

Analyse de données d'épidémie de malaria par un modèle à fragilités univariées à corrélations spatiales

Ajmal Oodally, Estelle Kuhn, Klara Goethals, Luc Duchateau

La malaria est une maladie avec un taux de mortalité encore très élevé en Afrique sub-saharienne. La transmission se fait via un moustique, dont le développement et la reproduction sont favorisés par la présence de plans d'eau. Afin d'étudier l'influence des plans d'eau sur le taux de transmission, un jeu de données décrit dans [1] a été constitué dans le secteur d'un barrage hydroélectrique en Éthiopie. Il est constitué de temps d'infection par la malaria d'enfants répartis en villages, ainsi que des covariables comme l'âge, le sexe, la structure du toit et la distance au barrage. Ces données ont déjà été analysées par un modèle de fragilité structuré selon les villages. Cependant, la proximité entre les enfants n'a pas été prise en compte. L'aspect spatial de ces données est d'une importance primordiale, car la maladie se transmet d'un enfant à l'autre. Par conséquent nous proposons un modèle à fragilités univariées spatialement corrélées qui modélise la corrélation entre les enfants en fonction de la distance qui les sépare. La nouveauté de ce modèle est que la corrélation spatiale est introduite au niveau de l'enfant, i.e, chaque observation possède un terme de fragilité qui lui est propre. À notre connaissance, les modèles spatiaux existants considèrent plutôt des modélisations au niveau des régions, villes ou pays avec une fragilité qui est donc au niveau d'un groupe, i.e, plusieurs observations qui partagent un même terme de fragilité. Les paramètres du modèles sont estimés par maximum de vraisemblance en utilisant une version stochastique de l'algorithme Expectation Maximization combiné à une méthode Monte-Carlo par chaînes de Markov (SAEM-MCMC). Nous prouvons la convergence presque sûre de cet algorithme vers un point critique de la vraisemblance marginale. Les performances de l'estimateur sont évaluées sur des données simulées et les résultats obtenus sont très concluants. Pour l'analyse des données de malaria, deux structures de corrélations sont comparées suivant un critère basé sur le maximum de vraisemblance. L'interprétation biologique des résultats obtenus s'avère intéressante et permet de mieux comprendre l'incidence de la malaria chez ces enfants.

[1] Getachew, Y., Janssen, P., Yewhalaw, D., Speybroeck N. et Duchateau L. (2013). Coping with time and space in modelling malaria incidence: a comparison of survival and count regression models, *Statistics in Medicine*, 32, pp. 3224-3233.