1

Méthodes mathématiques pour l'analyse des réseaux biologiques

Session organisée par Franck PICARD et Sophie SCHBATH

Les technologies à haut débit en biologie, comme celle des microarrays permettent désormais la collecte d'une grande quantité d'information concernant l'expression des gènes et les interactions entre protéines. Il est clair que les fonctions biologiques complexes régissant le fonctionnement d'une cellule sont portées par ces milliers de gènes dont les produits interagissent sous la forme de réseaux (réseaux d'interaction, de régulation). Par conséquent les nouveaux objectifs des méthodes mathématiques appliquées en biologie moléculaire sont doubles : comment reconstruire un réseau, et comment décrire un réseau après construction ?

Dans cette session, nous proposons 5 exposés traitant de ce nouveau problème qu'est l'analyse mathématiques des réseaux biologiques. Dans un premier temps, les exposés seront centrés autour de l'inférence de réseaux, c'est-à-dire la reconstruction des relations de dépendance ou des interactions existant entre les gènes. Dans un deuxième temps nous nous intéresserons à l'utilisation d'un réseau de gènes partiellement connu pour aider à la description de l'évolution temporelle de l'expression des gènes. Enfin le dernier exposé sera consacré à généralisation du modèle d'Erdos Rényi pour les graphes aléatoires. Ce modèle permet de mieux décrire la topologie modulaire des graphes observés, avec une application aux réseaux de réactions enzymatiques.

Adresse de l'organisateur :

Franck PICARD
Laboratoire Statistique et Génome - UMR CNRS 8071
Tour Evry 2 - 523 place des Terrasses
91000 Evry, France
E-mail : `picard@inapg.fr`

Apprentissage d'un modèle dynamique non linéaire de réseau de régulation.

par Pierre Geurts, Ming Quach et **Florence d'Alché Buc**

Nous nous intéressons à un modèle non linéaire de la dynamique des réseaux de régulation fondé sur les équations de Michaelis-Menten où gènes et protéines interviennent. Nous montrons comment, à structure fixée, l'apprentissage des paramètres peut être opéré par une variante de l'algorithme Unscented Kalman Filter. Lorsque la structure est partiellement connue, cet aprentissage est couplé à une recherche gloutonnne de cette structure.

Dans ce cadre, nous proposons différentes fonctions de coût incluant des donnaissances a priori sur la nature des interactions. Ce travail est illustré sur des réseaux synthétiques bio-réalistes.

Adresses :

Pierre Geurts
LaMI UMR 8042 Université d'Evry - CNRS, Genopole
Tour Evry 2 - 523 place des Terrasses
91000 Evry, France
E-mail : pgeurts@lami.univ-evry.fr

Ming Quach
LaMI UMR 8042 Université d'Evry - CNRS, Genopole
Tour Evry 2 - 523 place des Terrasses
91000 Evry, France
E-mail : mquach@lami.univ-evry.fr

Florence d'Alché Buc
LaMI UMR 8042 Université d'Evry - CNRS, Genopole
Tour Evry 2 - 523 place des Terrasses
91000 Evry, France
E-mail : dalche@lami.univ-evry.fr
<http://www.lami.univ-evry.fr/~dalche>

Inférence statistique sur les graphes

par Kevin Bleakley et **Gérard Biau**

Le problème de l'inférence, ou reconstruction, sur les graphes consiste à prédire les relations éventuelles entre un ensemble de points donnés supposés former les sommets d'un graphe dont on aurait perdu les arêtes. En motivant notre approche par des considérations biologiques, nous proposons tout d'abord un modèle mathématique original permettant de modéliser une structure graphe évoluant dans un espace de Banach séparable. Nous définissons ensuite formellement le problème de la reconstruction, en montrant plus particulièrement comment ce dernier peut être appréhendé en termes d'apprentissage statistique supervisé.

Mots-clés — Inférence sur les graphes, réseaux d'accélération, apprentissage statistique, Vapnik-Chervonenkis, mesures faiblement convergentes.

Adresses :

Kevin Bleakley
Institut de Mathématiques et de Modélisation de Montpellier – UMR CNRS 5149
Equipe de Probabilités et Statistique
Université Montpellier II
CC 51 – Place Eugène Bataillon,
34095 Montpellier Cedex 5, France
E-mail : `bleakley@math.univ-montp2.fr`
<`http://www.math.univ-montp2.fr/~bleakley`>

Gérard Biau
Institut de Mathématiques et de Modélisation de Montpellier – UMR CNRS 5149
Equipe de Probabilités et Statistique
Université Montpellier II
CC 51 – Place Eugène Bataillon,
34095 Montpellier Cedex 5, France
E-mail : `biau@math.univ-montp2.fr`
<`http://www.math.univ-montp2.fr/~biau` >

