

Quatrième Journée Normandie-Mathématiques
13 juin 2012 – INSA de Rouen

**Chaînes de Markov régulées
pour l'analyse de séquences biologiques**

Nicolas VERGNE

Laboratoire de Mathématiques Raphaël Salem

Résumé

Nous présentons le développement, en vue de l'analyse statistique des séquences d'ADN, de nouveaux modèles permettant de prendre en compte l'hétérogénéité de ces séquences : les chaînes de Markov régulées (DMM pour Drifting Markov Model).

Afin d'éviter l'homogénéité supposé par les modèles de Markov et de Markov cachés, nous permettons à la matrice de transition de varier du début à la fin de la séquence. Pour chaque position, nous avons une matrice de transition différente. Ces modèles peuvent être vus comme une alternative mais aussi comme un outil complémentaire aux modèles de Markov cachés. Nous avons considéré des dérives polynomiales ainsi que des dérives par splines polynomiales. Nous avons estimé nos modèles de multiples manières puis évalué la qualité de ces estimateurs avant de les utiliser en vue d'applications telle la recherche de mots exceptionnels. Nous avons mis en oeuvre le software DRIMM, dédié à l'estimation de nos modèles.

Référence :

VERGNE Nicolas, Drifting Markov Models with Polynomial Drift and Applications to DNA Sequences. *Statistical Applications in Genetics Molecular Biology* (2008), **7**(1)