

Sujet stage M2 Statistiques et applications

Intitulé :

Régression Beta par inférence bayésienne sur une très grande base de données et parallélisation des calculs.

Résumé :

La maladie d'Alzheimer est une pathologie neurodégénérative responsable en France de plus de 70% des démences. Son impact sur les fonctions cognitives est majeur : le Mini Mental State Examination (MMSE, score variant de 0 à 30) montre une décroissance annuelle moyenne de 3,4 points. En termes de traitements, si on sait que certaines molécules ont une certaine efficacité (inhibiteurs de l'acétylcholine estérase, antagonistes des récepteurs dits NMDA), les sujets atteints de la maladie ont aussi souvent d'autres types de traitement (seuls 15% des sujets n'ont aucun traitement) dont les effets ont peu été étudiés ni lorsqu'ils sont pris seuls, ni en association (antidépresseurs, anxiolytiques, ...).

Le but de l'étude est de d'estimer l'effet des traitements seuls ou en association sur l'évolution du MMSE. Les données viennent de la Base de données Nationale française Alzheimer, fondée par le plan national Alzheimer 2008 – 2012. Cette base vise à fournir des données épidémiologiques ainsi que des indicateurs d'activité et enregistre de nombreuses variables lors des actes médicaux effectués par les Centres Mémoire de Ressources et de Recherche. Sur une année, la base compte 220.000 patients (la moitié nés avant 1930), près de 70% de femmes, en moyenne 4 visites soit plus de 900.000 visites. Le MMSE moyen est à 18.

Parce que le MMSE ne prend que des valeurs discrètes entre 0 et 30, sa modélisation, une fois amené sur $(0,1)$, repose sur une régression Beta. Seule sa moyenne est modélisée et dans le prédicteur linéaire (fonction de lien logit) seront ajoutés les effets du temps, des traitements et leurs interactions. Des premiers résultats (sur une base Grand-Est) ont déjà été publiés (Blanc F. *et al.* Long-term cognitive outcome of Alzheimer's disease and dementia with Lewy bodies: dual disease is worse. *Alzheimer's Research & Therapy* 2017, 9:47). La complexité de la régression Beta et la taille de la base de données rend impossible l'estimation des paramètres du modèle sur un PC « classique » et en utilisant les logiciels statistiques habituel (R). Il est donc nécessaire de s'affranchir des outils habituels pour que l'estimation deviennent possible : écrire explicitement le mécanisme d'inférence, le choix de l'algorithme de mise à jour des chaînes de Markov, décider des possibilités de parallélisation des calculs, ...

Rôle de l'étudiant :

Dans le cadre de ce stage, l'étudiant aura à :

- si besoin, parfaire ses connaissances sur la régression Beta et l'inférence bayésienne ;
- écrire explicitement les étapes de l'estimation des paramètres dans la régression Beta ;
- argumenter du choix des algorithmes d'estimations ;
- proposer les grandes lignes de la parallélisation.

Son travail se fera en lien avec des chercheurs du laboratoire, mathématiciens et informaticiens, spécialistes en approches bayésienne, McMC et optimisation.

Durée du stage : 6 Mois

Lieu du stage : Laboratoire de Biostatistique et Informatique Médicale (iCube UMR CNRS 7357), Faculté de Médecine (locaux sur le site de l'Hôpital Civil).

Année 2017-2018