



# Utilisation de modèles multi-états en assurance de personnes

L'exemple de la modélisation des  
garanties arrêt de travail

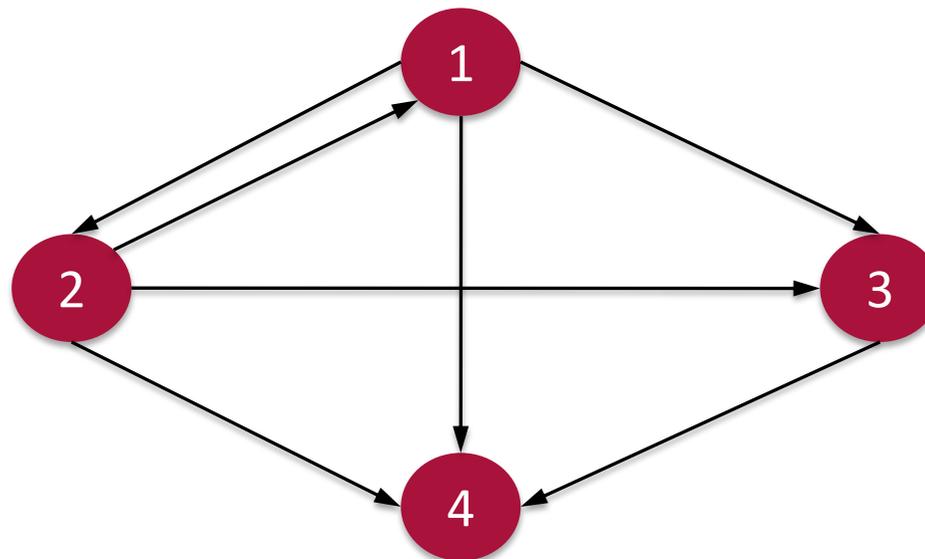
**Frédéric PLANCHET**  
frederic.planchet@univ-lyon1.fr

**Quentin GUIBERT**  
qguibert@winter-associes.fr



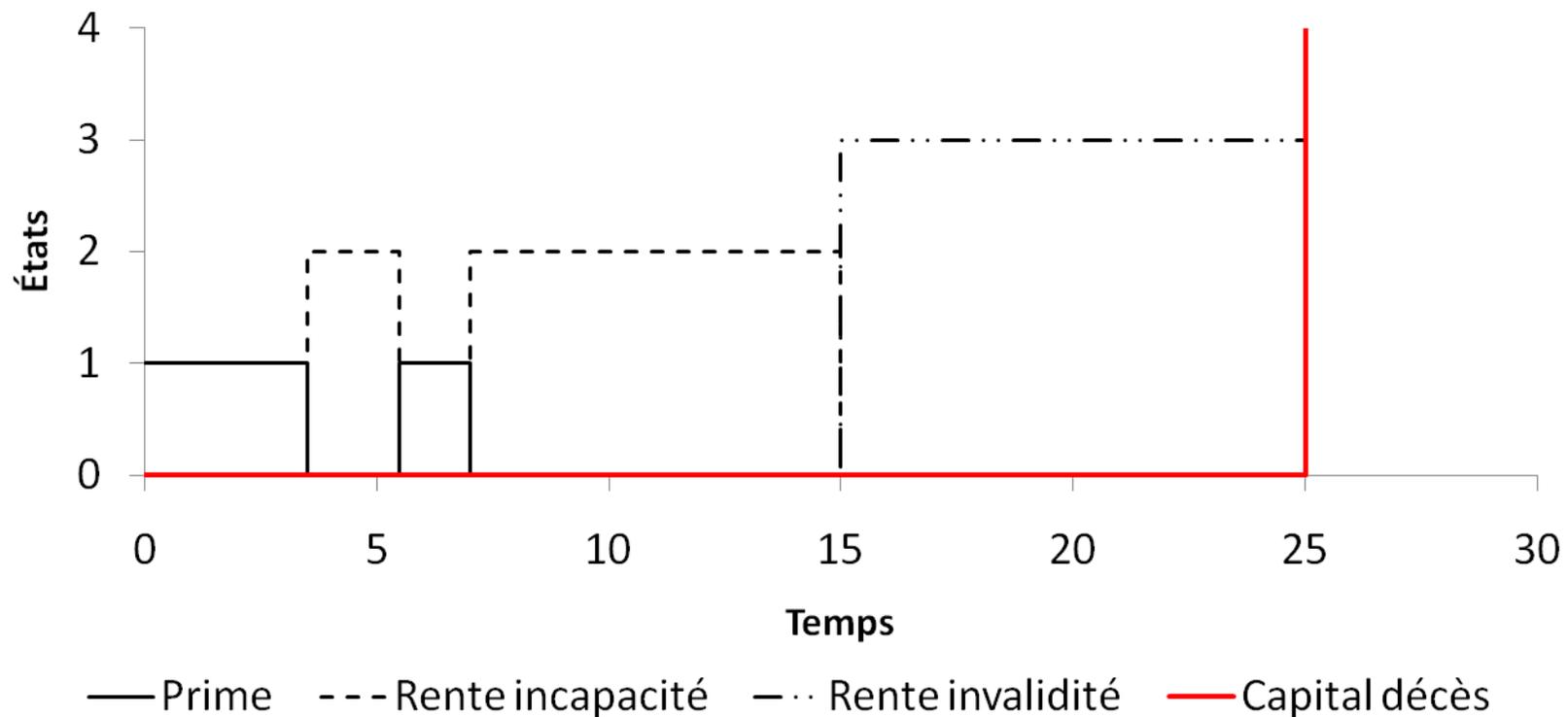
En arrêt de travail, le calcul d'une provision *best estimate* et le suivi des risques nécessitent de s'intéresser aux différents **états** parcourus par l'assuré pendant la période de couverture. Quatre états peuvent être identifiés :

1. actif ;
2. incapacité de travail ;
3. invalidité de travail ;
4. décès.





Chaque changement d'états donne lieu à des flux financiers différents soit pour l'assureur, soit pour l'assuré.





Pour un assuré actif (état 1) d'âge  $x$  en  $t$ , le *best estimate* s'écrit si  $t < T$  :

$$BEL(x) = -\int_t^T v(t, \tau) \cdot {}_{\tau-t}\bar{p}_x^{11} \cdot \pi d\tau \\ + \sum_{j \neq 1} \int_t^T v(t, \tau) \cdot {}_{\tau-t}\bar{p}_x^{11} \cdot \mu_{x+\tau-t}^{1j} \cdot [K_{1j} + BEL_j(x + \tau, 0)] d\tau$$

Pour un assuré incapable ou invalide (états 2 et 3) d'âge  $x$  en  $t$  et présent depuis une durée  $u$ , le *best estimate* s'écrit si  $t - u < T$  :

$$BEL_i(x, u) = \int_{t-u+f \wedge t}^{t-u+f+d_{\max}} v(t, \tau) \cdot {}_{\tau-t}\bar{p}_{x,u}^{ii} \cdot r_i d\tau \\ + \sum_{j \neq i} \int_t^{t-u+f+d_{\max}} v(t, \tau) \cdot {}_{\tau-t}\bar{p}_x^{ii} \cdot \mu_{x+\tau-t}^{ij} \cdot [K_{ij} + BEL_j(x + \tau, 0)] d\tau$$



Les approches classiques évaluent le coût d'une garantie arrêt de travail en considérant :

- des lois d'incidence pour décrire la survenance de l'incapacité, de l'invalidité et du décès ;
- des lois de maintien en incapacité et en invalidité.

Cependant, pour construire ces lois, les estimations sont réalisées de manière marginale en ne considérant que deux états à la fois (ex : Kaplan-Meier) ce qui biaise l'estimation réalisée. Par exemple, pour l'incidence en incapacité :

$$P_{12}(0,t) = (1 - P_{\underline{11}}(t)) \neq \int_0^t P_{11}(0,u) \mu_{12}(u) P_{\underline{22}}(u,t) du$$

Ici, l'**interdépendance** entre les états est ignorée.



La construction de lois d'expérience *best estimate* se devrait de tenir compte de l'interaction entre les différents états, sous peine de générer un biais dans l'estimation des engagements.

L'objet de cette présentation est de décrire un cadre de travail pour l'estimation de lois d'expérience permettant de tenir compte de ces interactions. Celui-ci s'appuie sur l'utilisation de modèles **multi-états** qui permettent d'avoir une représentation détaillée de chaque loi de transition.



