

ESTIMATION DE PROBABILITES DE CHANGEMENT D'ETAT EN PRESENCE DE DONNEES INCOMPLETES ET APPLICATIONS ACTUARIELLES

Didier RULLIERE, Actuaire Ecureuil Vie
Daniel SERANT, ISFA

*Laboratoire de Sciences Actuarielle et Financière
Université Claude Bernard, Lyon I
Travail soutenu par la société Ecureuil Vie*

Résumé

Les expressions des probabilités de changement d'état, sont, que ce soit de manière explicite ou de manière implicite, largement utilisées dans l'élaboration de tarifs, de calculs de provisions et dans la mise au point de procédures de suivi de portefeuille. Elles sont notamment exploitées dans les secteurs de l'assurance vie et de la prévoyance (états de *vie* ou *décès* pour l'assurance vie, états d'*actif*, *retraité*, *incapable*, *invalidé*, *décédé*, etc. en prévoyance). L'estimation de ces probabilités est un problème crucial dont va dépendre la fiabilité des calculs actuariels déduits.

Cet article propose une estimation des probabilités de transition d'un processus de Markov non homogène, en présence de données incomplètes. Cette estimation se fait à l'aide d'une extension naturelle de l'estimateur de KAPLAN-MEIER [[8]], proposée originellement par AALEN [[1]].

En présence d'observations évoluant entre plusieurs états, on estime, en fonction du temps, les probabilités de passage d'un état à un autre. Les observations sont ici incomplètes, soit qu'un individu apparaisse dans un état donné à une date donnée sans qu'il ait été observé jusqu'alors (troncatures gauches), soit qu'il ne soit plus observé à partir d'un instant donné (censures droites).

Une propriété intéressante du point de vue actuariel est donnée pour cet estimateur, qui fait également l'objet d'une application: un fichier est créé qui simule des changements d'état d'individus, partiellement observés, évoluant au cours du temps. L'estimation des probabilités de se trouver dans un état ε_j à une date t , sachant que l'on était dans un état ε_i à une date s est alors opérée. Des graphiques permettent enfin de constater la conformité des estimations aux lois qui ont généré le fichier.

Quelques applications actuarielles sont également développées, illustrant les très nombreuses utilisations possibles de cet estimateur.

Quelques mots-clés: estimateur de Kaplan-Meier, données tronquées et censurées, processus de Markov non homogène, probabilités de transition, changements d'états.

BULLETIN FRANÇAIS D'ACTUARIAT, Vol. 2, N° 3, 1998, pp. 71-88

1. INTRODUCTION

L'élaboration d'un tarif, le calcul d'une provision s'appuient sur la prévision de flux futurs aléatoires, pour lesquels sont généralement étudiés les probabilités de survenance d'une part (renseignant la fréquence des sinistres) et les montants engagés d'autre part (renseignant les montants des sinistres).

Cet article traite spécifiquement de l'estimation des probabilités de survenance, étudiées comme probabilités de transition dans un cadre Markovien.

Dans le cadre Markovien, le futur des individus n'est envisagé qu'à la lumière de leur état présent, en ignorant leur passé. Cette hypothèse, certes restrictive, est communément employée par les actuaires (par exemple par l'utilisation de tables de mortalité, où seul l'état de l'individu à un âge donné est considéré). De plus, une partition plus complexe des états présents peut permettre une prise en compte partielle du passé de chaque individu (on peut ainsi considérer l'état « malade pour la deuxième fois » etc.). Enfin, de nombreuses modélisations Markoviennes sont apparues dans la littérature actuarielle ces dernières années, pour retranscrire notamment les phénomènes transitoires comme l'incapacité et l'invalidité (cf. HABERMAN, BLOOMFIELD [[5]], PITACCO [[10]], JANSSEN, MANCA [[7]]).

Si les calculs tels que l'espérance de durée de présence dans un état pendant une période donnée, l'espérance de durée de présence continue dans un état, etc. déduits de la connaissance des probabilités de transition sont largement développés dans la littérature actuarielle citée ci-dessus, l'estimation des probabilités de transition n'est pas toujours détaillée.

Nous présenterons une méthode d'estimation non paramétrique relativement simple. Les probabilités de transition vont s'obtenir en fonction d'une variable abscisse (généralement temporelle), elles permettront une visualisation du bruit pesant sur ces estimations, et éventuellement des phénomènes brusques (par exemple, soudaines résiliations à un âge de changement de tarif) qui n'apparaîtraient pas forcément dans une modélisation paramétrique.

A titre d'illustration, des résultats classiques ont été calculés à l'aide de l'estimateur proposé dans cet article. En l'absence de données exploitables, ces résultats ont été extraits à partir de données fictives, ce qui présente néanmoins l'avantage de pouvoir comparer les résultats obtenus avec ceux que l'on était en droit d'attendre suivant le processus créé.

Les résultats présentés sont les suivants: une table de probabilités de présence dans un état, une table de probabilités de changement d'état dans l'année, des durées moyennes de présence future dans un état, la valorisation d'une rente de présence dans un état (cf. dernier paragraphe).

2. NOTATIONS

2.1. PROCESSUS ETUDIE

Soit $E = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k\}$ l'ensemble des k états d'un processus de Markov noté $(E_t)_{t \in R^+}$ engendré par un générateur infinitésimal $\{\mu_{ij}(t) | i \neq j\}$ tel que, pour tout triplet (i, j, t) , les probabilités de transitions locales soient définies par les relations:

$$\begin{cases} p_{ij}^{\Delta t}(t) = Pr[E_{t+\Delta t} = \varepsilon_j | E_t = \varepsilon_i] = \mu_{ij}(t) \cdot \Delta t + o(\Delta t) \text{ pour } i \neq j \\ p_{ii}^{\Delta t}(t) = Pr[E_{t+\Delta t} = \varepsilon_i | E_t = \varepsilon_i] = 1 - \sum_{j \neq i} \mu_{ij}(t) \cdot \Delta t + o(\Delta t) \end{cases}$$

Ainsi, pour une population i.i.d de n processus, la probabilité que deux (ou plus) changements interviennent en même temps est nulle.

