Frailty models

November 30, 2004

Programme

9.30h-10.30h Veronique Rondeau (INSERM, Bordeaux)

Title: **A three-level nested frailty model for clustered survival data using a penalized likelihood**

10.30h-11.00h Coffee Break

11.00h-11.30h Arnost Komarek (KULeuven)

Title: **Bayesian accelerated failure time model for correlated censored data with a normal mixture as an error distribution**

11.30h-12.00h Luc Duchateau (RUGent)

Title: **Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows**

12.00h-12.30h Philippe Lambert (Université de Louvain)

Title: **Bayesian survival models with smooth time-varying coefficients: a penalized Poisson regression approach**

Lecture Room: UZRA 00.55 at Uz St Rafaël, Kapucijnenvoer 35, Leuven

Costs: Free of charge

Further information can be obtained from Jeannine Rongy (*jeannine.rongy@med.kuleuven.ac.be*)

A three-level nested frailty model for clustered survival data using a penalized likelihood.

V. RONDEAU[1], L. FILLEUL[2], P. JOLY[1]

[1] : INSERM EMI 0338(Biostatistic) , Université Victor Segalen Bordeaux2,
146 rue Léo Saignat, 33076 Bordeaux

[2] : Institut de Veille Sanitaire, Saint-Maurice, France.

Nested frailty models are particularly appropriate when data are clustered at several hierarchical levels naturally or by design. In such cases it is important to estimate the parameters of interest as accurately as possible. We present a maximum penalized likelihood estimation (MPnLE) to estimate nonparametrically a continuous hazard function in a nested gamma-frailty model with right-censored and left-truncated data. The estimators for the regression coefficients and the variance components of the random effects will be obtained simultaneously. Simulations for the proposed estimation procedure are presented. In order to illustrate the MPnLE method and the nested frailty model, we present an application for modeling the effect of particulate air pollution on mortality in different geographical areas. The application is based on a cohort of 6973 subjects from 18 areas of 6 French towns.

A Bayesian Accelerated Failure Time Model for Correlated Censored Data with a Normal Mixture as an Error Distribution

Arnošt Komárek and Emmanuel Lesaffre

Katholieke Universiteit Leuven, Biostatistical Centre
Kapucijnenvoer 35, B–3000, Leuven, Belgium

Key words: Censored Data, Clustered Data, Regression, Reversible Jump Markov Chain Monte Carlo

A Bayesian accelerated failure time model for possibly clustered data with a mixture of normals as an error distribution is proposed. I.e.

$$\log(T_{il}) = b_i + \boldsymbol{\beta}^T \mathbf{x}_{il} + \varepsilon_{il}, \quad i = 1, \ldots, N,\ l = 1, \ldots, n_i,$$

where T_{il} is time to event of interest for the lth observation of the ith cluster, b_i is a cluster specific random effect, $\mathbf{x}_{il}$ is a vector of fixed covariates, $\boldsymbol{\beta}$ is a vector of regression parameters and ε_{il} the error term with a density $f(e)$.

To avoid strict parametric assumptions concerning the distribution of ε_{il} we adapted a method of Richardson and Green [1] into the regression context, specifying $f(e)$ in a flexible way as

$$f(e) = \sum_{j=1}^{k} w_j\ \varphi(e \mid \mu_j,\ \sigma_j^2),$$

with $\varphi(\cdot \mid \mu,\ \sigma^2)$ being a density of a normal distribution with mean μ and variance σ^2. Note that the number of mixture components, k, is unknown as well as mixture weights $\boldsymbol{w} = (w_1, \ldots, w_k)^T$, means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$, and variances $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)^T$. The random effect b_i is assumed to follow a normal distribution with zero mean and variance σ_b^2.

Unknown quantities (mixture parameters k, $\boldsymbol{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$, regression parameter $\boldsymbol{\beta}$ and a variance of random effects σ_b^2) are estimated in a Bayesian manner by applying Markov chain Monte Carlo techniques. The method allows not only for usual right or left censoring but also for interval censoring where the observations consist of intervals $(T_{il}^L, T_{il}^U]$ of which is known to contain the true event time T_{il}.

A software for the suggested model is provided as a set of R functions with McMC sampling performed by a C++ compiled code available upon request from the first author under the GNU licence.

References

[1] Richardson, S., and Green, P. J. (1997). *On Bayesian analysis of mixtures with unknown number of components (with Discussion).* Journal of the Royal Statistical Society, Series B, **59,** 731 – 792.

Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows

Luc Duchateau[1] and Paul Janssen[2]

[1]Department of Physiology, Biochemistry and Biometrics, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium.

[2]Limburgs Universitair Centrum, Center for Statistics, Universitaire Campus, 3590 Diepenbeek, Belgium.

In many epidemiological studies time to event data are clustered and the physiological relationship between (time dependent) covariates and the log hazard is often not linear as assumed in the Cox model. Introducing frailties in the Cox model can account for the clustering of the data and smoothing splines can be used to describe nonlinear relations. These two extensions of the Cox model are introduced jointly and it is shown how penalized partial likelihood techniques can be used to fit the extended model. We demonstrate the need for such a model to study the relation between the physiological covariates milk ureum and protein concentration and the log hazard of first insemination in dairy cows, with the farms as clusters.

Bayesian survival models with smooth time-varying coefficients: a penalized Poisson regression approach

P. Lambert & Paul Eilers

It is rather well known that one can approach survival problems without covariates in an actuarial way. The time axis is divided into intervals (named bins), and in each bin the number of people at risk is counted as well as the number of events. The relationship between time and probability of an event can then be estimated with a parametric or semi-parametric model (Efron, JASA 1988).

Here, we consider a subdivision of the time scale into a large number of bins. The number of events observed in each bin is described using a Poisson distribution with the log mean specified using a flexible penalized B-splines model (Eilers and Marx, Statistical Science 1996) with knots located at the bins limits.

Regression on pertinent covariates can easily be performed using the same log-linear model, leading to the classical proportional hazard model. We propose to extend that model by allowing the regression coefficients to vary in a smooth way with time. Penalized B-splines models will be proposed for each of these coefficients.

We shall show how the regression and the penalty coefficients can be estimated using Bayesian inference tools. More specifically, we propose to use the Metropolis-adjusted Langevin algorithm (Roberts and Tweedie, Bernoulli 1996) on blocks of B-splines parameters together with Gibbs steps for the penalty variance parameters.

Frailty components can also be added to the model if desired. Their inclusion in a Bayesian framework does not complicate the inference process.

The usefulness of the approach will be illustrated with the analysis of a clinical study.