

Modèle conjoint de survie pour l'analyse de données corrélées spatialement : application à l'étude de la propagation d'une maladie dans le contexte du palmier à huile

Que cela soit dans les domaines de la santé, de l'agronomie ou de la génétique, de nombreuses études permettent le suivi d'individus jusqu'à l'apparition ou non d'un événement d'intérêt, comme la mort ou l'apparition d'une maladie. L'analyse de ces données, appelées données de survie, permet de comprendre la relation entre l'apparition d'un événement et différentes variables explicatives. D'autre part, dans de nombreuses études, des variables d'intérêt mesurées à plusieurs pas de temps sur un même individu (données longitudinales) ainsi que les coordonnées géographiques des données sont disponibles. L'intégration de ce type d'information permet de modéliser la dépendance entre les individus appartenant à une même unité géographique, de comprendre l'influence des unités à proximité dans l'apparition de l'évènement, et d'étudier simultanément l'effet de variables explicatives sur la survie et les variables d'intérêts.

En statistique la prise en compte de la composante spatiale se traduit par l'introduction d'un effet aléatoire spatial dans le modèle d'analyse de survie (Banerjee et al., 2016). La structure de la matrice de variance-covariance associée permet de comprendre l'influence et les relations entre les unités géographiques. L'analyse simultanée de temps de survie et de mesures longitudinales se fait au travers d'une modélisation jointe (Brown and Ibrahim, 2003; Zhou et al., 2008).

Dans ce cadre, l'objectif de ce travail de recherche appliquée consistera à mettre en œuvre dans le cadre Bayésien des modèles conjoints de survie intégrant un effet spatial où la structure de matrice de variance-covariance se basera sur les coordonnées exactes des individus (Banerjee et al., 2003 ; Banerjee et al., 2016). Les modèles mis en œuvre permettront d'analyser les relations entre un caractère d'intérêt mesuré à plusieurs pas de temps et l'apparition de la maladie.

Les différentes approches développées seront appliquées dans un contexte d'amélioration génétique du palmier à huile sur des données issues d'essais génétiques mis en place par le CIRAD (Centre de coopération internationale en recherche agronomique pour le développement) et PalmElit, sa filiale pour la sélection et la production de semences.

Contacts encadrants : Marie Denis (marie.denis@cirad.fr)

Durée : 4 mois

Indemnité mensuelle : 550 euros

Thématiques : Analyse de survie, modèle spatial, méthodes Bayésiennes

Lieu du stage : Montpellier, CIRAD Lavalette, UMR AGAP

Banerjee, S., Wall, M. M., Carlin, B. P., Jan. 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostat* 4 (1), 123-142.
<http://dx.doi.org/10.1093/biostatistics/4.1.123>

Banerjee, S., 2016. . *Spatial Data Analysis*. *Annu Rev Public Health* 37, 47-60. 2016. <http://dx.doi.org/10.1146/annurev-publhealth-032315-021711>

Brown, E.R. and Ibrahim, J.G. (2003). Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics* 59, 686-693.

Zhou H, Lawson AB, Hebert JR, Slate EH, Hill EG., 2008. Joint spatial survival modeling for the age at diagnosis and the vital outcome of prostate cancer. *Statistics in medicine* 27 (18),3612-3628.
<http://dx.doi.org/10.1002/sim.3232>.