
Test non-paramétrique de comparaison d'échantillon basé sur les méthodes à noyaux: application au Single-Cell RNA sequencing

Anthony Ozier-Lafontaine* , Bertrand Michel¹, and Franck Picard

¹Ecole Centrale De Nantes (ECN) – Ecole Centrale de Nantes, École Centrale de Nantes – France

Résumé

Le single-cell RNA sequencing (scRNAseq) est une technique de séquençage génétique à haut débit. Les données générées contiennent beaucoup d'observations en grande dimension, elles sont aussi sparses (beaucoup de zéros) et hétérogènes. Ces propriétés nécessitent le développement de méthodes statistiques adaptées. Mon objectif est de mettre à disposition des bioinformaticiens des tests de comparaison d'échantillons capables de comparer des populations de cellules séquençées par cette technologie. Je m'intéresse particulièrement à des tests à noyaux non-paramétriques, dont le test Maximum Mean Discrepancy (MMD) est le représentant le plus connu. La méthode que j'utilise, basée sur l'analyse discriminante de Fisher (KFDA), se distingue du MMD en prenant en compte la structure de covariance des données. En contrepartie, la statistique KFDA est plus coûteuse que la statistique MMD. L'un des aspects fondamentaux de mon projet consiste donc à combiner des méthodes d'approximation et de factorisation matricielles pour réduire le coût de calcul de la statistique KFDA en préservant les performances du test. J'utilise par exemple des approches de Nystrom et des méthodes par quantization, ces approches ne sont pas spécifiques au calcul de la statistique KFDA, et peuvent aussi être appliquées à d'autres statistiques de test, comme la statistique MMD.

*Intervenant