

Offre de stage de Master 2 Statistique

Sujet

Extension d'une modélisation Poisson Log-Normal, appliquée à l'étude de l'émergence de maladies intestinales chez les jeunes ruminants

Contexte

Microbiote des jeunes ruminants. Chez les jeunes ruminants, outre son rôle dans la dégradation des aliments, le microbiote digestif joue un rôle crucial dans le développement de la physiologie intestinale et du système immunitaire. La colonisation initiale commence à la naissance et constitue une interaction à double sens entre l'hôte et les microbes. Chez les ruminants, il n'y a pas de transfert d'anticorps in utero et l'immunité passive est obtenue par la tétée. Pour permettre l'acquisition d'immunoglobulines, l'intestin du nouveau-né est perméable aux macromolécules, ce qui ouvre également la porte aux agents pathogènes. Pour cette raison et à cause de l'immaturation de leur microbiote, les jeunes ruminants sont vulnérables aux maladies pendant une période cruciale de leur développement. Par exemple, les agneaux diarrhéiques présentent des structures de population bactériennes différentes de celles des agneaux sains, ainsi qu'une richesse/diversité plus faible. A contrario, une population microbienne plus riche et plus diversifiée est souvent associée à une plus grande résilience pour maintenir la santé et les performances de l'animal ; qui est une caractéristique des jeunes agneaux élevés sous la mère. Cependant, la relation entre la diversité du microbiote et l'incidence de la diarrhée n'est à l'heure actuelle pas aussi évidente. De plus, la présence et la proximité avec la mère est une question majeure pour comprendre l'émergence de ce type de maladies. Dans ce contexte, la problématique se pose de la manière suivante : Pourquoi les agneaux élevés dans les mêmes conditions présentent-ils des prévalences différentes de diarrhée ?

Méthodologie statistique. Grâce à des approches modernes de génomiques, l'abondance de chaque espèce microbienne peut se déterminer pour un écosystème donné (ce qui donne lieu à du Big Data). Dans le cadre de cette problématique agronomique, l'objectif est de mettre en relation le statut de maladie de l'animal avec les compositions de son microbiote intestinal, et notamment en tenant compte de facteurs exogènes relatifs à l'environnement de l'animal. Dans ce cas, une modélisation adaptée pour ce genre de données et de problématiques, repose sur l'utilisation d'un modèle Poisson Log-Normal (PLN) permettant d'expliquer les abondances (multivariées) de différentes espèces microbiennes avec une couche de régression latente faisant intervenir les facteurs exogènes. Avec cette problématique, il devient important de pouvoir effectuer cette modélisation conjointement avec 1) une sélection de variables parmi un nombre assez grand de facteurs exogènes possibles et 2) potentiellement une couche de modélisation supplémentaire permettant d'expliquer le statut de maladie de l'animal.

Approche proposée. Pour cela, le travail consistera en partie à utiliser une méthodologie développée [2] dans l'équipe de recherche par le doctorant Jean Yves Kioye consistant à combiner une approche récente de sélection de variables [3] avec un l'approche PLN [1]. Des travaux de modélisation et d'implémentation, issus de questionnements autour de la donnée, sont à mener successivement pour répondre aux différents axes de ce sujet. A noter que la portée de ces travaux à vocation à être applicable ou généralisable à différents contextes d'applications (en termes méthodologiques).

Travail à effectuer

Parmi les tâches qui pourront être envisagées par le stagiaire, voici ci-dessous une liste des points (ordonnés chronologiquement) qui paraissent pertinents :

- prise en main d'une base de données qui sera mise à disposition,
- récupérer et comprendre le code et les approches existantes au sein de l'équipe de recherche,
- se former à utiliser un git et un serveur de calcul (via `ssh`),
- s'appropriier les éléments bibliographiques relatifs au sujet du stage,
- mettre au point une méthodologie et l'implémenter, et
- comparer la méthodologie proposée avec d'autres approches existantes.

A noter que ce stage pourrait donner lieu à une embauche sur un contrat à durée déterminée d'ingénieur, de sorte à poursuivre les travaux et implémenter la solution.

Pour candidater

Envoyez un mail à paul_marie.grollemund@uca.fr avant le 15 janvier avec les informations suivantes :

- CV,
- relevé de notes du M1, et
- indiquez vos disponibilités pour une visio.

Profil recherché

- Niveau équivalent Master 2 en mathématiques appliquées, orienté statistique ou probabilités.
- Connaissances en analyse de données et en modélisation.
- Une bonne maîtrise de l'environnement Linux sera un plus non négligeable.
- Une connaissance des domaines suivants sera un plus : modèle pour données de comptage, modèle hiérarchique, algorithme EM.
- Bonne maîtrise de *R* et/ou *python*.

Superviseurs

Paul-Marie Grollemund (paul_marie.grollemund@uca.fr) et Milka Popova (microbiologiste, milka.popova@inrae.fr)

Organisme d'accueil

Ce stage se fera à l'Université Clermont Auvergne au Laboratoire de Mathématiques Blaise Pascal, spécifiquement sur le site délocalisé de Aurillac.

Indemnisation

Taux légal : une gratification de 536 euros net par mois.

Références

- [1] Julien Chiquet, Mahendra Mariadassou, and Stéphane Robin. The poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *Frontiers in Ecology and Evolution*, 9 :588292, 2021.
- [2] Togo Jean Yves Kioye, Jocelyn Grollemund, Paul-Marie Chauvet, and Christophe Chassard. Sélection de variables par approximation de la norme l_0 dans un modèle poisson log-normal. *Journées de la statistique*, 2023.
- [3] Meadhbh O'Neill and Kevin Burke. Variable selection using a smooth information criterion for multi-parameter regression models. *arXiv preprint arXiv :2110.02643*, 2023.