

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{t \leq T_x}$$

[ressources-actuarielles.net](http://ressources-actuarielles.net)



## MODÈLES DE DURÉE

Support de cours 2023-2024

Méthodes de lissage et d'ajustement

Frédéric PLANCHET

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t:\infty} [T_x]$$

## SOMMAIRE

<b>1. Introduction .....</b>	<b>3</b>
1.1. Régularité et précision .....	3
1.2. Les différentes approches .....	4
1.3. Définitions et notations .....	4
<b>2. Les méthodes d'ajustement paramétrique .....</b>	<b>5</b>
2.1. Lissage par splines.....	5
2.1.1. Spline cubique à deux arcs .....	6
2.1.2. Splines cubiques : le cas général .....	7
2.1.3. Choix des paramètres .....	7
2.1.4. Généralisation.....	7
2.2. Interpolations à jonctions lisses .....	7
2.2.1. La formule d'interpolation d'Everett .....	8
2.2.2. Les formules à quatre points.....	8
<b>3. Les méthodes de lissage non paramétrique.....</b>	<b>9</b>
3.1. Les moyennes mobiles.....	9
3.2. Les méthodes de noyau.....	10
3.3. Méthode de Whittaker-Henderson .....	11
3.3.1. Cas de la dimension un.....	11
3.3.2. Extension en dimension deux .....	12
3.3.3. La méthode de Whittaker-Henderson comme lissage bayésien.....	15
3.4. Les lissages bayésiens .....	16
3.4.1. Présentation générale.....	16
3.4.2. La méthode de Kimeldorf-Jones .....	17
<b>4. Validation de l'ajustement .....</b>	<b>18</b>
4.1. Test du Khi-deux.....	18
4.2. <i>Standardized mortality ratio</i> (SMR).....	19
4.3. Test des changements de signe .....	19
<b>5. Références.....</b>	<b>20</b>

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{t \leq T_x}$$

## 1. Introduction

Lors de la construction d'une loi de survenance ou d'une loi de maintien, la première étape consiste en l'estimation de taux bruts<sup>1</sup>; cette étape est incontournable, que l'approche retenue soit une approche paramétrique ou une approche non paramétrique.

Les valeurs ainsi obtenues présentent une certaine irrégularité, dont il est légitime de penser qu'elle ne reflète pas le phénomène sous-jacent que l'on cherche à mesurer, mais qu'elle est le résultat de l'imperfection des conditions de l'expérience; les fluctuations d'échantillonnage induisent ainsi une variabilité « parasite » dans les valeurs estimées.

On souhaite alors « ajuster » ou « lisser » les valeurs brutes pour représenter de manière plus fidèle la loi (inconnue) que l'on veut estimer.

De manière formelle, et en se plaçant dans le cas standard de l'estimation de taux d'incidence (taux de mortalité, taux de sortie de l'incapacité, taux d'entrée en dépendance, etc.), la procédure d'estimation initiale a conduit à une valeur  $\hat{q}_x$  pour estimer  $q_x$ , et donc à une erreur  $e_x = \hat{q}_x - q_x$ ; l'objectif de la révision de l'estimation que l'on va conduire est de diminuer cette erreur, tout en construisant une courbe des taux en fonction de  $x$  plus « lisse » que la courbe des taux bruts.

On voit également que le processus d'estimation des taux bruts considère en général les différents âges (ou anciennetés) indépendamment les uns des autres, et ne tient donc pas compte des relations qui existent manifestement entre eux: par exemple dans le cas de la mortalité, et au moins à partir d'un certain âge, il semble naturel que la série théorique des  $q_x$  soit croissante. La série des  $\hat{q}_x$  n'est en général pas croissante.

On présente dans ce support les méthodes de base. Des extensions peuvent être trouvées dans PLANCHET et TOMAS [2013], avec les lissages par vraisemblance locale<sup>2</sup>.

### 1.1. Régularité et précision

Le choix d'une procédure de révision des données brutes fait intervenir deux types de contraintes qui devront être prises en considération de manière conjointe:

- ✓ la précision (ou fidélité): il est naturel d'attendre des taux révisés qu'ils soient proches des taux initiaux;
- ✓ la régularité: la suite des taux ajustés sera recherchée aussi régulière que possible.

L'arbitrage entre précision et régularité dépend du phénomène observé. Par exemple, dans un contexte d'arrêt de travail, on peut observer des taux bruts ayant l'allure suivante:

---

<sup>1</sup> Ou l'estimation empirique d'une caractéristique de la distribution de survie: fonction de hasard cumulée, fonction de survie, etc.

<sup>2</sup> Voir par exemple sur ce sujet [cette présentation](#).

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [ (T_x) ]$$

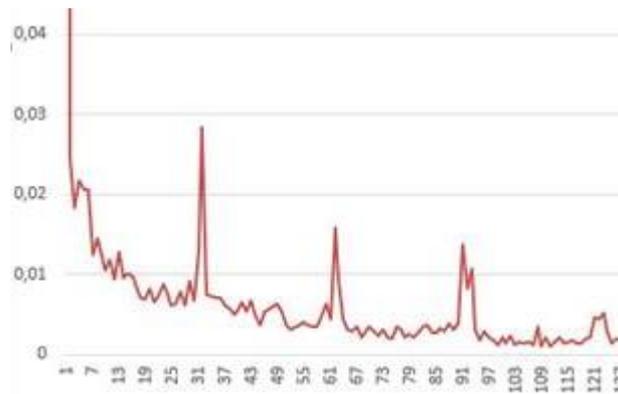


Fig. 1: Taux de sortie bruts quotidiens

Les pics observés les fins de mois correspondent en pratique à des fins d'arrêts conséquence de décisions d'une commission médicale et ne doivent donc pas être éliminés. En revanche, les fluctuations d'un jour sur l'autre aux autres dates sont la conséquence de fluctuations d'échantillonnage et doivent être lissées.

Comme lors de toute démarche statistique, la pertinence de la procédure d'ajustement devra être validée par des tests.

