

Offre de stage Master 2

Inférence statistique post-sélection de variables et impact de la prise d'antibiotiques sur la croissance des enfants dans les pays à bas revenus

Contexte de l'étude

La malnutrition infantile, caractérisée par un retard de croissance en terme de taille ou de poids, est une cause majeure de décès chez les enfants de moins de 5 ans à l'échelle mondiale. Les stratégies visant à améliorer la croissance des enfants se sont principalement concentrées sur la nutrition et la prévention de la diarrhée. L'observation des effets positifs des antibiotiques sur la croissance, initialement constatée chez les animaux d'élevage dans les années 1940, a soulevé l'hypothèse qu'une exposition continue aux microbes et aux agents pathogènes environnementaux peut entraver la croissance.

Chez les êtres humains, des liens ont été établis entre les infections, la malnutrition et la composition de la flore intestinale. Des études ont montré que l'utilisation d'antibiotiques peut favoriser la croissance chez des enfants malnutris, mais ces observations proviennent principalement de populations spécifiques et d'essais cliniques contrôlés (Gough et al. (2014)), où la prise d'antibiotiques est supervisée. Dans les pays à faible revenu, les maladies infectieuses, en particulier chez les enfants, sont fréquentes, et l'accès aux antibiotiques est mal régulé, ce qui suggère une exposition élevée des enfants à ces médicaments. Cependant, on ne dispose pas d'informations sur l'impact des antibiotiques sur la croissance des enfants "sains" en communautaire où la prise d'antibiotiques n'est pas contrôlée.

Ce stage se basera sur l'analyse de données de la cohorte multicentrique et internationale **BIRDY** (Bacterial Infections and antibiotic Resistant Diseases in Young children among low-income countries). Cette étude a pour objectif principal d'estimer l'incidence des infections bactériennes résistantes aux antibiotiques chez les enfants dans les pays en développement. Grâce à une surveillance à la fois passive et active, environ 4000 enfants ont été suivis en communautaire à partir de la naissance jusqu'à l'âge de 2 ans dans 3 pays (à Madagascar, au Cambodge et au Sénégal) et 2 sites par pays (rural et urbain). Les données concernant les infections présentées, la prise d'antibiotique ainsi que les mesures anthropométriques ont été collectées pendant tout le suivi.

D'un point de vue statistique, l'analyse des données d'épidémiologie passe généralement par une première étape de sélection des variables pertinentes, par exemple par méthode stepwise, régression lasso, ou l'utilisation de méthodes de machine learning comme les forêts aléatoires. Cette étape de sélection rend cependant difficile l'inférence statistique : les intervalles de confiance et les tests d'hypothèse réalisés sont incorrects lorsqu'ils sont réalisés sur les mêmes données que celles sur lesquelles a été réalisée la sélection de variables, et des approches adaptées doivent être alors développées (par exemple approches de type validation croisée, mais aussi des approches spécifiques aux modèles statistiques choisies, Lee et al. (2016), Mentch and Hooker (2016)).

Objectif du stage

Ce stage a pour objectif d'appliquer des méthodes d'inférence statistique valides malgré l'étape de sélection de variables, pour estimer l'impact de la prise d'antibiotiques sur la croissance des enfants en communautaire dans 3 pays à bas revenus (Madagascar, Sénégal et Cambodge). Dans ce cadre, plusieurs approches (validation croisée, régression lasso, forêts aléatoires, etc.) seront explorées et comparées aux méthodes communément utilisées en santé publique. Cette démarche statistique sera complétée par une réflexion sur la valorisation de ces méthodes en santé publique, pour aboutir à une publication scientifique à l'interface entre épidémiologie et biostatistique.

Profil recherché

- Master2 ou équivalent en statistiques appliquées, biostatistiques ou sciences des données
- Intérêt pour l'épidémiologie et la santé publique
- Connaissance du contexte des pays à bas revenus valorisée
- Aptitude de programmation avec le logiciel R ou Python
- Savoir rédiger en anglais

Encadrement

Le stage sera encadré par Félix Cheysson (chercheur au LAMA, Université Gustave Eiffel, Champs-sur-Marne) et le Dr Bich-Tram Huynh (chercheuse au sein de l'équipe EMEA, Institut Pasteur, Paris). L'étudiant sera hébergé environ 3 jours par semaine au LAMA et 1 jour par semaine dans l'équipe EMEA, pour une durée de 6 mois (date de début à la convenance selon la fin des cours du master). Gratification standard de stage (~614€ / mois).

Contacts

Merci d'envoyer votre candidature (CV et lettre de motivation) aux deux adresses suivantes, avec comme objet du mail *Candidature de stage M2* :

- felix.cheysson@univ-eiffel.fr
- bich-tram.huynh@pasteur.fr

Références bibliographiques

- Gough, E. K., Moodie, E. E., Prendergast, A. J., Johnson, S. M., Humphrey, J. H., Stoltzfus, R. J., Walker, A. S., Trehan, I., Gibb, D. M., Goto, R., Tahan, S., De Moraes, M. B., and Manges, A. R. (2014). The impact of antibiotics on growth in children in low and middle income countries: Systematic review and meta-analysis of randomised controlled trials. *BMJ (Online)*, 348(April):1–13.
- Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *Annals of Statistics*, 44(3):907–927, 1311.6238.
- Mentch, L. and Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *Journal of Machine Learning Research*, 17:1–41, 1404.6473.