Modèles de durée / Examen / Janvier 2022

Durée 2h – aucun document n'est autorisé

Corrigé

La qualité de la rédaction, des justifications apportées et de la présentation de la copie seront prises en compte dans la notation. Chaque réponse doit être correctement justifiée : une réponse juste sans justification sera considérée comme fausse.

Prise en compte d'un effet « température » dans un modèle prospectif de mortalité

On se place dans le cadre d'un modèle tronqué à gauche et censuré à droite avec une censure aléatoire droite non informative, avec les notations suivantes :

$$T_i = X_i \wedge C_i \text{ et } D_i = \begin{cases} 1 \ si \ X_i \leq C_i \\ 0 \ si \ X_i > C_i \end{cases}$$

On note $E_i,\ i = 1,\ldots,n$ les instants de troncature gauche.

Question n°1 (2 points) : Rappelez les définitions des termes ci-dessus. : censure aléatoire droite, censure non informative, troncature gauche.

Question n°2 (4 points) : Démontrez que la vraisemblance peut s'écrire

$$\ln L(\theta) = cste + \sum_{i=1}^{n} d_i \ln\left(h_\theta(t_i)\right) + \ln S_\theta(t_i) - \ln S_\theta(e_i).$$

Question n°3 (4 points) : Montrez à l'aide de l'expression ci-dessus, qu'en supposant la fonction de hasard constante μ_x sur un intervalle $[x, x+1[$, on peut, pour estimer μ_x, supposer que le nombre D_x de sorties observées non censurées sur $[x, x+1[$ suit une loi de Poisson dont on précisera le paramètre en fonction de μ_x et de l'exposition au risque E_x. Vous préciserez la définition des instants de troncature gauche et de censure droite pour l'observation de l'intervalle $[x, x+1[$ en fonction des instants e_i et t_i.

Pour la prise en compte de l'effet du temps sur les taux de mortalité, il est donc possible de proposer la spécification suivante :[1]

$$D_{x,t} \approx P\left(E_{x,t} \times \mu_{x,t}\right) \text{ avec } \ln\left(\mu_{x,t}\right) = \alpha_x + \beta_x \times k_t$$

[1] Voir la section 4.4 de ce support.

Question n°4 (2 points) : précisez les contraintes à ajouter sur les paramètres pour que la relation ci-dessus soit suffisante pour définir de manière univoque le lien entre la force de mortalité et les variables du modèle.

Question n°5 (2 points) : Indiquez deux manières d'estimer les paramètres du modèle prospectif de mortalité

On souhaite intégrer une variable explicative supplémentaire dépendant du temps, pour tenir compte de l'effet des températures observées sur la mortalité et on propose donc la modèle suivant :

$$\mathbf{ln}\left(\mu_{x,t}\right) = \alpha_x + \beta_x \times k_t + w_t$$

Question n°6 (4 points) : Montrez que $D_t = \sum_x D_{x,t}$ suit une loi de Poisson dont le paramètre λ_t vérifie une équation $\mathbf{ln}\left(\lambda_t\right) = u_t + w_t$ avec un u_t que l'on précisera.

Pour modéliser l'influence des variations de température sur la mortalité, le modèle suivant, appelé CSDL (*Constrained Segmented Distributed Lag Model*), a été proposé[2] :

$$\mathbf{ln}\left(\lambda_t\right) = \eta\left(t\right) + \sum_{i=0}^{L_1} \beta_{1,i} \times \left(z_{t-i} - \Psi_1\right) + \sum_{i=0}^{L_2} \beta_{2,i} \times \left(z_{t-i} - \Psi_2\right)$$

avec z_t la température moyenne du jour t et $\eta\left(t\right)$ une fonction intégrant les variables explicatives supplémentaires telles que l'année, le jour de la semaine ou le mois.

Question n°7 (2 points) : montrez que l'expression ci-dessus peut se mettre sous la forme $\mathbf{ln}\left(\lambda_t\right) = u_t + w_t$ avec des coefficients *u* et *w* que l'on précisera. En déduire que le modèle CSDL est un cas particulier du modèle $\mathbf{ln}\left(\mu_{x,t}\right) = \alpha_x + \beta_x \times k_t + w_t$ avec un choix adapté de $\eta\left(t\right)$.

