

Nom du laboratoire d'accueil

UR 1404 INRA MaIAGE « Mathématiques et Informatiques Appliquées au Génome et à l'Environnement », Jouy-en-Josas.

Directeur : Sophie Schbath (MaIAGE)

Encadrants :

Estelle Kuhn (INRA, MaIAGE)

Charlotte Baey (Université de Lille, Laboratoire Paul Painlevé)

Sujet de stage : Procédures non asymptotiques de test des composantes de la variance dans les modèles à effets mixtes. Application à la caractérisation de la variabilité génotypiques des paramètres d'un modèle de culture chez *Arabidopsis thaliana* et *Brassica napus*.

Contexte : Les modèles à effets mixtes sont largement utilisés pour modéliser différents niveaux de variabilité pour des mesures répétées dans de nombreux domaines tels que la pharmacodynamique, l'écologie ou encore l'épidémiologie. Ces modèles tiennent compte à la fois des variabilités intra- et inter-individuelles en combinant deux types d'effets : d'une part les *effets fixes*, communs à tous les individus de la population, d'autre part les *effets aléatoires* qui varient d'un individu à l'autre. Les effets aléatoires sont des variables aléatoires latentes du modèle et ne sont donc pas observés. Modéliser les différents types d'effets est primordial, à la fois pour l'inférence des paramètres et pour pouvoir utiliser le modèle à des fins prédictives. Ainsi, distinguer parmi tous les effets ceux qui peuvent être modélisés par des effets fixes peut permettre de réduire le nombre de paramètres du modèle, et peut également aider à mieux identifier les processus, par exemple biologiques dans notre contexte, qui seraient à l'origine de la variabilité observée dans la population.

Du point de vue statistique, tester si des effets sont aléatoires au regard des données considérées revient à tester si la matrice de covariance associée à ces effets est nulle. La difficulté réside dans le fait que les valeurs des paramètres à tester sont sur la frontière de l'espace des paramètres. La loi asymptotique du test de rapport de vraisemblance a été identifiée [1]. Cependant, dans de nombreux cas pratiques, le nombre d'individus observés est petit, rendant la mise en oeuvre des résultats asymptotiques discutables.

Objectifs du stage :

- (i) proposer une procédure de test alternative non asymptotique.
- (ii) implémenter la méthode sur des données simulées pour comparer les performances avec le test asymptotique en utilisant un modèle de croissance logistique.
- (iii) implémenter la méthode sur des données réelles de *Arabidopsis thaliana* et *Brassica napus* en utilisant un modèle de culture.

Aspects mathématiques : Deux approches sont envisagées pour construire une procédure de test non asymptotique, d'une part une approche basée sur une version bootstrap du test de rapport de vraisemblance, d'autre part une approche basée sur la construction d'une statistique de test ajustée, incluant un terme correctif pour les petits échantillons s'inspirant des travaux de [2].

Compétences recherchées : formation niveau BAC+5 (Master 2 ou école d'ingénieurs), connaissance en statistiques théoriques et appliquées, ayant un fort intérêt pour les applications en sciences du vivant. La maîtrise d'un langage de programmation est indispensable.

Qualités personnelles : rigueur scientifique, curiosité intellectuelle, qualités de communication.

Modalités pratiques : le stage se déroulera à Jouy-en-Josas. La durée du stage sera de cinq ou six mois, entre février et septembre 2019. La gratification mensuelle est d'environ 540 euro. Le stage pourra éventuellement déboucher sur un sujet de thèse.

Contact : Estelle Kuhn estelle.kuhn@inra.fr

Références bibliographiques :

[1] C. Baey, P.H. Cournède, E. Kuhn. Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, Vol. 135, pp 107--122, 2019.

[2] A. C. Monti, M. Taniguchi. Adjustments for a class of tests under nonstandard conditions. *Statistica Sinica*, Vol. 28, pp 1437--1458, 2018.