Modèles de durée / Examen du 31 janvier 2007

Durée 2h – aucun document n'est autorisé

Exercice n°1 (estimation non paramétrique)

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 censure 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 0,333.

Ce taux peut également être estimé en rapportant le nombre de décès observés sur le nombre d'années risque pendant la période d'observation $= \frac{2}{5{,}17} = 0{,}387$.

Indicateur	Durée d'exposition au risque
1	1
2	0,08
3	1
4	0,83
5	0
6	0,75
7	0,501026694
8	1
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.

Indication : vous pourrez vous aider du tableau

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

Exercice n°2 (Calcul du nombre de décès théorique sur un portefeuille)

On considère un portefeuille observé pendant une durée d'un an ; entre les dates de début et de fin d'observation t et $t+1$ on observe des individus qui entrent dans la période d'observation à l'âge x_i (à une date $t_i^e \geq t$) et qui en sortent à l'âge $x_i + d_i$, d_i étant la durée d'observation de l'individu $i \in I$ (avec la contrainte $t_i^s = t_i^e + d_i \leq t+1$).

On s'intéresse au nombre de décès espéré (« théorique ») observé sur l'exercice pour un âge entier x fixé.

Question n°1

Exprimer en fonction de x_i, d_i et x l'intervalle de temps J_i pendant lequel l'individu i est exposé au risque de décès à l'âge x.

La contribution de l'individu i à ce nombre est conditionnée par le fait que l'intervalle

$$J_i = \left[x_i, x_i + d_i\right[\cap \left[x, x+1\right[$$

soit non vide. On peut écrire de manière équivalente $J_i = \left[x_i \vee x, \left(x_i + d_i\right) \wedge \left(x+1\right)\right[$.

Question n°2

On note $D_x = D\left(\left[x, x+1\right[\right)$ le nombre de décès observés à l'âge x (par convention le décès a lieu à l'âge x si la partie entière de l'âge exact au décès est égale à x).

Exprimez la variable aléatoire D_x en fonction des variables aléatoires T_{x_i} représentant les durées de survie des individus de la population et des intervalles calculés à la question précédente.

On a trivialement :

$$D_x = \sum_{i \in I} \mathbf{1}_{J_i}\left(T_{x_i}\right).$$

Question n°3

Calculez à partir de l'expression établie à la question précédente l'espérance de D_x en fonction de la fonction de survie S sous-jacente au modèle.

On déduit de l'expression précédente que $E\left(D_x\right)=\sum_{i \in I} P\left(T_{x_i} \in J_i\right)$; Si le modèle de durée sous-jacent est associé à une fonction de survie S, alors on trouve :

$$E\left(D_x\right)=\sum_{i \in I} \frac{S\left(x_i \vee x\right)-S\left(\left(x_i+d_i\right) \wedge(x+1)\right)}{S\left(x_i\right)} \mathbf{1}_{\left\{J_i \neq \varnothing\right\}}$$

Question n°4

En pratique on utilise souvent l'approximation $E\left(D_x\right) \approx q_x \times \sum_{i \in I} d_i \mathbf{1}_{\left\{J_i \neq \varnothing\right\}}$ qui s'interprète comme le produit du taux de décès à l'âge x et de l'exposition au risque à cet âge.

Après avoir justifié l'expression « exposition au risque » pour le terme $\sum_{i \in I} d_i \mathbf{1}_{\left\{J_i \neq \varnothing\right\}}$, vous expliciterez les hypothèses qu'il convient de faire pour que cette approximation soit valide. Que pensez-vous de ces hypothèses ?

Cette approximation repose sur l'hypothèse de linéarité de la fonction de survie entre deux âges entiers et sur le fait de confondre x et x_i.

Question n°5

Vous disposez d'observations et vous comptabilisez d_x décès à l'âge x. Proposez un modèle simple vous permettant de décider si les décès observés sont compatibles avec la loi S évoquée ci-dessus.

Indication : vous pourrez chercher à comparer les nombres de décès théoriques et observés dans le cadre d'un modèle binomial.

Cf. support de cours.