

Stage M2 : Comparaison de modèles à blocs latents et réseaux écologiques

Sarah Ouadah (LPSM, Sorbonne Université), Marie Perrot-Dockès (MAP5, Université Paris Cité)
et François Massol (iEES, CNRS-Sorbonne Université)

LBM Les modèles à blocs latents [LBM] (Govaert & Nadif, 2003) sont des modèles de mélanges permettant d'établir des classifications simultanées des lignes et des colonnes d'un tableau de données. Ces modèles sont notamment utilisés en écologie pour déterminer des groupes de plantes, d'une part, et de pollinisateurs, d'autre part, à partir des visites observées de différentes espèces de pollinisateurs sur différentes espèces de plantes.

L'inférence statistique de ces modèles pose des problèmes spécifiques. La première raison en est que la répartition des plantes et des pollinisateurs dans leurs groupes respectifs est inconnue. La seconde est que la complexité de la structure de dépendance entre les différentes variables aléatoires impliquées dans le modèle (appartenance aux groupes, présence de visites) nécessite le recours à des approximations variationnelles (Blei & al, 2017; Latouche & al, 2012).

Le premier objectif de ce stage est la compréhension des modèles à blocs latents et de leur inférence au moyen d'approximations variationnelles bayésiennes.

Comparaison Les réseaux écologiques plante - pollinisateur peuvent être collectés dans différentes zones géographiques (De Manincor, & al, 2020). Les écologues sont alors amenés à se demander si les réseaux deux à deux ont la même structure de groupes, autrement dit s'ils sont décrits par un même modèle à blocs latents.

Le second objectif de ce stage est de proposer un test statistique permettant de comparer deux réseaux régis par des modèles à blocs latents. On considèrera plusieurs hypothèses sous lesquelles un ou plusieurs paramètres des LBM diffèrent. On estimera les Bayes Factor (Kass & al, 1995) correspondant qui permettront de choisir l'hypothèse la plus pertinente. Cette partie s'inspirera de notions et principes de preuve utilisés dans Latouche & al, 2018.

Le troisième objectif de ce stage est d'explorer d'autres modèles considérés en écologie, i.e ajustés aux réseaux observés, pour in fine construire des tests de comparaison de réseaux dans l'esprit de ce qui aura été fait précédemment.

Contact sarah.ouadah@sorbonne-universite.fr, marie.perrot-dockees@u-paris.fr

Références

- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518), 859-877.
- De Manincor, N., Hautekeete, N., Piquot, Y., Schatz, B., Vanappelghem, C., & Massol, F. (2020). Does phenology explain plant-pollinator interactions at different latitudes? An assessment of its explanatory power in plant-hoverfly networks in French calcareous grasslands. *Oikos*, 129(5), 753-765.
- Govaert, G., & Nadif, M. (2003). Clustering with block mixture models. *Pattern Recognition*, 36(2), 463-473.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430), 773-795.
- Latouche, P., Birmele, E., & Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1), 93-115.
- Latouche, P., Robin, S., & Ouadah, S. (2018). Goodness of fit of logistic regression models for random graphs. *Journal of Computational and Graphical Statistics*, 27(1), 98-109.