un sous-espace d'un espace de Banach Ψ et à valeurs dans un espace de Banach.*

Supposons qu'il existe deux éléments ψ_0 et $\hat{\psi}_n$ de Ψ vérifiant $S(\psi_0) = 0$, $S_n(\hat{\psi}_n) = 0$ et $\hat{\psi}_n \xrightarrow{P} \psi_0$. Supposons que

$$\sqrt{n}S(\hat{\psi}_n) + \sqrt{n}S_n(\psi_0) = o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|), \quad (6.2)$$

et que $\sqrt{n}(S_n(\psi_0) - S(\psi_0)) \xrightarrow{d} Z$. Supposons que $\psi \mapsto S(\psi)$ soit différentiable au sens de Fréchet en ψ_0 , que sa différentielle \dot{S}_{ψ_0} soit inversible et que l'inverse $\dot{S}_{\psi_0}^{-1}$ soit continue sur l'image de \dot{S}_{ψ_0} . Alors

$$-\sqrt{n}\dot{S}_{\psi_0}(\hat{\psi}_n - \psi_0) = \sqrt{n}(S_n - S)(\psi_0) + o_P(1).$$

D'où

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} -\dot{S}_{\psi_0}^{-1}Z.$$

Une condition importante est que l'inverse $\dot{S}_{\psi_0}^{-1}$ de la différentielle \dot{S}_{ψ_0} existe et soit continue. Dans le cas particulier de l'estimation par maximum de vraisemblance dans un modèle paramétrique $(\mathcal{Y}, P_\psi : \psi \in \Psi \subset \mathbb{R}^p)$, cette condition se ramène à supposer seulement que la matrice d'information de Fisher $E_{\psi_0} \left[-\frac{\partial^2 \ln f_Y(y; \psi_0)}{\partial \psi \partial \psi'} \right]$ existe et est inversible (où ψ_0 désigne la “vraie” valeur du paramètre). En effet, une application linéaire sur un espace de dimension finie est continue. Dans le cas d'un paramètre fonctionnel, le problème sera plus compliqué.

Les paramètres du modèle conjoint proposé sont le paramètre α de la loi de Z et les paramètres β et Λ_T du modèle de Cox. L'étude de la normalité asymptotique de l'estimateur semiparamétrique $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{T,n})$ (où $\hat{\Lambda}_{T,n} = (\hat{\Lambda}_{n,1}, \dots, \hat{\Lambda}_{n,p(n)})$) se heurte aux problèmes de la nature fonctionnelle du paramètre Λ_T et de la dimension de $\hat{\theta}_n$ qui augmente avec le nombre d'observations.

Nous proposons d'adapter à notre contexte l'approche développée par Murphy [66] pour établir la théorie asymptotique dans le modèle de fragilité. Cette approche nous permet d'obtenir une fonction score pour les paramètres et de lui choisir un espace de définition approprié. Après cette étape préliminaire, nous obtenons un résultat de normalité asymptotique pour l'estimateur semiparamétrique $\hat{\theta}_n$ en vérifiant que cette fonction score satisfait les conditions énoncées par Van der Vaart et Wellner [88] dans le théorème 6.4.

La première étape est le calcul du score. La démarche habituelle pour obtenir l'estimateur du maximum de vraisemblance $\hat{\phi}$ du paramètre q -dimensionnel ϕ d'un modèle paramétrique consiste à calculer les dérivées partielles de la log-vraisemblance $l(\phi)$ par rapport aux composantes ϕ_i ($i = 1, \dots, q$) du vecteur ϕ . Dans le cas où ϕ est un vecteur de grande dimension, il peut être plus pratique de considérer la fonction $t \mapsto l(\hat{\phi} + th)$, qui traduit la restriction de l à une droite de direction h passant par $\hat{\phi}$ (où $h = (h_1, \dots, h_q)$). Ceci permet de calculer les dérivées partielles tout en se ramenant au cas plus simple d'une fonction de variable réelle car

$$\frac{\partial}{\partial t} l(\hat{\phi} + th) \Big|_{t=0} = \sum_{i=1}^q \frac{\partial l}{\partial \phi_i}(\hat{\phi}) h_i.$$

Cette démarche est utilisée par Murphy [66] dans le modèle de fragilité.

Dans notre contexte, le paramètre est $(\alpha_i \ (i = 1, \dots, p), \beta_1, \beta_2, \Lambda_{n,1}, \dots, \Lambda_{n,p(n)})$. Nous allons utiliser la démarche décrite ci-dessus pour obtenir le score, en procédant à partir de la log-vraisemblance EM $L_{n,\hat{\theta}}(\theta)$ obtenue au chapitre 5.

Démonstration de la convergence en loi

Soit H l'espace des directions h . H sera défini comme l'espace produit de \mathbb{R}^p , \mathbb{R}^2 , et de l'espace noté $VB([0, \tau])$ des applications bornées et à variation bornée de $[0, \tau]$ dans \mathbb{R} :

$$H = \{h = (h_1, h_2, h_3) : h_1 \in \mathbb{R}^p, h_2 \in \mathbb{R}^2, h_3 \in VB([0, \tau])\}.$$

Le choix de cet espace sera justifié par la suite. Nous le munissons de la norme $\|\cdot\|_H$ définie par

$$\|h\|_H = |h_1| + |h_2| + \|h_3\|_\nu,$$

où $\|h_3\|_\nu = |h_3(0)| + \int_0^\tau |dh_3(u)| \cdot \int_0^\tau |df(u)|$ est la variation totale d'une fonction f sur $[0, \tau]$, nous la noterons encore $V_{[0,\tau]}(f)$. Nous noterons de plus par H_p et H_∞ les sous-espaces suivants de H : $H_p = \{h \in H : \|h\|_H \leq p\}$ ($0 < p < \infty$) et $H_\infty = \{h \in H : \|h\|_H < \infty\}$. Enfin, VB_p désignera l'espace des fonctions de $[0, \tau]$ dans \mathbb{R} , bornées par p et à variation bornée par p ($p < \infty$).

Notre démonstration de la convergence en loi de l'estimateur semiparamétrique s'appuiera sur la log-vraisemblance EM.

Nous obtenons d'abord le score, noté $S_{n,\tilde{\theta}}(\theta)(h)$, à partir des observations incomplètes y_i fournies par n réalisations indépendantes de (X, Δ, Z) , et de la log-vraisemblance EM.

Pour $\theta = (\alpha, \beta, \Lambda_T)$, où Λ_T est une fonction positive, croissante et bornée sur $[0, \infty[$, et pour une valeur $\tilde{\theta}$ de θ , le score se calcule de la manière suivante :

$$S_{n,\tilde{\theta}}(\theta)(h) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial t} L_{\tilde{\theta}}^{(i)}(\theta + th) \Big|_{t=0}.$$

Nous définissons les notations :

$$\begin{aligned} S_{n,\tilde{\theta},1}(\theta) &= \frac{1}{n} \sum_{i=1}^n E_{\tilde{\theta}} \left[\frac{\partial}{\partial \alpha} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y_i \right] \\ S_{n,\tilde{\theta},2}(\theta) &= \frac{1}{n} \sum_{i=1}^n E_{\tilde{\theta}} \left[\Delta W(X) - \int_0^X W(u) e^{\beta' W(u)} d\Lambda_T(u) | y_i \right] \\ S_{n,\tilde{\theta},3}(\theta)(h_3) &= \frac{1}{n} \sum_{i=1}^n \left[\delta_i h_3(x_i) - E_{\tilde{\theta}} \left[\int_0^X h_3(u) e^{\beta' W(u)} d\Lambda_T(u) | y_i \right] \right] \\ S_{n,\tilde{\theta},12}^T(\theta) &= (S_{n,\tilde{\theta},1}^T, S_{n,\tilde{\theta},2}^T)(\theta) \text{ et } h'_{12} = (h'_1, h'_2), \end{aligned}$$

et nous démontrons la proposition suivante :

Proposition 6.1 *Soit $h \in H_p$ ($0 < p < \infty$). Le score $S_{n,\tilde{\theta}}(\theta)(h)$ est donné par :*

$$S_{n,\tilde{\theta}}(\theta)(h) = h'_{12} S_{n,\tilde{\theta},12}(\theta) + S_{n,\tilde{\theta},3}(\theta)(h_3).$$

Démonstration. Rappelons que

$$\begin{aligned} L_{\tilde{\theta}}^{(j)}(\theta) &= E_{\tilde{\theta}} [\ln l(Y, Z; \theta) | y_j] \\ &= E_{\tilde{\theta}} \left[\Delta \ln d\Lambda_T(X) + \Delta \beta' W(X) - \int_0^X e^{\beta' W(u)} d\Lambda_T(u) \right. \\ &\quad \left. + \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y_j \right]. \end{aligned}$$

Soit $h = (h_1, h_2, h_3) \in H_p$ ($0 < p < \infty$). Soit $\theta_t = (\alpha_t, \beta_t, \Lambda_t)$, où $\theta_t = \theta + th$, $\alpha_t = \alpha + th_1$, $\beta_t = \beta + th_2$, $\Lambda_t(\cdot) = \int_0^\cdot (1 + th_3(u)) d\Lambda_T(u)$.

La fonction $L_{\tilde{\theta}}(\theta_t)$ est égale à (dans les calculs qui suivent, nous supprimons l'indice j de la notation, pour ne pas alourdir les formules) :

$$L_{\tilde{\theta}}(\theta_t) = \delta \ln[(1 + th_3(x))d\Lambda_T(x)] - E_{\tilde{\theta}} \left[\int_0^X e^{[\beta + th_2]'W(u)} (1 + th_3(u)) d\Lambda_T(u) | y \right] \\ + E_{\tilde{\theta}} [\Delta[\beta + th_2]'W(X) | y] + E_{\tilde{\theta}} [\ln f(Z_0, \dots, Z_{a_X}, Z; \alpha + th_1) | y].$$

Nous calculons sa dérivée $\frac{\partial}{\partial t} L_{\tilde{\theta}}(\theta_t)$:

$$\frac{\partial}{\partial t} L_{\tilde{\theta}}(\theta_t) = \delta \frac{h_3(x)}{1 + th_3(x)} - E_{\tilde{\theta}} \left[\int_0^X h_3(u) e^{[\beta + th_2]'W(u)} d\Lambda_T(u) | y \right] \\ + h'_2 E_{\tilde{\theta}} \left[\Delta W(X) - \int_0^X W(u) e^{[\beta + th_2]'W(u)} (1 + th_3(u)) d\Lambda_T(u) | y \right] \\ + E_{\tilde{\theta}} \left[\frac{\partial}{\partial t} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha + th_1) | y \right],$$

et nous en déduisons

$$\left. \frac{\partial}{\partial t} L_{\tilde{\theta}}(\theta_t) \right|_{t=0} = \delta h_3(x) - E_{\tilde{\theta}} \left[\int_0^X h_3(u) e^{\beta'W(u)} d\Lambda_T(u) | y \right] \\ + h'_2 E_{\tilde{\theta}} \left[\Delta W(X) - \int_0^X W(u) e^{\beta'W(u)} d\Lambda_T(u) | y \right] \\ + h'_1 E_{\tilde{\theta}} \left[\frac{\partial}{\partial \alpha} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y \right].$$

En utilisant les notations définies ci-dessus, nous déduisons de cette expression l'équation

$$S_{n,\tilde{\theta}}(\theta)(h) = h'_{12} S_{n,\tilde{\theta},12}(\theta) + S_{n,\tilde{\theta},3}(\theta)(h_3).$$

■

Par la suite, nous noterons par $S_{\tilde{\theta},12}^{(i)}(\theta)$ ($i = 1, \dots, n$) les contributions à la somme $S_{n,\tilde{\theta},12}(\theta) = n^{-1} \sum_{i=1}^n S_{\tilde{\theta},12}^{(i)}(\theta)$. Nous noterons également

$$s_{\tilde{\theta}}(y; \theta)(h) = \frac{\partial}{\partial t} L_{\tilde{\theta}}(\theta_t) |_{t=0}$$

et

$$S_{\tilde{\theta}}(\theta)(h) = E_{\theta_0} [s_{\tilde{\theta}}(Y; \theta)(h)].$$

Le score donné par la proposition précédente vérifie les propriétés suivantes :

Proposition 6.2 *Soit $h \in H_p$ ($0 < p < \infty$). Alors $S_{n,\hat{\theta}_n}(\hat{\theta}_n)(h) = 0$ et $S_{\theta_0}(\theta_0)(h) = 0$.*

Démonstration. Rappelons que d'après la proposition 5.2, la solution $\hat{\theta}_n$ du problème $\max_{\theta \in \Theta_n} \prod_{i=1}^n f_Y(Y_i; \theta)$ vérifie

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} [E_{\hat{\theta}_n} [\ln f_{Y,Z}(Y, Z; \theta) | y_i]] \Big|_{\theta=\hat{\theta}_n} = 0,$$

ou, de manière équivalente :

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} L_{\hat{\theta}_n}^{(i)}(\theta) \Big|_{\theta=\hat{\theta}_n} = 0.$$

Ceci revient à :

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial t} L_{\hat{\theta}_n}^{(i)}(\hat{\theta}_n + th) \Big|_{t=0} = S_{n,\hat{\theta}_n}(\hat{\theta}_n)(h) = 0.$$

Nous avons montré dans la proposition 4.4 que le paramètre $\theta = (\alpha, \beta, \Lambda_T)$ est identifiable, c'est-à-dire que l'application $\theta \mapsto E_{\theta_0}[\ln f_Y(Y; \theta)]$ est maximale pour la seule valeur $\theta = \theta_0$ (Proposition 4.3). D'où l'application $t \mapsto E_{\theta_0}[\ln f_Y(Y; \theta_{0,t})]$ a un unique maximum en $t = 0$ (avec $\theta_{0,t} = (\alpha_{0,t}, \beta_{0,t}, \Lambda_{T,0,t})$ et $\alpha_{0,t} = \alpha_0 + th_1, \beta_{0,t} = \beta_0 + th_2, \Lambda_{T,0,t}(\cdot) = \int_0^\cdot (1 + th_3(u)) d\Lambda_{T,0}(u), h \in H_p, p < \infty$). D'où

$$E_{\theta_0} \left[\frac{\partial}{\partial t} \ln f_Y(Y; \theta_{0,t}) \Big|_{t=0} \right] = 0. \quad (6.3)$$

Notons que

$$\frac{\partial}{\partial t} \ln f_Y(y; \theta_{0,t}) \Big|_{t=0} = \frac{\partial}{\partial t} [E_{\theta_0} [\ln f_{Y,Z}(Y, Z; \theta_{0,t}) | y]] \Big|_{t=0}.$$

Nous pouvons donc réécrire (6.3) :

$$E_{\theta_0} \left[\frac{\partial}{\partial t} [E_{\theta_0} [\ln f_{Y,Z}(Y, Z; \theta_{0,t}) | y]] \Big|_{t=0} \right] = 0.$$

En retirant de $\ln f_{Y,Z}(Y, Z; \theta_{0,t})$ les termes relatifs à la loi de la censure, ceci se réécrit

$$\begin{aligned} E_{\theta_0} \left[\frac{\partial}{\partial t} [E_{\theta_0} [\ln l(Y, Z; \theta_{0,t}) | y]] \Big|_{t=0} \right] &= E_{\theta_0} \left[\frac{\partial}{\partial t} L_{\theta_0}(\theta_{0,t}) \Big|_{t=0} \right] \\ &= E_{\theta_0} [s_{\theta_0}(Y; \theta_0)(h)] \\ &= S_{\theta_0}(\theta_0)(h) \\ &= 0. \end{aligned}$$

D'où $S_{\theta_0}(\theta_0)(h) = 0$ pour $h \in H_p$ ($p < \infty$).

■

La deuxième étape des préliminaires à la démonstration de la convergence en loi de l'estimateur semiparamétrique $\hat{\theta}_n$ va permettre de gérer le problème de la dimension de l'estimateur $\hat{\theta}_n$ et du paramètre $\theta = (\alpha, \beta, \Lambda_T)$, en choisissant un espace de définition approprié pour $S_{n,\tilde{\theta}}$. Nous reprenons l'approche de Murphy [66] et considérons le paramètre θ comme une application de H_∞ dans \mathbb{R} définie par

$$\theta : h \longmapsto \theta(h) = (\alpha, \beta, \Lambda_T)(h) = h'_1 \alpha + h'_2 \beta + \int_0^\tau h_3(u) d\Lambda_T(u).$$

Pour des choix appropriés de h_1 , h_2 , et h_3 , $\theta(h)$ se ramène aux paramètres α, β et $\Lambda_T(\cdot)$. Par exemple, si $h_2 = 0$, $h_3 = 0$, $h_{1i} = 0$ et $h_{1j} = 1$ ($j \in \{1, \dots, p\}$, $i = 1, \dots, p$, $i \neq j$), alors $\theta(h)$ est la composante α_j du vecteur α . Si $h_1 = 0$, $h_2 = 0$, et $h_3 = 1_{\{u \leq t\}}$ (pour $t \leq \tau$), alors $\theta(h) = \Lambda_T(t)$.

L'espace Θ des paramètres est maintenant un espace de fonctions θ définies sur H_∞ , bornées, à valeurs dans \mathbb{R} . Si $l^\infty(T)$ désigne l'espace des fonctions à valeurs réelles, définies sur T et bornées, nous pouvons considérer : $\Theta \subset l^\infty(H_\infty)$. Le score $S_{n,\tilde{\theta}}$ est donc considéré comme une application de Θ dans $l^\infty(H_\infty)$. Nous munissons les espaces $l^\infty(H_p)$ ($p < \infty$) de la norme $\|U\|_p = \sup_{h \in H_p} |U(h)|$.

Nous identifions $S_{n,\hat{\theta}_n}$ à l'application S_n du théorème 6.4 et l'espace des paramètres Θ ($\Theta \in l^\infty(H_\infty)$) à l'espace Ψ . Nous identifions S_{θ_0} à S . L'estimateur semiparamétrique $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{T,n})$ et la "vraie" valeur $\theta_0 = (\alpha_0, \beta_0, \Lambda_{T,0})$ sont considérés comme des éléments de $\Theta \subset l^\infty(H_\infty)$.

Nous vérifions que $S_{n,\hat{\theta}_n} : \Theta \longrightarrow l^\infty(H_\infty)$ et S_{θ_0} satisfont aux conditions énoncées par Van der Vaart et Wellner [88], et nous établissons le résultat suivant, sous les conditions C1 à C6 :

Théorème 6.5 *Pour tout p ($0 < p < \infty$), la suite*

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))$$

converge en loi dans $l^\infty(H_p)$ vers un processus gaussien centré G , de fonction de covariance

$$\text{cov}[G(g), G(g^*)] = \int_0^\tau g_3(u) \sigma_{3,\theta_0}^{-1}(g^*)(u) d\Lambda_{T,0}(u) + \sigma_{2,\theta_0}^{-1}(g^*)' g_2 + \sigma_{1,\theta_0}^{-1}(g^*)' g_1,$$

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 de H_∞ dans H_∞ $\sigma_{\theta_0} = (\sigma_{1,\theta_0}, \sigma_{2,\theta_0}, \sigma_{3,\theta_0})$ défini par

$$\begin{aligned} \sigma_{1,\theta_0}(h) &= -E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha_0) \right] h_1, \\ \sigma_{2,\theta_0}(h) &= E_{\theta_0} \left[\int_0^X W(u) e^{\beta_0' W(u)} (W(u)' h_2 + h_3(u)) d\Lambda_{T,0}(u) \right], \\ \sigma_{3,\theta_0}(h)(u) &= E_{\theta_0} \left[(h_2' W(u) + h_3(u)) e^{\beta_0' W(u)} 1_{\{u \leq X\}} \right]. \end{aligned}$$

Démonstration. Pour démontrer ce théorème, nous vérifions que les conditions du théorème 6.4 sont satisfaites par $S_{n,\hat{\theta}_n}$ et S_{θ_0} . Nous montrons tout d'abord que $S_{n,\hat{\theta}_n}$ et S_{θ_0} vérifient la condition (6.2) :

Lemme 6.1 *Pour tout q ($0 < q < \infty$),*

$$\|\sqrt{n}(S_{\theta_0}(\hat{\theta}_n) + S_{n,\hat{\theta}_n}(\theta_0))\|_q = o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|_q).$$

Démonstration.

