Computational Statistics and Data Analysis 74 (2014) 142–156

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

On the maximum penalized likelihood approach for proportional hazard models with right censored survival data

Jun Ma [a,*], Stephane Heritier [c], Serigne N. Lô [b,c]

[a] Macquarie University, Australia
[b] The George Institute for Global Health, Australia
[c] The University of Sydney, Australia

ARTICLE INFO

Article history:
Received 28 September 2012
Received in revised form 19 September 2013
Accepted 7 January 2014
Available online 21 January 2014

Keywords:
Proportional hazard model
Maximum penalized likelihood
Multiplicative iterative algorithm
Newton's algorithm
Independent censoring

ABSTRACT

This paper considers simultaneous estimation of the regression coefficients and baseline hazard in proportional hazard models using the maximum penalized likelihood (MPL) method where a penalty function is used to smooth the baseline hazard estimate. Although MPL methods exist to fit proportional hazard models, they suffer from the following deficiencies: (i) the positivity constraint on the baseline hazard estimate is either avoided or poorly treated leading to efficiency loss, (ii) the asymptotic properties of the MPL estimator are lacking, and (iii) simulation studies comparing the performance of MPL to that of the partial likelihood have not been conducted. In this paper we propose a new approach and aim to address these issues. We first model baseline hazard using basis functions, then estimate this approximate baseline hazard and the regression coefficients simultaneously. The penalty function included in the likelihood is quite general but typically assumes prior knowledge about the smoothness of the baseline hazard. A new iterative optimization algorithm, which combines Newton's method and a multiplicative iterative algorithm, is developed and its convergence properties studied. We show that if the smoothing parameter tends to zero sufficiently fast, the new estimator is consistent, asymptotically normal and retains full efficiency under independent censoring. A simulation study reveals that this method can be more efficient than the partial likelihood method, particularly for small to moderate samples. In addition, our simulation shows that the new estimator is substantially less biased under informative censoring.

1. Introduction

The proportional hazard (or Cox's regression) model (e.g. Cox (1972) and Kalbfleisch and Prentice (2002, Chapter 4)) is one of the most widely used model for analyzing time to event (survival time) data. Regression coefficients of proportional hazard models, however, are usually estimated by maximizing the Cox's partial likelihood (PL) function (Cox, 1975), where the baseline hazard is not required when estimating the regression coefficients. The popularity of PL is mainly due to: (i) its ability to avoid estimating the baseline hazard function when the main focus is on estimating the regression coefficients, and (ii) asymptotic properties of the maximum partial likelihood estimate, such as in Tsiatis (1981) and Andersen and Gill (1982). When the baseline hazard is also of interest it can be estimated from the Breslow estimate (Breslow, 1972) or smoothed versions of it based on a kernel (Gray, 1990) or penalized likelihood smoothing (Anderson and Senthilselvan, 1980). The PL approach, however, is not always easy to extend to more general models or complex censoring schemes. For instance,

* Correspondence to: Department of Statistics, Macquarie University, NSW 2109, Australia. Tel.: +61 2 98508548; fax: +61 2 98507669.
E-mail address: jun.ma@mq.edu.au (J. Ma).

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2014.01.005

such a partial likelihood is not available in a closed form for interval censored data (Sun, 2006) unless a PL-based estimating equation approximation (Datta et al., 2000) is used. Direct extensions of PL to additive or accelerated failure time models are not easy.

In this paper we adopt the MPL method to fit the Cox model. Earlier relevant references include Gray (1994), Joly et al. (1998) and Cai and Betensky (2003) where splines were used to approximate the baseline hazard for left or right censoring. Gray (1994) focused on testing hypotheses on covariate effects in the Cox model while others considered the estimation problem. Different penalty functions, such as roughness penalties, were considered.

