ISFA GRANDE ÉCOLE D'ACTUARIAT ET DE GESTION DES RISQUES
INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

MODÈLES DE DURÉE

Exercices – Modèles non paramétriques

Corrigé

Exercice n°1

A partir des données ci-dessous on se propose d'estimer le taux de mortalité annuel q_{80}.

La période d'observation s'étend du 01/01/1990 au 01/01/1992.

Indicateur	Naissance	Entrée	Sortie	Etat lors de la sortie
1	01/03/1910	03/12/1988	01/01/1992	Vivant
2	01/05/1909	03/12/1988	01/02/1990	Décédé
3	01/02/1910	03/06/1990	01/01/1992	Vivant
4	01/05/1910	03/06/1988	01/03/1991	Vivant
5	01/03/1905	03/12/1980	01/01/1992	Vivant
6	01/04/1911	01/06/1991	01/01/1992	Vivant
7	01/04/1910	03/01/1980	01/10/1990	Décédé
8	01/01/1910	03/12/1981	01/01/1992	Vivant

1. Certaines observations sont incomplètes. Quels sont les individus correspondants à ces observations incomplètes ? Explicitez la nature de ces données incomplètes.

> Comme le montre le tableau suivant, l'analyse des données fait apparaître une troncature gauche (l'individu 2 né le 1er mai 1909 qui n'est observé qu'à partir de 80,67 ans) et une censure droite pendant la période d'observation (l'individu 4 né le 1er mai 1910 qui n'est plus observé après 80,83 ans).

Indicateur	Naissance	Entrée	Sortie	Etat	Age début observation	Age fin observation
1	01/03/1910	03/12/1988	01/01/1992	V	79,84	81,84
2	01/05/1909	03/12/1988	01/02/1990	D	80,67	80,76
3	01/02/1910	03/06/1990	01/01/1992	V	79,92	81,91
4	01/05/1910	03/06/1988	01/03/1991	C	79,67	80,83
5	01/03/1905	03/12/1980	01/01/1992	V	84,84	86,84
6	01/04/1911	01/06/1991	01/01/1992	V	78,75	80,75
7	01/04/1910	03/01/1980	01/10/1990	D	79,75	80,50
8	01/01/1910	03/12/1981	01/01/1992	V	80,00	82,00

2. Donnez des estimateurs (dont celui de Kaplan-Meier) du taux de mortalité annuel à 80 ans q_{80} (on suivra notamment pour cela l'évolution de l'exposition au risque entre les âges 80 et 81 ans).

L'estimateur de Kaplan-Meier est obtenu à l'aide du tableau suivant qui retrace l'évolution de l'exposition au risque et les sorties pendant l'âge 80 ans.

Age	Nature	Exposition	Individus	Skm
80,00	E	6	1, 3, 4, 6, 7, 8	100%
80,50	D	5	1, 3, 4, 6, 8	83%
80,67	E	6	1, 2, 3, 4, 6, 8	83%
80,75	C	5	1, 2, 3, 4, 8	83%
80,76	D	4	1, 3, 4, 8	67%
80,83	C	3	1, 3, 8	67%
81,00				**67%**

Il en ressort que l'estimateur de Kaplan-Meier de q_{80} vaut 0,33.

Ce taux peut également être estimé en rapportant le nombre de décès observés sur le nombre d'années risque pendant la période d'observation $= \frac{2}{5,17} = 0,39$.

Indicateur	Durée d'exposition au risque
1	1,00
2	0,08
3	1,00
4	0,83
5	0,00
6	0,75
7	0,50
8	1,00
Total	**5,17**

Certains estimateurs sont moins précis. Regardons par exemple le rapport entre le nombre de décès parmi les individus de 80 ans en date de début d'observation et le nombre de personnes ayant 80 ans à cette date.

Date	Nombre indiv. 80 ans - début année	Individus	Nombre de décès - cours d'année	Individus
01/01/1990	2	2, 8	1	2
01/01/1991	3	1, 3, 4	0	-

Pour la 1re année d'observation $= \frac{1}{2} = 0{,}5$.

Pour la 2e année d'observation $= \frac{0}{3} = 0$.

