Statistiques appliquées à la biologie et à la santé

Journée thématique

1 février 2018,
Laboratoire de Mathématiques et Applications
UMR 7348 Université de Poitiers—CNRS (France)

Lors de cette journée thématique, différents modèles et méthodes statistiques seront présentés en lien avec des motivations provenant de la biologie et de la médecine et avec des applications à ces domaines en perspective. Les étudiants de Licence et Master de mathématiques sont les bienvenus.

Organisteurs Pierre-Yves LOUIS (LMA, Univ. Poitiers) ; Stéphanie RAGOT (CIC/CHU de Poitiers, Univ. Poitiers) ; Yousri SLAOUI (LMA, Univ. Poitiers)

Ci-après, l'ordre est celui des exposés de la journée.

1 R. Azais (inria Nancy)

Arbres de Galton-Watson conditionnés par la taille ou la hauteur : estimation via les modèles limites

Les données arborescentes occupent une place centrale en biologie puisqu'elles permettent de modéliser des situations variées : plantes, réseaux sanguins, relations de parenté entre êtres vivants... Après avoir donné quelques motivations applicatives, je présenterai une stratégie d'estimation pour ces données lorsqu'elles sont issues d'un modèle de Galton-Watson conditionné par la taille ou la hauteur. Les résultats seront illustrés par des simulations numériques.

2 C. Proust-Lima (INSERM Bordeaux)

Dynamic modelling of multiple domains involved in Alzheimer's disease: two approaches based on multivariate latent processes

collaboration avec Bachirou Taddé, Viviane Philipps
INSERM U1219, Université de Bordeaux, 33076 Bordeaux, France

Alzheimer's disease, the most frequent dementia in the elderly, is characterized by multiple progressive impairments in the brain structure and in clinical functions such as cognitive functioning and functional disability. Until recently, these domains were mostly studied independently while they are fundamentally inter-related in the degradation process towards dementia. We propose two statistical approaches to jointly model the dynamics of the multivariate domains involved in Alzheimer's disease. In both approaches, a domain is defined as a latent process for which measures of one or several markers, possibly non Gaussian, are available at discrete visits in a cohort. In the first approach, the main objective is to understand the link between the dimensions and the diagnosis of dementia. We thus propose a joint model in which the trajectories of the latent processes are described through a multivariate linear mixed model. Rather than considering the associated time to dementia as in standard

joint models, we assume dementia diagnosis corresponds to the passing above a covariate-specific threshold of a pathological process modeled as a combination of the domain-specific latent processes. This definition captures the clinical complexity of dementia diagnosis but also benefits from an inference via Maximum Likelihood which does not suffer from the usual complications of joint models estimation. The model and the estimation procedure can also handle competing clinical endpoints, such as the competing death in Alzheimer's disease. The method is illustrated on a large French population-based cohort of cerebral aging in which we study the clinical manifestations (cognitive functioning, physical dependency and depressive symptoms) in link with repeated clinical diagnoses of dementia and death. One limit of this approach is that the link between processes is only captured by correlations. In a second approach, we aim to model the dynamic influences between processes to understand the mechanisms underlying the dementia process. We define for this a dynamic causal model in discrete time based on both the linear mixed model theory to capture the correlation within a dimension and equations of difference to capture the temporal influences between dimensions. Parameters are estimated in the maximum likelihood framework enjoying a closed form for the likelihood. As causal relationships fundamentally lie in the continuous time framework, we evaluate the impact of the time discretization in simulations. The model is then applied to the data of the Alzheimer's Disease Neuro-imaging Initiative. Three longitudinal general domains (cerebral anatomy, cognitive ability and functional autonomy) are analyzed and their causal structure is contrasted between different clinical stages of Alzheimer's disease.

