A Cox-type regression model with change-points in the covariates

Uwe Jensen · Constanze Lütkebohmert

Received: 16 May 2007 / Accepted: 7 January 2008 / Published online: 26 January 2008
© Springer Science+Business Media, LLC 2008

Abstract We consider a Cox-type regression model with 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. At the end of this paper we apply our model to an actuarial dataset, the PBC dataset of Fleming and Harrington (Counting processes and survival analysis, 1991) and to a dataset of electric motors.

Keywords Survival analysis · Cox model · Change-point · Consistency · Asymptotic normality

1 Introduction

Consider a multivariate counting process $N(t) = (N_1(t), \ldots, N_n(t))$, where $N_i(t)$ counts observed events in the life of the $i$th individual, $i = 1, \ldots, n$, over the time interval $[0, \tau]$. The sample paths of $N(t)$ 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 \( N(t) \) admits an intensity \( \lambda(t) = (\lambda_1(t), \ldots, \lambda_n(t)) \) such that the processes \( M_i(t) = N_i(t) - \int_0^t \lambda_i(u) \, du, \quad i = 1, \ldots, n, \) 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 \( Z(t) \) is given by \( \lambda(t) = \lambda_0(t) \exp\{\beta^T Z(t)\} \).

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 an actuarial dataset 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 \( \xi \) (Gandy et al. 2005).

In the literature several extensions of the Cox model have been investigated. One has to distinguish between change-points which occur in time and those in the covariate. Luo and Boyett (1997) studied the following model \( \lambda(t) = \lambda_0(t) \exp\{\beta_0 \mathbb{I}\{X \leq \theta_0\} + \alpha_0 Z\} \) with one-dimensional covariates \( X \) and \( Z \), where a constant is added after reaching a threshold. They proved consistency of the maximum likelihood estimators. Liang et al. (1990) considered the model \( \lambda(t) = \lambda_0(t) \exp\{(\beta + \theta \mathbb{I}\{t \leq \tau\})Z + \gamma X\} \) with a change-point at an unknown time \( \tau \). Pons (2003) introduced the model

\[
\lambda(t) = \lambda_0(t) \exp[\alpha^T Z_1(t) + \beta^T Z_2(t) \mathbb{I}\{Z_{3\leq\xi}\} + \gamma^T Z_2(t) \mathbb{I}\{Z_{3>\xi}\}],
\]

where the influence of a covariate jumps at a certain threshold \( \xi \). The estimate of the jump change-point parameter was shown to be \( n \)-consistent in this case. In a recent paper Kosorok and Song (2007) considered linear transformation models with a change-point in the regression coefficient based on a covariate threshold. They established consistency and weak convergence of the non-parametric maximum likelihood estimates. The model we propose is a further extension of that one suggested in Gandy et al. (2005). Now we allow that more than only one change-point is in the model, that the covariates are time-dependent and that 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:

\[
\lambda_i(t, \theta) = \lambda_0(t) R_i(t) \exp\left\{\beta_1^T Z_{1i}(t) + \beta_2^T Z_{2i}(t) + \beta_3^T (Z_{2i}(t) - \xi)^+\right\},
\]

where \( \theta = (\xi^T, \beta^T)^T \) with \( \xi \in \Xi \subset \mathbb{R}^m \) and \( \beta = (\beta_1^T, \beta_2^T, \beta_3^T)^T \in \mathcal{B} \subset \mathbb{R}^{p+2m} \). Here \( \xi \) and \( \beta \) 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. For brevity, we consider

\[
\lambda_i(t, \theta) = \lambda_0(t) R_i(t) \exp\{\beta^T \tilde{Z}_i(t; \xi)\}, \quad (1)
\]

where

\[
\tilde{Z}_i(t; \xi) = \left(Z_{1i}^T(t), Z_{2i}^T(t), ((Z_{2i}(t) - \xi)^+)^T\right)^T.
\]