Pour démontrer ce lemme, nous utilisons le résultat suivant qui énonce une condition suffisante pour que (6.2) soit satisfaite [88].

Reprenons les notations du théorème 6.4. Si $S(\psi)$ est une fonction définie sur un ensemble \mathcal{H} de h , et peut s'écrire sous la forme $E_{\psi_0} \varphi_\psi(h)$ où $\varphi_\psi(h)$ est une fonction indexée par $\psi \in \Psi$ et définie sur \mathcal{H} , une condition suffisante pour que la condition (6.2) soit remplie est que la classe de fonctions $\{\varphi_{\psi_0}(h) - \varphi_\psi(h) : h \in \mathcal{H}, \|\psi - \psi_0\| < \epsilon\}$ (pour un $\epsilon > 0$) soit Donsker, que $\sup_{h \in \mathcal{H}} E_{\psi_0} [\varphi_{\psi_0}(h) - \varphi_\psi(h)]^2 \xrightarrow{P} 0$ quand $\psi \rightarrow \psi_0$, et que $\hat{\psi}_n \xrightarrow{P} \psi_0$ [88].

Rappelons que

$$\begin{aligned} S_{\theta_0}(\theta)(h) &= E_{\theta_0}[s_{\theta_0}(Y; \theta)(h)] \\ &= E_{\theta_0} \left[\Delta h_3(X) - \int_0^X h_3(u) e^{\beta' W(u)} d\Lambda_T(u) + h'_2 \Delta W(X) \right. \\ &\quad \left. - \int_0^X h'_2 W(u) e^{\beta' W(u)} d\Lambda_T(u) + h'_1 \frac{\partial}{\partial \alpha} \ln f(Z_0, \dots, Z_{a_x}, Z; \alpha) \right]. \end{aligned}$$

Soit q positif fini. Nous montrons que l'ensemble $\{\varphi_{\theta_0}(h) - \varphi_{\theta}(h) : \|\theta - \theta_0\|_q < \epsilon, h \in H_q\}$ égal à

$$\begin{aligned} &\left\{ \int_0^x h_3(u) e^{\beta'_0 w(u)} d\Lambda_{T,0}(u) - \int_0^x h_3(u) e^{\beta' w(u)} d\Lambda_T(u) \right. \\ &\quad + \int_0^x h'_2 w(u) e^{\beta'_0 w(u)} d\Lambda_{T,0}(u) - \int_0^x h'_2 w(u) e^{\beta' w(u)} d\Lambda_T(u) \\ &\quad \left. + h'_1 \frac{\partial}{\partial \alpha} \ln f(z_0, \dots, z_{a_x}, z; \alpha) - h'_1 \frac{\partial}{\partial \alpha} \ln f(z_0, \dots, z_{a_x}, z; \alpha_0) : \right. \\ &\quad \left. \|\theta - \theta_0\|_q < \epsilon, h \in H_q \right\} \end{aligned}$$

est une classe de Donsker (pour un $\epsilon > 0$). Nous montrons qu'elle s'obtient à partir de classes de Donsker plus simples sur lesquelles sont appliquées des transformations qui conservent cette propriété.

Soit $\epsilon > 0$ fini. Nous illustrons la méthode en montrant que les classes

$$\left\{ \int_0^\tau h_3(u) e^{\beta' w(u)} 1_{\{u \leq x\}} d\Lambda_T(u) : \beta \in \tilde{B}, h_3 \in VB_q, \Lambda_T \in VB_n \right\}$$

et

$$\left\{ h'_1 \frac{\partial}{\partial \alpha} \ln f(z_0, \dots, z_{a_x}, z; \alpha) : \alpha \in \tilde{A}, h_1 \in \mathbb{R}^p, |h_1| \leq q \right\}$$

sont Donsker (où $\tilde{A} = \{\alpha : |\alpha - \alpha_0| < \sqrt{p} \frac{\epsilon}{q}\}$, $\tilde{B} = \{\beta : |\beta - \beta_0| < \sqrt{2} \frac{\epsilon}{q}\}$ et $\frac{\epsilon}{q} + \Lambda_{T,0}(\tau) = n$).

D'après le lemme 2 de Parner [72], les classes $f_u : 1_{[\cdot, \infty[} \mapsto 1_{\{u, \infty[}$ et $g_u : \beta' w(\cdot) \mapsto \beta' w(u)$ ($u \in [0, \tau], \beta \in B$) sont Donsker. Puis $h_u : e^{\beta' w(\cdot)} \mapsto e^{\beta' w(u)}$ est Donsker (la fonction exponentielle est lipschitzienne sur tout ensemble borné de \mathbb{R} et $w(\cdot)$ est uniformément borné). Finalement, en multipliant deux classes de Donsker uniformément bornées, $i_u : e^{\beta' w(\cdot)} 1_{[\cdot, \infty[} \mapsto e^{\beta' w(u)} 1_{\{u, \infty[}$ ($u \in [0, \tau], \beta \in B$) est Donsker.

Soit $y_b : u \mapsto y_b(u)$ une application définie sur $[0, \tau]$ et à valeurs dans \mathbb{R} (avec $b \in B' \subset \mathbb{R}^d, d \geq 1$) et soit $\phi : y_b \mapsto \phi(y_b)$ l'application linéaire et continue de $l^\infty([0, \tau])$ dans $l^\infty(VB_q \times VB_m)$ ($m > 0$) définie par

$$\phi(y_b)(l, k) = \int_0^\tau l(u) y_b(u) dk(u).$$

Supposons que $j_u : y_b(\cdot) \mapsto y_b(u)$ ($u \in [0, \tau], \beta \in B'$) soit Donsker. Alors

$$\mathbb{G}_n j_u = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n y_{b,i}(u) - E y_b(u) \right)$$

converge en loi. Par continuité, $\phi(\mathbb{G}_n j_u)$ converge aussi en loi. Remarquons que

$$\begin{aligned} \phi(\mathbb{G}_n j_u)(l, k) &= \phi \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n y_{b,i}(u) - E y_b(u) \right) \right) (l, k) \\ &= \int_0^\tau l(u) \left(\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n y_{b,i}(u) - E y_b(u) \right) \right) dk(u) \\ &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \int_0^\tau l(u) y_{b,i}(u) dk(u) - E \int_0^\tau l(u) y_b(u) dk(u) \right) \\ &= \mathbb{G}_n(\phi(y_b)(l, k)). \end{aligned}$$

D'où $y_b(\cdot) \mapsto \int_0^\tau l(u) y_b(u) dk(u)$ est Donsker ($b \in B', l \in VB_q, k \in VB_m$).

Si nous appliquons ce raisonnement avec $y_b(\cdot) = e^{\beta' w(\cdot)} 1_{\{[\cdot, \infty]\}}$, nous concluons que $e^{\beta' w(\cdot)} 1_{\{[\cdot, \infty]\}} \mapsto \int_0^\tau h_3(u) e^{\beta' w(u)} 1_{\{[u, \infty]\}} d\Lambda_T(u)$ est Donsker ($\beta \in \tilde{B}, h_3 \in VB_q, \Lambda_T \in VB_n$).

Nous avons supposé que f est deux fois dérivable par rapport à α et que ses dérivées secondes sont continues en α , d'où $h'_1 \partial \ln f / \partial \alpha$ est lipschitzienne en α . D'après le résultat rappelé dans les préliminaires de cette section, la classe $\{h'_1 \frac{\partial}{\partial \alpha} \ln f(z_0, \dots, z_{a_x}, z; \alpha)\}$ est Donsker ($\alpha \in \tilde{A}, h_1 \in \mathbb{R}^p, |h_1| \leq q$).

Nous pouvons ainsi montrer que l'ensemble $\{\varphi_{\theta_0}(h) - \varphi_\theta(h) : \alpha \in \tilde{A}, \beta \in \tilde{B}, \Lambda_T \in VB_n, h_1 \in \mathbb{R}^p, |h_1| \leq q, h_2 \in \mathbb{R}^2, |h_2| \leq q, h_3 \in VB_q\}$ est une classe de Donsker. Nous montrons que les conditions $\{\|\theta - \theta_0\|_q < \epsilon, h \in H_q\}$ en définissent un sous-ensemble.

Soit $\theta = (\alpha, \beta, \Lambda_T) \in \{\theta : \|\theta - \theta_0\|_q < \epsilon\}$. Pour tout $h \in H_q$, $|(\theta - \theta_0)(h)| \leq \|\theta - \theta_0\|_q$.

Soit $\bar{h} = (\bar{h}_1, 0, 0)$ tel que $\bar{h}_{1j} = q$ ($j \in \{1, \dots, p\}$) et $\bar{h}_{1i} = 0$ ($i = 1, \dots, p, i \neq j$). $\bar{h} \in H_q$ car $\|\bar{h}\|_H = q$. Notons que $|(\theta - \theta_0)(\bar{h})| = q|\alpha_j - \alpha_{j0}| < \epsilon$, d'où $|\alpha_j - \alpha_{j0}| < \frac{\epsilon}{q}$, $j = 1, \dots, p$. Il vient : $|\alpha - \alpha_0| < \sqrt{p}\frac{\epsilon}{q}$. De même, $|\beta - \beta_0| < \sqrt{2}\frac{\epsilon}{q}$. Donc si $\theta \in \{\theta : \|\theta - \theta_0\|_q < \epsilon\}$, $\alpha \in \tilde{A}$ et $\beta \in \tilde{B}$.

Soit $\bar{h} = (0, 0, \bar{h}_3)$ et $\bar{h}_3 = q$. $\bar{h} \in H_q$ car $\|\bar{h}\|_H = q$. Notons que $|(\theta - \theta_0)(\bar{h})| = q|\Lambda_T(\tau) - \Lambda_{T,0}(\tau)| < \epsilon$, d'où $\Lambda_T(\tau) < \Lambda_{T,0}(\tau) + \frac{\epsilon}{q}$. Donc $\|\Lambda_T\|_\nu < \Lambda_{T,0}(\tau) + \frac{\epsilon}{q}$. Finalement $\Lambda_T \in VB_n$ où $n = \Lambda_{T,0}(\tau) + \frac{\epsilon}{q}$.

L'ensemble $\{\varphi_{\theta_0}(h) - \varphi_\theta(h) : \|\theta - \theta_0\|_q < \epsilon, h \in H_q\}$ est inclus dans $\{\varphi_{\theta_0}(h) - \varphi_\theta(h) : \alpha \in \tilde{A}, \beta \in \tilde{B}, \Lambda_T \in VB_n, h_1 \in \mathbb{R}^p, |h_1| \leq q, h_2 \in \mathbb{R}^2, |h_2| \leq q, h_3 \in VB_q\}$, c'est donc une classe de Donsker.

La condition $\sup_{h \in H_q} E_{\theta_0}[\varphi_{\theta_0}(h) - \varphi_\theta(h)]^2 \rightarrow 0$ découle du théorème de convergence dominé.

■

Notons :

$$\begin{aligned} \sigma_{1,\theta}(h) &= -E_{\theta_0} \left[E_\theta \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y \right] \right] h_1, \\ \sigma_{2,\theta}(h) &= E_{\theta_0} \left[E_\theta \left[\int_0^X W(u) e^{\beta' W(u)} [h'_2 W(u) + h_3(u)] d\Lambda_T(u) | y \right] \right], \\ \sigma_{3,\theta}(h)(u) &= E_{\theta_0} \left[E_\theta \left[[h'_2 W(u) + h_3(u)] e^{\beta' W(u)} 1_{\{u \leq X\}} | y \right] \right]. \end{aligned} \quad (6.4)$$

Nous montrons maintenant que l'application de Θ dans $l^\infty(H_p) : \theta \mapsto S_{\theta_0}(\theta)$ est différentiable au sens de Fréchet au point θ_0 , et nous calculons sa différentielle $\dot{S}_{\theta_0}(\theta_0)$ en θ_0 .

Lemme 6.2 *Pour tout p ($0 < p < \infty$), l'application $S_{\theta_0} : \theta \mapsto S_{\theta_0}(\theta)$ de $\Theta \subset l^\infty(H_p)$ dans $l^\infty(H_p)$ est différentiable au sens de Fréchet en θ_0 , et sa différentielle $\dot{S}_{\theta_0}(\theta_0)$ au point θ_0 est donnée par :*

$$\dot{S}_{\theta_0}(\theta_0)(\theta)(h) = - \int_0^\tau \sigma_{3,\theta_0}(h)(u) d\Lambda_T(u) - \beta' \sigma_{2,\theta_0}(h) - \alpha' \sigma_{1,\theta_0}(h).$$

Démonstration. Rappelons que

$$\begin{aligned} S_{\theta_0}(\theta)(h) &= E_{\theta_0} \left[\Delta h_3(X) - \int_0^X h_3(u) e^{\beta' W(u)} d\Lambda_T(u) + h'_2 \Delta W(X) \right. \\ &\quad \left. - \int_0^X h'_2 W(u) e^{\beta' W(u)} d\Lambda_T(u) + h'_1 \frac{\partial}{\partial \alpha} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) \right]. \end{aligned}$$

Pour montrer que S_{θ_0} est différentiable au sens de Fréchet en θ_0 , nous adoptons la démarche décrite dans les rappels de calcul différentiel figurant dans les préliminaires. Notons $\theta_0 + t\theta = (\alpha_0 + t\alpha, \beta_0 + t\beta, \Lambda_{T,0}(\cdot) + t\Lambda_T(\cdot))$.

Nous calculons tout d'abord la différentielle de Gâteaux $DS_{\theta_0}(\theta_0)$ de S_{θ_0} au point θ_0 de la manière suivante :

$$D_{\theta}S_{\theta_0}(\theta_0) = \left. \frac{\partial}{\partial t} S_{\theta_0}(\theta_0 + t\theta) \right|_{t=0}.$$

Pour tout $h = (h_1, h_2, h_3(\cdot)) \in H_p$ ($p < \infty$),

$$\begin{aligned} & \frac{\partial}{\partial t} S_{\theta_0}(\theta_0 + t\theta)(h) \\ &= \frac{\partial}{\partial t} E_{\theta_0}[s_{\theta_0}(Y, \theta_0 + t\theta)(h)] \\ &= \frac{\partial}{\partial t} E_{\theta_0} \left[\Delta h_3(X) - \int_0^X h_3(u) e^{[\beta_0 + t\beta]'W(u)} [d\Lambda_{T,0} + td\Lambda_T](u) \right. \\ & \quad \left. - \int_0^X h'_2 W(u) e^{[\beta_0 + t\beta]'W(u)} [d\Lambda_{T,0}(u) + td\Lambda_T](u) + h'_2 \Delta W(X) \right. \\ & \quad \left. + h'_1 \frac{\partial}{\partial \alpha} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha_0 + t\alpha) \right] \\ &= E_{\theta_0} \left[- \int_0^X [h_3(u) + h'_2 W(u)] \beta' W(u) e^{[\beta_0 + t\beta]'W(u)} [d\Lambda_{T,0} + td\Lambda_T](u) \right. \\ & \quad \left. - \int_0^X [h_3(u) + h'_2 W(u)] e^{[\beta_0 + t\beta]'W(u)} d\Lambda_T(u) \right. \\ & \quad \left. + h'_1 \frac{\partial^2}{\partial \alpha \partial t} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha_0 + t\alpha) \right]. \end{aligned}$$

Pour $t = 0$, nous en déduisons

$$\begin{aligned} D_{\theta}S_{\theta_0}(\theta_0)(h) &= -E_{\theta_0} \left[\int_0^X [h_3(u) + h'_2 W(u)] \beta' W(u) e^{\beta'_0 W(u)} d\Lambda_{T,0}(u) \right] \\ & \quad - E_{\theta_0} \left[\int_0^X [h_3(u) + h'_2 W(u)] e^{\beta'_0 W(u)} d\Lambda_T(u) \right] \\ & \quad + h'_1 E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha_0) \right] \alpha. \end{aligned}$$

Cette expression s'identifie à $-\int_0^{\tau} \sigma_{3,\theta_0}(h)(u) d\Lambda_T(u) - \beta' \sigma_{2,\theta_0}(h) - \alpha' \sigma_{1,\theta_0}(h)$, linéaire et continue en h .

Nous vérifions ensuite que $D_\theta S_{\theta_0}(\theta_0)$ satisfait à la définition 6.1 de la différentiabilité de Fréchet, c'est-à-dire que

$$\lim_{\epsilon \rightarrow 0} \frac{\|S_{\theta_0}(\theta_0 + \epsilon\theta) - S_{\theta_0}(\theta_0) - D_{\epsilon\theta}S_{\theta_0}(\theta_0)\|_p}{\epsilon} = 0,$$

uniformément en $\theta \in S$ pour tout $S \in \mathcal{P}_b(\Theta)$.

En utilisant un développement de Taylor de la fonction exponentielle au point $\beta'_0 W(u)$ dans $S_{\theta_0}(\theta_0 + \epsilon\theta)$, nous pouvons montrer que pour tout $h \in H_p$,

$$\begin{aligned} & \frac{S_{\theta_0}(\theta_0 + \epsilon\theta)(h) - S_{\theta_0}(\theta_0)(h) - D_{\epsilon\theta}S_{\theta_0}(\theta_0)(h)}{\epsilon} \\ & \approx -\epsilon E_{\theta_0} \left[\int_0^X [h_3(u) + h'_2 W(u)] \beta' W(u) e^{\beta'_0 W(u)} d\Lambda_T(u) \right]. \end{aligned}$$

Soit θ un élément quelconque dans S , pour $S \in \mathcal{P}_b(\Theta)$. Pour $h \in H_p$ et sous les conditions C1 à C6, $\epsilon E_{\theta_0} \left[\int_0^X [h_3(u) + h'_2 W(u)] \beta' W(u) e^{\beta'_0 W(u)} d\Lambda_T(u) \right] \rightarrow 0$ quand $\epsilon \rightarrow 0$.

L'application $\theta \mapsto S_{\theta_0}(\theta)$ est donc différentiable au sens de Fréchet au point θ_0 et sa différentielle en θ_0 , que nous noterons $\dot{S}_{\theta_0}(\theta_0)$, est égale à la différentielle de Gâteaux $D_\theta S_{\theta_0}(\theta_0)(h)$ donnée ci-dessus, d'où :

$$\dot{S}_{\theta_0}(\theta_0)(\theta)(h) = - \int_0^\tau \sigma_{3,\theta_0}(h)(u) d\Lambda_T(u) - \beta' \sigma_{2,\theta_0}(h) - \alpha' \sigma_{1,\theta_0}(h).$$

■

Nous nous intéressons maintenant à la distribution asymptotique de $\sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))$.