However, existing methods suffer from some of the following deficiencies. Firstly, the way the positivity constraint on the baseline hazard is dealt with is suboptimal. It is either ignored (Gray, 1994) or ensured through "ad-hoc" techniques like the use of squares of splines coefficients (Joly et al., 1998), which potentially creates convergence issues (especially when some coefficients are zero). Others like Cai and Betensky (2003) prefer to model the log-baseline hazard as a linear spline, leading to difficulties in obtaining the cumulative baseline hazard in closed form, and also linear spline is inflexible. Secondly, the asymptotic properties are rarely discussed or rigorously presented. Thirdly, simulation studies comparing their performance to PL's are lacking, casting doubt on the real advantages of these approaches. This paper attempts to address these issues for a new MPL proposal. Specifically the non-negativity constraint is ensured by directly restricting the coefficients of the basis functions, and we develop a new algorithm computing *simultaneously* the MPL estimates of the regression coefficients and baseline hazard. The asymptotic properties of the MPL estimator are derived when (a) the smoothing parameter converges to zero sufficiently fast, or (b) when it does not converge to zero. Its performance with respect to PL is assessed by simulations.

We develop our method in the context of proportional hazard model and right censoring in this paper. But it is, in principle, general enough to accommodate a large class of censoring schemes including interval censoring. Extensions to different survival models such as additive or accelerate failure time models are clearly possible but will be reported elsewhere.

This paper is organized as follows. In Section 2 we review the standard formulation of the Cox model and introduce the MPL approach. In Section 3 we propose a new alternating algorithm for computing the MPL estimates of the regression coefficients and baseline hazard and prove its convergence. The asymptotic properties of the MLP estimates are developed in Section 4, where consistency and asymptotic normality of the MPL estimator are obtained when the smoothing parameter is of order $o(n^{-1/2})$. In Section 5 a simulation study comparing the performance of PL and MPL under various levels of censoring is conducted. Under independent censoring MPL provides more efficient estimates than PL for small to moderate sample sizes, with the gain vanishing for large samples. Moreover, we find from the simulation that the MPL estimates are consistently less biased than the PL's under dependent censoring, particularly for large censoring proportions and sample sizes. In some cases the gain in efficiency is substantial. The new approach is also illustrated using a critical care dataset in Section 5. Finally, concluding remarks are given in Section 6.

2. Fitting proportional hazard model by maximum penalized likelihood

2.1. Model and notation

For simplicity, we assume continuous survival times throughout this paper. We denote the independent survival times by $t_1, \ldots, t_n$ (obtained from n individuals), where some of these observations are right censored. To simplify notations we use $\mathcal{C}$ and $\mathcal{O}$ to denote, respectively, the index set for censored and fully observed survival (or event) times; thus t_i with $i \in \mathcal{C}$ represents a right censored time, while t_i with $i \in \mathcal{O}$ presents an event time. Corresponding to each t_i, there are observations on a set of p covariates, and these observations are denoted by $x_{i1}, \ldots, x_{ip}$. For each observation, the censoring time is assumed to be independent of the survival time. Let $h_i(t)$ be the hazard function of object i. The proportional hazard model specifies $h_i(t)$ according to

$$h_i(t) = h_0(t) \exp\{X_i\beta\}, \tag{1}$$

where $h_0(t)$ is the baseline hazard, $X_i = (x_{i1}, \ldots, x_{ip})$ is the ith row of the design matrix X, and $\beta = (\beta_1, \ldots, \beta_p)^T$ is the regression coefficient vector. Here the superscript T represents matrix transpose. We aim at estimating the coefficient vector β and the baseline hazard $h_0(t)$.

The traditional approach of estimating β, proposed by Cox (1972), is achieved by maximizing the PL function, given by

$$\mathcal{L}(\beta) = \prod_{i=1}^{K} \frac{\exp\{X_{(i)}\beta\}}{\sum_{r \in R(t_{(i)})} \exp(X_r\beta)}, \tag{2}$$

where $t_{(i)}$ is the ith ordered statistic of the uncensored t_i's, K is the total number of event times, $X_{(i)}$ is X_i corresponding to $t_{(i)}$ and $R(t_{(i)})$ is the set of individuals at risk at $t_{(i)}^-$. Note that this partial likelihood can be explained as the conditional probability of an individual who failed at $t_{(i)}$ given the risk set $R(t_{(i)})$ (e.g. Kalbfleisch and Prentice (2002, Chapter 4)). Once the estimate of β is available, the baseline hazard $h_0(t)$ can be obtained from, for example, the Breslow method; see Breslow (1972).

Although the PL estimate of β is asymptotically unbiased (Tsiatis (1981) and Andersen and Gill (1982)) in the case of independent censoring, it is known to be biased under dependent or informative censoring; see also the simulation results