ENSAE 2022/2023

Actuariat de l'assurance non-vie

2h00

Pas de document, pas de calculatrice.

Exercice 1 - Modèles linéaires généralisés (6 pts). Soit $(\mathbb{P}_\theta)_{\theta\in\mathbb{R}}$ une famille de mesures de probabilité. On suppose qu'elle appartient à la famille exponentielle.

1. Soit $\theta \in \mathbb{R}$, rappelez la densité (par rapport à une mesure dominante) d'une telle loi en introduisant les fonctions mesurables a, b, c et le paramètre de dispersion $\phi > 0$.
2. On suppose b deux fois dérivables, montrez que b est convexe.
3. On introduit $(Y_i)_{1\leq i\leq n}$ des variables aléatoires indépendantes, $(\mathbb{P}_{\theta_i})_{1\leq i\leq n}$ des mesures de probabilité issues d'une même famille exponentielle, et $Y_i \sim \mathbb{P}_{\theta_i}$ pour tout $1 \leq i \leq n$. On se place dans le cadre des modèles linéaires généralisés avec fonction de lien canonique et on pose $\theta_i = x_i'\beta$ où $x_i' \in \mathbb{R}^d$ pour tout $1 \leq i \leq n$ et $\beta \in \mathbb{R}^d$. On introduit la log-vraisemblance $\ell(\beta, y)$ où $y := (y_i)_{1\leq i\leq n}$ sont les observations des $(Y_i)_{1\leq i\leq n}$. Montrez que
$$\ell(\beta, y) \propto \sum_{i=1}^{n} (y_i x_i'\beta - b(x_i'\beta)) .$$
4. On note
$$\bar{\ell}(\beta) := \sum_{i=1}^{n} (y_i x_i'\beta - b(x_i'\beta)) ,$$
puis $\nabla^2\bar{\ell}(\beta)$ la matrice Hessienne de $\bar{\ell}$ par rapport à β. Montrez que
$$\nabla^2\bar{\ell}(\beta) = -\sum_{i=1}^{n} x_i x_i' b''(x_i'\beta).$$
5. En déduire que, si on note $\nabla\bar{\ell}(\beta)$ le gradient de $\bar{\ell}$ par rapport à β, l'équation
$$\nabla\bar{\ell}(\beta) = 0$$
admet une unique solution et qu'il s'agit du maximum global.

Exercice 2 - Tarification a posteriori (7 pts) Pour un assuré $i \in \{1, \ldots, n\}$, on introduit $(N_t^i)_{t\geq 0}$ le nombre de sinistres survenus sur $[0, t]$. Le nombre moyen de sinistres a priori sur une année est $\mathbb{E}_{x_i}\left[N_1^i\right]$ où $\mathbb{E}_{x_i}$ est l'espérance sachant les variables caractéristiques.

1. En introduisant $(z_t^i)_{t\geq 0}$ (sans le définir explicitement), donnez la forme générale du nombre de sinistres moyen estimé a posteriori sur $[t, t + 1]$ avec les variables aléatoires et quantités introduites observées sur $[0, t]$.
2. Que doivent valoir z_0^i et $\lim_{t\to+\infty} z_t^i$? Justifiez brièvement.
3. On introduit Θ_i un paramètre aléatoire et inconnu dont dépend $(N_t^i)_{t\geq 0}$ et $P(\theta_i) := \mathbb{E}_{x_i}\left[N_1^i \mid \Theta_i = \theta_i\right]$. Quelle est la relation entre $\mathbb{E}_{x_i}(N_1^i)$ et $P(\Theta_i)$?
4. On admet que $Cov_{x_i}\left(\frac{N_t^i}{t}, P(\Theta_i)\right) = Var_{x_i}\left[P(\Theta_i)\right]$. En crédibilité non paramétrique, on cherche à minimiser
$$P_t^{i,\star} := \mathbb{E}\left[\left(P(\Theta_i) - a - b\frac{N_t^i}{t}\right)^2\right].$$
Donnez $a^\star$ et $b^\star$ en fonction des moments d'ordre 1 et 2 de $P(\Theta_i)$ et de $\frac{N_t^i}{t}$. Montrez que
$$P_t^{i,\star} = (1 - b^\star)\mathbb{E}_{x_i}\left[P\left(\Theta_i\right)\right] + b^\star\frac{N_t^i}{t}.$$