Pour information, le taux de mortalité à 80 ans est de 8,24 % sur la TD 88/90 et de 1,39 % sur la TPRV 93.

Exercice n°2

Un portefeuille de rentiers présente à une date t_0 la répartition d'âge décrite ci-dessous :

âge	**effectif**
80	*500*
81	*450*
82	*800*

Aux dates t_0+1, t_0+2, t_0+3 (unité de temps annuelle) on a observé respectivement : 45 décès ; 62 décès ; 58 décès.

La mortalité du portefeuille est-elle conforme à la mortalité de référence (tables TPRV 93) donnée ci-dessous ?

âge x	**« l_x »**
.	.
78	77462
79	75480
80	74030
81	72016
82	69780
83	67306
84	64621
85	61719
86	58596
87	55255
88	51700
.	.

Le tableau suivant reprend pour chaque génération, sur les trois années de projection, le nombre de décès prédit par la table.

Age en t0	**t0+1**	**t0+2**	**t0+3**
80	13,60	15,10	16,71
81	13,97	15,46	16,78
82	28,36	30,78	33,27
Table	**55,94**	**61,34**	**66,76**
Observations	**45**	**62**	**58**

Un test d'adéquation du Khi-deux conduit à une statistique de 3,29 correspondant à une p-valeur égale à 0,2. Cette statistique ne conduit pas à rejeter l'hypothèse de conformité avec la table (on rappelle que la p-valeur du test du Khi-deux (à deux degrés de liberté) à 95 % vaut 5,99).

Exercice n°3

On rappelle qu'une table de mortalité « partielle » peut être représentée par la donnée de la fonction $e(x, f)$ représente, pour un individu, l'espérance de maintien en vie dans la tranche d'âges $[x, f]$ sachant que l'individu est en vie à l'âge *x*.

Pour *f* égal à 60 ans on donne le tableau suivant :

âge x	e(x,60)
55	4.81
56	3.86
57	2.91
58	1.95
59	0.98
60	0

1. On demande de calculer les taux conditionnels« q_x » de mortalité aux âges 55, 56, 57, 58 et 59 ans ainsi que les « l_x » pour *x* égal à 56, 57, 58, 59 et 60 ans en choisissant pour l_{55} la valeur 50000.

En remarquant que $e(x, f) = \sum_{n=1}^{f-x} {}_n p_x$, on en déduit que

$$e(x, f) = \left(1 + e(x+1, f)\right) p_x .$$

Donc en commençant par déterminer $p_{59} = e(59, 60)$ puis par récurrence, on obtient le tableau suivant.

ages x	e(x,60)	px	qx	lx
55	4,81	0,9897	0,0103	50 000
56	3,86	0,9872	0,0128	49 486
57	2,91	0,9864	0,0136	48 853
58	1,95	0,9848	0,0152	48 190
59	0,98	0,9800	0,0200	47 460
60	0	-	-	46 511

2. On considère un portefeuille constitué de 1500 personnes âgées de 55 ans, de 1500 personnes âgées de 56 ans et de 5000 âgées de 57 ans. Ce portefeuille est observé pendant un an. Donner l'espérance et la variance de la variable aléatoire : nombre de décès.

Sous l'hypothèse d'indépendance des décès, le nombre de décès sur une année *N* est la somme de trois variables aléatoires distribuées selon des lois binomiales de paramètres respectifs $(1500;0{,}0103)$, $(1500;0{,}0128)$ et $(5000;0{,}0136)$.

$\mathbf{E}[N]=1500*0,0103+1500*0,0128+5000*0,0136=102,41$.

Grâce à l'hypothèse d'indépendance,

$$\mathbf{Var}[N]=\left\{\begin{matrix}1500\times\left(0,0103\times(1-0,0103)+0,0128\times(1-0,0128)\right)\\ +5000\times 0,0136\times(1-0,0136)\end{matrix}\right\}=101,09.$$

3. Même question si le portefeuille est observé deux ans.

Comme ${}_2q_x = q_x + p_x q_{x+1}$, le nombre de décès est toujours une somme de binomiale avec de nouveaux paramètres : $(1500;0{,}023)$, $(1500;0{,}026)$ et $(5000;0{,}029)$.

