

# ANALYSE DE SURVIE, MODÈLES MULTI-ÉTATS ET PROCESSUS DE COMPTAGE

PHILIPPE SAINT PIERRE <sup>1</sup>

*Avril 2021*

1. Institut de Mathématiques de Toulouse, Université Paul Sabatier - Toulouse III



# Table des matières

<b>I</b>	<b>Processus de comptage</b>	<b>3</b>
1	Introduction . . . . .	3
2	Processus aléatoires, martingales et intégrales stochastiques . . . . .	3
3	Produit intégral (ou infini) . . . . .	7
4	Processus de comptage et intensité . . . . .	8
5	Vraisemblance d'un processus de comptage . . . . .	10
<b>II</b>	<b>Analyse des durées de vie et processus de comptage</b>	<b>11</b>
1	Processus agrégé . . . . .	11
2	Processus de comptage et censure à droite . . . . .	12
	2.1 Notations . . . . .	12
	2.2 Censure à droite indépendante . . . . .	13
3	Estimateur de Nelson-Aalen . . . . .	15
	3.1 Hypothèses et notations . . . . .	15
	3.2 Estimateur de $A$ . . . . .	16
	3.3 Biais et variance . . . . .	17
	3.4 Propriétés asymptotiques . . . . .	18
4	Estimateur de Kaplan-Meier . . . . .	19
	4.1 Estimateur de $S$ . . . . .	19
	4.2 Biais et variance . . . . .	21
	4.3 Propriétés asymptotiques . . . . .	23
5	Estimation semi-paramétrique . . . . .	23
	5.1 Estimation du risque cumulé de base . . . . .	24
	5.2 Estimation des coefficients de régression . . . . .	24
	5.3 Estimation de la survie . . . . .	26
	5.4 Propriétés asymptotiques . . . . .	26
<b>III</b>	<b>Modèles multi-états Markovien</b>	<b>29</b>
1	Processus de Markov . . . . .	29
2	Modèle Markovien homogène . . . . .	32
	2.1 Processus Markovien homogène . . . . .	32
	2.2 Vraisemblance . . . . .	33
	2.3 Modèle à intensités proportionnelles . . . . .	33
	2.4 Modèle de Markov homogène par périodes . . . . .	33
	2.5 Tests d'hypothèses et d'adéquation . . . . .	34
3	Modèle Markovien non-homogène . . . . .	35

3.1	Processus de comptage . . . . .	35
3.2	Vraisemblance . . . . .	36
3.3	Processus de comptage et censure à droite . . . . .	37
3.4	Estimation non-paramétrique . . . . .	38
3.4.1	Observations et notations . . . . .	38
3.4.2	Estimation des intensités cumulées . . . . .	39
3.4.3	Estimation des probabilités de transition . . . . .	40
3.4.4	Test des intensités de transition . . . . .	42
3.4.5	Cas particulier: données de survie . . . . .	43
3.5	Estimation semi-paramétrique . . . . .	44
3.5.1	Définitions et notations . . . . .	44
3.5.2	Estimation des intensités de base . . . . .	45
3.5.3	Estimation des coefficients de régression . . . . .	45
3.5.4	Estimation des probabilités de transition . . . . .	46
3.5.5	Cas particulier des durées de vie . . . . .	48
3.5.6	Tests et extensions . . . . .	48
	Covariables dépendantes du temps . . . . .	48
	Tests des coefficients . . . . .	48
	Test de l'hypothèse de proportionnalité des risques . . . . .	49
<b>IV</b>	<b>Modèle semi-Markovien homogène</b> . . . . .	<b>51</b>
1	Modèle semi-Markovien homogène . . . . .	51
2	Introduction . . . . .	51
2.1	Définitions . . . . .	51
2.2	Probabilités de transition du processus semi-Markovien . . . . .	54
2.3	Fonction de vraisemblance . . . . .	55
3	Estimation paramétrique des temps de séjour . . . . .	56
3.1	Introduction . . . . .	56
3.2	Modèle à risques proportionnels . . . . .	57
3.3	Modélisation paramétrique de la loi de séjour dans l'état . . . . .	58
3.4	Extension à un modèle semi-Markovien non-homogène . . . . .	59
<b>V</b>	<b>Bibliographie</b> . . . . .	<b>61</b>



# Chapitre I

## Processus de comptage

### 1 Introduction

On considère une variable aléatoire positive ou nulle  $X$  (de fonction de répartition  $F$  et de densité  $f$ ) qui représente la durée de vie. On suppose que  $X$  est censurée aléatoirement à droite par la variable  $C$ , c'est-à-dire que pour chaque individu ( $i = 1, \dots, n$ ) on observe les variables

$$\begin{cases} T_i = X_i \wedge C_i, & \text{où } X_i \text{ et } C_i \text{ sont } i.i.d \\ \delta_i = \mathbb{1}_{\{X_i \leq C_i\}} \end{cases}$$

On suppose également que les variables  $X$  et  $C$  sont indépendantes.

A partir de ces observations, on peut construire deux processus aléatoires :

$$N(t) = \sum_{i=1}^n \mathbb{1}_{\{T_i \leq t, \delta_i = 1\}},$$

qui compte le nombre de morts au temps  $t$  (*i.e.* le nombre d'événements qui ont eu lieu avant le temps  $t$ ),

$$Y(t) = \sum_{i=1}^n \mathbb{1}_{\{T_i \geq t\}},$$

qui compte le nombre d'individus à risque au temps  $t$  (individu ni mort ni censuré).

Dans ce qui suit, on trouvera une présentation succincte de la théorie des processus stochastiques. En effet, on introduira uniquement les notions nécessaires pour la suite de ce cours.

### 2 Processus aléatoires, martingales et intégrales stochastiques

On s'intéresse à la modélisation des occurrences dans le temps d'événements aléatoires. On étudiera particulièrement le cas où les événements ont lieu en **temps continu** et sont

de nature discrète (ils appartiennent à un **espace d'états dénombrable**). On considère donc un intervalle de temps continu

$$\mathcal{T} = [0, \tau[,$$

où  $\tau$  temps terminal tel que  $0 \leq \tau \leq \infty$ .

Exemples de différents types de processus dans un service de chirurgie :

- Espace d'états dénombrable et temps discret : le nombre d'opérations en attente aux temps de la  $t^{\text{ième}}$  opération.
- Espace d'états dénombrable et temps continu : le nombre d'opérations au cours du temps.
- Espace d'états continu et temps discret : temps d'attente pour la  $t^{\text{ième}}$  opération.
- Espace d'états continu et temps continu : temps total cumulé de toutes les opérations au cours du temps.

**Définition 1** Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ , une **filtration** (ou histoire)  $\{\mathcal{F}_t : t \in \mathcal{T}\}$  est une famille croissante et continue à droite de sous-tribus de  $\mathcal{A}$ . On supposera que la filtration est complète dans le sens où  $\forall t \in \mathcal{T}$ ,  $\mathcal{F}_t$  contient tous les ensembles négligeables de  $\mathcal{A}$ . Les filtrations  $\{\mathcal{F}_t : t \in \mathcal{T}\}$  considérées dans ce cours satisferont les conditions suivantes

$$\begin{aligned} \mathcal{F}_s &\subset \mathcal{F}_t \subset \mathcal{A}, & \forall s \leq t \text{ (croissance)}, \\ \mathcal{F}_s &= \bigcap_{t > s} \mathcal{F}_t, & \forall s \text{ (càd)}, \\ A \subset B \in \mathcal{A}, \mathbb{P}(B) = 0 &\implies A \in \mathcal{A} & \text{(complète)}. \end{aligned}$$

Une filtration permet de définir ce qui appartient passé du processus (qui est observé et connu) et ce qui appartient au futur (pas encore accessible).

Remarques :

- Les filtrations sont souvent définies comme étant engendrées par un processus stochastique  $X$ . La tribu engendrée par  $(X_s, s \leq t)$  est notée  $\mathcal{F}_t = \sigma(X_s, s \leq t)$ . Dans ce cas,  $X$  est automatiquement adapté à la filtration  $\sigma(X_s, s \leq t)$ .  $\mathcal{F}_t = \sigma(X_s, s \leq t)$  représente l'histoire du processus entre l'origine et la date  $t$ . Les variables  $\mathcal{F}_t$ -mesurables seront donc les fonctions des  $X_s$  pour  $s \leq t$ . Si, pour  $\omega \in \Omega$ , on a observé le processus jusqu'à la date  $t$ , on dispose de toute l'information nécessaire pour connaître exactement la valeur de  $Z(\omega)$  quelle que soit  $Z$  une variable  $\mathcal{F}_t$ -mesurable (à plus forte raison une variable  $\mathcal{F}_s$ -mesurable pour  $s \leq t$ ).
- On dispose souvent de plus d'information que la seule histoire du processus. A chaque instant, on observe également d'autres processus qui peuvent apporter de l'information sur le processus étudié. Dans ce cas, la filtration est élargie pour intégrer cette information (par exemple, prise en compte de covariables pour améliorer la prévision).
- On peut montrer qu'une filtration engendrée par un processus de saut ( $X$  tel que  $\forall s$  et  $\omega$ ,  $X_s(\omega)$  est constant pour  $s \in [t, t + \varepsilon[$ ,  $\varepsilon > 0$ ) est càd. De plus, si on ajoute les ensembles négligeables à une filtration càd, elle reste càd (par conséquent,  $\sigma(N_s, s < t)$  est càd).

**Définition 2** Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  où  $\mathbb{P}$  est la mesure de probabilité sur  $(\Omega, \mathcal{A})$ . Un **processus aléatoire**, ou encore une **fonction aléatoire réelle** (f.a.r.) est une fonction à deux variables

$$\begin{aligned} X : \mathcal{T} \times \Omega &\longrightarrow \mathbb{R} \\ (t, \omega) &\longmapsto X(t, \omega) \end{aligned}$$

où  $t$  représente le temps et  $\omega$  le hasard. Pour tout  $t \geq 0$ , la fonction  $X_t : \omega \longmapsto X(t, \omega)$  est une variable aléatoire réelle appelée **coordonnée** à l'instant  $t$ . Pour tout  $\omega \in \Omega$ , la **trajectoire** (chemin) est la fonction  $t \longmapsto (X(t, \omega), t \in \mathcal{T})$ . Si, pour presque tout  $\omega$ , la trajectoire est continue alors  $X(t, \omega)$  est une f.a.r. à trajectoire continue.

**Définition 3** Soient un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et une filtration  $\mathcal{F}_t$ . Le processus  $X$  est dit continu à droite (ou à gauche), à variation bornée, croissant (décroissant), possédant des limites à droite (à gauche) si l'ensemble des trajectoires qui satisfont ces propriétés est de probabilité 1.

**Définition 4** Soient un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $\mathcal{F}_t$  une filtration. Le processus  $X = X(t, \omega)$  est dit  $\mathcal{F}_t$ -**adapté** si  $\forall t, X_t$  est  $\mathcal{F}_t$ -mesurable.

**Définition 5** Un processus  $(X_t, t \in \mathcal{T})$  est intégrable si  $\sup_{t \in \mathcal{T}} \mathbb{E}(|X_t|) < \infty$ .

Un processus  $(X_t, t \in \mathcal{T})$  est de carré intégrable si  $\sup_{t \in \mathcal{T}} \mathbb{E}(X_t^2) < \infty$ .

**Définition 6** Un processus  $(M_t, t \in \mathcal{T})$  intégrable et adapté à une filtration  $\{\mathcal{F}_t : t \in \mathcal{T}\}$  est une **martingale** (resp. une sur-martingale, resp. une sous-martingale) si

$$\forall (s, t), 0 \leq s \leq t, \mathbb{E}(M_t | \mathcal{F}_s) = M_s \text{ p.s (resp. } \leq, \text{ resp. } \geq \text{)}.$$

**Définition 7** Un **temps d'arrêt**  $T$  par rapport à  $\mathcal{F}_t$  est une v.a. prenant ses valeurs dans  $[0, \infty]$  telle que  $\{T \leq t\} \in \mathcal{F}_t$ , pour tout  $t \in [0, \infty[$ .

Un **processus arrêté**  $X^T$  est défini par  $X_t^T(\omega) = \begin{cases} X_t(\omega) & \text{si } T(\omega) > t \\ X_{T(\omega)}(\omega) & \text{sinon} \end{cases}$

Exemple : si  $X$  est adapté et càdlàg alors  $T_C = \inf\{t, |X_t| \geq C\}$  est un temps d'arrêt. Soit  $X_t$  le volume d'eau dans un seau et  $C$  le volume du seau, l'eau déborde si son volume est supérieur ou égal au volume du seau. Le volume d'eau dans le seau n'évolue plus :  $X_{T_C}$  est un processus arrêté. On étudie la variable  $T_C$  pour minimiser le gaspillage d'eau.

**Proposition 1** Si  $X$  est un processus càdlàg et adapté alors le processus arrêté  $X^T$  l'est aussi.

**Définition 8** Un processus stochastique  $X$  est dit  $\mathcal{F}_t$ -**prévisible** si

(i) comme fonction de  $(t, \omega) \in \mathcal{T} \times \Omega \mapsto \mathbb{R}$  il est mesurable par rapport à la tribu sur  $\mathcal{T} \times \Omega$  engendrée par les processus adaptés et continu à gauche.

ou si

(ii)  $X(T)$  est  $\mathcal{F}_t$ -mesurable pour tout temps d'arrêt  $T$ .

En quelque sorte, la valeur de  $X$  en  $t$  est entièrement déterminé à partir des valeurs passées de  $X(s)$ ,  $s < t$ . Intuitivement, un processus prévisible est un processus dont la valeur en  $t$  découle des valeurs observés avant  $t$ .

**Proposition 2** Soit  $X$  un processus  $\mathcal{F}_t$ -prévisible alors pour tout  $t$ ,  $X_t$  est  $\mathcal{F}_t$ -mesurable.

**Proposition 3** Tout processus adapté à  $\mathcal{F}_t$  et continu à gauche est  $\mathcal{F}_t$ -prévisible.

### **Théorème 1 Décomposition de Doob-Meyer**

Soit  $X$  une sous-martingale positive continue à droite et adaptée à la filtration  $\mathcal{F}_t$ . Il existe une martingale continue à droite  $M$  et un processus prévisible croissant continu à droite  $\Lambda$  vérifiant  $\mathbb{E}(\Lambda_t) < \infty$  tel que presque sûrement on a

$$X_t = M_t + \Lambda_t \text{ pour tout } t \geq 0.$$

Si  $\Lambda_0 = 0$  presque sûrement et si  $X = M' + \Lambda'$  est une autre décomposition avec  $\Lambda'_0 = 0$ , alors pour tout  $t \geq 0$ ,  $\mathbb{P}(M'_t \neq M_t) = \mathbb{P}(\Lambda'_t \neq \Lambda_t) = 0$  (unicité de la décomposition).

**Définition 9** Le processus prévisible  $\Lambda$  est appelé **compensateur** de  $X$ .

**Définition 10** Une suite de temps d'arrêts  $(T_n)$  est dite **localisante** si elle est non décroissante et si pour tout  $t \in \mathcal{T}$ ,  $\mathbb{P}(T_n \geq t) \xrightarrow[n \rightarrow \infty]{} 1$ .

On dira que  $X$  a une **propriété locale** s'il existe une suite localisante  $(T_n)$  telle que pour tout  $n$  le processus  $\mathbf{1}_{\{T_n > 0\}} X^{T_n}$  a la propriété.

En particulier, un processus  $M$  est une **martingale locale** (resp. sous-martingale locale) pour  $\mathcal{F}_t$  s'il existe une suite localisante  $(T_n)$  telle que, pour tout  $n$ , le processus arrêté  $M^{T_n}$  est une martingale (resp. sous-martingale).

Un processus  $X$  est **localement borné** s'il existe une suite localisante  $(T_n)$  telle que  $X^{T_n}$  est un processus borné pour tout  $n$  (pour une suite de constante  $(c_n)$ , on a presque sûrement pour tout  $n$ ,  $\mathbf{1}_{\{T_n > 0\}} \sup_{t \leq T_n} |X_t| \leq c_n$ ).

Pour l'obtention de résultats théoriques, comme le calcul des variances, on considère souvent la décomposition de Doob-Meyer associée aux carrés des martingales. En particulier, comme le carré d'une martingale est une sous-martingale (inégalité de Jensen), la décomposition de Doob-Meyer permet d'obtenir le résultat suivant,

**Proposition 4** Soit  $M$  une martingale (resp. locale) adaptée à  $\mathcal{F}_t$ , continue à droite, telle que  $\mathbb{E}(M^2(t)) < \infty$  pour tout  $t$ . Alors il existe un unique processus continu à droite prévisible noté  $\langle M \rangle = \langle M, M \rangle$  tel que  $\langle M \rangle(0) = 0$  presque sûrement,  $\mathbb{E}(\langle M \rangle(t)) < \infty$  pour tout  $t$  et

$$M^2(t) - \langle M \rangle(t) \quad t \geq 0,$$

est une martingale (resp. locale) continue à droite.  $\langle M \rangle$  est le compensateur de  $M^2$ .

**Définition 11** Le processus  $\langle M \rangle = \langle M, M \rangle$  est appelé processus de variation prévisible de  $M$ .

**Définition 12** L'intégrale stochastique d'un processus  $X_t$  par rapport à un processus  $B_t$  est décrite par l'intégrale :

$$\int_0^t X_t dB_t = \lim_{\delta_n \rightarrow 0} \sum_{t_i \in \Delta_n} X_{t_i} (B_{t_{i+1}} - B_{t_i}),$$

où  $\Delta_n = (t_1, \dots, t_n)$  est une partition de  $[0, t]$  de pas  $\delta_n$ . La notation  $dB_t$  signifie un changement infinitésimal dans le processus aléatoire qui est une variable aléatoire.

**Remarque 1** Si  $f$  est une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$ , éventuellement aléatoire, la quantité  $\int_s^t f(u) dB_u$  ou  $\int_s^t f dB$  est une variable aléatoire, qui associe à toute réalisation  $\omega$  l'intégrale de Lebesgue-Stieltjes de  $f$  par rapport à la trajectoire  $B(\cdot, \omega)$ .

### 3 Produit intégral (ou infini)

**Définition 13** Soit  $\mathbf{X}(t)$ ,  $t \in \mathcal{T}$ , une matrice  $p \times p$  de processus càdlàg, nul en 0 et à variation localement bornée. On appelle **produit intégral** de  $X$  sur  $[s, t]$  (ou **produit infini**)

$$\mathcal{P}_{s \leq u \leq t} (\mathbf{Id} + d\mathbf{X}(u)) = \lim_{\max |t_i - t_{i-1}| \rightarrow 0} \prod_{i=1}^n [\mathbf{Id} + \mathbf{X}(t_i) - \mathbf{X}(t_{i-1})]$$

où  $t_0 = s < t_1 < \dots < t_n = t$  est une partition de  $[s, t]$  et  $\mathbf{Id}$  représente la matrice identité  $p \times p$ .

Dans le cas scalaire ( $p = 1$ ), le produit intégral vérifie

$$\mathcal{P}_{s \leq u \leq t} (1 + dX(u)) = \prod_{s \leq u \leq t} (1 + \Delta X(u)) \exp \left( \int_s^t dX^C(u) \right)$$

où  $X^C$  désigne la partie continue de  $X$  et  $\Delta X(u) = X(u) - X(u-)$ .

En effet, si  $X$  est continue et  $t_i - t_{i-1}$  petit alors  $X(t_i) - X(t_{i-1})$  est petit et  $\exp(X(t_i) - X(t_{i-1})) \simeq 1 + X(t_i) - X(t_{i-1})$ .

Notons que le produit intégral est introduit par Volterra (1887) comme l'unique solution de certaines équations intégrales.

**Théorème 2** Soient  $\mathbf{Z}$  une matrice  $p \times p$  de fonctions càdlàg et  $\mathbf{Id}$  la matrice identité. L'unique solution  $\mathbf{Z}$  de l'équation

$$\mathbf{Z}(t) = \mathbf{Id} + \int_0^t \mathbf{Z}(s-) \mathbf{X}(ds).$$

est  $\mathcal{P}_{0 \leq u \leq t}(\mathbf{Id} + d\mathbf{X}(\mathbf{u}))$ .

## 4 Processus de comptage et intensité

Un processus de comptage peut être vu comme l'enregistrement des occurrences dans le temps d'événements disjoints et discrets. Plus précisément,

**Définition 14** Soit  $\{\mathcal{F}_t : t \in \mathcal{T}\}$  une filtration sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ . Un **processus de comptage**  $N(\cdot)$  sur  $\mathcal{T}$  est un processus continu à droite avec une limite à gauche,  $\mathcal{F}_t$ -adaptée, nul en zéro, croissant, à trajectoire constante par morceaux, ayant des sauts d'amplitude 1 et tel que  $N(t)$  est p.s. fini pour tout  $t \in \mathcal{T}$ .

On considérera la **filtration naturelle** engendrée par  $N$ ,  $\mathcal{N} = \sigma(N(s), s \leq t)$ . On pourra également ajouter les ensembles négligeables. On montre que cette filtration est càd et croissante.