## SOMMAIRE

1. *Le cadre multi-états*
2. Estimation des lois de transition
3. Prise en compte de l'hétérogénéité
4. Validation

# 1. *Le cadre multi-états*



Les modèles **multi-états** constituent une classe particulière de modèle de durée qui s'utilise de manière naturelle dans les situations où l'assuré est susceptible d'évoluer entre plusieurs états. Il permet de déterminer les lois de transition d'un état vers un autre.

Toutefois, on distingue plusieurs **type** de modèles multi-états :

- **markovien homogène** : les intensités de transition sont **constantes** ;
- **markovien inhomogène** : les intensités de transition dépendent du **temps** ;
- **semi markovien homogène** : les intensités de transition dépendent de la **durée** passée dans l'état ;
- **semi markovien inhomogène** : les intensités de transition dépendent de la **durée** passée dans l'état et du **temps**.

# 1. Le cadre multi-états



En France, l'utilisation des modèles multi-états est peu rependue (*cf.* Rullière et Serant [1998] ou Denuit et Robert [2007]), notamment pour la construction de lois d'expérience. Il s'agit de modèles qui sont plus largement utilisés dans le cadre de l'invalidité (PHI) en Grande Bretagne (*cf.* Haberman et Pitacco [1998]).

Par ailleurs, ce type de modèle est couramment utilisé :

- en épidémiologie (*cf.* Saint Pierre [2005] ou Andersen et Keiding [2002]) ;
- pour la construction de matrices de transition d'un *rating* à l'autre pour les problématiques de crédit (*cf.* Janssen et Manca [2007], Moteiro et *al.* [2006] ou Lando et Skodeberg [2002]).

Dans le cadre de **l'arrêt de travail**, les modèles **markoviens inhomogènes** ou **semi markovien inhomogènes** constituent les classes de modèle la plus souvent retenues.

# 1. Le cadre multi-états



## 1.1. Modélisation d'un contrat arrêt de travail

Pour un assuré d'âge  $x$ , sa police peut être modélisée par un processus multi-états contenant **4 états**. L'état et la durée de transition  $(X(t), D(t))$  sont supposés formés un processus **markovien**.

États	Probabilité de transition
Sain (âge)	$P_{1j}(s, t) = \Pr\{X(t) = j   X(s) = 1\} = {}_{t-s}P_{x+s}^{1j}$
Incapable / Invalide (âge et durée)	$P_{ij}(s, t, u, v) = \Pr\{X(t) = j \wedge D(t) \leq v   X(s) = i \wedge D(s) = u\}$ $= {}_{v, t-s}P_{x+t, u}^{ij}$

# 1. Le cadre multi-états



## 1.2. Intensité de transition (ou fonction de hasard)

De la même manière que pour un modèle de durée classique, on définit des **intensités de transition** et **d'intensité cumulée**.

États	Intensité de transition	Intensité cumulée
Sain (âge)	$\mu_x^{1j} = \lim_{\Delta t \rightarrow 0} \frac{\Delta t P_x^{1j}}{\Delta t}$	$A_x^{1j} = \int_0^x \mu_t^{1j} dt$
Incapable / Invalide (âge et durée)	$\mu_{x,u}^{ij} = \lim_{\Delta t \rightarrow 0} \frac{\Delta t P_{x,u}^{ij}}{\Delta t}$	$A_{x,u}^{ij} = \int_0^x \mu_{t,u}^{ij} dt$
Par ailleurs, on note :	$\mu_x^{i\bullet} = -\sum_{j \neq i} \mu_x^{ij}$	$\mu_{x,u}^{i\bullet} = -\sum_{j \neq i} \mu_{x,u}^{ij}$

# 1. *Le cadre multi-états*



## 1.2. Intensité de transition (ou fonction de hasard)

Le fait de considérer que les états **d'incapacité** et **d'invalidité** dépendent de la **durée** rend toutefois le **problème particulièrement complexe** (cf. Janssen et Manca [2006] ou D'Amico et *al.* [2010]).

En pratique, cette hypothèse est le plus souvent relâchée de manière à ce que l'ensemble du problème devienne uniquement **markovien inhomogène**.

La dépendance en fonction de la durée peut être réintroduite dans un deuxième temps par d'autres moyens.

Par conséquent, les intensité de transition depuis les états 2 et 3 deviennent :

$$\mu_x^{ij} = \lim_{\Delta t \rightarrow 0} \frac{\Delta t P_x^{ij}}{\Delta t}$$

# 1. *Le cadre multi-états*



## 1.3. Expression des probabilités de transition

Le cadre **markovien** permet de définir une relation fondamentale entre les probabilités de transition (équation de Chapman-Kolmogorov) :

$$\forall i, j \text{ et } \forall s \leq w \leq t, \quad P_{ij}(s, t) = \sum_k P_{ik}(s, w) \cdot P_{kj}(w, t)$$

Cette expression peut se mettre sous forme **matricielle** :

$$\forall s \leq w \leq t, \quad P(s, t) = P(s, w) \cdot P(w, t)$$

# 1. Le cadre multi-états



## 1.3. Expression des probabilités de transition

En utilisant les intensités de transition, les probabilités de transition sont exprimées à partir **d'équations différentielles** (Kolmogorov *forward*). **Par exemple :**

$$\frac{\partial}{\partial t} {}_{t-s}p_{x+s}^{11} = {}_{t-s}p_{x+s}^{12} \cdot \mu_{x+t}^{21} + {}_{t-s}p_{x+s}^{11} \cdot \mu_{x+t}^{1\bullet}$$

$$\frac{\partial}{\partial t} {}_{t-s}p_{x+s}^{12} = {}_{t-s}p_{x+s}^{11} \cdot \mu_{x+t}^{12} + {}_{t-s}p_{x+s}^{12} \cdot \mu_{x+t}^{2\bullet}$$

$$\frac{\partial}{\partial t} {}_{t-s}p_{x+s}^{13} = {}_{t-s}p_{x+s}^{11} \cdot \mu_{x+t}^{13} + {}_{t-s}p_{x+s}^{12} \cdot \mu_{x+t}^{23} - {}_{t-s}p_{x+s}^{13} \cdot \mu_{x+t}^{34}$$

$$\frac{\partial}{\partial t} {}_{t-s}p_{x+s}^{14} = {}_{t-s}p_{x+s}^{11} \cdot \mu_{x+t}^{14} + {}_{t-s}p_{x+s}^{12} \cdot \mu_{x+t}^{24} + {}_{t-s}p_{x+s}^{13} \cdot \mu_{x+t}^{34}$$

# 1. *Le cadre multi-états*



## 1.3. Expression des probabilités de transition

Le cadre multi-états **markovien** prévoit une forme générale pour résoudre ces équations :

$$P(s, t) = \mathcal{P}_{u \in ]s, t]} (\text{Id} + dA_u)$$

L'opération ci-dessus utilise la notion de **produit intégral**. Si les intensités cumulées sont des fonctions étagées, l'expression se simplifie :

$$P(s, t) = \prod_{k=1}^n (\text{Id} + \Delta A_{T_k})$$

où  $(T_k)_{k=1, \dots, n}$  représentent les instants de sauts de la matrice d'intensité cumulée.

# 1. Le cadre multi-états



## 1.3. Expression des probabilités de transition

Au global, il est possible d'obtenir des relations exprimant les **probabilités de survie** dans un état. Par exemple, pour l'état **Actif** :

$$\frac{\partial}{\partial t} {}_{t-s}P_{x+s}^{11} = {}_{t-s}P_{x+s}^{11} \cdot \mu_{x+t}^{1\bullet} \Rightarrow {}_{t-s}P_{x+s}^{11} = \exp \left[ \int_s^t \mu_{x+u}^{1\bullet} du \right]$$

Par ailleurs, il est possible d'obtenir une expression des **probabilités de transition**. Par exemple, pour la transition de l'état **Actif** vers l'état **Incapable** :

$${}_{t-s}P_{x+s}^{12} = \int_s^t {}_uP_{x+s}^{11} \cdot \mu_{x+u}^{12} \cdot {}_{t-u}P_{x+u}^{22} du$$

# 1. *Le cadre multi-états*



## 1.4. Prendre en compte la durée de séjour dans un état

Considérer un cadre de travail **semi-markovien** complexifie les expressions des probabilités de transition et des engagements. Plusieurs solutions permettent néanmoins de contourner cette difficulté :

- se placer dans le cas particulier d'un modèle semi-markovien inhomogène mais en supposant que les intensités de transition **sont constantes par morceaux** (modèle de Poisson) ce qui simplifie les calculs (*cf.* Monteiro *et al.* [2006]);
- intégrer l'hétérogénéité induite par la **durée de séjour** à partir d'un modèle à hasard proportionnel (éventuellement dépendance avec l'âge) (Andersen et Keiting [2002]) ;
- augmenter le nombre d'états en décomposant les états dépendants de la durée (*cf.* Haberman et Pitacco [1998]).