$\mathbf{E}[N]=216,2$.

Toujours grâce à l'hypothèse d'indépendance, $\mathbf{Var}[N]=210,3$.

Exercice n°4

Un portefeuille de rentiers a été scindé en deux sous ensembles : assurés fumeurs et assurés non fumeurs. On a pour la tranche 80 ans-85 ans estimé les taux de mortalité des deux populations à partir des informations suivantes :

Age	Effectif exposé Non fumeurs	Nombre de décès Non fumeurs	Effectif exposé Fumeurs	Nombre de décès Fumeurs
80-81	250	6	500	15
81-82	400	8	700	16
82-83	150	4	1000	45
83-84	500	10	800	40
84-85	600	18	600	38

1. La mortalité des non fumeurs peut–elle être regardée comme conforme à la mortalité de référence caractérisée par les taux de mortalité suivants :

q_{80}=2,7 % ; q_{81}=3,1 % ; q_{82}=3,5 % ; q_{83}=4 % ; q_{84}=4,5 % .

Un test d'adéquation du khi-deux peut nous aider à répondre à cette question.

Age	Effectif exposé Non fumeurs	Nombre de décès observés Non fumeurs	Nombre de décès prédits Non fumeurs
80-81	250	6	6,75
81-82	400	8	12,4
82-83	150	4	5,25
83-84	500	10	20
84-85	600	18	27

La statistique constatée est égale à 9,93 (p-valeur de 4,14 %) : elle est supérieure à 9,49 qui est le quantile d'ordre 0,95 d'une loi χ_4^2. Ce test nous conduit à rejeter l'hypothèse nulle de conformité de la mortalité des non-fumeurs à la mortalité de référence.

2. Peut-on considérer que la population des fumeurs a une mortalité supérieure à la mortalité de la population des non fumeurs ?

Les taux de mortalité observés sont nettement supérieurs pour les fumeurs que pour les non-fumeurs comme le fait apparaître le tableau suivant.

Age	Taux mortalité des fumeurs	Taux de mortalité des non-fumeurs
80-81	3,00%	2,40%
81-82	2,29%	2,00%
82-83	4,50%	2,67%
83-84	5,00%	2,00%
84-85	6,33%	3,00%

Ces différences sont-elles dues aux fluctuations d'échantillonnages ou reflètent elles une différence de structure de la mortalité ?

Les tests du log-rank et de Gehan peuvent nous aider à répondre à cette question.

On étudie la statistique

$$\varphi_j = \frac{\left[\sum_{i=1}^{N} w_i \left(d_{ij} - d_i \frac{n_{ij}}{n_i} \right)\right]^2}{\sum_{i=1}^{N} w_i^2 d_i \frac{n_i - d_i}{n_i - 1} \frac{n_{i1} n_{i2}}{n_i^2}} := \frac{\left[\sum_{i=1}^{N} w_i A_{ij}\right]^2}{\sum_{i=1}^{N} w_i^2 B_{ij}},$$

qui sous l'hypothèses que les deux groupes sont issus de la même population, est distribuée asymptotiquement selon à khi-deux à 1 degré de liberté. Notons que $\mathbf{Pr}\left[\chi_1^2 \leq 3{,}84\right] = 0{,}95$.

Le test du log-rank consiste à choisir les pondérations $w_i = 1$ et celui de Gehan $w_i = i$.

Age	Effectif exposé Non fumeurs	Nombre de décès Non fumeurs	Effectif exposé Fumeurs	Nombre de décès Fumeurs	**Effectif exposé Total**	**Nombre de décès Total**	**Aij**	**Bij**	**Poids log-rank**	**Poids Gehan**
80-81	250	6	500	15	**750**	**21**	-1,00	4,54	1	750
81-82	400	8	700	16	**1100**	**24**	-0,73	5,44	1	1100
82-83	150	4	1000	45	**1150**	**49**	-2,39	5,33	1	1150
83-84	500	10	800	40	**1300**	**50**	-9,23	11,39	1	1300
84-85	600	18	600	38	**1200**	**56**	-10,00	13,36	1	1200
									13,61	**14,65**

On obtient les statistiques 13,61 et 14,65 qui sont nettement supérieures à 3,84 ce qui conduit à rejeter l'hypothèse nulle.

