Modelling long term survival with non-proportional hazards

Perperoglou, A.

Summary

The application of statistical models in medical research is a very important part of statistical science. In this work we consider models for survival data. This thesis consists of an Introduction, five papers (three published and two submitted), a Discussion and an Appendix. The Introduction presents some technical information about the core of survival analysis, namely the Cox model. Moreover, it introduces the main ideas behind the extensions of the model proposed later on.

In Chapter 2, reduced-rank methods for modelling non-proportional hazards are presented. The main idea of reduced-rank regression is extended to apply in survival data and the methodology is explained in detail. The chapter illustrates how to estimate the model and discusses how to choose the rank and the appropriate time functions. The methods are then applied to a small data set of 358 ovarian cancer patients and the results are compared to those from a proportional hazards and a gamma frailty model.

Chapter 3 presents an efficient algorithm for estimating reduced-rank models as well as Cox models with time varying effects of the covariates. The computational challenges that these model present are discussed in the introduction and then an algorithm for estimation is suggested. Based on that algorithm a program is written in `R` that is distributed as a package from the authors website. The program is compared with the already available routines in `S-plus` and `SAS` on simulated data sets. For illustration purposes reduced-rank models are fitted to the same ovarian data set as in Chapter 2, and a larger data set of 2433 breast cancer patients.

The next Chapter deals with the gamma frailty (Burr) model, and the limitations that arise from the assumption of constant frailties. A small discussion is presented on the plausibility of that assumption. The chapter discuss alternative models with time dependent frailties. Most of these models make use of complicated mathematical expressions that makes their estimation difficult. As a result it is hard to incorporate their use in applied statistical research. As a solution the relaxed Burr model is introduced providing a flexible way to model converging and non-converging hazards. The model is presented in detail and evaluated using a series of simulated data sets.

The relaxed Burr model is then fitted to different data sets. In a study of patients suffering from lung cancer the relaxed Burr model was fitted to the data revealing converging hazards. In such a case the model estimates were very similar to the simple Burr model. In a second application, coming from registry data of patients suffering from Acute Myeloid Leucemia, the relaxed Burr model showed that the hazards had a more complex pattern, converging for the first four years and then diverging to proportionality.

An important aspect of statistical research is which model is the most appropriate for the data and how to evaluate the performance of fitted models. In Chapter 5 models with time varying effects of the covariates, frailty models and cure rate models are considered as different modelling approaches to the data set of 2433 breast cancer patients, introduced before in Chapter 3. The data are presented in detail, and seven different models are fitted. The usefulness of each one of these models is discussed and their results are compared. To asses the predictive ability of these approaches pseudo-observations and Brier scores are used. These two measures are then combined to come up with a new way of assessing the predictive ability of the model.

The sixth Chapter of the thesis discusses ways of dealing with overdispersion when using generalized linear models. A class of new models are introduced, which include a parameter vector in the linear component of the model, one for each individual unit of observation. For estimation, a ridge penalty is subtracted from the likelihood of the model. The penalty is useful to maintain identifiability, remove collinearity in the estimating equations and reduce the effective dimensions of the model. The theory is presented for Poisson and Binomial data, as well as smoothing of life tables and several applications are presented. An efficient algorithm for computing the deviance effects is given in detail.

The Discussion is about future directions of the research presented in this thesis. Finally, there is an Appendix about the use of `coxvc` a package written in `R` for fitting Cox models with time varying effects of the covariates and reduced-rank models.