---

[2] Cf. MUGGEO, VMR [2010] Analyzing Temperature Effects on Mortality Within the R Environment : The Constrained Segmented Distributed Lag Parameterization. *Journal of Statistical Software* ainsi que la presentation détaillée qui en est faite dans
PINCEMIN G. [2021] Risques climatiques et mortalité, impact du risque canicule à l'horizon 2070, Mémoire d'actuaire, EURIA.

Prise en compte d'un effet "température" dans un modèle prospectif de mortalité

① *** Censure aléatoire droite :**

Soit un échantillon de durées de vie $(X_1, \dots, X_n)$ et un second échantillon indep composé de variables positives $(C_1, \dots, C_n)$.

On dit qu'il y a Censure de type III / aléatoire droite si au lieu d'observer directem^t $(X_1, \dots, X_n)$ on observe $(T_1, D_1), \dots, (T_n, D_n)$

avec

$$T_i = X_i \wedge C_i \quad \text{et} \quad D_i = \mathbb{1}_{X_i \leq C_i}$$

*** Censure non informative :** Cela signifie que la loi de censure est indépendante du paramètre θ.

$\iff$ Les lois de X et C n'ont pas de paramètres commun.

*** Troncature gauche :** $\quad X_i \longrightarrow X_i \mid X_i > E_i$

On dit qu'il y a troncature gauche lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieur à un seuil $c>0$.

$X_i \longrightarrow X_i \mid X_i > c$

② On se place dans le cadre d'un modèle tronqué à gauche et censuré à droite avec censure aléat. droite non informative.

$$\longrightarrow \quad T_i = X_i \wedge C_i \quad \text{et} \quad D_i = \mathbb{1}_{\{X_i \leq C_i\}}$$

On note E_i les instants de troncature gauche.

$$S_T(\theta, t) = \mathbb{P}(T>t, D=1 \mid X_i > E_i)$$
$$= \frac{\mathbb{P}(T>t, D=1, X_i > E_i)}{\mathbb{P}(X_i > E_i)} = \frac{\mathbb{P}(T>t, D=1, X_i > E_i)}{S_X(E_i)}$$

Troncature gauche E.

$$X_i \longrightarrow X_i \mid X_i > E_i$$

$$S_\theta(t_i) \longrightarrow \frac{S_\theta(t_i)}{S_\theta(e_i)}$$

$$h_\theta(t_i) \longrightarrow h_\theta(t_i)$$

$(X_1, \dots, X_n)$ iid

$$\ln \alpha_\theta(x_1, \dots, x_n) = \sum_{i=1}^{n} \ln(f_\theta(x_i))$$

Censure $\longrightarrow$ on n'observe pas X_i mais on observe $\quad Y_i \begin{cases} T_i = X_i \wedge C_i \\ D_i = \mathbb{1}_{X_i \leq C_i} \end{cases} \qquad (C_1, \dots, C_n) \perp\!\!\!\perp (X_1, \dots, X_n)$.

$(Y_1, \dots, Y_n)$ iid

$$\begin{aligned}
\mathbb{P}(T>t, D=1) &= \mathbb{P}(X \wedge C > t, X \leq c) \\
&= \mathbb{P}(X > t, X \leq c) \\
&= \mathbb{P}(t < X \leq c) \\
&= \mathbb{E}[\mathbb{P}(t < X \leq c \mid c)] \\
&= \int_t^\infty \int_t^c f_X(u)\,du\, f_C(c)\,dc \overset{\text{Fubini}}{=} \int_t^\infty \int_u^\infty f_C(c) f_X(u)\,dc\,du \\
&= \int_t^\infty f_X(u) S_C(u)\,du = S_X(t) S_C(t)
\end{aligned}$$

$t \leq c < \infty$

$t \leq u \leq c$

$$\longrightarrow -\frac{d}{dt}\mathbb{P}(T>t, D=1) = f_X(t) S_C(t)$$

$$\mathbb{P}(T>t, D=0) = f_C(t) S_X(t)$$

$$\mathcal{L}(Y_1, \dots, Y_n) = \prod_{i=1}^{n} [f_X(t_i) S_C(t_i)]^{d_i} [f_C(t_i) S_X(t_i)]^{1-d_i}$$

$$\boxed{\mathcal{L} = \prod_{i=1}^{n} h_X(t_i)^{d_i} S_X(t_i) \prod_{i=1}^{n} S_C(t_i) h_C(t_i)^{1-d_i}}$$