Lemme 6.3 *Pour tout p ($0 < p < \infty$), $\sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))$ converge en loi dans $l^\infty(H_p)$ vers un processus gaussien centré \mathcal{G} de fonction de covariance*

$$\text{cov}(\mathcal{G}(h), \mathcal{G}(h^*)) = \int_0^\tau h_3(u) \sigma_{3,\theta_0}(h^*)(u) d\Lambda_{T,0}(u) + h'_2 \sigma_{2,\theta_0}(h^*) + h'_1 \sigma_{1,\theta_0}(h^*).$$

Démonstration. Remarquons tout d'abord que $\sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))$ peut être réécrit de la manière suivante :

$$\begin{aligned} \sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[h'_{12} S_{\hat{\theta}_n,12}^{(i)}(\theta_0) + \delta_i h_3(x_i) \right. \\ & \quad \left. - \int_0^{x_i} h_3(u) E_{\hat{\theta}_n} [e^{\beta'_0 W(u)} | y_i] d\Lambda_{T,0}(u) \right]. \end{aligned}$$

Notons que $\{h'_{12}S_{\hat{\theta}_{n,12}}(\theta_0) : h_1 \in \mathbb{R}^p, |h_1| \leq p, h_2 \in \mathbb{R}^2, |h_2| \leq p\}$ est une classe de Donsker, que $\{\delta h_3(x) : h_3 \in VB_p\}$ est Donsker (l'ensemble des fonctions uniformément bornées et de variation uniformément bornée est une classe de Donsker) et que $\{\int_0^\tau h_3(u)E_{\hat{\theta}_n}[e^{\beta'_0 w(u)}|y]1_{\{u \leq x\}}d\Lambda_{T,0}(u) : h_3 \in VB_p\}$ est Donsker (le même raisonnement que dans la démonstration du lemme 6.1 s'applique). La classe $\{h'_{12}S_{\hat{\theta}_{n,12}}(\theta_0) + \delta h_3(x) + \int_0^\tau h_3(u)E_{\hat{\theta}_n}[e^{\beta'_0 w(u)}|y]1_{\{u \leq x\}}d\Lambda_{T,0}(u) : h \in H_p\}$ est Donsker.

D'où $\sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))$ converge en loi dans $l^\infty(H_p)$ vers un processus gaussien \mathcal{G} centré dont la fonction de covariance se calcule de la manière suivante.

$$\begin{aligned} \text{cov}(\mathcal{G}(h), \mathcal{G}(h^*)) &= -E_{\theta_0} \left[\frac{\partial^2}{\partial s \partial t} L_{\theta_0}(\theta_{0,s,t}) \Big|_{s=t=0} \right] \\ &= -E_{\theta_0} \left[\frac{\partial}{\partial s} s_{\theta_0}(Y, \theta_{0,s})(h) \Big|_{s=0} \right] \\ &= - \frac{\partial}{\partial s} S_{\theta_0}(\theta_{0,s})(h) \Big|_{s=0} \\ &= -\dot{S}_{\theta_0}(\theta_0)(h^*)(h). \end{aligned}$$

Soit $\theta_s = (\alpha_s, \beta_s, \Lambda_{T,s})$, où $\alpha_s = \alpha + sh_1^*$, $\beta_s = \beta + sh_2^*$ et $\Lambda_{T,s}(\cdot) = \int_0^\cdot (1 + sh_3^*(u)) d\Lambda_T(u)$.

$\partial s \tilde{s}(y, \theta_s)(h) / \partial s$ se calcule comme suit :

$$\begin{aligned} &\frac{\partial}{\partial s} s_{\tilde{\theta}}(y, \theta_s)(h) \\ &= \frac{\partial}{\partial s} \left[\frac{\partial}{\partial t} L_{\tilde{\theta}}(\theta_{s,t}) \Big|_{t=0} \right], \\ &= \frac{\partial}{\partial s} \left[\delta h_3(x) - E_{\tilde{\theta}} \left[\int_0^X h_3(u) e^{\beta'_s W(u)} (1 + sh_3^*(u)) d\Lambda_T(u) | y \right] \right. \\ &\quad \left. + h'_2 E_{\tilde{\theta}} \left[\Delta W(X) - \int_0^X W(u) e^{\beta'_s W(u)} (1 + sh_3^*(u)) d\Lambda_T(u) | y \right] \right. \\ &\quad \left. + h'_1 E_{\tilde{\theta}} \left[\frac{\partial}{\partial \alpha} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha + sh_1^*) | y \right] \right], \\ &= -E_{\tilde{\theta}} \left[\int_0^X h_3(u) e^{\beta'_s W(u)} [h_2^{*'} W(u) (1 + sh_3^*(u)) + h_3^*(u)] d\Lambda_T(u) | y \right] \\ &\quad - h'_2 E_{\tilde{\theta}} \left[\int_0^X W(u) e^{\beta'_s W(u)} [h_2^{*'} W(u) (1 + sh_3^*(u)) + h_3^*(u)] d\Lambda_T(u) | y \right] \\ &\quad + h'_1 E_{\tilde{\theta}} \left[\frac{\partial^2}{\partial s \partial \alpha} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha + sh_1^*) | y \right]. \end{aligned}$$

D'où

$$\begin{aligned} \left. \frac{\partial}{\partial s} s_{\tilde{\theta}}(y, \theta_s)(h) \right|_{s=0} &= -E_{\tilde{\theta}} \left[\int_0^X h_3(u) e^{\beta' W(u)} [h_2^{*'} W(u) + h_3^*(u)] d\Lambda_T(u) | y \right] \\ &\quad - h_2' E_{\tilde{\theta}} \left[\int_0^X W(u) e^{\beta' W(u)} [h_2^{*'} W(u) + h_3^*(u)] d\Lambda_T(u) | y \right] \\ &\quad + h_1' E_{\tilde{\theta}} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y \right] h_1^*. \end{aligned}$$

Finalement, $\text{cov}(\mathcal{G}(h), \mathcal{G}(h^*))$ s'identifie à

$$\text{cov}(\mathcal{G}(h), \mathcal{G}(h^*)) = \int_0^\tau h_3(u) \sigma_{3, \theta_0}(h^*)(u) d\Lambda_{T,0}(u) + h_2' \sigma_{2, \theta_0}(h^*) + h_1' \sigma_{1, \theta_0}(h^*).$$

Notons que

$$\begin{aligned} \text{var}(\mathcal{G}(h)) &= -\dot{S}_{\theta_0}(\theta_0)(h)(h) \\ &= \int_0^\tau h_3(u) \sigma_{3, \theta_0}(h)(u) d\Lambda_{T,0}(u) + h_2' \sigma_{2, \theta_0}(h) + h_1' \sigma_{1, \theta_0}(h). \end{aligned}$$

■

Nous allons maintenant montrer que la différentielle de Fréchet $\dot{S}_{\theta_0}(\theta_0)$ de l'application de Θ dans $l^\infty(H_p) : \theta \mapsto S_{\theta_0}(\theta)$ est inversible et que son inverse $\dot{S}_{\theta_0}(\theta_0)^{-1}$ est continue.

Nous montrons d'abord trois lemmes qui seront utiles dans la suite des démonstrations.

Lemme 6.4 *Soit $0 < p < \infty$ et soit $\theta \in l^\infty(H_p)$. Alors*

$$p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda_T)) \leq \|\theta\|_p \leq 3p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda_T)).$$

Démonstration. Soit $p < \infty$ et $\theta = (\alpha, \beta, \Lambda_T) \in l^\infty(H_p)$. Rappelons que

$$\|\theta\|_p = \sup_{h \in H_p} |\theta(h)| = \sup_{h \in H_p} \left| h_1' \alpha + h_2' \beta + \int_0^\tau h_3(u) d\Lambda_T(u) \right|.$$

Soit $h \in H_p$. Notons que $|h_1' \alpha| \leq |h_1| |\alpha|$. Puisque $h \in H_p$, alors $\|h\|_H = |h_1| + |h_2| + \|h_3\|_\nu \leq p$, d'où $|h_1' \alpha| \leq p|\alpha|$. De même $|h_2' \beta| \leq p|\beta|$.

Remarquons ensuite que

$$\begin{aligned} \left| \int_0^\tau h_3(u) d\Lambda_T(u) \right| &\leq \int_0^\tau \sup_{u \in [0, \tau]} |h_3(u)| d\Lambda_T(u) \\ &\leq \|h_3\|_\infty V_{[0, \tau]}(\Lambda_T). \end{aligned}$$

Notons que $\|h_3\|_\infty \leq V_{[0, \tau]}(h_3) + |h_3(0)| = \|h_3\|_\nu$. D'où $\|h_3\|_\infty \leq p$ et finalement $\left| \int_0^\tau h_3(u) d\Lambda_T(u) \right| \leq p V_{[0, \tau]}(\Lambda_T)$.

Nous obtenons la majoration suivante de $|h'_1 \alpha + h'_2 \beta + \int_0^\tau h_3(u) d\Lambda_T(u)|$ (pour tout $h \in H_p$) :

$$\begin{aligned} \left| h'_1 \alpha + h'_2 \beta + \int_0^\tau h_3(u) d\Lambda_T(u) \right| &\leq |h'_1 \alpha| + |h'_2 \beta| + \left| \int_0^\tau h_3(u) d\Lambda_T(u) \right| \\ &\leq p(|\alpha| + |\beta| + V_{[0, \tau]}(\Lambda_T)) \\ &\leq 3p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda_T)). \end{aligned}$$

Finalement, nous obtenons : $\|\theta\|_p \leq 3p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda))$.

Soit

$$h^{(1)} = \left(p \frac{\alpha}{|\alpha|}, 0, 0 \right), \quad h^{(2)} = \left(0, p \frac{\beta}{|\beta|}, 0 \right), \quad h^{(3)} = (0, 0, p).$$

Nous remarquons que $\|h^{(1)}\|_H = \|h^{(2)}\|_H = \|h^{(3)}\|_H = p$, donc $h^{(1)}, h^{(2)}$ et $h^{(3)}$ appartiennent à H_p .

Nous pouvons écrire : $|\theta(h^{(i)})| \leq \sup_{h \in H_p} |\theta(h)|$ ($i = 1, 2, 3$), d'où

$$|\theta(h^{(1)})| \vee |\theta(h^{(2)})| \vee |\theta(h^{(3)})| \leq \|\theta\|_p.$$

Remarquons que

$$\begin{aligned} |\theta(h^{(1)})| \vee |\theta(h^{(2)})| \vee |\theta(h^{(3)})| &= |h_1^{(1)T} \alpha| \vee |h_2^{(2)T} \beta| \vee \left| \int_0^\tau h_3^{(3)}(u) d\Lambda_T(u) \right| \\ &= p|\alpha| \vee p|\beta| \vee pV_{[0, \tau]}(\Lambda_T). \end{aligned}$$

Nous obtenons donc la majoration $p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda)) \leq \|\theta\|_p$.

Finalement,

$$p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda_T)) \leq \|\theta\|_p \leq 3p(|\alpha| \vee |\beta| \vee V_{[0, \tau]}(\Lambda_T)).$$

■

Lemme 6.5 *L'opérateur de H_∞ dans H_∞ $\sigma_{\theta_0} = (\sigma_{1,\theta_0}(h), \sigma_{2,\theta_0}(h), \sigma_{3,\theta_0}(h))$ est injectif.*

Démonstration. Pour montrer que σ_{θ_0} est injectif, nous montrons que si $\sigma_{\theta_0}(h) = 0$ ($h = (h_1, h_2, h_3) \in H_\infty$), alors $h = 0$; c'est-à-dire que $h_1 = 0$, $h_2 = 0$ et $h_3(u) = 0$ pour tout $u \in [0, \tau]$.

Si $\sigma_{\theta_0}(h) = 0$, alors $\sigma_{1,\theta_0}(h) = 0$, d'où $h'_1 \sigma_{1,\theta_0}(h) = 0$. Ceci s'écrit

$$h'_1 \left[-E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z; \alpha_0) \right] \right] h_1 = 0.$$

D'après l'hypothèse C6, $h_1 = 0$.

Si $\sigma_{\theta_0}(h) = 0$, alors

$$\int_0^\tau h_3(u) \sigma_{3,\theta_0}(h)(u) d\Lambda_{T,0}(u) + h'_2 \sigma_{2,\theta_0}(h) + h'_1 \sigma_{1,\theta_0}(h) = 0,$$

qui peut être réécrit : $E_{\theta_0} [(s_{\theta_0}(Y; \theta_0))^2] = 0$, d'où $s_{\theta_0}(y; \theta_0) = 0$ presque partout. Une décomposition de $s_{\theta_0}(Y; \theta_0)$ analogue à celle utilisée dans la démonstration du lemme 4.2 permet de montrer que $h_2 = 0$.

Si $\sigma_{\theta_0}(h) = 0$, alors $h_1 = 0$ et $h_2 = 0$. Soit $h = (0, 0, h_3)$. $\sigma_{3,\theta_0}(h)(u) = 0 \forall u \in [0, \tau]$ s'écrit

$$\sigma_{3,\theta_0}(h)(u) = E_{\theta_0} \left[h_3(u) e^{\beta'_0 W(u)} 1_{\{u \leq X\}} \right] = 0, \forall u \in [0, \tau],$$

ou, de manière équivalente

$$h_3(u) E_{\theta_0} \left[e^{\beta'_0 W(u)} 1_{\{u \leq X\}} \right] = 0, \forall u \in [0, \tau].$$

D'après l'hypothèse C6, il existe un q ($q > 0$) tel que $E_{\theta_0} [e^{\beta'_0 W(u)} 1_{\{u \leq X\}}] > q$ pour tout $u \in [0, \tau]$. Donc $h_3(u) = 0$ pour tout $u \in [0, \tau]$. L'opérateur σ_{θ_0} est donc injectif.

■

Lemme 6.6 *L'opérateur $\sigma_{\theta_0} : H_\infty \longrightarrow H_\infty$ est inversible et son opérateur inverse $\sigma_{\theta_0}^{-1}$ est continu.*

Démonstration. Pour montrer que σ_{θ_0} est inversible et que son inverse est continue, nous utilisons les résultats suivants.

Soit $(X, \|\cdot\|)$ un espace normé, $S : X \rightarrow X$ un opérateur linéaire borné inversible et d'inverse borné et $A : X \rightarrow X$ un opérateur linéaire compact. Kress [52] rappelle que si l'opérateur $S + A$ est injectif, alors il est inversible.

D'autre part, Luenberger [59] rappelle que si $T : X \rightarrow X$ est un opérateur linéaire borné sur un Banach X , et que T est inversible, alors l'inverse T^{-1} est continue.

Rudin [76] rappelle qu'un opérateur $T : X \rightarrow X$ est compact si et seulement si toute suite bornée (x_n) dans X contient une sous-suite (x_{n_i}) telle que $T(x_{n_i})$ converge dans X .

σ_{θ_0} est un opérateur linéaire borné et injectif sur H_∞ . H_∞ est un espace de Banach [66]. Nous montrons que σ_{θ_0} est inversible.

Pour cela, nous montrons que σ_{θ_0} peut se décomposer comme la somme $S + A$ d'un opérateur linéaire borné inversible d'inverse borné et d'un opérateur linéaire compact.

Soit l'opérateur $\Sigma : H_\infty \rightarrow H_\infty$ défini par

$$\begin{aligned} \Sigma(h) = & \left(-E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha_0) \right] h_1, \right. \\ & E_{\theta_0} \left[\int_0^X W(u) W(u)' e^{\beta'_0 W(u)} d\Lambda_{T,0}(u) \right] h_2, \\ & \left. E_{\theta_0} \left[e^{\beta'_0 W(u)} 1_{\{u \leq X\}} \right] h_3(u) \right). \end{aligned}$$

Son opérateur inverse

$$\begin{aligned} \Sigma(h)^{-1} = & \left(-E_{\theta_0} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \alpha_0) \right]^{-1} h_1, \right. \\ & E_{\theta_0} \left[\int_0^X W(u) W(u)' e^{\beta'_0 W(u)} d\Lambda_{T,0}(u) \right]^{-1} h_2, \\ & \left. E_{\theta_0} \left[e^{\beta'_0 W(u)} 1_{\{u \leq X\}} \right]^{-1} h_3(u) \right). \end{aligned}$$

est linéaire et borné d'après les hypothèses C1 à C6. Nous montrons que l'opérateur linéaire $\sigma_{\theta_0} - \Sigma$ est compact.

Soit $(h_n)_{n \in \mathbb{N}} = (h_{n1}, h_{n2}, h_{n3})_{n \in \mathbb{N}}$ une suite d'éléments de H_∞ . Nous devons montrer qu'il existe une sous-suite convergente de $\sigma_{\theta_0}(h_n) - \Sigma(h_n)$.

$(h_{3n})_{n \in \mathbb{N}}$ est à variation bornée donc nous pouvons l'écrire comme la différence de deux fonctions bornées et croissantes $(h_{3n}^{(1)})_{n \in \mathbb{N}}$ et $(h_{3n}^{(2)})_{n \in \mathbb{N}}$. D'après le théorème de Helly, il existe une sous-suite $(h_{3\phi(n)}^{(1)})_{n \in \mathbb{N}}$ de $(h_{3n}^{(1)})_{n \in \mathbb{N}}$ qui converge simplement vers une limite $h_3^{(1)*}$. Il existe alors une sous-suite $(h_{3\eta(\phi(n))}^{(2)})_{n \in \mathbb{N}}$ de $(h_{3\phi(n)}^{(2)})_{n \in \mathbb{N}}$ qui converge vers une limite $h_3^{(2)*}$. La sous-suite $(h_{3\eta(\phi(n))}^{(1)}, h_{3\eta(\phi(n))}^{(2)})_{n \in \mathbb{N}}$ converge donc vers $(h_3^{(2)*}, h_3^{(2)*})$.

En utilisant le même argument avec le théorème de Bolzano-Weierstrass, nous pouvons finalement trouver une sous-suite de $(h_n)_{n \in \mathbb{N}}$ (pour simplifier les notations, nous notons $(h_{g(n)})_{n \in \mathbb{N}}$ cette sous-suite) qui converge vers une limite $h^* = (h_1^*, h_2^*, h_3^*)$. Nous voulons montrer que $\sigma_{\theta_0}(h_{g(n)}) - \Sigma(h_{g(n)})$ converge vers $\sigma_{\theta_0}(h^*) - \Sigma(h^*)$ dans H_∞ . Notons que l'opérateur $\sigma_{\theta_0} - \Sigma$ s'écrit

$$\left(0, E_{\theta_0} \left[\int_0^X W(u) e^{\beta'_0 W(u)} h_3(u) d\Lambda_{T,0}(u) \right], E_{\theta_0} \left[h'_2 W(u) e^{\beta'_0 W(u)} 1_{\{u \leq X\}} \right] \right).$$

Nous pouvons donc calculer $\|\sigma_{\theta_0}(h_{g(n)}) - \Sigma(h_{g(n)}) - \sigma_{\theta_0}(h^*) + \Sigma(h^*)\|_H$ comme suit :

$$\begin{aligned}
 & \left\| E_{\theta_0} \left[(h_{2g(n)} - h_2^*)' W(u) e^{\beta'_0 W(u)} 1_{\{u \leq X\}} \right] \right\|_\nu \\
 & + \left\| E_{\theta_0} \left[\int_0^X W(u) e^{\beta'_0 W(u)} (h_{3g(n)}(u) - h_3^*(u)) d\Lambda_{T,0}(u) \right] \right\|.
 \end{aligned}$$

Sous les hypothèses C1 à C6, cette expression est majorée par

$$\sqrt{2}[ce^{2bc} + K]|h_{2g(n)} - h_2^*| + \sqrt{2}ce^{2bc} \int_0^\tau |h_{3g(n)}(u) - h_3^*(u)| d\Lambda_{T,0}(u). \quad (6.5)$$

Lorsque $h_{g(n)}$ converge vers h^* le premier terme de (6.5) converge vers 0 par continuité de la norme, le second tend vers 0 en appliquant le théorème de convergence dominé.

Finalement, $\|\sigma_{\theta_0}(h_{g(n)}) - \Sigma(h_{g(n)}) - \sigma_{\theta_0}(h^*) + \Sigma(h^*)\|_H \rightarrow 0$ lorsque $h_{g(n)} \rightarrow h^*$ dans H_∞ et $\sigma_{\theta_0} - \Sigma$ est un opérateur compact.