# 1. Le cadre multi-états



## 1.4. Prendre en compte la durée de séjour dans un état

Pour ces dernières approches, en supposant pour des raisons de lisibilité que les états Incapable et Invalide ne forment qu'un seul état Incapable, on retient le schéma suivant :

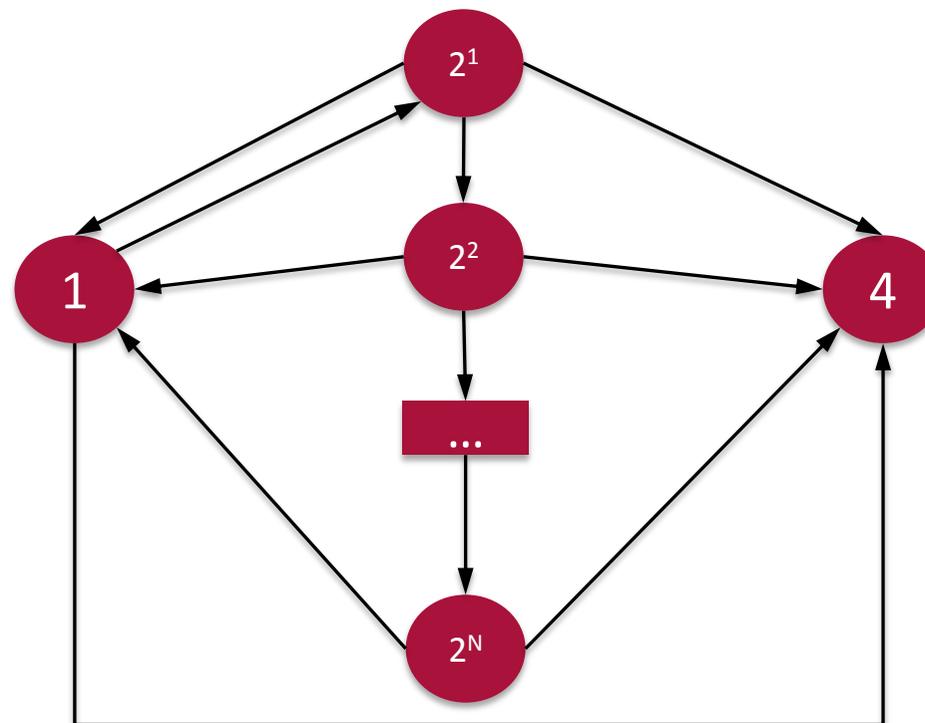
1. Actif

$2^1$ . Incapable depuis moins d'un mois

$2^2$ . Incapable depuis moins de 2 mois mais plus d'un mois

...

$2^N$ . Incapable depuis moins de N mois mais plus de N-1 mois





## SOMMAIRE

1. Le cadre multi-états
2. *Estimation des lois de transition*
3. Prise en compte de l'hétérogénéité
4. Validation

## 2. Estimation des lois de transition



Une fois le cadre du modèle multi états mis en place, les lois de transition peuvent être représentées à partir de plusieurs types de modèles :

- modèles **non paramétriques** : l'estimateur choisi ne s'appuie sur aucune hypothèse *a priori* pour décrire les intensités de transition et doit prendre en compte les censures ;
- modèles **paramétriques** : les intensités de transition ont une forme prédéterminées ;

Les critères de choix de tels modèles sont identiques à ceux retenus pour les modèles de durée classiques (qualité et quantité des données).

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

L'approche la plus couramment retenue pour ajuster un modèle multi-états consiste à utiliser une estimation **non paramétrique**. Un cadre formel existe pour le modèle de type markovien (*cf.* Andersen et *al.* [1993]) et permet d'obtenir une estimation des lois de transition brutes.

Ce cadre d'estimation peut être vu comme une **extension** du modèle de **Kaplan-Meier** (*cf.* Kaplan et Meier [1958]), les modèles de survie classiques étant finalement des modèles multi-états à 2 états.

L'estimation se déroule en 2 étapes :

- utiliser l'estimateur de Nelson-Aalen pour ajuster les intensités cumulée ;
- utiliser la formule reliant les intensité de transition et les probabilité de transition ainsi que l'estimateur de Nelson-Aalen pour avoir l'estimation des probabilités de transition.

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

Pour mener l'estimation non paramétrique des lois de transition, les données (sous l'hypothèse que la volumétrie suffisante) doivent bénéficier d'une attention particulière. Il est en particulier indispensable que **l'ensemble des dates de transition** soient identifiées.

Data:

	id	from	to	trans	Tstart	Tstop	time	status	Sexe	CSP	Region
1	1297	1	2	1	41.32317	47.73687	6.41369863	1	Homme	CADRE	Sud-Est
2	1297	1	3	2	41.32317	47.73687	6.41369863	0	Homme	CADRE	Sud-Est
3	1297	1	4	3	41.32317	47.73687	6.41369863	0	Homme	CADRE	Sud-Est
4	1297	2	1	4	47.73687	47.84932	0.11244292	1	Homme	CADRE	Sud-Est
5	1297	2	3	5	47.73687	47.84932	0.11244292	0	Homme	CADRE	Sud-Est
6	1297	2	4	6	47.73687	47.84932	0.11244292	0	Homme	CADRE	Sud-Est
7	1297	1	2	1	47.84932	49.37523	1.52591324	1	Homme	CADRE	Sud-Est
8	1297	1	3	2	47.84932	49.37523	1.52591324	0	Homme	CADRE	Sud-Est
9	1297	1	4	3	47.84932	49.37523	1.52591324	0	Homme	CADRE	Sud-Est
10	1297	2	1	4	49.37523	49.45742	0.08219178	1	Homme	CADRE	Sud-Est

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

L'estimateur de **Nelson-Aalen** permet dans un premier temps d'estimer les intensités de transition cumulées en prenant en compte la **censure à droite et la troncature à gauche** :

$$A_t^{ij} = \sum_{\{k: T_k \leq t\}} \frac{d_{ij}(T_k)}{n_i(T_k)}$$

où :

- $d_{ij}(s)$  représente le nombre de transition de l'état  $i$  vers  $j$  **exactement** en  $s$  (en pratique égale à 1) ;
- $n_i(s)$  représente le nombre d'individus dans l'état  $i$  en date  $s^-$  (hors censure à droite et troncature à gauche).

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

L'estimateur de **Nelson-Aalen** présente un certain nombre de bonnes propriétés :

- l'estimateur est **très faiblement biaisé** (cas rare où l'exposition est nulle) ;
- l'estimateur de sa **variance** est utile pour la construction d'intervalles ou de bandes de confiance :

$$\sigma_t^{ij} = \sum_{\{k: T_k \leq t\}} \frac{d_{ij}(T_k)}{(n_i(T_k))^2}$$

- l'estimateur est **asymptotiquement gaussien** :

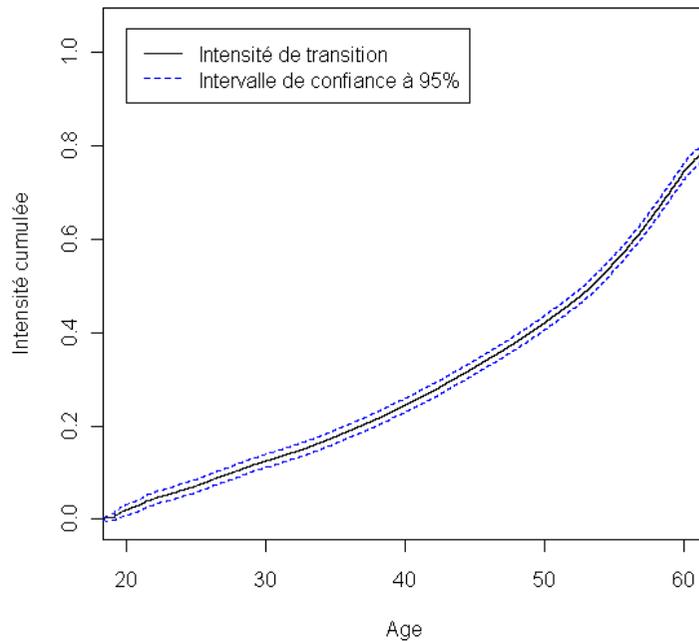
$$\sqrt{n} \left( A_t^{ij} - \Lambda_t^{ij} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V^{ij})$$

