Annale type :

① Rappelez les définitions des termes : censure aléatoire droite, censure non informative, troncature gauche

Censure aléatoire droite : Soit un échantillon de durées de vie $(X_1, \dots, X_n)$ et un second échantillon indépendant composé de variables positives $(C_1, \dots, C_n)$.

On dit qu'il y a censure aléatoire droite si au lieu d'observer directement $(X_1, \dots, X_n)$ on observe $(T_1, D_1), \dots, (T_n, D_n)$

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

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

$\Longleftrightarrow$ les lois de X et C n'ont pas de paramètre commun.

$\rightarrow$ Exemple de censure qui n'est pas non informative : Censure au $r^{ième}$ décès $\Rightarrow C = X_{(r)}$

Troncature gauche : On dit qu'il y a troncature gauche lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieure à un seuil $E_i > 0$

$\rightarrow X_i \rightarrow X_i \mid X_i > E_i$

$\rightarrow$ Vraisemblances :

Dans le cas de la censure aléatoire droite on a :

$$\begin{aligned}
\mathbb{P}(T_i > t_i, D_i = 1) &= \mathbb{P}(X_i \wedge C_i > t_i, X_i \leq C_i) \\
&= \mathbb{P}(X_i > t_i, X_i \leq C_i) \\
&= \mathbb{P}(t_i < X_i \leq C_i) \\
&= \mathbb{E}[\mathbb{P}(t_i < X_i \leq c) \mid C_i = c] \\
&= \int_{t_i}^{\infty} \left(\int_{t_i}^{c} f_X(u)\, du \right) f_C(c)\, dc \overset{\text{Fubini}}{=} \int_{t_i}^{\infty} \int_{u}^{\infty} f_C(c) f_X(u)\, dc\, du \\
&= \int_{t_i}^{\infty} f_X(u)\, S_C(u)\, du = S_X(t_i)\, S_C(t_i)
\end{aligned}$$

$t_i \leq c < \infty$
$t_i \leq u \leq c$

$$\Rightarrow -\frac{d}{dt_i} \mathbb{P}(T_i > t_i, D_i = 1) = f_X(t_i)\, S_C(t_i)$$

De même, $-\frac{d}{dt_i} \mathbb{P}(T_i > t_i, D_i = 0) = f_C(t_i)\, S_X(t_i)$

$$\Rightarrow \mathcal{L}(\theta) = \prod_{i=1}^{n} [f_X(t_i) S_C(t_i)]^{D_i} [f_C(t_i) S_X(t_i)]^{1-D_i}$$

Avec l'hypothèse de censure non informative, X et C n'ont pas de paramètre commun.

$$\Rightarrow \mathcal{L}(\theta) = \text{cte} \times \prod_{i=1}^{n} f_X(t_i;\theta)^{D_i}\, S_X(t_i;\theta)^{1-D_i}$$

contient les infos relatives à la censure.

De plus, $f(t) = h(t) S(t)$

$$\Rightarrow \mathcal{L}(\theta) = \text{cte} \times \prod_{i=1}^{n} h_X(t_i;\theta)^{D_i}\, S_X(t_i;\theta)$$

Avec l'hypothèse de troncature gauche $h_{X|E_i}(t_i;\theta) = h_X(t_i;\theta)$ et $S_{X|E_i}(t_i;\theta) = \dfrac{S_X(t_i;\theta)}{S_X(e_i;\theta)}$

$$\Rightarrow \mathcal{L}(\theta) = cte \times \prod_{i=1}^{m} h_x(t_i;\theta)^{D_i} \frac{S_x(t_i;\theta)}{S_x(e_i;\theta)}$$

$$\Rightarrow \ln(\mathcal{L}(\theta)) = cste + \sum_{i=1}^{m} \left[d_i \ln(h_x(t_i;\theta)) + \ln(S_\theta(t_i)) - \ln(S_\theta(e_i)) \right]$$

② Si la fonct° de hasard est cte sur $[x, x+1[$ donner vraisemblance et instants e_i et t_i.

