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

Young Researchers' Day

31 January, 2014

9^{00}	**Nicolas Asin**	Local Likelihood and Transformation Models in Survival Analysis
9^{30}	**Sylvie Scolas**	Accelerated Failure time model with interval censored data and cure
10^{00}	**Fabian Bocart**	Applied statistics and research

Coffee Break

11^{00}	**Anne Benoit**	Taking into Account Strains Heterogeneity in the Estimation of Vaccine Efficacy Against Seasonal Influenza
11^{30}	**Cedric Taverne**	Extended Beta Regressions for (Inflated) Discrete Scales
12^{00}	**Marco Munda**	Modelling spatio–temporal variation in malaria incidence with the spatial frailty model

The seminar is followed by the annual lunch of the ISBA.

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.

Local Likelihood and Transformation Models in Survival Analysis

Nicolas Asin (Nicolas.Asin@uclouvain.be)

In survival analysis we are often interested by the relationship between a survival time T and a covariate X. Many existing models in the literature can be written as :

$$\phi(S(t|x)) = \beta_0 + \beta_1 x \quad \text{or} \quad \phi(\mathbb{E}(I(T > t)|X = x)) = \beta_0 + \beta_1 x$$

where ϕ is a link-function (i.e., logit, probit) and $S(t|x)$ corresponds to the conditional survival function of T given $X = x$. The binary model introduced before can be viewed as a special case of a generalized linear model and the maximum likelihood estimator of $\boldsymbol{\beta}$ can be found by maximizing the log likelihood function (see [1]). The linearity condition is often hard to satisfy in practice. That is why an other estimator can be obtained using a local approximation approach :

$$\phi(S(t|X)) \approx \beta_0 + \beta_1(X - x) \equiv \boldsymbol{\beta}^T \boldsymbol{X}$$

An estimator of $\phi(S(t|x))$ is obtained by maximizing the local kernel-weighted likelihood:

$$\sum_{i=1}^{n} \left[I(T_i > t) \ln \pi(\boldsymbol{\beta}^T \boldsymbol{X}_i) + (1 - I(T_i > t)) \ln \bar{\pi}(\boldsymbol{\beta}^T \boldsymbol{X}_i) \right] K\left(\tfrac{X_i - x}{h}\right)$$

where $\pi = \phi^{-1}$.
In this presentation we will study the asymptotic result in the uncensored case and we will extend the result for the censored case. In this latter situation we will compare the local constant estimator with the Beran estimator [2].

References

[1] Jung, S. H. (1996). Regression analysis for long-term survival rate. *Biometrika*, 83, 227–232.

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

Accelerated Failure time model with interval censored data and cure

Sylvie Scolas (Sylvie.Scolas@uclouvain.be)

Mild cognitive impairment (MCI) may be a precursor of Alzheimer disease or other dementia. Studying the time until conversion to MCI [1] makes use of survival analysis theory. Generally, within this field, it is assumed that if the follow-up time is long enough, then the event of interest will be observed for each individual. In our case, not everybody will show signs of impairment. We then say that a proportion of the population is "cured", or "long-term survivor".

Also, patients come to scheduled interviews and thus we can only detect MCI to have appeared between two visits. That is, the database contains interval censored data. Thus, we propose to extend the existing survival models to the case where interval censored data and cure may be present.

In this talk, we present the method we want to use: to model event times (i.e. the latency part), we utilize an accelerated failure time (AFT) regression model, adapted to interval censored data, together with an extended generalized gamma (EGG) distribution for the error term of the AFT. In addition, modeling the cure proportion (i.e. the incidence part) is made by a logistic regression.

Furthermore we show the good behavior of the method thanks to results of simulations. Then, we address some issues concerning variable selection in such a model and finally, if time permits, we apply this method to our Alzheimer disease database, which consist in 241 at-risk patients followed-up between 1998 and 2008 with regular checks for the appearance of MCI.

References

[1] Oulhaj, A., Wilcock, G. K., Smith, A. D, and de Jager, C. A. (2009), Predicting the time of conversion to MCI in the elderly: role of verbal expression and learning. *Neurology*, 73, 1436–42.

Taking into Account Strains Heterogeneity in the Estimation of Vaccine Efficacy Against Seasonal Influenza

Anne Benoit (Anne.Benoit@uclouvain.be)

Influenza is an infectious disease caused by several virus strains whose repartition varies between geographical regions and seasons. Typically, a vaccine contains 3 or 4 influenza strains and the antigen content is annually reconsidered based on the WHO recommendation. For the same vaccine formulation, pharmaceutical regulations only require efficacy again clinical disease to be shown for a single season, which is performed through a large phase III trial. Subsequent annual modifications of the strain related portion of the vaccine only have to be validated through immunogenicity trials.

Classically, influenza vaccine efficacy (VE) trials take place over a single season but over several geographical regions assuming common VE. However, depending on the circulating strains characteristics such as their immunogenicity and their matching levels with the vaccine strains, the vaccinal protection level may vary from one season/region to another. As a result, the same vaccine can sometimes be proved to be significantly efficacious in one trial but not the other, all other things being equal.

We argue that not taking this into account provides incomplete and unreliable response as for the benefit of the vaccine in the future. We therefore propose to run phase III VE trials over several geographical regions and seasons in order to characterize the VE heterogeneity. We consider VE as the sum of a common quantity to all clusters (season and geographical region) and of a random cluster-specific part.

We also argue that in the specific context of VE against a heterogeneous disease such as seasonal influenza the decision about VE shouldn't be taken based on a confidence interval for the mean estimated VE but based on a tolerance interval, taking into account the heterogeneity across clusters. Such information provides insight on the range of future VE across seasons and geographical regions, instead of a mean past season specific vaccine effect.

We analyse trial data with a hierarchical survival model taking into account risk heterogeneity and VE heterogeneity across clusters. Our model parameters and the tolerance interval for the cluster-specific VE are estimated using Bayesian statistics.

Extended Beta Regressions for (Inflated) Discrete Scales

Cedric Taverne (Cedric.Taverne@uclouvain.be)

Whatever the subject of interest, discrete scales are used everywhere in questionnaires. It could be Likert scales from *Strongly disagree* to *Strongly agree* or rating scales from 0 to 10. In applications, these kinds of scales are generally used as response variable in linear regression with normally distributed error terms. However, this approach has several statistical limits: Rates on such scales are regularly (a) skewed, always (b) bounded and often (c) inflated at one of the bounds of the scale. Last but not least, the (d) discreteness of the scale is not recognized in estimation and predictions.

In response to those problems, we adapt the beta regression proposed by [2] and [1]. We consider the discrete scales as the visible part of an underlying beta distribution and we jointly model the mean and the dispersion of this distribution. We have adapted our likelihood to suit the discreteness of the scale (d). Skewed (a) and bounded (b) rates are naturally supported by the beta distribution. Finally, when one of the bound is inflated (c), we switch to a bound-inflated discrete beta regression by jointly modeling the additional mass of probability on the inflated end.

References

[1] Simas, A., Barreto-Souza, W. and Rocha, A. (2010), Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis*, 54(2), 348–366.

[2] Ferrari S. and Cribari-Neto F. (2004), Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 31(7), 799–815.

Modelling spatio–temporal variation in malaria incidence with the spatial frailty model

Marco Munda (Marco.Munda@uclouvain.be)

I have data on the time to malaria (possibly right-censored) for 2000 children living close to a dam in Ethiopia. Starting from the Cox model, I will introduce a model that enables to estimate the effect of distance to the dam on the risk of malaria while accounting for the correlation between children.