

Programme détaillé YSP 2017

Accueil (9h15) : Benjamin Guedj (Inria, président du groupe "Jeunes" de la SFdS)

Ouverture de YSP5 (9h30): Gérard Biau (UMPC, IUF, président de la SFdS)

Session 1 (9h45-10h45) : Détection de ruptures (Modérateur: Charlotte Dion)

Exposé 1 : Thibault Allart (Ubisoft, CNAM, LSTA Paris 6) - Analyse des données longitudinales issues des jeux vidéo.

Historiquement issues des études cliniques où elles restaient cantonnées à des échantillons de petite taille, les données longitudinales sont aujourd'hui de plus en plus nombreuses. Des traces de navigation internet des utilisateurs d'e-commerce, aux objets connectés, en passant par nos parcours de santé enregistrés sur nos cartes vitales; leur particularité est qu'elles sont désormais de grande taille. Ces grands échantillons permettent d'envisager des modèles plus complexes, tels que des coefficients variationnels en temps. En s'appuyant sur la théorie des processus de comptage, nous présenterons un modèle de survie adapté à ces données. Nous montrerons que, contrairement au modèle de Cox, l'estimation de ce modèle peut être effectuée "en ligne", offrant dès lors la possibilité d'utiliser des algorithmes de type descente de gradient stochastique, plus performants sur ces échantillons de grande taille. Nous illustrerons cette approche sur des simulations en nous comparant aux méthodes existantes. L'ensemble de l'exposé se fera sur le thème des jeux vidéo et nous présenterons une mise en application industrielle visant à améliorer le design des jeux vidéo.

Exposé 2 : Loïc Schwallier (AgroParisTech/INRA) - Détection de ruptures au sein de la structure d'un modèle graphique arborescent dans un cadre bayésien.

Nous nous plaçons dans le cadre de la détection de ruptures au sein d'une série temporelle multivariée. La distribution des observations est donnée par un modèle graphique dont la structure et les paramètres sont soumis à de brusques changements au cours du temps. En considérant que les structures possibles sont des arbres, nous montrons que l'inférence du modèle peut être effectuée de manière exacte et efficace dans un cadre bayésien.

Pause (10h45-11h20)

Session 2 (11h20-13h) : Matrices aléatoires (Modérateur: Laure Sansonnet)

Tutoriel : Benjamin Groux (Versailles et Lille 1) - Matrices aléatoires : généralités et interactions.

Dans ce tutoriel, on présentera les principaux modèles et les résultats fondamentaux de la théorie des matrices aléatoires, ainsi que quelques-unes des nombreuses interactions de cette théorie. On donnera dans ce tutoriel tous les éléments nécessaires à la compréhension des trois exposés de la session.

Exposé 1 : Alkéos Michaïl (MAP5) - Perturbations by random matrices.

We provide a perturbative expansion for the empirical spectral distribution of a Hermitian matrix with large size perturbed by a random matrix with small operator norm whose entries in the eigenvector basis of the first one are independent with a variance profile. We prove that, depending on the order of magnitude of the perturbation, several regimes can appear (called

perturbative and semi-perturbative regimes): the leading terms of the expansion are either related to free probability theory or to the one-dimensional Gaussian free field.

Exposé 2 : Jean Lafond (Telecom ParisTech) Low Rank Matrix Completion with Exponential Family Noise

The matrix completion problem consists in reconstructing a matrix from a sample of entries, possibly observed with noise. A popular class of estimators, known as nuclear norm penalized estimators, are based on minimizing the sum of a data fitting term and a nuclear norm penalization.

In this talk, we consider the case where the noise distribution belongs to the exponential family and is sub-exponential and study the nuclear norm penalized loglikelihood estimator.

We prove that we can extend the results previously obtained for this estimator in the Gaussian and noiseless settings.

In particular, we show that the quadratic reconstruction error is upper bounded with high probability and that the upper bound is minimax optimal up to a logarithmic factor.

Exposé 3 : Fanny Augeri (Institut Mathématiques de Toulouse) : Problèmes de grandes déviations en matrices aléatoires

Le phénomène de répulsions des valeurs propres observé dans de nombreux modèles de matrices aléatoires pose de nombreux obstacles à la compréhension des grandes déviations du spectre. Les outils classiques développés dans le cadre de variables indépendantes n'étant plus directement disponibles, les principaux principes de grandes déviations connus sont obtenus dans le contexte des modèles intégrables (ensembles Gaussiens classiques, et plus généralement beta-ensembles), ou bien pour des modèles de matrices aléatoires qui présentent un phénomène de queues lourdes.