### 1.2. Les différentes approches

Le processus de révision de l'estimation initiale peut être conduit de deux manières :

- ✓ on peut se fixer une forme a priori pour la loi sous-jacente, en supposant par exemple que la fonction de hasard est une fonction de Makeham ; il s'agit là d'une démarche d'ajustement à une distribution définie par un certain paramètre  $\theta$  ; on trouve dans cette classe de procédures les lissages par splines et les interpolations à « jonction lisse » ;
- ✓ on peut ne pas chercher une représentation paramétrique, mais simplement définir un certain nombre de traitement à appliquer aux données brutes initiales pour les rendre plus « lisses » ; on trouve dans ces méthodes non paramétriques les moyennes mobiles<sup>3</sup>, les lissages de type Whittaker-Henderson et leur extension dans un cadre bayésien plus général.

D'un point de vue sémantique, on aura tendance par la suite à parler d'« ajustement » dans le cadre paramétrique et de « lissage » dans le cadre non paramétrique, quoi que cette terminologie ne soit pas fixée<sup>4</sup>.

Les deux approches sont développées ci-après.

### 1.3. Définitions et notations

Les méthodes de lissage font souvent appel à la série des différences terme à terme de la série d'origine. On distingue en pratique trois opérateurs de différentiation discrète :

<sup>3</sup> Qui sont d'une certaine manière les ancêtres des procédures de lissage.

<sup>4</sup> Les anglo-saxons parlent de *graduation*.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

ressources-actuarielles.net

- ✓ différence avant :  $\Delta u(x) = u(x+1) - u(x)$
- ✓ différence arrière :  $\nabla u(x) = u(x) - u(x-1)$
- ✓ différence centrale :  $\delta u(x) = u\left(x + \frac{1}{2}\right) - u\left(x - \frac{1}{2}\right)$

Ces différents opérateurs peuvent être appliqués de manière récursive :  $\Delta^2 u = \Delta(\Delta u)$  ; ainsi  $\Delta^2 u(x) = u(x+2) - 2u(x+1) + u(x)$  ; plus généralement,  $\Delta^n u(x)$  s'écrit en fonction des coefficients binomiaux :

$$\Delta^n u(x) = \sum_{j=0}^n \binom{n}{j} (-1)^{n-j} u(x+j)$$

On obtient des expressions analogues avec  $\nabla$  et  $\delta$ .

## 2. Les méthodes d'ajustement paramétrique

L'approche usuelle décrite dans le support « statistique des modèles paramétriques » consiste à effectuer un ajustement à une loi paramétrique en général continue ; elle ne sera donc pas redéveloppée ici.

On peut d'ailleurs noter que dans ce cas la méthode du maximum de vraisemblance conduit directement des observations aux valeurs ajustées *via* l'obtention des paramètres de la loi sous-jacente ; le raisonnement en deux étapes, obtention de taux bruts, puis ajustement de ces valeurs, peut donc être simplifié.

Dans le cas de l'arrêt de travail, des méthodes initialement développées pour la construction de tables de mortalité prospectives, comme la méthode de Lee-Carter, peuvent être adaptées à l'ajustement de taux instantanés de sortie  $\mu_{xt}$  où  $x$  est l'ancienneté de l'arrêt et  $t$  l'âge à l'entrée<sup>5</sup>.

### 2.1. Lissage par splines

L'ajustement à une loi continue suppose implicitement que la courbe des taux d'incidence peut être représentée sur toute la plage d'âges ou d'ancienneté considérée par une seule fonction paramétrique. En pratique, du fait par exemple de ruptures dans l'évolution des taux bruts, cette condition est assez restrictive.

L'idée du lissage par splines est de découper la plage de la fonction à ajuster en sous-intervalles, puis d'ajuster sur chaque sous-intervalle une fonction simple, en prenant des précautions pour le raccordement aux points de jonction. Un découpage bien choisi doit en effet permettre d'utiliser sur chaque sous-intervalle une fonction sensiblement plus simple que la fonction qu'il aurait fallu ajuster globalement.

Les polynômes sont des fonctions simples et peuvent à ce titre être utilisés pour construire des lissages par spline ; en pratique, on considère à présent uniquement des

<sup>5</sup> Voir le support de cours « [tables de mortalité](#) » pour la présentation de ces méthodes

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

ressources-actuarielles.net

polynômes de degré 3 qui vont nous permettre de construire des splines cubiques. Le raccordement de ces arcs se fera en imposant aux points de jonction la continuité ainsi que l'égalité des pentes et des courbures.

### 2.1.1. Spline cubique à deux arcs

Dans un premier temps, on suppose qu'un découpage en 2 parties de la plage de variation des âges (ou anciennetés) est suffisant, et on pose donc :

$$q_x = \begin{cases} p_0(x) & x_0 \leq x \leq x_1 \\ p_1(x) & x_1 \leq x \leq x_2 \end{cases}$$

$p_i(x)$  étant un polynôme de degré 3, avec les contraintes au point de jonction :

$$p_0(x_1) = p_1(x_1) \quad \frac{d}{dx} p_0(x_1) = \frac{d}{dx} p_1(x_1) \quad \frac{d^2}{dx^2} p_0(x_1) = \frac{d^2}{dx^2} p_1(x_1)$$

Cela conduit à poser  $p_0(x) = c_1 + c_2x + c_3x^2 + c_4x^3$  et  $p_1(x) = p_0(x) + c_5(x - x_1)^3$ . Le problème comporte donc 5 inconnues (les 8 coefficients des polynômes moins les 3 contraintes de régularité). On utilise pour le résoudre un critère de moindres carrés pondérés, sur la base de poids ( $w_x$ ), qui conduit à rechercher les paramètres qui

$$\text{minimisent } M = \sum_{x=x_0}^{x_2} w_x (q_x - \hat{q}_x)^2.$$

On peut noter que dans cette approche il n'est pas nécessaire de disposer de toutes les valeurs brutes  $\hat{q}_x$  et que le spline pourra être interpolant pour les valeurs manquantes. Si on note alors  $\bar{x}_1$  la plus grande valeur de  $x$  inférieure ou égale à  $x_1$  pour laquelle on dispose d'une valeur de  $\hat{q}_x$ , on décompose la somme intervenant dans le critère  $M$  en deux sommes puis on écrit les équations normales en annulant les dérivées par rapport aux paramètres :  $\frac{\partial M}{\partial c_i} = 0$ .

