

Proposition de stage de Master 2 / Dernière année d'école Ingénieur en Biostatistique / Apprentissage statistique / Science de données ou Domaine similaire

Etablissement d'accueil : Equipe SISTM (Statistics In System biology and Translational Medicine), commune à l'INRIA et l'INSERM U1219, Bordeaux, France

Encadrement : Marta Avalos (enseignante-chercheur) en Biostatistique, Institut de Santé Publique d'Epidémiologie et Développement, Université de Bordeaux

En collaboration avec : Laurence Delhaes (PU-PH), Service de Parasitologie-Mycologie, CHU de Bordeaux et équipe "Remodelage bronchique" de l'Inserm U1045, Université de Bordeaux

Sujet : Analyse de données compositionnelles : comparaison de méthodes et application à l'étude du microbiote pulmonaire chez des patients asthmatiques

Au cours des dernières années, un intérêt croissant a été porté au *microbiote* (l'ensemble des bactéries, champignons et autres micro-organismes de l'organisme humain). En effet, les progrès en bioinformatique et en séquençage de la dernière génération ont permis des avancées importantes dans la compréhension du rôle du microbiote en santé. Le nombre de publications scientifiques sur le sujet a augmenté exponentiellement. Un grand nombre d'articles de vulgarisation scientifique ont suivi, témoignant de l'importance croissante accordée au sujet autant au niveau de la recherche que du grand public : « Le bonheur est dans la variété ». J Doré, D Ehrlich et al. *Cerveau & Psycho*, avril 2017 ; « Dossier : le microbiote respiratoire, kezaço ? » H Joubert. *BPCO Association*, avril 2018. En particulier, le microbiote a été associé à certaines pathologies, telles que des maladies inflammatoires de l'intestin, diabète, cancers, ou encore des pathologies respiratoires chroniques. Pour ces dernières, pourtant, pendant très longtemps, les voies aériennes inférieures étaient considérées comme stériles. De multiples questions concernant le rôle du microbiote respiratoire, ses interactions avec l'hôte ou les interactions entre les différentes communautés le constituant (bactérienne, fongique, virale,...) restent néanmoins en suspens.

Les données microbiote sont usuellement mesurées en tant qu'abondance relative des espèces, c'est-à-dire, il s'agit de données compositionnelles (CoDa pour *Compositional Data*), dont la somme vaut 1. Puisqu'une composante peut être déterminée à partir de la somme du reste de la composition, les composantes sont mathématiquement et statistiquement dépendantes. Cette structure complique l'analyse et ne permet pas d'effectuer des inférences valides à partir d'analyses statistiques standard, telles que la corrélation. Pourtant, de nombreuses techniques d'analyse de données, telle que l'analyse en composantes principales (ACP) ou l'inférence de réseaux d'interactions, reposent sur cette mesure.

Des niveaux de difficulté supplémentaires sont rajoutés à l'étude des données microbiote, lorsqu'on tient compte de la grande dimension des données provenant de séquençage à haut débit, d'une part, ou de l'hétérogénéité des données intégrant le microbiote (communautés bactérienne et fongique, notamment), d'autre part.

Le premier objectif du stage est la mise en œuvre et comparaison des méthodes d'analyse de données multidimensionnelles (telles que les ACP). Le deuxième objectif du stage est d'explorer les différents critères de sélection de modèles graphiques permettant de déterminer le réseau d'interaction le plus pertinent.

Les résultats de ces comparaisons permettront de guider le choix des méthodes à appliquer dans l'analyse des données du projet COBRA-ENV. Cette étude s'intéresse aux interactions entre les composantes bactériennes et fongiques du microbiote pulmonaire ainsi qu'environnemental chez des patients asthmatiques.

Le troisième objectif du stage est d'analyser les données du projet COBRA-ENV afin de répondre aux questions posées par les cliniciens.

Durée de stage : Entre 4 et 6 mois pendant l'année académique 2019-2020

Gratification de stage : Selon les taux en vigueur (~550 €/mois)

Adresse et contact :

Marta Avalos (marta.avalos-fernandez@inria.fr)

Institut de Santé Publique, Epidémiologie et Développement (ISPED) - Université de Bordeaux

146, Rue Léo-Saignat ; 33076 Bordeaux Cedex. Tél. : 05 57 57 15 34