## 2. Estimation des lois de transition

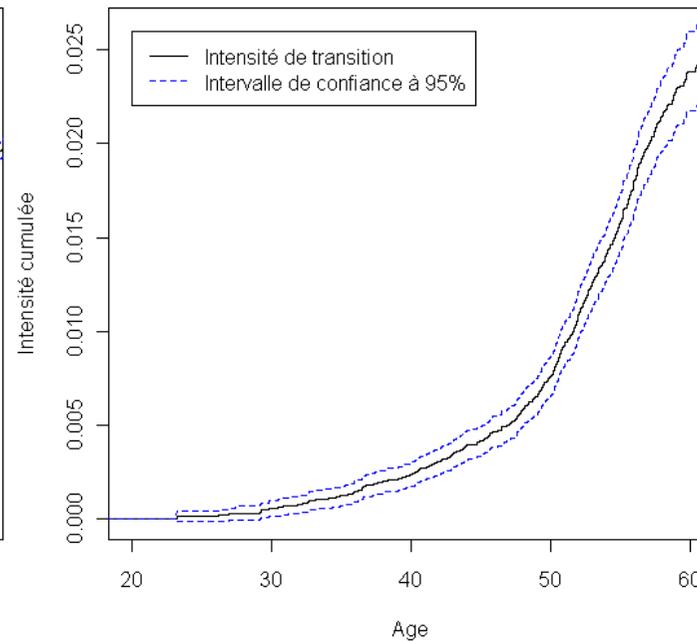


### 2.1. Cadre non paramétrique

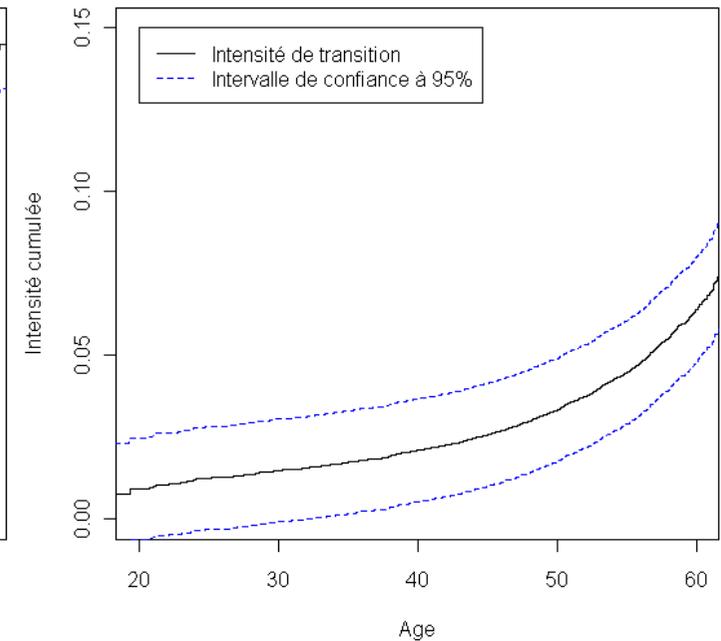
Intensité cumulée de Sain vers Incapable



Intensité cumulée de Sain vers Invalide



Intensité cumulée de Sain vers Décès

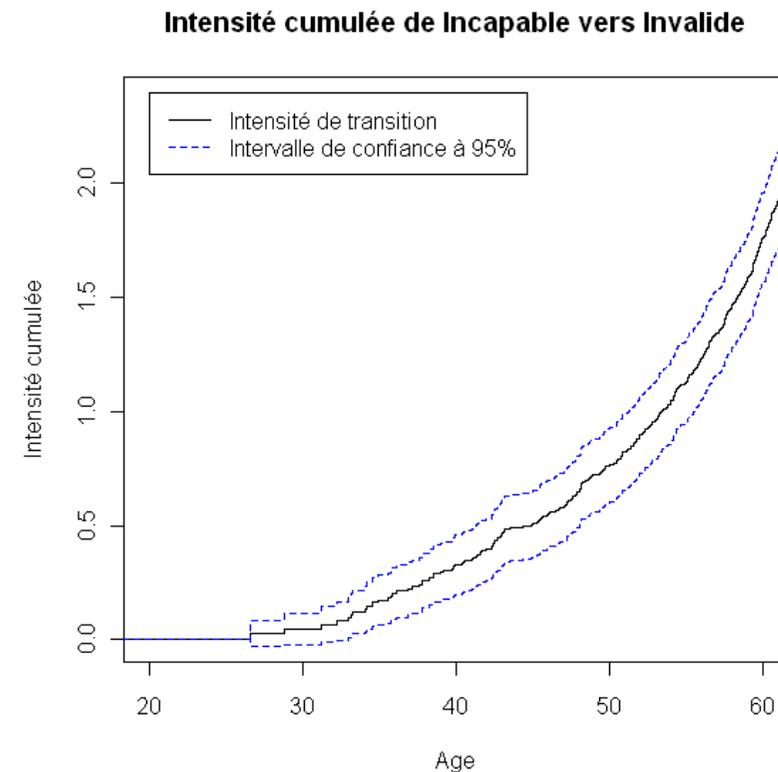
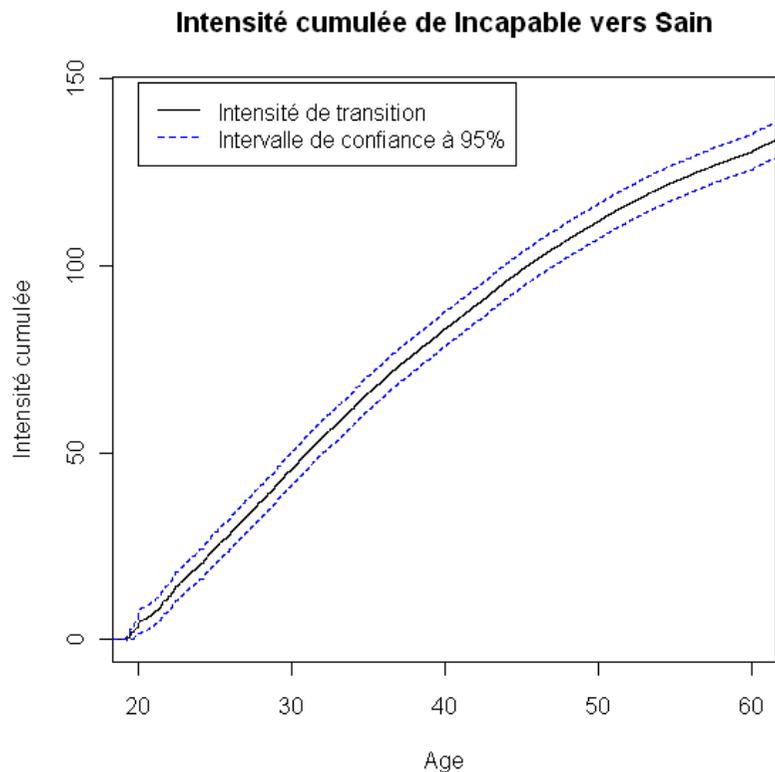


```
> library(mstate)
> c0=coxph(Surv(Tstart,Tstop,status)~strata(trans),data=t_data)
> fit=msfit(object=c0,vartype="aalen",trans=trans,variance=TRUE)
```

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique



```
> library(mstate)
> c0=coxph(Surv(Tstart,Tstop,status)~strata(trans),data=t_data)
> fit=msfit(object=c0,vartype="aalen",trans=trans,variance=TRUE)
```

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

L'estimateur d'**Aalen-Johansen** (cf. Aalen et Johansen [1978]) permet dans un second temps d'estimer les **probabilités de transition**. Cet estimateur s'exprime sous une forme matricielle :

$$P(s, t) = \prod_{\{k: T_k \leq t\}} (\text{Id} + \Delta A_{T_k})$$

$$\text{Id} + \Delta A_{T_k} = \begin{bmatrix} 1 - \frac{\sum_j d_{1j}(T_k)}{n_1(T_k)} & \frac{d_{12}(T_k)}{n_1(T_k)} & \frac{d_{13}(T_k)}{n_1(T_k)} & \frac{d_{14}(T_k)}{n_1(T_k)} \\ \frac{d_{21}(T_k)}{n_2(T_k)} & 1 - \frac{\sum_j d_{2j}(T_k)}{n_2(T_k)} & \frac{d_{23}(T_k)}{n_2(T_k)} & \frac{d_{24}(T_k)}{n_2(T_k)} \\ 0 & 0 & 1 - \frac{d_{34}(T_k)}{n_1(T_k)} & \frac{d_{34}(T_k)}{n_3(T_k)} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

L'estimateur d'**Aalen-Johansen** constitue une généralisation de l'estimateur de Kaplan-Meier. Au surplus, l'estimateur des **lois de survie** reste dans ce cadre égale à l'estimateur de Kaplan-Meier. L'estimateur dispose des propriétés suivantes :

- l'estimateur est **biaisé** positivement mais uniformément **consistant** ;
- il existe un estimateur de la covariance, utile en particulier pour construire des intervalles de confiance asymptotiques ;
- l'estimateur est asymptotiquement normal.