L'ensemble des indices des éléments de E sera noté $I = \{1, 2, \dots, k\}$.

$t \in R^+$ est la variable en fonction de laquelle on étudie les changements d'état: t représentera donc le plus souvent une variable temporelle (date calendaire, âge, durée...), bien qu'elle puisse représenter toute autre variable (montant...). Dans la suite de cet article, cette variable sera appelée *date*.

On cherche à estimer soit le générateur du processus, soit les matrices $M(s, t)$ définies par:

$$M(s, t) = (p_{ij}(s, t))_{(i, j) \in I^2} \text{ où } p_{ij}(s, t) = Pr[E_t = \varepsilon_j | E_s = \varepsilon_i], (i, j) \in I^2, (s, t) \in R^{+2}$$

Note: on rappelle que l'on passe des matrices $M(s, t)$ au générateur par les équations de CHAPMAN-KOLMOGOROV:

$$\frac{d}{dt} M(0, t) = M(0, t) \cdot G(t) \text{ où } G(t) \text{ est égal à la matrice: } G(t) = (g_{ij}(t))_{(i, j) \in I^2}$$

$$\begin{cases} g_{ij}(t) = \mu_{ij}(t) & i \neq j \\ g_{ii}(t) = 1 - \sum_{i \neq j} \mu_{ij}(t) \end{cases}$$

Dans la pratique on peut, à partir de l'estimation des matrices $M(0, t)$:

- soit régulariser, par des techniques d'estimation fonctionnelle, les fonctions $\hat{p}_{ij}(0, t)$ en fonctions dérivables de t .

- soit discrétiser le temps avec une échelle suffisamment fine pour que, à toute date, on ne puisse enregistrer qu'au plus un changement d'état. On peut alors se limiter à l'estimation des matrices $M(s, t)$, le processus étant alors traité comme une chaîne de Markov.

2.2. FICHER OBSERVE

Soit une population de n individus, $n \in \mathbb{N}$. Chaque individu, noté $l \in \{1..n\}$, est observé durant une période $[e_l, t_l]$, et est supposé suivre le processus exposé en 2.1.

On observe alors les changements d'état de l'ensemble des individus à des dates notées τ_r , $r \in \{1, \dots, r_{max}\}$.

Les indices des états de l'individu l avant et après une date τ de changement, $\tau \in [e_l, t_l]$, sont notés $E_l(\tau^-)$ et $E_l(\tau^+)$. L'indice de l'état de l'individu l à une date t quelconque est noté $E_l(t)$, et l'on pose $E_l(\tau) = E_l(\tau^-)$. Lorsque $t \notin [e_l, t_l]$, on note $E_l(t) = 0$.

Enfin, $n_i(t)$ représente le nombre de personnes observées dans l'état ε_i à la date t , $i \in I, t \in \mathbb{R}^+$. Le vecteur des $n_i(t)$ pour tout état sera noté $\bar{n}(t)$.

On a $n_i(t) = \sum_{l \in \{1, \dots, n\}} 1_{E_l(t)=i}$.

3. ESTIMATION EN PRESENCE DE PLUSIEURS ETATS

3.1. ESTIMATION EN PRESENCE DE DEUX ETATS, DONT UN ABSORBANT

En présence de deux états, dont un absorbant, on retrouve le cadre originel, unidimensionnel, de l'estimateur de KAPLAN-MEIER [[8]].

Ainsi, lorsque l'un des deux états, indicé n° 2, est absorbant, l'estimateur de Kaplan-Meier de $p_{11}(0, t)$ est donné par:

$$\hat{p}_{11}(0, t) = \prod_{\tau_r < t} \left(1 - \frac{1}{n_1(\tau_r)} \right) \text{ et dès lors, } \hat{p}_{12}(0, t) = 1 - \hat{p}_{11}(0, t)$$

Cet estimateur peut s'analyser sous plusieurs angles: on peut le concevoir respectivement comme l'estimateur qui maximise la vraisemblance du fichier, comme l'estimateur qui vérifie la cohérence, c'est à dire qui égalise les nombres « espérés » de changements d'état sur un intervalle de temps donné et la somme sur cet intervalle du nombre de changements effectivement observés et du nombre de changements « espérés »

et non observés à cause des censures, ou encore plus naturellement comme l'estimateur considérant les changements d'état par rapport aux expositions au risque de changement (cf. DROESBEKE, FICHET, TASSI [[3]] et RULLIERE, SERANT [[11]]).

Cette dernière analyse, très intuitive, est reprise ici: on note τ^- et τ^+ deux dates entre lesquelles survient au plus un changement d'état, on considère que tous les individus présents dans l'état n° 1 à la date τ^- sont égaux face au risque de passer dans l'état n° 2. L'estimation qui en découle est l'estimation binomiale classique:

$$p_{11}(\tau^-, \tau^+) = \frac{n_1(\tau^+)}{n_1(\tau^-)} \text{ et } p_{12}(\tau^-, \tau^+) = \frac{n_1(\tau^-) - n_1(\tau^+)}{n_1(\tau^-)}$$

Et l'estimateur utilisé est le suivant: $\hat{M}(0, \tau^+) = \hat{M}(0, \tau^-) \cdot \hat{M}(\tau^-, \tau^+)$.

On retrouve bien l'estimation actuarielle classique:

Soit $E = \{Vie, Mort\} = \{\varepsilon_1, \varepsilon_2\}$, x et $x+k$ deux dates clés, entre lesquelles on observe un décès.

