

Programme des journées Statistiques et Santé

Jeudi 10 et vendredi 11 octobre 2019

Conservatoire national des Arts et Métiers (CNAM)
292 rue Saint-Martin
75003 Paris

Amphithéâtre Jean-Baptiste Say

Version du 30/09/2019

GDR Statistiques et Santé
Société Française de Biométrie
Groupe Biopharmacie de la SFDS

Programme du 10 octobre 2019

8h30 - 9h10 Accueil des participants

9h10 - 9h30 N. Savy Présentation du GDR S&S
M. Hocine Présentation de la SFB
E. Pham Présentation du groupe Biopharmacie de la SFDS
M. Hocine Présentation des journées

9h30 – 10h30 **Conférencier invité : Guillem RIGAILL**
LaMME & IPS2, INRA.

Changepoint detection in the presence of outliers and constraints.

Abstract: One approach to detecting multiple changepoints is by minimising an objective function. In some cases, dynamic programming can be used to minimise this function. The standard dynamic programming algorithm use a recursion that conditions on the time of the last changepoint. I will present an alternative recursion that conditions on the value of the last segment parameter. This algorithm is in some cases much faster. It can exactly minimize the cost (i) in situations where the loss function is not convex and (ii) in situations where there are dependencies in parameters associated with successive segments. I will present an application of this algorithm to the detection of changes in the presence of outliers. For this application, I will explain how we can consistently estimate the number of changepoints using the biweight loss function.

10h30 – 11h00 Pause Café

11h00 – 11h30 **JL. Foulley**
IMAG, Université de Montpellier.

A propos d'une formule magique de conversion de la p-value en facteur de Bayes.

11h30 - 12h00 **V. Laizet**
Laboratoire Servier

Nouveau challenge des essais cliniques.

12h00 – 12h30 **C. Castel**
Santé publique France, Direction Appui, Traitements et Analyses de données (DATA) et
Université Paris-Est.

Modèle Markovien pour estimer l'incidence du VIH en France.

12h30 – 13h30 Buffet

13h30 – 14h00 Visite de la chapelle du musée

14h00 – 14h30 **L. Villain**
Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, INRIA Bordeaux
Sud Ouest, SISTM team, Vaccine Research Institute, Créteil.

Détermination de la taille d'échantillon nécessaire afin d'identifier une signature prédictive de la survie de Gliomes.

14h30 – 15h00 **M. Antigny**
Laboratoires Servier.

Des approches bayésiennes pour la prédiction d'événements en essais cliniques pour des études de survie.

15h00 – 15h30 **SL. Rincourt**
Université Paris-Saclay, Université de Paris-Sud, UVSQ, CESP, INSERM U1018
Oncostat et Service de Biostatistique et d'Epidémiologie, Institut Gustave Roussy.

Modélisation de la structure génomique tumorale du cancer du sein par analyse factorielle parcimonieuse non paramétrique.

15h30 – 16h00 Pause Café

16h00 – 17h00 **Conférencier invité : Emmanuel BACRY**
Directeur de Recherche au CNRS, Université Paris-Dauphine
Directeur Scientifique du Health Data Hub
Directeur projets Data/Santé Ecole Polytechnique

IA et Système National des Données de Santé - Le Health Data Hub.

Description : Le Système National des Données de Santé (SNDS) historique est une des plus grosses base de données médico-administratives du monde. Elle regroupe les données de remboursements de santé (Carte Vitale) de plus de 65 millions de personnes et pèse plus de 200To. Bien que ne contenant aucune données cliniques, c'est une base d'une grande richesse dont l'analyse peut avoir un impact sociétal important. Dans un premier temps, je présenterai le partenariat de recherche entre l'Ecole Polytechnique et la Caisse Nationale d'Assurance Maladie (CNAM) en charge du SNDS et certains résultats en découlant. Dans un second temps, je présenterai le Health Data Hub nouveau portail national du SNDS instauré par la nouvelle loi « Ma santé 2022 » (votée en juillet dernier) qui élargit le SNDS à toutes les données de santé issues de soins partiellement remboursés par l'assurance maladie (donc quasiment l'ensemble de toutes les données de santé françaises).

17h00 – 18h00 **Conférencier invité : Niel HENS**
Interuniversity Institute of Biostatistics and statistical Bioinformatics, Data Science Institute, Hasselt University, Hasselt, and Centre for Health Economic Research and Modelling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium.