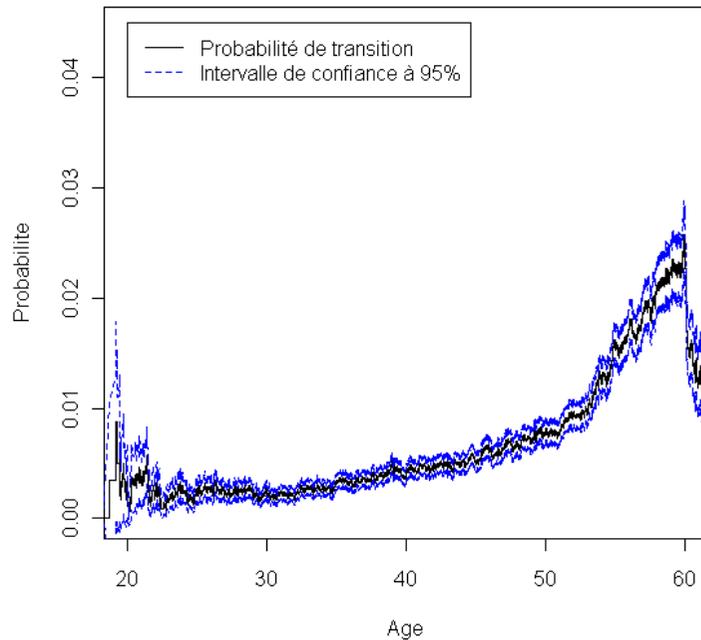
$$\sqrt{n} \left( \hat{P}(s,t) - P(s,t) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

## 2. Estimation des lois de transition

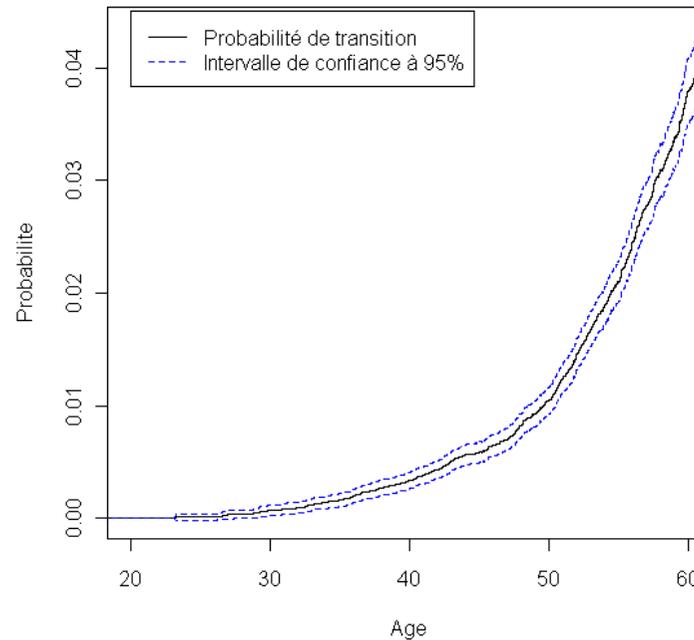


### 2.1. Cadre non paramétrique

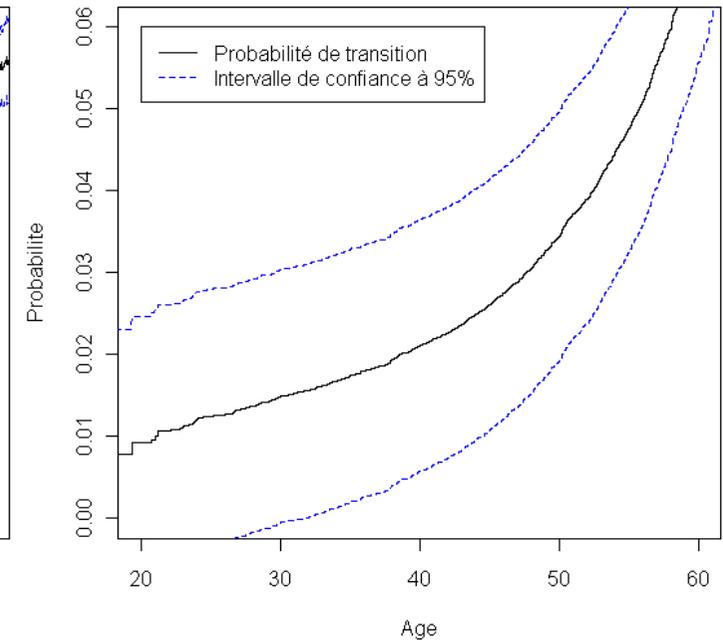
Probabilité de transition de Sain vers Incapable



Probabilité de transition de Sain vers Invalide



Probabilité de transition de Sain vers Décès



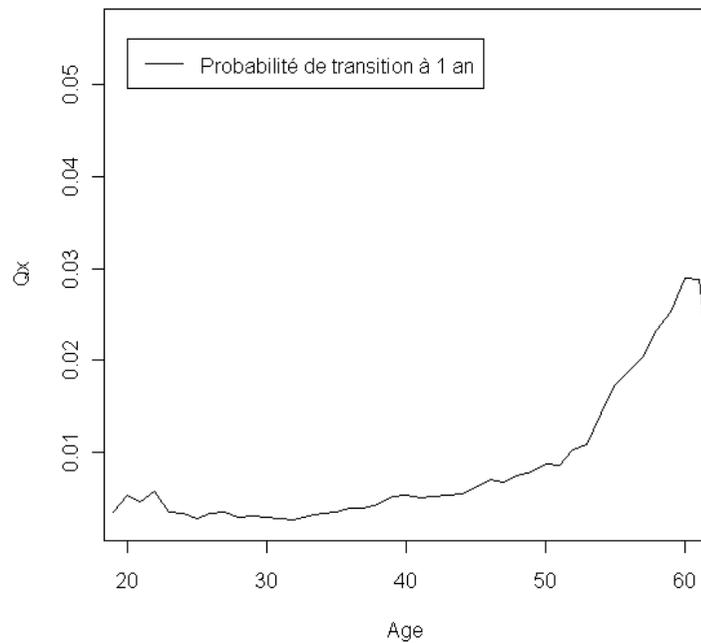
```
> pt=probtrans (fit, predt=AgeMin, method="aalen", variance=TRUE)
```