Après calculs, ces équations peuvent se mettre sous la forme :

$$X'wXc = X'w\hat{q}$$

la matrice  $X$  de taille  $(m,5)$  pour  $m$  valeurs de  $\hat{q}_x$  disponibles sur  $[x_0, x_2]$  étant définie par :

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \bar{x}_1 & \bar{x}_1^2 & \bar{x}_1^3 & 0 \\ 1 & \bar{x}_1 & (\bar{x}_1)^2 & (\bar{x}_1)^3 & (\bar{x}_1 - x_1)^3 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_2 & x_2^2 & x_2^3 & (x_2 - x_1)^3 \end{bmatrix}$$

avec  $\bar{x}_1^1$  la valeur de l'indice postérieure à  $\bar{x}_1$  pour laquelle  $\hat{q}_x$  est connue.

### 2.1.2. Splines cubiques : le cas général

Les expressions ci-dessus se généralisent aisément au cas de  $n$  nœuds  $x_1, \dots, x_n$ , avec une matrice de taille  $(m, n+4)$  ; les coefficients sont obtenus par l'estimateur des moindres carrés ordinaires :

$$c = (X'wX)^{-1} X'w\hat{q}.$$

### 2.1.3. Choix des paramètres

Le choix des paramètres se ramène ici au choix de l'emplacement des nœuds. Comme la fonction d'ajustement est un polynôme de degré 3, on remarque que si on ne dispose que de 4 valeurs de  $\hat{q}_x$  dans l'intervalle  $[x_i, x_{i+1}]$ , les valeurs ajustées seront égales aux valeurs brutes : il convient donc de choisir des intervalles suffisamment larges ; en pratique les changements de forme de la courbe brute pourront donner des indications sur l'emplacement des nœuds.

### 2.1.4. Généralisation

La méthode de lissage par splines peut être présentée dans un cadre plus général de régression non paramétrique, qui permet notamment d'introduire un arbitrage fidélité / régularité au travers d'un critère analogue à celui de Whittaker-Henderson (voir la section 3.3 ci-dessous).

Cette présentation ne sera pas développée ici, mais le lecteur intéressé pourra consulter sur ce sujet BESSE et CARDOT [2001].

## 2.2. Interpolations à jonctions lisses

Si au lieu de calculer une estimation brute à chaque âge, on est amené à regrouper les individus par groupe d'âges, on obtient ainsi des estimations brutes par « paliers » (un palier correspond à un regroupement d'âges). Cette situation se rencontre par exemple dans le cas de la construction de lois de maintien en arrêt de travail, avec regroupement par âge pour l'âge à l'entrée.

On souhaite alors se ramener à des estimations pour chaque âge. La technique des « interpolations à jonctions lisses » répond à cet objectif. On notera que l'objectif n'est

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

pas ici de lisser ou d'ajuster des valeurs irrégulières, mais de compléter des valeurs manquantes. La technique de l'ajustement pas splines cubiques permet également d'interpoler des valeurs manquantes, et elle devra être utilisée de manière préférentielle à ces interpolations.

### 2.2.1. La formule d'interpolation d'Everett

On s'intéresse ici aux procédures d'interpolation symétriques, au sens où le sens de l'interpolation ne modifie pas le résultat. Il est alors possible d'écrire la formule d'interpolation sous la forme générale suivante :

$$q_{x+s} = F(s) \hat{q}_{x+1} + F(1-s) \hat{q}_x$$

pour  $0 \leq s \leq 1$  et  $F(s)$  un opérateur de différentiation de la forme :

$$F(s) = p_0(s) + p_1(s) \delta^2 + p_2(s) \delta^4 + \dots$$

avec  $\delta$  l'opérateur de différence centrale et les  $p_i$  des polynômes. On peut par exemple imaginer que l'on a des taux bruts quinquennaux, et on va utiliser cette approche avec  $s = 0, 2; 0, 4; \dots$  pour obtenir les taux annuels par interpolation. Si dans l'expression de  $F(s)$  ci-dessus le dernier terme est en  $\delta^{2m}$ , alors  $q_{x+s}$  est déterminé à partir des  $2(m+1)$  valeurs  $\hat{q}_{x-m}, \dots, \hat{q}_{x+m+1}$ . On parle alors de « formule d'interpolation à  $2(m+1)$  points ». Lorsque  $m = 0$  et  $p_0(s) = s$  on obtient la formule d'interpolation linéaire classique. Enfin, on peut noter que  $q_{x+s}$  est un polynôme en  $s$ . On retrouve donc une interpolation de type spline, comme en 2.1 ci-dessus.

On impose des conditions de régularité à la fonction d'interpolation ainsi obtenue :

- ✓ la continuité implique que  $p_i(0) = 0$  ;
- ✓ l'égalité des dérivées aux points de jonction, qui implique l'égalité (formelle)  $2F'(1) = F'(0)(2 + \delta^2)$  ;
- ✓ l'égalité des dérivées secondes aux points de jonction, qui implique  $p_i''(0) = 0$ .

Les conditions ci-dessus n'imposent pas que la formule d'interpolation redonne les valeurs d'origine aux points entiers, *i.e.* on peut avoir  $q_x \neq \hat{q}_x$ . Pour que l'on retrouve les valeurs d'origine aux points entiers, il faut de plus que  $p_0(1) = 1$  et  $p_i(1) = 0, i \geq 1$ .

On peut de même déterminer des conditions sur les polynômes  $p_i$  pour que la formule soit exacte pour interpoler les polynômes de degré  $z$  fixé<sup>6</sup>.