σ_{θ_0} est donc inversible et l'opérateur $\sigma_{\theta_0}^{-1}$ est continu.

■

Nous montrons maintenant que la différentielle de Fréchet $\dot{S}_{\theta_0}(\theta_0)$ de $\theta \mapsto S_{\theta_0}(\theta)$ est inversible et que son inverse est continue.

Lemme 6.7 *Pour tout p ($0 < p < \infty$), l'opérateur $\dot{S}_{\theta_0}(\theta_0) : \Theta \rightarrow l^\infty(H_p)$ (pour $\Theta \subset H_p$) est inversible et son inverse est continue sur l'image de $\dot{S}_{\theta_0}(\theta_0)$.*

Démonstration. Bickel, Klaassen, Ritov et Wellner [11] rappellent qu'un opérateur linéaire $A : X \longrightarrow Y$ est inversible et que son inverse A^{-1} est continue de l'image de A dans A si et seulement si il existe c ($c > 0$) tel que

$$\inf_{x \in X} \frac{\|Ax\|_Y}{\|x\|_X} > c.$$

L'opérateur $\dot{S}_{\theta_0}(\theta_0) : \Theta \longrightarrow l^\infty(H_p)$ est inversible et son inverse est continue si et seulement si il existe c ($c > 0$) tel que

$$\inf_{\theta \in \Theta} \frac{\|\dot{S}_{\theta_0}(\theta_0)(\theta)\|_p}{\|\theta\|_p} > c. \quad (6.6)$$

Nous avons montré que σ_{θ_0} est inversible et que l'opérateur $\sigma_{\theta_0}^{-1}$ est continu, donc pour tout p ($0 < p < \infty$), il existe q ($0 < q < \infty$) tel que $\sigma_{\theta_0}^{-1}(H_q) \subset H_p$. Donc (6.6) est minoré par

$$\begin{aligned} & \inf_{\theta \in \Theta} \frac{\sup_{h \in \sigma_{\theta_0}^{-1}(H_q)} \left| \dot{S}_{\theta_0}(\theta_0)(\theta)(h) \right|}{\|\theta\|_p} \\ &= \inf_{\theta \in \Theta} \frac{\sup_{h \in H_q} \left| \dot{S}_{\theta_0}(\theta_0)(\theta)(\sigma_{\theta_0}^{-1}(h)) \right|}{\|\theta\|_p} \\ &= \inf_{\theta \in \Theta} \frac{\sup_{h \in H_q} \left| \int_0^\tau \sigma_{3,\theta_0}(\sigma_{\theta_0}^{-1}(h))(u) d\Lambda_T(u) + \beta' \sigma_{2,\theta_0}(\sigma_{\theta_0}^{-1}(h)) + \alpha' \sigma_{1,\theta_0}(\sigma_{\theta_0}^{-1}(h)) \right|}{\|\theta\|_p}. \end{aligned}$$

Rappelons que σ_{θ_0} est inversible, donc $\sigma_{\theta_0}(\sigma_{\theta_0}^{-1}(h)) = h$ pour tout $h = (h_1, h_2, h_3) \in H_q$, d'où $\sigma_{i,\theta_0}(\sigma_{\theta_0}^{-1}(h)) = h_i$, $i = 1, 2, 3$.

Le minorant précédent est donc égal à

$$\inf_{\theta \in \Theta} \frac{\sup_{h \in H_q} \left| \int_0^\tau h_3(u) d\Lambda_T(u) + \beta' h_2 + \alpha' h_1 \right|}{\|\theta\|_p} = \inf_{\theta \in \Theta} \frac{\|\theta\|_q}{\|\theta\|_p}. \quad (6.7)$$

D'après le lemme 6.4, $\|\theta\|_q$ est minoré par $q(|\alpha| \vee |\beta| \vee V_{[0,\tau]}(\Lambda_T))$ et $\|\theta\|_p$ est majoré par $3p(|\alpha| \vee |\beta| \vee V_{[0,\tau]}(\Lambda_T))$. Finalement, (6.7) est minoré par $\frac{q}{3p}$. L'opérateur $\dot{S}_{\theta_0}(\theta_0)$ est donc inversible et son inverse $\dot{S}_{\theta_0}(\theta_0)^{-1}$ est continue sur l'image de $\dot{S}_{\theta_0}(\theta_0)$.

■

D'après le théorème 6.4, pour tout p ($0 < p < \infty$),

$$-\dot{S}_{\theta_0}(\theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0)(h) = \sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))(h) + o_P(1),$$

où, d'après le lemme 6.2,

$$\begin{aligned} -\dot{S}_{\theta_0}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)(h) &= \int_0^\tau \sigma_{3,\theta_0}(h)(u)\sqrt{n}d(\hat{\Lambda}_{T,n} - \Lambda_{T,0})(u) \\ &\quad + \sqrt{n}(\hat{\beta}_n - \beta_0)'\sigma_{2,\theta_0}(h) + \sqrt{n}(\hat{\alpha}_n - \alpha_0)'\sigma_{1,\theta_0}(h). \end{aligned}$$

D'après les lemmes 6.3 et 6.7, $\sqrt{n}(\hat{\theta}_n - \theta_0)(h)$ converge en loi vers $-\dot{S}_{\theta_0}(\theta_0)^{-1}\mathcal{G}(h)$. Nous voulons identifier la limite $-\dot{S}_{\theta_0}(\theta_0)^{-1}\mathcal{G}(h)$.

L'opérateur σ_{θ_0} est continuellement inversible donc pour tout q ($0 < q < \infty$), il existe p ($0 < p < \infty$) tel que $\sigma_{\theta_0}^{-1}(g) \in H_q$ si $g \in H_p$. Posons $h = \sigma_{\theta_0}^{-1}(g)$.

Avec ces notations, $-\dot{S}_{\theta_0}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)(h)$ s'écrit sous les formes suivantes :

$$\begin{aligned} -\dot{S}_{\theta_0}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)(h) &= \int_0^\tau g_3(u)\sqrt{n}d(\hat{\Lambda}_{t,n} - \Lambda_{T,0})(u) + \sqrt{n}(\hat{\beta}_n - \beta_0)'g_2 \\ &\quad + \sqrt{n}(\hat{\alpha}_n - \alpha_0)'g_1, \end{aligned} \tag{6.8}$$

et

$$-\dot{S}_{\theta_0}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)(h) = \sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))(\sigma_{\theta_0}^{-1}(g)) + o_P(1). \tag{6.9}$$

Remarquons que le membre droit de l'égalité (6.8) est égal à

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))(g),$$

qui converge en loi vers $-\dot{S}_{\theta_0}(\theta_0)^{-1}\mathcal{G}(g)$. Remarquons également que le membre droit de (6.9) converge en loi vers $\mathcal{G}(\sigma_{\theta_0}^{-1}(g))$. D'où $-\dot{S}_{\theta_0}(\theta_0)^{-1}\mathcal{G} = \mathcal{G}(\sigma_{\theta_0}^{-1})$.

Finalement,

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0})),$$

converge en loi dans $l^\infty(H_p)$ vers un processus gaussien centré G de fonction de covariance

$$\text{cov}[G(g), G(g^*)] = \int_0^\tau g_3(u)\sigma_{3,\theta_0}^{-1}(g^*)(u)d\Lambda_{T,0}(u) + \sigma_{2,\theta_0}^{-1}(g^*)'g_2 + \sigma_{1,\theta_0}^{-1}(g^*)'g_1.$$

■

6.4 Estimation de la variance asymptotique des estimateurs

Nous nous intéressons maintenant au problème de l'estimation de la variance asymptotique des estimateurs semiparamétriques des paramètres du modèle conjoint, qui intervient dans la construction d'intervalles de confiance et de tests d'hypothèses.

Rappelons que

$$\begin{aligned} & (\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))(h) \\ &= h'_1 \sqrt{n}(\hat{\alpha}_n - \alpha_0) + h'_2 \sqrt{n}(\hat{\beta}_n - \beta_0) + \int_0^\tau h_3(u) d\sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0})(u). \end{aligned}$$

D'après le théorème 6.5, la variance asymptotique de

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))(h)$$

est

$$\int_0^\tau h_3(u) \sigma_{3,\theta_0}^{-1}(h)(u) d\Lambda_{T,0}(u) + \sigma_{2,\theta_0}^{-1}(h)' h_2 + \sigma_{1,\theta_0}^{-1}(h)' h_1. \quad (6.10)$$

Il est assez naturel de proposer d'estimer la variance (6.10) par :

$$\int_0^\tau h_3(u) \sigma_{3,\hat{\theta}_n}^{-1}(h)(u) d\hat{\Lambda}_{T,n}(u) + \sigma_{2,\hat{\theta}_n}^{-1}(h)' h_2 + \sigma_{1,\hat{\theta}_n}^{-1}(h)' h_1, \quad (6.11)$$

où $\sigma_{1,\hat{\theta}_n}$, $\sigma_{2,\hat{\theta}_n}$ et $\sigma_{3,\hat{\theta}_n}$ sont obtenus en remplaçant $\theta_0 = (\alpha_0, \beta_0, \Lambda_{T,0})$ par $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{T,n})$ dans les σ_{1,θ_0} , σ_{2,θ_0} , σ_{3,θ_0} donnés dans le théorème 6.5.

Nous illustrons cette proposition et ses limites sur l'exemple de l'estimation de la variance asymptotique de $\sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})$.

Posons $h_{\beta_1} = (0, (1, 0)', 0)$. Il vient

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))(h_{\beta_1}) = \sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0}).$$

La variance asymptotique de $\sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})$ est alors estimée par l'estimateur (6.11) dans lequel $h = h_{\beta_1}$, c'est-à-dire par :

$$\begin{aligned} & (1, 0) \sigma_{2,\hat{\theta}_n}^{-1}(h_{\beta_1}) \\ &= (1, 0) E_{\hat{\theta}_n} \left[\int_0^X W(u) W(u)' e^{\hat{\beta}_n' W(u)} d\hat{\Lambda}_{T,n}(u) \right]^{-1} (1, 0)' \\ &= (1, 0) E_{\hat{\theta}_n} \left[\sum_{k=1}^{p(n)} W(x_{(k)}) W(x_{(k)})' 1_{\{x_{(k)} \leq X\}} e^{\hat{\beta}_n' W(x_{(k)})} \Delta \hat{\Lambda}_{n,k} \right]^{-1} (1, 0)'. \end{aligned}$$

Il apparaît néanmoins que les calculs d'espérances requis par cette solution impliquent des intégrales multiples difficiles à calculer. Approcher ces espérances par des moyennes empiriques, et estimer par exemple la variance asymptotique de $\sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})$ par

$$(1, 0) \left[\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{p(n)} w_i(x_{(k)}) w_i(x_{(k)})' 1_{\{x_{(k)} \leq x_i\}} e^{\hat{\beta}_n' w_i(x_{(k)})} \Delta \hat{\Lambda}_{n,k} \right]^{-1} (1, 0)'.$$

n'est pas possible car les $Z_i(X_i)$ nécessaires aux calculs de ces moyennes ne sont pas observés.

Nous proposons donc, pour des raisons théoriques et pratiques que nous expliquons ci-dessous, d'estimer la variance (6.10) à partir des formules (6.4). Nous estimons tout d'abord $\sigma_{\theta_0} = (\sigma_{1,\theta_0}, \sigma_{2,\theta_0}, \sigma_{3,\theta_0})$ par $\hat{\sigma}_{\hat{\theta}_n} = (\hat{\sigma}_{1,\hat{\theta}_n}, \hat{\sigma}_{2,\hat{\theta}_n}, \hat{\sigma}_{3,\hat{\theta}_n})$, où

$$\begin{aligned} \hat{\sigma}_{1,\hat{\theta}_n}(h) &= -\frac{1}{n} \sum_{i=1}^n E_{\hat{\theta}_n} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \hat{\alpha}_n) | y_i \right] h_1, \\ \hat{\sigma}_{2,\hat{\theta}_n}(h) &= \frac{1}{n} \sum_{i=1}^n E_{\hat{\theta}_n} \left[\int_0^X W(u) e^{\hat{\beta}_n' W(u)} [h_2' W(u) + h_3(u)] d\hat{\Lambda}_{T,n}(u) | y_i \right], \\ \hat{\sigma}_{3,\hat{\theta}_n}(h)(u) &= \frac{1}{n} \sum_{i=1}^n E_{\hat{\theta}_n} \left[[h_2' W(u) + h_3(u)] e^{\hat{\beta}_n' W(u)} 1_{\{u \leq X\}} | y_i \right]. \end{aligned}$$

Puis nous proposons d'estimer la variance asymptotique (6.10) par

$$\int_0^\tau h_3(u) \hat{\sigma}_{3,\hat{\theta}_n}^{-1}(h)(u) d\hat{\Lambda}_{T,n}(u) + \hat{\sigma}_{2,\hat{\theta}_n}^{-1}(h)' h_2 + \hat{\sigma}_{1,\hat{\theta}_n}^{-1}(h)' h_1. \quad (6.12)$$

En utilisant le principe de démonstration proposé par Murphy [66], nous montrons tout d'abord que l'estimateur (6.12) est un estimateur convergent de la variance (6.10). Puis nous expliquons comment (6.12) peut être utilisé en pratique pour obtenir des estimations des variances asymptotiques des estimateurs des paramètres du modèle conjoint.

Utilisant les mêmes raisonnements que dans la démonstration du lemme 6.1, les fonctions sous le signe \sum dans $\hat{\sigma}_{1,\hat{\theta}_n}, \hat{\sigma}_{2,\hat{\theta}_n}, \hat{\sigma}_{3,\hat{\theta}_n}$ forment une classe de Donsker (pour $h \in H_p, p > 0$) d'où $\sup_{h \in H_p} \|\hat{\sigma}_{\hat{\theta}_n}(h) - \sigma_{\theta_0}(h)\|_H \rightarrow 0$. $\hat{\sigma}_{\hat{\theta}_n}$ est inversible et son inverse continue, donc pour tout $H_p \subset H_\infty$, il existe un

$H_q \subset H_\infty$ tel que $\hat{\sigma}_{\hat{\theta}_n}^{-1}(H_q) \subset H_p$, et pour tout $g \in H_q$, il existe $h \in H_p$ tel que $h = \hat{\sigma}_{\hat{\theta}_n}^{-1}(g)$.

Alors

$$\begin{aligned} \|\hat{\sigma}_{\hat{\theta}_n}^{-1}(g) - \sigma_{\theta_0}^{-1}(g)\|_H &= \|\sigma_{\theta_0}^{-1}(\sigma_{\theta_0}(h)) - \sigma_{\theta_0}^{-1}(\hat{\sigma}_{\hat{\theta}_n}(h))\|_H \\ &\leq \sup_{h \in H_q} \frac{\|\sigma_{\theta_0}^{-1}(h)\|_H}{\|h\|_H} \|\sigma_{\theta_0}(h) - \hat{\sigma}_{\hat{\theta}_n}(h)\|_H \\ &\leq \sup_{h \in H_q} \frac{\|\sigma_{\theta_0}^{-1}(h)\|_H}{\|h\|_H} \sup_{h \in H_p} \|\sigma_{\theta_0}(h) - \hat{\sigma}_{\hat{\theta}_n}(h)\|_H. \end{aligned}$$

D'où $\sup_{g \in H_q} \|\hat{\sigma}_{\hat{\theta}_n}^{-1}(g) - \sigma_{\theta_0}^{-1}(g)\|_H \rightarrow 0$. Il s'ensuit que l'estimateur (6.12) donné par $\int_0^\tau h_3(u) \hat{\sigma}_{3, \hat{\theta}_n}^{-1}(h)(u) d\hat{\Lambda}_{T,n}(u) + \hat{\sigma}_{2, \hat{\theta}_n}^{-1}(h)' h_2 + \hat{\sigma}_{1, \hat{\theta}_n}^{-1}(h)' h_1$ converge vers la variance asymptotique $\int_0^\tau h_3(u) \sigma_{3, \theta_0}^{-1}(h)(u) d\Lambda_{T,0}(u) + \sigma_{2, \theta_0}^{-1}(h)' h_2 + \sigma_{1, \theta_0}^{-1}(h)' h_1$.

En particulier, la matrice de covariance asymptotique de l'estimateur semi-paramétrique du paramètre d'intérêt β du modèle conjoint peut être estimée de manière convergente par (6.12) pour un choix approprié de h .

Reprenons par exemple $h_{\beta_1} = (0, (1, 0)', 0)$. La variance asymptotique de $\sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})$ est estimée par l'estimateur (6.12) dans lequel $h = h_{\beta_1}$, c'est-à-dire par :

$$\begin{aligned} &(1, 0) \hat{\sigma}_{2, \hat{\theta}_n}^{-1}(h_{\beta_1}) \\ &= (1, 0) \left[\frac{1}{n} \sum_{i=1}^n \int_0^{x_i} E_{\hat{\theta}_n} \left[W(u) W(u)' e^{\hat{\beta}_n' W(u)} | y_i \right] d\hat{\Lambda}_{T,n} \right]^{-1} (1, 0)' \\ &= (1, 0) \left[\frac{1}{n} \sum_{i=1}^n \sum_{k=1: x_{(k)} \leq x_i}^{p(n)} E_{\hat{\theta}_n} \left[W(x_{(k)}) W(x_{(k)})' e^{\hat{\beta}_n' W(x_{(k)})} | y_i \right] \Delta \hat{\Lambda}_{n,k} \right]^{-1} (1, 0)'. \end{aligned}$$

De même, la variance asymptotique de $\sqrt{n}(\hat{\beta}_{2,n} - \beta_{2,0})$ sera estimée par

$$(0, 1) \left[\frac{1}{n} \sum_{i=1}^n \sum_{k=1: x_{(k)} \leq x_i}^{p(n)} E_{\hat{\theta}_n} \left[W(x_{(k)}) W(x_{(k)})' e^{\hat{\beta}_n' W(x_{(k)})} | y_i \right] \Delta \hat{\Lambda}_{n,k} \right]^{-1} (0, 1)'.$$

Enfin, il pourra également être utile, pour certaines applications, d'estimer la variance asymptotique de l'estimateur du paramètre α . La variance asymptotique de $\sqrt{n}(\hat{\alpha}_{l,n} - \alpha_{l,0})$, où $\hat{\alpha}_{l,n}$ (respectivement $\alpha_{l,0}$) désigne la l -ième composante de l'estimateur $\hat{\alpha}_n$ (respectivement de α_0) peut être estimée par

$$J_l' \left[-\frac{1}{n} \sum_{i=1}^n E_{\hat{\theta}_n} \left[\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f(Z_0, \dots, Z_{a_X}, Z; \hat{\alpha}_n) | y_i \right] \right] J_l,$$

où J_l est le p -vecteur $(0, \dots, 1, 0, \dots)'$ dont toutes les composantes sont nulles sauf la l -ième, qui est égale à 1. Des versions spécifiques de cette dernière formule seront obtenues dans le chapitre 7, pour une distribution de la covariable $Z(\cdot)$ particulière, choisie pour les applications du modèle conjoint.

Une méthode d'approximation numérique sera utilisée pour approcher les intégrales $E_{\hat{\theta}_n}$. Elle est décrite en Annexe A.

Chapitre 7

Mise en oeuvre de l'algorithme EM dans le modèle de Cox avec covariable manquante

Dans ce chapitre, nous nous intéressons à la détermination des estimations semiparamétriques des paramètres du modèle conjoint proposé. Ces paramètres sont le paramètre d'intérêt β (paramètre de régression du modèle marginal de Cox), la fonction de risque $\Lambda_T = \int_0^t \lambda_T(s) ds$, et le paramètre α de la loi du processus de covariable. Nous montrons comment obtenir les estimations semiparamétriques de β , Λ_T et α à l'aide de l'algorithme itératif Espérance-Maximisation (EM) [26]. La mise en oeuvre de cet algorithme nécessitera d'approcher numériquement des intégrales (nous utiliserons une quadrature de Gauss-Hermite). A chaque itération de l'EM, les estimations de β seront calculées à l'aide d'un algorithme de Newton-Raphson.