## 2. Estimation des lois de transition

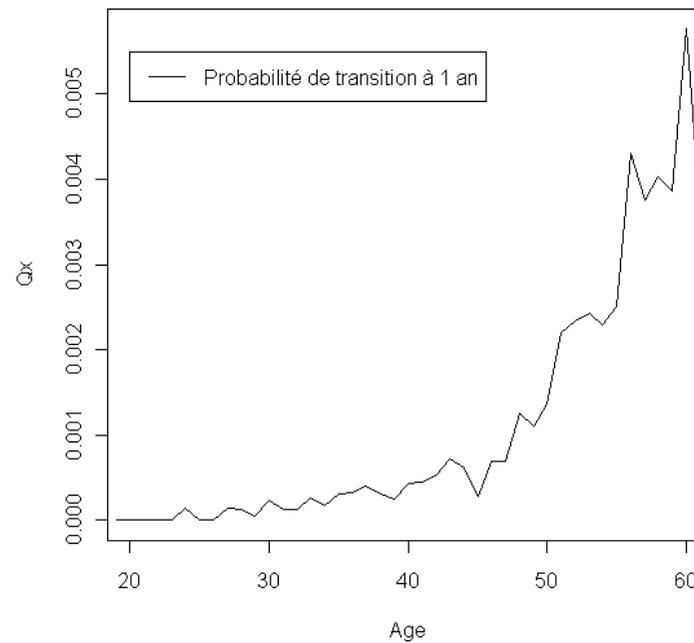


### 2.1. Cadre non paramétrique

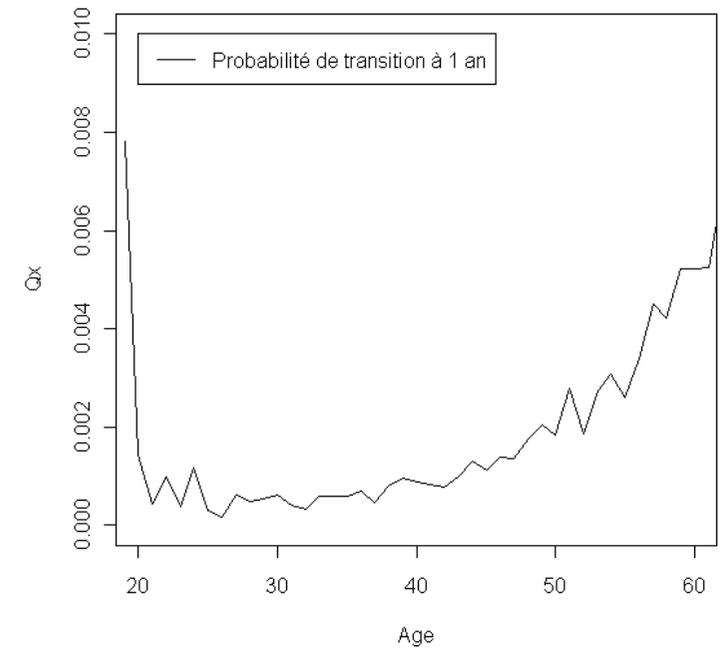
Probabilité de transition de Sain vers Incapable



Probabilité de transition de Sain vers Invalide



Probabilité de transition de Sain vers Décès



Il faut ensuite **lisser** les taux de transition ( $Q_x$ ) à un an.

## 2. Estimation des lois de transition



### 2.1. Cadre non paramétrique

Les propriétés des estimateurs non paramétriques permettent de proposer des **tests statistiques** permettant de comparer les lois de transition de plusieurs échantillons, estimés indépendamment.

On retrouve dans ce cadre les tests classiques de **log-rank** et de **Gehan-Wilcoxon** permettant de tester l'hypothèse suivante :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_N$$

## 2. Estimation des lois de transition



### 2.2. Cadre paramétrique

Dans contexte paramétrique, il est important de proposer un cadre de modélisation **cohérent** pour l'ensemble des transitions.

Ainsi, les modèles retenus sont en général **identiques pour chaque transition** et relativement **basiques** pour permettre d'avoir une cohérence d'ensemble.

Des approches dans lesquelles les intensités de transition suivent par exemple une loi de Gompertz-Makeham ou se positionnent par rapport à des lois de référence, existent mais il conviendra de vérifier qu'elles soient bien adaptées au cadre de l'arrêt de travail.

## 2. Estimation des lois de transition



### 2.3. Cas particulier du modèle de Poisson

Pour un modèle multi-états en présence de censure à droite et de troncature à gauche, la log vraisemblance du modèle à utiliser est :

$$\ln L(\boldsymbol{\theta}, t_1, \dots, t_n) = cste + \sum_{k=1}^n \sum_{i,j} \left[ d_k \ln(\mu_{t_k}^{ij}(\boldsymbol{\theta})) - \int_{v_k}^{t_k} \mu_t^{ij}(\boldsymbol{\theta}) dt \right]$$

où :

- $t_k$  représente les instants où une transition ou une censure survient pour l'individu  $k$  ;
- $v_k$  et  $d_k$  représentent la date de troncature et l'indicatrice de non censure pour l'individu  $k$  ;

## 2. Estimation des lois de transition



### 2.3. Cas particulier du modèle de Poisson

En particulier, si les intensités de transition sont **constantes par morceaux** (modèle de Poisson) :

$$\mu_t^{ij} = \mu_l^{ij} \text{ pour } \tau_{l-1}^{ij} \leq t \leq \tau_l^{ij}$$

L'estimateur du **maximum de vraisemblance** est facilement obtenue (dans un cadre markovien ) et est asymptotiquement gaussien :

$$\mu_l^{ij} = \frac{d_l^{ij}}{er_l^i} \quad \sqrt{n} \left( \mu_l^{ij} - \mu_l^{ij} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left( 0, \frac{\sqrt{d_l^{ij}}}{er_l^i} \right)$$

Au surplus, un résultat équivalent est obtenue dans le cas **semi-markovien** (homogène et inhomogène). Les estimateurs des **probabilités** de transition sont alors obtenus par des méthodes numériques.

## 2. Estimation des lois de transition



### 2.3. Cas particulier du modèle de Poisson

De nombreux outils statistiques permettent de valider les paramètres estimés. On retrouve par exemple les **tests classiques** de **Wald, du score ou du rapport de vraisemblance** pour tester l'hypothèse  $H_0 : \theta = \theta_0$  s'appuyant sur une statistique du Khi 2.

Dans le cadre du modèle de Poisson, le score et l'information observée, utiles à ces tests, sont données par :

$$U_l^{ij} = \frac{d_l^{ij}}{\mu_l^{ij}} - er_l^i \qquad I_l^{ij} = \frac{d_l^{ij}}{(\mu_l^{ij})^2}$$

Par ailleurs, l'hypothèse d'homogénéité par morceaux du modèle peut être comparée à l'hypothèse d'homogénéité générale en testant laquelle des deux s'ajuste le mieux.



## SOMMAIRE

1. Le cadre multi-états
2. Estimation des lois de transition
- 3. *Prise en compte de l'hétérogénéité***
4. Validation

### 3. *Prise en compte de l'hétérogénéité*



Pour prendre en compte une source d'hétérogénéité, les volumes de données sont souvent insuffisants pour permettre des estimations indépendantes (stratification).

Les sources d'hétérogénéité **explicites** sont alors prise en compte à partir de modèles **régressifs** et les sources **implicites** à partir de modèle de **fragilité**. Dans ces conditions, les différents sous-groupes étudiés sont positionnés par rapport à une **population de référence**.

Dans le cadre de l'arrêt de travail, ce sont essentiellement des modèles de type **Cox** (multiplicatif) ou **Aalen** (additif) qui sont utilisés du fait de leur facilité d'utilisation et de leurs qualités naturelles. Le modèle de Poisson peut également être considéré pour décrire les intensités de base d'un modèle à hasard proportionnel.

Dans la suite, seul les modèles multiplicatifs sont présentés.

## 3. *Prise en compte de l'hétérogénéité*



### 3.1. Modèles semi-paramétriques

Le modèle de semi-paramétrique de Cox (*cf.* Cox [1972]) constitue le modèle le plus utilisé en pratique :

$$\mu_{t|Z^{ij}, \beta^{ij}}^{ij} = \mu_{t,0}^{ij} \exp(-\beta^{ij} \cdot T Z^{ij})$$

L'estimation des paramètres s'appuient comme pour les modèles de durée classiques sur la **vraisemblance partielle de Cox** (*cf.* Andersen et *al.* [1993]) :

$$L_{Cox}(\beta, t_1, \dots, t_n) = \prod_{k=1}^n \prod_{i,j} \left[ \frac{\exp(-\beta^{ij} \cdot T Z_k^{ij})}{\sum_{l=1}^n \exp(-\beta^{ij} \cdot T Z_l^{ij}) \cdot 1_{\{t_l \leq t_k\}}} \right]^{d_k}$$

### 3. *Prise en compte de l'hétérogénéité*



#### 3.1. Modèles semi-paramétriques

Pour le modèle de Cox, une **méthode parcimonieuse** consiste à considérer un vecteur de coefficients, relativement long, applicable à toutes les transitions de manière à avoir :

$$\mu_{t|Z^{ij},\beta}^{ij} = \mu_{t,0}^{ij} \exp(-\beta \cdot {}^T Z^{ij})$$

L'estimateur du maximum de vraisemblance  $\hat{\beta}$  est **asymptotiquement gaussien** et est obtenu à l'aide de méthode numérique.

