On Change-Point Models in Survival Analysis with Applications in Reliability

Uwe Jensen *Department of Applied Mathematics and Statistics, University of Hohenheim, Stuttgart, Germany*

Abstract: A Cox-type regression model is considered which admits change-points in the covariates. A change-point specifies the unknown threshold at which the influence of a covariate shifts smoothly, i.e. the regression parameter may change over the range of a covariate and the underlying regression function is continuous but not differentiable. The model can be used to describe change-points in different covariates but also to model more than one change-point in a single covariate. Estimates of the change-points and of the regression parameters are derived and their properties are investigated. It is shown that not only the estimates of the regression parameters are $\sqrt{n}$-consistent but also the estimates of the change-points in contrast to the conjecture of other authors. Asymptotic normality is shown by using results developed for M-estimators. Finally the model is applied to data sets which stem from engineering experiments.

1 Introduction

Consider a multivariate counting process $\mathbf{N}(t) = (N_1(t), \dots, N_n(t))$, where $N_i(t)$ counts observed events in the life of the ith individual, $i = 1, \dots, n$, over the time interval $[0, \tau]$. The sample paths of $\mathbf{N}(t)$ are step functions, zero at time zero with jumps of size one only and no two components jump at the same time. The counting process $\mathbf{N}(t)$ admits an intensity $\boldsymbol{\lambda}(t) = (\lambda_1(t), \dots, \lambda_n(t))$ such that the processes $M_i(t) = N_i(t) - \int_0^t \lambda_i(u)du$, $i = 1, \dots, n$, and $t \in [0, \tau]$ are martingales. Different models are determined by their intensities. The intensity of the basic Cox model with baseline hazard $\lambda_0(t)$ and covariate vector $\mathbf{Z}(t)$ is given by $\lambda(t) = \lambda_0(t) \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}(t)\}$. In this model it is assumed that the influence of a covariate is constant in time and over the range of the covariate. By analyzing different datasets we found out that some covariates exhibit deviations from this assumption. Therefore, we proposed a new variant of the Cox model with a smooth change at an unknown threshold ξ (see [1]).

In the literature several extensions of the Cox model have been investigated. Among these extensions is a model introduced by Pons [2]

$$\lambda(t) = \lambda_0(t) \exp\{\boldsymbol{\alpha}^T \boldsymbol{Z}_1(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_2(t) I_{\{Z_3 \le \zeta\}} + \boldsymbol{\gamma}^T \boldsymbol{Z}_2(t) I_{\{Z_3 > \zeta\}}\},$$

where the influence of a covariate jumps at a certain threshold ζ. The estimate of the jump change-point parameter was shown to be n-consistent in this case.

2 The Model

In our model we allow for more than only one smooth change-point, the covariates may be time-dependent and the counting process may jump more than once. The model involving m change-points and p ordinary covariates (without change-points) is given as follows:

$$\lambda_i(t,\boldsymbol{\theta}) = \lambda_0(t)R_i(t)\exp\left\{\boldsymbol{\beta}_1{}^T\mathbf{Z}_{1i}(t) + \boldsymbol{\beta}_2^T\mathbf{Z}_{2i}(t) + \boldsymbol{\beta}_3^T(\mathbf{Z}_{2i}(t) - \boldsymbol{\xi})^+\right\},$$

where $\boldsymbol{\theta} = (\boldsymbol{\xi}^T, \boldsymbol{\beta}^T)^T$ with $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^m$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T)^T \in \mathcal{B} \subset \mathbb{R}^{p+2m}$. Here $\boldsymbol{\xi}$ and $\boldsymbol{\beta}$ are the vectors of change-points and regression parameters respectively, $\lambda_0(t)$ is the baseline intensity and $R_i(t)$ is a process taking only values 1 or 0 to indicate whether a subject is at risk or not.

The parameter vector $\boldsymbol{\theta}_0$ is estimated by the value $\hat{\boldsymbol{\theta}}_n$ that maximizes the logarithm of the partial likelihood. In particular, we show that the estimates of the change-points are only $\sqrt{n}$-consistent and not n-consistent as one might have guessed. As usual, the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is estimated by the Breslow estimator $\hat{\Lambda}_n(t)$.

Besides consistency we can show asymptotic normality of the estimators by means of results developed for M-estimators (see [3]).

3 Applications

We applied our model to some datasets. Among these are an actuarial dataset and the well known PBC dataset. It can be shown by means of a goodness-of-fit test that we get a better fit of the data using our change-point model instead of using the classical models. In addition we analysed datasets consisting of survival times and covariates recorded for electric motors and transmissions.

References

[1] Jensen U., Lütkebohmert, C. (2008). A Cox-Type Regression Model with Change-Points in the Covariates, *Lifetime Data Analysis* 14: 267 - 285.

[2] Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate, *Ann. Stat.* 31: 442 - 463.

[3] Van der Vaart, A. (1998). *Asymptotic statistics*, Cambridge University Press, New York.