
Modèles stochastiques de réseaux de gènes

Ulysse Herbach^{*1,2,3}

¹Laboratoire de Biologie et Modélisation de la Cellule (LBMC) – CNRS : UMR5239, Institut national de la recherche agronomique (INRA) : UR5239, Université Claude Bernard - Lyon I (UCBL), École Normale Supérieure (ENS) - Lyon – ENS de Lyon 46 allée d'Italie 69364 LYON Cedex 07, France

²Inria team DRACULA (Inria Grenoble Rhône-Alpes / Institut Camille Jordan) – Université Claude Bernard - Lyon I (UCBL), CNRS : UMR5534, CNRS : UMR5208, INRIA, Institut Camille Jordan – Institut Camille Jordan Université Claude Bernard Lyon 1 43 boulevard du 11 novembre 1918 69622 Villeurbanne cedex France, France

³Institut Camille Jordan (ICJ) – Institut National des Sciences Appliquées [INSA], Ecole Centrale de Lyon, Université Claude Bernard - Lyon I (UCBL), CNRS : UMR5208, Université Jean Monnet - Saint-Etienne – Bât. Jean Braconnier n 101 43 Bd du 11 novembre 1918 69622 VILLEURBANNE CEDEX, France

Résumé

L'expression génétique des cellules a longtemps été observable uniquement via des quantités moyennes mesurées sur des populations. L'arrivée récente des techniques "single-cell" permet aujourd'hui d'observer des niveaux d'ARN et de protéines dans des cellules individuelles : il s'avère que même dans une population de génome identique, la variabilité entre les cellules est souvent très forte et diffère clairement de la simple perturbation autour d'une valeur moyenne.

Ce constat incite à aborder sous un nouvel angle le problème statistique de l'inférence de réseaux de régulation. En effet, il met en défaut les méthodes classiques basées sur des modèles linéaires gaussiens, mais offre en même temps l'opportunité d'utiliser une description physiquement fondamentalement stochastique de l'expression des gènes. Une telle approche pourrait permettre, en exploitant pleinement l'information contenue dans les données de cellules uniques, d'obtenir des modèles statistiques plus robustes, dans un contexte où le nombre de paramètres est généralement bien supérieur au nombre de données.

Je m'intéresserai ici à la construction de modèles stochastiques de réseaux de gènes qui puissent décrire les observations à partir d'arguments biologiques plutôt qu'empiriques. Mathématiquement, il s'agira de voir les gènes comme un système de particules en interaction formant un processus de Markov déterministe par morceaux. J'exposerai ensuite une méthode simple pour obtenir, à partir de la loi stationnaire de ce processus, un modèle statistique intéressant de type Markov caché.

*Intervenant