Journal de la Société Française de Statistique
Vol. 155 No. 1 (2014)

Editorial du numéro spécial "Données longitudinales quantitatives, événementielles, incomplètement observées"

Title: Editorial of the Special Issue "Longitudinal and incompletely observed data of quantitative measurements and events"

Laurent Bordes[1] et Daniel Commenges[2]

Ce numéro spécial du Journal de la Société Française de la Statistique intitulé *Données longitudinales quantitatives, événementielles, incomplètement observées* est consacré à un thème large qui regroupe l'analyse de données quantitative et événementielles. Le mot clé ici est *longitudinal*, ce qui signifie que nous disposons d'observations indexées par le temps, qui permettent d'estimer la dynamique des phénomènes d'intérêt. Il faut rajouter *événementielles* car l'expression *données longitudinales* a longtemps été synonyme de données longitudinales quantitatives, également connues sous le nom de *mesures répétées*. Les données événementielles sont aussi de nature longitudinale dans le sens où elles sont également indexées par le temps. Le cas le plus classique est celui des données de survie, qui peut être généralisé par la considération d'événements ou d'états multiples. Pour recueillir les deux types de données il faut des *études longitudinales*. De plus depuis la fin du siècle dernier la modélisation conjointe des deux types de données s'est développée. Finalement les observations sont souvent incomplètes car d'une part il est difficile d'observer les phénomènes en temps continu et d'autre part la période d'observation se réduit à un intervalle de temps calendaire.

Ce numéro comporte cinq articles qui abordent des sujets différents tous reliés à cette thématique. Celui de Donnet et Samson traite du problème de l'inférence des paramètres d'une équation différentielle stochastique indirectement observée. La dynamique de nombreux phénomènes biologiques peut être modélisée par de telles équations. Les phénomènes sont observés en temps discret et l'inférence dans ce contexte est particulièrement difficile. Donnet et Samson proposent un nouvel algorithme obtenu par une combinaison de l'algorithme SAEM (Stochastic Approximation EM) et d'un algorithme MCMC particulaire. Dedieu *et coll.* proposent un modèle de Markov caché pour traiter des données longitudinales hétérogènes avec des données manquantes et des erreurs. Pour l'inférence ils proposent un algorithme de type SEM (Stochastic EM) ; ils appliquent leur approche à l'épidémiologie du cancer à partir de données de la cohorte britannique

[1] Laboratoire de Mathématiques et de leurs Applications – UMR CNRS 5142, Université de Pau et des Pays de l'Adour, F-64013, France.
[2] Centre INSERM U-897-Epidemiologie-Biostatistique, Bordeaux, F-33000, France.
Université de Bordeaux, ISPED, F-33000, France.
E-mail : daniel.commenges@isped.u-bordeaux2.fr

 ISSN: 2102-6238

"NCDS 58". Paroissin *et coll.* développent un modèle multi-état en temps discret en présence de variables explicatives pour décrire l'usure (ou la dégradation) d'un composant industriel. L'inférence est réalisée par maximum de vraisemblance couplée à la sélection de variables. Le modèle est appliqué à des composants installés dans des centrales de production d'électricité EDF. Les modèles multi-états sont également utilisés par Gillaizeau *et coll.* qui développent un modèle semi-markovien paramétrique pour étudier l'issue de transplantations rénales. Finalement Séne *et coll.* développent un modèle conjoint pour les concentrations d'antigène spécifique de la prostate et les rechutes cliniques de cancer de la prostate après radiothérapie. Le modèle conjoint est basé sur des effets aléatoires partagés et les paramètres estimés par maximum de vraisemblance.

L'ensemble de ces articles donne un aperçu de ce qu'il est possible de faire dans ce domaine de recherche très actif, à la fois sur le plan de la modélisation, des algorithmes et des applications. La plupart des applications de ce numéro concernent le domaine médical ou biologique. Toutefois une application concerne le domaine industriel où des problématiques semblables trouvent des solutions en fiabilité. Il nous apparaît essentiel de faire communiquer ces deux domaines de la statistique.

 ISSN: 2102-6238

Editorial of the Special Issue "Longitudinal and incompletely observed data of quantitative measurements and events"

Titre : Editorial du numéro spécial "Données longitudinales quantitatives, événementielles, incomplètement observées"

Laurent Bordes[1] and Daniel Commenges[2]

This special issue entitled “Longitudinal and incompletely observed data of quantitative measurements and events” aims at covering a broader than usual spectrum as it combines the analysis of both quantitative data and events. The keyword here is “longitudinal” meaning that we have observations indexed by time allowing one to estimate the dynamics of the phenomena of interest. We felt the need to add “event” because the term “longitudinal data” has long been synonymous with quantitative longitudinal data, also known as “repeated measurements”. Event data are also longitudinal in nature in the sense that they are also indexed by time : the most classical case is that of survival data, which can be generalized by considering events or multiple states. To collect the two types of data, “longitudinal studies” are needed. In addition, since the end of last century the joint modeling of both types of data has been developed. Finally, the observations are often incomplete because it is difficult to observe phenomena in continuous time, and the observation period starts at a certain calendar time and has a finite duration.

This issue contains five articles that address various problems related to this topic. The work of Donnet and Samson concerns the inference of the parameters of a stochastic differential equation whose trajectories are not directly observed. Many phenomena in biology can be modeled by such equations. The observations are in discrete time and inference in this context is particularly difficult. Donnet and Samson propose a new estimation algorithm obtained by combining the SAEM (Stochastic Approximation EM) algorithm with a particular MCMC algorithm. Dedieu *et al.* propose a hidden Markov model for dealing with heterogeneous longitudinal data with missing observations and errors. For inference they also use a SEM algorithm (stochastic EM) and they apply their approach to the epidemiology of cancer using data from the British Cohort “NCDS 58”. Paroissin *et al.* develop a multi-state discrete-time model including covariates to describe the degradation of an industrial component. The inference is based on maximum likelihood estimation with variable selection, and the model is applied to components used in power plants. Multistate models are also used by Gillaizeau *et al.* who consider a parametric semi-Markov model in order to study the outcome of kidney transplants. Finally Sene *et al.* develop a joint model for concentrations of the prostate specific antigen (PSA) and clinical relapses of prostate cancer after radiotherapy. The joint model is based on shared random effects and parameters are estimated by maximum likelihood.

This special issue gives an overview of a very active research topic, both in terms of modeling and algorithms, and applications. Most applications are of a medical or biological nature. One application however stems from the industry, and is based on an estimation method relying on reliability theory. We believe that there is a lot to be gained by connecting the latter field with the areas on which this special issue focuses.

Journal de la Société Française de Statistique, **Vol. 155** No. 1 46-48
http://www.sfds.asso.fr/journal
 ISSN: 2102-6238