Nous décrivons tout d'abord l'algorithme EM utilisé pour l'estimation semiparamétrique dans le modèle de Cox avec covariable dépendant du temps manquante. Puis, nous présentons deux applications du modèle conjoint sur des données réelles. Nous proposons tout d'abord d'appliquer ce modèle à l'analyse statistique des abandons informatifs dans les études longitudinales, et nous comparons ses résultats à ceux obtenus par le modèle de Diggle et Kenward [28]. Nous comparons ensuite les estimations du paramètre de régression du modèle de Cox obtenues avec le modèle conjoint et deux méthodes d'imputation d'une covariable dépendant du temps.

7.1 Description de l'algorithme EM pour le modèle conjoint

Nous rappelons brièvement le principe de l'algorithme EM [26] puis nous l'appliquons à la recherche des estimations semiparamétriques dans le modèle conjoint.

Principe de l'algorithme EM

Supposons que l'on observe incomplètement un vecteur aléatoire X de loi admettant une densité $f_X(x; \phi)$ ($\phi \in \Phi \subseteq \mathbb{R}^p$) sur \mathbb{R}^d ($d \geq 2$). X représente la donnée complète qui n'est pas entièrement observée. Notons Y la donnée incomplète réellement observée.

Soit $f_Y(y; \phi)$ la densité de Y , et soit $f_{X|Y}(x; \phi)$ la densité conditionnelle de X sachant $Y = y$. Notons ϕ_0 la "vraie" valeur du paramètre ϕ .

Si X était observable, l'estimateur du maximum de vraisemblance de ϕ_0 serait obtenu comme solution du problème de maximisation de $\ln f_X(x; \phi)$. Dans le cas où X n'est pas complètement observé, l'algorithme EM consiste à maximiser à la place l'espérance conditionnelle de la log-vraisemblance complète $\ln f_X(x; \phi)$ sachant les observations $Y = y$. Plus précisément, on définit la fonction de deux arguments ϕ et $\tilde{\phi}$:

$$\begin{aligned} Q(\phi, \tilde{\phi}) &= E_{\tilde{\phi}} [\ln f_X(X; \phi) | y] \\ &= \int \ln f_X(x; \phi) f_{X|Y}(x; \tilde{\phi}) dx. \end{aligned}$$

L'espérance conditionnelle est calculée avec une valeur $\tilde{\phi}$ du paramètre qui peut être différente de la valeur ϕ , ce qui conduit à envisager un algorithme itératif qui, partant d'une valeur initiale $\phi^{(0)}$, génère une suite $\phi^{(m)}$ telle que $\phi^{(m+1)} = \operatorname{argmax}_{\phi \in \Phi} Q(\phi, \phi^{(m)})$, où $\phi^{(m)}$ désigne la m^e valeur itérée et $\phi^{(m+1)}$ la valeur obtenue à l'itération suivante. Chaque itération de l'algorithme comporte deux étapes, un calcul d'espérance et une maximisation :

1. Etape E : calcul de l'espérance mathématique de la log-vraisemblance (fonction d'une valeur courante du paramètre) des données complètes conditionnellement à l'observation : $Q(\phi, \phi^{(m)})$.
2. Etape M : recherche de $\phi^{(m+1)}$ comme solution de $\operatorname{argmax}_{\phi \in \Phi} Q(\phi, \phi^{(m)})$.

Les étapes E et M sont alternativement répétées jusqu'à ce que la différence $\ln f_Y(y; \theta^{(m+1)}) - \ln f_Y(y; \theta^{(m)})$ soit suffisamment petite.

Le lecteur intéressé par l'algorithme EM pourra se référer aux traitements détaillés de Dempster, Laird et Rubin [26] et McLachlan et Krishnan [62].

Algorithme EM dans le modèle conjoint

Nous avons introduit au chapitre 5 la log-vraisemblance EM $E_{\tilde{\theta}}[\ln l(Y, Z; \theta)|y]$, où $E_{\tilde{\theta}}[\ln l(Y, Z; \theta)|y]$ représente l'espérance conditionnelle de $\ln l(Y, Z; \theta)$ sachant $Y = y$ (i.e. de la log-vraisemblance des données complètes (Y, Z) sachant l'observation des données incomplètes y).

Dans notre problème, la fonction $Q(\theta, \tilde{\theta})$ est

$$\begin{aligned} Q(\theta, \tilde{\theta}) &= E_{\tilde{\theta}}[\ln l(Y, Z; \theta)|y] \\ &= E_{\tilde{\theta}} \left[\Delta \beta' W(X) - \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} e^{\beta' W(x_{(k)})} 1_{\{x_{(k)} \leq X\}} | y \right] \\ &\quad + \sum_{k=1}^{p(n)} 1_{\{X=x_{(k)}\}} \Delta \ln \Delta \Lambda_{n,k} + E_{\theta^{(m)}} [\ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y]. \end{aligned}$$

A la $(m+1)^e$ itération, nous sommes conduits à maximiser $L_{n,\tilde{\theta}}(\theta) = \sum_{i=1}^n Q_i(\theta, \tilde{\theta})$ par rapport à $\theta = (\alpha, \beta, \Delta \Lambda_{n,1}, \dots, \Delta \Lambda_{n,p(n)})$ (où $p(n)$ désigne le nombre d'instantants d'événement distincts).

Nous supposons un modèle de la forme $Z_j = \mu Z_{j-1} + \epsilon_j$ pour la covariable Z , où ϵ_j est un terme d'erreur centré et de variance supposée égale à une constante σ^2 . Pour simplifier les procédures numériques nécessaires aux calculs des estimations, nous supposons que l'erreur ϵ_j suit une loi normale. Le paramètre de la loi de Z est $\alpha = (\mu, \sigma^2)'$.

Nous obtenons d'abord les valeurs itérées $\mu^{(m+1)}, \sigma^{2(m+1)}, \Delta \Lambda_{n,1}^{(m+1)}, \dots, \Delta \Lambda_{n,p(n)}^{(m+1)}$ grâce aux formules explicites de calcul données par la proposition suivante.

Proposition 7.1 *Les valeurs $\mu^{(m+1)}, \sigma^{2(m+1)}$ et $\Delta \Lambda_{n,i}^{(m+1)}$ ($i = 1, \dots, p(n)$) obtenues à la $(m+1)^e$ itération de l'algorithme EM à partir des données incomplètes y_i ($i = 1, \dots, n$) sont données par*

$$\begin{aligned} \mu^{(m+1)} &= \frac{\sum_{i=1}^n [\sum_{j=1}^{a_{x_i}} z_{ij} z_{i,j-1} + z_{a_{x_i}} E_{\theta^{(m)}} [Z | y_i]]}{\sum_{i=1}^n \sum_{j=0}^{a_{x_i}} z_{ij}^2}, \\ \sigma^{2(m+1)} &= \frac{\sum_{i=1}^n [\sum_{j=1}^{a_{x_i}} (z_{ij} - \mu^{(m)} z_{i,j-1})^2 + E_{\theta^{(m)}} [(Z - \mu^{(m)} Z_{a_X})^2 | y_i]]}{\sum_{i=1}^n (a_{x_i} + 1)}, \end{aligned}$$

$$\Delta\Lambda_{n,i}^{(m+1)} = \frac{\sum_{j=1}^n \delta_j 1_{\{x_j=u_i\}}}{\sum_{j=1}^n E_{\theta^{(m)}} [e^{\beta^{(m)T} W(u_i)} 1_{\{u_i \leq X\}} | y_j]}, \quad i = 1, \dots, p(n).$$

Démonstration. A la $(m+1)^e$ itération de l'algorithme, la dérivée de $E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y]$ par rapport à α se calcule comme suit :

$$\begin{aligned} \frac{\partial}{\partial \alpha} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y] &= \frac{\partial}{\partial \alpha} E_{\theta^{(m)}} [\ln f(Z_0, \dots, Z_{a_X}, Z; \alpha) | y] \\ &= \frac{\partial}{\partial \alpha} E_{\theta^{(m)}} \left[\ln \left[\prod_{j=1}^{a_X} f(Z_j | Z_{j-1}; \alpha) \right] + \ln f(Z | Z_{a_X}; \alpha) | y \right] \\ &= \frac{\partial}{\partial \alpha} \left[\sum_{j=1}^{a_X} \ln f(z_j | z_{j-1}; \alpha) + E_{\theta^{(m)}} [\ln f(Z | Z_{a_X}; \alpha)] \right] \\ &= -\frac{\partial}{\partial \alpha} \left[\sum_{j=1}^{a_X} \left[\ln(\sigma \sqrt{2\pi}) + \frac{1}{2\sigma^2} (z_j - \mu z_{j-1})^2 \right] \right. \\ &\quad \left. + E_{\theta^{(m)}} \left[\ln(\sigma \sqrt{2\pi}) + \frac{1}{2\sigma^2} (Z - \mu Z_{a_X})^2 | y \right] \right]. \end{aligned}$$

Nous en déduisons les dérivées de $E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y]$ par rapport à μ et σ^2 :

$$\begin{aligned} \frac{\partial}{\partial \mu} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y] &= \frac{1}{\sigma^2} \left[\sum_{j=1}^{a_X} z_{j-1} (z_j - \mu z_{j-1}) + z_{a_X} E_{\theta^{(m)}} [Z - \mu Z_{a_X} | y] \right] \\ &= \frac{1}{\sigma^2} \left[\sum_{j=1}^{a_X} z_j z_{j-1} - \mu \sum_{j=0}^{a_X} z_j^2 + z_{a_X} E_{\theta^{(m)}} [Z | y] \right], \\ \frac{\partial}{\partial \sigma^2} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y] &= - \left[\sum_{j=1}^{a_X} \left[\frac{1}{\sigma^2} - \frac{1}{2\sigma^4} (z_j - \mu z_{j-1})^2 \right] \right. \\ &\quad \left. + E_{\theta^{(m)}} \left[\frac{1}{\sigma^2} - \frac{1}{2\sigma^4} (Z - \mu Z_{a_X})^2 | y \right] \right]. \end{aligned}$$

Finalement, en résolvant les équations

$$\sum_{i=1}^n \frac{\partial}{\partial \mu} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y_i] = 0 \quad \text{et} \quad \sum_{i=1}^n \frac{\partial}{\partial \sigma^2} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y_i] = 0,$$

nous obtenons les expressions suivantes pour $\mu^{(m+1)}$ et $\sigma^{2(m+1)}$:

$$\begin{aligned} \mu^{(m+1)} &= \frac{\sum_{i=1}^n [\sum_{j=1}^{a_{x_i}} z_{ij} z_{i,j-1} + z_{a_{x_i}} E_{\theta^{(m)}} [Z | y_i]]}{\sum_{i=1}^n \sum_{j=0}^{a_{x_i}} z_{ij}^2}, \\ \sigma^{2(m+1)} &= \frac{\sum_{i=1}^n [\sum_{j=1}^{a_{x_i}} (z_{ij} - \mu^{(m)} z_{i,j-1})^2 + E_{\theta^{(m)}} [(Z - \mu^{(m)} Z_{a_X})^2 | y_i]]}{\sum_{i=1}^n (a_{x_i} + 1)}. \end{aligned}$$

Dans la proposition 5.3, nous avons dérivé la log-vraisemblance EM par rapport à $\Delta\Lambda_{n,i}$ pour obtenir une caractérisation de $\hat{\theta}_n$. Le résultat obtenu nous permet de déduire

$$\Delta\Lambda_{n,i}^{(m+1)} = \frac{\sum_{j=1}^n \delta_j 1_{\{x_j=u_i\}}}{\sum_{j=1}^n E_{\theta^{(m)}} [e^{\beta^{(m)T} W(u_i)} 1_{\{u_i \leq x\}} | y_j]}.$$

■

Remarques

1. Pour la mise en oeuvre informatique de l'algorithme EM, il sera utile de décomposer le dénominateur de $\Delta\Lambda_{n,i}^{(m+1)}$ de la manière suivante, afin de faire clairement apparaître les espérances conditionnelles que nous aurons besoin de calculer :

$$\Delta\Lambda_{n,i}^{(m+1)} = \frac{\sum_{j=1}^n \delta_j 1_{\{x_j=u_i\}}}{\sum_{j=1}^n [E_{\theta^{(m)}} [e^{\beta^{(m)T} W(u_i)} | y_j] V_j(u_i) + e^{\beta^{(m)T} w_j(u_i)} U_j(u_i)]},$$

où $V_j(u) = 1_{\{t_{ax_j} < u \leq x_j\}}$ et $U_j(u) = 1_{\{u \leq t_{ax_j}\}}$.

2. Le calcul des valeurs itérées de μ, σ^2 et $\Delta\Lambda_{n,i}$ nécessite l'utilisation d'une méthode d'approximation numérique pour approcher les intégrales $E_{\theta^{(m)}}$. Dans le cas gaussien, l'une des techniques les plus utilisées est la quadrature de Gauss-Hermite. Une description de cette technique, ainsi que les formules d'approximation nécessaires pour $\mu^{(m+1)}, \sigma^{2(m+1)}, \Delta\Lambda_{n,i}^{(m+1)}$ sont données en Annexe A. Crouch et Spiegelman [23] donnent de plus amples détails sur la quadrature de Gauss-Hermite.

Il n'est pas possible de résoudre analytiquement l'équation $\partial E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y] / \partial \beta = 0$ pour obtenir une formule explicite de calcul de la valeur itérée $\beta^{(m+1)}$. Nous résolvons cette équation, à chaque itération de l'algorithme EM, à l'aide de l'algorithme de Newton-Raphson.

Au cours de la $(m+1)^e$ itération de l'algorithme EM, la formule d'itération de l'algorithme de Newton-Raphson permettant de calculer l'approximation $\beta_{(k+1)}$ de $\beta^{(m+1)}$ à partir de l'approximation $\beta_{(k)}$ de la k^e itération est obtenue de la façon suivante. La dérivée de la fonction $E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y]$

$$\frac{\partial}{\partial \beta} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y]$$

est remplacée par son développement de Taylor à l'ordre 1 au voisinage de $\beta_{(k)}$:

$$\begin{aligned} & \frac{\partial}{\partial \beta} E_{\theta^{(m)}} \left[\ln l(Y, Z; \alpha^{(m)}, \beta_{(k)}, \Delta \Lambda_{T,n}^{(m)}) | y \right] + \\ & \frac{\partial^2}{\partial \beta \partial \beta'} E_{\theta^{(m)}} \left[\ln l(Y, Z; \alpha^{(m)}, \beta_{(k)}, \Delta \Lambda_{T,n}^{(m)}) | y \right] (\beta - \beta_{(k)}). \end{aligned}$$

Nous sommions cette expression sur l'ensemble des sujets i ($i = 1, \dots, n$). La valeur de β qui annule cette somme fournit la valeur $\beta_{(k+1)}$, soit :

$$\begin{aligned} \beta_{(k+1)} &= \beta_{(k)} - \left[\frac{\partial^2}{\partial \beta \partial \beta'} \sum_{i=1}^n E_{\theta^{(m)}} \left[\ln l(Y, Z; \alpha^{(m)}, \beta_{(k)}, \Delta \Lambda_{T,n}^{(m)}) | y_i \right] \right]^{-1} \\ & \quad \times \left[\frac{\partial}{\partial \beta} \sum_{i=1}^n E_{\theta^{(m)}} \left[\ln l(Y, Z; \alpha^{(m)}, \beta_{(k)}, \Delta \Lambda_{T,n}^{(m)}) | y_i \right] \right] \\ &= \beta_{(k)} + [I^{(m)}(\beta_{(k)})]^{-1} U^{(m)}(\beta_{(k)}), \end{aligned} \quad (7.1)$$

où la valeur d'initialisation $\beta_{(0)}$ est choisie égale à $\beta^{(m)}$. On arrête l'algorithme lorsque $|\beta_{(k+1)} - \beta_{(k)}| \leq 10^{-4}$.

La proposition suivante donne les expressions de $U^{(m)}(\beta)$ et $I^{(m)}(\beta)$ nécessaires au calcul (7.1).

Proposition 7.2 *A la $(m+1)^e$ itération de l'algorithme EM, $U^{(m)}(\beta)$ et $I^{(m)}(\beta)$ sont données par*

$$\begin{aligned} U^{(m)}(\beta) &= \sum_{i=1}^n \delta_i [E_{\theta^{(m)}} [W(X) | y_i] - \mathbb{E}_m(\beta, x_i)], \\ I^{(m)}(\beta) &= \sum_{i=1}^n \delta_i \left[\frac{S_m^{(2)}(\beta, x_i)}{S_m^{(0)}(\beta, x_i)} - \mathbb{E}_m^{\otimes 2}(\beta, x_i) \right], \end{aligned}$$

où

$$S_m^{(r)}(\beta, s) = \sum_{i=1}^n \left[E_{\theta^{(m)}} \left[W^{\otimes r}(s) e^{\beta' W(s)} | y_i \right] V_i(s) + w_i^{\otimes r}(s) e^{\beta' w_i(s)} U_i(s) \right], \quad r = 0, 1, 2,$$

et

$$\mathbb{E}_m(\beta, s) = \frac{S_m^{(1)}(\beta, s)}{S_m^{(0)}(\beta, s)}.$$

Démonstration. $U^{(m)}(\beta)$ se calcule de la façon suivante :

$$\begin{aligned}
& \frac{\partial}{\partial \beta} E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y] \\
&= E_{\theta^{(m)}} \left[\Delta W(X) - \sum_{k=1}^{p(n)} \Delta \Lambda_{n,k} W(x_{(k)}) e^{\beta' W(x_{(k)})} 1_{\{x_{(k)} \leq X\}} | y \right] \\
&= E_{\theta^{(m)}} \left[\Delta W(X) - \sum_{k=1}^n \frac{\Delta \Lambda_n(x_k) \delta_k}{\sum_{j=1}^n \delta_j 1_{\{x_j = x_k\}}} W(x_k) e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y \right] \\
&= E_{\theta^{(m)}} \left[\Delta W(X) - \sum_{k=1}^n \frac{\delta_k}{\sum_{j=1}^n E_{\theta^{(m)}} [e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_j]} \right. \\
&\quad \left. \times W(x_k) e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y \right].
\end{aligned}$$

Nous sommions cette expression sur l'ensemble des sujets i ($i = 1, \dots, n$) et nous remarquons que

$$\begin{aligned}
& \sum_{i=1}^n \sum_{k=1}^n \frac{\delta_k}{\sum_{j=1}^n E_{\theta^{(m)}} [e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_j]} E_{\theta^{(m)}} [W(x_k) e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i] \\
&= \sum_{k=1}^n \delta_k \frac{\sum_{i=1}^n E_{\theta^{(m)}} [W(x_k) e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]}{\sum_{i=1}^n E_{\theta^{(m)}} [e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]},
\end{aligned}$$

d'où

$$U^{(m)}(\beta) = \sum_{k=1}^n \delta_k \left[E_{\theta^{(m)}} [W(X) | y_k] - \frac{\sum_{i=1}^n E_{\theta^{(m)}} [W(x_k) e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]}{\sum_{i=1}^n E_{\theta^{(m)}} [e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]} \right].$$

De nouveau, la mise en oeuvre informatique de ces procédures numériques sera facilitée en décomposant le dénominateur dans l'expression précédente de la manière suivante :

$$\begin{aligned}
U^{(m)}(\beta) &= \sum_{k=1}^n \delta_k \left[E_{\theta^{(m)}} [W(X) | y_k] \right. \\
&\quad \left. - \frac{\sum_{i=1}^n [E_{\theta^{(m)}} [W(x_k) e^{\beta' W(x_k)} | y_i] V_i(x_k) + w_i(x_k) e^{\beta' w_i(x_k)} U_i(x_k)]}{\sum_{i=1}^n [E_{\theta^{(m)}} [e^{\beta' W(x_k)} | y_i] V_i(x_k) + e^{\beta' w_i(x_k)} U_i(x_k)]} \right] \\
&= \sum_{k=1}^n \delta_k [E_{\theta^{(m)}} [W(X) | y_k] - \mathbb{E}_m(\beta, x_k)].
\end{aligned}$$

De même, nous calculons :

$$\begin{aligned}
I^{(m)}(\beta) &= -\frac{\partial^2}{\partial\beta\partial\beta'} \sum_{i=1}^n E_{\theta^{(m)}} [\ln l(Y, Z; \theta) | y_i] \\
&= \sum_{k=1}^n \delta_k \left[\frac{\sum_{i=1}^n E_{\theta^{(m)}} [W(x_k) W(x_k)' e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]}{\sum_{i=1}^n E_{\theta^{(m)}} [e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]} \right. \\
&\quad \left. - \frac{[\sum_{i=1}^n E_{\theta^{(m)}} [W(x_k) e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]]^{\otimes 2}}{\sum_{i=1}^n E_{\theta^{(m)}} [e^{\beta' W(x_k)} 1_{\{x_k \leq X\}} | y_i]^2} \right] \\
&= \sum_{k=1}^n \delta_k \left[\frac{S_m^{(2)}(\beta, x_k)}{S_m^{(0)}(\beta, x_k)} - \mathbb{E}_m^{\otimes 2}(\beta, x_k) \right].
\end{aligned}$$

■

Le calcul des valeurs itérées $\beta^{(m)}$ nécessitera également de calculer des approximations des intégrales $E_{\theta^{(m)}}$. Ces approximations sont décrites en Annexe A.

