

Apprentissage statistique pour tester l'influence du microbiote intestinal sur la réponse au traitement par immunothérapie

Lieu du stage : Sorbonne Université (Jussieu), éventuellement TelecomParisTech (Palaiseau)

Début du stage : A partir d'avril 2020

Encadrement : Anna Bonnet (Sorbonne Université), Sylvain Le Corff (TelecomParisTech), Harry Sokol (APHP, Sorbonne Université)

De nombreux patients atteints d'un cancer du colon peuvent bénéficier de l'immunothérapie, un nouveau traitement qui est en train de révolutionner la thérapie anti-cancéreuse. Cependant, l'efficacité de l'immunothérapie varie fortement entre les individus et il est difficile de prévoir avant le début du traitement si la réponse d'un patient sera satisfaisante. La qualité de la réponse prend en compte deux critères fondamentaux : la progression de la maladie et la toxicité du traitement. Des études récentes suggèrent que le microbiote intestinal influence de façon majeure ces deux aspects de la réponse au traitement, mais les mécanismes sous-jacents sont encore peu connus. Le but du stage est :

- de déterminer si la composition du microbiote intestinal contribue à la réponse à l'immunothérapie chez des patients atteints d'un cancer du colon ;
- d'identifier parmi les données disponibles sur le microbiote les variables participant de façon significative à la qualité de la prédiction.

Les données cliniques, obtenues sur une soixantaine de patients, proviennent d'une collaboration entre les services d'oncologie (qui traite les patients et mesure leur réponse à l'immunothérapie) et d'hépatogastro-entérologie (qui fournit et analyse les échantillons de microbiote) de l'hôpital Saint-Antoine à Paris. Pour une soixantaine de patients, les variables explicatives sont issues des échantillons de microbiote prélevés avant le début du traitement. Les observations (catégorielles) obtenues pour chaque échantillon sont de deux types : efficacité du traitement et toxicité.

Une première partie du stage sera consacrée à l'application d'outils classiques de machine learning telles que les forêts aléatoires et de modèles statistiques simples (modèles linéaires et modèles linéaires généralisés). En fonction des résultats obtenus avec ces techniques standard, nous développerons de nouveaux algorithmes adaptés à l'étude de notre jeu de données, l'une des difficultés étant de trouver une méthode performante malgré le faible nombre d'observations disponibles. Une attention particulière sera portée à la quantification des erreurs d'estimation et de prédiction, sans laquelle les différents outils statistiques apportent peu de garanties. Dans ce but, le développement d'approches du type Monte Carlo (par chaînes de Markov) pour la sélection de variable offre des perspectives intéressantes lorsque de très nombreuses variables sont disponibles pour peu d'observations, [1].

[1] **Adaptive Monte Carlo for Bayesian Variable Selection in Regression Models**, D. Lamnissos et al., Journal of Computational and Graphical Statistics, Volume 22, 2013 - [Issue 3](#), p. 729-748

Profil recherché : Etudiant.e en M2 de statistique, machine learning, science des données ou équivalent, avec de solides connaissances en programmation (de préférence Python, éventuellement R).

Contact : anna.bonnet@upmc.fr