Rencontre printanière du laboratoire de statistique 2017

*Spring Meeting of Statlab 2017*

12 Mai 2017/ *May 12, 2017*

Résumés / *Abstracts*

Thierry Duchesne

Faire du neuf avec du vieux : méthode d’estimation à deux étapes (*two-step*) pour ajuster des modèles de régression à effets mixtes à des jeux de données complexes ou de grande dimension

Résumé : Lorsque les chercheurs ont commencé à vouloir ajuster certains modèles linéaires généralisés mixtes à des données groupées ou longitudinales à la fin des années 70, la puissance de calcul limitée qui était disponible à l’époque les a forcés à développer des méthodes d’inférence qui contournaient l’intégration numérique impliquée dans le calcul de la vraisemblance. Des approches en « deux étapes » ont alors été mises sur pied. Une trentaine d’années plus tard, ce type de méthode est redevenu populaire, en particulier pour certains devis d’échantillonnage complexes comme les suivis longitudinaux de famille en statistique génétique ou les études de déplacement d’animaux munis de colliers GPS. Et de nos jours, les approches à deux étapes ont un potentiel intéressant pour analyser les cubes de données (*OLAP cubes*). Dans cet exposé, je présenterai nos travaux sur les méthodes en deux étapes pour les modèles mixtes et illustrerai leur application à des données de suivi GPS d’animaux et des données sur la vente de maisons.

Mylène Bédard

Locally generated candidates in Metropolis-Hastings algorithms

Random walk Metropolis (RWM) and Metropolis-adjusted Langevin (MALA) algorithms within Gibbs are popular tools for sampling from hierarchical models. In this talk, we derive asymptotically optimal tunings for the RWM-within-Gibbs used in conjunction with Gaussian proposal distributions featuring a diagonal covariance matrix. The principal difference with traditional optimal scaling results lies in the local character of the optimal proposal variances obtained, meaning that they vary from one iteration to the next.  These variances may be found analytically in several cases, leading to a personalized version of the proposal variance in a given iteration. Similar ideas are applied to MALA-within-Gibbs, leading to efficient yet computationally affordable algorithms.

The new approach is predicated on the tractability of the distribution of the conditionally i.i.d. components given the mixing parameters and (in practice) the observations; it is thus well suited to some hierarchical models.  As an alternative to the case where local proposal variances have to be obtained numerically, we propose a fixed optimal proposal variance that is computationally more affordable (but less efficient than the local approach). We present numerical illustrations that quantify the benefit of using local proposal variances rather than a fixed one in the RWM-within-Gibbs and MALA-within-Gibbs. To add some perspective, we compare these samplers to single-block RWM and MALA, along with some of their variants that include correlation among candidates. In several cases, local versions of RWM-within-Gibbs and MALA-within-Gibbs can outperform fancy variants included in the MCMC toolbox.

Juli Atherton

Utiliser les statistiques d'ordre pour lier la fonction de risque et

le processus d'intensité

Juli Atherton, Geneviève Lefebvre et René Ferland

A un niveau d'introduction, l'analyse de survie est souvent présentée en termes

du modèle de censure aléatoire ou, pour chaque individu i, les variables aléatoires

Ti et Ci représentent les temps de survie et de censure, respectivement, et les

données observées sont décrites a la fois par le temps observe ~ Ti = min(Ti;Ci) et

l'indicateur de censure 4i = I(Ti Ci). Les fonctions de densité et de survie

des variables Ti et Ci continues peuvent être dénies et, en présumant une censure

aléatoire simple ou Ti et Ci sont statistiquement indépendants, un enseignant peut

alors facilement initier le développement de l'inférence, souvent exécute à travers

la fonction de risque. La présentation décrite plus haut permet aux étudiants

ayant une formation de base en statistique mathématique d'apprendre rapidement

les rudiments de l'analyse de survie. Cependant, elle ne permet pas facilement

l'introduction des techniques liées aux processus de comptage, omniprésentes dans

l'analyse de l'historique des évènements.

Au cœur des techniques liées aux processus de comptage est le processus d'intensité

qui, sous certaines hypothèses simplicatrices, est égal au processus des individus

a risque multiplie par la fonction de risque. Habituellement, cette forme multiplicative

de l'intensité est démontrée sous les conditions les plus faibles possibles

en utilisant des calculs probabilistes. Dans ce travail, nous évitons la notation

probabiliste et présentons le modèle de censure aléatoire en dérivant l'intensité

multiplicative par l'utilisation des statistiques d'ordre. Nous examinons également

le processus de comptage de survie sous-jacent.

*Using Order Statistics to Link the Hazard Function and the*

*Intensity Process*

*Juli Atherton, Genevieve Lefebvre and Rene Ferland*

*Survival analysis, at an introductory level, is often presented in terms of the*

*random censorship model where, for each individual i, random variables Ti and Ci*

*represent the survival and censoring times, respectively, and the observed data is*

*described by both the observed time ~ Ti = min(Ti;Ci) and the censoring indicator*

*4i = I(Ti Ci). Densities and survival functions for continuous Ti and Ci can*

*be dened and by assuming simple random right censoring where Ti and Ci are*

*statistically independent, one can easily begin making inference which is often*

*done through the hazard function. The presentation above allows students with a*

*basis in mathematical statistics to quickly learn survival analysis; however it does*

*not introduce counting methods techniques which are ubiquitous in event history*

*analysis.*

Erica Moodie

Finding Methodological Inspiration in Collaborative HIV Research

In this talk, I will give an overview of my some of my main collaborations in HIV research, and point to methodological gaps that became apparent in trying to answer epidemiological and clinical questions. I will briefly touch on the proposed solutions and the methods that have been developed as a result, while view to highlighting the synergy that true collaboration – rather than simply statistical consulting – can bring to a statistician’s career.

Taoufik Bouezmarni

Copula-based regression function estimation

In this talk, the copula-based approach of estimating a regression function will be presented. The main idea behind this approach is to write the regression function in terms of a copula and marginal distributions. Once the copula and the marginal distributions are estimated, we use the plug-in method to construct the new estimator. Because various methods are available in the literature for estimating both a copula and a distribution, this idea provides a rich and flexible alternative to many existing regression estimators. I provide some asymptotic results related to this copula-based regression modeling when the copula is estimated via profile likelihood and the marginals are estimated nonparametrically. The misspecification problem will be discussed and a nonparametric method will be presented to resolve the problem. Finally, the extension of the mean regression copula estimation to censored data and quantile regression will be introduced, and their asymptotic behavior will be elicited.