Comme un processus de comptage (croissant, nul en zéro et borné) est une sous-martingale positive, on déduit de la décomposition de Doob-Meyer,

**Proposition 5** Soit  $N = \{N(t), t \geq 0\}$  un processus de comptage adapté à  $\mathcal{F}_t$  tel que  $\mathbb{E}(N(t)) < \infty$  pour tout  $t$ . Alors il existe un unique processus  $\Lambda(\cdot)$   $\mathcal{F}_t$ -prévisible, croissant, càdlàg et nul en zéro p.s. vérifiant  $\mathbb{E}(\Lambda(t)) < \infty$  pour tout  $t$ , tel que

$$M(t) = N(t) - \Lambda(t), t \in \mathcal{T} \tag{I.1}$$

soit une martingale continue à droite.

L'idée de la décomposition est de diviser l'information initiale (le processus de comptage) en une composante  $\Lambda$  facile à contrôler (proche de ce qu'on cherche à obtenir), et une composante  $M$  qui s'apparente à un résidu (la structure de martingale permet de contrôler cette composante par l'utilisation des théorèmes asymptotiques).

En pratique, les processus de comptage ne vérifient pas toujours la condition  $\mathbb{E}(N(t)) < \infty$  pour tout  $t$ . Cette propriété n'est parfois vérifiée que pour des processus tronqués par les valeurs prises par une suite de temps d'arrêts qui tend p.s. vers  $+\infty$  (principe de localisation). La décomposition de Doob-Meyer peut être généralisée à des sous-martingales locales. On obtient le résultat suivant pour les processus de comptage,

**Proposition 6** Soit  $N = \{N(t), t \geq 0\}$  un processus de comptage adapté à  $\mathcal{F}_t$ , càdlàg, croissant et localement borné (c'est une sous-martingale locale). Alors il existe une unique processus  $\Lambda(\cdot)$   $\mathcal{F}_t$ -prévisible, croissant, càdlàg et nul en zéro p.s. vérifiant  $\mathbb{E}(\Lambda(t)) < \infty$  p.s. pour tout  $t$ , tel que

$$M(t) = N(t) - \Lambda(t), t \in \mathcal{T} \quad (\text{I.2})$$

soit une martingale locale continue à droite.  $\Lambda(\cdot)$  est le compensateur de  $N$ .

De plus,  $\Lambda$  est localement borné,  $M$  est localement de carré intégrable et  $\Lambda(t) = \lim_{u \rightarrow t, u < t} \Lambda(u) \leq 1$  p.s. pour tout  $t$ .

Dans les cas simples, il est possible de deviner la forme du compensateur  $\Lambda$ , et de montrer que  $N - \Lambda$  est une martingale. On peut ensuite conclure que  $\Lambda$  est bien le compensateur de  $N$  par unicité de la décomposition de Doob-Meyer. Le résultat suivant permet de déterminer le compensateur d'un processus de comptage

**Proposition 7** Soit  $N = \{N(t), t \geq 0\}$  un processus de comptage adapté à  $\mathcal{F}_t$ , càdlàg, croissant et localement borné et soit  $\Lambda(\cdot)$  son compensateur tel que  $N - \Lambda = M$ . Si toutes les trajectoires de  $\Lambda$  sont continues et si  $\mathbb{E}(M^2(t)) < \infty$  pour tout  $t$  alors on montre que  $M^2 - \Lambda$  est une martingale continue à droite. Par conséquent, par unicité de la décomposition et la proposition (4) on a

$$\langle M, M \rangle = \Lambda.$$

De nombreuses quantités rencontrées ne s'écrivent pas immédiatement comme des martingales mais sous la forme  $\int H dM$ , où  $M$  est une martingale et  $H$  est un processus prévisible. Sous certaines conditions les quantités  $\int H dM$  sont des martingales ou des martingales locales,

**Proposition 8** Soit  $M$  une martingale locale de carré intégrable et soit  $H$  un processus prévisible et localement borné alors

$$U(t) = \int_0^t H(u) dM(u)$$

est une martingale locale de carré intégrable telle que  $\mathbb{E}(U(t)) = 0$  et

$$\langle \int H dM \rangle = \int H^2 d \langle M \rangle.$$

Cette proposition s'applique notamment à la martingale de décomposition Doob-Meyer d'un processus de comptage.

**Proposition 9** Soient  $N$  un processus comptage et  $\Lambda$  son compensateur obtenu par la décomposition de Doob-Meyer. On dit que  $N$  a une **intensité**  $\lambda$  si  $\lambda$  est prévisible et si

$$\Lambda(t) = \int_0^t \lambda(u) du, \quad \forall t.$$

Si  $N$  intégrable et  $\lambda$  càdlàg on montre que

$$\lambda(u^+) = \lim_{\varepsilon \rightarrow 0} \mathbb{E}(N(u + \varepsilon) - N(u) \mid \mathcal{F}_u)$$

où  $\mathcal{F}_u = \sigma\{N(s) : s \leq u\}$ . Le paramètre  $\lambda$  représente, sachant le passé, la probabilité infinitésimale d'avoir un saut du processus de comptage.

**Définition 15** Un processus de comptage  $k$ -dimensionnel  $\mathbf{N} = (N_1, N_2, \dots, N_k)$  est appelé **processus de comptage multivarié** si chacune de ses composantes est un processus de comptage univarié et s'il ne peut y avoir simultanément des sauts de deux (ou plus) de ses composantes.

## 5 Vraisemblance d'un processus de comptage

Soit  $N = (N_1, N_2, \dots, N_k)$  un processus de comptage multivarié défini sur  $\mathcal{T}$  admettant comme compensateur  $\Lambda = (\Lambda_1, \Lambda_2, \dots, \Lambda_k)$  et comme intensité  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$  avec la mesure de probabilité  $\mathbb{P}$ .

**Proposition 10** La vraisemblance associée au processus  $\mathbf{N}$  est donnée par

$$d\mathbb{P} = d\mathbb{P}|_{\mathcal{F}_0} \mathcal{P}_{t \in [0, \tau]} \left( \prod_{h=1}^k d\Lambda_h(t)^{\Delta N_h(t)} (1 - d\Lambda.(t))^{1 - \Delta N.(t)} \right).$$

On définit la vraisemblance partielle

$$L_p = \mathcal{P}_{t \in [0, \tau]} \left( \prod_{h=1}^k d\Lambda_h(t)^{\Delta N_h(t)} (1 - d\Lambda.(t))^{1 - \Delta N.(t)} \right).$$

Si  $\Lambda$  est presque sûrement continue, la vraisemblance est

$$d\mathbb{P} = d\mathbb{P}|_{\mathcal{F}_0} \mathcal{P}_{t \in [0, \tau]} \left( \prod_{h=1}^k d\Lambda_h(t)^{\Delta N_h(t)} \right) \exp[-\Lambda.(\tau)],$$

Si  $\Lambda$  est presque sûrement absolument continue, alors

$$d\mathbb{P} = d\mathbb{P}|_{\mathcal{F}_0} \mathcal{P}_{t \in [0, \tau]} \left( \prod_{h=1}^k \lambda_h(t)^{\Delta N_h(t)} \right) \exp[-\Lambda.(\tau)],$$

où  $\Lambda. = \sum_{h=1}^k \Lambda_h$  et  $N. = \sum_{h=1}^k N_h$ .

La vraisemblance partielle conserve les mêmes propriétés que la vraisemblance complète pour l'inférence statistique.

## Chapitre II

# Analyse des durées de vie et processus de comptage

On considère  $T_1, \dots, T_n$  des v.a. positives indépendantes telles que tous les  $T_i$  aient la même fonction de survie  $S$ , fonction de répartition  $F$ , la densité  $f$  et la fonction de risque instantané  $\alpha = \frac{f}{S}$ . Les variables  $T_1, \dots, T_n$  représentent les durées de vie de  $n$  individus.

Pour tout  $i = 1, \dots, n$ , on peut définir

- le processus de comptage  $N_i(t) = \mathbf{1}_{\{T_i \leq t\}}$ ,
- le processus prévisible (càg et adapté)  $Y_i(t) = \mathbf{1}_{\{T_i \geq t\}}$ ,
- la filtration  $\mathcal{F}_i = (\mathcal{F}_{it})_{t \geq 0}$  où  $\mathcal{F}_{it} = \sigma(N_i(s), s \leq t)$ .

**Proposition 11** Le compensateur du processus de comptage  $N_i$  est  $\Lambda_i$  avec

$$\Lambda_i(t) = \int_0^t \alpha(s) Y_i(s) ds \quad \forall t \geq 0,$$

et  $N_i$  admet pour intensité

$$\lambda_i(t) = \alpha(t) Y_i(t) \quad \forall t \geq 0,$$

$N_i$  satisfait un modèle d'intensité multiplicative.

En effet, on peut montrer que  $\Lambda_i$  est un processus càdlàg, prévisible, croissant, nul en zéro et continu et que  $M_i = N_i - \Lambda_i$  est une martingale locale.

## 1 Processus agrégé

Comme les  $T_i$  sont des variables continues, on a  $\mathbb{P}(T_i = T_j) = 0$  pour tout  $i \neq j$ , on peut définir les processus agrégés  $N$  et  $Y$  et la filtration  $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ ,

$$N.(t) = \sum_{i=1}^n N_i(t), \quad Y.(t) = \sum_{i=1}^n Y_i(t), \quad \text{et} \quad \mathcal{F}_t = \sigma(N.(s), s \leq t).$$

**Proposition 12**  $N$  est un processus de comptage appelé processus de comptage agrégé.

Comme somme de processus càdlàg et adapté,  $N$  est un processus càdlàg et adapté ( $\mathcal{F}_{it} \subset \mathcal{F}_t$ ). De plus,  $N$  est nul en zéro et à trajectoire constante sur les intervalles  $[T_{(i)}, T_{(i+1)})$  où  $T_{(i)}$   $i = 1, \dots, n$  sont les statistiques d'ordre de  $T_1, \dots, T_n$ . Enfin,  $N$  saute de  $+1$  à chaque point  $T_{(i)}$  et  $N \leq n$ .

On peut remarquer que les  $T_{(i)}$   $i = 1, \dots, n$  vérifient

$$T_{(i)} = \inf\{t, N.(t) \geq i\},$$

et sont donc des temps d'arrêts pour  $N$ .

**Proposition 13** Le compensateur du processus de comptage  $N$  est  $\Lambda$ , avec

$$\Lambda.(t) = \int_0^t \alpha(s) Y.(s) ds \quad \forall t \geq 0,$$

et  $N$  admet pour intensité

$$\lambda(t) = \alpha(t) Y.(t) \quad \forall t \geq 0.$$

$N$  satisfait un modèle d'intensité multiplicative.

Notons que  $M = \sum_{i=1}^n M_i = N - \Lambda$  est une martingale locale et que  $\Lambda = \sum_{i=1}^n \Lambda_i$  est un processus càdlàg, prévisible, croissant, nul en zéro et continu.

## 2 Processus de comptage et censure à droite

### 2.1 Notations

Considérons  $n$  individus et  $X_i$  la durée de vie de l'individu  $i$ . On suppose que les durées  $X_1, \dots, X_n$  sont *i.i.d.* On se place dans le cas d'une censure à droite aléatoire de type I. Pour  $i = 1, \dots, n$ , on observe les variables

$$\begin{cases} T_i = X_i \wedge C_i, & \text{où les } C_i \text{ sont } i.i.d \\ \delta_i = \mathbb{1}_{\{X_i \leq C_i\}} \end{cases}$$

et on suppose que  $X_i$  et  $C_i$  sont indépendantes.

Dans le cas où la durée de vie  $X_i$  est censurée à droite par  $C_i$ , l'observation des processus de comptage  $N_i(t) = \mathbb{1}_{\{X_i \leq t\}}$  et  $Y_i(t) = \mathbb{1}_{\{X_i \geq t\}}$  est stoppée au temps aléatoire  $C_i$  :  $N_i$  est observé uniquement sur l'ensemble aléatoire  $E_i = \{t \leq C_i\}$  où quand

$$C_i(t) = \mathbb{1}_{\{t \in E_i\}} = \mathbb{1}_{\{t \leq C_i\}} = 1.$$

$C_i(t)$  est le processus de censure à droite (prévisible) de l'individu  $i$ .

Pour  $i = 1, \dots, n$ , les processus réellement observés sont définis par

—  $N_i^c(\cdot)$  le processus de comptage censuré à droite représentant la partie observable de  $N_i(\cdot)$  :

$$\begin{aligned} N_i^c(t) &= \int_0^t C_i(s) dN_i(s) \\ &= \mathbb{1}_{\{X_i \leq t, \delta_i=1\}} = \mathbb{1}_{\{T_i \leq t, \delta_i=1\}}, \end{aligned}$$

—  $Y_i^c(t)$  tel que,

$$\begin{aligned} Y_i^c(t) &= \mathbb{1}_{\{T_i \geq t\}} = \mathbb{1}_{\{C_i \geq t\}} \mathbb{1}_{\{X_i \geq t\}} \\ &= C_i(t) Y_i(t). \end{aligned}$$

L'introduction de la censure dans le modèle du paragraphe précédent apporte une variation aléatoire supplémentaire. En effet, les  $C_i(t)$  ne sont pas forcément adaptés à la filtration  $\mathcal{F}_t = \sigma(N_i(s), s \leq t)$ . On travaille donc avec une filtration élargie  $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(C_i(s), i = 1, \dots, n, s \leq t)$  pour laquelle  $C_i$  est prévisible. Cependant, on ne peut pas supposer que les processus  $C_i(t)$  sont entièrement observés (car on n'observe pas  $C_i$  si  $X_i \leq C_i$ ). On réduit donc la filtration  $\mathcal{G}_t$  pour ne prendre en compte que les données disponibles. Définissons

$$\mathcal{F}^c = (\mathcal{F}_t^c)_{t \geq 0} \quad \text{avec } \mathcal{F}_t^c = \sigma((N_i^c(s), Y_i^c(s)), i = 1, \dots, n, s \leq t).$$

## 2.2 Censure à droite indépendante

**Définition 16** Soit  $N(\cdot)$  un processus de comptage de compensateur  $\Lambda(\cdot)$  par rapport à la filtration  $\mathcal{F}_t$ . Soit  $\mathcal{C}(\cdot)$  un processus de censure à droite prévisible par rapport à  $\mathcal{F}_t^c$ . La censure à droite générée par  $\mathcal{C}(\cdot)$  est **indépendante** si le compensateur de  $N(\cdot)$  par rapport à  $\mathcal{F}_t^c$  est aussi  $\Lambda(\cdot)$ .

Autrement dit, la connaissance des temps de censure juste avant  $t$  ne modifie pas l'intensité du processus  $N$  au temps  $t$ . Lorsque la censure est indépendante, la répartition des temps de décès est la même pour les patients censurés et pour les patients non censurés.

Considérons les processus agrégés

$$N^c(t) = \sum_{i=1}^n N_i^c(t) \quad \text{et} \quad Y^c(t) = \sum_{i=1}^n Y_i^c(t).$$

**Proposition 14** Sous l'hypothèse de censure à droite indépendante, Le compensateur  $\Lambda^c(\cdot)$  par rapport à  $\mathcal{F}_t^c$  de  $N^c(\cdot)$  est

$$\Lambda^c(t) = \sum_{i=1}^n \int_0^t C_i(s) d\Lambda_i(s) = \int_0^t \alpha(s) Y^c(s) ds.$$

$N^c(\cdot)$  satisfait un modèle d'intensité multiplicative par rapport à  $\mathcal{F}_t^c$  avec

$$\lambda^c(t) = \alpha(t) Y^c(t) \quad \forall t \geq 0,$$

et  $Y_i^c(t) = C_i(t) Y_i(t)$ .

**Preuve** On sait que  $\forall i = 1, \dots, n$ ,  $N_i = \Lambda_i + M_i$  où  $M_i$  est une martingale locale et  $\Lambda_i$  le compensateur de  $N_i$  où  $\Lambda_i(t) = \int_0^t \alpha(s) Y_i(s) ds$ .

Comme  $\mathcal{F}_t \subset \mathcal{F}_t^c$ ,  $M_i$  est également une martingale locale par rapport à  $\mathcal{F}_t^c$ .

$$\begin{aligned} N_i^c(t) &= \int_0^t \mathcal{C}_i(s) dN_i(s) = \int_0^t \mathcal{C}_i(s) d\Lambda_i(s) + \int_0^t \mathcal{C}_i(s) dM_i(s) \\ &= \Lambda_i^c(t) + M_i^c(t). \end{aligned}$$

Comme  $\Lambda_i$  est prévisible croissant et  $\mathcal{C}_i$  est prévisible,  $\Lambda_i^c$  est prévisible croissant. De plus  $M_i$  est une martingale locale de carré intégrable et  $\mathcal{C}_i$  est prévisible borné donc  $M_i^c$  est une martingale locale. En sommant  $N^c(\cdot) = \sum_{i=1}^n N_i^c(\cdot)$ , on obtient le résultat. De plus,

$$\begin{aligned} \Lambda^c(t) &= \sum_{i=1}^n \Lambda_i^c(t) \\ &= \sum_{i=1}^n \int_0^t \alpha(s) \mathcal{C}_i(s) Y_i(s) ds = \sum_{i=1}^n \int_0^t \alpha(s) Y_i^c(s) ds \\ &= \int_0^t \alpha(s) \sum_{i=1}^n Y_i^c(s) ds. \end{aligned}$$

■

**Proposition 15** Soit  $\mathbf{N}^c(t) = (N_i^c(t); i = 1, \dots, n)$  un processus de comptage multivarié censuré. Soient  $\Lambda_i^c$  et  $\lambda_i^c$  le compensateur et l'intensité du processus  $N_i$ . Si la censure à droite est indépendante alors la vraisemblance partielle associée au processus  $\mathbf{N}^c$  s'écrit

$$\begin{aligned} \mathcal{L}^c &= \mathcal{P}_{t \in [0, \tau]} \left( \prod_{i=1}^n d\Lambda_i^c(t)^{\Delta N_i^c(t)} [1 - d\Lambda^c(t)]^{1 - \Delta N^c(t)} \right) \\ &= \mathcal{P}_{t \in [0, \tau]} \left( \prod_{i=1}^n \lambda_i^c(t)^{\Delta N_i^c(t)} \right) \exp[-\Lambda^c(\tau)]. \end{aligned} \quad (\text{II.1})$$

où  $\Lambda^c = \sum_{i=1}^n \Lambda_i^c$  et  $N^c = \sum_{i=1}^n N_i^c$ .

Sous hypothèse d'indépendance, un processus de comptage censuré à droite satisfait toujours un modèle d'intensité multiplicative. De plus, la vraisemblance partielle associée à ce processus a la même forme que la vraisemblance du processus non censuré. Ces résultats impliquent que l'inférence et la théorie asymptotique s'appliquent de la même manière pour les données censurées.

Désormais, nous nous placerons dans le cadre d'un **mécanisme de censure indépendante**. Dans tout ce qui suit, les processus considérés seront des versions censurées mêmes si les notations ne le feront plus apparaître.

### 3 Estimateur de Nelson-Aalen

#### 3.1 Hypothèses et notations

Soient  $X_1, \dots, X_n$  les durées de vie de  $n$  individus. On se place dans le cas d'une censure à droite aléatoire de type I. Pour  $i = 1, \dots, n$ , on observe les variables

$$\begin{cases} T_i = X_i \wedge C_i, \\ \delta_i = \mathbb{1}_{\{X_i \leq C_i\}} \end{cases}$$

On fait les hypothèses suivantes,

- les variables  $X_i$  et  $C_i$  sont indépendantes,
- les variables  $X_1, \dots, X_n$  sont *i.i.d* de fonction de répartition  $F$ , de fonction de survie  $S = \bar{F}$ , de densité  $f$ , de fonction de risque instantané  $\alpha = \frac{f}{S}$  et de risque cumulé  $A = \int \alpha$ ,
- les variables  $C_1, \dots, C_n$  sont *i.i.d* de fonction de répartition  $G$  et de fonction de survie  $\bar{G}$ ,
- on travaille sur  $[0, \tau_H]$  où  $\tau_H = \tau_F \wedge \tau_G = \sup\{t : \bar{F}(t)\bar{G}(t) > 0\}$ ,  $\tau_F = \sup\{t : \bar{F}(t) > 0\}$  et  $\tau_G = \sup\{t : \bar{G}(t) > 0\}$ .

Dans la section précédente on a vu que, sous une hypothèse de censure à droite indépendante, le processus de comptage

$$N(t) = \sum_{i=1}^n N_i(t) = \sum_{i=1}^n \mathbb{1}_{\{T_i \leq t, \delta_i=1\}}$$

vérifie

$$N(t) = \Lambda(t) + M(t)$$

où  $M(\cdot)$  est une martingale locale de carré intégrable et où  $\Lambda(\cdot)$  est le compensateur de  $N(\cdot)$  tel que

$$\Lambda(t) = \sum_{i=1}^n \Lambda_i(t) = \int_0^t \alpha(s) Y(s) ds \quad \text{où} \quad Y(t) = \sum_{i=1}^n Y_i(t) = \sum_{i=1}^n \mathbb{1}_{\{T_i \geq t\}}.$$

De plus, le processus de comptage  $N(\cdot)$  a pour intensité multiplicative

$$\lambda(t) = \alpha(t) Y(t) \quad \forall t \geq 0,$$

avec  $Y(t)$  est un processus prévisible et  $\alpha$  est une fonction de risque instantané.