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

## 3. *Prise en compte de l'hétérogénéité*



### 3.1. Modèles semi-paramétriques

Les **intensités de base** sont quant à elles estimées dans un second temps à partir de l'estimateur de **Breslow** qui s'adapte au cas multi-états :

$$\mathbf{A}_{t,0}^{ij} = \frac{d_{ij}(T_k)}{\sum_{\{k:T_k \leq t\}} \sum_{l=1}^n \exp(-\hat{\beta} \cdot {}^T \mathbf{Z}_l^{ij}) \cdot 1_{\{T_l \leq T_k\}}}$$

Les **probabilités** de transition individuelles sont enfin déduites à partir de l'estimateur d'**Aalen-Johansen** appliqué aux intensités cumulés prenant en compte les covariables.

$$\mathbf{P}(s, t | \mathbf{Z}_l) = \prod_{\{k:T_k \leq t\}} \left( \text{Id} + \Delta \mathbf{A}_{T_k} | \mathbf{Z}_l \right)$$

## 3. *Prise en compte de l'hétérogénéité*



### 3.1. Modèles semi-paramétriques

La significativité des **coefficients** est estimée à partir des **tests classiques** de **Wald, du score ou du rapport de vraisemblance**.

L'hypothèse de proportionnalité des risques est quant à elle testée de manière classique à partir des résidus de Schoenfeld .

Par ailleurs, si la durée est prise en compte via un modèle à hasard proportionnel dépendant de l'âge, l'ensemble des résultats vus ci-avant restent valables (hormis bien sur l'hypothèse de proportionnalité avec le temps).

# 3. Prise en compte de l'hétérogénéité



## 3.1. Modèles semi-paramétriques

	coef	exp(coef)	se(coef)	z	Pr(> z )		rho	chisq	p	
SexeHomme.1	-1.707e-01	8.431e-01	1.582e-02	-10.787	< 2e-16	***	SexeHomme.1	0.036770	4.62e+01	1.06e-11
SexeHomme.2	1.554e-01	1.168e+00	9.003e-02	1.726	0.084367	.	SexeHomme.2	-0.004074	6.01e-01	4.38e-01
SexeHomme.3	9.483e-01	2.581e+00	7.020e-02	13.509	< 2e-16	***	SexeHomme.3	-0.003675	4.71e-01	4.93e-01
SexeHomme.4	-2.929e-02	9.711e-01	1.695e-02	-1.728	0.084018	.	SexeHomme.4	-0.002555	2.25e-01	6.35e-01
SexeHomme.5	2.224e-01	1.249e+00	1.053e-01	2.111	0.034734	*	SexeHomme.5	0.003004	3.13e-01	5.76e-01
SexeHomme.6	5.270e-01	1.694e+00	2.051e-01	2.569	0.010194	*	SexeHomme.6	-0.001910	1.25e-01	7.23e-01
SexeHomme.7	8.806e-01	2.412e+00	3.470e-01	2.538	0.011162	*	SexeHomme.7	-0.001123	4.38e-02	8.34e-01
CSP_4NON_CADRE.1	6.974e-01	2.009e+00	6.450e-02	10.813	< 2e-16	***	CSP_4NON_CADRE.1	0.000716	1.78e-02	8.94e-01
CSP_4NON_CADRE.2	5.660e-01	1.761e+00	2.723e-01	2.079	0.037663	*	CSP_4NON_CADRE.2	0.003915	5.37e-01	4.64e-01
CSP_4NON_CADRE.3	1.636e-01	1.178e+00	1.374e-01	1.191	0.233647	.	CSP_4NON_CADRE.3	-0.009734	3.26e+00	7.10e-02
CSP_4NON_CADRE.4	-2.399e-01	7.867e-01	6.840e-02	-3.507	0.000453	***	CSP_4NON_CADRE.4	-0.003652	4.61e-01	4.97e-01
CSP_4NON_CADRE.5	3.894e-01	1.476e+00	4.621e-01	0.843	0.399366	.	CSP_4NON_CADRE.5	-0.004879	8.16e-01	3.66e-01
CSP_4NON_CADRE.6	9.164e+00	9.542e+03	1.049e+02	0.087	0.930372	.	CSP_4NON_CADRE.6	0.020462	1.23e-04	9.91e-01
CSP_4NON_CADRE.7	9.687e+00	1.610e+04	1.253e+02	0.077	0.938380	.	CSP_4NON_CADRE.7	0.027408	1.66e-04	9.90e-01
CSP_4PATRONS.PROF_LIB.1	1.074e+00	2.927e+00	6.263e-02	17.147	< 2e-16	***	CSP_4PATRONS.PROF_LIB.1	0.017919	1.11e+01	8.52e-04
CSP_4PATRONS.PROF_LIB.2	5.161e-01	1.676e+00	2.628e-01	1.964	0.049535	*	CSP_4PATRONS.PROF_LIB.2	0.006321	1.38e+00	2.40e-01
CSP_4PATRONS.PROF_LIB.3	-1.486e-01	8.619e-01	1.308e-01	-1.136	0.255755	.	CSP_4PATRONS.PROF_LIB.3	-0.011948	4.88e+00	2.72e-02
CSP_4PATRONS.PROF_LIB.4	1.120e-01	1.119e+00	6.645e-02	1.686	0.091870	.	CSP_4PATRONS.PROF_LIB.4	0.001955	1.32e-01	7.16e-01
CSP_4PATRONS.PROF_LIB.5	3.976e-01	1.488e+00	4.513e-01	0.881	0.378289	.	CSP_4PATRONS.PROF_LIB.5	-0.005498	1.04e+00	3.07e-01
CSP_4PATRONS.PROF_LIB.6	1.023e+01	2.785e+04	1.049e+02	0.098	0.922259	.	CSP_4PATRONS.PROF_LIB.6	0.034559	3.75e-05	9.95e-01
CSP_4PATRONS.PROF_LIB.7	1.005e+01	2.326e+04	1.253e+02	0.080	0.936046	.	CSP_4PATRONS.PROF_LIB.7	0.039230	7.82e-05	9.93e-01
---							GLOBAL	NA	2.03e+02	0.00e+00

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

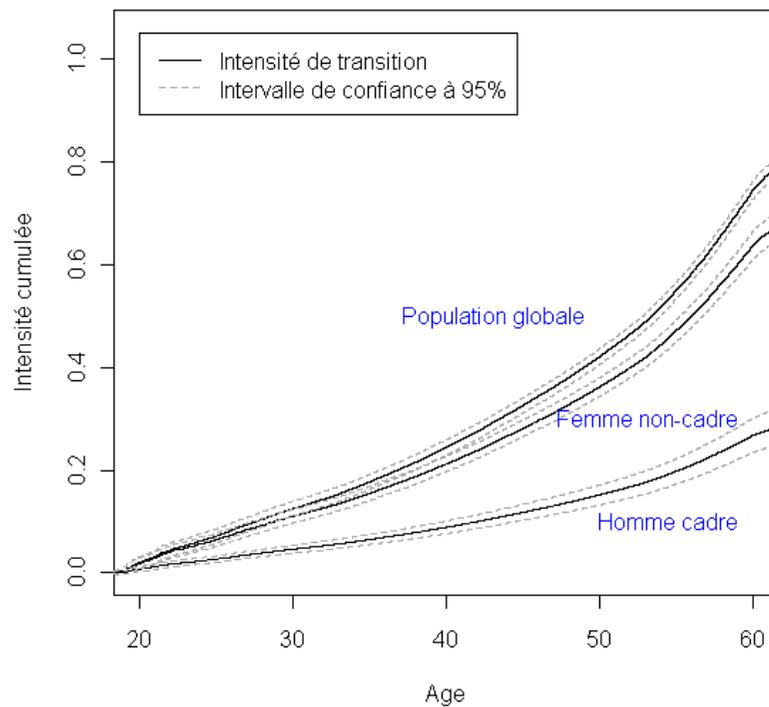
```
> c1=coxph(Surv(Tstart,Tstop,status)~strata(trans)+Var1+Var2+...,data=t_data)
> cox.zph(c1)
```