3 V. Audigier (CNAM Paris)

Multiple imputation for multilevel data with continuous and binary variables

Vincent Audigier (1), Ian R. White (2;3), Shahab Jolani (4), Thomas P. A. Debray (5;6), Matteo Quartagno (7), James Carpenter (7), Stef van Buuren (8) and Matthieu Resche-Rigon (9;10;11)

Individual participant data (IPD) meta-analysis is often considered to be the gold standard method for systematic reviews. The aim is to consider several studies, sharing the same outcome, to obtain better inference than could be obtained from any one study. However, studies typically differ in their data collection and availability of confounders typically varies. Consequently, by merging the studies, systematically missing data, i.e. missing for all individuals in a study, could be introduced. In addition, missing data (called sporadically missing) can occur within each study.

Multiple imputation (MI) is a common strategy to overcome the missing data issue. The imputation model used can be an explicit joint model (JM), specifying the distribution of all variables, or it can be defined only by conditional densities (fully conditional specification, FCS). The choice of the imputation model is a task crucial but difficult a priori.

We investigate MI methods to overcome systematically and sporadically missing data in a multilevel setting, such as meta-analysis, in the context of mixed data (continuous and binary). The methods compared are JM imputation of clustered data proposed by Quartagno and Carpenter (2016), FCS using generalized mixed models proposed by Jolani et al. (2015), FCS using a two-stage meta-analysis estimation procedure (Resche-Rigon and White, 2016).

First, methods are compared from a methodological point of view and through a simulation study. The study highlights the benefit to use such methods compared to reference multiple imputation methods for multilevel missing data. However, this work also shows that performances need to be nuanced according to the missing data pattern, the multilevel structure and the type of missing variables. Then, the MI methods are applied to an IPD meta-analysis in cardiovascular disease consisting of 28 observational cohorts in which systematically missing and sporadically missing data occur. Finally, practical recommendations are provided.

1-Cedric, Equipe MSDMA, Conservatoire National des Arts et Métiers, Paris
2-MRC Biostatistics Unit, Cambridge Institute of Public Health, U.K
3-MRC Clinical Trials Unit at UCL, London, U.K
4-Department of Methodology and Statistics, School of Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands
5-Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
6-Cochrane Netherlands, University Medical Center Utrecht, Utrecht, The Netherlands
7-Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, U.K.
8-Department of Statistics, TNO Prevention and Health, Leiden, The Netherlands

9-Service de Biostatistique et Information Médicale, Hôpital Saint-Louis, AP-HP, Paris, France
10-Université Paris Diderot - Paris 7, Sorbonne Paris Cité, UMR-S 1153, Paris, France
11-INSERM, UMR 1153, Equipe ECSTRA, Hôpital Saint-Louis, Paris, France

4 A. Ounajim (Prismatics, CHU de Poitiers et LMA)

L'utilisation des algorithmes d'apprentissage automatique pour une prise en charge personnalisée des patients douloureux chronique

Failed back surgery syndrome is a clinical pathology in which patients present with a set of symptoms encountered after they have had one or more technically, anatomically successful surgical procedures on the lumbar spine for correcting their disc related pathology. The Principal symptom of this pathology is a persistent recurrent pain mainly in the region of the lower back and legs that is generally resistant to physiotherapy and pharmacological treatment. An alternative proposed treatment to FBSS patients is Spinal Cord Stimulation which is becoming a widely used treatment for a number of pain conditions, and it is frequently considered as a last resort pain management option when conservative or less invasive techniques have proven ineffective. While research on SCS is growing, the SCS success rates are at best modest. It is clear that substantial variation exists in the degree of benefit obtained from SCS, and the procedure does not come without risks; thus focused patient selection is becoming very important. The current method used in forecasting who may benefit most from SCS consists of a 5 to 10 days trial period which requires an invasive surgical procedure that may lead to complications such as electrode migration, dural puncture during electrode placement and/or infection. In order to propose an alternative to this trial procedure, we used available data to develop and test eight machine learning binary classification algorithms which can be used as screening tools of SCS efficacy.