

Quelques modèles statistiques pour les analyses génomiques

S. Robin

1er cours. Nous considérerons tout d'abord l'analyse des puces CGH ("comparative genomic hybridization") qui sont utilisées pour détecter des aberrations chromosomiques. Nous présenterons des modèles de segmentations et traiterons de leur inférence aussi bien d'un point de vue algorithmique que statistique. Nous aborderons également les problèmes de sélection de modèles induits par ces analyses.

1er TD. Un premier TD sera consacré à l'analyse de quelques profiles CGH.

2ème cours. Nous considérerons ensuite l'analyse des puces à "haute densité" qui pose souvent des problèmes de classification des sondes. Les modèles de mélange fournissent un moyen naturel d'aborder de telles questions de classification non supervisée. Ces modèles permettent également de prendre en compte une dépendance spatiale.

3ème cours. Les problèmes des tests multiples se posent dans l'analyse des données issues de la plupart des technologies à "haut débit". On souhaite typiquement contrôler l'erreur de première espèce afin d'éviter un trop grand nombre de faux positifs. Nous rappellerons les procédures classiques telles que Bonferroni ou FDR. Nous aborderons aussi l'estimation du FDR local qui peut être posée en termes de mélange.

2nd TD. Un second TD sera consacré à l'implémentation de l'algorithme E-M en vue de l'inférence des modèles de mélange.

4ème cours. Les réseaux constituent une façon naturelle de décrire les interactions entre les éléments présents dans la cellule. La compréhension de l'organisation et de la topologie de ces réseaux requière de définir des modèles de graphes aléatoires réalistes. Nous présenterons quelques uns de ces modèles et discuterons de leur inférence.