Avec les notations actuarielles classiques ${}_k p_x = \hat{p}_{11}(x, x+k)$ et ${}_k q_x = \hat{p}_{12}(x, x+k)$ on a:

$$\hat{M}(x, x+k) = \begin{pmatrix} 1 - 1/n_1(x) & 1/n_1(x) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \hat{p}_{11}(x, x+k) & \hat{p}_{12}(x, x+k) \\ \hat{p}_{21}(x, x+k) & \hat{p}_{22}(x, x+k) \end{pmatrix} = \begin{pmatrix} {}_k p_x & {}_k q_x \\ 0 & 1 \end{pmatrix}$$

$$\hat{M}(0, x+k) = \begin{pmatrix} {}_x p_0 & {}_x q_0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 - 1/n_1(x) & 1/n_1(x) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} {}_x p_0 \cdot (1 - 1/n_1(x)) & {}_x p_0 \cdot (1/n_1(x)) + {}_x q_0 \\ 0 & 1 \end{pmatrix}$$

on vérifie bien: ${}_{x+k} q_0 = 1 - {}_x p_0 \cdot (1 - 1/n_1(x)) = {}_x q_0 + {}_x p_0 \cdot (1/n_1(x))$

dès lors: ${}_x p_0 = \prod_{\tau_r < x} \left(1 - \frac{1}{n_1(\tau_r)} \right)$ on retrouve l'estimateur usuel de Kaplan-Meier.

3.2. ESTIMATION EN PRESENCE DE PLUSIEURS ETATS

L'estimation dans ce cadre peut, dans un premier temps, s'appuyer sur un raisonnement simplement intuitif. Un cadre plus théorique sera proposé par la suite.

Pour des phénomènes se déroulant au cours du temps, considérant que jamais deux événements ne surviennent exactement au même instant, on peut choisir un découpage du temps suffisamment fin pour que n'intervienne qu'un seul événement entre deux instants donnés. Cet événement étant alors modélisé comme le passage d'un état initial à un état final, on va pouvoir ignorer sur cette période les autres états.

On va donc se ramener localement à une estimation en présence de deux états. Sur une longue période, les probabilités de transition peuvent alors s'obtenir par la multiplication des matrices locales de passage.

A un instant donné, on considère que tous les individus présents dans un état donné sont égaux devant le hasard. Ainsi, la probabilité pour un individu de passer d'un état dans un autre à un instant t ne dépend-elle que de l'état dans lequel l'individu se trouve en t , indépendamment de son histoire jusqu'à cette date: le passé de l'individu se résume dans le seul instant présent, et la modélisation adoptée revêt bien un caractère Markovien.

Il faut noter que l'hypothèse forte d'évolution d'un individu indépendamment de son histoire est contournable en augmentant le nombre d'états considérés: on peut par exemple, lors de l'étude d'une maladie, étudier les probabilités de transition de la rechute n° k à la rechute n° $k+1$ si l'on estime que le fait de tomber malade n'est pas indépendant du nombre de rechutes passées... L'augmentation du nombre d'états pèse néanmoins sur la précision des résultats.

Dans le cadre Markovien non homogène décrit plus haut, on cherche à estimer les matrices $M(s,t)$ pour $(s,t) \in \mathbb{R}^+{}^2$. La recherche de ces matrices peut se restreindre à la recherche des matrices de survie $M(0,t)$ pour $t \in \mathbb{R}^+$. Le produit des probabilités conditionnelles s'écrit en effet de manière tout à fait classique $M(0,x+k) = M(0,x) \cdot M(x,x+k)$. Dès lors, lorsque la matrice $M(0,x)$ est inversible: $M(x,x+k) = M^{-1}(0,x) \cdot M(0,x+k)$.

L'estimation des matrices de passage va s'effectuer de manière parfaitement similaire à ce qui a été exposé en partie 2.1. Soient deux dates τ^- et τ^+ entre lesquelles ne survient qu'un seul événement, l'unicité de ce dernier va permettre d'estimer la matrice de passage de τ^- à τ^+ : $M(\tau^-, \tau^+)$.

Supposons qu'entre ces dates, l'événement observé soit un passage de l'état ε_{i^*} à l'état ε_{j^*} . Alors, de la même manière que dans la partie 2.1, les probabilités de passage de l'état ε_{i^*} à l'état ε_{i^*} et à l'état ε_{j^*} vont être modifiées: parmi les $n_{i^*}(\tau)$ individus présents à l'état ε_{i^*} en τ , $n_{i^*}(\tau) - 1$ se retrouvent dans l'état ε_{i^*} et 1 se retrouve dans l'état ε_{j^*} . On pose donc:

$$p_{i^*i^*}(\tau^-, \tau^+) = 1 - \frac{1}{n_{i^*}(\tau^-)} \text{ et } p_{i^*j^*}(\tau^-, \tau^+) = \frac{1}{n_{i^*}(\tau^-)}$$

Pour toutes les autres probabilités, les individus étant restés dans leur état,
 $p_{ij}(\tau^-, \tau^+) = 1_{\{i=j\}}$.

Matriciellement, $\hat{M}(\tau^-, \tau^+) = Id - \hat{h}(\tau^-, \tau^+)$, où

Id représente la matrice identité de dimension k , le nombre d'états.

$$\hat{h}(\tau^-, \tau^+) = (h_{ij}(\tau))_{(i,j) \in I^2} \text{ avec } h_{ij}(\tau) = \frac{1}{n_{i^*}(\tau)} \cdot \left(1_{\{i=i^*, j=i^*\}} - 1_{\{i=i^*, j=j^*\}} \right)$$

Naturellement entre deux dates où il ne se passe aucun événement, la matrice \hat{M} reste égale à l'identité.

