

Graphe de recombinaison ancestral contraint : modélisation et inférence

Problématique : Au centre de mes préoccupations est la modélisation de la transmission des séquences d'ADN, en tenant compte de plusieurs contraintes rencontrées dans la réalité. Tout d'abord, le processus de méiose (mélange du matériel génétique hérité des grands-parents) induit une dépendance spatiale entre des nucléotides rapprochés sur le chromosome (hérités par blocs). Ensuite, les principes biologiques de l'hérédité causent une dépendance due à la parenté des individus. Pour un échantillon donné, ces relations s'expriment à travers le graphe de recombinaison ancestral [1] qui généralise le processus de coalescence introduit par [2] : un ensemble d'arbres aléatoires corrélés, chacun décrivant l'historique génétique d'un point sur le génome. Plusieurs phénomènes réels compliquent sa structure, par exemple lorsque les membres de l'échantillon proviennent d'un mélange de populations originellement isolées («admixture»).

Le grand mérite du processus de coalescence est qu'il peut être associé à une grande classe de modèles de population à générations discrètes qui évoluent du passé vers le présent, en tant que processus limite, lorsque la population est suffisamment grande par rapport à la taille d'échantillon. Ce processus à temps continu est une approximation du processus markovien à temps discret sous-jacent qui décrit l'ascendance d'une génération à la précédente. Une des suppositions de ce modèle est que les parents sont choisis au hasard, de façon indépendante, ce qui donne une topologie particulière au graphe. Dans la littérature, certaines généralisations permettent de poser des contraintes sur la topologie du graphe de recombinaison ancestral, par exemple en incluant, comme dans [3], des modèles de migration. Le projet actuel, dans un esprit semblable, propose de développer une nouvelle classe de processus où la contrainte, sous forme d'une loi *a priori*, est placée sur un sous graphe de la population. Ainsi, le processus à générations discrètes est perturbé : l'échangeabilité des membres de la population est perdue et les résultats limite ne sont plus valides dans la formulation courante. Il s'agirait d'une généralisation du processus de coalescence sous des suppositions additionnelles sur les arêtes du graphe.

Objectifs : Un premier objectif est de modéliser le graphe de recombinaison ancestral en combinant le processus ancestral (où les choix des parents sont des événements indépendants) et un processus markovien à temps discret qui est contraint par la structure d'un sous graphe, par exemple lorsque les relations familiales entre les membres de l'échantillon sont spécifiées (pedigree). Ensuite, on se propose de trouver sous quelles conditions la théorie limite peut toujours être appliquée. De même, mon intention est de développer une nouvelle méthode d'inférence à partir de séquences d'ADN génotypées, qui pourra servir par exemple en cartographie génétique, et de perfectionner les méthodes d'optimisation qui y sont rattachées. Finalement, un objectif plus lointain serait de développer des plans d'échantillonnage permettant d'exploiter la méthodologie à son plein potentiel.

Contribution : L'utilisation de l'information disponible, par exemple un pedigree marqué par des événements de migration, peut possiblement améliorer l'estimation de généalogies dans des contextes difficiles comme en présence d'«admixture». La statistique génétique sur des données familiales redevient prisée aujourd'hui dans le contexte de maladies reliées à des variants génétiques rares. Par ailleurs, tous ces concepts s'inscrivent dans la thématique plus large de l'analyse des graphes aléatoires ; pour ce qui est des champs d'application, en plus de la génétique humaine, la phylogénétique et l'épidémiologie sont envisagés.

- [1] Griffiths, R. C. (1997). An ancestral recombination graph. *Progress in Population Genetics and Human Evolution*, 257-270.
- [2] Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13(3) : 235–248.
- [3] Wakeley, J. (2001). The coalescent in an island model of population subdivision with variation among demes. *Theoretical Population Biology*, 59(2) : 133-144.