Session 3 (14h30-15h45) : Processus stochastiques (Modérateur: Erwan Scornet)

Tutoriel : Adelaïde Olivier (Paris 11) - Quelques processus stochastiques qui se rencontrent dans la nature.

Nous présenterons certains processus stochastiques classiques, aussi bien en temps discret (X_0, \dots, X_n, \dots) qu'en temps continu $(X_t, t \geq 0)$. Nous étudierons notamment les chaînes de Markov ; les diffusions avec pour principal représentant le mouvement brownien, utilisé aussi bien en biologie qu'en finance. La modélisation de la dynamique des populations sera également abordée à travers le processus de Galton-Watson.

Exposé 1: Julien Chevallier (Université de Cergy-Pontoise) - Méthode d'approximation de champ-moyen : couplage à la Sznitman sur un exemple jouet.

Le but des approches de champ-moyen est d'étudier un système de n particules en interaction lorsque n tend vers l'infini. La méthode de preuve introduite par Sznitman en 1991 est basée sur un couplage entre les n particules en interaction du système et n particules indépendantes bien choisies. Nous verrons les grandes lignes de cette méthode sur un exemple très simple (mouvement Brownien, processus d'Ornstein-Uhlenbeck) et le résultat théorique ainsi obtenu sera illustré par des simulations.

Exposé 2: Olga Lopusanschi (LPMA) - Le drift d'aire dans un processus limite: exemple de construction et calcul.

En topologie uniforme, le passage à la limite « efface » une partie des informations sur la suite des processus qu'on considère, ce qui empêche souvent, par exemple, d'approcher la solution d'une équation différentielle par une suite de solutions d'équations différentielles.

Introduite par Terry Lyons dans les années '90, la théorie de chemins rugueux résout ce problème en « enrichissant » un chemin ordinaire de plusieurs niveaux, en fonction de la régularité de celui-ci. Notamment, dans le cas de la convergence des trajectoires vers un mouvement brownien, on ne retient que deux niveaux : les incréments du chemin et l'aire qu'il accumule. C'est au niveau de cette aire (l'aire de Lévy) que la différence entre les limites de deux processus, invisible dans le cadre de la topologie uniforme, peut apparaître sous forme d'un drift non nul.

On verra que, lorsqu'il s'agit de convergence de chaînes de Markov sur graphes périodiques, on peut donner une expression explicite pour ce drift. On présentera un modèle pour illustrer sa construction et, si le temps le permet, des méthodes qui permettent de le calculer.

Pause (15h45-16h15)

Session 4 (16h15 - 17h30) : Clustering (*Modérateur: Vincent Brault*)

Tutoriel : Charlotte Laclau (LIG – Grenoble) Introduction au clustering.

Exposé 1 : Rawya Zreik (Paris 1) - The Stochastic Topic Block Model.

Due to the significant increase of communications between individuals via social media (Facebook, Twitter, LinkedIn) or electronic formats (email, web, e- publication) in the past two decades, network analysis has become a unavoidable discipline. Many random graph models have been proposed to extract information from networks based on person-to-person links only, without taking into account information on the contents. In this work we introduce the stochastic topic block model (STBM), a probabilistic model for networks with textual edges. We address here the problem of discovering meaningful clusters of vertices that are coherent from both the network interactions and the text contents. A classification variational expectation-maximization (C-DEM) algorithm is proposed to perform inference.

Exposé 2 : Le LI (INRIA Lille, Université d'Angers) - Pac Bayesian Online Clustering.

We address the online clustering problem. When faced with high frequency streams of data, clustering raises theoretical and algorithmic pitfalls. Working under a sparsity assumption, a new online clustering algorithm is introduced. Our procedure relies on the PAC-Bayesian approach, allowing for a dynamic (i.e., time-dependent) estimation of the number of clusters. Its theoretical merits are supported by sparsity regret bounds, and an RJMCMC-flavored implementation called Paco is proposed along with numerical experiments to assess its potential.

Présentation d'Opération Postes (17h40-17h55) : tout sur les campagnes MCF, CR, ATER, ...

Présentation des derniers chiffres de parité dans la recherche (18h-18h15) : Laurence Broze pour l'association Femmes et Mathématiques.