Fonct° de hasard μ_x est cte sur $[x, x+1[\Rightarrow \forall t \in [x, x+1[, \mu_x(t) = \theta \Rightarrow S_x(t) = e^{-\int_0^t \mu_x(u)du} = e^{-\theta t}$

$$\Rightarrow \ln(\mathcal{L}(\theta)) = \sum_{i=1}^{m} \left[d_i \ln(\theta) - \theta(t_i(x) - e_i(x)) \right]$$

$$= d_x \ln(\theta) - \theta E_x$$

avec $d_x = \sum_{i=1}^{m} d_i$ et $E_x = \sum_{i=1}^{m} (t_i(x) - e_i(x))$.

Tout se passe comme si $D_x \sim \mathcal{P}(\theta E_x)$

Et $e_i(x) = e_i \vee x$ et $t_i(x) = t_i \wedge (x+1)$

On considère la spécificat° suivante :

$$\mu(x,t) = Z_t \, \mu_0(x,t)$$

avec $\ln(\mu_0(x,t)) = \alpha_x + \beta_x k_t$

Z_t des chocs en moyenne centrés $\mathbb{E}[Z_t] = 1$

③ Identifier les coefficients du modèle

* α_x : niveau moyen du log des taux instantanés à chaque âge (ils sont généralem^t ↗ sauf aux âges jeunes).

* k_t : Ils décrivent l'évolut°, en log, de la mortalité au cours du temps. Les coeffs k_t sont globalement ↘ et suivent une tendance approximativement linéaire. Cette évolut° du coeff traduit l'accroissement régulier de l'espérance de vie avec le temps (amélioration de l'hygiène, du syst. de santé...)

* β_x : sensibilité par âge de cette tendance d'évolut° (coeff le + souvent positifs).

④ Contraintes et pourquoi ?

Sous sa formulat° non contrainte, le modèle étudié n'est pas identifiable.

En effet, la modification $(\alpha_x, \beta_x, k_t) \longrightarrow (\alpha_x + c\beta_x, \frac{\beta_x}{d}, d(k_t - c))$ n'affecte pas la valeur de $\mu(x,t)$. $(d,c) \in \mathbb{R}$.

Nous devons alors imposer les contraintes $\sum_{x=x_m}^{x_M} \beta_x = 1$ et $\sum_{t=t_m}^{t_M} k_t = 0$

⑤ Poser le problème d'optimisat° considéré dans le modèle traditionnel de Lee-Carter (i.e $\mu(x,t) = \exp(\alpha_x + \beta_x k_t)$) pour déterminer les paramètres (α, β, k). Inconvénient majeur de cette approche ? → Reformulat° pour corriger cet inconvénient ?

Ici on fait par moindres carrés pour garantir unicité

$$\Rightarrow (\hat{\alpha}_x, \hat{\beta}_x, \hat{k}_t) = \underset{(\alpha_x, \beta_x, k_t)}{\arg\min} \sum_{x=x_m}^{x_M} \sum_{t=t_m}^{t_M} \left(\ln(\hat{\mu}(x,t)) - (\alpha_x + \beta_x k_t) \right)^2$$

→ Approche classique : décomposit° en valeurs singulières
Newton Raphson possible

Inconvénient → s'appuie sur l'hyp d'homoscédasticité des erreurs. En pratique pas vérifié.

→ On utilise lien log → $D_{xt} \sim \mathcal{P}(E_{xt}\, \mu(x,t))$

facteur de l'exposition au risque de la population étudiée.

→ On peut utiliser Newton-Raphson multivarié pour estimat° des params.

On considère donc $D_{x,t} \sim \mathcal{P}(E_{x,t}\, \mu(x,t)) \Rightarrow D_{x,t} \sim \mathcal{P}(Z_t \lambda_{x,t})$ avec $\lambda_{x,t} = E_{x,t}\, \mu_0(x,t)$

$$\Rightarrow \mathbb{P}(D=d \mid Z) = e^{-\lambda z} \frac{\lambda^d}{d!} z^d$$

$$\Rightarrow \mathbb{P}(D=d) = \mathbb{E}[\mathbb{P}(D=d \mid Z)] = \int e^{-\lambda z} z^d \frac{\lambda^d}{d!} \, dF_Z(z)$$