### 2.2.2. Les formules à quatre points

En pratique les formules à 4 points, avec donc  $m = 2$  sont les plus utilisées. On cherche donc à exprimer  $q_{x+s}$  en fonction de  $\hat{q}_{x-1}, \hat{q}_x, \hat{q}_{x+1}, \hat{q}_{x+2}$  ; on a  $F(s) = p_0(s) + p_1(s) \delta^2$  et il

<sup>6</sup> On pourra se reporter à LONDON [1995].

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

faut donc déterminer les polynômes  $p_0$  et  $p_1$ . Les contraintes de continuité, de dérivabilité et d'invariance pour les polynômes de degré un implique aisément que :  $p_0(s) = s$ ,  $p_1(0) = 0$ ,  $p_1'(0) = 0$  et  $p_1'(1) = \frac{1}{2}$  ; on contrôle la régularité via  $l = p_1(1)$ . Au global, on a donc 4 contraintes et le polynôme de degré minimal qui les satisfait est de degré 3, avec l'expression suivante :

$$p_1(s) = \left(3l - \frac{1}{2}\right)s^2 + \left(\frac{1}{2} - 2l\right)s^3$$

Le cas particulier  $l = 0$  conduit à la formule de Karup-King, souvent utilisée en pratique :

$$p_1(s) = \frac{1}{2}s^2(s-1).$$

### 3. Les méthodes de lissage non paramétrique

#### 3.1. Les moyennes mobiles

Les moyennes mobiles ont l'avantage de la simplicité de mise en œuvre ; toutefois elles présentent un certain nombre d'inconvénients, liés pour l'essentiel à la sensibilité de la moyenne arithmétique aux valeurs extrêmes, qui conduiront à ne pas les utiliser souvent. À tout le moins, la moyenne mobile ne sera pas alors le seul moyen de révision des taux bruts mis en œuvre.

Pour autant que l'on se restreigne aux moyennes mobiles symétriques, la formule de base est ici :

$$q_x = \sum_{i=-r}^{+r} a_i \hat{q}_{x+i}$$

avec  $a_{-i} = a_i$ . La limitation majeure des moyennes mobiles est que leur utilisation aux bords pose un problème.

Dans l'optique de diminuer une erreur d'estimation dans le cadre de la mesure de taux théoriques  $q_x$ , on pourra se fixer des contraintes consistant à exiger que si la série des  $q_x$  présente la régularité d'un polynôme, par exemple de degré 3, alors on souhaite que l'application de la moyenne mobile ne modifie pas les valeurs de  $q_x$  ; en d'autres termes,

on écrit  $q_x = \sum_{i=-r}^{+r} a_i q_{x+i}$  ce qui conduit à :

$$\sum_{i=-r}^{+r} a_i = 1 \quad \sum_{i=-r}^{+r} i^2 a_i = 0$$

Dans certains cas (arrêt de travail, avec les dimensions âge à l'entrée et ancienneté de l'arrêt, ou mortalité prospective avec les dimensions âge et année), la double dimension des tables amène à rechercher la méthode de lissage la mieux adaptée dans chacune des dimensions.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} (T_x)$$

L'idée est que, si on choisit de lisser dans un sens puis dans l'autre, on risque de détériorer le premier lissage. On cherche donc à lisser la série dans les deux sens en une seule fois. La méthode des moyennes mobiles est bien adaptée pour cela. Elle est par exemple utilisée pour lisser les provisions calculées à partir des tables réglementaires brutes. Dans ce cas le BCAC utilise :

$$q(i, j) = \frac{1}{2} \hat{q}(i, j) + \frac{1}{16} \sum_{k,l} \hat{q}(k, l),$$

la sommation étant étendue aux 8 points entourant  $(i, j)$ .

Cette démarche peut se généraliser à n'importe quelle moyenne mobile ; cependant, ce procédé est mal adapté au traitement des bords du tableau, et conduit en pratique à laisser subsister des irrégularités à ce niveau.

### 3.2. Les méthodes de noyau

Lorsqu'on estime la fonction de survie par une méthode non-paramétrique, la fonction obtenue n'est pas dérivable, et il n'est donc pas possible de recalculer simplement la fonction de hasard. Lorsqu'on souhaite régulariser la fonction de répartition empirique

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}},$$

une méthode classique consiste à choisir une fonction  $K$  régulière, positive et d'intégrale unité, appelé noyau, puis à poser :

$$\hat{f}_{h,n}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right).$$

Lorsque  $h \rightarrow 0$  et  $nh \rightarrow +\infty$ ,  $\hat{f}_{h,n}(x)$  converge vers  $f(x)$  et on a :

$$\sqrt{nh}(\hat{f}_{h,n}(x) - f(x)) \rightarrow N\left(0, f(x) \int K(u)^2 du\right).$$

Parmi les noyaux souvent utilisés on peut citer le noyau gaussien  $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$

et le noyau d'Epanechnikov  $K(u) = \frac{3}{4}(1-u^2)1_{\{|u| \leq 1\}}$ . La principale difficulté pratique

consiste à déterminer la « bande passante »  $h$  optimale.

Dans le contexte d'un modèle de durée censuré, l'application directe de ces formules introduit un biais du fait de la censure (droite). MARRON et PADGETT [1987] ont proposé d'utiliser l'estimateur :

$$\hat{f}_{h,n}(t) = \frac{1}{h} \sum_{i=1}^n \frac{d_i}{r_i} K\left(\frac{T_i - t}{h}\right)$$

dans lequel le saut uniforme  $\frac{1}{n}$  du cas non censuré est remplacé par le saut de

l'estimateur de Kaplan-Meier lors d'une sortie non censurée.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

ressources-actuarielles.net

Pour une présentation complète de ces méthodes, le lecteur pourra se reporter à WAND et JONES [1995].

### 3.3. Méthode de Whittaker-Henderson

Le principe de la méthode de Whittaker-Henderson est de combiner un critère de fidélité et un critère de régularité et de rechercher les valeurs ajustées qui minimisent la somme des deux critères.

