Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA)

Young Researchers' Day

23 September, 2016

C115, ISBA

9^{00}
Nathalie Lucas
Actuarial models for health and disability insurance

9^{25}
Florian Pechon
Multiline dynamic P&C insurance loss models

9^{50}
Mailis Amico
The single-index / Cox mixture cure model

10^{20} : Coffee Break

10^{45}
Nicolas Asin
Non-parametric instrumental regression: Adaptive estimation in presence of dependence

11^{15}
Oswaldo Gressani
Approximate Bayesian methods in cure survival models: Coupling P-splines with Laplace approximations for fast inference

11^{40}
Cheikh Ndour
Multiblock Redundancy Analysis : Application to dataset from an adjuvanted vaccine

12^{10}
Gauthier Dierickx
On the Darling-Erdös Theorem

12^{40} : Lunch sandwich in the cafeteria

Nathatlie Lucas

Title : Actuarial models for health and disability insurance

Abstract
The objective of this research project is to develop efficient actuarial models for pricing and reserving in health insurance, with special emphasis on medical expenses cover and loss of autonomy cover. Several issues need to be considered, including indexing mechanism for medical expenses insurance premiums, accounting for the accumulated reserve, prediction of medical costs inflation, and the determination of a surrender or transfer value. A new multistate model is also proposed for the loss of autonomy cover.

Multiline dynamic P&C insurance loss models

Florian Pechon

Abstract

Actuarial risk classification studies are typically confined to univariate, policy-based analyses: individual claim frequencies or severities are modelled for a single product, without accounting for the interactions between the different coverages bought by the same policyholder. We present, in the context of this research funded by the AXA Research Fund, an idea on how to combine all products subscribed by policyholders inside a household.

Maïlis Amico

Title :The single-index / Cox mixture cure model

Abstract
Survival analysis is based on the assumption that if the follow-up period would be long enough, all observations will experience the event of interest. In some situations however, this assumption is not realistic, and survival data can contain a "cure" fraction that will never experience the event of interest. In order to take into account for such a situation, classical survival models have been extended to cure models. The mixture cure model is one approach proposed in the literature. Considering that the population of interest is a mixture of cured and uncured individuals, the model is composed of two elements, the incidence part referring to probability of being uncured, and the latency part corresponding to the survival function for uncured observations. Most often, the probability of being uncured is modeled parametrically assuming a logistic regression model. In this research, we propose to consider a semiparametric modeling through a single-index structure, which offers more flexibility than a parametric approach but avoids the curse of dimensionality phenomenon encountered in nonparametric modeling. We use a kernel estimator for the unknown link function in the single-index and develop an estimation method based on the EM algorithm. Based on simulations, we demonstrate the performance of the proposed method. We also present an application of our methodology on a breast cancer dataset and contrast our result with those obtained assuming a logistic regression model for the incidence part.

Non-parametric instrumental regression: Adaptive estimation in presence of dependence

N. Asin and J. Johannes

Abstract

We consider the estimation of a structural function which models a non-parametric relationship between a response and an endogenous regressor given an instrument in presence of dependence in the data generating process. Assuming an independent and identically distributed (iid.) sample it has been shown in *Johannes and Schwarz [2011]* that a least squares estimator based on dimension reduction and thresholding can attain minimax-optimal rates of convergence up to a constant. As this estimation procedure requires an optimal choice of a dimension parameter with regard amongst others to certain characteristics of the unknown structural function we investigate its fully data-driven choice based on a combination of model selection and Lepskis method inspired by *Goldenshluger and Lepski [2011]*. For the resulting fully data-driven thresholded least squares estimator a non-asymptotic oracle risk bound is derived by considering either an iid. sample or by dismissing the independence assumption. In both cases the derived risk bounds coincide up to a constant assuming sufficiently weak dependence characterised by a fast decay of the mixing coefficients. Employing the risk bounds the minimax optimality up to constant of the estimator is established over a variety of classes of structural functions.

References

Goldenshluger A. and Lepski O., Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, **39**, 1608–1632, 2011.

Johannes J. and Schwarz M., Partially adaptive nonparametric instrumental regression. *Journal of the Indian Statistical Association*, **49(2)**:149–175, 2011.