Nous avons vu que  $M = \sum_{i=1}^n M_i = N - \Lambda$  est une martingale locale et que  $\Lambda$  est un processus càdlàg, prévisible, croissant, nul en zéro et continu. Ce processus  $M$  joue un rôle important dans l'obtention des résultats théoriques. On peut notamment remarquer que

$\mathbb{E}(M) = 0$  car,  $\forall i = 1, \dots, n$

$$\begin{aligned} \mathbb{E}(N_i(t)) &= \mathbb{P}(T_i \leq t, \delta_i = 1) = \mathbb{P}(X_i \leq t, X_i \leq C_i) = \mathbb{E}(\mathbb{E}(\mathbb{1}_{\{X_i \leq t\}} \mathbb{1}_{\{X_i \leq C_i\}} \mid X_i)) \\ &= \mathbb{E}(\mathbb{1}_{\{X_i \leq t\}} \bar{G}(X_i-)) = \int_0^t \bar{G}(s-) f(s) ds = \int_0^t \bar{G}(s-) S(s-) \frac{f(s)}{S(s)} ds \\ &= \int_0^t \mathbb{P}(C_i \geq s, X_i \geq s) \frac{f(s)}{S(s)} ds = \int_0^t \mathbb{P}(T_i \geq s) \alpha(s) ds \\ &= \int_0^t \mathbb{E}(Y_i(s)) \alpha(s) ds = \mathbb{E} \left( \int_0^t Y_i(t) \alpha(s) ds \right) = \mathbb{E}(\Lambda_i(t)). \end{aligned}$$

### 3.2 Estimateur de A

On cherche à estimer le risque cumulé A

$$A(t) = \int_0^t \alpha(s) ds \quad \forall t \geq 0.$$

D'après la décomposition de Doob-Meyer

$$M(t) = N(t) - \Lambda(t) = N(t) - \int_0^t \alpha(s) Y(s) ds$$

est une martingale locale de carré intégrable. Heuristiquement, on a

$$dN(t) = dM(t) + \alpha(t) Y(t) dt,$$

où  $dM(t)$  peut être considéré comme un "bruit aléatoire". On déduit alors que  $\frac{dN(t)}{Y(t)} \simeq \alpha(t) dt$  et que

$$\int_0^t \frac{dN(s)}{Y(s)} \simeq \int_0^t \alpha(s) ds = A(t).$$

**Définition 17** L'estimateur de **Nelson-Aalen** du risque cumulé est

$$\hat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s), \quad (\text{II.2})$$

où  $J(t) = \mathbb{1}_{\{Y(t) > 0\}}$ . On utilise la convention  $J(t)/Y(t) = 0$  si  $J(t) = 0$ .

**Remarque 2** Soient  $\left( \begin{matrix} T_{(1)} \\ \delta_{(1)} \end{matrix} \right) < \dots < \left( \begin{matrix} T_{(n)} \\ \delta_{(n)} \end{matrix} \right)$  les statistiques d'ordre de l'échantillon ordonnées selon les valeurs de  $T$ , alors

$$\hat{A}(t) = \sum_{j: T_{(j)} \leq t} \frac{\delta_{(j)}}{Y(T_{(j)})} = \sum_{j: T_{(j)} \leq t} \frac{\delta_{(j)}}{n+1-j}.$$

La fonction  $\hat{A}$  est constante par morceaux, croissante, continue à droite, de saut  $1/Y(T_{(j)})$  au temps de saut  $T_{(j)}$  de  $N$ .

**Remarque 3** On peut montrer que l'estimateur de Nelson-Aalen peut être vu comme un estimateur non-paramétrique du maximum de vraisemblance. La vraisemblance peut s'écrire

$$\mathcal{L} = \prod_{t \leq \tau} \left( \prod_{i=1}^n (Y_i(t) \Delta A(t))^{\Delta N_i(t)} \right) \exp \left( - \sum_{i=1}^n \sum_{t \leq \tau} Y_i(t) \Delta A(t) \right).$$

Le maximum de la vraisemblance est obtenu pour

$$\Delta \hat{A}(t) = \frac{\sum_{i=1}^n \Delta N_i(t)}{\sum_{i=1}^n Y_i(t)} = \frac{\Delta N(t)}{Y(t)}.$$

On en déduit un estimateur du risque cumulé équivalent à l'estimateur de Nelson-Aalen,

$$\hat{A}(t) = \sum_{s \leq t} \Delta \hat{A}(s) = \sum_{s \leq t} \frac{\Delta N(s)}{Y(s)}.$$

### 3.3 Biais et variance

Lorsque  $n$  devient grand, la limite de  $\hat{A}(t)$  doit être une fonction de risque cumulé si  $t$  est plus petit que la dernière durée observée. Ainsi, on définit

$$A^*(t) = \int_0^t J(s) \alpha(s) ds$$

où  $J(t) = \mathbf{1}_{\{Y(t) > 0\}}$ . On a ensuite la proposition suivante

**Proposition 16** On a pour tout  $t \geq 0$

$$\begin{aligned} \hat{A}(t) - A^*(t) &= \int_0^t \frac{J(s)}{Y(s)} dN(s) - \int_0^t J(s) \alpha(s) ds \\ &= \int_0^t \frac{J(s)}{Y(s)} (dN(s) - \alpha(s) Y(s) ds) = \int_0^t \frac{J(s)}{Y(s)} dM(s), \end{aligned}$$

où  $\frac{J}{Y}$  est un processus prévisible localement borné (car  $Y$  est prévisible) et  $M$  une martingale locale de carré intégrable. De plus, on a

- $\hat{A} - A^*$  est une martingale locale de carré intégrable et  $A^*$  est le compensateur de  $\hat{A}$ ,
- $\mathbb{E}(\hat{A}(t) - A^*(t)) = 0$ .

**Proposition 17** Le biais de  $\hat{A}$  est

$$\begin{aligned} \mathbb{E}(\hat{A}(t)) - A(t) &= - \int_0^t \alpha(s) \mathbb{P}(Y(s) = 0) ds \\ &= - \int_0^t \alpha(s) (1 - \bar{F}(s^-) \bar{G}(s^-))^n ds. \end{aligned}$$

Le biais de  $\hat{A}$  est asymptotiquement nul.

On a vu que  $\mathbb{E}(\hat{A}(t)) = \mathbb{E}(A^*(t))$ . De plus  $\mathbb{E}(A^*(t)) = \int_0^t \mathbb{E}(J(s))\alpha(s)ds = \int_0^t \mathbb{P}(Y(s) > 0)\alpha(s)ds$ , on peut donc écrire

$$\mathbb{E}(\hat{A}(t)) - A(t) = \int_0^t (\mathbb{P}(Y(s) > 0) - 1)\alpha(s)ds = - \int_0^t \mathbb{P}(Y(s) = 0)\alpha(s)ds.$$

De plus

$$\begin{aligned} \mathbb{P}(Y(s) = 0) &= \mathbb{P}(T_i < s, \forall i = 1, \dots, n) = \mathbb{P}(T_i < s)^n = (1 - \mathbb{P}(T_i \geq s))^n \\ &= (1 - \mathbb{P}(X_i \geq s, C_i \geq s))^n = (1 - \bar{F}(s)\bar{G}(s))^n. \end{aligned}$$

On considère les  $t \in [0, \tau_H]$  donc  $\bar{F}(t)\bar{G}(t) > 0$  et  $0 < 1 - \bar{F}(s)\bar{G}(s) < 1$ .

On peut montrer les résultats suivant,

**Proposition 18**  $\langle \hat{A} - A^* \rangle (t) = \int_0^t \frac{J(s)}{Y^2(s)} d\langle M \rangle (s) = \int_0^t \frac{J(s)}{Y^2(s)} d\Lambda(s) = \int_0^t \frac{J(s)}{Y(s)} \alpha(s) ds.$

**Proposition 19**  $\mathbb{E} \left( (\hat{A}(t) - A^*(t))^2 \right) = \int_0^t \mathbb{E} \left( \frac{J(s)}{Y(s)} \right) \alpha(s) ds.$

En effet,  $\langle \hat{A} - A^* \rangle$  est le compensateur de  $(\hat{A} - A^*)^2$ . On a donc  $\mathbb{E} \left( (\hat{A}(t) - A^*(t))^2 \right) = \mathbb{E} \left( \langle \hat{A} - A^* \rangle (t) \right) = \int_0^t \mathbb{E} \left( \frac{J(s)}{Y(s)} \right) \alpha(s) ds.$

On a montré que  $\mathbb{E}(\hat{A}(t)) \simeq A(t)$  quand  $n$  est grand. De plus, on peut montrer que  $\forall t \in [0, \tau_H], A^*(t) - A(t) \xrightarrow[n \rightarrow \infty]{} 0$  car  $\mathbb{P}(J(t) = 1) \xrightarrow[n \rightarrow \infty]{} 1$ . Finalement, la variance de  $\hat{A}$  vérifie

$$\begin{aligned} \mathbb{V}(\hat{A}(t)) &= \mathbb{E} \left( (\hat{A}(t) - \mathbb{E}(\hat{A}(t)))^2 \right) \underset{n \rightarrow \infty}{\simeq} \mathbb{E} \left( (\hat{A}(t) - A(t))^2 \right) \underset{n \rightarrow \infty}{\simeq} \mathbb{E} \left( (\hat{A}(t) - A^*(t))^2 \right) \\ &= \int_0^t \mathbb{E} \left( \frac{J(s)}{Y(s)} \right) \alpha(s) ds = \int_0^t \mathbb{E} \left( \frac{J(s)}{Y(s)} \right) \frac{1}{Y(s)} d\Lambda(s). \end{aligned}$$

Comme  $\int_0^t d\Lambda(s) = \int_0^t \alpha(s)Y(s)ds = N(t) - M(s)$  où  $M$  est une martingale, on propose l'estimateur suivant

**Proposition 20** Un estimateur de la variance de  $\hat{A}$  est

$$\hat{\sigma}_{\hat{A}}^2(t) = \int_0^t \frac{J(s)}{[Y(s)]^2} dN(s).$$

### 3.4 Propriétés asymptotiques

Ces estimateurs vérifient les propriétés asymptotiques ci-dessous

**Théorème 3 Consistance uniforme** de  $\hat{A}$ . On a

$$\sup_{s \in [0, t]} \left| \hat{A}(s) - A(s) \right| \xrightarrow{\mathbb{P}} 0,$$

car les conditions suivantes sont vérifiées

$$\int_0^t \frac{J(s)}{Y(s)} \alpha(s) ds \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (1)$$

$$\int_0^t (1 - J(s)) \alpha(s) ds \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (2)$$

On peut montrer, en appliquant l'inégalité de Lengart, que les conditions (1) et (2) impliquent la consistance uniforme.

**Théorème 4 Convergence faible** de  $\hat{A}$ . Soit  $t \in [0, \tau_H]$ , on pose  $\sigma^2(s) = \int_0^s \frac{\alpha(u)}{F(u)G(u)} du$ . Comme les conditions suivantes, sont vérifiées

$$\forall s \in [0, t], \quad n \int_0^s \frac{J(u)}{Y(u)} \alpha(u) du \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2(s) \quad (1)$$

$$\forall s \in [0, t], \quad \sqrt{n} \int_0^s (1 - J(u)) \alpha(u) du \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (2)$$

$$\forall \varepsilon > 0, \quad n \int_0^s \frac{J(u)}{Y(u)} \alpha(u) \mathbb{1}_{\left\{ \sqrt{n} \frac{J(u)}{Y(u)} > \varepsilon \right\}} du \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (3)$$

on a

$$\sqrt{n}(\hat{A} - A) \xrightarrow{D} U$$

où  $U$  est une martingale gaussienne telle que  $U(0) = 0$  et  $Cov(U(s), U(t)) = \sigma^2(s \wedge t)$ . De plus, on a

$$\sup_{s \in [0, t]} \left| n \hat{\sigma}_{\hat{A}}^2(t) - \sigma^2(s) \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

On peut montrer, en appliquant notamment le théorème de Rebolledo, que les conditions (1), (2) et (3) impliquent la convergence faible.

## 4 Estimateur de Kaplan-Meier

### 4.1 Estimateur de $S$

**Proposition 21** Soit  $\forall t \geq 0$ ,  $S(t) = \bar{F}(t) = \mathbb{P}(X > t)$  et  $A(t) = - \int_0^t \frac{d\bar{F}(s)}{F(s-)}$ , alors

$$\bar{F}(t) = \mathcal{P}_{0 \leq s \leq t} (1 - dA(s)), \quad \forall t \text{ tel que } A(t) < \infty \text{ (i.e. } \forall t \in [0, \tau_F]).$$

**Preuve** On remarque que  $\forall t \in [0, \tau_F[, \bar{F}(t-) > 0$  et  $A(t) < \infty$ . En différenciant  $A$ ,

$$\begin{aligned} dA(t) = -\frac{d\bar{F}(t)}{\bar{F}(t-)} &\iff \bar{F}(t-)dA(t) = -d\bar{F}(t) \implies \int_0^t d\bar{F}(s) = -\int_0^t \bar{F}(s-)dA(s) \\ &\implies \bar{F}(t) = 1 - \int_0^t \bar{F}(s-)dA(s). \end{aligned}$$

L'équation précédente est une équation de Volterra où  $Z = \bar{F}$  et  $X = -A$ . Donc  $\forall t \in [0, \tau_F[, \bar{F}(t-) > 0$ , la solution de cette équation est donnée par le produit intégral. ■

**Remarque 4** Si  $T$  est une variable continue, elle admet une densité  $f$ . Dans ce cas,  $d\bar{F}(s) = -f(s)ds$  et  $\bar{F}(s-) = \bar{F}(s)$ , on a alors que

$$A(t) = \int_0^t \frac{f(s)}{\bar{F}(s-)} ds = \int_0^t \alpha(s) ds$$

où  $A$  est la fonction de risque cumulé. De plus, d'après la définition du produit intégral, on a

$$\bar{F}(t) = \mathcal{P}_{0 \leq s \leq t} (1 - dA(s)) = \exp\left(-\int_0^t dA^C(s)\right) = \exp\left(-\int_0^t \alpha(s) ds\right).$$

On cherche à estimer la fonction  $S(t) = \bar{F}(t) = \mathbb{P}(T > t) = \mathcal{P}_{0 \leq s \leq t} (1 - dA(s))$  où  $A(t) = \int_0^t \alpha(s) ds$  et  $A(t) < \infty$  sur  $[0, \tau_F[$  à partir des données observables

$$\begin{cases} T_i = X_i \wedge C_i, & \forall i = 1, \dots, n, X_i \text{ indépendant de } C_i \\ \delta_i = \mathbb{1}_{\{X_i \leq C_i\}} \end{cases}$$

On va utiliser l'estimateur de Nelson-Aalen et la propriété précédente pour obtenir l'estimateur de Kaplan-Meier. En effet, d'après la définition du produit intégral, comme  $\hat{A}$  est une fonction càdlàg, croissante et à variation localement bornée ( $\hat{A}(t) - \hat{A}(t-) \leq n$ ), on peut donc définir

**Définition 18** L'estimateur de **Kaplan-Meier** de la fonction de survie est

$$\hat{S}(t) = \mathcal{P}_{0 \leq s \leq t} (1 - d\hat{A}(s)), \quad \forall t \in [0, \tau_F[ \tag{II.3}$$

où  $\hat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s)$  est l'estimateur de Nelson-Aalen ;  $J(t) = \mathbb{1}_{\{Y(t) > 0\}}$ . On utilise la convention  $J(t)/Y(t) = 0$  si  $J(t) = 0$ .

**Remarque 5** Soit  $\left(\begin{smallmatrix} T_{(1)} \\ \delta_{(1)} \end{smallmatrix}\right) < \dots < \left(\begin{smallmatrix} T_{(n)} \\ \delta_{(n)} \end{smallmatrix}\right)$  les statistiques d'ordre de l'échantillon ordonnées selon les valeurs de  $T$ . L'estimateur  $\hat{A}$  étant une fonction constante par morceaux,

*l'estimateur de Kaplan-Meier devient un produit fini*

$$\begin{aligned}\hat{S}(t) &= \mathcal{P}_{0 \leq s \leq t} (1 - \Delta \hat{A}(s)) = \prod_{0 \leq s \leq t} \left(1 - \frac{\Delta N(s)}{Y(s)}\right) \\ &= \prod_{j: T(j) \leq t} \left(1 - \frac{\Delta N(T(j))}{Y(T(j))}\right) = \prod_{j: T(j) \leq t} \left(1 - \frac{\delta_{(j)}}{Y(T(j))}\right) = \prod_{j: T(j) \leq t} \left(1 - \frac{\delta_{(j)}}{n+1-j}\right), \quad \forall t \geq 0.\end{aligned}$$

**Remarque 6** *On montre que l'estimateur de Kaplan-Meier est un estimateur non-paramétrique du maximum de vraisemblance. La modélisation des durées de vie peut s'exprimer en fonction des deux paramètres suivants qui sont liés par les relations suivantes*

$$\begin{aligned}A(t) &= - \int_0^t \frac{dS(s)}{S(s-)}, \\ S(t) &= \mathcal{P}_{0 \leq s \leq t} (1 - dA(s)).\end{aligned}$$

*Comme  $\hat{A}$  est un estimateur non paramétrique du maximum de vraisemblance de  $A$ , la propriété d'invariance des estimateurs du maximum de vraisemblance permet de conclure que  $\hat{S}$  est également un estimateur non paramétrique du maximum de vraisemblance de  $S$ .*

## 4.2 Biais et variance

On introduit la quantité

$$\bar{F}^*(t) = S^*(t) = \mathcal{P}_{0 \leq s \leq t} (1 - dA^*(s)), \quad \forall t \geq 0,$$

où  $A^*(s) = \int_0^s J(u)\alpha(u)du$ .

**Proposition 22** On a pour tout  $t \geq 0$

$$\begin{aligned}\frac{\hat{S}(t)}{S^*(t)} - 1 &= - \int_0^t \frac{\hat{S}(s-)}{S^*(s)} d(\hat{A} - A^*)(s) \\ &= - \int_0^t \frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y(s)} dM(s),\end{aligned}$$

où  $M$  est une martingale locale de carré intégrable telle que pour tout  $s \geq 0$ ,  $M(t) = N(t) - \int_0^t \alpha(s)Y(s)ds$ . De plus,  $\frac{\hat{S}(t)}{S^*(t)} - 1$  est une martingale locale de carré intégrable sur  $[0, \tau_F[$  telle que

$$\mathbb{E} \left( \frac{\hat{S}(t)}{S^*(t)} \right) = 1.$$

La démonstration de cette propriété utilise l'équation de Duhamel.

**Proposition 23** Le biais de  $\hat{S}$  vérifie les inégalités suivantes

$$0 \leq \mathbb{E}(\hat{S}(t)) - S(t) \leq F(t) (1 - S(t)\bar{G}(t))^n \leq F(t) \exp(-\mathbb{E}(Y(t))).$$

La dernière inégalité est obtenue en remarquant que  $\mathbb{E}(Y(t)) = n\mathbb{P}(T_i \geq t) = nS(t)\bar{G}(t)$  et que  $\log(1-x) \leq -x$ . Le biais de l'estimateur de Kaplan-Meier est faible quand  $\mathbb{E}(Y(t))$  est grande, *i.e.* au niveau des premières observations.

On cherche à obtenir un estimateur de la variance de  $\hat{S}$ . Comme  $\frac{\hat{S}(\cdot)}{S^*(\cdot)} - 1$  est une martingale locale de carré intégrable, on peut considérer la décomposition suivante

$$\left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right)^2 = \left\langle \frac{\hat{S}(t)}{S^*(t)} - 1 \right\rangle + \xi_t,$$

où  $\xi_t$  est une martingale locale de carré intégrable et  $\mathbb{E} \left( \left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right)^2 \right) = \mathbb{E} \left( \left\langle \frac{\hat{S}(t)}{S^*(t)} - 1 \right\rangle \right)$ .

On a donc

$$\mathbb{V} \left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right) = \mathbb{E} \left( \left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right)^2 \right) = \mathbb{V} \left( \frac{\hat{S}(t)}{S^*(t)} \right)$$

avec

$$\begin{aligned} \mathbb{E} \left( \left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right)^2 \right) &= \mathbb{E} \left( \left\langle - \int_0^t \frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y(s)} dM(s) \right\rangle \right) \\ &= \mathbb{E} \left( \int_0^t \left( \frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y(s)} \right)^2 d \langle M \rangle (s) \right) \\ &= \mathbb{E} \left( \int_0^t \left( \frac{\hat{S}(s-)}{S^*(s)} \frac{J(s)}{Y(s)} \right)^2 J(s) Y(s) \alpha(s) ds \right). \end{aligned}$$

En remplaçant  $\hat{S}(s-)$  par  $\hat{S}(s)$  et  $\alpha(s)Y(s)ds$  par  $dN(s)$  on montre que

$$\mathbb{V} \left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right) \simeq (S(t))^2 \int_0^t \frac{J(s)}{(Y(s))^2} dN(s).$$

**Proposition 24** Un estimateur de la variance de  $\hat{S}$  est

$$\hat{\sigma}_{\hat{S}}^2(t) = \left( \frac{\hat{S}(t)}{S^*(t)} - 1 \right)^2 \int_0^t \frac{J(s)}{(Y(s))^2} dN(s).$$

### 4.3 Propriétés asymptotiques

Ces estimateurs vérifient les propriétés asymptotiques ci-dessous

**Théorème 5 Consistance uniforme** de  $\hat{S}$ . Soit  $t \in [0, \tau_H]$ , on a

$$\sup_{s \in [0, t]} \left| \hat{S}(s) - S(s) \right| \xrightarrow{\mathbb{P}} 0,$$

car les conditions suivantes sont vérifiées

$$\int_0^t \frac{J(s)}{Y(s)} \alpha(s) ds \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (1)$$

$$\int_0^t (1 - J(s)) \alpha(s) ds \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (2)$$

On peut montrer, en appliquant l'inégalité de Lengart, que les conditions (1) et (2) impliquent la consistance uniforme.