On fait le choix que $Z \sim \Gamma(a,b)$

⑥ Donne l'expression de $\mathbb{P}(D=d)$ sachant que $Z \sim \Gamma(a,b)$

$Z \sim \Gamma(a,b) \to f_Z(z) = \frac{b^a}{\Gamma(a)} z^{a-1} e^{-bz}$

$$\Rightarrow \mathbb{P}(D=d) = \int_0^\infty e^{-\lambda z} z^d \frac{\lambda^d}{d!} \times \frac{b^a}{\Gamma(a)} z^{a-1} e^{-bz} \, dz$$

$$= \frac{\lambda^d}{d!} \frac{b^a}{\Gamma(a)} \int_0^\infty z^{(a+d)-1} e^{-(\lambda+b)z} \, dz$$

$u = (\lambda+b)z$
$du = (\lambda+b)dz$

$$= \frac{\lambda^d}{d!} \frac{b^a}{\Gamma(a)} \left(\frac{1}{\lambda+b} \right)^{a+d} \int_0^\infty u^{(a+d)-1} e^{-u} \, du$$

$$= \frac{\lambda^d}{d!} \left(\frac{1}{\lambda+b} \right)^{a+d} \frac{b^a}{\Gamma(a)} \Gamma(a+d) = \frac{b^a \, \Gamma(a+d)}{\Gamma(d+1)\Gamma(a)} \frac{\lambda^d}{(\lambda+b)^{a+d}}$$

$$\Rightarrow \ln(\mathbb{P}(D=d)) = f(a,b) + d \ln(\lambda) + (a+d) \ln(\lambda+b) \qquad \text{avec } f(a,b) = \ln\left(b^a \frac{\Gamma(a+d)}{\Gamma(d+1)\Gamma(a)} \right)$$

$$\Rightarrow \mathcal{L}(\alpha_x, \beta_x, k_t) = \ln(\mathbb{P}(D_{x,t} = d_{x,t}))$$

⑦ Dérivées partielles p/r^t aux ≠^ts paramètres.

Soit $p \in \{\alpha_x, \beta_x, k_t\}$, on a :

$$\frac{\partial}{\partial p} \ell(P(D=d)) = d \times \frac{1}{\lambda}\frac{\partial \lambda}{\partial p} - (d+a) \times \frac{1}{\lambda+b}\frac{\partial \lambda}{\partial p}$$

$$= \left[\frac{d}{\lambda} - \frac{d+a}{\lambda+b} \right] \frac{\partial \lambda}{\partial p}$$

où $\frac{\partial \lambda}{\partial p} = \begin{cases} \lambda & \text{si } p = \alpha_x \\ k\lambda & \text{si } p = \beta_x \\ \beta\lambda & \text{si } p = k_t \end{cases}$

On a donc conditionnellement à (a,b), $\ell(\mathcal{L}) = \sum_{x=x_m}^{x_n} \sum_{t=t_m}^{t_n} \ell(P(D=d|a,b)) = \sum_{x=x_m}^{x_n} \sum_{t=t_m}^{t_n} \left[d_{x,t} \, \ell(\lambda_{x,t}) - (d_{x,t}+a) \ell(\lambda_{x,t}+b) \right]$

Il s'agit donc de maximiser $\ell(\mathcal{L})$ en (α, β, k) sous les contraintes.

⑧ Estimez les paramètres

* Estimat° du paramètre de fragilité Z_t

→ On a supposé que $Z_t \sim \Gamma(a,b)$. De plus $\mathbb{E}[Z_t] = 1$

$\Rightarrow \mathbb{E}[Z_t] = \frac{a}{b} = 1 \Rightarrow a = b \Rightarrow Z_t \sim \Gamma(a,a)$

De plus $\mathbb{V}[Z_t] = \frac{a}{b^2} = \frac{1}{a}$

Donc le paramètre de contrôle de la perturbation Z est l'inverse de la variance $a = (\sigma_Z^2)^{-1}$

De plus $\mu(x,t) = Z_t \, \mu_0(x,t) \Rightarrow Z_t = \frac{\mu(x,t)}{\mu_0(x,t)}$

On peut directem^t estimer ce paramètre en observant que les intensités annuelles moyennes sont tq :

$\bar{\mu}(t) = Z_t \, \bar{\mu}_0(t)$ avec $\bar{\mu}_0(x,t) = \frac{\sum_{x=x_m}^{x_n} E_{x,t} \, \mu_0(x,t)}{\sum_{x=x_m}^{x_n} E_{x,t}}$

