



INRA

Institut National de la Recherche Agronomique

OFFRE DE STAGE EN STATISTIQUE niveau BAC+5 5 ou 6 mois de mars à septembre 2018

Unités d'accueil : Unité MaIAGE (Mathématiques et Informatique Appliquée du Génome à l'Environnement),
Centre INRA, 78352 Jouy en Josas

Encadrant : Estelle Kuhn (estelle.kuhn@jouy.inra.fr)

Titre : **Optimisation des temps d'observations de données de type current status data pour l'estimation dans des modèles de fragilité**

Discipline(s) scientifique(s) (en français) : **Statistique, modélisation, algorithmique.**

Présentation du projet (contexte, objectifs, méthodologies à mettre en œuvre) :

Les données recueillies en agronomie ou en épidémiologie sont souvent partiellement observées. Au cours de ce stage, on considère des données de type current status data : on dispose d'information sur l'état de l'individu au temps d'observation t . Par exemple, en agronomie, on s'intéresse à la date de floraison d'un plant de maïs. Le technicien récolte les données en plein champ au temps d'observation t . Il observe chaque plant et constate s'il a fleuri ou non. On sait donc si la date de floraison est antérieure ou postérieure au temps t , mais on ne connaît pas exactement cette date. Ce type de données se rencontre également en épidémiologie.

On souhaite par ailleurs analyser l'influence de covariables sur ces données. En analyse de survie, le modèle de Cox est très fréquemment utilisé et permet de modéliser la survenue de l'événement considéré en fonction de covariables. Plus précisément, le risque instantané de survenue de l'événement s'exprime comme produit d'une fonction de base dépendant du temps et d'une fonction de régression des covariables. Une hypothèse implicite à ce modèle est la proportionnalité des risques pour des groupes d'individus qui partagent les mêmes valeurs de covariables. Cependant de nombreux jeux de données réelles ne satisfont pas cette hypothèse du fait d'une forte hétérogénéité intrinsèque aux données. Pour relâcher l'hypothèse de proportionnalité des risques, on peut considérer des modèles de fragilité (frailty models). Ces modèles prennent en compte l'hétérogénéité présente dans les données via des effets aléatoires qui ne sont pas observés.

L'objectif du stage est de fournir un critère et un outil pour déterminer des temps d'observations optimaux pour l'estimation dans un modèle de fragilité à partir de données de type current status data. Les performances de la méthode seront évaluées numériquement sur données simulées et réelles de transition florale chez le maïs en collaboration avec Adrienne Ressayre et Christine Dillmann (UMR Génétique Quantitative et Evolution Le Moulon).

Modalités pratiques :

Le stage se déroulera à Jouy en Josas. La durée sera de 5 ou 6 mois, entre mars et septembre 2018. La gratification mensuelle est de 554,40 euro. Le stage pourra éventuellement se poursuivre par une thèse.

Compétences requises :

Formation : niveau Bac+5 (master ou école d'ingénieur), en statistique, avec un goût prononcé pour les applications en sciences du vivant.

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