#### 3.3.1. Cas de la dimension un

On se fixe des poids  $(w_i)$  et on pose pour le critère de fidélité<sup>7</sup> :

$$F = \sum_{i=1}^p w_i (q_i - \hat{q}_i)^2$$

et pour le critère de régularité<sup>8</sup> :

$$S = \sum_{i=1}^{p-z} (\Delta^z q_i)^2$$

$z$  étant un paramètre du modèle. Le critère à minimiser est une combinaison linéaire de la fidélité et de la régularité, le poids de chacun des deux termes étant contrôlé par un second paramètre  $h$  :

$$M = F + h \times S$$

La solution de ce problème d'optimisation satisfait aux conditions  $\frac{\partial M}{\partial q_i} = 0, 1 \leq i \leq p$  ; la résolution de ce système d'équations peut être effectuée au moyen de quelques manipulations matricielles. À cet effet, on pose  $q = (q_i)_{1 \leq i \leq p}$ ,  $\hat{q} = (\hat{q}_i)_{1 \leq i \leq p}$  et  $w = \text{diag}(w_i)_{1 \leq i \leq p}$  ; avec ces notations on peut écrire  $F = (q - \hat{q})' w (q - \hat{q})$  ; pour ce qui concerne le critère de régularité, si on note  $\Delta^z q = (\Delta^z q_i)_{1 \leq i \leq p-z}$ , alors  $S = (\Delta^z q)' \Delta^z q$ . Pour détailler cette écriture, on introduit la matrice  $K_z$  de taille  $(p-z, p)$ , dont les termes sont les coefficients binomiaux d'ordre  $z$  dont le signe alterne et commence positivement pour  $z$  pair<sup>9</sup> :

$$\Delta^z q(i) = \sum_{j=0}^z \binom{z}{j} (-1)^{z-j} q(j+i).$$

Par exemple pour  $z = 2$  et  $p = 5$  on a :

<sup>7</sup> « fit » en anglais.

<sup>8</sup> « smoothness » en anglais.

<sup>9</sup> Voir la section 1.3

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{t;\infty} (T_x)$$

$$K_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

Si  $p = 3$  et  $z = 1$ , on obtient  $K_2 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$ . On vérifie aisément que  $\Delta^z q = K_z q$ , ce qui permet finalement d'écrire le critère  $M$  sous la forme :

$$M = (q - \hat{q})' w (q - \hat{q}) + h q' K_z' K_z q$$

En développant l'expression ci-dessus on trouve que :

$$M = q' w q - 2q' w \hat{q} + \hat{q}' w \hat{q} + h q' K_z' K_z q$$

ce qui conduit à :

$$\frac{\partial M}{\partial q} = 2w q - 2w \hat{q} + 2h K_z' K_z q.$$

La résolution de  $\frac{\partial M}{\partial q} = 0$  conduit alors à l'expression des taux ajustés :

$$q^* = (w + h K_z' K_z)^{-1} w \hat{q}$$

L'inversion de la matrice  $C = w + h K_z' K_z$  nécessite toutefois certaines précautions, car  $h K_z' K_z$  n'est pas inversible, et l'addition du terme  $w$  rend  $C$  inversible, mais de ce fait l'inversion de  $C$  peut être délicate. On peut en pratique utiliser la décomposition de Cholesky de la matrice symétrique positive  $C$  pour l'inverser.

### 3.3.2. Extension en dimension deux

L'extension de la méthode de Whittaker-Henderson en dimension 2 (ou plus) ne pose pas de problème majeur. On dispose donc d'estimations  $\hat{q} = (\hat{q}_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$  ; le critère de fidélité se généralise immédiatement :

$$F = \sum_{i=1}^p \sum_{j=1}^q w_{ij} (q_{ij} - \hat{q}_{ij})^2.$$

L'extension en dimension 2 du critère de régularité est un peu plus délicate ; on distingue d'abord la régularité verticale via l'opérateur  $\Delta_v^z$  (qui agit sur  $q_{ij}$  à  $j$  fixé vu comme une série indexée par  $i$ ) qui permet de calculer un indice de régularité verticale :

$$S_v = \sum_{j=1}^q \sum_{i=1}^{p-z} (\Delta_v^z q_{ij})^2.$$

De la même manière on calcule l'indice de régularité horizontale  $S_h$ , puis on pose :

$$M = F + \alpha \times S_v + \beta \times S_h$$

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{t;\infty} (T_x)$$

qui doit être minimisé. La résolution du problème d'optimisation s'effectue en réarrangeant les éléments pour se ramener au cas unidimensionnel. Pour cela on définit le vecteur de taille  $p \times q$ ,  $u$  tel que :  $u_{q(i-1)+j} = \hat{q}_{ij}$  ; cela revient à prendre pour les  $q$  premiers éléments du vecteur  $u$  la première ligne de la matrice  $\hat{q}$ , puis les éléments de la seconde ligne, et ainsi de suite. De même on fabrique une matrice de poids en copiant sur la diagonale les lignes de la matrice  $(w_{ij})$ . On pose donc  $w_{q(i-1)+j, q(i-1)+j}^* = w_{ij}$ . On procède de la même manière pour définir les matrices  $K_z^v$  et  $K_y^h$  (cf. KNORR [1984]). Les valeurs lissées s'obtiennent alors par :

$$q^* = (w^* + \alpha K_z^v {}^t K_z^v + \beta K_y^h {}^t K_y^h)^{-1} w^* u.$$

### Un exemple

Voici un cas concret simple qui illustre cette méthode. Les taux bruts forment une matrice  $p \times q$  avec  $p = 4$  et  $q = 3$ . On choisit  $z = 2$  (resp.  $y = 1$ ) comme degré de régularité verticale (resp. horizontale), on aura  $K_v^z$ , de dimensions  $(q(p-z), m) = (6, 12)$  et  $K_h^y$ , de dimensions  $(p(q-z), m) = (8, 12)$ . La construction en dimension un a déjà été faite précédemment. La ventilation de la matrice verticale donne :

$$K_h^2 = \begin{pmatrix} 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -2 & 0 & 1 \end{pmatrix}$$

La ventilation de la matrice horizontale conduit à :