On détermine donc les $\hat{M}(0, t)$ pour chaque t de la manière suivante:

1. $\hat{M}(0, 0) = Id$
2. entre deux dates $\left] \tau^-, \tau^+ \right]$ encadrant un passage de ε_{i^*} à ε_{j^*} ,
 $\hat{M}(\tau^-, \tau^+) = Id - \hat{h}(\tau^-, \tau^+)$
3. par récurrence on a $\hat{M}(0, \tau^+) = \hat{M}(0, \tau^-) \hat{M}(\tau^-, \tau^+)$.

Ainsi, avec la convention d'effectuer le produit matriciel pour des τ_r croissants, on retrouve une expression généralisée de l'estimateur de Kaplan-Meier:

$$\hat{M}(s, t) = \prod_{\{\tau_r, s < \tau_r \leq t\}} \left(Id - \hat{h}(\tau_r^-, \tau_r^+) \right)$$

En pratique, cette expression peut être calculée directement pour des $s = k \cdot pas$ et $t = s + pas$, lorsque l'on désire ne pas avoir de matrice à inverser.

Le résultat intuitif présenté ci-dessus correspond à l'estimateur proposé par AALEN [[1]], et présenté dans JACOBSEN [[6]] par exemple, où de nombreuses propriétés (de non-biais, de convergence...) sont démontrées. En pratique, les actuaires peuvent être également intéressés par une autre propriété, dite de cohérence, énoncée ci-dessous:

On définit ici un estimateur cohérent comme un estimateur qui, pour tout état ε_j donné et pour toute date t donnée, égalise les nombres espérés d'individus présents dans cet état à cette date, pour des nombres calculés a priori (i.e. en connaissance des seules dates d'entrées des individus) et a posteriori (i.e. en connaissance de la totalité des observations). Cette égalité correspond donc à l'égalité suivante:

$$\forall j, \forall t \sum_{e_l \leq t} \hat{p}_{E(e_l)j}(e_l, t) = n_j(t) + \sum_{t_l \leq t} \hat{p}_{E(t_l)j}(t_l, t)$$

Dès lors, pour tout état ε_j , pour toute partie du temps A , on peut égaliser les espérances de durée future dans cet état, calculées a priori et a posteriori:

$$\forall j, \sum_{e_l \leq t} \int_A \hat{p}_{E(e_l)j}(e_l, t) dt = \sum_l \int_A 1_{E_l(t)=j} dt + \sum_{t_l \leq t} \int_A \hat{p}_{E(t_l)j}(t_l, t) dt$$

Cette dernière propriété est particulièrement intéressante au sens actuariel; elle égalise les durées moyennes « observées » sur le fichier et les durées moyennes « estimées » à l'aide de l'estimateur.

Pour une cohérence écrite entre deux dates t^- et t^+ , encadrant un passage de ε_i à ε_j , on peut formuler une autre écriture à la cohérence. On écrit par exemple la cohérence en t^+ :

$$\forall j, \sum_{e_l \leq t^-} \hat{p}_{E(e_l)j}(e_l, t^+) - \sum_{t_l \leq t^-} \hat{p}_{E(t_l)j}(t_l, t^+) = n_j(t^+)$$

puis, à l'aide des équations de CHAPMAN-KOLMOGOROV,

$$\forall j, \sum_{e_l \leq t^-} \sum_{k \in I} \hat{p}_{E(e_l)k}(e_l, t^-) \cdot \hat{p}_{kj}(t^-, t^+) - \sum_{t_l \leq t^-} \sum_{k \in I} \hat{p}_{E(t_l)k}(t_l, t^-) \cdot \hat{p}_{kj}(t^-, t^+) = n_j(t^+)$$

enfin, par interversion des sommes, en utilisant la propriété de cohérence en t^-

$$\forall j, \sum_{k \in I} n_k(t^-) \cdot \hat{p}_{kj}(t^-, t^+) = n_j(t^+)$$

Cette dernière propriété s'écrit de manière matricielle:

$${}^t \bar{n}(t^-) \cdot \hat{M}(t^-, t^+) = {}^t \bar{n}(t^+)$$

Et l'on vérifie aisément que l'estimateur proposé satisfait cette nouvelle équation de cohérence.

Dans de nombreux cas particuliers, on retrouve des résultats connus:

Dans le cas unidimensionnel classique, la cohérence revêt la même écriture et permet directement d'étendre l'estimateur de Kaplan-Meier (l'estimateur cohérent est alors maximum de vraisemblance).

D'autre part, on montre par récurrence que pour toutes dates s, t avec $s < t$, et lorsque les données sont cette fois-ci complètes, ${}^t \bar{n}(s) \cdot \hat{M}(s, t) = {}^t \bar{n}(t)$, qui s'interprète comme la conservation par l'estimateur des effectifs observés lorsque l'unique cause de variation des effectifs est le passage d'un état à l'autre.

Comme nous l'avons signalé plus haut, les matrices de probabilités $M(s, t)$ définissent entièrement le processus. Ainsi, il sera possible de déduire des estimations \hat{M}

de M des estimations des expressions actuarielles déduites de la connaissance des matrices M . Lorsque la formulation théorique de la mesure recherchée apparaîtra trop ardue, un recours à une simulation aléatoire pourra être envisagé.

4. APPLICATION

4.1. SIMULATION D'UN FICHIER

Calcul des probabilités de passage

Les taux instantanés de passage de l'état ε_i à l'état ε_j à l'instant t sont notés $\mu_{ij}(t)$.

On suppose que les passages ne sont enregistrés que lors des inventaires, espacés de Δt .

Dès lors, pour simuler ces passages, on prend l'hypothèse, lorsque Δt est petit, que les probabilités de passage consigné à l'inventaire t sont:

$$p_{ij}^{\Delta}(t) = \mu_{ij}(t) \cdot \Delta t + o(\Delta t) \quad i \neq j \quad \text{et} \quad p_{ii}^{\Delta}(t) = 1 - \sum_{j \neq i} \mu_{ij}(t) \cdot \Delta t + o(\Delta t)$$

Pour les besoins de la simulation, on pourra faire des hypothèses directement sur les $p_{ij}^{\Delta}(t) = f(i, j, t)$. Les probabilités de passage dépendent du temps et forment une chaîne de Markov non homogène.

