



APPRENTISSAGE STATISTIQUE POUR L'ANALYSE DE DONNÉES DE COMPTAGE MULTIVARIÉS

Les données de comptage multivariés résultent de l'observation conjointe de plusieurs variables prenant des valeurs entières (0, 1, 2, ...) parmi les individus d'une population.

Leur étude nécessite une réflexion approfondie tant sur le choix d'un modèle probabiliste adapté que sur le développement et l'implémentation de nouvelles méthodes statistiques pour l'ajustement des modèles aux données, un enjeu majeur étant de faire un choix de modèle pertinent, qui permette de développer des méthodes statistique puissantes ayant une complexité de calcul raisonnable pour une application à des jeux de données de grande taille.

Parmi les modèles probabilistes les plus utilisés pour le traitement statistique des données de comptage multivariées, on trouve le modèle multinomial et le modèle de Poisson multivarié. Le modèle multinomial, extension multivarié du modèle binomial, présente certaines restrictions qui limitent son utilisation à des données spécifiques ou la valeur de chacune des marges de la variable multivariée est bornée, avec des contraintes fortes sur leur variabilité et sur le domaine des corrélations possibles entre les marges.

Le modèle de Poisson multivarié permet quant à lui de prendre en charge des variables de comptage non bornées à priori et de modéliser avec plus de flexibilité les corrélations entre les différentes marges.

L'objectif de cette thèse est de s'appuyer sur les progrès récents réalisés sur le modèle probabiliste représenté par la loi de Poisson multivariée pour proposer d'abord des techniques d'ajustement statistique de ce modèle aux données dans toutes les configurations possibles des paramètres qui définissent le modèle et d'évaluer les performances de ces techniques et ensuite utiliser des techniques d'ajustement de modèles paramétriques de régression lorsque la variable réponse est multivariée et distribuée suivant une loi de Poisson multivariée conditionnellement à des covariables.