**Théorème 6 Convergence faible** de  $\hat{A}$ . Soit  $t \in [0, \tau_H]$ , on pose  $\sigma^2(s) = \int_0^s \frac{\alpha(u)}{S(u)G(u)} du$ . Comme les conditions suivantes, sont vérifiées

$$\forall s \in [0, t], \quad n \int_0^s \frac{J(u)}{Y(u)} \alpha(u) du \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2(s) \quad (1)$$

$$\forall s \in [0, t], \quad \sqrt{n} \int_0^s (1 - J(u)) \alpha(u) du \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (2)$$

$$\forall \varepsilon > 0, \quad n \int_0^s \frac{J(u)}{Y(u)} \alpha(u) \mathbb{1}_{\left\{ \sqrt{n} \frac{J(u)}{Y(u)} > \varepsilon \right\}} du \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (3)$$

on a

$$\sqrt{n}(\hat{S} - S) \xrightarrow{D} -SU$$

où  $U$  est une martingale gaussienne telle que  $U(0) = 0$  et  $Cov(U(s), U(t)) = \sigma^2(s \wedge t)$ . De plus, on a

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

On peut montrer que les conditions (1), (2) et (3) sont vérifiées. On montre ensuite, en appliquant notamment le théorème de Rebolledo, que ces conditions impliquent la convergence faible.

## 5 Estimation semi-paramétrique

Cette section, présente un modèle de régression permettant d'ajuster l'intensité sur des covariables.

En plus des notations des sections précédentes on définit pour  $i \in \{1, \dots, n\}$ ,  
 —  $\mathbf{Z}_i = (Z_{1i}, \dots, Z_{pi})$ , le vecteur de covariables de dimension  $p$ ,

—  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ , le vecteur de dimension  $p$  des coefficients de régression,

Pour tout  $i = 1, \dots, n$ , l'intensité  $\lambda_i$  du processus  $N_i$  est

$$\lambda_i(t) = Y_i(t) \alpha_i(t),$$

où  $\alpha_i$  est le risque instantané spécifique à l'individu  $i$ . Pour prendre en compte les covariables  $\mathbf{Z}_i$  on suppose que les risques instantanés  $\alpha_i$  suivent un modèle semi-paramétrique à risque multiplicatif de "Cox", c'est-à-dire,

$$\alpha_i(t | \mathbf{Z}_i) = \alpha_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i), \quad i = 1, \dots, n.$$

où  $\alpha_0(t)$  est le risque de base. Notons que le modèle fait des hypothèses de risques proportionnels et de log-linéarité.

## 5.1 Estimation du risque cumulé de base

Dans le cadre d'un modèle semi-paramétrique multiplicatif, la vraisemblance partielle peut s'écrire

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{t \leq \tau} \prod_{i=1}^n (\Delta A_0(t) Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i))^{\Delta N_i(t)} \exp \left[ - \sum_{i=1}^n \sum_{t \leq \tau} Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i) \Delta A_0(t) \right].$$

où  $A_0(t) = \int_0^t \alpha_0(s) ds$ . En considérant  $\boldsymbol{\beta}$  fixé, la maximisation de la vraisemblance par rapport à  $\Delta A_0(\cdot)$  conduit à

$$\Delta \hat{A}_0(t) = \frac{\Delta N(t)}{\sum_{i=1}^n Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}.$$

**Proposition 25** Pour  $\boldsymbol{\beta}$  fixé,  $A_0(t) = \int_0^t \alpha_0(s) ds$  est estimé par l'estimateur de Breslow,

$$\hat{A}_0(t) = \int_0^t \frac{J(s)}{\sum_{i=1}^n Y_i(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)} dN(s), \quad (\text{II.4})$$

avec  $J(u) = \mathbf{1}_{\{Y(u) > 0\}}$ .

## 5.2 Estimation des coefficients de régression

En remplaçant,  $A_0(t)$  par son estimateur, la vraisemblance partielle devient,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}) &= \prod_{t \leq \tau} \prod_{i=1}^n \left( \frac{Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{\sum_{j=1}^n Y_j(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_j)} \right)^{\Delta N_i(t)} \prod_{t \leq \tau} \prod_{i=1}^n (J(t) dN(t))^{\Delta N_i(t)} \exp \left[ - \int_0^\tau J(s) dN(s) \right] \\ &= \mathcal{L}_{Cox}(\boldsymbol{\beta}) \times \prod_{t \leq \tau} \prod_{i=1}^n (J(t) dN(t))^{\Delta N_i(t)} \exp \left[ - \int_0^\tau J(s) dN(s) \right]. \end{aligned}$$

Par définition,  $\mathcal{L}_{Cox}(\boldsymbol{\beta})$  est la **vraisemblance partielle de Cox**,

$$\mathcal{L}_{Cox}(\boldsymbol{\beta}) = \prod_{t \leq \tau} \prod_{i=1}^n \left( \frac{Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{\sum_{j=1}^n Y_j(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_j)} \right)^{\Delta N_i(t)}. \quad (\text{II.5})$$

Cette vraisemblance est introduite par Cox (1972) mais par une approche différente. La vraisemblance partielle de Cox n'est pas une vraisemblance dans le sens statistique du terme, mais il est établi qu'elle peut être utilisée comme telle pour estimer les coefficients de régression.

La fonction de log-vraisemblance partielle de Cox est,

$$\log \mathcal{L}_{Cox}(\boldsymbol{\beta}) = \sum_i \int_0^\tau \left[ \boldsymbol{\beta}^T \mathbf{Z}_i - \log \left( \sum_{j=1}^n Y_j(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_j) \right) \right] dN_i(t) + cste.$$

Le vecteur score (dérivées de la Log-Vraisemblance par rapport à  $\boldsymbol{\beta}$ ) est donné par

$$U(\boldsymbol{\beta}) = \frac{\partial \log \mathcal{L}_{Cox}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_i \int_0^\tau \left[ \mathbf{Z}_i - \frac{\sum_{j=1}^n Y_j(t) \mathbf{Z}_j \exp(\boldsymbol{\beta}^T \mathbf{Z}_j)}{\sum_{j=1}^n Y_j(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_j)} \right] dN_i(t),$$

**Proposition 26** L'estimateur  $\hat{\boldsymbol{\beta}}$  du maximum de la vraisemblance partielle de Cox vérifie

$$U(\hat{\boldsymbol{\beta}}) = 0.$$

En pratique, les estimations des coefficients de régression sont obtenues par maximisation de la log-vraisemblance à l'aide d'algorithmes itératifs, comme par exemple l'algorithme de quasi-Newton.

**Remarque 7** Pour simplifier les écritures, on considère des covariables indépendantes du temps néanmoins le modèle de Cox permet de prendre en compte des covariables dépendante du temps. En effet, les raisonnements précédents peuvent facilement être adaptés à des covariables dépendantes du temps.

**Remarque 8** Les vraisemblances successivement utilisées sont résumées par le schéma suivant :

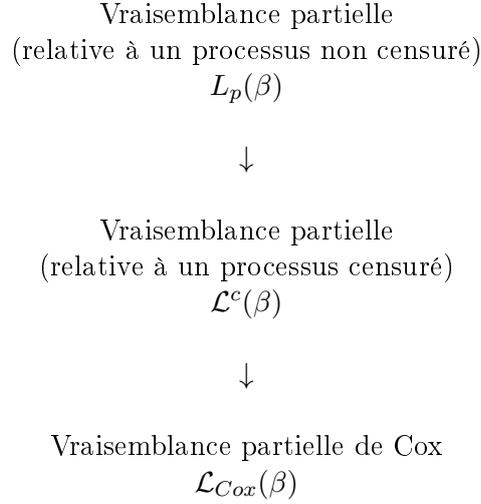


TABLE II.1 – Les vraisemblances successives.

### 5.3 Estimation de la survie

A partir des estimateurs précédents (du risque de base et des coefficients), la fonction de survie s'obtient par le produit intégral. Soit  $\mathbf{Z}_0$ , la valeur des covariables pour un individu.

**Proposition 27** Un estimateur de la fonction de survie est donné par le produit intégral

$$\hat{S}(t | \mathbf{Z}_0) = \mathcal{P}_{s \in ]0, t]} \left( 1 - d\hat{A}(s | \mathbf{Z}_0) \right), \quad \forall t \in [0, \tau[,$$

avec

$$\hat{A}(t | \mathbf{Z}_0) = \int_0^t \frac{J(s) \times \exp(\hat{\boldsymbol{\beta}}^T \mathbf{Z}_0)}{\sum_{i=1}^n \exp(\hat{\boldsymbol{\beta}}^T \mathbf{Z}_i) Y_i(s)} dN(s).$$

### 5.4 Propriétés asymptotiques

Pour démontrer les propriétés asymptotiques, on utilise le fait que  $M_i(t) = N_i(t) - \int_0^t Y_i(s) \alpha_o(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i) ds$  est une martingale locale par rapport à  $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$  avec  $\mathcal{F}_t = \sigma((N_i(s), Y_i(s), \mathbf{Z}_i), i = 1, \dots, n, s \leq t)$ .

**Proposition 28** L'estimateur  $\hat{\boldsymbol{\beta}}$  du maximum de la vraisemblance partielle de Cox vérifie

$$\sqrt{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} N(0, nI^{-1}(\hat{\boldsymbol{\beta}})),$$

où  $\beta_0$  est le vrai paramètre et  $I(\beta)$  est la matrice d'information de Fisher :

$$\begin{aligned} I(\beta) &= -\mathbb{E} \left( \frac{\partial^2 \log \mathcal{L}_{Cox}(\beta)}{\partial \beta^2} \right) = - \left\{ \mathbb{E} \left( \frac{\partial^2 \log \mathcal{L}_{Cox}(\beta)}{\partial \beta_i \partial \beta_j} \right) \right\}_{p \times p} \\ &= -\mathbb{E} \left( \int_0^\tau \frac{[\sum_{i=1}^n Y_i(t) \mathbf{Z}_i \mathbf{Z}_i^T \exp(\beta^T \mathbf{Z}_i)] - [\sum_{i=1}^n Y_i(t) \mathbf{Z}_i \exp(\beta^T \mathbf{Z}_i)] [\sum_{i=1}^n Y_i(t) \mathbf{Z}_i \exp(\beta^T \mathbf{Z}_i)]^T}{[\sum_{i=1}^n Y_i(t) \exp(\beta^T \mathbf{Z}_i)]^2} dN(t) \right) \end{aligned}$$

L'inverse de la matrice d'information de Fisher étant égale à la matrice de variance-covariance, elle fournit une estimation de la variance de  $\hat{\beta}$  en considérant la matrice d'information observée au point  $\beta = \hat{\beta}$ , *i.e.*  $\mathcal{I}(\hat{\beta}) = - \left[ \frac{\partial^2 \log \mathcal{L}_{Cox}(\beta)}{\partial \beta^2} \right]_{\beta=\hat{\beta}}$ .

On peut ensuite obtenir un estimateur de la variance de  $\hat{S}$  et les propriétés asymptotiques de  $\hat{S}$  à partir de celles de  $\hat{\beta}$  et  $\hat{A}_0(t)$  et de l'utilisation de la delta-méthode.



## Chapitre III

# Modèles multi-états Markovien

Dans l'analyse de données longitudinales (mesures répétées dans le temps pour un même individu), les sujets sont souvent suivis par intermittence et l'information recueillie se présente sous la forme de mesures en plusieurs temps discrets. Ces mesures pour un même individu ne peuvent pas être considérées comme étant indépendantes puisque la même variable est mesurée plusieurs fois pour un même individu. Une autre spécificité de ces données est que le temps entre deux mesures est variable et que les temps de mesures ne sont pas les mêmes pour tous les individus.

Les modèles multi-états permettent de modéliser ce type de données. En effet, les variables mesurées pour chaque individu permettent de définir un "état" de l'individu à chaque temps d'observation. A partir d'une séquence d'"état" et d'une séquence de temps, les modèles multi-états permettent de représenter le suivi d'un individu comme une succession dans le temps d'"état". En épidémiologie par exemple, ils permettent de représenter l'évolution d'un patient à travers les différents stades d'une maladie. On cherche ensuite à estimer les paramètres liés aux changements d'"état", *i.e.* les paramètres qui renseignent sur l'évolution du processus.

Dans les modèles multi-états, on fait souvent une hypothèse Markovienne qui permet de résumer le passé des individus. En effet, dans de nombreuses situations le passé de l'individu est trop compliqué et on est amené à le résumer (c'est le cas quand un individu peut revenir dans un état déjà visité). L'hypothèse de Markov considère que l'évolution future du processus dépend uniquement de l'état du processus au temps  $t$ , autrement dit, l'histoire du processus est résumée par l'état au temps  $t$ . Dans ce qui suit, on s'intéressera aux modèles multi-états à temps continu qui permettent de travailler avec des temps d'observation différents d'un individu à l'autre et des durées entre deux observations variables.

## 1 Processus de Markov

Un processus de Markov  $\{X(t); t \in \mathcal{T}\}$  à temps continu et à espace d'états fini est un processus dont l'évolution future  $\{X(t); t \geq s\}$  ne dépend de son passé qu'à travers son état à l'instant  $s$ , pour tout  $t \geq s$

$$\mathbb{P}\{X(t) = j \mid X(r) = x_r, r \leq s\} = \mathbb{P}\{X(t) = j \mid X(s) = x_s\}.$$

Cette définition signifie que tout le passé du processus est résumé dans l'état précédent ou encore, le présent étant connu, le futur est indépendant du passé.

**Définition 19** Un **processus de Markov** à temps continu et à espace d'états fini  $S = \{1, \dots, k\}$  est complètement défini par

1. Son vecteur des probabilités initiales, noté  $\mathbf{P}_0$  tel que

$$\mathbf{P}_0[j] = \mathbb{P}\{X(0) = j\}, \quad j = 1, \dots, k$$

avec  $\sum_{j=1}^k \mathbb{P}\{X(0) = j\} = 1$ ,

2. Sa matrice de **probabilités de transition** :  $\mathbf{P}(s, t) = (p_{hj}(s, t))_{h,j}$  telle que

$$p_{hj}(s, t) = \mathbb{P}\{X(t) = j \mid X(s) = h\} \quad \forall s \leq t \in \mathcal{T} \text{ et } h, j \in S,$$

avec

$$\mathbf{P}(s, s) = \mathbf{Id}, \text{ et } \sum_{j=1}^k p_{hj}(s, t) = 1 \text{ pour tout } h \text{ et } 0 \leq s \leq t.$$

Les **intensités de transition** permettent également de définir un processus de Markov. Soit  $\mathbf{Q} = (\alpha_{hj}(\cdot))_{i,j}$  la matrice  $k \times k$  des intensités de transition,

$$\begin{aligned} \alpha_{hj}(t) &= \lim_{\Delta t \rightarrow 0} \frac{p_{hj}(t, t + \Delta t) - \delta_{hj}}{\Delta t}, \quad h \neq j, \\ \alpha_{hh}(t) &= - \sum_{j \neq h} \alpha_{hj}(t), \quad h = 1, \dots, k. \end{aligned}$$

La matrice des **intensités cumulées**  $\mathbf{A}$  permet également de définir un processus de Markov,

$$\mathbf{A} = \{A_{hj}\}_{h,j} \text{ avec } A_{hh}(t) = - \sum_{j \neq h} A_{hj}(t), \quad \text{pour tout } t,$$

et pour tout  $h \neq j$ ,  $A_{hj}(\cdot)$  est une fonction càdlàg (continue à droite avec une limite à gauche) non décroissante, nulle en zéro.

On suppose généralement que les fonctions  $A_{hj}(\cdot)$  sont absolument continues, dans ce cas, on a la relation suivante

$$A_{hj}(t) = \int_0^t \alpha_{hj}(u) du.$$

où les  $\alpha_{hj}(\cdot)$  sont les intensités de transition.

Les probabilités de transition d'un processus Markovien vérifient la relation suivante,  $\forall i, j \in S = \{1, \dots, k\}$  et  $\forall 0 < s < u < t$ ,

$$p_{hj}(s, t) = \sum_{k \in S} p_{hk}(s, u) p_{kj}(u, t), \quad (\text{III.1})$$

Cette propriété est appelée équation de Chapman-Kolmogorov. Sous forme matricielle, la relation (III.1) devient

$$\mathbf{P}(s, t) = \mathbf{P}(s, u)\mathbf{P}(u, t) \quad \forall s \leq u \leq t.$$

On peut en déduire les équations différentielles de Kolmogorov qui définissent le lien entre la matrice de probabilité de transition et la matrice d'intensité cumulée

— équation « forward » de Kolmogorov

$$\frac{\partial \mathbf{P}(s, t)}{\partial t} = \mathbf{P}(s, t) \mathbf{A}(dt), \quad (\text{III.2})$$

— équation « backward » de Kolmogorov

$$\frac{\partial \mathbf{P}(s, t)}{\partial s} = \mathbf{A}(ds) \mathbf{P}(s, t). \quad (\text{III.3})$$

La notion de **produit intégral** permet de relier la matrice des intensités de transition et la matrice des probabilités de transition. Dans ce qui suit, la notation  $\mathcal{P}_{]s,t]}$  suggère une version continue du produit ordinaire comme l'intégrale  $\int$  généralise la somme  $\sum$ .

**Théorème 7** Soit  $\mathbf{A}$  une matrice de fonctions de dimension  $k \times k$  correspondant à une mesure d'intensité (matrice des intensités cumulées).

Alors la matrice,

$$\mathbf{P}(s, t) = \mathcal{P}_{u \in ]s,t]} (\mathbf{Id} + \mathbf{A}(du)), \quad s \leq t, \quad t, s \in \mathcal{T},$$

est la matrice de probabilité de transition d'un processus de Markov à espace d'états fini  $\{1, \dots, k\}$ .

Le processus peut être construit de la façon suivante

— Sachant que le processus est dans  $h$  au temps  $t_0$ , il reste dans l'état  $h$  pour une durée avec une intensité cumulée

$$-(A_{hh}(t) - A_{hh}(t_0)), \quad t_0 \leq t \leq \inf \{u \geq t_0 : \Delta A_{hh}(u) = -1\}.$$

— Sachant qu'il quitte  $h$  au temps  $t$ , il transite vers l'état  $j$  ( $j \neq h$ ) avec une probabilité

$$-\frac{dA_{hj}(t)}{dA_{hh}(t)}.$$

Quand la fonction  $\mathbf{A}$  est une fonction en escalier (cadlag), le produit intégral devient un produit fini sur les temps de sauts de  $\mathbf{A}$ ; ainsi

$$\mathbf{P}(s, t) = \prod_{i=1}^n (\mathbf{Id} + \Delta \mathbf{A}(T_i)), \quad s \leq t,$$

où  $s < T_1 < \dots < T_n = t$  sont les temps de sauts et  $\Delta \mathbf{A}(T_k) = \mathbf{A}(T_k) - \mathbf{A}(T_{k-1})$  et  $T_0 = 0$ .

## 2 Modèle Markovien homogène

Dans les modèles de Markov, on peut faire différentes hypothèses sur l'évolution du processus. En particulier, les paramètres du modèle peuvent dépendre ou non de la durée totale du suivi : on parle alors de modèle non-homogène et de modèle homogène. L'hypothèse d'homogénéité qui rend les forces de transition constantes au cours du temps simplifie la modélisation et la méthodologie. Néanmoins, les intensités sont supposées constantes sur une longue période ce qui est restrictif dans de nombreuses applications. On peut alors ajuster un modèle à intensités constantes par périodes qui permet de considérer un modèle non-homogène tout en conservant l'hypothèse d'homogénéité au sein d'une même période.

### 2.1 Processus Markovien homogène

Un processus de Markov est **homogène** si la probabilité de transition de l'état  $h$  vers  $j$  vérifie

$$\begin{aligned} p_{hj}(s, t) &= \mathbb{P}\{X(t) = j \mid X(s) = h\}, \\ &= p_{hj}(0, t - s), \\ &= P_{hj}(t - s). \end{aligned}$$

Les probabilités de transition dépendent uniquement du temps entre deux transitions et non du moment où se produisent ces transitions. Dans ce cas, les intensités de transition sont indépendantes du temps, pour tout  $h \neq j$ ,

$$\alpha_{hj}(s) = \lim_{\Delta t \rightarrow 0} \frac{p_{hj}(s, s + \Delta t) - p_{hj}(s, s)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(\Delta t) - P_{hj}(0)}{\Delta t} = \alpha_{hj}.$$

Les équations différentielles de Kolmogorov permettent d'obtenir la relation suivante

$$\frac{\partial \mathbf{P}(0, t)}{\partial t} = \mathbf{P}(0, t) \mathbf{Q}.$$

Sachant que  $P_{hh}(0) = 1$  et  $P_{hj}(0) = 0$ , la solution de cette équation est donnée par,

$$\mathbf{P}(0, t) = \exp(\mathbf{Q} \times t). \quad (\text{III.4})$$

Le terme  $P_{hh}(\cdot) = p_{hh}(0, t)$  s'écrit

$$\begin{aligned} P_{hh}(t) &= \exp(\alpha_{hh} \times t) \\ &= \exp\left(-\sum_{j \neq h} \alpha_{hj} \times t\right). \end{aligned}$$

Ainsi on montre que la variable aléatoire  $T = \ll \text{temps passé dans l'état } h \text{ avant de le quitter} \gg$  (*i.e.* le temps de séjour dans l'état  $h$ ) suit une loi exponentielle de paramètre  $\sum_{j \neq h} \alpha_{hj}$  (car  $\mathbb{P}(T > t) = P_{hh}(t)$ ). Le temps moyen passé dans l'état  $h$  avant de le quitter vaut  $\mathbb{E}(T) = 1 / \sum_{j \neq h} \alpha_{hj}$ . Dans les modèles Markoviens homogènes, les temps d'attente dans un état suivent des lois exponentielles. Ces lois sont dites sans mémoire car les fonctions de risque associées sont constantes au cours du temps.