Simulation d'un individu du portefeuille

Pour un individu $n^{\circ} l$, sont tout d'abord créées les dates d'entrée e_l et de sortie t_l du portefeuille.

La date d'entrée est choisie suivant la loi hypothèse E, de la date d'entrée dans le portefeuille.

La date de sortie s'obtient en ajoutant à cette date de la durée d'observation, qui suit la loi hypothèse D. Ces deux tirages peuvent s'obtenir par la fonction de répartition inverse d'une loi uniforme ou encore par la méthode des rejets (cf. COCOZZA-THIVENT [[2]] chapitre 1.3).

On mène ensuite l'étude des éventuels changements d'état, pour chaque date t variant de e_l à t_l . Pour chaque date t d'inventaire, connaissant les probabilités de changement d'état de l'individu, on va effectuer un tirage aléatoire pour savoir si l'individu change d'état à cette date.

A chaque changement d'état, la date t et les états avant et après ce changement sont consignés dans un journal des changements d'état. D'un point de vue pratique, le cas le

plus commun est supposé être celui où il n'y a pas de changement: on essaie donc de l'écartier tout de suite.

Les entrées et sorties de l'observation aux dates d'entrée e_l et de sortie t_l sont également consignées dans le journal des changements d'état, comme aller ou retour vers l'état fictif indicé 0, correspondant à l'état « non observé ».

Enfin, on renouvelle l'opération: tant que le nombre d'individus (ou de changements d'état) désiré n'est pas atteint, on simule de nouveaux individus.

Données utilisées pour la simulation

Une simulation a été effectuée à l'aide de trois états. Tous les paramètres ont été choisis arbitrairement.

La matrice de passage, contenant les $p_{ij}^\Delta(t)$, probabilités d'être dans l'état ε_j en $t + \Delta t$ sachant que l'on est dans l'état ε_i en t , est la suivante:

$$\left(p_{ij}^\Delta(t) \right)_{(i,j) \in I^2} = \begin{pmatrix} 1 - \dots & \Delta t \cdot \frac{e^{-t}}{2} & \Delta t \cdot \frac{\sqrt{t}}{5} \\ \Delta t \cdot \frac{e^{-(1-t)^2}}{5} & 1 - \dots & \Delta t \cdot \frac{|t-4|}{10} \\ \Delta t \cdot \frac{t}{10} & \Delta t \cdot \frac{1}{10} & 1 - \dots \end{pmatrix}$$

La date d'entrée d'un individu suit une loi Exponentielle de paramètre 1.

La durée d'observation d'un individu suit le triple d'une loi Exponentielle de paramètre 1.

Pour chaque individu, ses changements d'état potentiels sont analysés tous les Δt , depuis sa date d'entrée jusqu'à sa date de sortie.

20 000 individus sont ainsi générés, pour un pas Δt égal à 0.001. L'effectif utilisé est donc un effectif de taille raisonnable, comparable aux effectifs disponibles dans les fichiers d'entreprise.

S'il arrivait que, malgré la finesse du pas, deux changements d'état surviennent en même temps, l'estimation proposée reviendrait à leur donner un ordre arbitraire de survenance.

4.2. CALCUL DE LA VALEUR D'UNE RENTE DE PRESENCE DANS UN ETAT

Soit une période d'étude $[0, \rho]$. Soit un assuré présent dans un état ε_i à l'âge x . Le calcul d'une rente de Δt francs, versée toutes les périodes (de longueur Δt) de l'étude en début desquelles l'assuré est dans un état ε_j , est particulièrement simple: il suffit de cumuler les

flux actualisés, pondérés par leur probabilité de survénance. Dès lors, la prime pure valorisant une telle rente est:

$$P^{(i)}(x) = \sum_{k=0}^{\rho/\Delta t} \Delta t \cdot v^{k \cdot \Delta t} \cdot p_{ij}(x, x + k \cdot \Delta t)$$

où $v = (1 + r)^{-1}$ représente le facteur d'actualisation pour un taux actuariel r .

4.3. CALCUL D'ESPERANCE DE DUREE DE PRESENCE FUTURE DANS UN ETAT

Le calcul de durées moyennes de présence future dans un état, dans la période d'étude $[0, \rho]$, s'assimile au calcul de la valeur d'une rente de montant dt , versée tous les dt en début desquels l'assuré est dans l'état, actualisée au taux 0: le montant versé est alors équivalent à la durée de présence dans l'état.

L'espérance de la durée de présence future dans l'état ε_j , pour un individu présent dans l'état ε_i à l'âge x s'écrit alors:

$$E[D_i^{(j)}(x)] = \sum_{k=0}^{\rho/\Delta t} \Delta t \cdot p_{ij}(x, x + k \cdot \Delta t)$$

En faisant tendre Δt vers 0, on obtient une approximation de la durée future de présence dans un état, tenant compte des allers et retours possibles dans cet état.

4.4. CALCUL D'ESPERANCE DE DUREE DE MAINTIEN CONTINU DANS UN ETAT

Par maintien continu dans un état, on entend ici le fait de rester dans un état entre deux dates, sans jamais en sortir, c'est à dire en excluant le maintien entre deux dates *via* un autre état.