# 3. Prise en compte de l'hétérogénéité

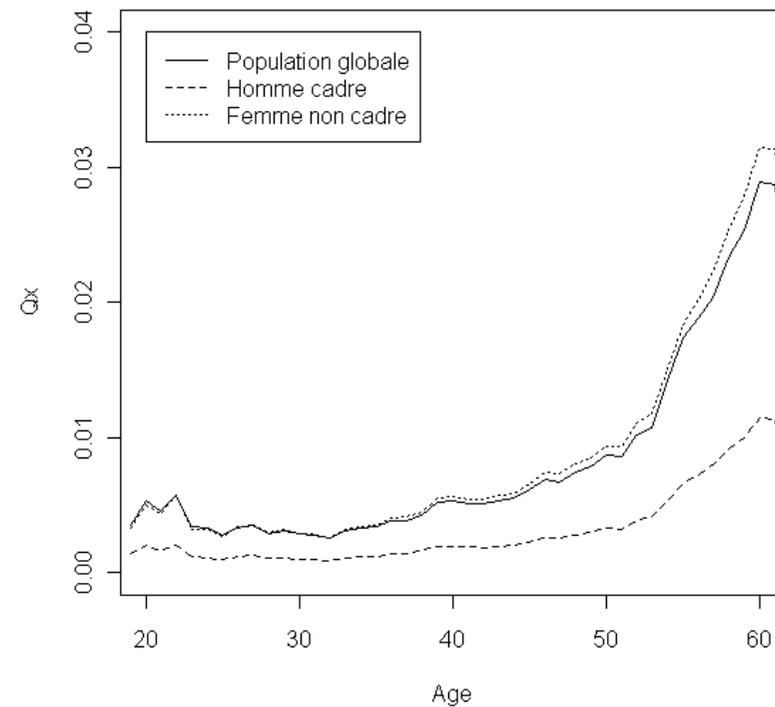


## 3.1. Modèles semi-paramétriques

Intensité cumulée de Sain vers Incapable



Probabilité de transition de Sain vers Incapable



### 3. *Prise en compte de l'hétérogénéité*



#### 3.2. Modèles paramétriques

Dans le cadre de modèles paramétriques, l'estimation des paramètres de l'intensité de base et de coefficients de régression doit être menée.

$$\mu_{t|Z^j, \beta^j}^{ij}(\theta) = \mu_{t,0}^{ij}(\theta) \exp(-\beta^j \cdot T Z^j)$$

Comme dans un modèle paramétrique classique, l'estimation est menée par maximum de vraisemblance. Les propriétés vérifiées de  $\hat{\theta}$  et de  $\hat{\beta}$  sont les mêmes que celles détaillées précédemment.



## SOMMAIRE

1. Le cadre multi-états
2. Estimation des lois de transition
3. Prise en compte de l'hétérogénéité
- 4. *Validation***

## 4. *Validation*



Une fois que les lois de transition sont estimées, il convient de valider la pertinence de la construction effectuée. Cette validation est effectuée à 2 niveaux :

- en réalisant des tests statistiques adaptés aux techniques de construction ;
- *backtesting* des résultats en contrôlant la pertinence de la prédiction des lois d'expérience avec les observations.

Cette étape de *backtesting* est importante car elle est réalisée sur des éléments différents des tests statistiques et en particulier pour lesquels l'expert peut avoir une appréciation à donner.

## 4. *Validation*



Cette étape de *backtesting* n'est pas spécifique au cadre multi-états. Il s'agira de comparer par exemple :

- la différence entre les transitions observées et les transitions prédites ;
- la différence entre les règlements observés et les règlements prédits par les lois d'expérience construites ;
- la différence entre les provisions observées et les provisions prédites par les lois d'expérience construites.

Ces analyses sont importantes car elles permettent de valider les lois construites et donner une mesure du risque d'estimation induit par les fluctuations d'échantillonnage.

## Conclusion



Le cadre introduit par les modèles multi-états permet d'estimer des lois de transition donnant une meilleure représentation de la structure complexe d'une garantie arrêt de travail. Ces modèles s'avèrent plus cohérents et permettent ainsi de contourner un certain nombre d'approximations utilisées dans la pratique courante pour pallier à l'absence de lois dédiées (loi de mortalité des incapables par exemple).

Ces modèles présentent également l'avantage de s'intégrer aisément à la liste des outils statistiques utilisés en assurance de personnes. Par exemple, les méthodes utilisées pour prendre en compte l'hétérogénéité d'une population restent valables. Elles permettent également de traiter d'autres types de garanties complexes comme l'estimation des lois d'expérience relatives à la dépendance ou des sorties (rachat, décès) en épargne.

L'utilisation de ces méthodes contribue à donner une meilleure représentation du tarif des garanties et du *best estimate*.

## Conclusion



Au delà, une réflexion est à mener pour mesurer les risques d'estimation relatifs à l'utilisation de ces méthodes. L'utilisation de méthodes de type *bootstrap* dans le cas non paramétrique et de techniques bayésiennes dans le cadre paramétrique devraient permettre de donner une quantification de cette erreur.

Par ailleurs, les risques associés au choix de la forme du modèle sont également ouvertes :

- comment estimer au mieux l'impact de la durée de transition ?
- doit-on tenir compte d'une tendance particulière pour estimer la mortalité des invalides ?
- quel traitement des risques de modèle et d'avis d'expert ?

En effet, il s'agit là de risques systématiques qui engendrent *a priori* un besoin en fonds propres et doivent par conséquent être quantifiés.



- AALLEN O. O., JOHANSEN S. [1978] « An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations », *Scandinavian Journal of Statistics*, Vol. 5, No. 3, p. 141-150.
- ANDERSEN P. K., BORGAN Ø., GILL R. D., KEIDING N. [1993] *Statistical Models Based on Counting Processes*, Springer-Verlag New York Inc., 767 p., (Springer Series in Statistics), ISBN : 0-378-97872-0.
- ANDERSEN P. K., KEIDING N. [2002] « Multi-state models for event history analysis », *Statistical Methods in Medical Research*, Vol. 11, No. 2, p. 91-115.
- BEYERSMANN J., SCHUMACHER M., ALLIGNOL A. [2011] *Competing Risks and Multistate Models with R*, Springer, 245 p., (Use R!), ISBN : 9781461420347.
- COX D. R. [1972] « Regression Models and Life-Tables », *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 2, p. 187-220.
- D'AMICO G., JANSSEN J., MANCA R. [2010] « Initial and Final Backward and Forward Discrete Time Non-homogeneous Semi-Markov Credit Risk Models », *Methodology and Computing in Applied Probability*, Vol. 12, No. 2, p. 215-225.
- DENUIT M., ROBERT C. [2007] *Actuariat des assurances de Personnes : Modélisation, tarification et provisionnement*, Paris : Economica, 405 p., (Assurance Audit Actuariat), ISBN : 978-2-7178-5329-2.
- HABERMAN S., PITACCO E. [1998] *Actuarial Models for Disability Insurance*, Chapman and Hall/CRC, 1er éd., 280 p., ISBN : 0-8493-0389-3.
- JANSSEN J., MANCA R.. [2006] *Applied Semi-Markov Processes*, Springer, 310 p., ISBN : 0-387-29548-8.
- JANSSEN J., MANCA R. [2007] *Semi-Markov Risk Models for Finance, Insurance and Reliability*, Springer-Verlag New York Inc., 448 p., ISBN : 978-0-387-70729-7.



KAPLAN E. L., MEIER P. [1958] « Nonparametric Estimation from Incomplete Observations », *Journal of the American Statistical Association*, Vol. 53, No. 282, p. 457-481.

LANDO D., SKØDEBERG T. M. [2002] « Analyzing rating transitions and rating drift with continuous observations », *Journal of Banking & Finance*, Vol. 26, No. 2–3, p. 423-444.

MONTEIRO A., SMIRNOV G. V., LUCAS A. [2006] *Non-parametric Estimation for Nonhomogeneous Semi-Markov Processes: An Application to Credit Risk*, Amsterdam : Tinbergen Institute

R DEVELOPMENT CORE TEAM [2011] *R: A Language and Environment for Statistical Computing*, Vienna, Austria, (R Foundation for Statistical Computing), ISBN : 3-900051-07-0. URL : < <http://www.R-project.org> >.

RULLIÈRE D., SERANT D. [1998] « Estimation de probabilités de changement d'état en présence de données incomplètes et applications actuarielles », *Bulletin Français d'Actuariat*, Vol. 2, No. 3, p. 71-88.

SAINT PIERRE P. [2005] *Modèles multi-états de type Markovien et application à l'asthme*, Thèse, Montpellier : Université Montpellier I, 202 p.

DE WREEDE L. C., FIOCCO M., PUTTER H. [2011] « mstate: An R Package for the Analysis of Competing Risks and Multi-State Models », *Journal of Statistical Software*, Vol. 38, No. 7.



### Frédéric PLANCHET

frederic.planchet@univ-lyon1.fr

### Quentin GUIBERT

qguibert@winter-associes.fr

#### WINTER & Associés

43-46 avenue de la Grande Armée

F - 75116 Paris

+33-1-45-72-63-00

#### ISFA

50 avenue Tony Garnier

F - 69007 Lyon

+33-4-37-38-74-37

<http://www.ressources-actuarielles.net/>

<http://blog.ressources-actuarielles.net/>