Examen 2025-2026

Mathématiques de l'assurance dommages : théorie et pratique

2h00

Aucun document autorisé, pas de calculatrice.

Exercice 1 - Modèles linéaires généralisés (7 pts). Soit $A, B \subset \mathbb{R}$ et $T : A \to B$ une fonction inversible, dérivable et d'inverse dérivable. Pour des lois continues, soit la famille de densité :

$$f(y, \theta, \phi) = e^{\frac{T(y)\theta - b(\theta)}{a(\phi)} + c(y,\phi)}, \quad y \in A,$$

avec a, b, c des fonctions mesurables.

1. Sous quelle condition cette famille fait-elle partie de la famille exponentielle ?
2. Soit Y une variable aléatoire de densité f. Soit $Z := T(Y)$, montrez que la loi de Z fait partie de la famille exponentielle avec des fonctions éventuellement nouvelles a, b, c qu'on précisera.
3. On souhaite appliquer un modèle linéaire généralisé à Y. Pour cela, on propose de passer par $Z = T(Y)$, de faire le modèle linéaire généralisé dessus ce qui permet d'obtenir l'estimation de

   $$E[Z \mid X] = g^{-1}(\beta_0 + X\beta),$$

   où g est la fonction de lien. Enfin, puisque $Y = T^{-1}(Z)$, on propose d'utiliser

   $$\widehat{E[Y \mid X]} := T^{-1}(E[Z \mid X]).$$

   Discutez de l'erreur de cette approximation en prenant pour exemple $T = \log$.
4. On se place dans le cas où $T = \log$ et $Z = \log(Y) \mid X \sim \mathcal{N}(\beta_0 + X\beta, a(\phi))$. On propose d'utiliser l'estimateur

   $$\widehat{E[Y \mid X]} := \exp\left(E[Z \mid X] + \frac{a(\phi)}{2}\right).$$

   Quel est l'avantage de ce nouvel estimateur comparé au précédent ? Vous pourrez considérer le cas où les paramètres sont connus (puisque qu'ils convergent, cela restera vrai asymptotiquement).

Exercice 2 - Tarification a posteriori (5 pts) Pour un assuré $i \in \{1, \dots, n\}$, on introduit N_i son nombre de sinistres. Pour X_i une variable aléatoire dans $\mathbb{R}^d$ et Θ_i une variable aléatoire dans $\mathbb{R}^q$, $x_i \in \mathbb{R}^d$, $\theta_i \in \mathbb{R}^q$, on suppose que

$$N_i \mid \{X_i = x_i, \Theta_i = \theta_i\} \overset{i.i.d.}{\sim} \mathcal{P}(\lambda(x_i, \theta_i)),$$

où $\lambda : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}_+$ est une fonction mesurable.

1. Que représente Θ_i ?
2. Donnez $\mathbb{E}[N_i \mid X_i = x_i]$ en fonction des moments de $\lambda(x_i, \Theta_i)$.
3. Donnez $Var[N_i \mid X_i = x_i]$ en fonction des moments de $\lambda(x_i, \Theta_i)$.
4. Quels sont les estimateurs de $\mathbb{E}[N_i \mid X_i = x_i]$ et $Var[N_i \mid X_i = x_i]$ obtenus par le modèle linéaire généralisé dans la famille de loi Poisson avec surdispersion en notant $\widehat{P}_i^N$ la prime pure estimée et $\widehat{\phi}$ la dispersion estimée ?

Exercice 3 - Provisionnement (9 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. Donnez $\widehat{f}_j$, l'estimateur de f_j, pour $1 \leq j \leq n-1$.
3. On note $\widehat{f}_j^n$ pour $\widehat{f}_j$ et $\mathcal{B}_j^n$ pour $\mathcal{B}_j$ afin d'exhiber la dépendance en n qu'on fera varier. Montrer que, sous la condition $\sum_{i=1}^{n-j} C_{i,j} \to +\infty$ quand $n \to +\infty$, on a
$$\widehat{f}_j^n \mid \mathcal{B}_j^n \xrightarrow[n\to+\infty]{L^2} f_j.$$
4. On souhaite montrer la convergence p.s., il suffit de montrer que
$$\widehat{f}_j^n - f_j \mid \mathcal{B}_j^n = \left.\frac{\sum_{i=1}^{n-j} \varepsilon_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}}\right| \mathcal{B}_j^n \xrightarrow[n\to+\infty]{p.s.} 0.$$
On utilisera le théorème des deux séries de Kolmogorov (cas particulier des variables aléatoires centrées) :
Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires centrées et indépendantes telles que $Var(X_i) < \infty$ pour $i \geq 1$. Alors
$$\sum_{i\geq i_0} Var(X_i) < +\infty \Longrightarrow \sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{p.s.} \xi,$$
où $i_0 \geq 1$ et ξ est une variable aléatoire réelle. On pose $\varepsilon_{i,j} := C_{i,j+1} - f_j C_{i,j}$. Montrez que
$$\left.\sum_{i=1}^{n-j} \frac{\varepsilon_{i,j}}{\sum_{k=1}^{i} C_{k,j}}\right| \mathcal{B}_j^n \xrightarrow[n\to+\infty]{p.s.} \xi,$$
où ξ est une variable aléatoire réelle (qu'on ne précisera pas).
Indice : remarquez que, pour $i \geq 2$,
$$\frac{C_{i,j}}{\left(\sum_{k=1}^{i-1} C_{k,j}\right)\left(\sum_{k=1}^{i} C_{k,j}\right)} = \frac{1}{\sum_{k=1}^{i-1} C_{k,j}} - \frac{1}{\sum_{k=1}^{i} C_{k,j}}.$$
5. En utilisant le lemme de Kronecker :
Si u_n est le terme général réel d'une série convergente, et si (b_n) est une suite croissante de réels positifs divergeant vers l'infini, alors :
$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^{n} b_i u_i = 0,$$
en déduire que
$$\left.\frac{\sum_{i=1}^{n-j} \varepsilon_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}}\right| \mathcal{B}_j^n \xrightarrow[n\to+\infty]{p.s.} 0,$$
et conclure.