Estimation de la variance asymptotique des estimateurs de μ et σ^2

Les formules d'estimation des variances asymptotiques des estimateurs des paramètres β_1 et β_2 du modèle conjoint sont données dans la section 6.4 de ce manuscrit. Ici, nous obtenons des formules de calcul des estimations des variances asymptotiques des estimateurs de μ et σ^2 , à partir de la formule générale donnée pour α dans la section 6.4. La variance asymptotique de $\sqrt{n}(\hat{\mu}_n - \mu_0)$ (respectivement $\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)$) est estimée en posant $J = (1, 0)'$ (respectivement $J = (0, 1)'$) dans $J' [-\frac{1}{n} \sum_{i=1}^n M_i] J$, où :

$$\begin{aligned}
M_{i,1,1} &= -\frac{1}{\hat{\sigma}_n^2} \sum_{j=0}^{a_{x_i}} z_{ij}^2, \\
M_{i,1,2} &= M_{i,2,1} = -\frac{1}{\hat{\sigma}_n^4} \left[\sum_{j=1}^{a_{x_i}} z_{ij} z_{i,j-1} - \hat{\mu}_n \sum_{j=0}^{a_{x_i}} z_{ij}^2 + z_{ia_{x_i}} E_{\hat{\theta}_n} [Z | y_i] \right], \\
M_{i,2,2} &= \frac{a_{x_i} + 1}{\hat{\sigma}_n^4} - \frac{1}{\hat{\sigma}_n^6} \left[\sum_{j=1}^{a_{x_i}} (z_{ij} - \hat{\mu}_n z_{i,j-1})^2 \right] + E_{\hat{\theta}_n} [(Z - \hat{\mu}_n Z_{a_X})^2 | y_i].
\end{aligned}$$

Les algorithmes décrits dans cette section ont été programmés avec le logiciel Scilab [80]. Un listing des parties les plus importantes de ce programme est donné en Annexe B.

7.2 Applications de l'approche par vraisemblance conjointe

Nous présentons deux applications du modèle conjoint proposé. Nous expliquons tout d'abord comment il peut être appliqué à l'analyse statistique des abandons de mesure informatifs dans les études longitudinales. L'un des modèles les plus utilisés dans ce domaine est proposé par Diggle et Kenward [28] (nous le noterons “modèle DK” par la suite). Nous comparons, sur des données réelles de qualité de vie, les résultats des deux approches. Le modèle DK suppose deux hypothèses (décrites plus loin) restrictives sur la nature des données étudiées. Lorsque ces hypothèses sont vérifiées, nous constatons que les résultats du modèle DK et de l'approche conjointe sont proches. En revanche, l'approche conjointe sera préférable dans des situations plus générales.

Nous comparons ensuite sur un jeu de données réelles (quantités de lymphocytes CD4 chez des sujets séropositifs) les estimations du paramètre de régression d'un modèle de Cox obtenues par l'approche conjointe et deux méthodes d'imputation. Nous constatons sur cet exemple que l'approche conjointe donne des estimations moins biaisées que celles obtenues par les méthodes d'imputation.

7.2.1 Application aux abandons informatifs dans une étude longitudinale

L'analyse des données longitudinales (mesures répétées dans le temps d'une variable sur des sujets) se heurte très souvent au problème de l'abandon de mesure. Un abandon se produit lorsque la série de mesures d'un sujet s'interrompt et ne reprend pas jusqu'à la fin de l'étude. Diggle, Liang et Zeger [29] présentent un traitement détaillé de l'analyse statistique des données longitudinales et consacrent un chapitre au traitement des abandons.

Modèles pour l'abandon de mesure dans un étude longitudinale

Supposons que des mesures répétées z_j d'une variable Z sont obtenues aux instants t_j ($j = 0, \dots, m$) sur chaque sujet inclus dans une étude longitudinale. Notons D la variable aléatoire qui indique l'instant d'abandon d'un sujet dans cette étude. Si la première valeur manquante de Z est z_j , on considère le plus souvent, par convention, que l'abandon est survenu en t_j et que D prend la valeur $D = j$.

Diggle, Liang et Zeger distinguent trois types d'abandon de mesure dans une étude longitudinale, correspondant aux trois mécanismes de données manquantes rappelés dans le chapitre 3 de ce manuscrit.

Notons 1_{Z_j} la fonction indicatrice valant 1 si la réalisation de Z_j est observée et 0 sinon. Notons que $D = \sum_{j=0}^m 1_{Z_j}$. Un abandon est dit complètement aléatoire si les variables D et 1_{Z_j} sont indépendantes (pour tout $j = 0, \dots, m$). Un abandon est dit aléatoire si D est indépendant de Z_j ($j \geq D$) mais dépend de Z_j ($j < D$), c'est-à-dire, si la probabilité d'abandon dépend de valeurs observées de Z mais ne dépend pas de valeurs inobservées de Z , en particulier, de sa valeur à l'instant d'abandon. Enfin, un abandon est dit informatif si D dépend de Z_D .

Nous nous intéressons à la loi de l'instant d'abandon conditionnelle aux réalisations de Z . Lorsque l'abandon est aléatoire ou complètement aléatoire, les paramètres de cette loi sont estimés par les méthodes usuelles. En revanche, lorsque l'abandon est informatif, nous sommes confrontés au problème des valeurs manquantes de Z .

La méthode la plus utilisée consiste alors à modéliser conjointement le processus Z et la durée jusqu'à abandon. Little [58] et Verbeke et Molenberghs [89] donnent des revues exhaustives de la récente littérature consacrée à ce sujet. Elle se répartit entre les approches par modèles de mélange et modèles de sélection.

Les modèles de sélection consistent à obtenir la distribution conjointe des Z_j et de l'instant d'abandon D en choisissant une loi conditionnelle pour D sachant Z , et une loi marginale pour Z . Les modèles de mélange consistent à choisir une loi conditionnelle pour Z sachant l'instant d'abandon et une loi marginale pour D . Un exemple de modèle de mélange est décrit par Hogan et Laird [42].

Nous nous intéressons ici au modèle de sélection le plus utilisé en pratique, il est proposé par Diggle et Kenward [28]. Le lecteur intéressé par d'autres exemples pourra se référer aux modèles proposés par De Gruttola et Tu [25], Schluchter [79], Wu et Carroll [94] (où la loi conjointe des Z_j et de D est choisie normale multivariée).

Diggle et Kenward [28] proposent de combiner une loi normale multivariée pour Z avec un modèle logistique pour la loi d'abandon. Ce modèle a la forme suivante :

$$\text{logit } \mathbb{P}[D = j] = \beta_0 + \beta_1 z_j + \sum_{k=1}^j \beta_{k+1} z_{j-k} + \gamma' V, \quad (7.2)$$

où V est un vecteur de covariables connu.

Si $\beta_1 \neq 0$, (7.2) formule un modèle d'abandon de mesure informatif car la probabilité d'abandon à l'instant t_j dépend de la valeur inobservée z_j . Si $\beta_1 = 0$, (7.2)

formule un modèle d'abandon aléatoire. Enfin, si tous les β_k sont nuls (à l'exception éventuellement de β_0), (7.2) est un modèle d'abandon complètement aléatoire. Ce modèle doit sa popularité notamment à l'existence du logiciel gratuit OSWALD [81] qui estime ses paramètres par la méthode du maximum de vraisemblance à l'aide de l'algorithme du simplexe [69].

Ce modèle a donc connu de nombreux développements : extension au cas où Z est une variable catégorielle (Molenberghs, Kenward et Lesaffre [64]), extension à des abandons temporaires avec reprises intermittentes des mesures de Z (Troxel, Lipsitz et Harrington [83]).

Néanmoins, ce modèle présente certaines contraintes. Tout d'abord, il ne prend pas en compte la possibilité d'une censure de l'événement abandon. Puis, la convention suivante est adoptée : si un abandon survient entre t_{j-1} et t_j , on considère qu'il s'est produit en t_j , et $D = j$. Le modèle DK considère des abandons à temps discrets, ce qui entraîne une perte d'information pour l'estimation de ses paramètres, lorsque la vraie nature de l'abandon est continue.

Nous montrons que le modèle conjoint que nous avons proposé peut s'appliquer à l'analyse des abandons informatifs dans les études longitudinales, dans des conditions plus générales que le modèle DK, et nous comparons leurs résultats.

Comparaison du modèle conjoint et du modèle de Diggle et Kenward

Le modèle décrit dans le chapitre 4 formule une loi conjointe pour un processus longitudinal avec données manquantes et des durées de vie censurées. Cette loi est définie par une loi conditionnelle des durées de vie sachant la covariable longitudinale (c'est le modèle de Cox avec covariable dépendant du temps) et une loi marginale pour cette covariable. Le modèle proposé peut donc être vu comme un modèle de sélection.

La valeur de Z à l'instant d'événement n'est pas observée. Si cet événement est un abandon de mesure dans un suivi longitudinal, le modèle conjoint développé peut être vu comme un modèle d'abandon informatif.

Ce modèle prend en compte une possible censure de l'instant d'abandon et considère l'abandon comme un événement de nature continue. Il prend donc en compte des situations plus générales que le modèle DK. Nous comparons ces deux modèles sur des données de qualité de vie (notée qdv par la suite).

Le jeu de données étudié concerne 120 sujets traités par une nouvelle thérapie contre le cancer. Parallèlement au suivi médical de ces sujets, des mesures de leur

qdv ont été obtenues par des questionnaires d'évaluation de qdv, distribués à chaque sujet aux instants t_j et remplis par eux avant la date de remise du questionnaire, à t_{j+1} .

Un sujet abandonnant l'étude dans l'intervalle $]t_j, t_{j+1}]$ prive donc l'analyste de sa valeur de qdv, considérée comme constante sur cet intervalle. Les questionnaires produisent un score quantitatif compris entre 0 et 7, un score élevé indiquant une bonne qdv. Ici, l'événement d'intérêt est l'abandon et la covariable longitudinale Z est la qdv.

Reprenant les notations utilisées dans ce manuscrit, chaque sujet i ($i = 1, \dots, 120$) fournit les observations $(X_i = T_i \wedge C_i, \Delta_i, Z_{i0}, \dots, Z_{ia_{X_i}})$, où T_i est l'instant d'abandon et C_i est l'instant de censure (retrait du sujet de l'étude sur décision des médecins pour raisons médicales autres que la qdv). La valeur $Z_i(X_i)$ est inobservée. On peut penser que l'abandon est informatif : il est en effet vraisemblable qu'un sujet renoncera à remplir un questionnaire si sa qdv à cet instant est dégradée. 89 sujets ont abandonné l'étude, 31 sont censurés. On compte entre 1 et 13 mesures répétées par sujet.

Une première méthode d'analyse de l'abandon informatif dans cette étude consiste à utiliser le modèle DK. Pour cela, nous considérons que si un sujet abandonne entre t_j et t_{j+1} , alors $D = j$. De plus, nous considérons que tous les événements observés sont des abandons car le modèle DK ne prend pas en compte la censure (par la suite, nous noterons par $C1$ l'ensemble de ces deux conditions d'application du modèle DK).

Pour établir une comparaison pertinente avec le modèle conjoint, nous supposons que le modèle (7.2) s'écrit, pour une observation i :

$$\text{logit } \mathbb{P}[D_i = j] = \beta_1 z_{i,j-1} + \beta_2 z_{ij}. \quad (7.3)$$

Le programme OSWALD estime les paramètres β_1 et β_2 :

$$\text{logit } \mathbb{P}[D_i = j] = -0.036 z_{i,j-1} - 0.170 z_{ij}.$$

Il sera utile de réécrire le modèle estimé sous la forme suivante :

$$\text{logit } \mathbb{P}[D_i = j] = -0.170(z_{ij} - z_{i,j-1}) - 0.206 z_{i,j-1}.$$

Nous concluons que la probabilité d'abandon d'un sujet est d'autant plus forte que sa qdv est faible, et augmente lorsque sa qdv décroît entre deux instants successifs.

Les estimations des paramètres du modèle conjoint sont obtenues sous les mêmes conditions *C1* par l'algorithme EM décrit précédemment. Elles sont données dans la première colonne du tableau (7.1).

TAB. 7.1 – Estimations et écart-types pour les paramètres du modèle conjoint sous plusieurs conditions de censure et de nature de l'abandon.

emv	<i>C1</i> : pas de cens.	<i>C2</i> : pas de cens.	<i>C3</i> : censure	<i>C4</i> : censure
e.t.	abandon discret	abandon continu	abandon discret	abandon continu
$\hat{\beta}_{1,n}$	-0.0107	0.0592	0.0137	0.0887
e.t.	0.0691	0.0687	0.0825	0.0780
$\hat{\beta}_{2,n}$	-0.2017	-0.2840	-0.2241	-0.3161
e.t.	0.0697	0.0723	0.0850	0.0870
$\hat{\mu}_n$	0.9506	0.9461	0.9510	0.9475
e.t.	0.0092	0.0098	0.0091	0.0095
$\hat{\sigma}_n^2$	0.5718	0.5756	0.5723	0.5757
e.t.	0.0392	0.0400	0.0369	0.0401

Deux questions se posent. Tout d'abord, il serait intéressant de pouvoir comparer les résultats de l'approche conjointe et du modèle DK sous les conditions *C1*. Supposons ensuite que l'on dispose de données où les abandons sont de nature continue et sont éventuellement censurés. Pour utiliser le modèle DK, nous nous ramenons en fait aux conditions *C1*. Quelle est donc l'influence de ces hypothèses simplificatrices sur l'estimation du modèle (7.3) ? Le modèle conjoint proposé peut-il offrir une solution de rechange pour l'analyse des abandons informatifs dans des situations plus générales que celles envisagées par le modèle DK ?

Une comparaison directe des estimations des paramètres β_1 et β_2 des deux modèles n'est pas possible. Nous proposons donc la stratégie suivante. Soit deux individus i et j . Supposons qu'à l'instant t_k , $z_{ik} = z_{jk} = z$. Supposons qu'entre t_k et t_{k+1} , la qdv du sujet i reste constante et celle du sujet j varie. Nous allons comparer entre les deux modèles une grandeur indépendante de leur formulation : le risque relatif d'abandon du sujet j par rapport au sujet i , pour différentes valeurs de la variation de z_j entre t_k et t_{k+1} .

Supposons que cette variation soit égale à v . A l'instant t_{k+1} , le risque relatif d'aban-

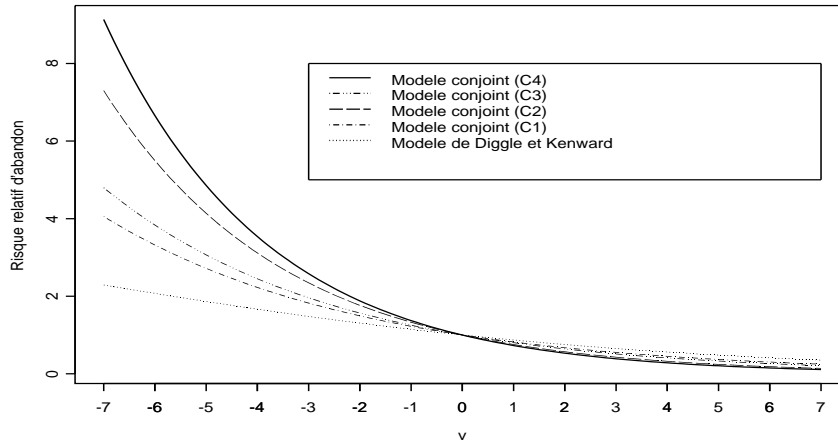
don du sujet j par rapport au sujet i , estimé par le modèle DK vaut :

$$\begin{aligned} rr_{j/i} &= \frac{e^{-0.170v-0.206z}}{1 + e^{-0.170v-0.206z}} \frac{1 + e^{-0.206z}}{e^{-0.206z}} \\ &= \frac{e^{-0.170v}(1 + e^{-0.206z})}{1 + e^{-0.170v-0.206z}}. \end{aligned}$$

D'après les résultats de la colonne 1 du tableau (7.1), le risque relatif d'abandon de j par rapport à i , estimé par le modèle conjoint est $e^{-0.201v}$.

Ces deux risques relatifs sont représentés en fonction de v sur le graphique (7.1) par les deux courbes les plus basses.

FIG. 7.1 – Estimation du risque relatif d'abandon en fonction de la variation v .



La variation v prend ses valeurs dans $[-7, 7]$. Néanmoins, les valeurs réellement interprétables se situent entre -3 et 3 , qui est l'étendue de valeurs dont peut réellement varier en pratique la qdv d'un sujet entre deux instants consécutifs. Notons qu'en effet, une variation de v égale à -5 par exemple, concernerait des sujets dont la qdv à t_k serait supérieure à 5 et diminuerait de 5 entre t_k et t_{k+1} , ce qui est improbable en pratique. Sur $[-3, 3]$, nous constatons la proximité des deux courbes sous $C1$.

Prenons maintenant en compte la nature réelle de nos données (présence de censure et continuité de l'abandon : conditions notées $C4$). Les estimations des

paramètres du modèle conjoint obtenues par l'algorithme EM sous ces conditions sont données dans le tableau (7.1). Il serait intéressant de comparer ces résultats à ceux du modèle DK, et d'évaluer l'effet de la perte d'information occasionnée en traitant par le modèle DK des données $C4$ sous les hypothèses $C1$. Sur le graphique, nous constatons que les courbes estimées sous $C1$ par le modèle conjoint et le modèle DK sont éloignées de la courbe obtenue par le modèle conjoint sous $C4$ pour les valeurs négatives de v . Deux autres courbes sont obtenues dans les deux cas intermédiaires suivants :

1. pas de prise en compte de la censure et abandon en temps continu ($C2$),
2. prise en compte de la censure et abandon en temps discret ($C3$).

Ces courbes se situent à des niveaux intermédiaires entre les courbes $C1$ et $C4$.