## 2.2 Vraisemblance

Soit  $\{X(t), t \in \mathcal{F}\}$  un processus de Markov homogène à espace d'états fini  $S = \{1, \dots, k\}$ . Supposons que pour chaque individu  $i = 1, \dots, n$ , on observe les données suivantes

- les temps successifs d'observation  $T_{i,0} < T_{i,1} < \dots < T_{i,n_i}$
- les états occupés  $x_{i,j} = X(T_{i,j})$ ,  $j = 0, 1, \dots, n_i$ .

La contribution à la vraisemblance pour un individu  $i$ , est le produit des probabilités associées à chaque transition observée,

$$L_i = \mathbb{P}\{X(T_{i,0}) = x_{i,0}\} \times \prod_{j=1}^{n_i} P_{x_{i,j-1}, x_{i,j}}(T_{i,j} - T_{i,j-1}).$$

L'équation (III.4) permet d'exprimer les probabilités en fonction des intensités de transition. La vraisemblance totale est le produit des contributions individuelles,

$$L = \prod_{i=1}^n L_i. \quad (\text{III.5})$$

On peut ensuite estimer les intensités de transitions par la méthode du maximum de vraisemblance.

## 2.3 Modèle à intensités proportionnelles

Considérons un modèle de régression à intensités proportionnelles de "Cox"

$$\alpha_{hj}(\mathbf{z}) = \alpha_{hj0} \exp(\beta'_{hj} \mathbf{Z}) \quad h \neq j, \quad (\text{III.6})$$

avec  $\mathbf{Z}$  un vecteur de covariables,  $\beta_{hj}$  un vecteur de coefficients de régression et  $\alpha_{hj0}$  l'intensité de transition de base associée à la transition de l'état  $h$  vers l'état  $j$ .

Comme dans le modèle de Cox, il est possible de considérer des covariables dépendantes du temps. Il est nécessaire de supposer que les valeurs des covariables dépendantes du temps restent constantes entre les deux temps d'observations consécutifs. Cette hypothèse peut être forte dans certaines situations, notamment, si le temps entre deux consultations est important.

## 2.4 Modèle de Markov homogène par périodes

Dans de nombreuses situations l'hypothèse d'homogénéité est souvent trop restrictive. L'utilisation d'une covariable artificielle dépendante du temps permet de considérer un modèle où les intensités de transition sont constantes au sein d'une même période mais sont différentes d'une période à une autre.

Considérons une subdivision de l'axe du temps  $[t_{l-1}, t_l[$  où  $t_0 = 0 < t_1 < \dots < t_{r+1} = +\infty$  et  $\mathbf{V}(t) = (V_1(t), \dots, V_r(t))'$  un vecteur de covariables défini par

$$V_l(t) = \begin{cases} 0 & \text{si } t_0 \leq t < t_l \\ 1 & \text{si } t \geq t_l \end{cases} \quad \text{pour } l = 1, 2, \dots, r.$$

En considérant un modèle semi-paramétrique de Cox on obtient des intensités constantes par intervalles. Pour tout  $h \neq j$ ,

$$\alpha_{hj}(t \mid \mathbf{V}(t)) = \alpha_{hj0} \exp\{\gamma'_{hj} \mathbf{V}(t)\} \\ = \begin{cases} \alpha_{hj0} = \alpha_{hj0} & \text{if } t_0 \leq t < t_1 \\ \alpha_{hj1} = \alpha_{hj0} \exp\{\gamma_{hj,1}\} & \text{if } t_1 \leq t < t_2 \\ \vdots & \\ \alpha_{hjr} = \alpha_{hj0} \exp\{\gamma_{hj,1} + \gamma_{hj,2} + \dots + \gamma_{hj,r}\} & \text{if } t \geq t_r. \end{cases}$$

La vraisemblance d'un modèle homogène par périodes doit tenir compte des intensités constantes par intervalles. En effet, dans la vraisemblance, les probabilités pour lesquelles les deux temps n'appartiennent pas aux mêmes intervalles doivent être réécrites à l'aide de l'équation de Chapman-Kolmogorov (III.1) Pour tout  $h, j \in S$ , on considère  $p_{hj}^{(l)}(t)$  la probabilité de transition associée à l'intervalle  $[t_{l-1}, t_l[$ . Pour tout  $l = 1, \dots, r + 1$ , on définit

$$p_{hj}(s, s + t) = p_{hj}^{(l)}(t) \quad \text{pour tout } t_{l-1} \leq s < s + t < t_l.$$

On introduit également, pour tout  $t$ , la variable  $I_t \in \{1, 2, \dots, r + 1\}$  qui renseigne sur l'intervalle de temps de la forme  $[t_{l-1}, t_l[$  contenant  $t$ . Considérons  $X_1$  et  $X_2$  les états occupés par un individu pour deux temps consécutifs d'observation  $T_1$  et  $T_2$ , alors la contribution à la vraisemblance s'écrit

$$p_{X_1 X_2}\{T_1, T_2 \mid \mathbf{V}(t), T_1 \leq t < T_2\} = \sum_{k_1 \in S} \dots \sum_{k_v \in S} \left[ p_{X_1 k_1}^{(I_{T_1})}\{t_{I_{T_1}} - T_1 \mid \mathbf{V}(T_1)\} \right. \\ \times p_{k_1 k_2}^{(I_{T_1} + 1)}\{t_{I_{T_1} + 1} - t_{I_{T_1}} \mid \mathbf{V}(t_{I_{T_1}})\} \times \dots \\ \left. \times p_{k_v X_2}^{(I_{T_2})}\{T_2 - t_{I_{T_2} - 1} \mid \mathbf{V}(t_{I_{T_2} - 1})\} \right],$$

où  $v = I_{T_2} - I_{T_1}$  est le nombre de points où l'intensité de transition change de valeur. Les estimations paramètres sont ensuite obtenues en maximisant la vraisemblance.

## 2.5 Tests d'hypothèses et d'adéquation

Au cours de l'analyse, on pourra tester des hypothèses de la forme  $H_0 : \alpha_{hj} = 0$ ,  $H_0 : \alpha_{hj} = \alpha_{kl}$  ou encore  $H_0 : \beta_{hj,k} = 0$  à l'aide des tests du rapport de vraisemblance, de Wald ou du score.

On peut tester l'hypothèse d'homogénéité en comparant le modèle homogène par périodes avec un modèle homogène : si le modèle homogène par périodes s'ajuste mieux, alors l'hypothèse d'homogénéité sera trop restrictive. L'hypothèse de Markov peut être testée en considérant par exemple une covariable « durée de séjour dans l'état ». En effet, dans un modèle Markovien, les intensités de transition ne dépendent pas du temps écoulé dans l'état avant de transiter. L'hypothèse de proportionnalité des risques doit également être vérifiée, par exemple, en considérant des covariables dépendantes du temps.

### 3 Modèle Markovien non-homogène

Les modèles de Markov homogènes sont de plus en plus appliqués. La méthodologie est bien connue et des programmes sont maintenant disponibles pour ajuster ce type de modèles. Cependant, dans de nombreuses applications, l'hypothèse d'homogénéité n'est pas adaptée et s'avère trop restrictive. L'utilisation d'un modèle de Markov non-homogène est une alternative permettant de considérer des intensités de transition qui dépendent du temps. Les méthodes présentées ici reposent sur la théorie des processus de comptage qui permet d'obtenir des estimateurs non-paramétriques. Les estimateurs obtenus constituent une généralisation aux modèles Markoviens des estimateurs de Kaplan-Meier et de la vraisemblance partielle de Cox.

#### 3.1 Processus de comptage

De manière similaire à l'analyse de survie, l'approche par les processus de comptage permet d'obtenir des estimateurs non-paramétriques dans les modèles de Markov non-homogènes.

Soit  $\{X(t), t \in \mathcal{T} = [0, \tau]\}$  un processus de Markov **non-homogène** à temps continu et à espace d'états fini  $S = \{1, \dots, k\}$  sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . À partir de l'observation du processus  $X(t)$ ,  $0 \leq t \leq \tau$ , on peut construire deux processus aléatoire

— la quantité  $Y_h(t)$  qui renseigne sur l'état du processus au temps  $t-$

$$Y_h(t) = \mathbf{1}_{\{X(t-)=h\}},$$

— le processus de comptage multivarié  $\mathbf{N}(t) = (N_{hj}(t); t \in \mathcal{T}, h \neq j)$

$$N_{hj}(t) = \# \{0 \leq s \leq t : X(s^-) = h, X(s) = j\} \quad \forall h \neq j.$$

où  $N_{hj}(t)$  compte le nombre de transitions observées de l'état  $h$  vers l'état  $j$  dans l'intervalle de temps  $[0, t]$ .

En effet, la connaissance du processus de comptage  $\mathbf{N} = (N_{hj}(t); t \in \mathcal{T}, h \neq j)$  apporte la même information que l'observation du processus  $\{X(t), t \in \mathcal{T}\}$ .

On peut montrer le résultat suivant en appliquant la décomposition de Doob-Meyer,

**Proposition 29** Soit  $A$  la mesure d'intensité d'un processus de Markov  $\{X(t), t \in \mathcal{T}\}$ , alors on a une décomposition de Doob-Meyer des processus de comptage  $N_{hj}$

$$N_{hj} = \Lambda_{hj} + M_{hj} \quad \forall h \neq j. \tag{III.7}$$

où les processus  $M_{hj}$  sont des martingales locales et les processus  $\Lambda_{hj}$  sont les compensateurs de  $N_{hj}$  par rapport à  $\mathcal{F}_t = \sigma(N_{hj}(u), t \in \mathcal{T})$  tels que

$$\Lambda_{hj}(t) = \int_0^t Y_h(s) dA_{hj}(s). \tag{III.8}$$

$\mathbf{\Lambda} = (\Lambda_{hj}; h \neq j)$  est appelé le **processus d'intensité cumulée**.

Les fonctions  $A_{hj}(\cdot)$  sont supposées absolument continues, ainsi

$$A_{hj}(t) = \int_0^t \alpha_{hj}(u) du.$$

où les fonctions  $\alpha_{hj}(\cdot)$  sont les intensités de transition. Dans ce cas, la proposition précédente implique que

— le compensateur  $\mathbf{\Lambda} = (\Lambda_{hj}; h \neq j)$  de  $\mathbf{N}$  vérifie

$$\Lambda_{hj}(t) = \int_0^t Y_h(s) \alpha_{hj}(ds),$$

— le processus  $\mathbf{N}$  a une **intensité multiplicative**  $\boldsymbol{\lambda} = (\lambda_{hj}; h \neq j)$  par rapport à  $\mathcal{F}_t = \sigma(\mathbf{N}(u), t \in \mathcal{T})$  qui vérifie

$$\lambda_{hj}(t) = \alpha_{hj}(t) Y_h(t) \quad \forall h \neq j.$$

Notons que  $\lambda_{hj}(\cdot)$  est une variable aléatoire car  $Y_h(t)$  est aléatoire et  $\alpha_{hj}(\cdot)$  est déterministe.

### 3.2 Vraisemblance

Considérons, le processus de comptage  $\mathbf{N} = \{N_{hj}(t), h, j \in S, h \neq j\}$ , sa filtration naturelle  $\mathcal{F}_t = \sigma(\mathbf{N}(u), t \in \mathcal{T})$  et  $\mathbf{\Lambda} = (\Lambda_{hj}; h, j \in S, h \neq j)$  son compensateur par rapport à  $\mathcal{F}_t$ .

La vraisemblance basée sur l'observation de  $\mathbf{N}(t)$  s'obtient par le théorème ?????. Conditionnellement à l'état initial, la vraisemblance est donnée par

$$\mathcal{L} = \mathcal{P}_{t \in [0, \tau]} \left\{ \prod_{h=1}^k \prod_{j \neq h} [d\Lambda_{hj}(t)]^{\Delta N_{hj}(t)} \left[ 1 - \sum_{h=1}^k \sum_{j \neq h} \Lambda_{hj}(t) \right]^{1 - \sum_{h=1}^k \sum_{j \neq h} \Delta N_{hj}(t)} \right\}. \quad (\text{III.9})$$

où  $\Delta Z(t) = Z(t) - Z(t^-)$ . Comme  $A_{hj}(t) = \int_0^t \alpha_{hj}(u) du$ , la vraisemblance s'écrit,

$$\mathcal{L} = \mathcal{P}_{t \in \mathcal{T}} \left\{ \prod_{h=1}^k \prod_{j \neq h} (\alpha_{hj}(t) Y_h(t))^{\Delta N_{hj}(t)} \left( 1 - \sum_{h=1}^k \sum_{j \neq h} \alpha_{hj}(t) Y_h(t) dt \right)^{1 - \sum_{h=1}^k \sum_{j \neq h} \Delta N_{hj}(t)} \right\} \quad (\text{III.10})$$

### 3.3 Processus de comptage et censure à droite

Le phénomène de censure à droite peut stopper l'observation du processus de Markov  $\{X(t), t \in \mathcal{T}\}$  et donc du processus de comptage  $\mathbf{N}$ . Soit  $\mathcal{C}(t)$  le processus de censure à droite,

$$\mathcal{C}(t) = I_{\{t \leq C\}},$$

où  $C$  est la variable aléatoire de censure à droite.

**Remarque 9** *On suppose ici que le processus de censure ne dépend pas du type d'événement. Néanmoins, les résultats obtenus peuvent facilement être adaptés au cas de différents mécanismes de censure pour différents types d'événements.*

Suite à la censure à droite, les processus réellement observés sont  
—  $\mathbf{N}^c(\cdot)$  le processus de comptage censuré à droite,

$$\mathbf{N}^c(t) = (N_{hj}^c(t), h, j = 1, \dots, k, h \neq j),$$

avec

$$N_{hj}^c(t) = \int_0^t \mathcal{C}(s) dN_{hj}(s),$$

—  $Y_h^c(t)$  tel que,

$$Y_h^c(t) = \mathcal{C}(t) Y_h(t), \quad h = 1, \dots, k.$$

On procède ensuite de manière similaire à l'analyse des données des durées de vie. On peut montrer que,

**Proposition 30** Sous l'hypothèse de censure indépendante, il existe une décomposition de Doob-Meyer du processus  $N_{hj}^c(\cdot)$

$$N_{hj}^c(t) = \Lambda_{hj}^c(t) + M_{hj}^c(t)$$

où  $M_{hj}^c$  est une martingale locale et  $\Lambda_{hj}^c$  est le compensateur tel que

$$\Lambda_{hj}^c(t) = \int_0^t \mathcal{C}(s) d\Lambda_{hj}(s) = \int_0^t \alpha_{hj}(s) Y_h^c(s) ds.$$

Le processus  $N_{hj}^c(\cdot)$  satisfait un modèle à intensité multiplicative par rapport à  $\sigma((N_{hj}^c(s), Y_h^c(s)), 0 \leq s \leq t)$ .

De plus on montre que la vraisemblance associé au processus censuré à la même forme que la vraisemblance du processus non censuré et qu'elle peut être utilisée pour l'inférence.

**Proposition 31** Soit  $\mathbf{N}^c(t) = (N_{hj}^c(t); h, j = 1, \dots, k, h \neq j)$  un processus de comptage censuré. Si la censure à droite est indépendante alors la vraisemblance du modèle s'écrit

$$\mathcal{L}^c = \mathcal{P} \prod_{t \in [0, \tau]} \prod_{h=1}^k \prod_{j \neq h} [d\Lambda_{hj}^c(t)]^{\Delta N_{hj}^c(t)} [1 - d\Lambda_{..}^c(t)]^{1 - \Delta N_{..}^c(t)}. \quad (\text{III.11})$$

où

$$\Lambda_{..}^c = \sum_{h=1}^k \sum_{j \neq h} \Lambda_{hj}^c \quad \text{et} \quad N_{..}^c(t) = \sum_{h=1}^k \sum_{j \neq h} N_{hj}^c(t).$$

Désormais, nous nous placerons dans le cadre d'un **mécanisme de censure indépendante**. Dans tout ce qui suit, les processus considérés seront des versions censurées mêmes si les notations ne le feront plus apparaître.

## 3.4 Estimation non-paramétrique

### 3.4.1 Observations et notations

Considérons un échantillon  $X_1(\cdot), \dots, X_n(\cdot)$  de processus de Markov indépendants et identiquement distribués à espace d'états fini  $S = \{1, \dots, k\}$ . Le processus  $X_i(\cdot)$  associé à l'individu  $i$  représente l'état de l'individu au temps  $t$ . On pose  $\mathcal{T} = [0, \tau]$  où  $\tau = \sup\{u : \int_0^u \alpha_{hj}(v)dv < \infty, h \neq j\}$  et on définit, pour  $i = 1, \dots, n$ ,

- Les processus de comptage  $N_{hji}(t)$  qui comptent le nombre de transitions de l'état  $h$  vers l'état  $j$  dans  $[0, t]$  pour l'individu  $i$ ,

$$N_{hji}(t) = \text{card} \{s \leq t : X_i(s^-) = h, X_i(s) = j\}, \quad \forall h, j = 1, \dots, k, h \neq j.$$

- $Y_{hi}(t)$  qui est un indicateur pour savoir si  $X_i$  est dans l'état  $h$  juste avant le temps  $t$ ,

$$Y_{hi}(t) = \mathbf{1}_{\{X_i(t^-) = h\}}, \quad h = 1, \dots, k.$$

- Le processus de comptage agrégé  $N_{hj}(t)$  qui compte le nombre total de transitions de l'état  $h$  vers l'état  $j$  dans  $[0, t]$  (dans toute la population),

$$N_{hj}(t) = \sum_{i=1}^n N_{hji}(t), \quad \forall h, j = 1, \dots, k, h \neq j,$$

- $Y_h(t)$  qui renseigne sur le nombre total de personne « à risque » dans l'état  $h$  juste avant l'instant  $t$ , c'est-à-dire le nombre de personnes susceptibles de subir une transition à partir de l'état  $h$ ,

$$Y_h(t) = \sum_{i=1}^n Y_{hi}(t) \quad h = 1, \dots, k.$$

**Proposition 32** Le processus agrégé  $N_{hj}(\cdot)$  admet la décomposition suivante

$$N_{hj}(t) = \Lambda_{hj}(t) + M_{hj}(t) \quad \forall t \geq 0,$$

où  $M_{hj}$  est une martingale locale et  $\Lambda_{hj}$  est le compensateur tel que

$$\Lambda_{hj}(t) = \int_0^t \alpha_{hj}(s) Y_h(s) ds \quad \forall t \geq 0.$$

$N_{hj}(\cdot)$  satisfait un modèle à intensité multiplicative,  $\forall i = 1, \dots, n; \forall h, j = 1, \dots, k, h \neq j$ ,

$$\lambda_{hj}(t) = \alpha_{hj}(t) Y_h(t). \quad (\text{III.12})$$

La vraisemblance associée au processus de comptage  $\mathbf{N} = \{N_{hji}(t), i = 1, \dots, n; h, j \in S, h \neq j\}$  conditionnellement aux données initiales est donnée par,

$$\mathcal{L} = \mathcal{P}_{t \in \mathcal{T}} \left\{ \prod_i \prod_{h \neq j} (\alpha_{hj}(t) Y_{hi}(t))^{\Delta N_{hji}(t)} \left( 1 - \sum_{h \neq j} \alpha_{hj}(t) Y_h(t) dt \right)^{1 - \sum_{h \neq j} \Delta N_{hj}(t)} \right\}.$$

### 3.4.2 Estimation des intensités cumulées

On a supposé que pour tout  $h \neq j$  et pour tout  $t \in \mathcal{T}$  les intensités de transition vérifient  $\int_0^t \alpha_{hj}(u) du < \infty$ . Un estimateur des intensités cumulées  $A_{hj}(t) = \int_0^t \alpha_{hj}(u) du$  est donné par la proposition suivante

**Définition 20** L'estimateur de **Nelson-Aalen** des fonctions d'intensité cumulée est défini

$$\hat{A}_{hj}(t) = \int_0^t \frac{J_h(u)}{Y_h(u)} dN_{hj}(u), \quad \forall h \neq j, \quad (\text{III.13})$$

où  $J_h(t) = \mathbf{1}_{\{Y_h(t) > 0\}}$ .