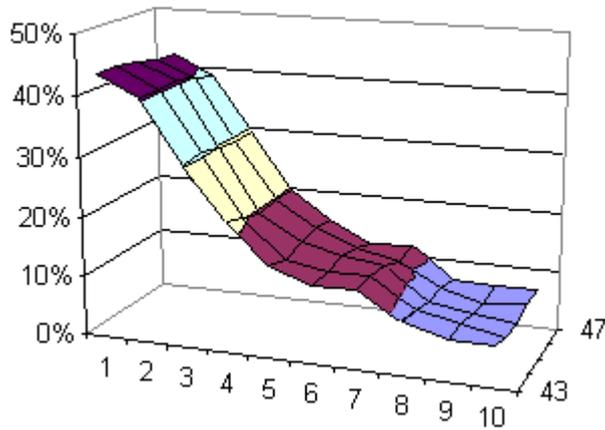
$$K_h^1 = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

La méthode de Whitaker-Henderson en dimension 2 peut notamment être appliquée sur des données de maintien en arrêt de travail, qui sont usuellement estimées en fonction de l'âge à l'entrée et de l'ancienneté de l'arrêt. Cette méthode permet un lissage conjoint dans les deux directions, plus efficace que le lissage séparé selon chaque variable. Elle est illustrée ci-dessous sur des données de ce type :

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

ressources-actuarielles.net

Taux bruts



Taux lissés

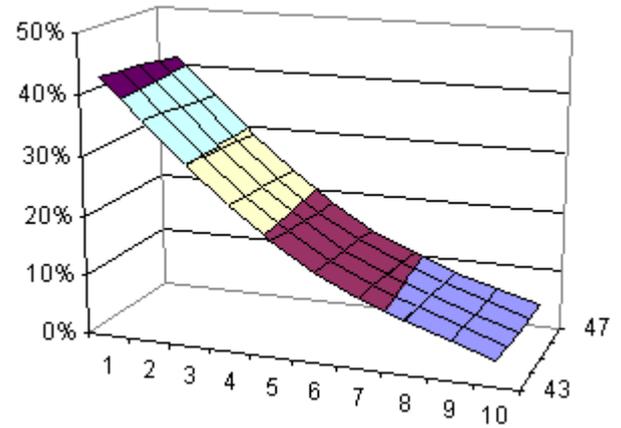


Fig. 2 : *Taux bruts VS lissés*

En pratique, en superposant les résultats après lissage et les taux bruts, on obtiendra des figures comme<sup>10</sup> :

<sup>10</sup> Une comparaison dans ce contexte entre les méthodes de Brass et de Whittaker-Henderson en fonction de la taille de l'échantillon est proposée dans [ce post de blog](#).

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{t \leq \omega} (T_x)$$

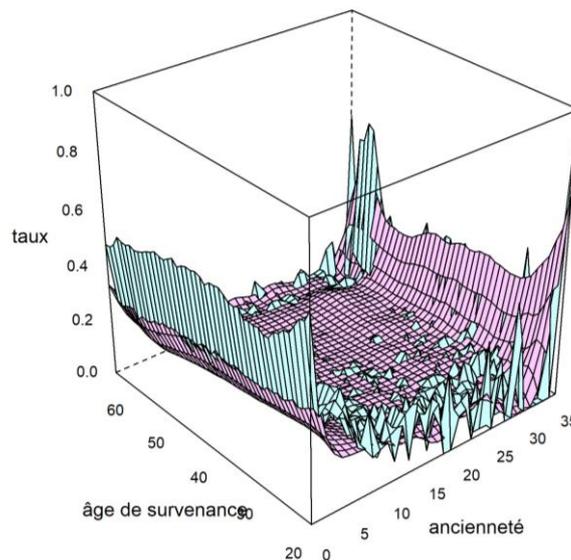


Fig. 3 : Probabilités conditionnelles de sortie brutes et lissées en incapacité de travail

### 3.3.3. La méthode de Whittaker-Henderson comme lissage bayésien<sup>11</sup>

La mesure de régularité  $S(q) = \sum_{i=1}^{p-z} (\Delta^z q_i)^2$  peut être utilisée pour définir une loi *a priori* pour le vecteur  $q = (q_x)$  ; en effet, on est naturellement conduit à supposer que de petites valeurs de  $S$  sont associées à des valeurs « plus probables » de  $q$ , ce qui amène à poser comme densité *a priori* :

$$f_q(q) = c \exp(-\lambda S)$$

avec  $c$  une constante de normalisation<sup>12</sup> et  $\lambda$  un paramètre. Si on fait de plus l'hypothèse que l'erreur d'estimation  $e_x = \hat{q}_x - q_x$  est distribuée selon une loi normale centrée de variance  $\sigma_x^2$ , et que les différents âges (ou anciennetés) sont indépendants, on trouve que :

$$f_{\hat{q}|q}(\hat{q}|q) = c \exp\left(-\frac{1}{2} \sum_{i=1}^p \frac{(\hat{q}_i - q_i)^2}{\sigma_i^2}\right)$$

On se trouve alors dans les conditions d'application du théorème de Bayes pour écrire la densité de  $q = (q_x)$  conditionnellement aux observations  $\hat{q} = (\hat{q}_x)$  :

$$f_{q|\hat{q}}(q|\hat{q}) = c \exp\left(-\lambda S - \frac{1}{2} \sum_{i=1}^p \frac{(q_i - \hat{q}_i)^2}{\sigma_i^2}\right)$$

<sup>11</sup> Voir notamment TAYLOR [1992].

<sup>12</sup> La notation  $c$  désigne dans la suite un terme constant qui peut changer en fonction de la formule considérée.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty}[T_x]$$

Maximiser cette expression en  $q$  est équivalent à minimiser  $M = \lambda S + \frac{1}{2} \sum_{i=1}^p \frac{(q_i - \hat{q}_i)^2}{\sigma_i^2}$  ; on reconnaît le critère de Whittaker-Henderson ; le lissage de Whittaker-Henderson a donc une interprétation probabiliste dans le cadre des lissages bayésiens, décrits de manière plus générale ci-après.