Il semble donc qu'ignorer une des deux caractéristiques suivantes : censure et continuité de l'abandon, sous-estime le risque d'abandon d'un individu dont la qdv décroît. Cette sous-estimation est maximale lorsque l'on ignore les deux caractéristiques en même temps. Le risque relatif estimé par le modèle conjoint est alors équivalent à celui estimé par le modèle DK. Nous expliquons ce résultat de la manière suivante.

Une censure est due à une décision de retrait prise par les médecins. Elle n'est pas liée directement à une baisse de la qdv, tandis qu'une baisse de la qdv est constatée chez les sujets qui abandonnent réellement. Considérer une censure comme étant un abandon tend à amoindrir cet impact de la baisse de qdv sur le risque d'abandon. Ceci explique la sous-estimation de ce risque constatée pour le modèle DK et le modèle conjoint estimé sous $C1$ et $C2$.

Lorsque la nature continue de l'abandon est ignorée, tous les sujets abandonnant dans un intervalle sont considérés comme ayant abandonné au même instant, à la fin de cet intervalle. Les sujets qui abandonnent le plus tôt sont ceux dont la qdv décroît le plus. La valeur pronostique de cette baisse pour l'abandon est affaiblie quand on considère que tous les sujets abandonnent au même instant. Ceci explique la sous-estimation constatée pour le modèle DK et le modèle conjoint estimé sous $C1$ et $C3$.

Conclusion

Le modèle de Diggle et Kenward (le plus utilisé des modèles d'abandons informatifs dans les études longitudinales) repose sur deux hypothèses contraignantes : pas de censure et nature discrète de l'abandon. Ignorer ces caractéristiques pour

des données qui les présentent conduit à une sous-estimation du risque d'abandon pour les sujets dont la qdv décroît. Nous avons proposé une utilisation originale du modèle conjoint dans le contexte des abandons informatifs, ce qui permet de prendre en compte pleinement la nature des données. Ce modèle fournit une alternative au modèle de Diggle et Kenward dans des situations où celui-ci ne devrait pas être utilisé.

7.2.2 Comparaison avec deux méthodes d'imputation de covariable dans le modèle de Cox

Il serait intéressant de comparer les estimations du paramètre de régression du modèle de Cox avec covariable dépendant du temps manquante obtenues par la vraisemblance conjointe et d'autres méthodes d'estimation.

Nous supposons que la fonction de risque instantané d'occurrence d'un événement d'intérêt est

$$\lambda_{T|Z}(t) = \lambda_T(t)e^{\beta Z(t)},$$

où la valeur de Z à l'instant d'événement n'est pas observée.

Dans le contexte de notre travail, où le mécanisme de données manquantes est informatif et où tous les sujets sont affectés par les valeurs manquantes, les méthodes proposées par Lin et Ying, Zhou et Pepe, Wang, Hsu, Feng et Prentice et Paik et Tsai ne sont pas applicables car elles supposent un mécanisme de données manquantes MAR ou MCAR, ou qu'il existe un sous-échantillon de sujets pour lesquels le processus de covariable est entièrement observé.

En revanche, nous avons programmé deux autres méthodes d'imputation très largement utilisées en pratique [19]. La méthode LCF (Last Carried Forward) consiste à remplacer la valeur inobservée $z(x)$ d'un sujet par sa dernière valeur observée (ici z_{a_x}). Cette méthode ne suppose pas que des sujets soient complètement observés, et n'utilise pas de valeurs observées sur d'autres sujets pour estimer des valeurs manquantes.

Nous programmons également une approche en deux étapes. Les paramètres du modèle longitudinal sont d'abord estimés dans chaque groupe constitué des sujets dont l'événement d'intérêt se produit dans un même intervalle $[t_k, t_{k+1}]$, et non sur l'ensemble des sujets. La valeur manquante pour un sujet est ensuite estimée à l'aide du modèle longitudinal estimé dans son groupe.

Nous disposons d'un jeu de données réelles (quantités de lymphocytes CD4 chez

101 patients séropositifs suivis jusqu'au décès, pas de censure), comprenant de 2 à 11 mesures par sujet (source : www.maths.lancs.ac.uk/diggle/lda/Datasets/). Nous estimons le paramètre β sur ces données complètes par la méthode du maximum de vraisemblance partielle : son estimation vaut -0.6840 et l'écart-type estimé est 0.1249 .

Puis nous retirons la dernière valeur observée pour chacun des sujets. Pour évaluer l'influence de la censure sur les méthodes d'estimation comparées, nous affectons aléatoirement une valeur de l'indicatrice de censure à chacun des sujets (pour 4 proportions de censure différentes : 0%, 10%, 25% et 50%). Les estimations obtenues sont données dans le tableau (7.2).

TAB. 7.2 – Comparaison de deux méthodes d'imputation et de l'approche conjointe pour l'estimation du paramètre β du modèle de Cox.

% censure		LCF	Approche en 2 étapes	Modèle conjoint
0	$\hat{\beta}_n$	-0.2634	-0.5190	-0.6498
	e.t.	0.1872	0.1829	0.1909
10	$\hat{\beta}_n$	-0.2629	-0.5211	-0.6569
	e.t.	0.1970	0.1923	0.1913
25	$\hat{\beta}_n$	-0.2931	-0.5433	-0.6727
	e.t.	0.2157	0.2112	0.2205
50	$\hat{\beta}_n$	-0.2359	-0.4943	-0.6314
	e.t.	0.2647	0.2597	0.2672

Nous constatons que l'approche par modélisation conjointe donne les estimations les moins biaisées, pour les quatre proportions de censure. L'estimation de la variance asymptotique augmente avec la proportion de censure pour les trois méthodes. La méthode LCF, qui ne tient pas compte de l'évolution de la covariable Z entre deux instants de mesure donne les estimations les plus biaisées. La méthode en deux étapes aboutit à des résultats intermédiaires. L'utilisation de l'approche conjointe devrait être préférée à ces deux méthodes.

Conclusion et perspectives

Dans ce travail, nous nous sommes intéressés au problème de l'estimation des paramètres du modèle de Cox en présence de valeurs manquantes informatives d'une covariable dépendant du temps.

Pour résoudre ce problème, nous avons proposé de modéliser conjointement les durées de vie censurées et le processus de covariable, pour en déduire une vraisemblance qui nous permette d'estimer les paramètres du modèle de Cox au vu des données incomplètes. Dans un premier temps, nous avons décrit le modèle conjoint proposé, puis nous avons établi sa propriété d'identifiabilité.

Nous avons ensuite considéré l'estimation des paramètres du modèle de Cox avec une covariable dépendant du temps manquante à partir du modèle conjoint. Nous avons adapté à ce modèle la méthode du maximum de vraisemblance semiparamétrique.

Nous avons tout d'abord montré l'existence d'estimateurs du maximum de vraisemblance semiparamétrique des paramètres du modèle conjoint, le paramètre de régression du modèle marginal de Cox étant considéré comme le paramètre d'intérêt de ce modèle.

Nous avons ensuite obtenu une caractérisation des estimateurs semiparamétriques en appliquant le principe de l'algorithme Espérance-Maximisation (EM) à une espérance conditionnelle de la log-vraisemblance des données complètes inobservées. En utilisant cette caractérisation, nous avons montré la consistance des estimateurs semiparamétriques des paramètres du modèle conjoint. Nous avons ensuite établi la convergence en loi de ces estimateurs vers un processus gaussien. Puis nous avons proposé un estimateur convergent de la variance asymptotique des estimateurs semiparamétriques.

D'un point de vue pratique, nous avons utilisé l'algorithme EM pour estimer les paramètres du modèle conjoint. Nous avons obtenu des formules explicites pour calculer, à chaque itération de l'algorithme, des estimations des paramètres de la loi du processus longitudinal et de la fonction de risque cumulé du modèle de Cox. Les estimations du paramètre de régression du modèle de Cox ont été obtenues, à chaque itération, par un algorithme numérique de type Newton. Nous avons obtenu des modifications de ces formules pour prendre en compte des

instants d'événement *ex aequo*.

Nous avons comparé, sur des données réelles, les estimations du paramètre de régression du modèle de Cox, obtenues par le modèle conjoint et deux méthodes d'imputation habituellement utilisées en cas de données manquantes. Cette comparaison suggère que la modélisation conjointe donne des estimations moins biaisées, y compris lorsque la proportion de censure des instants d'événements est importante. Nous avons aussi proposé d'appliquer le modèle conjoint à l'analyse statistique des abandons informatifs dans les études longitudinales. Ce modèle constitue une solution alternative au modèle proposé par Diggle et Kenward, qui repose sur des hypothèses restrictives sur la nature des données analysées. Un exemple d'application sur des données réelles suggère que ces deux modèles donnent des résultats similaires lorsque ces hypothèses sont vérifiées, mais que le modèle conjoint est préférable dans le cas contraire.

Il serait maintenant intéressant de réaliser des simulations plus poussées pour évaluer précisément les performances de la modélisation conjointe dans ces deux applications. En particulier, il serait intéressant de la comparer avec une approche récente du problème des covariables manquantes dans le modèle de Cox, basée sur une analyse de sensibilité [77].

Enfin, il serait intéressant d'étudier comment l'approche conjointe proposée peut permettre de généraliser les tests d'ajustement au modèle de Cox existants (une revue bibliographique de ces tests est donnée par Bagdonavičius et Nikulin [7, 8] et Devarajan et Ebrahimi [27]) à une situation de données manquantes d'une covariable dépendant du temps.

Annexe A

Approximation des intégrales intervenant dans la vraisemblance conjointe

Quadrature de Gauss-Hermite

Le calcul explicite d'une intégrale $I(f) = \int_0^b f(x) dx$ peut être difficile, voire impossible. On appelle formule de quadrature toute formule permettant de calculer une approximation de $I(f)$. En particulier, la quadrature de Gauss-Hermite permet l'approximation des intégrales de la forme

$$\int_a^b f(x) e^{-x^2} dx,$$

où a et b peuvent être infinis. Cette méthode est donc particulièrement intéressante pour calculer l'espérance d'une fonction d'une variable gaussienne.

Sa formule d'approximation est

$$\int_a^b f(x) e^{-x^2} dx \approx \sum_{l=1}^L \omega_l f(a_l),$$

où a_1, \dots, a_L sont des réels et $\omega_1, \dots, \omega_L$ sont les poids associés. Les points de quadrature a_l et les poids associés ω_l sont choisis de telle sorte que l'approximation soit exacte pour tout polynôme de degré inférieur ou égal à $2L - 1$.

Le principe du calcul de ces points et poids est expliqué dans [75]. Nous l'avons programmé à l'aide d'une fonction Scilab, en adaptant une routine de calcul proposée dans l'ouvrage de Press, Teukolsky, Vetterling et Flannery [75].

Approximation de $\mathbf{E}_{\theta^{(m)}}[\mathbf{g}(\mathbf{Z})|\mathbf{y}]$

Le calcul des valeurs itérées $\theta^{(m)}$ nécessite le calcul des espérances conditionnelles $E_{\theta^{(m)}}[g(Z)|y]$, pour les fonctions $g(Z)$ suivantes : $g(Z) = Z$, $g(Z) = Z^2$, $g(Z) = e^{\beta_1 Z}$, $g(Z) = Ze^{\beta_1 Z}$ et $g(Z) = Z^2 e^{\beta_1 Z}$. Nous utilisons la quadrature de Gauss-Hermite pour calculer des approximations de ces intégrales.

Les espérances $E_{\theta^{(m)}}[g(Z)|y]$ sont calculées par rapport à la densité conditionnelle $f_{Z|Y}(z|y; \theta^{(m)})$, qui peut se calculer comme

$$\frac{f(x, \delta | z_0, \dots, z_{a_x}, z; \beta^{(m)}, \Lambda_{T,n}^{(m)}) f(z_0, \dots, z_{a_x}, z; \alpha^{(m)})}{\int f(x, \delta | z_0, \dots, z_{a_x}, z; \beta^{(m)}, \Lambda_{T,n}^{(m)}) f(z_0, \dots, z_{a_x}, z; \alpha^{(m)}) dz},$$

où $f(x, \delta | z_0, \dots, z_{a_x}, z; \beta^{(m)}, \Lambda_{T,n}^{(m)})$ désigne la densité conditionnelle de (X, Δ) sachant Z .

L'espérance $E_{\theta^{(m)}}[g(Z)|y]$ est donc égale à

$$\frac{\int g(z) f(x, \delta | z_0, \dots, z_{a_x}, z; \beta^{(m)}, \Lambda_{T,n}^{(m)}) f(z_0, \dots, z_{a_x}, z; \alpha^{(m)}) dz}{\int f(x, \delta | z_0, \dots, z_{a_x}, z; \beta^{(m)}, \Lambda_{T,n}^{(m)}) f(z_0, \dots, z_{a_x}, z; \alpha^{(m)}) dz}. \quad (7.4)$$

Procédons au changement de variable suivant : $u(x) = (\sigma\sqrt{2})^{-1}(z(x) - \mu z_{a_x})$ et notons $\tilde{z}(x) = u(x)\sigma\sqrt{2} + \mu z_{a_x}$ et $w(x) = (z_{a_x}, \tilde{z}(x))$.

Avec ces notations, (7.4) devient :

$$\frac{\int g(\tilde{z}(x)) e^{\delta\beta_1\tilde{z}(x) - \sum_{i=1}^{p(n)} \Delta\Lambda_{n,i}^{(m)} \exp(\beta^{(m)T} w(x_i)) 1_{\{x_i \leq x\}}} e^{-u^2(x)} du(x)}{\int e^{\delta\beta_1\tilde{z}(x) - \sum_{i=1}^{p(n)} \Delta\Lambda_{n,i}^{(m)} \exp(\beta^{(m)T} w(x_i)) 1_{\{x_i \leq x\}}} e^{-u^2(x)} du(x)}.$$

Nous obtenons une approximation de cette expression par

$$\frac{\sum_{l=1}^L \left[g(\tilde{z}(x)) e^{\delta\beta_1\tilde{z}(x) - \sum_{i=1}^{p(n)} \Delta\Lambda_{n,i}^{(m)} \exp(\beta^{(m)T} w(x_i)) 1_{\{x_i \leq x\}}} \right] \omega_l}{\sum_{l=1}^L \left[e^{\delta\beta_1\tilde{z}(x) - \sum_{i=1}^{p(n)} \Delta\Lambda_{n,i}^{(m)} \exp(\beta^{(m)T} w(x_i)) 1_{\{x_i \leq x\}}} \right] \omega_l},$$

où $u(x)$ prend les valeurs a_l ($l = 1, \dots, L$) et ω_l sont les poids correspondants.

Notons que $\tilde{z}(x)$ et $w(x)$ dépendent de $u(x)$ donc des a_l ($l = 1, \dots, L$).

Nous avons programmé ces approximations à l'aide du logiciel Scilab [80].

Annexe B

Listing Scilab

Nous présentons ci-dessous les routines de calcul les plus importantes du programme implémentant l'algorithme EM décrit dans le chapitre 7.

Le programme suivant crée un vecteur c qui contient pour chaque sujet i le nombre d'observations de Z , et un vecteur Y qui contient pour chaque i la dernière valeur observée $z_{ia_{x_i}}$.

```
[nbrowX,nbcolX]=size(X) ; \\ X : matrice des données
c=zeros(nbrowX) ;
for i=1 :nbrowX ;
    k=1 ;
    for j=2 :nbcolX-1 ;
        if (X(i,j)<>0 & X(i,j-1)<>0) then
            k=k+1 ;
        end ;
        c(i)=k ;
    end ;
end ;
Y=ones(nbrowX,1) ;
for i=1 :nbrowX ;
    Y(i)=X(i,c(i)) ;
end ;
```

A la $(m+1)^e$ itération de l'algorithme EM, le programme ci-dessous calcule les k^e valeurs itérées $U^{(m)}(\beta_{(k)})$ et $I^{(m)}(\beta_{(k)})$.

```
E1b=zeros(2,1) ;
E1bE1bt=zeros(2,2) ;
E2b=zeros(2,2) ;
s=zeros(2,1) ;
for i=1 :nbrowX ;
    E1=zeros(2,1) ;
    E2=zeros(2,2) ;
```

```

d=0 ;
for l=i :nbrowX ;
    if (delta(i)==1 & X(l,c(i)+1)<>0) then
        ebX=exp(b0*X(l,c(i))+b1*X(l,c(i)+1)) ;
        XebX=[X(l,c(i)) ;X(l,c(i)+1)]*ebX ;
        E1=E1+XebX ;
        E2=E2+XebX*[X(l,c(i)),X(l,c(i)+1)] ;
        d=d+ebX ;
    elseif (delta(i)==1 & X(l,c(i)+1)==0 ) then
        [Eey,Eyey,Ey2ey]=esi(delta(l),H(l),Y(l)) ;
        eb0X=exp(b0*X(l,c(i))) ;
        eb0Xey=eb0X*Eey ;
        eb0Xyey=eb0X*Eyey ;
        Xeb0Xey=X(l,c(i))*eb0Xey ;
        E1=E1+[Xeb0Xey ;eb0Xyey] ;
        E2=E2+[X(l,c(i))*Xeb0Xey, X(l,c(i))*eb0Xyey ;
        X(l,c(i))*eb0Xyey,eb0X*Ey2ey] ;
        d=d+eb0Xey ;
    end ;
end ;
E1b=E1b+E1/d ;
E1bE1bt=E1bE1bt+(E1/d)*(E1'/d) ;
E2b=E2b+E2/d ;
if delta(i)==1 then
    s=s+[Y(i) ;et(delta(i),H(i),Y(i)))] ;
end ;
end ;
s=s-E1b ; \\ calcul de  $U^{(m)}(\beta_{(k)})$ 
i=E2b-E1bE1bt ; \\ calcul de  $I^{(m)}(\beta_{(k)})$ 

```

Ce programme fait appel aux fonctions **et** et **esi** qui calculent les approximations de $E_{\theta(m)}[g(Z)|y]$ pour $g(Z) = Z$, $g(Z) = e^{\beta_1 Z}$, $g(Z) = Ze^{\beta_1 Z}$ et $g(Z) = Z^2 e^{\beta_1 Z}$.

function[Ey]=et(d,has,y) ;

\\ En entrée : $\Delta\Lambda_{n,j}^{(m)}$, $j = 1, \dots, p(n)$, et pour chaque individu i : δ_i et $z_{ia_{x_i}}$

\\ En sortie : $E_{\theta^{(m)}}[Z|y_i]$

```
f=exp(d*b1t*((sqrt(2*st))*vec+tt*y)
      -has*exp(((sqrt(2*st))*vec+tt*y)*b1t+b0t*y));
g=(sqrt(2*scarre)*vec+theta*y).*f;
Ey=(weig'*g)/(weig'*f); \\ calcul de  $E_{\theta^{(m)}}[Z|y_i]$ 
```

function[Eey,Eyey,Ey2ey]=esi(d,has,y);

\\ En entrée : $\Delta\Lambda_{n,j}^{(m)}$, $j = 1, \dots, p(n)$, et pour chaque individu i : δ_i et $z_{ia_{x_i}}$
 \\ En sortie : $E_{\theta^{(m)}}[e^{\beta_1 Z}|y_i]$, $E_{\theta^{(m)}}[Ze^{\beta_1 Z}|y_i]$, $E_{\theta^{(m)}}[Z^2 e^{\beta_1 Z}|y_i]$

```
f=exp(d*b1t*((sqrt(2*st))*vec+tt*y)
      -has*exp(((sqrt(2*st))*vec+tt*y)*b1t+b0t*y));
yc=sqrt(2*scarre)*vec+theta*y;
eyc=exp(b1*yc);
ey=eyc.*f;
denomquad=weig'*f;
Eey=weig'*ey/denomquad; \\ calcul de  $E_{\theta}[e^{\beta_1 Z}|y_i]$ 
Eyey=weig'*(yc.*ey)/denomquad; \\ calcul de  $E_{\theta}[Ze^{\beta_1 Z}|y_i]$ 
Ey2ey=weig'*(yc.*(yc.*ey))/denomquad; \\ calcul de  $E_{\theta}[Z^2 e^{\beta_1 Z}|y_i]$ 
```