Notons que  $J_h(s)/Y_h(s)$  est interprété comme étant 0 quand  $Y_h(s) = 0$ . L'origine « naturelle » de cet estimateur provient de la décomposition de Doob-Meyer du processus agrégé qui permet d'écrire heuristiquement

$$dN_{hj}(t) = Y_h(t) \alpha_{hj}(t) dt + dM_{hj}(t),$$

où  $dM_{hj}(t)$  peut être considéré comme un bruit aléatoire.

**Remarque 10** Soit  $T_{hj,1} < T_{hj,2} < \dots$  les temps de saut successifs de  $N_{hj+}$ .  $N_{hj+}$  attribue une masse 1 à chacun de ces temps de saut et une masse 0 ailleurs. Il s'ensuit qu'on peut écrire  $\hat{A}_{hj}(t)$  comme une simple somme

$$\hat{A}_{hj}(t) = \sum_{\{k: T_{hj,k} \leq t\}} \frac{1}{Y_h(T_{hj,k})}.$$

Ainsi  $\hat{A}_{hj}$  est une fonction en escalier, croissante, continue à droite, de saut  $1/Y_h(T_{hj,k})$  au temps de saut  $T_{hj,k}$  de  $N_{hj}$ .

**Proposition 33** Un estimateur de la variance de  $\hat{A}_{hj}(t)$  est

$$\hat{\sigma}_{hj}^2(t) = \int_0^t \frac{J_h(u)}{(Y_h(u))^2} dN_{hj}(u), \quad \forall h \neq j.$$

**Proposition 34** Sous certaines conditions (pages 190-191 de Andersen et al. (1993)),  $\hat{A}_{hj}(t)$  est un estimateur

— biaisé, tel que

$$\mathbb{E} \left( \hat{A}_{hj}(t) \right) - A_{hj}(t) = - \int_0^t \alpha_{hj}(u) \mathbb{P}(Y_h(u) = 0) du,$$

— uniformément consistant,

$$\sup_{s \in [0, t]} \left| \hat{A}_{hj}(s) - A_{hj}(s) \right| \xrightarrow{\mathbb{P}} 0,$$

— et asymptotiquement normal, tel que

$$\left( \sqrt{n} \left( \hat{A}_{hj}(t) - A_{hj}(t) \right); h \neq j \right) \xrightarrow{\mathcal{L}} (U_{hj}; h \neq j),$$

avec  $U_{hj}$  martingales gaussiennes indépendantes telle que

$$\begin{cases} U_{hj}(0) = 0 \\ \mathbb{V}(U_{hj}(t)) = \int_0^t \frac{\alpha_{hj}(u)}{\sum_{j=1}^k p_{jh}(0, u)} du. \end{cases}$$

Le biais de cet estimateur est très faible si la probabilité qu'à un instant  $t$  aucun individu ne soit à risque est proche de zéro. Les hypothèses sont discutées et les démonstrations sont données pages 179 et 190-199 de Andersen et al. (1993).

### 3.4.3 Estimation des probabilités de transition

Rappelons que la matrice des probabilités de transition est définie en fonction de la matrice des intensités cumulées par le produit intégral,

$$\mathbf{P}(s, t) = \mathcal{P}_{u \in ]s, t]} (\mathbf{Id} + d\mathbf{A}(u)) \quad 0 < s \leq t, \quad t, s \in \mathcal{T}.$$

avec  $\mathbf{P}(s, t) = \{p_{hj}(s, t)\}_{h,j}$  et  $\mathbf{A}(t) = \{A_{hj}(t)\}_{h,j}$ . L'estimateur introduit par Aalen et Johansen en 1978 utilise cette relation et l'estimateur de Nelson-Aalen pour obtenir une estimation de la matrice des probabilités de transition.

**Définition 21** L'estimateur de **Aalen-Johansen** de la matrice des probabilités de transition est défini par

$$\hat{\mathbf{P}}(s, t) = \mathcal{P}_{u \in ]s, t]} (\mathbf{Id} + d\hat{\mathbf{A}}(u)), \quad 0 < s \leq t, \quad t, s \in \mathcal{T}, \quad (\text{III.14})$$

où  $\hat{\mathbf{P}}(s, t) = \{\hat{p}_{hj}(s, t)\}_{h,j}$  et  $\hat{\mathbf{A}}(t) = \{\hat{A}_{hj}(t)\}_{h,j}$  est l'estimateur de Nelson-Aalen.

**Remarque 11**  $p_{hj}(s, t)$  correspond à la probabilité de transiter dans l'état  $j$  à l'instant  $t$  sachant que le patient est dans l'état  $h$  à l'instant  $s$ . En pratique, on interprètera souvent la quantité  $p_{hj}(0, t)$ .

**Remarque 12** En pratique, il y a un nombre fini de transitions. Soient  $s < T_1 < \dots < T_k < t$ , les temps de transition entre deux états. L'estimateur de Aalen-Johansen devient un produit fini de matrice tel que

$$\hat{\mathbf{P}}(s, t) = \prod_{l=1}^k \left( \mathbf{Id} + \Delta \hat{\mathbf{A}}(T_l) \right).$$

Dans le cas d'un modèle à trois états, où toutes les transitions entre états sont possibles,

$$\mathbf{I} + \Delta \hat{\mathbf{A}}(T_l) = \begin{pmatrix} 1 - \frac{\Delta N_{12}(T_l) + \Delta N_{13}(T_l)}{Y_1(T_l)} & \frac{\Delta N_{12}(T_l)}{Y_1(T_l)} & \frac{\Delta N_{13}(T_l)}{Y_1(T_l)} \\ \frac{\Delta N_{21}(T_l)}{Y_2(T_l)} & 1 - \frac{\Delta N_{21}(T_l) + \Delta N_{23}(T_l)}{Y_2(T_l)} & \frac{\Delta N_{23}(T_l)}{Y_2(T_l)} \\ \frac{\Delta N_{31}(T_l)}{Y_3(T_l)} & \frac{\Delta N_{32}(T_l)}{Y_3(T_l)} & 1 - \frac{\Delta N_{31}(T_l) + \Delta N_{32}(T_l)}{Y_3(T_l)} \end{pmatrix},$$

car  $A_{hh}(t) = -\sum_{j \neq h} A_{hj}(t)$  et où pour un processus  $Z(\cdot)$  càdlàg,  $\Delta Z(T_l) = Z(T_l) - Z(T_{l-1})$ , pour  $l = 2, \dots, k$ , et  $\Delta Z(T_1) = Z(T_1)$ .

**Proposition 35** Un premier estimateur de la covariance de  $\hat{p}_{hj}(s, t)$  et  $\hat{p}_{mr}(s, t)$  est

$$\begin{aligned} \widehat{Cov}(\hat{p}_{hj}(s, t), \hat{p}_{mr}(s, t)) &= \sum_{l=1}^k \sum_{g \neq l} \int_s^t \hat{p}_{hg}(s, u) \hat{p}_{mg}(s, u) \{ \hat{p}_{lj}(u, t) - \hat{p}_{gj}(u, t) \} \\ &\quad \times \{ \hat{p}_{lr}(u, t) - \hat{p}_{gr}(u, t) \} J_g(u) (Y_{g+}(u))^{-2} dN_{gl}(u). \end{aligned}$$

L'estimateur de la variance de  $\hat{p}_{hj}(s, t)$  est

$$\hat{V}(\hat{p}_{hj}(s, t)) = \sum_{l=1}^k \sum_{g \neq l} \int_s^t (\hat{p}_{hg}(s, u))^2 \{ \hat{p}_{lj}(u, t) - \hat{p}_{gj}(u, t) \}^2 J_g(u) (Y_g(u))^{-2} dN_{gl}(u).$$

**Proposition 36** Un deuxième estimateur, de type Greenwood, de la covariance de  $\hat{p}_{hj}(s, t)$  et  $\hat{p}_{mr}(s, t)$  est

$$\begin{aligned} \widehat{Cov}(\hat{p}_{hj}(s, t), \hat{p}_{mr}(s, t)) &= \sum_{l=1}^k \sum_{g \neq l} \int_s^t \hat{p}_{hg}(s, u^-) \hat{p}_{mg}(s, u^-) \{ \hat{p}_{lj}(u, t) - \hat{p}_{gj}(u, t) \} \\ &\quad \times \{ \hat{p}_{lr}(u, t) - \hat{p}_{gr}(u, t) \} J_g(u) (Y_g(u) - 1) (Y_g(u))^{-3} dN_{gl}(u). \end{aligned}$$

L'estimateur correspondant de la variance de  $\hat{p}_{hj}(s, t)$  est

$$\hat{V}(\hat{p}_{hj}(s, t)) = \sum_{l=1}^k \sum_{g \neq l} \int_s^t (\hat{p}_{hg}(s, u^-))^2 \{ \hat{p}_{lj}(u, t) - \hat{p}_{gj}(u, t) \}^2 J_g(u) \frac{(Y_g(u) - 1)}{(Y_g(u))^3} dN_{gl}(u).$$

Aalen et Johansen (1978) utilisent notamment la décomposition de Doob-Meyer et le théorème de limite centrale pour martingales pour obtenir les propriétés asymptotiques de  $\hat{\mathbf{P}}$  quand  $n \rightarrow \infty$ .

**Proposition 37** Sous le même type de condition que pour l'estimateur de Nelson-Aalen (pages 317-319 de Andersen et al. (1993)),  $\hat{p}_{hj}(s, t)$  est un estimateur

— presque sans biais quand  $\mathbb{P}(Y_h(u) = 0)$ ,  $s < u \leq t$  est petite, de plus on montre que

$$\mathbb{E} \left( \hat{\mathbf{P}}(s, t) \right) \geq \mathbf{P}(s, t),$$

— uniformément consistant,

$$\sup_{t \in [s, v]} \left\| \hat{\mathbf{P}}(s, t) - \mathbf{P}(s, t) \right\| \xrightarrow{\mathbb{P}} 0 \Leftrightarrow \sup_{t \in [s, v]} \sup_h \sum_j |\hat{p}_{hj}(s, t) - p_{hj}(s, t)| \xrightarrow{\mathbb{P}} 0$$

— et asymptotiquement normal, tel que

$$\sqrt{n} \left( \hat{\mathbf{P}}(s, \cdot) - \mathbf{P}(s, \cdot) \right) \xrightarrow{\mathcal{L}} \int_0^{\cdot s} \mathbf{P}(s, u) d\mathbf{U}(u) \mathbf{P}(u, \cdot),$$

avec  $\mathbf{U} = \{U_{hj}\}_{hj}$  où  $U_{hj}$  martingales gaussiennes indépendantes telle que

$$\begin{cases} U_{hj}(0) = 0 \\ \mathbb{V}(U_{hj}(t)) = \int_0^t \frac{\alpha_{hj}(u)}{\sum_{j=1}^k p_{jh}(0, u)} du. \end{cases}$$

Les démonstrations de ces propositions sont données pages 197; 289-290 et 317-321 de Andersen et al. (1993).

### 3.4.4 Test des intensités de transition

Cette méthode d'estimation ne prend pas en compte l'effet des covariables dans la modélisation. On peut cependant comparer les intensités de transition obtenues dans deux groupes.

**Proposition 38** Soit un processus de comptage  $\mathbf{N} = (N_1, N_2)$  avec un processus d'intensité  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$  tel que  $\lambda_h(t) = \alpha_h(t)Y_h(t)$ ,  $h = 1, 2$ . Soit l'hypothèse nulle,

$$H_0 : \alpha_1 = \alpha_2, \tag{III.15}$$

Alors la statistique de test

$$X^2(t) = \frac{H^2(t)}{\sigma(t)},$$

suit asymptotiquement une loi du  $\chi^2$  à un degré de liberté sous l'hypothèse  $H_0$  avec,

$$\begin{aligned} H(t) &= \int_0^t L(u) d\hat{A}_1(u) - \int_0^t L(u) d\hat{A}_2(u), \\ \sigma(t) &= \int_0^t L^2(u) \{Y_1(u)Y_2(u)\}^{-1} d(N_1 + N_2)(u), \\ L(t) &= K(t)Y_1(t)Y_2(t) \{Y_1(u) + Y_2(u)\}^{-1}. \end{aligned}$$

Le test du **Log-rank** est obtenu avec

$$K(t) = \mathbf{1}_{\{Y_1(t)+Y_2(t)>0\}}. \quad (\text{III.16})$$

Le test de **Gehan-Wilcoxon** est obtenu avec

$$K(t) = Y_1(t) + Y_2(t). \quad (\text{III.17})$$

La proposition précédente est un cas particulier ( $k = 2$ ) d'une classe de statistiques de test pour l'hypothèse  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k$ , (cf. page 345 de Andersen et al. (1993)).

### 3.4.5 Cas particulier : données de survie

Le modèle de survie peut être considéré comme un modèle de Markov non-homogène particulier comportant deux états (vivant et décès) avec une seule transition possible. Le processus est Markovien dans le sens où le passé du processus se résume (évidemment) à l'état présent. Le processus de comptage associé est un processus dont les marques sont : état 0 (vivant), état 1 (décès). La matrice des probabilités de transition associée et la matrice des intensités cumulées sont définies par

$$\mathbf{P}(s, t) = \begin{pmatrix} p_{00}(s, t) & p_{01}(s, t) \\ 0 & 1 \end{pmatrix}, \quad \text{et} \quad \mathbf{A}(t) = \begin{pmatrix} A_{00}(t) & A_{01}(t) \\ 0 & 0 \end{pmatrix}.$$

L'équation différentielle « forward » de Kolmogorov donne l'équation suivante

$$\frac{\partial P_{00}(s, t)}{\partial t} = P_{00}(s, t^-) dA_{00}(t).$$

En intégrant par rapport à  $t$ , nous obtenons alors la probabilité de rester dans l'état 0 entre les instants  $s$  et  $t$  sous la forme d'une équation de type Volterra

$$P_{00}(s, t) = P_{00}(s, s) + \int_s^t \frac{\partial P_{00}(s, v)}{\partial v} dv = 1 + \int_s^t P_{00}(s, v^-) dA_{00}(v).$$

où  $dA_{00}(v) = -dA_{01}(v)$ . La probabilité  $P_{00}(s, t)$  est de la forme

$$Z(s, t) = 1 + \int_s^t Z(s, v^-) dX(v),$$

avec  $Z(s, t) = P_{00}(s, t)$ , et  $X(v) = -A_{01}(v)$ . D'après l'équation de Volterra, il existe une unique solution de la forme

$$P_{00}(s, t) = \mathcal{P}_{u \in ]s, t]} \{1 - dA_{01}(u)\}. \quad (\text{III.18})$$

$P_{00}(0, t)$  est la probabilité d'être vivant au temps  $t$  (sachant qu'on est vivant au temps 0), c'est-à-dire que  $P_{00}(0, t)$  correspond à la probabilité de survivre après  $t$ , i.e. la fonction de survie. Ainsi, on peut déduire de l'équation (III.18) et de l'estimateur de Nelson-Aalen, l'estimateur bien connu Kaplan-Meier.

Soient, pour tout  $i = 1, \dots, n$ ,

- $X_i$  la date de survenue de l'événement et  $C_i$  la date de censure,
- $T_i = \min(X_i, C_i)$  et  $\delta_i = \mathbf{1}_{\{X_i \leq C_i\}}$
- $Y_i(t) = \mathbf{1}_{\{T_i \geq t\}}$ , vaut 1 si l'individu est à risque l'instant  $t$ , 0 sinon.
- $N_i(t) = \mathbf{1}_{\{T_i \leq t, \delta_i = 1\}}$ , vaut 1 si l'individu est décédé à l'instant  $t$ , 0 sinon.
- Enfin, définissons
- $Y(t) = \sum_{i=1}^n Y_i(t)$ , le nombre total de survenues de l'événement à l'instant  $t$ ,
- $N(t) = \sum_{i=1}^n N_i(t)$ , le nombre total d'individus à risque juste avant l'instant  $t$ ,
- $J(t) = \mathbf{1}_{\{Y(t) > 0\}}$ .

Alors l'estimateur de **Kaplan-Meier** de la fonction de survie est défini par

$$\hat{S}(t) = \mathcal{P}_{u \leq t} (1 - dA(u)),$$

où  $A$  est l'estimateur de Nelson-Aalen,

$$dA(u) = \frac{J(u)}{Y(u)} dN(u).$$

## 3.5 Estimation semi-paramétrique

### 3.5.1 Définitions et notations

Considérons un échantillon  $X_1(\cdot), \dots, X_n(\cdot)$  de processus de Markov indépendants et identiquement distribués à espace d'états fini  $S = \{1, \dots, k\}$ . Considérons,

- $\mathbf{Z}_i = (Z_{1i}, \dots, Z_{pi})$ ,  $i = 1, \dots, n$ , le vecteur de covariables de dimension  $p$ .
- $\beta_{hj} = (\beta_{hj,1}, \dots, \beta_{hj,p})$ , le vecteur de dimension  $p$  des coefficients de régression ;

Le modèle considéré se caractérise par une structure multiplicative des processus d'intensité individuelle

$$\lambda_{hji}(t) = Y_{hi}(t) \alpha_{hj}(t | \mathbf{Z}_i),$$

où les intensités de transition  $\alpha_{hj}$  suivent un modèle semi-paramétrique à risque multiplicatif de "Cox", c'est-à-dire,

$$\alpha_{hj}(t | \mathbf{Z}_i) = \alpha_{hj0}(t) \exp(\beta_{hj}^T \mathbf{Z}_i), \quad h, j \in S, h \neq j, i = 1, \dots, n. \quad (\text{III.19})$$

où  $\alpha_{hj0}(t)$  est l'intensité de transition de base associée à la transition de l'état  $h$  vers l'état  $j$ .

Ce modèle est dit à **risques proportionnels** car, par définition, quels que soient deux individus (1 et 2), le rapport des intensités de transition ne varie pas au cours du temps

$$\frac{\alpha_{hj}(t | \mathbf{Z}_1)}{\alpha_{hj}(t | \mathbf{Z}_2)} = \exp(\beta_{hj}^T (\mathbf{Z}_1 - \mathbf{Z}_2)). \quad (\text{III.20})$$

Notons également que le modèle (III.19) est **log-linéaire**, en effet

$$\log \alpha_{hj}(t | \mathbf{Z}_i) = \log \alpha_{hj0}(t) + \beta_{hj}^T \mathbf{Z}_i.$$

L'hypothèse de log-linéarité entraîne que le risque relatif est constant pour une augmentation d'une unité quelle que soit la valeur de la covariable explicative.

### 3.5.2 Estimation des intensités de base

Supposons pour l'estimation que les fonctions  $\alpha_{hj0}(\cdot)$  sont positives et que  $\forall t \in \mathcal{T}$ ,  $A_{hj0}(t) = \int_0^t \alpha_{hj0}(u)du < \infty$ ,  $\forall h \neq j$ . La vraisemblance associée au processus de comptage est

$$\mathcal{L}(\boldsymbol{\beta}) = \mathcal{P} \prod_{t \leq \tau} \prod_i \prod_{h \neq j} \left\{ (Y_{hi}(t) dA_{hj0}(t) \exp(\boldsymbol{\beta}_{hj}^T \mathbf{Z}_i))^{\Delta N_{hji}(t)} \right\} \times \exp \left( - \sum_{h \neq j} \int_0^\tau S_{hj}^0(\boldsymbol{\beta}_{hj}, s) dA_{hj0}(s) \right),$$

avec

$$S_{hj}^0(\boldsymbol{\beta}_{hj}, t) = \sum_{i=1}^n \exp(\boldsymbol{\beta}_{hj}^T \mathbf{Z}_i) Y_{hi}(t).$$

En considérant  $\boldsymbol{\beta} = (\boldsymbol{\beta}_{hj}, h \neq j)$  fixé, la maximisation de la vraisemblance fournit l'estimateur suivant

$$\Delta \hat{A}_{hj0}(t) = \frac{\Delta N_{hj}(t)}{S_{hj}^0(\boldsymbol{\beta}_{hj}, t)}.$$

**Proposition 39** Pour  $\boldsymbol{\beta} = (\boldsymbol{\beta}_{hj}, h \neq j)$  fixé,  $A_{hj0}(t) = \int_0^t \alpha_{hj0}(u)du$  est estimé par l'estimateur de Breslow (Breslow [1974]),

$$\hat{A}_{hj0}(t) = \int_0^t \frac{J_h(u)}{\sum_{i=1}^n \exp(\boldsymbol{\beta}_{hj}^T \mathbf{Z}_i) Y_{hi}(t)} dN_{hj}(u), \quad (\text{III.21})$$

avec  $J_h(u) = \mathbb{1}_{\{Y_h(t) > 0\}}$ .

### 3.5.3 Estimation des coefficients de régression

En remplaçant,  $A_{hj0}(t)$  par son estimateur la vraisemblance devient,

$$\mathcal{L}(\boldsymbol{\beta}) = \mathcal{L}_{Cox}(\boldsymbol{\beta}) \times \mathcal{P} \prod_{t \leq \tau} \prod_i \prod_{h \neq j} (J_h(t) dN_{hj}(t))^{\Delta N_{hji}(t)} \times \exp \left( - \sum_{h \neq j} \int_0^\tau J_h(s) dN_{hj}(s) \right),$$

avec

$$\mathcal{L}_{Cox}(\boldsymbol{\beta}) = \mathcal{P} \prod_{t \leq \tau} \prod_{i=1}^n \prod_{h \neq j} \left[ \frac{Y_{hi}(t) \exp(\boldsymbol{\beta}_{hj}^T \mathbf{Z}_i)}{S_{hj}^0(\boldsymbol{\beta}_{hj}, t)} \right]^{\Delta N_{hji}(t)}. \quad (\text{III.22})$$

$\mathcal{L}_{Cox}(\boldsymbol{\beta})$  est la **vraisemblance partielle de Cox**. Cette vraisemblance est introduite par Cox (1972) mais par une approche différente. Ce n'est pas une vraisemblance dans le sens statistique du terme, mais il est établi qu'elle peut être utilisée comme telle pour estimer les coefficients de régression.

