Modèles de durée / Examen du 23 mai 2006

Durée 1h – aucun document n'est autorisé

Estimation de taux de décès par maximum de vraisemblance discret

Dans le cadre de la construction d'une table d'expérience pour des contrats en cas de décès, on s'intéresse ici à l'ajustement de taux de mortalité bruts $\hat{q}_x$ par un modèle paramétrique $q_x(\theta)$ pour $\theta \in R^p$.

Question n°1

On note D_x le nombre de décès observés à l'âge x et N_x l'effectif soumis au risque à cet âge.

Quelle est la loi de D_x ?

Le nombre de décès observés à l'âge x, D_x, suit une loi binomiale de paramètres $(N_x, q_x(\theta))$

Question n°2

Déduire de la question précédente la log-vraisemblance du nombre de décès pour une plage d'âges $x_0, .., x_1$.

La vraisemblance associée à la réalisation d'un nombre d_x de décès est donc égale à :

$$P(D_x = d_x) = C_{N_x}^{d_x} \, q_x^{d_x} \, (1 - q_x)^{N_x - d_x}.$$

Pour l'ensemble des observations on obtient donc la log-vraisemblance suivante (à une constante indépendante du paramètre près) :

$$\ln L(\theta) = \sum_x d_x \ln q_x(\theta) + \sum_x (N_x - d_x) \ln(1 - q_x(\theta)).$$

Question n°3

En supposant les effectifs suffisamment grands indiquez la forme de l'approximation normale de la loi du nombre de décès ; on rappelera les conditions d'application de cette approximation.

En déduire la forme de la log-vraisemblance dans ce cas.

L'approximation normale de la loi binômiale est $N\left(\hat{q}_x; \sigma^2(\theta) = \frac{q_x(\theta)(1-q_x(\theta))}{N_x}\right)$.

La log-vraisemblance associée est :

$$\ln(L(\theta)) = \sum_x \ln\left(\frac{1}{\sigma(\theta)\sqrt{2\pi}}\right) - \sum_x \frac{1}{2}\frac{(q_x(\theta)-\hat{q}_x)^2}{\sigma^2(\theta)}.$$

Pour déterminer le domaine de validité de cette approximation, on pourra par exemple utiliser le critère de Cochran, qui consiste à vérifier que $N_x \times \hat{q}_x \geq 5$ et $N_x \times (1-\hat{q}_x) \geq 5$.

Question n°4

Proposez une approximation de la log-vraisemblance déterminée ci-dessus qui ramène le problème de maximisation de la vraisemblance à un problème de moindres carrés pondérés.

On va utiliser la vraisemblance approchée dans laquelle on remplace la variance théorique par la variance estimée. La maximisation de la vraisemblance est alors équivalente à la minimisation de :

$$\sum_x \frac{1}{2}\frac{(q_x(\theta)-\hat{q}_x)^2}{\hat{\sigma}^2} = \sum_x \frac{N_x}{\hat{q}_x(1-\hat{q}_x)}(q_x(\theta)-\hat{q}_x)^2$$

Question n°5

Proposez une statistique permettant de tester la qualité de l'adéquation du modèle. Vous donnerez la loi limite de cette statistique.

On utilise des tests du Chi2, sur la base de la statistique $W = \sum N_x \frac{(\hat{q}_x - q_x)^2}{q_x}$, q_x étant le taux de décès théorique du modèle à l'âge x. La loi asymptotique de W est une loi $\chi^2(p-k-1)$, où p désigne le nombre d'âges intervenants dans la somme. Il convient en pratique de manipuler avec précaution le test du Chi-2, la loi asymptotique n'étant un $\chi^2(p-k-1)$, p étant le nombre de classes et k le nombre de paramètres du modèle que parce qu'ici l'estimateur est du maximum de vraisemblance. Pour d'autres méthodes de détermination du paramètre, ce résultat n'est plus vrai en général (voir FISCHER [1924]).

Question n°6

On se propose d'appliquer la démarche ci-dessus avec un modèle de Makham dont la fonction de hasard est définie par $\mu_x = a + b \times c^x$.

Donnez l'expression de $q_x(\theta)$ dans ce cas et écrivez la log-vraisemblance du modèle.

$$p_x = \mathbf{exp}\left[-\int_x^{x+1} \mu_y . dy\right] \mathbf{. exp} = \left[-\int_x^{x+1} \left(a + b \mathbf{.} c^y\right) dy\right] = \mathbf{exp(}-a\mathbf{)}\ \mathbf{exp}\left[-\frac{b}{\mathbf{ln(}c\mathbf{)}} c^x (c-1)\right]$$

Posons $s = \mathbf{exp}(-a)$ et $g = \mathbf{exp}\left(-\frac{b}{\mathbf{ln}(c)}\right)$, la fonction utilisée pour l'ajustement des taux de décès discrets est donc :

$$q_x = 1 - p_x = 1 - s \mathbf{.} g^{c^x(c-1)} .$$