

Les problèmes statistiques de la mise en relation du génome et du comportement

Depuis quelques années les chercheurs arrivent à obtenir de larges quantités de données génétiques (de l'ordre du million) et comportementales sur des populations assez grandes de sujets (plusieurs milliers). Souvent le but de ces études est de mettre en rapport les données de génétiques avec le comportement de façon à trouver les corrélats génétiques de certains comportements (e.g., propensions aux comportements à risques, ou addictifs) ou de certaines populations (e.g., prédisposition à la schizophrénie ou à la maladie d'Alzheimer). Ces larges ensembles de données qualitatives se doivent d'être étudiés avec des techniques appropriées. Ici nous proposons d'utiliser une généralisation de la méthode des moindres carrés partiels et de l'analyse des correspondances appelée « partial least square correspondence analysis » (PLSCA). Nous présentons la méthode avec un petits exemple réel puis avec un exemple plus large montrant les relations entre les données génomiques, les mesure neuropsychologiques et la prédisposition à la maladie d'Alzheimer