Considérons la fonction de log-vraisemblance partielle de Cox,

$$\log \mathcal{L}_{Cox}(\boldsymbol{\beta}) = \sum_i \sum_{h \neq j} \int_0^\tau [\boldsymbol{\beta}_{hj}^T \mathbf{Z}_i - \log S_{hj}^0(\boldsymbol{\beta}_{hj}, t)] dN_{hji}(t).$$

Les vecteurs scores sont donnés par

$$\mathbf{U}_{hj}(\boldsymbol{\beta}_{hj}) = \frac{\partial \log \mathcal{L}_{Cox}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{hj}} = \sum_i \int_0^\tau \left[ \mathbf{Z}_i - \frac{S_{hj}^1(\boldsymbol{\beta}_{hj}, t)}{S_{hj}^0(\boldsymbol{\beta}_{hj}, t)} \right] dN_{hji}(t),$$

avec

$$S_{hj}^1(\boldsymbol{\beta}_{hj}, t) = \sum_{i=1}^n Y_{hi}(t) \mathbf{Z}_i \exp(\boldsymbol{\beta}_{hj}^T \mathbf{Z}_i).$$

**Proposition 40** L'estimateur  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{hj}, h \neq j)$  du maximum de la vraisemblance partielle de Cox vérifie

$$\mathbf{U}_{hj}(\hat{\boldsymbol{\beta}}_{hj}) = 0 \quad \forall h \neq j.$$

De plus,

$$\sqrt{n} \left( \hat{\boldsymbol{\beta}}_{hj} - \boldsymbol{\beta}_{hj0} \right) \xrightarrow{\mathcal{L}} N(0, nI^{-1}(\hat{\boldsymbol{\beta}}_{hj})),$$

où  $\boldsymbol{\beta}_{hj0}$  est le vrai paramètre et où  $I(\boldsymbol{\beta}_{hj})$  est la matrice d'information de Fisher,

$$\begin{aligned} I(\boldsymbol{\beta}_{hj}) &= -\mathbb{E} \left( \frac{\partial^2 \log \mathcal{L}_{Cox}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{hj}^2} \right) = - \left\{ \mathbb{E} \left( \frac{\partial^2 \log \mathcal{L}_{Cox}(\boldsymbol{\beta})}{\partial \beta_{hj,k} \partial \beta_{hj,l}} \right) \right\}_{p \times p} \\ &= \mathbb{E} \left( \int_0^\tau \frac{[S_{hj}^2(\boldsymbol{\beta}_{hj}, t) S_{hj}^0(\boldsymbol{\beta}_{hj}, t)] - [S_{hj}^1(\boldsymbol{\beta}_{hj}, t)] [S_{hj}^1(\boldsymbol{\beta}_{hj}, t)]^T}{(S_{hj}^0(\boldsymbol{\beta}_{hj}, t))^2} dN_{hj}(t) \right), \end{aligned}$$

avec

$$S_{hj}^2(\boldsymbol{\beta}_{hj}, t) = \sum_{i=1}^n Y_{hi}(t) \mathbf{Z}_i \mathbf{Z}_i^T \exp(\boldsymbol{\beta}_{hj} \mathbf{Z}_i).$$

L'inverse de la matrice d'information de Fisher étant égale à la matrice de variance-covariance, elle fournit une estimation de la variance de  $\hat{\boldsymbol{\beta}}_{hj}$  en considérant la matrice d'information observée au point  $\boldsymbol{\beta}_{hj} = \hat{\boldsymbol{\beta}}_{hj}$ , *i.e.*  $\mathcal{I}(\hat{\boldsymbol{\beta}}_{hj}) = - \left[ \frac{\partial^2 \log \mathcal{L}_{Cox}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{hj}^2} \right]_{\boldsymbol{\beta}_{hj} = \hat{\boldsymbol{\beta}}_{hj}}$ . En pratique, les estimations des coefficients de régression sont obtenues à l'aide d'algorithmes itératifs.

### 3.5.4 Estimation des probabilités de transition

Un estimateur de la matrice des probabilités de transition s'obtient par le produit intégral. Considérons  $\mathbf{Z}_0$ , la valeur des covariables pour un individu.

**Proposition 41** Un estimateur de la matrice des probabilités de transition est donné par le produit intégral

$$\hat{\mathbf{P}}(s, t | \mathbf{Z}_0) = \mathcal{P}_{u \in ]s, t]} \left( \mathbf{Id} + d\hat{\mathbf{A}}(u | \mathbf{Z}_0) \right), \quad s \leq t \leq \tau,$$

avec  $\hat{\mathbf{A}} = \{\hat{A}_{hj}\}_{hj}$ ,

$$\begin{aligned} d\hat{A}_{hj}(t | \mathbf{Z}_0) &= d\hat{A}_{hj0}(t | \mathbf{Z}_0) \times \exp(\hat{\beta}_{hj}^T \mathbf{Z}_0) \\ &= \frac{J_h(t) \times \exp(\hat{\beta}_{hj}^T \mathbf{Z}_0)}{\sum_{i=1}^n \exp(\hat{\beta}_{hj}^T \mathbf{Z}_i) Y_{hi}(t)} \times \Delta N_{hj}(t), \quad h \neq j, \end{aligned}$$

et

$$d\hat{A}_{hh}(t | \mathbf{Z}_0) = - \sum_{j \neq h} d\hat{A}_{hj}(t | \mathbf{Z}_0), \quad h = 1, \dots, s.$$

Les propriétés asymptotiques de  $\hat{\mathbf{P}}(s, t | \mathbf{Z}_0)$  s'obtiennent notamment à partir de celles de  $\hat{\beta}_{hj}$  et  $\hat{A}_{hj0}(\cdot)$  et de l'utilisation de la delta-méthode,

**Proposition 42** Un estimateur de la covariance de  $\hat{p}_{hj}(s, t | \mathbf{Z}_0)$  et  $\hat{p}_{mr}(s, t | \mathbf{Z}_0)$  est

$$\widehat{Cov}(\hat{p}_{hj}(s, t | \mathbf{Z}_0), \hat{p}_{mr}(s, t | \mathbf{Z}_0)) = \widehat{Cov}_1 + \widehat{Cov}_2,$$

avec

$$\begin{aligned} \widehat{Cov}_1 &= \sum_{g \neq l} \int_s^t \hat{p}_{hg}(s, u | \mathbf{Z}_0) \hat{p}_{mg}(s, u | \mathbf{Z}_0) \{ \hat{p}_{lj}(u, t | \mathbf{Z}_0) - \hat{p}_{gj}(u, t | \mathbf{Z}_0) \} \\ &\quad \times \{ \hat{p}_{lr}(u, t | \mathbf{Z}_0) - \hat{p}_{gr}(u, t | \mathbf{Z}_0) \} J_g(u) \left\{ \exp(\hat{\beta}_{gl}^T \mathbf{Z}_0) \right\}^2 \left\{ S_{gl}^0(\hat{\beta}_{gl}, u) \right\}^{-2} dN_{gl}(u), \end{aligned}$$

et

$$\begin{aligned} \widehat{Cov}_2 &= \int_s^t \sum_{g,l} \hat{p}_{hg}(s, u | \mathbf{Z}_0) dW_{gl}(u) \hat{p}_{lj}(u, t | \mathbf{Z}_0) \times \mathcal{I}^{-1}(\hat{\beta}_{gl}) \\ &\quad \times \int_s^t \sum_{g,l} \hat{p}_{mg}(s, u | \mathbf{Z}_0) dW_{gl}(u) \hat{p}_{lr}(u, t | \mathbf{Z}_0), \end{aligned}$$

où  $\mathcal{I}^{-1}(\hat{\beta}_{gl})$  est un estimateur de la matrice de variance-covariance de  $\hat{\beta}_{gl}$  et pour tout  $g \neq l$ ,

$$W_{gl}(t) = \exp(\hat{\beta}_{gl}^T \mathbf{Z}_0) \int_0^t \left( \mathbf{Z}_0 - \frac{S_{hj}^1(\hat{\beta}_{hj}, t)}{S_{hj}^0(\hat{\beta}_{hj}, t)} \right) J_g(u) \left\{ S_{gl}^0(\hat{\beta}_{gl}, u) \right\}^{-1} dN_{gl}(u),$$

et  $W_{gg} = - \sum_{g \neq l} W_{gl}$ .

### 3.5.5 Cas particulier des durées de vie

La méthodologie précédente comprend le cas particulier des données de survie. Le modèle de survie est un modèle à deux états où une seule transition est possible. Considérons deux états  $S = \{0, 1\}$  (vivant et décès), la vraisemblance partielle de Cox s'écrit

$$\mathcal{L}_{Cox}(\boldsymbol{\beta}) = \prod_t \prod_{i=1} \left[ \frac{Y_{0i}(t) \exp(\boldsymbol{\beta}_{01}^T \mathbf{Z}_i)}{S_{01}^0(\boldsymbol{\beta}_{01}, t)} \right]^{\Delta N_{0i}(t)},$$

Cette expression est bien la vraisemblance obtenue par Cox pour des données de survie. De même, l'estimateur des intensités cumulées de base généralise l'estimateur dans le cas des données de survie,

$$\hat{A}_{010}(t) = \int_0^t \frac{J_0(u)}{\sum_{i=1}^n \exp(\boldsymbol{\beta}_{01}^T \mathbf{Z}_i) Y_{0i}(t)} dN_{01}(u),$$

où  $\hat{A}_{010}(t)$  est l'intensité cumulée de base associée au « décès ».

### 3.5.6 Tests et extensions

**Covariables dépendantes du temps** L'utilisation de covariables dépendantes du temps est pertinente dans certaines situations. Par exemple, pour prendre en compte des covariables comme le type de traitement, un dosage biologique ou le temps de séjour.

L'approche par la vraisemblance partielle et les résultats précédents restent vrais pour des covariables dépendantes du temps. De plus, l'hypothèse de proportionnalité des risques est toujours vérifiée car ce n'est pas l'effet de la variable qui varie avec le temps mais la variable elle-même.

Notons cependant quelques différences avec l'utilisation de covariables dépendantes du temps.

- Afin de calculer la contribution à la vraisemblance partielle de Cox à chaque temps d'événement, il est nécessaire de connaître la valeur de ces covariables pour tous ces temps. Le plus souvent, ces valeurs ne sont pas disponibles. Ce problème de données manquantes peut se résoudre en supposant par exemple que la valeur de la covariable à l'instant  $t$  est la valeur enregistrée à la dernière consultation (avant  $t^-$ ).
- L'interprétation des résultats est moins aisée même si celle des coefficients de régression demeure identique. En effet, l'interprétation des probabilités de transition pour des covariables dépendantes du temps devient délicate car il y a un estimateur des probabilités de transition pour chaque « histoire » de covariable.
- Les temps de calcul peuvent singulièrement augmenter.

**Tests des coefficients** Les trois statistiques de test suivantes permettent de tester l'hypothèse  $H_0 : \hat{\boldsymbol{\beta}}_{hj} = \boldsymbol{\beta}_0$ ,

- Test du rapport de vraisemblance : ce test mesure les différences entre le logarithme de la vraisemblance en  $\hat{\boldsymbol{\beta}}_{hj}$  et  $\boldsymbol{\beta}_0$ ,

$$2 \left[ \log \mathcal{L}_{Cox}(\hat{\boldsymbol{\beta}}) - \log \mathcal{L}_{Cox}(\boldsymbol{\beta}_0) \right] \stackrel{H_0}{\rightsquigarrow} \chi^2(p).$$

— Test de Wald :

Il mesure l'écart entre  $\hat{\beta}_{hj}$  et  $\beta_0$ ,

$$(\hat{\beta} - \beta_0)^T \mathcal{I}(\hat{\beta}_{hj})(\hat{\beta} - \beta_0) \stackrel{H_0}{\rightsquigarrow} \chi^2(p).$$

— Test du score :

Ce test mesure la pente de la tangente en  $\beta_0$ ,

$$\mathbf{U}_{hj}(\beta_0)^T \mathcal{I}(\beta_0)^{-1} \mathbf{U}_{hj}(\beta_0) \stackrel{H_0}{\rightsquigarrow} \chi^2(p).$$

**Test de l'hypothèse de proportionnalité des risques** La méthode d'estimation semi-paramétrique suppose que les risques sont à hasards proportionnels (III.20). Les méthodes pour vérifier ce type d'hypothèses sont bien connues dans les modèles de survie. Certaines peuvent être adaptées aux modèles multi-états comme l'utilisation d'une covariables dépendantes du temps ou les méthodes graphiques (cf. chapitre Analyse des durées de vie).



# Chapitre IV

## Modèle semi-Markovien homogène

### 1 Modèle semi-Markovien homogène

Les modèles Markoviens supposent que l'évolution du processus est indépendante du temps déjà passé dans l'état actuel. Dans de nombreuses applications, cette hypothèse ne correspond pas à la réalité du phénomène étudié. Les processus semi-Markoviens constituent alors une alternative intéressante puisqu'ils intègrent dans la définition du modèle les distributions des temps de séjour dans un état. Dans les modèles semi-Markovien, les intensités de transition peuvent dépendre du temps passé dans un état.

### 2 Introduction

#### 2.1 Définitions

On considère  $(J_n, S_n)_{n \geq 0}$  un processus semi-Markovien, où  $0 = S_0 < S_1 < \dots < S_n < \dots$  sont les temps consécutifs d'entrée dans les états  $J_0, J_1, \dots, J_n, \dots$  avec  $J_n \neq J_{n+1}, \forall n \geq 0$ .  $(J_n)_{n \geq 0}$  est une chaîne de Markov homogène à valeurs dans l'espace d'état fini  $E = \{1, \dots, s\}$ .  $X_0, X_1, X_2, \dots$  définis par  $X_0 = 0$  et  $X_n = S_n - S_{n-1}$ , représentent les temps de séjour dans ces états.

Un processus semi-Markovien homogène peut être entièrement déterminé par

- (i) sa loi initiale :  $\Pr(J_0 = k) = p(k)$
- (ii) le noyau semi-Markovien

$$\begin{aligned} Q_{ij}(d) &= \Pr(J_{n+1} = j, X_{n+1} \leq d \mid J_0, \dots, J_n = i, X_1, \dots, X_n) \\ &= \Pr(J_{n+1} = j, X_{n+1} \leq d \mid J_n = i), \end{aligned} \tag{IV.1}$$

Les probabilités de transition de la chaîne de Markov  $(J_n)_{n \geq 0}$  sont définis par

$$\begin{aligned} p_{ij} &= \lim_{d \rightarrow \infty} Q_{ij}(d) \\ &= \Pr(J_{n+1} = j \mid J_0, J_1, \dots, J_n = i) \\ &= \Pr(J_{n+1} = j \mid J_n = i) \end{aligned} \tag{IV.2}$$

Le processus  $(J_n)_{n \geq 0}$  est une chaîne de Markov sous-jacente qui ne gère pas le temps, mais seulement la séquence des états. Les temps de séjour dans les états sont renseignés par le processus  $(X_n)_{n \geq 0}$ . D'après (IV.1), le passé de l'individu est résumé uniquement par le dernier état visité (Markov d'ordre 1). Notons que dans les définitions (IV.1) et (IV.2), l'indice  $n$  n'a pas d'importance car le processus est homogène sur le temps chronologique. Grâce à l'hypothèse d'homogénéité, ces quantités dépendent de l'état précédent et non du couple (état, temps d'entrée). Notons que  $J_n \neq J_{n+1}$ , c'est-à-dire qu'une transition vers le même état est impossible, ainsi,  $Q_{ii}(d) = 0 \forall i \in E$ , ( $p_{ii} = 0, \forall i \in E$ ).

La fonction de distribution du temps de séjour dans l'état  $i$  avant d'aller dans l'état  $j$ , est définie par

$$\begin{aligned} F_{ij}(d) &= \Pr(X_{n+1} \leq d \mid J_n = i, J_{n+1} = j) \\ &= \begin{cases} \frac{\Pr(X_{n+1} \leq d, J_{n+1} = j \mid J_n = i)}{\Pr(J_{n+1} = j \mid J_n = i)} & \text{si } \Pr(J_{n+1} = j \mid J_n = i) > 0 \\ 0 & \text{sinon} \end{cases} \\ &= \begin{cases} \frac{Q_{ij}(d)}{p_{ij}} & \text{si } p_{ij} > 0 \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

$S_{ij}(\cdot) = 1 - F_{ij}(\cdot)$  est la fonction de survie correspondant à  $F_{ij}(\cdot)$ . Dans la pratique, l'état dans lequel va passer le processus est incertain. Ainsi, on définit la fonction de distribution du temps de séjour dans l'état  $i$ ,

$$\begin{aligned} H_i(d) &= \Pr(X_{n+1} \leq d \mid J_n = i) \\ &= \sum_{j=1}^s \Pr(X_{n+1} \leq d, J_{n+1} = j \mid J_n = i) \\ &= \sum_{j=1}^s Q_{ij}(d) \\ &= \sum_{j=1}^s F_{ij}(d)p_{ij} \end{aligned} \tag{IV.3}$$

On peut définir les fonctions de survie correspondantes,  $S_i(\cdot) = 1 - H_i(\cdot)$  et  $S_{ij}(\cdot) = 1 - F_{ij}(\cdot)$ ,

$$\begin{aligned} S_i(d) &= \Pr(X_{n+1} > d \mid J_n = i) \\ &= \sum_{j=1}^s p_{ij} S_{ij}(d) \end{aligned} \tag{IV.4}$$

$S_i(d)$  représente la probabilité que l'individu survive dans l'état  $J_n = i$  jusqu'au temps  $S_n + d$ .

En supposant que  $Q_{ij}(\cdot)$  est absolument continue par rapport à la mesure de Lebesgue, on peut définir  $q_{ij}(\cdot)$  la densité de  $Q_{ij}(\cdot)$ ,

$$q_{ij}(d) = \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta t} \Pr(J_{n+1} = j, d < X_{n+1} \leq d + \Delta d \mid J_n = i),$$

et  $f_{ij}(\cdot)$  la densité de  $F_{ij}(\cdot)$ ,

$$\begin{aligned} f_{ij}(d) &= \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \Pr(d < X_{n+1} \leq d + \Delta d \mid J_n = i, J_{n+1} = j) \\ &= \begin{cases} \frac{q_{ij}(d)}{p_{ij}} & \text{si } p_{ij} > 0 \\ 0 & \text{sinon.} \end{cases} \end{aligned} \quad (\text{IV.5})$$

Par définition, les intensités de transition instantanées du noyau semi-Markovien,  $\forall i, j \in E$ , sont données par l'expression suivante

$$\begin{aligned} \lambda_{ij}(d) &= \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \Pr(J_{n+1} = j, d < X_{n+1} \leq d + \Delta d \mid J_n = i, X_{n+1} > d) \\ &= \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \frac{\Pr(J_{n+1} = j, d < X_{n+1} \leq d + \Delta d, X_{n+1} > d \mid J_n = i)}{\Pr(X_{n+1} > d \mid J_n = i)} \\ &= \begin{cases} \frac{q_{ij}(d)}{1-H_i(d)} & \text{si } p_{ij} > 0 \text{ et } H_i(d) < 1 \\ 0 & \text{sinon.} \end{cases} \\ &= \begin{cases} \frac{p_{ij}f_{ij}(d)}{S_i(d)} & \text{si } p_{ij} > 0 \text{ et } S_i(d) > 0 \\ 0 & \text{sinon.} \end{cases} \end{aligned} \quad (\text{IV.6})$$

En fait,  $\lambda_{ij}(d)\Delta d + o(\Delta d)$ ,  $i \neq j$ , représente la probabilité que le processus transite dans l'état  $j$  dans  $]d, d + \Delta d]$  sachant qu'il est resté un temps  $d$  dans l'état  $i$ .

Cette fonction de risque du processus semi-Markovien ne doit pas être confondue avec la fonction de risque de la loi des temps de séjour qui suppose, par définition, que l'état d'arrivée est connu. Les fonctions de risque des temps de séjour sont données par

$$\tilde{\alpha}_{ij}(d) = \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \Pr(d < X_{n+1} \leq d + \Delta d \mid X_{n+1} > d, J_n = i, J_{n+1} = j). \quad (\text{IV.7})$$

**Remarque 13** Dans les définitions, l'indice  $n$  renseigne sur le numéro de la transition,  $S_n$  sur les temps de sauts,  $J_n$  sur les états et  $X_{n+1}$  sur la durée écoulé dans l'état  $J_n$ . D'autres notations issues des processus de comptage peuvent être utilisées. On considère  $N_{ij}(t)$  un processus de comptage du nombre de transitions observées de l'état  $i$  vers l'état  $j$  dans  $[0, t]$ , où  $t$  est le temps chronologique,

$$N_{ij}(t) = \sum_{n \geq 0} \mathbb{1}_{\{S_n \leq t, J_n = i, J_{n+1} = j\}}(t).$$

$N_{ij}(\cdot)$  est un processus càdlàg avec des sauts de 1,  $N_{ij}(0) = 0$ . On définit ensuite

$$N(t) = \sum_{i, j \in E} N_{ij}(t),$$

comptant le nombre total de transitions observées dans  $[0, t]$ . Ainsi, l'état du processus au temps  $t$  (souvent noté  $J(t)$  dans la littérature) peut s'écrire  $J_{N(t)} : (J_{N(t)}, t \in \mathbb{R}^+)$  contient la même information que le processus  $Z = (J_n, S_n)_{n \geq 0}$ .