The statistical analysis of serological and social contact data to inform infectious disease modelling.

Abstract: The analysis of infectious disease data dates its origin back to the work of Daniel Bernoulli in 1760 for estimating the effect of eradicating smallpox from the population. Many of the methods developed within this setting have resulted in important contributions to the general field of statistics. In this talk I will give several examples of state-of-the-art statistical methodology for the analysis of cross-sectional serological data, i.e. cross sectionally collected pathogen-specific antibody test results. I will also discuss some recent advances on the analysis of social contact data, i.e. data on social encounters deemed to be good markers of those contacts that are responsible for the transmission of

respiratory pathogens. The statistical methodology used, is diverse and well linked to both mathematical as well as computational concepts.

18h00 – 19h30 Assemblée Générale du groupe Biopharmacie de la SFDS
Réunion du bureau du GDR Statistiques et Santé

19h30 – 19h45 Visite de la bibliothèque du Cnam

20h00 Restaurant

Programme du 11 octobre 2019

8h00 – 9h00 **Conférencier invité : Catherine MATIAS**
CNRS - Sorbonne Université et Université de Paris.

Classification pour des données d'interaction dynamiques

Résumé : Nous aborderons 2 types de données d'interaction dynamiques : les séquences de graphes et les flots de liens. Pour chacun de ces types, je présenterai des modèles à blocs stochastiques qui permettent de faire de la classification des individus et illustrerait ces approches sur des données réelles.

9h00 – 10h00 **Conférencier invité : Clémence LEYRAT**
Department of Medical Statistics & Cancer Survival Group, London School of Hygiene and Tropical Medicine, United Kingdom.

Emulation d'essais cliniques à partir de données observationnelles: promesses et enjeux méthodologiques

Résumé : Bien que les essais randomisés soient considérés comme le design de choix pour évaluer l'effet causal de traitements ou interventions, il est parfois impossible de conduire de tels essais, pour des questions éthiques ou logistiques. Les essais cliniques peuvent aussi manquer de pragmatisme, incitant les autorités de santé à réclamer, en complément des essais randomisés, des évaluations en contexte réel. Ces évaluations utilisent de plus en plus de données observationnelles, issues de registres ou collectées en routine en soins primaires et secondaires. Cependant, de nombreux biais peuvent être présents lors de l'analyse de ces données, en particulier les biais de confusion, de sélection ou de temps immortel. Afin de limiter ces biais, un cadre théorique pour l'émulation d'essais cliniques à partir de données observationnelles a été récemment proposé. Une des manières d'émuler un essai est de « cloner » les patients à l'entrée de l'étude et d'allouer chacun des clones à un bras de l'essai. Les patients sont ensuite suivis au cours du temps et censurés lorsque leur traitement dévie du protocole (e.g. lorsqu'ils reçoivent le traitement dans le bras contrôle). Afin de corriger le biais lié à cette censure informative, l'analyse implique l'estimation de poids de censure et l'utilisation de modèles de régression pondérés, sur lesquels repose la validité des estimations obtenues. L'objectif de cette présentation est de discuter les principes et atouts de cette approche à l'aide d'une application utilisant des données du registre des cancers anglais, visant à évaluer les bénéfices de la chirurgie chez des patients âgés atteints de cancer du poumon. Cet exemple illustre également les enjeux méthodologiques et les difficultés statistiques posés par cette approche. En particulier, nous

montrons l'intérêt de modèles de régression flexibles du taux de mortalité, ainsi que des forêts aléatoires pour l'estimation des poids de censure. Enfin, nous présenterons différents outils permettant de s'assurer de la validité interne de l'essai émulé et discuterons la comparabilité de ces résultats avec ceux obtenus à partir d'essais randomisés.

