

Évaluation des hypothèses relatives à l'optimisation des moindres carrés ordinaires dans un contexte de déconvolution de données multi-omiques

vendredi 16 mai 2025 11:40 (20 minutes)

Dans un contexte de données multi-omiques de benchmark relatives au cancer du pancreas, l'objectif est d'inférer l'hétérogénéité cellulaire. Les données de benchmark sont des données de référence obtenues par expérimentation in vitro. Dans le cas présent, les données sont de nature multi-omique. En l'occurrence, transcriptomique et méthylomique. L'inférence de l'hétérogénéité cellulaire, ou déconvolution cellulaire, est un moyen d'estimer des proportions de types cellulaires dans un bulk de cellules à partir de données de référence.

En théorie, un bulk est la résultante de la somme d'un signal et d'un bruit. Le signal correspond aux proportions des types cellulaires attendues dans le tissu bulk, multipliées par des valeurs de référence spécifiques à chaque type cellulaire. Dans ce contexte, les individus statistiques considérés sont des gènes pour lesquels des données de méthylome et de transcriptome ont été récoltées.

Les méthodes de déconvolution sont, pour la plupart, basées sur les moindres carrés ordinaires. Or, les moindres carrés ordinaires supposent que les données respectent des hypothèses de normalité, d'homoscédasticité et de centralité du bruit ainsi que d'indépendance des individus.

Toutes les hypothèses sont remises en question par la nature intrinsèque des données de méthylome et de transcriptome. Cette présentation porte sur l'hypothèse d'homoscédasticité. L'idée est de voir la relation de dépendance, si elle existe, entre l'expression d'un gène et les variations du bruit observé dans nos données expérimentales. Aussi sera-t-il possible d'observer la part de signal et la part de bruit dans un bulk de cellules. Observer une relation de dépendance entre le gène et les variations du bruit témoigneraient de l'hétéroscédasticité du bruit. Comprendre cette relation entre le gène et le bruit et comprendre la part de bruit dans un bulk de cellules mènerait au développement de méthodes de déconvolution plus adaptées aux données omiques.

Thématiques

Biostatistiques ; déconvolution ; simulations

Authors: CAUSEUR, David (Institut Agro Rennes Angers - departement of statistic IRMAR); DEFFAINS, Elouan (Institut agro Rennes Angers); BARBOT, Hugo (IRMAR)

Orateur: DEFFAINS, Elouan (Institut agro Rennes Angers)

Classification de Session: Matin