### 3.4. Les lissages bayésiens

On a vu que la méthode de Whittaker-Henderson pouvait être justifiée par un raisonnement probabiliste en faisant une hypothèse sur la loi *a priori* des vraies valeurs et sur la forme de la loi conditionnelle des erreurs. Cette démarche se généralise et conduit à la définition des lissages bayésiens, dont le plus connu est la méthode de Kimeldorf-Jones.

#### 3.4.1. Présentation générale

On considère donc  $q = (q_x)$  le vecteur des vraies valeurs que l'on suppose issu d'une loi *a priori*  $f_q(\cdot)$ . Dans le cadre de l'expérience de mesure, on réalise des observations issues de la loi conditionnelle sachant  $q$ ,  $f_{\hat{q}|q}(\cdot)$ . Cette loi est dans le cas de l'estimation de taux de mortalité un produit de lois binomiales de paramètres<sup>13</sup>  $(n_x, q_x)$  ; d'une manière générale cette loi est déterminée par le contexte de l'expérience<sup>14</sup>.

À partir de ces deux distributions, on utilise la formule de Bayes pour déterminer la distribution *a posteriori* de  $q$  :

$$f_{q|\hat{q}}(q|\hat{q}) = \frac{f_{\hat{q}|q}(\hat{q}|q) f_q(q)}{f_{\hat{q}}(\hat{q})}$$

Enfin, à partir de cette distribution *a posteriori* on peut définir une version révisée de l'estimateur  $\hat{q}$ , par exemple par la méthode du « maximum de vraisemblance » en retenant le mode de la densité *a posteriori* comme valeur ajustée.

La difficulté de cette approche est de définir les caractéristiques de la loi *a priori*. La méthode de Kimeldorf-Jones, présentée ci-dessous, fournit un cadre de mise en œuvre de l'approche bayésienne relativement général et bien adapté au contexte des lois d'incidence (ou de maintien), dans lequel certaines indications utiles pour définir les caractéristiques de la loi *a priori* sont fournies.

<sup>13</sup> Que l'on peut approcher par une loi normale multidimensionnelle lorsque le nombre d'observations est important.

<sup>14</sup> C'est à dire de la procédure d'estimation des taux bruts.

### 3.4.2. La méthode de Kimeldorf-Jones<sup>15</sup>

On se place dans le contexte de l'estimation de taux d'incidence. On fait l'hypothèse que  $f_q(\cdot)$  est une loi normale multidimensionnelle de paramètres  $(m, \Sigma)$  ; on a donc :

$$f_q(q) = \left[ (2\pi)^p |\Sigma| \right]^{-1/2} \exp \left( -\frac{1}{2} (q-m)' \Sigma^{-1} (q-m) \right)$$

avec  $q = (q_1, \dots, q_p)$  les  $p$  taux d'incidence à estimer. À ce stade de la démarche, on suppose donné, et connu,  $(m, \Sigma)$ . Dans ce contexte, et comme on l'a vu en 3.4.1 ci-dessus, la loi de l'expérience est un produit de lois binomiales, qui peuvent chacune être approchées par une loi normale, et on peut donc écrire :

$$f_{\hat{q}|q}(\hat{q}|q) = \left[ (2\pi)^p |V| \right]^{-1/2} \exp \left( -\frac{1}{2} (\hat{q}-q)' V^{-1} (\hat{q}-q) \right)$$

car les estimateurs bruts initiaux sont sans biais ; la matrice  $V$  est ici diagonale, avec les éléments  $n_i q_i (1 - q_i)$ . La densité *a posteriori* est donc de la forme :

$$f_{q|\hat{q}}(q|\hat{q}) = C \exp \left( -\frac{1}{2} \left[ (q-m)' \Sigma^{-1} (q-m) + (\hat{q}-q)' V^{-1} (\hat{q}-q) \right] \right),$$

$C$  étant une constante ; après quelques manipulations on peut mettre cette expression sous la forme :

$$f_{q|\hat{q}}(q|\hat{q}) = C' \exp \left( -\frac{1}{2} (q-z)' W^{-1} (q-z) \right)$$

avec  $z = (\Sigma^{-1} + V^{-1})^{-1} (V^{-1} \hat{q} + \Sigma^{-1} m)$  et  $W^{-1} = \Sigma^{-1} + V^{-1}$ . Cela montre que cette loi est également normale. On en déduit l'expression naturelle des taux ajustés<sup>16</sup> :

$$q^* = (\Sigma^{-1} + V^{-1})^{-1} (V^{-1} \hat{q} + \Sigma^{-1} m).$$

Cette expression signifie que les taux ajustés sont une moyenne pondérée entre les taux bruts et les taux *a priori*. Cette expression peut être écrite de la manière suivante, utile pour les calculs :

$$q^* = \hat{q} + (I_p + \Sigma V^{-1})^{-1} (m - \hat{q})$$

**Remarque :** en toute rigueur les éléments de la matrice diagonale  $V$  sont les  $n_i q_i (1 - q_i)$  qui sont inconnus ; on peut dans la formule ci-dessus les remplacer par l'estimateur  $n_i \hat{q}_i (1 - \hat{q}_i)$ , soit par les estimations *a priori*  $n_i m_i (1 - m_i)$ .

<sup>15</sup> Voir KIMELDORF et JONES [1967].

<sup>16</sup> L'espérance d'une loi normale étant également son mode.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [ (T_x) ]$$

Au-delà de l'obtention d'une version lissée des taux bruts, cette approche offre au surplus l'intérêt de fournir un cadre simple de modèle stochastique de mortalité, en fournissant une loi (normale) pour les taux de sortie du modèle.

Il reste toutefois à expliciter les paramètres de lissage  $(m, \Sigma)$  ; les valeurs de  $m$  sont l'idée que l'on peut se faire des taux d'incidence sans observations ; on peut par exemple utiliser une référence comme les tables du BCAC en arrêt de travail, une table de mortalité INSEE pour des taux de décès, etc. La situation est plus délicate pour la matrice de variances-covariances  $\Sigma$ .