La mesure des probabilités de passage pour le phénomène de maintien continu se dérive très simplement de celle des probabilités de passage dans le cadre général: on cherche désormais à estimer les matrices $M^*(s, t)$ définies par:

$$M^*(s, t) = \left(p_{ij}^*(s, t) \right)_{(i, j) \in I^2}$$

où $p_{ij}^*(s, t) = Pr[E_x = \varepsilon_j \forall x \in [s, t] | E_s = \varepsilon_i]$, $(i, j) \in I^2$, $(s, t) \in R^{+2}$

On a bien entendu $p_{ij}^*(s, t) = 0$ pour $i \neq j$, et les $p_{ii}^*(s, t)$ s'obtiennent par l'estimateur classique, unidimensionnel, de Kaplan-Meier, en présence de données tronquées et censurées.

Les matrices $M^*(s, t)$ sont donc diagonales, et leur estimation peut être menée comme indiqué dans le paragraphe 3.2; on s'intéresse ici au seul phénomène de sortie d'un état, non au phénomène d'entrée, et le raisonnement d'exposition au risque mené plus haut pourra être reconduit, en considérant les individus qui se maintiennent continûment dans chaque état. Dès lors, avec les notations du paragraphe 3.2, on obtient les matrices locales de passage suivantes, pour le maintien continu:

$$\hat{h}^*(\tau^-, \tau^+) = (h_{ij}^*)_{(i,j) \in E^2} \text{ avec } h_{ij}^* = \frac{1}{n_i^*(\tau)} \cdot \left(1_{\{i=i^*, j=i^*\}} \right)$$

Chaque élément de la diagonale de $M^*(s, t)$ correspond bien alors à l'estimateur classique, unidimensionnel, de Kaplan-Meier.

En pratique, il suffira donc d'ajouter une indicatrice pour savoir si l'on calcule les probabilités de passage dans un cadre de maintien continu ou de maintien avec aller-retour possible. Les résultats tels que durée future moyenne de présence dans un état, valorisation d'une rente, etc. sont alors les mêmes que ceux présentés dans les paragraphes précédents, ils s'appuient simplement sur des estimations différentes de probabilités de passage.

D'autres applications peuvent être envisagées: par exemple, dans les cas où l'arrivée dans un état engendre des flux financiers seulement après une durée de maintien continu d appelée franchise, on peut estimer à l'aide des probabilités dont l'estimation est détaillée plus haut, les probabilités suivantes:

$$p_{ij}^d(s, t) = Pr[E_x = \varepsilon_j \forall x \in [t-d, t] | E_s = \varepsilon_i]$$

En effet, $p_{ij}^d(s, t) = p_{ij}(s, t-d) \cdot p_{jj}^*(t-d, d)$ pour $\Delta < t-d$.

L'estimation de ces probabilités découle alors directement de celle des $p_{ij}(s, t)$ et des $p_{ij}^*(s, t)$.

Les résultats tels que durées moyennes, valorisations de rentes, etc. tenant compte de cette franchise se calculent alors comme précédemment.

4.5. RESULTATS

Quelques résultats numériques sont exposés dans ce paragraphe. Ils permettent notamment de comparer les résultats « estimés », obtenus à l'aide de l'estimateur proposé au vu du fichier créé, et les résultats « théoriques », obtenus directement avec les $p_{ij}^\Delta(t)$ exposés dans le paragraphe 4.1, à l'origine du fichier observé des changements d'état.

Les graphiques et tableaux exposés ci-après indiquent respectivement:

- le nombre d'individus en cours d'observation à chaque instant t
 - les probabilités théoriques $p_{ij}(0,t)$ et estimées $\hat{p}_{ij}(0,t)$
 - les probabilités théoriques $p_{ij}(1,t)$ et estimées $\hat{p}_{ij}(1,t)$
 - les probabilités théoriques $p_{ij}(t,t+1)$ et estimées $\hat{p}_{ij}(t,t+1)$
 - les durées de présence future dans chacun des états, pour un individu d'état n° 1, 2 ou 3 en $t=1$
 - les durées de maintien continu dans chacun des états, pour un individu présent à la date $t=1$
 - la valorisation d'une rente de présence dans un état, pour un individu d'état n° 2 en $t=1$
- Pour des raisons de commodité de lecture, seules apparaissent sur les graphiques les probabilités de passage ayant pour état initial $i = \varepsilon_2$.