En conclusion, on peut considérer que la mortalité des fumeurs est supérieure à celle des non-fumeurs.

Exercice n°5

On considère une table de maintien en activité d'un salarié âgé de 60 ans avec, pour causes de sortie de l'activité, le décès et le départ à la retraite. L'age maximal de départ à la retraite est de 65 ans.
r_x désigne le « nombre » de sorties pour cause de départ à la retraite et d_x celui pour cause de décès entre les âges *x* et $x+1$.

âges	**« l_x »**	**« r_x »**	**« d_x »**
60	100 000	18 892	?
61	79 508	18 892	?
62	59 185	18 892	?
63	39 109	18 892	?
64	19 356	18 892	?
65	0		

1. Remplir la colonne des d_x de la table ci-dessus.

âges	**« l_x »**	**« r_x »**	**« d_x »**
60	100 000	18 892	1 600
61	79 508	18 892	1 431
62	59 185	18 892	1 184
63	39 109	18 892	861

64	19 356	18 892	464
65	0		

2. Donner pour ce salarié, les probabilités de sortir de l'état de salarié par décès et par départ à la retraite.

Probabilité de sortir pour cause de décès $= \dfrac{\sum_{k=0}^{4} d_{x+k}}{l_x} = 5{,}54\%$.

Probabilité de sortir pour cause de retraite $= \dfrac{\sum_{k=0}^{4} r_{x+k}}{l_x} = 94{,}46\%$

3. En faisant l'hypothèse que les décès interviennent en milieu d'âge et que les départs à la retraite interviennent à la fin de l'âge concerné, donnez l'espérance de maintien du salarié dans son état quel que soit son âge $x \in \{60,\ldots,64\}$.

Effectuez le même calcul avec une convention différente (exemple : en ne comptant que les années entières passées dans l'entreprise).

Remarquons que $e(64) = \dfrac{r_{64} \times 1 + d_{64} \times 0{,}5}{l_{64}}$ puis que

$$e(x-1) = \frac{l_x \times (1 + e(x)) + r_{x-1} \times 1 + d_{x-1} \times 0{,}5}{l_{x-1}}.$$

On en déduit :

x	e(x)
60	2,94
61	2,45
62	1,97
63	1,48
64	0,99

Si l'on ne compte que les années entières de maintien dans l'état de salarié, on obtient les espérances résiduelles résumées dans le tableau suivant :

x	e(x)
60	1,97
61	1,48
62	0,99
63	0,49
64	0,00

4. Continuons de faire l'hypothèse que les départs à la retraite ne sont possibles qu'à la fin de l'âge concerné (par exemple pour la période [60-61] un départ à la retraite n'est possible qu'à la fin de la 60^{e} année). La mortalité qui a été utilisée pour construire la table ci-dessus a été empruntée à une table de mortalité usuelle dont l'effectif à 60 ans était de l_{60}=44 724. Reconstituez les effectifs de survivants de cette table aux âges 61 62 63 et 64 puis donnez la probabilité de décès entre 60 et 65 ans calculée sur cette table. Faites un rapprochement avec la probabilité calculée en 1. et commentez.

L'hypothèse faite nous permet que les nombres de décès observés le sont sur l'ensemble de l'effectif de début d'âge l_x. On peut ainsi calculer les taux de mortalité et reconstituer la table d'origine.

âges	« l_x »	« r_x »	« d_x »	qx	Lx
60	100 000	18 892	1 600	1,60%	44 724
61	79 508	18 892	1 431	1,80%	43 919
62	59 185	18 892	1 184	2,00%	43 040
63	39 109	18 892	861	2,20%	42 093
64	19 356	18 892	464	2,40%	41 084

On en déduit que la probabilité de mourir entre 60 et 65 ans vaut 8,14 %. Ce taux est naturellement supérieur au de 5,54 % de la question 1. En effet ce dernier taux correspond à la probabilité de décéder entre 60 et 65 ans alors que l'on est encore salarié, or une partie des salariés à 60 ans va prendre sa retraite puis décéder.