## 2.2 Probabilités de transition du processus semi-Markovien

Dans cette section, on s'intéresse à la probabilité de transition du processus semi-Markovien. Comme nous l'avons vu dans la section précédente (*cf.* Remarque 13), le processus semi-markovien  $Z = (J_n, S_n)_{n \geq 0}$  peut aussi s'écrire  $Z(t) = J_{N(t)}$  où  $J_{N(t)}$  représente l'état du processus au temps  $t$  (*i.e.* l'état du processus après  $N(t)$  sauts). Les probabilités de transition du processus  $P_{ij}(s, t)$  sont définies par

$$P_{ij}(s, t) = \Pr(J_{N(t)} = j \mid J_{N(s)} = i)$$

Par la propriété d'homogénéité de la chaîne de Markov sous-jacente  $(J_n)_{n \geq 0}$ , les probabilités de transition vérifient la propriété suivante,

$$\begin{aligned} P_{ij}(t, t+d) &= \Pr(J_{N(t+d)} = j \mid J_{N(t)} = i) \\ &= \Pr(J_{N(d)} = j \mid J_{N(0)} = i) \\ &= P_{ij}(0, d) \\ &= \Psi_{ij}(d). \end{aligned}$$

La propriété d'homogénéité de la chaîne de Markov est transmise au processus semi-Markovien.

Les probabilités de transition du processus semi-Markovien homogène sont définies par l'équation suivante,

$$\begin{aligned} \Psi_{ij}(t) &= \Pr(Z(t) = j \mid Z(0) = i) \\ &= \sum_{r=1}^s \int_0^t p_{ir} f_{ir}(u) \Psi_{rj}(t-u) du + \delta_{ij} \sum_{r=1}^s p_{ir} S_{ir}(t) \end{aligned} \quad (\text{IV.8})$$

où  $\delta_{ij} = 1$  si  $i = j$  et 0 sinon. Afin d'obtenir l'écriture des probabilités  $\Psi_{ij}(t)$ , on peut considérer les deux cas suivants :

- Cas  $i \neq j$  : dans ce cas, au moins un événement s'est produit entre  $[0, t]$  (car une transition vers le même état est impossible). Pour écrire la probabilité de transition, on considère un conditionnement sur le premier événement qui est couramment utilisé dans la théorie du renouvellement. Ainsi la probabilité  $\Psi_{ij}(t)$  peut s'écrire,

$$\begin{aligned} \Psi_{ij}(t) &= \Pr(J_{N(t)} = j \mid J_{N(0)} = i) \\ &= \Pr(J_{N(t)} = j, X_1 \leq t \mid J_0 = i) \\ &= \sum_{r=1}^s \Pr(J_{N(t)} = j, X_1 \leq t, J_1 = r \mid J_0 = i) \end{aligned} \quad (\text{IV.9})$$

Il faut déterminer la probabilité que le processus soit dans l'état  $j$  sachant que l'état initial est  $i$  et que le premier nouvel état est  $r$ . Le calcul de cette probabilité peut s'obtenir dans un cadre formel à l'aide du produit de convolution. Cependant, l'écriture de cette probabilité peut aussi s'obtenir de manière plus intuitive. En effet, en considérant  $x$  le temps d'entrée dans le premier état, la probabilité cherchée s'écrit comme le produit des probabilités suivantes :

- la probabilité que le sujet reste dans l'état  $i$  un temps  $x$ , *i.e.*,  $S_i(x)$ ,
- la probabilité qu'il transite de l'état  $i$  vers l'état  $r$  au temps  $x$ , *i.e.*,  $\lambda_{ir}(x)$ ,
- la probabilité que l'individu soit dans l'état  $j$  au temps  $t$  sachant qu'il était dans l'état  $r$  au temps  $x$ , *i.e.*,  $P_{rj}(x, t) = \Psi_{rj}(t - x)$ .

La probabilité  $\Psi_{ij}(t)$  s'exprime ensuite à l'aide d'une intégrale afin de prendre en compte tous les temps  $x \in [0, t]$ ,

$$\begin{aligned} \Pr(J_{N(t)} = j, X_1 \leq t, J_1 = r \mid J_0 = i) &= \int_0^t S_i(u) \lambda_{ir}(u) \Psi_{rj}(t - u) du \\ &= \int_0^t p_{ir} f_{ir}(u) \Psi_{rj}(t - u) du. \end{aligned}$$

L'équation (IV.9) permet ensuite d'obtenir la probabilité cherchée.

- Cas  $i = j$  : dans ce cas, soit au moins deux événements se produisent dans  $[0, t]$  soit aucun événement ne se produit.  $\Psi_{ij}(t)$  peut alors s'écrire

$$\begin{aligned} \Psi_{ij}(t) &= \Pr(J_{N(t)} = i \mid J_{N(0)} = i) \\ &= \underbrace{\Pr(J_{N(t)} = i, N(t) > 0 \mid J_0 = i)}_A + \underbrace{\Pr(J_{N(t)} = i, N(t) = 0 \mid J_0 = i)}_B \end{aligned}$$

$A$  se calcule de manière identique au cas  $i \neq j$ . La probabilité  $B$  s'exprime en terme de survie (IV.4),

$$B = \Pr(X_1 > t \mid J_{N(0)} = i) = S_i(t) = \sum_{r=1}^s p_{ir} S_{ir}(t).$$

Ainsi la probabilité cherchée est,

$$\Psi_{ij}(t) = \sum_{r=1}^s \int_0^t p_{ir} f_{ir}(u) \Psi_{rj}(t - u) du + \sum_{r=1}^s p_{ir} S_{ir}(t).$$

L'utilisation du symbole de Kroneker  $\delta_{ij}$  permet de généraliser l'écriture des  $\Psi_{ij}(t)$  aux deux cas possibles.

### 2.3 Fonction de vraisemblance

Considérons un échantillon de  $n$  individus ( $h = 1, \dots, n$ ). Pour l'individu  $h$ , considérons  $0 = S_{h,0} < S_{h,1} < \dots < S_{h,N_h}$  les temps de sauts. A ces temps, l'individu a successivement occupé les états  $J_{h,0}, J_{h,1}, \dots, J_{h,N_h}$  avec  $J_{h,p} \neq J_{h,p+1}$ , où  $N_h$  est le nombre de sauts du processus associé à l'individu  $h$  à la date de point ( $N_h \geq 0$ ). En utilisant ces notations,  $S_{h,N_h}$  représente le dernier temps d'entrée dans un état et  $J_{h,N_h}$  le dernier état occupé par l'individu  $h$ . Considérons  $X_{h,p} = S_{h,p} - S_{h,p-1}$ , le temps de séjour dans l'état  $J_{h,p-1}$ ,  $p = 1, \dots, N_h$ .

Dans tout ce chapitre, on considère le cas de données censurées à droite. De plus, on suppose que la censure n'apporte aucune information sur l'événement étudié (censure indépendante). Dans le cas semi-Markovien, la censure à droite empêche l'observation du temps de séjour dans le dernier état visité. Notons qu'il est supposé que les temps de transition entre les états correspondent aux temps d'observation (de la base de données). Dans le cas des modèles avec états absorbants, on rencontre de deux types d'« histoires ».

— Soit l'individu  $h$  rentre dans un état absorbant, son « histoire » n'est pas censurée,

$$\mathcal{H}_h(t) = (J_{h,0}, J_{h,1}, \dots, J_{h,N_h}, X_{h,1}, X_{h,2}, \dots, X_{h,N_h}).$$

— Soit l'individu  $h$  n'entre pas dans un état absorbant. Dans ce cas, le temps de séjour dans le dernier état visité ( $J_{h,N_h}$ ) est censuré à la date de fin d'étude. L'« histoire » de l'individu est

$$\mathcal{H}_h(t) = (J_{h,0}, J_{h,1}, \dots, J_{h,N_h}, X_{h,1}, X_{h,2}, \dots, X_{h,N_h}, U_h).$$

où  $U_h$  représente la durée écoulée entre  $S_{h,N_h}$  et la date de fin d'étude.

Ainsi, deux types de contribution à la vraisemblance sont possibles.

(i) Soit l'individu reste un temps  $d$  dans l'état  $i$  et ensuite, il transite dans l'état  $j$  ( $j \neq i$ ), dans ce cas où la transition est observée, la contribution à la vraisemblance est

$$S_i(d)\lambda_{ij}(d) = p_{ij}f_{ij}(d) = q_{ij}(d).$$

(ii) Soit l'individu reste un temps  $d$  dans l'état  $i$  et il est censuré. Dans ce cas de censure à droite où le temps de séjour est censuré, la contribution à la vraisemblance s'exprime en terme de survie par

$$S_i(d) = \sum_{j=1}^s p_{ij}S_{ij}(d).$$

Ainsi, en considérant  $\delta_h$  qui vaut 1 si l'individu est censuré et 0 sinon, la contribution à la vraisemblance de l'individu  $h$  s'écrit

$$L_h = \prod_{k=1}^{N_h} p_{J_{h,k-1}J_{h,k}} f_{J_{h,k-1}J_{h,k}}(X_{h,k}) \times [S_{J_{h,N_h}}(U_h)]^{\delta_h}$$

La vraisemblance totale est obtenue en faisant le produit des contributions individuelles,

$$L = \prod_{h=1}^n L_h. \quad (\text{IV.10})$$

## 3 Estimation paramétrique des temps de séjour

### 3.1 Introduction

La méthode d'estimation paramétrique repose sur une estimation des lois des temps de séjour par des fonctions paramétriques. Rappelons la définition des fonctions de risque des temps d'attente dans les états,

$$\alpha_{ij}(d) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta d} \Pr(d < X_{n+1} \leq d + \Delta d \mid X_{n+1} > d, J_n = i, J_{n+1} = j)$$

L'estimation paramétrique va supposer que les fonctions de risque  $\alpha_{ij}(\cdot)$  appartiennent à une classe de fonctions paramétriques. Les fonctions  $S_{ij}(\cdot)$  et  $f_{ij}(\cdot)$  correspondant respectivement aux fonctions de survie et de densité associées peuvent s'écrire en fonction de  $\alpha_{ij}(\cdot)$ .

$$\begin{aligned} \frac{\partial S_{ij}(d)}{\partial d} &= - \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \Pr(d < X_{n+1} \leq d + \Delta d \mid J_n = i, J_{n+1} = j) \\ &= - \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \Pr(d < X_{n+1} \leq d + \Delta d \mid X_{n+1} > d, J_n = i, J_{n+1} = j) \\ &\quad \times \Pr(X_{n+1} > d \mid J_n = i, J_{n+1} = j) \\ &= -S_{ij}(d) \times \lim_{\Delta d \rightarrow 0} \frac{1}{\Delta d} \Pr(d < X_{n+1} \leq d + \Delta d \mid X_{n+1} > d, J_n = i, J_{n+1} = j) \\ &= -S_{ij}(d) \times \alpha_{ij}(d). \end{aligned}$$

La résolution cette équation différentielle sachant que  $S_{ij}(0) = 1$ , donne

$$S_{ij}(d) = \exp\left(- \int_0^d \alpha_{ij}(u) du\right). \quad (\text{IV.11})$$

De plus, comme  $S_{ij}(\cdot) = 1 - F_{ij}(\cdot)$  et comme  $f_{ij}(\cdot)$  est la densité de  $F_{ij}(\cdot)$ , on a

$$f_{ij}(d) = - \frac{\partial S_{ij}(d)}{\partial d} = S_{ij}(d) \alpha_{ij}(d). \quad (\text{IV.12})$$

## 3.2 Modèle à risques proportionnels

Afin d'introduire des covariables on considère un modèle à risques proportionnels de "Cox".

Considérons  $(J_n, S_n)_{n \geq 0}$  un processus semi-Markovien. Les covariables sont introduites dans les fonctions de risque des temps d'attente dans les états. La chaîne de Markov sous-jacente  $(J_n)_{n \geq 0}$  ne dépend pas du vecteur de covariables et ainsi la chaîne conserve la probabilité de transition  $p_{ij} = \Pr(J_{n+1} = j \mid J_n = i)$ . Les fonctions de risque du processus semi-Markovien ne dépendent pas directement des covariables. Cependant comme ils sont définis à l'aide des risques de temps d'attente, l'effet des covariables sera quand même répercuté même s'il ne pourra pas s'interpréter en termes de risque relatif.

Considérons  $\mathbf{z}_{ij}(\cdot) = (z_{ij}^1(\cdot), z_{ij}^2(\cdot), \dots, z_{ij}^{n_{ij}}(\cdot))$  un vecteur de covariables associé à la transition de l'état  $i$  vers l'état  $j$ , tel que  $n_{ij}$  soit le nombre de covariables pour cette transition. Cette notation permet d'ajuster un modèle où le nombre de covariables est spécifique à chaque transition. Les covariables peuvent être dépendantes du temps, cependant il est nécessaire de supposer que la valeur des covariables ne change pas entre deux consultations. Afin de simplifier les calculs et les écritures, on supposera dans ce qui suit que les covariables sont fixées au cours du temps d'attente :  $\mathbf{z}_{ij}(t) = \mathbf{z}_{ij}$ . On suppose que les covariables vont modifier les fonctions d'intensité en suivant un modèle à risques proportionnels de Cox.

$$\alpha_{ij}(t - S_{N(t^-)}) = \alpha_{ij,0}(t - S_{N(t^-)}) \exp(\boldsymbol{\beta}_{ij}^T \mathbf{z}_{ij}), \quad (\text{IV.13})$$

où  $\beta_{ij}$  est le vecteur des coefficients de régression associés à  $\mathbf{z}_{ij}$  et  $\alpha_{ij,0}(\cdot)$  est le risque de base. Dans ce modèle, la proportionnalité des risques est supposée au sein d'une même transition.

Intéressons nous maintenant aux fonctions de survie et de densité correspondant à des fonctions de risque de temps d'attente dépendantes de covariables. Considérons  $\forall i, j \in E$ ,  $\alpha_{ij}(d, \mathbf{z}) = \alpha_{ij,0}(d)e^{\beta_{ij}^T \mathbf{z}_{ij}}$ , alors d'après les équations (IV.11) et (IV.12) les fonctions de survie correspondantes sont données par

$$\begin{aligned} S_{ij}(d, \mathbf{z}) &= \exp\left(-\int_0^d \alpha_{ij}(u) du\right) \\ &= \exp\left(-\int_0^d \alpha_{ij,0}(u) e^{\beta_{ij}^T \mathbf{z}_{ij}} du\right) \\ &= S_{ij,0}(d) e^{\beta_{ij}^T \mathbf{z}_{ij}} \end{aligned} \quad (\text{IV.14})$$

où  $S_{ij,0}(d) = \exp\left(-\int_0^d \alpha_{ij,0}(u) du\right)$  et les fonctions de densité sont

$$\begin{aligned} f_{ij}(d, \mathbf{z}) &= S_{ij}(d, \mathbf{z}) \alpha_{ij}(d, \mathbf{z}) \\ &= \alpha_{ij,0}(d) e^{\beta_{ij}^T \mathbf{z}_{ij}} S_{ij,0}(d) e^{\beta_{ij}^T \mathbf{z}_{ij}}. \end{aligned} \quad (\text{IV.15})$$

### 3.3 Modélisation paramétrique de la loi de séjour dans l'état

La modélisation paramétrique consiste à estimer les fonctions de risque des temps d'attente dans les états par des fonctions paramétriques. Ainsi, on suppose que  $\alpha_{ij}(t) = g_{ij}(t, \boldsymbol{\theta}_{ij})$ , où  $g_{ij}$  est une fonction paramétrique intégrable. L'estimation de  $\alpha_{ij}(\cdot)$  consiste à estimer le vecteur de paramètres  $\boldsymbol{\theta}_{ij}$  (qui inclut les coefficients de régression).

D'après l'équation (IV.4), la vraisemblance (IV.10) peut s'écrire

$$L = \prod_{h=1}^n \left\{ \prod_{k=1}^{N_h} p_{J_{h,k-1} J_{h,k}} f_{J_{h,k-1} J_{h,k}}(X_{h,k}) \times \left[ \sum_{j=1}^s p_{J_{h,N_h} j} S_{J_{h,N_h} j}(U_h) \right]^{\delta_h} \right\}. \quad (\text{IV.16})$$

A partir des équations (IV.14) et (IV.15), la vraisemblance (IV.16) peut s'écrire en fonction des paramètres  $p_{ij}$  et  $\boldsymbol{\theta}_{ij}$ . L'estimation des paramètres se fait ensuite par maximisation de la vraisemblance. On obtient ainsi les estimations  $\hat{p}_{ij}$  des probabilités de la chaîne de Markov et les estimations  $\hat{\alpha}_{ij}(\cdot)$  des fonctions de risque des temps de séjour. On en déduit les estimateurs  $\hat{f}_{ij}(\cdot)$  des fonctions de densité et  $\hat{S}_{ij}(\cdot)$  des fonctions de survie. De plus, d'après (IV.6) et (IV.4), il est possible de déduire les estimateurs  $\hat{\lambda}_{ij}(\cdot)$  des intensités du processus semi-Markovien par la formule suivante

$$\hat{\lambda}_{ij}(d) = \frac{\hat{p}_{ij} \hat{f}_{ij}(d)}{\sum_{j=1}^s \hat{p}_{ij} \hat{S}_{ij}(d)}. \quad (\text{IV.17})$$

On peut noter que la notation  $\alpha_{ij}(d) = g_{ij}(d, \theta_{ij})$  permet de considérer des natures de fonctions différentes suivant la transition étudiée. De plus, le nombre de paramètres qui définissent la fonction est spécifique à chaque transition. En pratique, cette écriture est très utile car elle permet d'adapter la modélisation à chaque transition et elle permet ainsi d'optimiser le nombre de paramètres.

Dans les études de survie, les familles de fonctions couramment utilisées pour modéliser les risques sont les lois exponentielles, les lois de Weibull et les lois de Weibull généralisées (qui ont l'avantage d'être «emboîtées»). D'autres choix de distributions sont possibles, mais il est important de faire un compromis entre la taille de la base et le nombre de paramètres qui définissent la loi.

### 3.4 Extension à un modèle semi-Markovien non-homogène

Dans le modèle semi-Markovien homogène, l'évolution du processus dépend uniquement de la durée écoulée dans l'état. En effet, les probabilités de transition du processus ne dépendent pas du temps (chronologique) où se produit la transition. Cette hypothèse d'homogénéité peut, dans certaines applications, être trop restrictive.

Afin de rendre le modèle non-homogène, on peut considérer que les probabilités de la chaîne de Markov sous-jacente dépendent de la durée du suivi  $t$ . On définit les probabilités de transition de la chaîne de Markov sous-jacente, qui dépendent du temps  $t$ ,

$$p_{ij}(t) = \Pr(J(t) = j \mid J(t^-) = i),$$

où  $J(t)$  représente l'état occupé par le processus au temps  $t$ . Il est ensuite possible de redéfinir les fonctions de survie, la fonction de vraisemblance et les probabilités de transition du processus de manière à prendre en compte la durée de suivi dans ces quantités.

La méthode d'estimation paramétrique présentée précédemment peut ensuite être étendue afin d'estimer de manière paramétrique les probabilités  $p_{ij}(\cdot)$ . On peut supposer, par exemple, que les probabilités sont de la forme

$$\hat{p}_{ij}(t) = \frac{V_{ij}(t)}{\sum_{\{(i,j) \in E \times E\}} V_{ij}(t)},$$

avec

$$V_{ij}(t) = \begin{cases} \frac{\exp(a_{ij} \times t + b_{ij})}{1 + \exp(a_{ij} \times t + b_{ij})} & i \neq j \\ 0 & i = j. \end{cases}$$

Le choix de ces fonctions est motivé par le fait que  $\hat{p}_{ij}(\cdot)$  est compris entre 0 et 1 et que  $\sum_{j=1}^n \hat{p}_{ij}(\cdot) = 1$ . L'estimation des  $p_{ij}(\cdot)$  consiste alors à estimer les paramètres  $a_{ij}$  et  $b_{ij}$  par maximisation de la vraisemblance.

En pratique, ce modèle et cette méthode d'estimation sont difficilement utilisables. En effet, le nombre de paramètres devient très important et les estimations deviennent difficiles. De plus, les résultats obtenus avec ce type de modèle dépendent de deux échelles de temps ce qui rend l'interprétation délicate.



## Chapitre V

# Bibliographie

- Andersen P. K., Borgan  $\emptyset$ ., Gill R. D., Keiding N. *Statistical Models Based on Counting Processes*. Springer, 1993.
- Fleming T. R., Harrington D. P. *Counting Processes and Survival Analysis*. Wiley, 1991.