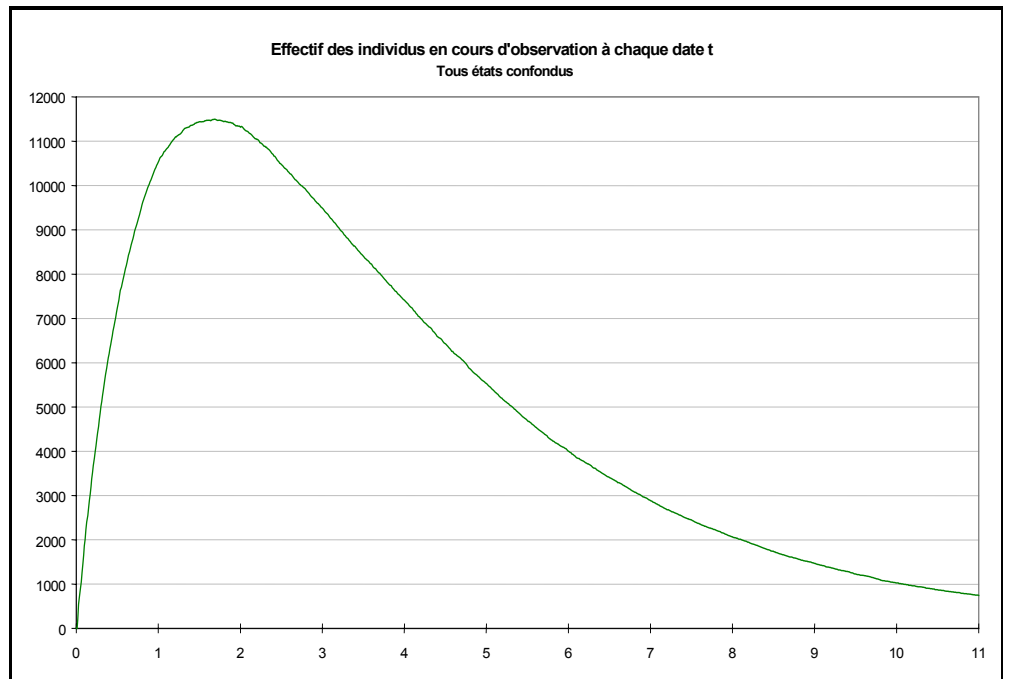
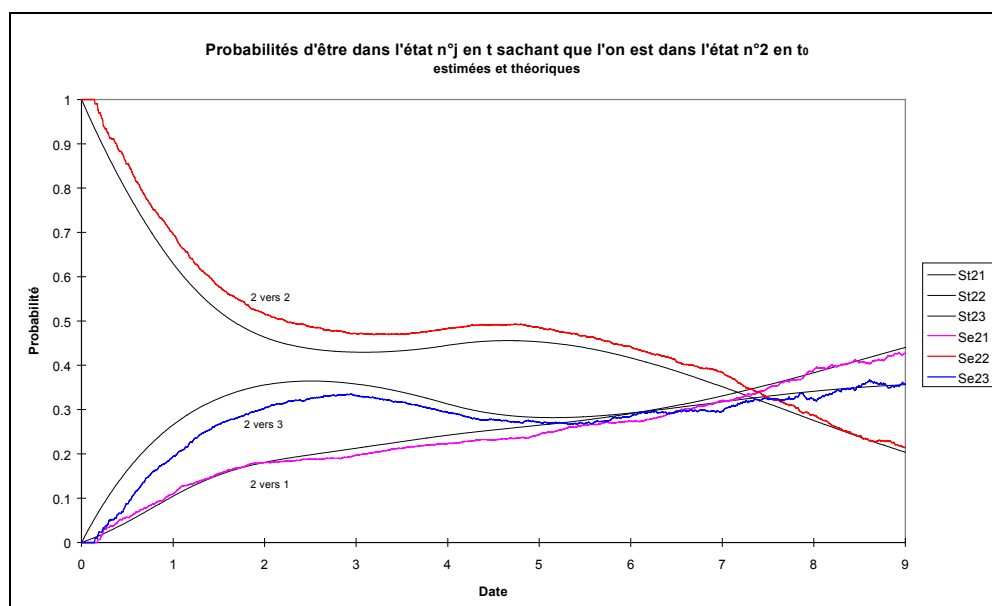


Figure 1

Effectif total en cours d'observation à chaque date t ,

pour le fichier simulé (tous états confondus).



*Figure 2. Probabilités de passage de l'état n° 2 à l'état n° j, entre $t_0=0$ et t .
Le faible effectif autour de la date 0 induit des écarts d'estimation dès l'origine,
qui se répercutent sur les dates postérieures.*

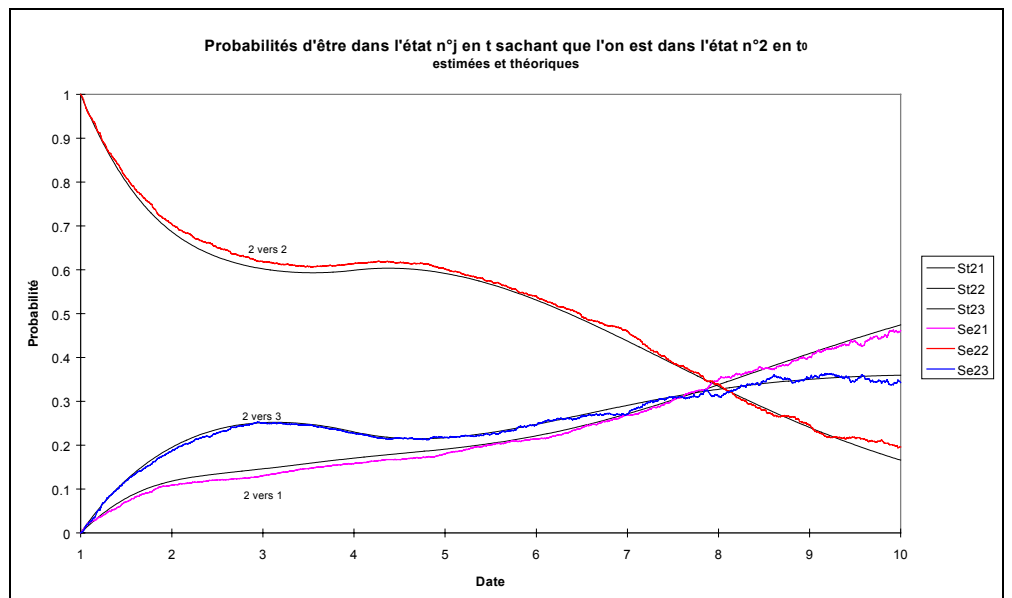


Figure 3. Probabilités de passage de l'état n° 2 à l'état n° j, entre $t_0=1$ et t . Par un conditionnement simple ($t_0=1$), on réduit considérablement l'incertitude des estimations.

