Université catholique de Louvain – Institut de statistique, biostatistique et sciences actuarielles

Young Researchers' Day

5 February, 2016

9^{00}	**Hervé Azobou Anantia**	Accelerated Failure Time Model with one covariate subject to random right censoring
9^{35}	**Joris Chau**	Mixed-effects modelling meets spectral analysis of time series
and	**Rainer von Sachs**	
10^{25}	***Coffee Break***	
10^{55}	**Samuel Maistre**	Spline backfitted kernel estimation of an additive model for censored data
11^{30}	**Josefine Hinkelmann**	More Negative Expectation Dependence: Key Properties and Hypothesis Testing

The seminar is followed by the annual lunch of ISBA at "Au coeur des sens".
Traverse d'Esope, 12, 1348 Louvain-la-Neuve

The Young Researchers' Day is held in room c.115 of the Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, Louvain-la-Neuve.

Accelerated Failure Time Model with one covariate subject to random right censoring

Hervé Azobou Anantia
(Herve.Azobou@uclouvain.be)

Survival analysis focuses on positive valued random variables and their relation to explanatory variables. Most of the existing parametric and semiparametric methods assume that the covariates are fully observed. However, less attention has been paid to the case where both the time-to-event response variable and at least one covariate are incompletely observed due to censoring, truncation or even missingness. In some studies, covariates are subject to censoring. For example:

- Measurement below a detection limit: it often happens when measuring concentration in blood sample or when trying to detect pollution in environmental studies. Then, the corresponding measurement is sometimes only known to be below a threshold.
- In studies on bone marrow transplantation: several variables are typically subject to random censoring. The response being the time to death after transplantation, one could be interested in including for example time to normal platelet levels or to acute graft-versus-host disease as covariates in the analysis.
- etc

Some methods are commonly used to handle censored covariates in survival analysis:

- Single imputation or the substitution method consist of replacing each unobserved or censored value by a fixed quantify value and then applying a standard statistical technique to obtain the estimates. This method is not satisfactory and does not have theoretical justification.
- The complete case method consists of removing all entries in the database having at least one missing or censored value for any covariate to be used in the analysis.
- Multiple imputation and data augmentation consist of replacing each missing or censored value by a set of $m > 1$ plausible values reflecting the uncertainty about the values to impute. In that way, One obtains m datasets on which standard techniques are applied.
- In the maximum likelihood based approach, the censoring pattern of the covariates is taken into account for the construction and maximization of the likelihood to reflect the uncertainty due to censoring.

In this work, we propose a simple method based on EM algorithm to estimate and make inference in the context of AFT models where both the response variable and one covariate are subject to random right censoring. We start with the simple case of lognormal AFT model and then extend it by using a smooth distribution for the error term. This is done by using the idea underlying P-splines (Penalized B-splines), but with each B-spline in the basis approximated by a Gaussian density. That approximation was first proposed by *Komarek et al.*(2005).

References

[1] Arnošt Komárek and Emmanuel Lesaffre and Joan F. Hilton (2005), Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, 14, 726 – 745.

[2] A. P. Dempster and N. M. Laird and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 71 – 38.

Mixed-effects modelling meets spectral analysis of time series

Joris Chau and Rainer von Sachs
(Joris.Chau@uclouvain.be and rvs@uclouvain.be)

In this talk we explain how to combine spectral analysis of *replicated* time series with *functional* mixed-effects modelling.
We start with a brain data example motivating our approach and briefly review traditional analysis of the variance-covariance structure of time series in the frequency domain. The existence of subject-replicated time series traces opens the ground for a new approach that is explained in the second part of this team presentation. The idea is to assume that the subject-replicated time series share a common population-mean spectrum (functional fixed effect), additional to some random subject-specific deviation (functional random effects), which models the variability within the population. In contrast to existing work, we allow this variability to be non-diagonal, i.e. there may exist explicit correlation between different subjects in the population.
In order to estimate the population-mean curve and variance-covariance structure within / between different curves, we project the subject-curves onto an appropriate orthonormal basis (such as a wavelet basis) and continue working in the coefficient domain instead of the functional domain. In a sampled data model, with discretely observed noisy subject-curves, the model in the coefficient domain reduces to a finite-dimensional linear mixed model, which allows us to apply traditional linear mixed model estimation methods combined with non-linear wavelet thresholding. We highlight the influence of the correlation in the subject population, as this is among the innovations of our approach, by showing some examples using simulated time series data. To conclude, we return to the analysis of the motivating brain data example.

Spline backfitted kernel estimation of an additive model for censored data

Samuel Maistre
(Samuel.Maistre@uclouvain.be)

Nonparametric additive models have been studied widely since the seminal work of Hastie and Tibshirani (1986). They provide a good compromise between parametric and fully nonparametric modelings. As each function of interest is one-dimensional, it can be pictured and therefore interpreted easily by practitioner. Moreover, a good estimation method should not suffer the curse of dimensionality.

Different techniques have been proposed to estimate the model, including marginal integration, splines, kernel backfitting, ... We present an adaptation of a relatively recent method that combines the ideas of the two latter, namely Spline Backfitted Kernel (SBK) estimation. We propose to adapt it when the response variable is right-censored. This method includes two steps :

1. estimate the additive model using splines for which the number of knots leads to undersmoothing ;
2. use the previous estimates to perform univariate kernel smoothing to estimate each function of interest.

We show that when we use synthetic data, the asymptotic results of this procedure are similar to those of the uncensored case.

References

[1] T. Hastie, R. Tibshirani Generalized additive models *with discussion. Statist. Sci.* 1(3):297–318, 1986.

[2] J. Wang, L. Yang. Efficient and fast spline-backfitted kernel smoothing of additive models. *Ann. Inst. Statist. Math.*, 61(3):663–690, 2009.

[3] H. Koul, V. Susarla, and J. Van Ryzin. Regression analysis with randomly right-censored data. *Ann. Statist.*, 9(6):1276–1288, 1981.

More Negative Expectation Dependence: Key Properties and Hypothesis Testing

Josefine Hinkelmann
(Josefine.Hinkelmann@uclouvain.be)

In this talk, we introduce the negative expectation dependence concept for pairs of random variables, defined after Wright (1987), i.e. when we discover Z is small, in the precise sense that we are given the truncation $Z \leq z$, the expectation of X is revised upward. After having established several of its key properties, we extend this concept to the more negative expectation dependence concept proposed by Dionne et al. (2012). Extending the recent work by Zhu et al. (2014), we construct a hypothesis test of Kolmogorov-Smirnov type, which is shown to control the type I error well and to be consistent under global alternatives. We propose and justify inference based on simulations motivated by the work of Scaillet (2005) and involving the copula approach by Egozcue et al. (2011).

References

[1] Dionne, G., Li, J., Okou, C. (2012). *An extension of the consumption-based CAPM model.* Working paper.

[2] Egozcue, M., García, L., Wong, W., Zitikis, R. (2011). Grüss-type bounds for covariances and the notion of quadrant dependence in expectation. *Central European Journal of Mathematics 9*, 1288-1297.

[3] Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 33(3)*, 415-427.

[4] Wright, R. (1987). Expectation dependence of random variables, with an application in portfolio theory. *Theory and Decision 22*, 111-124.

[5] Zhu, X., Xu, G. Lu, L., Lixing, Z. (2014). Testing for positive expectation dependence. *Annals of the Institute of Statistical Mathematics 68*, 135-153