L'idée est de décrire une dépendance en fonction de l'écart d'âge ou d'ancienneté décroissante entre les taux ; cela conduit à proposer la forme suivante pour les coefficients de la matrice  $\Sigma$  :

$$\Sigma_{ij} = \sigma_i \sigma_j \rho^{|i-j|}$$

avec  $\sigma_i^2 = \text{var}(q_i)$  et  $\rho$  le coefficient de corrélation de 2 termes consécutifs. Il reste à fixer des valeurs numériques pour ces  $p+1$  éléments.

#### 4. Validation de l'ajustement

Une fois l'ajustement terminé, il convient de vérifier sa validité par quelques tests statistiques classiques, qui permettront notamment de s'assurer que les taux révisés ne sont pas trop éloignés des taux bruts. Toutefois, dans certaines situations (construction d'une table de mortalité d'expérience par exemple), l'application de règles de prudence peut conduire à des valeurs des taux ajustés délibérément différentes des taux bruts (plus élevées par exemple pour un contrat en cas de décès), et les tests statistiques devront alors être adaptés à cette situation. Concrètement, il faudra abandonner les tests bilatéraux au profit de tests unilatéraux mieux adaptés à la situation.

L'objectif de cette section n'est pas de décrire de manière exhaustive les tests à conduire, mais plutôt de donner quelques pistes et illustrations pour conduire cette démarche de validation.

##### 4.1. Test du Khi-deux

Une fois l'ajustement (ou le lissage) effectué, le test du Khi-2 permet de vérifier la qualité globale des taux révisés en s'assurant qu'ils ne sont pas « trop loin » des taux estimés. On calcule la statistique :

$$Z = \sum_{i=1}^p n_i \frac{(\hat{q}_i - q_i)^2}{q_i(1 - q_i)}$$

Dans le cas d'un ajustement paramétrique par maximum de vraisemblance avec  $r$  paramètres, alors la distribution (asymptotique) de  $Z$  est un  $\chi^2(p-r-1)$  ; dans le cas d'un lissage non paramétrique le nombre de degrés de libertés est moins facile à déterminer (voir par exemple GIESECKE [1981] pour le cas de Whittaker-Henderson).

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} [T_x]$$

#### 4.2. Standardized mortality ratio (SMR)

Le SMR est défini comme le rapport du nombre de décès observé au nombre de décès prédits dans une population de référence, avec l'objectif de décider si la mortalité du groupe observé est identique à celle du groupe de référence ; on a ainsi :

$$SMR = \frac{D}{E} = \frac{\sum_{i=1}^p D_i}{\sum_{i=1}^p n_i q_i}$$

Dans cette expression,  $E$  est une constante et  $D$  une variable aléatoire binomiale que l'on peut approcher par une loi de Poisson, ce qui conduit à l'intervalle de confiance suivant pour la variable  $SMR$ , dans le cas où la mortalité de référence ne dépend pas du groupe étudié (cf. LIDDELL [1984]) :

$$\frac{D}{E} \left( 1 - \frac{1}{9 \times D} - \frac{u_{1-\frac{\alpha}{2}}}{3\sqrt{D}} \right)^3 \leq SMR \leq \frac{D+1}{E} \left( 1 - \frac{1}{9 \times (D+1)} + \frac{u_{1-\frac{\alpha}{2}}}{3\sqrt{D+1}} \right)^3$$

Dans le cas d'une mortalité de référence dépendant des données (à la suite donc d'un ajustement), comme pour le Khi-2, la formule ci-dessus doit être adaptée en fonction du contexte.

#### 4.3. Test des changements de signe

Le signe de la différence  $d_x = q_x - \hat{q}_x$  a, dans les conditions d'application de l'approximation normale, une probabilité  $\frac{1}{2}$  d'être positif. Si les âges sont indépendants, la probabilité que  $d_x$  et  $d_{x+1}$  n'aient pas le même signe est donc aussi égale à  $\frac{1}{2}$ . Si la plage d'âge considérée contient  $p$  valeurs, alors on en déduit que le nombre de changement de signe dans la série des  $d_x$  est distribué suivant une loi binomiale de paramètres  $\left( p-1, \frac{1}{2} \right)$ . En particulier, le nombre moyen de changements de signe est  $\frac{p-1}{2}$ . On en déduit aisément un test, ou bien à distance fini<sup>17</sup> ou bien en faisant l'approximation gaussienne, en s'appuyant sur la statistique :

$$S = \frac{2n - (p-1)}{\sqrt{p-1}}$$

où  $n$  est le nombre de changements de signe constaté.

<sup>17</sup> Voir l'algorithme de construction d'intervalles de confiance pour une loi binomiale.

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} I_{t;\infty} (T_x)$$

ressources-actuarielles.net

## 5. Références

- BESSE P., CARDOT H. [2001] « *Modélisation statistique de données fonctionnelles* », Université de Toulouse, support de cours
- GIESECKE L. [1981] [Use of the Chi-square Statistic to set Whittaker-Henderson Smoothing Coefficients](#), AMS.
- KIMELDORF G.S, JONES D.A. [1967] « *Bayesian graduation* », TSA, XIX
- KNORR F.E. [1984] « *Multidimensional Whittaker-Henderson Graduation* », TSA, vol. 36.
- LIDDELL F.D.K. [1984] « [Simple exact analysis of the standardised mortality ratio](#) », *Journal of Epidemiology and Community Health*, 38(1): 85–88.
- LONDON D. [1995] « *The revision of estimates* », Actex
- PLANCHET F., TOMAS J. [2013] « [Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance](#) », *Insurance: Mathematics and Economics*, Vol. 52, pp. 573–589.  
<http://dx.doi.org/10.1016/j.insmatheco.2013.03.009>
- SAPORTA G., [1990] *Probabilités, analyse des données et statistiques*», Technip
- TAYLOR G. [1992] « *A bayesian interpretation of Whittaker-Henderson graduation* », IME, 11, 7-16
- WAND M.P., JONES M.C. [1995] *Kernel Smoothing*. London: Chapman and Hall.