5. On suppose que $(N_t^i \mid \Theta_i)_{t\geq 0}$ est un processus de Poisson de paramètre Θ_i, en particulier $N_1^i \mid \Theta_i \sim \mathcal{P}(\Theta_i)$ et plus généralement $N_t^i \mid \Theta_i \sim \mathcal{P}(\Theta_i t)$ pour tout $t \geq 0$. Dans le modèle linéaire généralisé de la famille de Poisson surdispersé, on a mesuré

$$\mathbb{E}_{x_i}(N_1^i) = \lambda(x_i), \qquad Var_{x_i}(N_1^i) = \phi\lambda(x_i),$$

où λ_i est une fonction continue et positive, et $\phi > 1$. Déduire $b^\star$ dans ce contexte, donnez la valeur de $(z_t^i)_{t\geq 0}$ associée et remarquez qu'elle ne dépend pas de i mais uniquement de t et de ϕ.

Exercice 3 - Provisionnement (7 pts) Soit $(C_{i,j})_{1\leq i,j\leq n}$ les montants cumulés des sinistres survenus en année i et à l'année de développement j. Soient $\mathcal{F}_k := \sigma(C_{i,j} \mid i+j \leq k+1)$ et $\mathcal{B}_k := \sigma(C_{i,j} \mid i+j \leq n+1, j \leq k)$. pour $1 \leq k \leq n$.

1. Rappelez les trois hypothèses *H1*, *H2* et *H3* du modèle de Chain Ladder - Mack.
2. On pose

$$\widehat{f}_{i,j} := \frac{C_{i,j+1}}{C_{i,j}},$$

et, pour des poids $(p_{i,j})$ qui sont $\mathcal{B}_j-$mesurables

$$\widehat{f}_j := \frac{\sum_{i=1}^{n-j} p_{i,j}\widehat{f}_{i,j}}{\sum_{i=1}^{n-j} p_{i,j}}.$$

Montrez que les $(\widehat{f}_j)$ ne sont pas correlés entre-eux, c'est à dire que pour $j_1 < \ldots < j_k$,

$$\mathbb{E}\left(\prod_{i=1}^{k} \widehat{f}_{j_i} \mid \mathcal{B}_{j_1}\right) = \prod_{i=1}^{k} f_{j_i},$$

et ainsi cette propriété ne dépend pas des poids choisis.

3. On propose de changer *H3* et de la remplacer par :

H3 Il existe $\alpha \in [1,2]$ et pour $1 \leq j \leq n-1$, il existe $\sigma_j \geq 0$ tels que

$$Var(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) = \sigma_j^2 C_{i,j}^{\alpha}, \quad 1 \leq i \leq n.$$

Le paramètre α sera considéré comme fixé et connu. Déterminez les $(p_{i,j})$ qui minimisent l'erreur quadratique moyenne conditionnellement à $\mathcal{B}_j$ et en déduire les $(\widehat{f}_j)$ optimaux.

4. Montrez que l'estimateur habituel de $\widehat{C}_{i,n}$ demeure sans biais ainsi que celui des provisions totales $\widehat{R}$.
5. L'erreur de $\widehat{R}$ se décompose en deux sources : lesquelles ?
6. L'estimateur de la variance des provisions n'est plus valable avec notre nouvelle hypothèse. Quelle autre méthode (qu'il faudrait adapter à notre nouvelle hypothèse) permettrait d'estimer l'écart-type de $\widehat{R}$?