
DREW NEISH, University of Guelph

Clustering Using Mixtures of Dirichlet-multinomial Regression Models

Compositional analysis of the human microbiome is made possible through advanced sequencing techniques, where the output consists of abundance of different bacterial taxa in each microbiome sample. Previously, a Dirichlet-multinomial mixture model has been used for modelling such microbial metagenomic data, where each mixture component represents distinct meta-communities that show similar biota compositions. However, identifying the association of environmental/biological covariates with abundance in different meta-communities remains an important problem. Here, a mixture of Dirichlet-multinomial regression models is proposed and illustrated. These models allow for a probabilistic investigation of the relationship between bacterial abundance and biological/environmental covariates within each inferred meta-community.