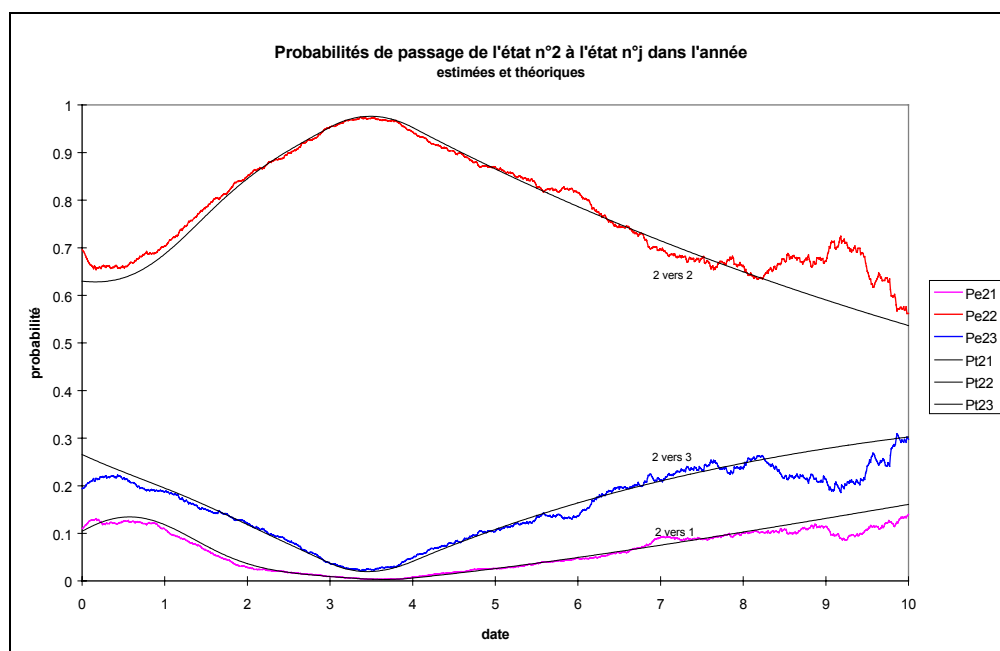


Figure 4.
Probabilités de passage de l'état n° 2 à l'état n° j, entre t et t+1.

Chaque estimation est indépendante des estimations en dehors de l'intervalle $[t, t+1[$.

Espérance de durée de présence future dans un état

	Théorique	Observée	écart
état 1	5.343	5.325	-0.3%
état 2	1.407	1.418	0.8%
état 3	3.251	3.257	0.2%
Total	10.000	10.000	-

Tableau 1

Espérance de durées futures de présence dans un état, pour un individu présent dans l'état n° 1 en $t=1$, sur la période 1-11, avec un pas égal à 1/100.

Espérance de durée de présence future dans un état

	Théorique	Observée	écart
état 1	2.615	2.533	-3.1%
état 2	4.712	4.839	2.7%
état 3	2.674	2.627	-1.7%
Total	10.000	10.000	-

Tableau 2

Même tableau, pour un individu présent dans l'état n° 2 en $t=1$;
Les individus entrant dans l'état n° 1, l'effectif observé dans l'état n° 2 est inférieur ce qui induit une moindre précision des mesures.

Espérance de durée future dans un état

	Théorique	Observée	écart
état 1	3.614	3.610	-0.1%
état 2	1.569	1.605	2.3%
état 3	4.817	4.785	-0.7%
Total	10.000	10.000	-

Tableau 3

Même tableau, pour un individu présent dans l'état n° 3 en $t=1$

Espérance de durée future de maintien continu

	Théorique	Observée	écart
état 1	2.508	2.492	-0.6%
état 2	4.021	4.147	3.1%
état 3	2.555	2.522	-1.3%
Total	9.083	9.161	0.9%

Tableau 4

Espérance de durée de maintien continu dans un état, sur la période [1-11],
pour un individu présent dans l'état indiqué à la date $t=1$.

Prix d'une rente de présence dans un état

	Théorique	Observée	écart
état 1	4.580	4.565	-0.3%
état 2	1.182	1.189	0.5%
état 3	2.701	2.709	0.3%
Total	8.462767	8.462767	-

Tableau 5

Valorisation d'une rente anticipée de un franc annuel, versée chaque centième d'année
lors de la présence dans un état, pour un individu présent dans l'état n° 1 en $t=1$,
sur la période 1-11. Rente actualisée en $t=1$, au taux de 3.5 %.

A titre de vérification, on a bien: $\ddot{a}_{10^{-1}}^{(100)} = \frac{1}{100} \cdot \frac{1 - v^{(1/100)^{10}}}{1 - v^{(1/100)}} \approx 8.462767$.

BIBLIOGRAPHIE

- [1] AALEN O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*. Vol. 6, N° 4.
- [2] COCOZZA-THIVENT C. (1997). Processus stochastiques et fiabilité des systèmes. *Springer*.
- [3] DROESBEKE J.J., FICHET B., TASSI P. (1989). Analyse statistique des durées de vie. *Economica*.
- [4] FLEMING T.R. (1978). Nonparametric estimation for non homogeneous Markov processes in the problem of competing risks. *The Annals of Statistics*. Vol. 6, N° 5.
- [5] HABERMAN S., BLOOMFIELD D.S.F. (1990). Work time lost to sickness, unemployment and stoppages: measurement and application. *Journal of the institute of actuaries*. Vol. 117, Part III.
- [6] JACOBSEN M. (1982). Statistical Analysis of Counting Processes. *Lecture notes in statistics*. 12. *Springer-Verlag*.
- [7] JANSSEN J., MANCA R. (1997). A realistic non-homogeneous stochastic pension fund model on scenario basis. *Scandinavian Actuarial Journal*. Vol. 2.
- [8] KAPLAN E.L., MEIER P. (1958). Nonparametric estimation from incomplete observations. *Journal of The American Statistical Association*. 53.
- [9] MOLLER C.M.. (1992). Numerical evaluation of Markov transition probabilities based on the discretized product integral. *Scandinavian Actuarial Journal*. Vol. 1.
- [10] PITACCO E. (1995). Actuarial models for pricing disability benefits: Towards a unifying approach. *Insurance: Mathematics and Economics*. 16.
- [11] RULLIERE D., SERANT D. (1997). Généralisation de l'estimateur Kaplan-Meier d'une loi de durée de maintien en présence d'observations tronquées à gauche. Extension à l'étude conjointe de deux durées de maintien. *Bulletin Français d'Actuariat*. Vol. 1, N° 2.