Inférence de Réseaux de Régulation à partir de profils temporels : estimation d'un graphe de dépendance sous un modèle autorégressif de dimension p très supérieure au nombre de mesures n.

par **Sophie Lebre**

L'inférence de réseaux de régulation de gènes constitue un enjeu majeur de la génomique. Nous proposons ici d'inférer un graphe de dépendance à partir de profils temporels d'expression de gènes. Nous considérons un processus stochastique multidimensionnel $(X_i(t))_{\{1 \leq i \leq p; 0 \leq t \leq n\}}$ décrivant le niveau d'expression de p gènes sur n instants successifs. Sous l'hypothèse d'un modèle autorégressif multidimensionnel, la matrice de transition du modèle définit un graphe $\mathcal{G}$ d'indépendances conditionnelles entre chaque variable $X_i(t)$ et $X_j(t-1)$ sachant l'ensemble des variables $\{X_j(t-1)\}_{1 \leq j \leq p}$ observées au temps précédent. Néanmoins, le nombre de mesures n est très inférieur au nombre p de gènes observés et les méthodes classiques ne permettent pas l'estimation de ce modèle.

Nous proposons de réduire la dimension du problème et d'estimer la topologie du graphe $\mathcal{G}$ en l'approchant par le graphe $\mathcal{G}^k$ d'indépendance conditionnelle d'ordre k $(k \geq 1)$.

Adresse :

Sophie Lebre
Laboratoire Statistique et Génome - UMR CNRS 8071
Tour Evry 2 - 523 place des Terrasses
91000 Evry, France
E-mail : lebre@genopole.cnrs.fr

Integrated Markov models for clustering gene expression data

par Florence Forbes et **Matthieu Vignes**

A major challenge in bioinformatics is to reveal interactions between components of living organisms and to discover the corresponding networks responsible for their biological complexity. In this framework, clustering of genes into groups sharing common characteristics is a useful exploratory technique. A wide range of clustering algorithms has been proposed to analyse gene expression data. But most of them consider genes as independent, do not treat gene expressions as individual or include relevant information on gene interactions a posteriori. We propose a probabilistic model that has the advantage to take into account individual features (e.g. expression) and pairwise data (e.g interaction information from biological networks) simultaneously. Our model is based on hidden Markov random fields models. Parametric probability distributions account for the distribution of individual data for each gene. Data on pairs are included through a graph where the nodes stand for the genes and the edges are weighted according to pair data, for instance in order to reflect distance or similarity measures between genes. This model has many interesting features. It leads to various possible statistical criteria to select automatically the number of clusters. It is also able to incorporate many types of data. It is flexible in the sens that its generalization to include missing data, that often occur when dealing with expression data, is straightforward. Its extension to overlapping clustering methods, to deal with more realistic situations where genes can belong to many groups at the same time, can also be considered. This approach was illustrated and validated on simulated data as well as on yeast expression data combined with pathways neighbourhoods.

Adresses :

Florence Forbes
INRIA Rhone-Alpes, Equipe MISTIS
655 avenue de l'Europe - Montbonnot
38334 Saint Ismier Cedex - France
E-mail : `florence.forbes@inrialpes.fr`

Matthieu Vignes
INRIA Rhone-Alpes, Equipe MISTIS
655 avenue de l'Europe - Montbonnot
38334 Saint Ismier Cedex - France
E-mail : `matthieu.vignes@inrialpes.fr`

Un modèle de mélange pour les graphes aléatoires

par Jean-Jacques Daudin, Franck Picard et **Stéphane Robin**

The Erdos-Rényi (ER) model is probably the oldest and simplest model for random graphs. It is simple and many explicit expressions for its average and asymptotic properties are known. However it does not fit well many real-word networks.

One reason is that the vertices of a network are often structured in a priori unknown clusters (functionally related proteins, social communities, etc.) with different connectivity properties. We define a generalization of ER named the Erdos-Rényi mixture for graphs (ERMG). We derive some of its theoretical properties and propose a pseudo-EM algorithm to estimate its parameters. We apply this modeling to uncover the modular structure of a network of enzymatic reactions.

Adresses :

Jean-Jacques DAUDIN
INA P-G, Département OMIP
16, rue Claude Bernard,
75231 Paris cedex 0.5
E-mail : `daudin@inapg.fr`

Franck PICARD
Laboratoire Statistique et Génome - UMR CNRS 8071
Tour Evry 2 - 523 place des Terrasses
91000 Evry, France
E-mail : `picard@inapg.fr`

Stéphane ROBIN
INA P-G, Département OMIP
16, rue Claude Bernard,
75231 Paris cedex 0.5
E-mail : `robin@inapg.fr`