10h00 – 10h30 Pause Café

10h30 – 11h00 **B. Cuer**

Institut du Cancer Montpellier, Université de Montpellier, - Plateforme Nationale Qualité de Vie et Cancer

Modèles conjoints à risques concurrents

11h00 – 11h30 **A. Meddis**

Institut Curie, PSL Research University, INSERM, U900

Un test pour la taille informative de grappes

11h30 – 12h00 **A. Chatton**

UMR INSERM 1246 SPHERE, Université de Nantes, Université de Tours et IDBC-A2COM, Nantes, France

G-computation et Inverse-Probability-Weighting en analyse de survie

12h00 – 13h30 Buffet

13h30 – 14h00 Assemblée Générale de la SFB

14h00 – 15h00 **Conférencier invité : Antoine CHAMBAZ**

MAP5 (UMR CNRS 8145), Université de Paris

Définition et estimation d'une méta-cible pour la prédiction d'épisodes hypotensifs en unité de soin aigus par monitoring continu

Résumé: Généralement défini comme une chute de la pression artérielle moyenne sous 65mmHg pendant au moins 5 minutes, un épisode hypotensif (EH) fait partie des événements délétères les plus fréquents en réanimation. Il est par ailleurs établi que les EHs sont (positivement) associés avec diverses issues indésirables chez les patients en réanimation. Ainsi, la prédiction fiable d'EHs présente un intérêt majeur car elle permettrait de mieux prévenir leur occurrence et, surtout, faciliterait la personnalisation des traitements.

Dans ce contexte, nous développons un outil de prédiction des EHs par "machine learning", qui exploite données cliniques, signaux biologiques et historique de traitement. Spécifiquement, nous ajustons notre algorithme de prédiction sur la base de données MIMIC-III (MIMIC-III Waveform Database Matched Subset version 1.0), qui contient les informations relatives aux séjours de 10,282 patients en réanimation.

Nous définissons un méta-objectif statistique qui se prête à la prédiction d'EH à tout moment du séjour en réanimation. Sa définition est adaptée à une situation (future) où le monitoring sera quasi-systématique, alors qu'il est à ce jour souvent haché. Nous élaborons une procédure statistique pour son estimation. Les performances sont évaluées en termes d'aire sous la courbe (AUC) et de score de Brier.

Je présenterai tout cela succinctement.

Ce travail s'inscrit dans le projet ACTERRÉA. Il est le fruit d'une collaboration avec Ményssa Cherifa, Romain Pirracchio et Matthieu Resche-Rigon

- 15h00 – 15h30 **L. Faucheux**
Equipe ECSTRRA, UMR 1153, Inserm, Paris, université Paris Diderot
- Données manquantes et classification non supervisée : regrouper les partitions après imputation multiple avec un clustering par consensus*
- 15h30 – 16h00 **L. Manitchoko**
Biostatistique, Biomathématique, Pharmaco-épidémiologie et Maladies Infectieuses (B2PHI) / Inserm / Université Paris-Saclay / Institut Pasteur
- Comparaison de méthodes d'estimation de l'effet d'une exposition médicamenteuse sur le risque d'évènement indésirable.*
- 16h00 – 16h30 **J. Seurat**
Université de Paris, IAME, INSERM
- Protocoles robustes dans les études longitudinales*
- 16h30 – 17h00 **M. Roux**
Laboratoire HIFIH, UPRES 3859, SFR 4208, Université d'Angers
- Etude de la significativité de multiples transformations d'une même variable d'intérêt dans le modèle de Cox et contrôle du FWER*
- 17h00 Clotûre des journées

Liste des posters

- M. Chion** Institut de Recherche Mathématique Avancée, UMR7501, CNRS – Université de Strasbourg et Laboratoire de Spectrométrie de Masse BioOrganique, Institut Pluridisciplinaire Hubert Curien, UMR7178, CNRS-Université de Strasbourg.
- Imputation multiple et prise en compte de l'incertitude pour les données de protéomique quantitative.*
- T. Delespierre** Système de surveillance syndromique
- Y. Dubuy** UMR INSERM 1246 SPHERE, Université de Nantes, Université de Tours.
- Validation of a Guttman errors-based method to detect response shift at an individual level.*
- G. Guillot** International Prevention Research Institute.
- Significativité des études d'interaction toxicologiques et pharmacologiques.*
- P. Lu** Sorbonne Université, Inserm, CNRS, Institut du cerveau et la moelle (ICM), AP-HP - Hôpital Pitié-Salpêtrière et INRIA Paris, ARAMIS équipe-projet.
- Modèles de survie multi-niveaux pour la prédiction de Alzheimer.*
- C. Moreau** Service d'Hépatogastroentérologie, Centre Hospitalier Universitaire d'Angers.
- Étude de l'impact de la progression de la fibrose hépatique chez les patients atteints de NAFLD fibrose hépatique.*