Asin N. and Johannes J., Adaptive non-parametric instrumental regression in the presence of dependence. Discussion paper, Universite catholique de Louvain, 2016. (arXiv:1604.01992)

Approximate Bayesian methods in cure survival models: Coupling P-splines with Laplace approximations for fast inference

Gressani Oswaldo and Philippe Lambert

Abstract

Standard Bayesian methods for time-to-event data rely on Markov chain Monte Carlo (MCMC) to sample from posterior distributions and perform statistical inference. When confronted with the increasing sophistication of survival models to cope with applied challenges, the MCMC toolbox exhibits a spectrum of practical issues such as slow mixing samplers, potential high posterior correlation between parameters and a strong computational burden. In an attempt to overcome the drawbacks inherent to MCMC sampling, an approximate Bayesian inference methodology has recently been proposed by Rue et al. [1] that delivers accurate posterior approximations at a fast computational speed. We extend the INLA methodology in a Cox proportional hazards model where the baseline log-hazard is specified as a linear combination of cubic B-splines whose coefficients are assigned a multivariate normal prior. This is motivated by the generic idea underlying penalized B-splines by Eilers and Marx [2] and their Bayesian adaptation following Lang and Brezger [3]. Simulation results suggest that our approximate inference method is a promising alternative to MCMC in Bayesian P-spline models. The computational speed is increased by a factor of at least twenty as compared to classic MCMC methods without losing on precision and accuracy. Furthermore, approximate pointwise credible intervals for the conditional survival function can be obtained in a streamlined way. We conclude the presentation with an extension of the methodology to cure survival models where an unknown proportion of unidentified subjects are not at risk for the monitored event.

References

[1] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319-392.

[2] Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89-102.

[3] Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1), 183-212.

Multiblock Redundancy Analysis : Application to dataset from an adjuvanted vaccine

C. Ndour, C. Legrand and B. Govaerts

Abstract

The use of alum as adjuvant was well shown in the Twenties and is largely used within the framework of human vaccine. Its use induces a humoral response but it is not indicated when an important cellular immune response is necessary to protect the individual effectively. It is the case for the treatment of pathogenic complexes, chronic infections, or old individuals or immunocompromising [2].
These last years, a better comprehension of the role of the adjuvants made it possible to develop a more rational approach of those. Using this knowledge, GSK Vaccines developed a series of families of adjuvants system. The adjuvants system AS01 was thus developed in particular for the vaccine against palladium and made it possible to obtain a significant effectiveness against the development of the disease [3]. In this context GSK Vaccinate led a clinical study named Early-Clinres-008 (ECR-008) in order to compare the adjuvant $AS01_B$ with the conventional adjuvant called Alum.
We analyze ECR-008 dataset with a multiblock redundancy method [1] to investigate relationships between the innate immune response and the reactogenocity also between the innate immune response and the adaptive response. Multiblock redundancy modeling provide us a large spectrum of interpretation indices for the investigation of the relationships among variables and among datasets.

References

[1] Stéphanie Bougeard, El Mostafa Qannari, and Nicolas Rose. Multiblock redundancy analysis: interpretation tools and application in epidemiology. *Journal of Chemometrics*, 25(9):467–475, 2011.

[2] Geert Leroux-Roels. Unmet needs in modern vaccinology: Adjuvants to improve the immune response. *Vaccine*, 28, Supplement 3:C25 – C36, 2010. Vaccines Educational Supplement.

[3] S Clinical Trials Partnership RTS and U. D'Alessandro. A phase 3 trial of rts,s/as01 malaria vaccine in african infants. *New England Journal of Medicine*, 367(24):2284–2295, 2012. ITG-B32A; DBM; U-MALAR; JIF; DOI; PDF; DSPACE.

On the Darling-Erdős Theorem

September 23, 2016

Abstract

In the mid fifties, Darling and Erdős obtained a limit in distribution for the maximum of normalized sums of independent random variables, when suitably normalized, under the assumption of a uniformly bounded absolute third moment.

Shorack weakened the moment condition of Darling and Erdős to a finite$(2+\epsilon)-$moment, for any $\epsilon > 0$, when the random variables are i.i.d.

Later on Horváth extended the Darling-Erdős theorem to the case of i.i.d. random vectors with covariance matrix the identity.

By a new strong approximation we show that a slightly modified Darling-Erdős theorem holds under a second moment condition, if the normalizing sequences are allowed to be slightly modified. Furthermore under an additional (weak) assumption, we can show that the theorem holds in the usual form too.

Whether this condition is also necessary in the multivariate case, is a question that is more involved than in the one dimensional, where it was shown by Einmahl to be necessary and sufficient. Other possible generalizations, which are currently investigated, will also be discussed.

References

[1] Darling, D. A. and Erdős, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

[2] Dierickx, G. and Einmahl (2016), U. A general Darling-Erdős theorem in Euclidean space. arXiv:1608.04549

[3] Einmahl, U. (1989). The Darling-Erdős theorem for sums of i.i.d. random variables. *Probab. Th. Rel. Fields.* **82**, 241–257.

[4] Horváth, L. (1994). Likelihood method for testing changes in the parameters of normal observations. *Ann. Stat.* **21**, 671–680.

[5] Shorack, G. (1979). Extension of the Darling and Erdős theorem on the maximum of normalized sums. *Ann. Probab.* **7**, 1092–1096.