Le programme suivant calcule les valeurs itérées $\Delta\Lambda_{n,j}^{(m+1)}$, $j = 1, \dots, p(n)$:

```
lambda=zeros(nbrowX,1);
dlambda=0;
for i=1 :nbrowX;
    dlambda=0;
    if delta(i)==1 then
        for l=i :nbrowX;
            if X(l,c(i)+1)<>0 then
                dlambda=dlambda+exp(b1*X(l,c(i)+1)+ b0*X(l,c(i)));
            elseif X(l,c(i)+1)==0 then
                dlambda=dlambda+exp(b0*X(l,c(i)))* el(delta(l),H(l),Y(l));
            end;
            lambda(i)=1/dlambda; \\ calcul de  $\Delta\Lambda_{T,n}^{(m+1)}(x_i)$  si  $\delta_i = 1$ 
        end;
    end;
```

```

elseif delta(i)==0 then lambda(i)=0 ; \\ si  $\delta_i = 0$ ,  $\Delta\Lambda_{T,n}^{(m+1)}(x_i) = 0$ 
end ;
end ;

```

Le programme suivant calcule les valeurs itérées $\mu^{(m+1)}$ et $\sigma^{2(m+1)}$:

```

for i=1 :nbrowX ;
  for j=1 :c(i) ;
    dm=dm+X(i,j)*X(i,j) ;
  end ;
  for j=2 :c(i) ;
    nm1=nm1+X(i,j)*X(i,j-1) ;
  end ;
  if X(i,c(i)+1)<>0 then
    nm2=nm2+(X(i,c(i)+1)*X(i,c(i))) ;
  elseif X(i,c(i)+1)==0 then
    nm2=nm2+(X(i,c(i))*et(delta(i),H(i),Y(i))) ;
  end ;
end ;
mu=(nm1+nm2)/dt ; \\ calcul de  $\mu^{(m+1)}$ 
for i=1 :nbrowX ;
  for j=2 :c(i) ;
    ns1=ns1+(X(i,j)-theta*X(i,j-1))^2
  end ;
  if X(i,c(i)+1)<>0 then
    ns2=ns2+(X(i,c(i)+1)-theta*X(i,c(i)))^2 ;
  elseif X(i,c(i)+1)==0 then
    [Ey,Ey2]=es(delta(i),H(i),Y(i)) ;
    ns2=ns2+Ey2-2*theta*X(i,c(i))* Ey+(theta*X(i,c(i)))^2 ;
  end ;
end ;
ds=sum(c)+nbrowX ;
scarre=(ns1+ns2)/ds ; \\ calcul de  $\sigma^{2(m+1)}$ 

```

Le programme suivant calcule $\sum_{i=1}^n \ln L^{(i)}(\theta)$:

```

[n1,nc]=size(vec) ;

```

```

s1=0 ;
somm=0 ;
som=0 ;
s5=0 ;
for i=1 :nbrowX ;
    if (delta(i)<>0 & lambda(i)<>0) then
        s1=s1+log(lambda(i)) ;
    end ;
    for j=2 :c(i) ;
        somm=somm+(X(i,j)-theta*X(i,j-1))^2 ;
    end ;
    for j=1 :i ;
        if (delta(j)==1 & c(i)>c(j)) then
            som=som+(lambda(j))*exp(b*X(i,c(j)+1)) ;
        end ;
    end ;
    f=exp(delta(i)*b*(sqrt(2*scarre) *vec+ theta*Y(i))
        -H(i)* exp((sqrt(2*scarre)*vec+theta*Y(i))*b)) ;
    s5=s5+log(weig'*f) ;
end ;
s2=-0.5*sum(c)*log(2*(%pi)*scarre) ;
s3=-somm/(2*scarre) ;
s4=-som ;
loglik=s1+s2+s3+s4+s5 ; \\ calcul de  $\sum_{i=1}^n \ln L^{(i)}(\theta)$ 

```


Bibliographie

- [1] P. K. Andersen and R. D. Gill, *Cox's Regression Model for Counting Processes : A Large Sample Study*, The Annals of Statistics **10** (1982), 1100–1120.
- [2] T. Aven and U. Jensen, *Stochastic Models in Reliability*, Applications of Mathematics : Stochastic Modelling and Applied Probability, Springer, 1999.
- [3] V. Bagdonavičius and M. S. Nikulin, *Analysis of Generalized Additive Semiparametric Models*, Comptes Rendus de l'Académie des Sciences de Paris **323**, **Série I** (1996), 1079–1084.
- [4] ———, *Asymptotic Analysis of Semiparametric Models in Survival Analysis and Accelerated Life testing*, Statistics **29** (1997), 261–281.
- [5] ———, *Statistical Analysis of the Generalized Additive Semiparametric Survival Model with Random Covariates*, Qüestiió **21** (1997), 273–291.
- [6] ———, *Modèle Statistique de Dégradation avec des Covariables Dépendant du Temps*, Comptes Rendus de l'Académie des Sciences de Paris **330**, **Série I** (2000), 131–134.
- [7] ———, *On Goodness-of-Fit for the Linear Transformation and Frailty Models*, Statistics and Probability Letters **47** (2000), 177–188.
- [8] ———, *Accelerated Life Models : Modeling and Statistical Analysis*, Monographs on Statistics and Applied Probability, Chapman & Hall, 2001.
- [9] ———, *Estimation in Degradation Models with Explanatory Variables*, Lifetime Data Analysis **7** (2001), 85–103.
- [10] R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, Wiley Series in Probability and Mathematical Statistics, Wiley, New-York, 1965.
- [11] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins Series in Mathematical Sciences, The Johns Hopkins University Press, Baltimore, 1993.
- [12] P. Billingsley, *Convergence of Probability Measures*, Wiley Series in Probability and Mathematical Statistics, Wiley, New-York, 1968.

- [13] N. E. Breslow, *Contribution to the Discussion on the Paper by DR Cox, Regression Models and Life Tables*, Journal of the Royal Statistical Society, Series B **34** (1972), 216–217.
- [14] ———, *Covariance Analysis of Censored Survival Data*, Biometrics **30** (1974), 89–99.
- [15] M. Carbon, *Les Outils Probabilistes, p. 19-47 de Analyse Statistique des Durées de Vie. Modélisation des Données Censurées*, ASU, Economica, J.-J Droesbeke, B. Fichet et P. Tassi ed., Paris, 1989.
- [16] H. Y. Chen and R. J. A. Little, *Proportional Hazards Regression with Missing Covariates*, Journal of the American Statistical Association **94** (1999), 896–908.
- [17] G. Christol, A. Cot, and C.-M. Marle, *Calcul Différentiel*, Mathématiques Pour le Deuxième Cycle, Ellipses, Paris, 1997.
- [18] C. Coccoza-Thivent, *Processus Stochastiques et Fiabilité des Systèmes*, Mathématiques et Applications, Springer, 1997.
- [19] D. Collett, *Modelling Survival Data in Medical Research*, Texts in Statistical Science, Chapman & Hall, London, 1994.
- [20] D. R. Cox, *Regression Models and Life Tables (with Discussion)*, Journal of the Royal Statistical Society, Series B **34** (1972), 187–220.
- [21] ———, *Partial Likelihood*, Biometrika **62** (1975), 269–276.
- [22] D. R. Cox and D. Oakes, *Analysis of Survival Data*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1984.
- [23] E. A. C. Crouch and D. Spiegelman, *The Evaluation of Integrals of the Form $\int_{-\infty}^{+\infty} f(t) \exp(-t^2) dt$: Application to Logistic-normal Models*, Journal of the American Statistical Association **85** (1990), 464–469.
- [24] U. G. Dafni and A. A. Tsiatis, *Evaluating Surrogate Markers of Clinical Outcome When Measured With Error*, Biometrics **54** (1998), 1445–1462.
- [25] V. De Gruttola and X. M. Tu, *Modelling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time*, Biometrics **50** (1994), 1003–1014.
- [26] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion)*, Journal of the Royal Statistical Society, Series B **39** (1977), 1–38.
- [27] K. Devarajan and N. Ebrahimi, *Goodness-of-Fit Testing for the Cox Proportional Hazards Model, p. 237-254 de Goodness-of-Fit Tests and Model Validity*,

- C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah ed., *Statistics for Industry and Technology*, Birkhäuser, Boston, 2002.
- [28] P. J. Diggle and M. G. Kenward, *Informative Drop-out in Longitudinal Data Analysis (with Discussion)*, *Applied Statistics* **43** (1994), 49–93.
 - [29] P. J. Diggle, K. Y. Liang, and S. L. Zeger, *Analysis of Longitudinal Data*, Oxford Science Publications, 1994.
 - [30] M. Donsker, *An Invariance Principle for Certain Probability Limit Theorems*, *Memoirs of the American Mathematical Society* **6** (1951).
 - [31] ———, *Justification and Extension of Doob's Heuristic Approach to the Kolmogorov-Smirnov Theorems*, *The Annals of Mathematical Statistics* **23** (1952), 277–281.
 - [32] J. L. Doob, *Heuristic Approach to the Kolmogorov-Smirnov Theorems*, *The Annals of Mathematical Statistics* **20** (1949), 393–403.
 - [33] J.-F. Dupuy, I. Grama, and M. Mesbah, *Asymptotic Theory for the Cox Model with Missing Time-dependent Covariate*, soumis (2002).
 - [34] ———, *Normalité Asymptotique des Estimateurs Semiparamétriques dans le Modèle de Cox avec Covariable Manquante Non-ignorable*, soumis (2002).
 - [35] J.-F. Dupuy and M. Mesbah, *Joint Modeling of Survival and Informatively Missing Continuous Longitudinal Data*, dans : Abstract's book of the Second International Conference on Mathematical Methods in Reliability, MMR'2000, Bordeaux, **1** (2000), 357–360.
 - [36] ———, *Joint Modeling of Event Time and Nonignorable Missing Longitudinal Data*, *Lifetime Data Analysis* **8** (2002), 99–115.
 - [37] T. R. Fleming and D. P. Harrington, *Counting Processes and Survival Analysis*, Wiley Series in Probability and Mathematical Statistics, Wiley, New-York, 1991.
 - [38] I. B. Gertsbakh, *Statistical Reliability Theory*, Probability : Pure and Applied, Marcel Dekker, 1989.
 - [39] C. Gouieroux, J.-P. Lecoutre, and P. Tassi, *L'Approche Semi-Paramétrique, p. 243-248 de Analyse Statistique des Durées de Vie. Modélisation des Données Censurées*, ASU, Economica, J.-J Dreesbeke, B. Fichet et P. Tassi ed., Paris, 1989.
 - [40] C. Gouieroux and A. Monfort, *Statistique et Modèles Econométriques. Volume 1 : Notions Générales, Estimation, Préviation, Algorithmes*, Economica ed., Collection Economie et Statistiques Avancées, Paris, 1996.

- [41] F. Hirsch and G. Lacombe, *Éléments d'Analyse Fonctionnelle*, Dunod, 1997.
- [42] J. W. Hogan and N. M. Laird, *Mixture Models for the Joint Distribution of Repeated Measures and Event Times*, *Statistics in Medicine* **16** (1997), 239–257.
- [43] P. Hu, A. A. Tsiatis, and M. Davidian, *Estimating the Parameters in the Cox Model when Covariate Variables are Measured with Error*, *Biometrics* **54** (1998), 1407–1419.
- [44] C. Huber, *Les Différents Types de Censures, p. 49-58 de Analyse Statistique des Durées de Vie. Modélisation des Données Censurées*, ASU, Economica, J.-J Dreesbeke, B. Fichet et P. Tassi ed., Paris, 1989.
- [45] ———, *Censored and Truncated Lifetime Data, p. 291-306 de Recent Advances in Reliability Theory : Methodology, Practice, and Inference*, N. Limnios and M. Nikulin ed., *Statistics for Industry and Technology*, Birkhäuser, Boston, 2000.
- [46] C. Huber and J.-P. Lecoutre, *Estimation Fonctionnelle dans les Modèles de Durée, p. 59-120 de Analyse Statistique des Durées de Vie. Modélisation des Données Censurées*, ASU, Economica, J.-J Dreesbeke, B. Fichet et P. Tassi ed., Paris, 1989.
- [47] S. Johansen, *An Extension of Cox's Regression Model*, *International Statistical Review* **51** (1983), 165–174.
- [48] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions. Volume 1*, second ed., *Wiley Series in Probability and Mathematical Statistics*, Wiley, New-York, 1994.
- [49] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley, New-York, 1980.
- [50] J. P. Klein and M. L. Moeschberger, *Survival Analysis. Techniques for Censored and Truncated Data*, *Statistics for Biology and Health*, Springer, New-York, 1997.
- [51] L. Korsholm, *Likelihood Ratio Test in the Correlated Gamma-Frailty Model*, Tech. report, University of Aarhus and the Aarhus School of Business, 1998.
- [52] R. Kress, *Linear Integral Equations*, *Applied Mathematical Sciences*, Springer, Berlin, 1989.
- [53] M. Kulich and D. Y. Lin, *Additive Hazards Regression with Covariate Measurement Error*, *Journal of the American Statistical Association* **95** (2000), 238–248.

- [54] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley Series in Probability and Mathematical Statistics, Wiley, New-York, 1982.
- [55] Y. Li and X. Lin, *Covariate Measurement Errors in Frailty Models for Clustered Survival Data*, *Biometrika* **87** (2000), 849–866.
- [56] D. Y. Lin and Z. Ying, *Cox Regression With Incomplete Covariate Measurements*, *Journal of the American Statistical Association* **88** (1993), 1341–1349.
- [57] ———, *Semiparametric Analysis of the Additive Risk Model*, *Biometrika* **81** (1994), 61–71.
- [58] R. J. A. Little, *Modeling the Drop-Out Mechanism in Repeated-Measures Studies*, *Journal of the American Statistical Association* **90** (1995), 1112–1121.
- [59] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley Series in Decision and Control, Wiley, New-York, 1969.
- [60] T. Martinussen, *Cox Regression with Incomplete Covariate Measurements Using the EM-algorithm*, *Scandinavian Journal of Statistics* **26** (1999), 479–491.
- [61] E. Marubini and G. M. Valsecchi, *Analysing Survival Data from Clinical Trials and Observational Studies*, Wiley, Chichester, 1982.
- [62] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley Series in Probability and Statistics, Wiley, New-York, 1997.
- [63] W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, Wiley Series in Probability and Statistics, Wiley, New-York, 1998.
- [64] G. Molenberghs, M. G. Kenward, and E. Lesaffre, *The Analysis of Longitudinal Ordinal Data with Nonrandom Drop-out*, *Biometrika* **84** (1997), 33–44.
- [65] S. A. Murphy, *Consistency in a Proportional Hazards Model Incorporating a Random Effect*, *The Annals of Statistics* **22** (1994), 712–731.
- [66] ———, *Asymptotic Theory for the Frailty Model*, *The Annals of Statistics* **23** (1995), 182–198.
- [67] S. A. Murphy, A. J. Rossini, and A. W. Van der Vaart, *Maximum Likelihood Estimation in the Proportional Odds Model*, *Journal of the American Statistical Association* **92** (1997), 968–976.
- [68] S. A. Murphy and A. W. Van der Vaart, *On Profile Likelihood*, *Journal of the American Statistical Association* **95** (2000), 449–485.
- [69] J. A. Nelder and R. Mead, *A Simplex Method for Function Minimisation*, *The Computer Journal* **7** (1965), 303–313.

- [70] G. G. Nielsen, R. D. Gill, P. K. Andersen, and T. I. A. Sørensen, *A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models*, Scandinavian Journal of Statistics **19** (1992), 25–43.
- [71] M. C. Paik and W.-Y. Tsai, *On Using the Cox Proportional Hazards Model with Missing Covariates*, Biometrika **84** (1997), 579–593.
- [72] E. Parner, *Asymptotic Theory for the Correlated Gamma-Frailty Model*, The Annals of Statistics **26** (1998), 183–214.
- [73] O. Pons and C. Huber, *Survie et Processus Ponctuels*, Tech. report, Université René Descartes - Paris V, 2000.
- [74] R. L. Prentice, *Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model*, Biometrika **69** (1982), 331–342.
- [75] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*, second ed., Cambridge University Press, Cambridge, 1992.
- [76] W. Rudin, *Analyse Fonctionnelle*, Ediscience International, Paris, 1995.
- [77] D. O. Scharfstein, A. Rotnitzky, and J. M. Robins, *Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models*, Journal of the American Statistical Association **94** (1999), 1096–1120.
- [78] D. O. Scharfstein, A. A. Tsiatis, and P. B. Gilbert, *Semiparametric Efficient Estimation in the Generalized Odds-Rate Class of Regression Models for Right-Censored Time-to-Event Data*, Lifetime Data Analysis **4** (1998), 355–391.
- [79] M. D. Schluchter, *Methods for the Analysis of Informatively Censored Longitudinal Data*, Statistics in Medicine **11** (1992), 1861–1870.
- [80] Group Scilab, *Introduction to Scilab. INRIA Meta2 Project/ENPC Cergrene, User's Guide*, 1998.
- [81] D. M. Smith, W. H. Robertson, and P. J. Diggle, *Oswald : Object-Oriented Software for the Analysis of Longitudinal Data in S*, Tech. report, Department of Mathematics and Statistics, University of Lancaster, LA1 4YF, United Kingdom, 1996.
- [82] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data. Extending the Cox Model*, Statistics for Biology and Health, Springer, New-York, 2000.
- [83] A. B. Troxel, S. R. Lipsitz, and D. P. Harrington, *Marginal Models for the Analysis of Longitudinal Measurements with Nonignorable Non-monotone Missing Data*, Biometrika **85** (1998), 661–672.

- [84] A. A. Tsiatis, *A Large Sample Study of Cox's Regression Model*, The Annals of Statistics **9** (1981), 93–108.
- [85] A. A. Tsiatis, V. De Gruttola, and M. S. Wulfsohn, *Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS*, Journal of the American Statistical Association **90** (1995), 27–37.
- [86] A. W. Van der Vaart, *Asymptotic Statistics*, Cambridge Series in Probability and Statistics, Cambridge University Press, New-York, 1998.
- [87] ———, *Semiparametric Statistics*, Vrije Universiteit Amsterdam, www.cs.vu.nl/afdelingen/stochastiek/publicaties/2000-1.eps, 2000.
- [88] A. W. Van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes. With Applications to Statistics*, Springer Series in Statistics, Springer, Berlin, 1996.
- [89] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer Series in Statistics, Springer, Berlin, 2000.
- [90] V. G. Voinov and M. S. Nikulin, *Unbiased Estimators and Their Applications. Volume 1 : Univariate Case*, Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, 1993.
- [91] ———, *Unbiased Estimators and Their Applications. Volume 2 : Multivariate Case*, Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, 1996.
- [92] C. Y. Wang, L. Hsu, Z. D. Feng, and R. L. Prentice, *Regression Calibration in Failure Time Regression*, Biometrics **53** (1997), 131–144.
- [93] W. H. Wong, *Theory of Partial Likelihood*, The Annals of Statistics **14** (1986), 88–123.
- [94] M. C. Wu and R. J. Carroll, *Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modelling the Censoring Process*, Biometrics **44** (1988), 175–188.
- [95] M. S. Wulfsohn and A. A. Tsiatis, *A Joint Model for Survival and Longitudinal Data Measured with Error*, Biometrics **53** (1997), 330–339.
- [96] H. Zhou and M. S. Pepe, *Auxiliary Covariate Data in Failure Time Regression*, Biometrika **82** (1995), 139–149.