$\Rightarrow \mathbb{E}[\bar{\mu}(t)] = \bar{\mu}_0(t)$

$\mathbb{V}[\bar{\mu}(t)] = \mathbb{V}[Z_t] \, \bar{\mu}_0(t)^2$

$\Rightarrow \mathbb{V}[Z_t] = \frac{\mathbb{V}[\bar{\mu}(t)]}{\mathbb{E}[\bar{\mu}(t)]^2}$

$$\Rightarrow \hat{\sigma}_Z^2 = \frac{\frac{1}{t_n - t_m + 1} \sum_{t=t_m}^{t_n} \left(\hat{\mu}(t) - \frac{1}{t_n - t_m + 1} \sum_{t=t_m}^{t_n} \hat{\mu}(t) \right)^2}{\left(\frac{1}{t_n - t_m + 1} \sum_{t=t_m}^{t_n} \hat{\mu}(t) \right)^2}$$

avec $\hat{\mu}(t) = \frac{\sum_{x=x_m}^{x_n} E_{x,t} \, \hat{\mu}(x,t)}{\sum_{x=x_m}^{x_n} E_{x,t}}$ et $\hat{\mu}(x,t) = \frac{D_{x,t}}{E_{x,t}}$ l'estimateur de Hoem.

* Estimat° de (α_x, β_x, k_t) solut° des CPO :

$\rightarrow \frac{\partial}{\partial \alpha_x} \ell(\mathcal{L}) = \sum_{t=t_m}^{t_n} \left[\frac{d}{\lambda_{x,t}} - \frac{d+a}{\lambda_{x,t}+b} \right] \lambda_{x,t} = 0$

$\rightarrow \frac{\partial}{\partial \beta_x} \ell(\mathcal{L}) = \sum_{t=t_m}^{t_n} \left[\frac{d}{\lambda_{x,t}} - \frac{d+a}{\lambda_{x,t}+b} \right] k_t \, \lambda_{x,t} = 0$

$\rightarrow \frac{\partial}{\partial k_t} \ell(\mathcal{L}) = \sum_{t=t_m}^{t_n} \left[\frac{d}{\lambda_{x,t}} - \frac{d+a}{\lambda_{x,t}+b} \right] \beta_x \, \lambda_{x,t} = 0$

Ce syst est non linéaire.

→ recours à Newton-Raphson multivarié.

Newton-Raphson multivarié :

$$\theta_{i+1} = \theta_i - \left[\frac{\partial^2 \ln(\mathcal{L}(\theta))}{\partial\theta_i \partial\theta_j}\right]^{-1} \frac{\partial \ln(\mathcal{L}(\theta))}{\partial\theta}$$

9 Calculer $e(x,t)$ l'espérance de vie prospective.

$$e(x,t) = \sum_{i \geq 0} \prod_{j=0}^{i} \mathbb{E}[\exp(-\mu_{x+j,t+j})]$$

Ici $\mu_{x,t} = Z_t \, \mu_0(x,t)$ avec $Z_t \sim \Gamma(a,b)$

$\Rightarrow \mathbb{E}[e^{-xZ_t}] = \left(\frac{b}{b+x}\right)^a$ (transformée de Laplace d'une $\Gamma(a,b)$)

$$\Rightarrow \mathbb{E}[e^{-\mu(x,t)}] = \mathbb{E}[e^{-\mu_0(x,t) Z_t}] = \left(\frac{b}{b+\mu_0(x,t)}\right)^a = \exp\left(-a \ln\left(1 + \frac{\mu_0(x,t)}{b}\right)\right) *$$

$$\Rightarrow e(x,t) = \sum_{i \geq 0} \prod_{j=0}^{i} \left(\frac{b}{b+\mu_0(x+j,t+j)}\right)^a$$

Donc par $*$, lorsque $b = a \to \infty$, $\mathbb{E}[e^{-\mu(x,t)}] = e^{-\mu_0(x,t)}$

$$\Rightarrow \text{lorsque } b = a \to \infty \quad e(x,t) = \sum_{i \geq 0} \prod_{j=0}^{i} e^{-\mu_0(x+j,t+j)}$$