ID de Contribution: **22**   Type: **Non spécifié**

Polya urn models for multivariate species abundance data: Properties and application

jeudi 20 juin 2024 15:00 (1 heure)

This talk focuses on models for multivariate count data, with emphasis on species abundance data. Two approches emerge in this framework: the Poisson log-normal (PLN) and the Tree Dirichlet multinomial (TDM) models. The first uses a latent gaussian vector to model dependencies between species whereas the second models dependencies directly on observed abundances. The TDM model makes the assumption that the total abundance is fixed, and is then often used for microbiome datasets since the sequencing depth (in RNA seq) varies from one observation to another, leading to a total abundance that is not really interpretable. We propose to generalize TDM models in two ways: by relaxing the fixed total abundance assumption and by using Polya distribution instead of Dirichlet multinomial. This family of models corresponds to Polya urn models with a random number of draws and will be named Polya splitting distributions. In a first part I will present the probabilistic properties of such models, with focus on marginals and probabilistic graphical model. Then it will be shown that these models emerge as stationary distributions of multivariate birth death process under simple parametric assumption on birth-detah rates. These assumptions are related to the neutral theory of biodiversity that assumes no biological interaction between species. Finally the statistical aspects of Polya splitting models will be presented: the regression framework, the inference, the consideration of a partition tree structure and two applications on real data.

Orateur: PEYHARDI, Jean (Institut Montpelliérain Alexander Grothendieck (IMAG))

Classification de Session: Exposé long

Classification de thématique: Exposés longs