

Méthode de vraisemblance par Monte Carlo pour un modèle de tessellation aléatoire

Offre de stage de 6 mois niveau BAC+5

Unité d'accueil : Unité MaIAGE (Mathématiques et Informatique Appliquées du Génome à L'Environnement), Centre INRA, Jouy-en-Josas (78)

Encadrant : Katarzyna Adamczyk (Katarzyna.Adamczyk@inra.fr)

En collaboration avec Radu Stoica, IECL (Institut Élie Cartan de Lorraine), Université de Lorraine, Nancy

Contexte La méthode de maximum de vraisemblance par Monte Carlo a été proposée par Geyer et Thompson [1] pour estimer les paramètres des distributions appartenant à la famille exponentielle, définies à une constante de normalisation près. Les auteurs proposent de remplacer la constante de normalisation par son approximation par Monte Carlo et de maximiser la vraisemblance approchée. Geyer [2] a présenté l'application de cette méthode aux processus ponctuels spatiaux, pour l'estimation des paramètres et la comparaison de modèles.

Le modèle de Gibbs de tessellation aléatoire en T^1 , proposé par Kiêu et al. [3], est un autre exemple du modèle spatial de la famille exponentielle non-normalisée. La figure (1) présente deux réalisations de ce modèle correspondant aux différents choix de statistiques canoniques. Un simulateur du modèle, basé sur un algorithme de Metropolis-Hastings-Green, a été implémenté dans une librairie C++ interfacée avec R [4]. La question d'inférence statistique pour ce modèle reste ouverte.

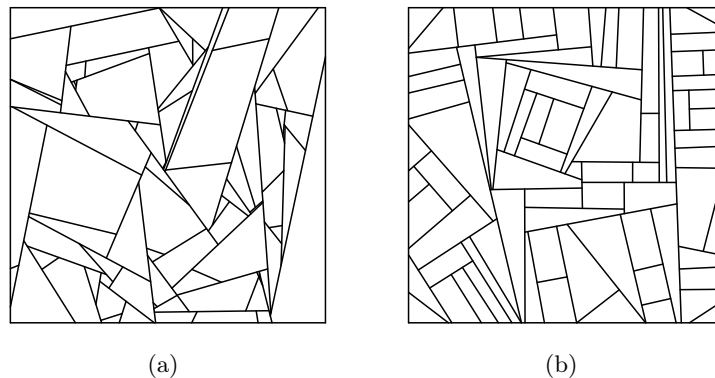


FIGURE 1 – Deux réalisations du modèle gibbsien de tessellation en T : (a) modèle complètement aléatoire, (b) modèle avec la pénalité pour les angles aigus entre des segments et pour la somme des carrés des aires des faces.

Déroulement de stage L'objectif de stage est d'appliquer la méthode de maximum de vraisemblance par Monte Carlo au modèle de tessellation en T , en développant en particulier des tests d'hypothèses pour les paramètres de ce modèle. Le stagiaire mettra en place le calcul des tests basés sur la vraisemblance, en s'inspirant de travaux de Geyer pour des processus ponctuels. Les outils programmés seront intégrés dans la librairie **RLiTe**. Pour déterminer leur capacité discriminante, les tests seront appliqués à des données simulées sous différents modèles de tessellation.

La suite de stage portera sur l'application du modèle de tessellation à la simulation des parcellaires agricoles. L'objectif de cette partie sera de sélectionner des variables à inclure dans le modèle pour obtenir des simulations réalistes. Le stagiaire déterminera une collection initiale de variables à partir de l'étude bibliographique. La sélection des variables sera faite selon une procédure ascendante, reposant sur les tests développés en première partie. Elle s'appuiera sur un jeu de données représentant un parcellaire agricole.

1. Une tessellation polygonale est dite une tessellation en T lorsque : 1) chaque sommet interne de la tessellation a trois arêtes incidentes 2) deux des trois arêtes incidentes sont alignées.

Profil recherché Nous recherchons un étudiant en statistique ou en mathématiques appliquées, familiarisé avec les méthodes MCMC et motivé par des applications (ici en agronomie). Des connaissances en statistique spatiale seraient un atout. La maîtrise du langage **R** est nécessaire, des compétences en programmation **C++** seront fortement utiles. Les connaissances relatives au modèle de tessellation devront être acquises pendant le stage.

Modalités pratiques Le stage commencera au mois de février/mars 2018 et durera 5 à 6 mois. La gratification mensuelle est de 554 euros, selon la grille officielle. Les personnes intéressées peuvent adresser leur CV à Katarzyna.Adamczyk@inra.fr.

Références

1. C.J. Geyer, E.A. Thompson (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Statist. Soc. B*, 54(3) :657-699.
2. O.E. Barndorff-Nielsen, W.S. Kendall, M.N.M.van Lieshout (1999). Stochastic geometry, likelihood and computation. *Chapman and Hall*
3. K. Kiêu, K. Adamczyk-Chauvat, H. Monod, R.S. Stoica (2013). A completely random T-tessellation model and Gibbsian extensions. *Spatial Statistics*, 6 :118-138.
4. <http://kien-kieu.github.io/lite>