Journal of Multivariate Analysis 123 (2014) 281–303

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Model assisted Cox regression

Shoubhik Mondal, Sundarraman Subramanian *

Center for Applied Mathematics and Statistics, Department of Mathematical Sciences, New Jersey Institute of Technology, USA

ARTICLE INFO

Article history:
Received 26 December 2012
Available online 11 October 2013

AMS subject classifications:
62E29
62H10
62N01
62N02
62N03

Keywords:
Empirical coverage
Event-time hazard
Gaussian process
Loewner ordering
Mean integrated squared error
Missing at random

ABSTRACT

Semiparametric random censorship (SRC) models (Dikta, 1998) [7], derive their rationale from their ability to utilize parametric ideas within the random censorship environment. An extension of this approach is developed for Cox regression, producing new estimators of the regression parameter and baseline cumulative hazard function. Under correct parametric specification, the proposed estimator of the regression parameter and the baseline cumulative hazard function are shown to be asymptotically as or more efficient than their standard Cox regression counterparts. Numerical studies are presented to showcase the efficacy of the proposed approach even under significant misspecification. Two real examples are provided. A further extension to the case of missing censoring indicators is also developed and an illustration with pseudo-real data is provided.

1. Introduction

The SRC framework to survival function estimation for a homogeneous population, introduced by Dikta [7], operates as follows: specify a good-fitting parametric model for $m(x)$, the conditional expectation of the censoring indicator given the observed (possibly censored) event time, and replace the censoring indicators with the estimated model thereafter. With correct parametric specification of $m(x)$, this leads to an estimator which is asymptotically more efficient than the Kaplan–Meier estimator. Subramanian [21] employed this idea to construct likelihood ratio based confidence intervals for survival functions and reported good performance even in the face of considerable misspecification. The SRC approach is more flexible than the nonparametric approach in the sense that it applies even when there are missing censoring indicators (MCIs). In fact, when the MCIs are missing at random (MAR), no additional effort need be expended to address estimation [22]. Here, we propose and implement an extension that incorporates the SRC approach into Cox regression.

Analogous to the homogeneous case, for Cox proportional hazards (PH) regression we propose to replace the censoring indicator with any good-fitting parametric model for the aforementioned conditional expectation, which, in addition to its dependence on the observed event time, may now also depend on a set of covariates $\boldsymbol{Z}$. In order to understand the rationale for tying SRC models to Cox regression, note that, under conditional independence of failure and censoring variables given the covariate $\boldsymbol{Z}$, $m(x, \boldsymbol{z}) = P(\delta = 1|X = x, \boldsymbol{Z} = \boldsymbol{z})$ is the ratio of the conditional event-time hazard to the conditional *total* hazard [7,27]. Specifically, for the Cox PH regression model, the conditional censoring hazard is linked to the event-time hazard through the multiplicative factor $\exp(-\text{logit}(m))$, which is a smooth function of the conditional odds of non-censoring given X and $\boldsymbol{Z}$. The standard Cox analysis ignores this relationship by leaving the conditional censoring

* Corresponding author.
E-mail address: sundars@njit.edu (S. Subramanian).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.09.013

hazard unspecified. With SRC models, however, we exploit the link, using a model for m. Although the relationship explicitly calls for employing the logit, other choices such as the complementary log–log, generalized proportional hazards (GPH) and the Cauchy link may also be explored for m. In Section 3, the logit and Cauchy links are shown to provide improved estimator performance over standard Cox PH regression, with the Cauchy performing better than the logit in the sensitivity study. Furthermore, the SRC framework adapts to MCIs readily, unlike standard Cox analysis. We expect that in practice the added flexibility and improved performance would justify the additional effort required in the search for a good-fitting model for m.

Yuan [27] extended the Koziol–Green model [12] to the subject-specific setting implicit in Cox regression, which subsumes an earlier approach [23] as well. In terms of finite sample performance, both the proposed and Yuan [27] estimators perform equally well, see Section 3. Our proposed method, however, offers a more attractive alternative for the following reasons. Yuan [27] developed a log profile likelihood function which, however, involves the censoring indicator δ and hence would be inapplicable when there are MCIs. Furthermore, his approach requires simultaneous estimation of the finite dimensional components $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, compromising to some extent the simplicity of standard Cox regression analysis. Indeed, for a logistic model, Yuan's [27] approach will not be able to take advantage of the available logit function in statistical software. Our proposed method retains the simplicity of standard Cox regression and applies readily even when the MCIs are MAR. We show that $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}(t)$, the proposed estimators of $\boldsymbol{\beta}$ and $\Lambda_0(t)$ respectively, are each asymptotically as or more efficient than the standard Cox regression estimators.

Liu and Wang [14] proposed two estimators of $\boldsymbol{\beta}$ to account for the MCIs. They only focused on estimation of $\boldsymbol{\beta}$, which is a limitation. Their first estimator denoted $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$, was based on a mixture, that reduces to the Cox partial likelihood estimator when there are *no* MCIs – and therefore *less efficient* than our proposed estimator when there are no MCIs. Their second estimator requires computation of kernel estimates which would be inefficient due to curse of dimensionality and the need for data-based optimal bandwidths. Numerical studies reported in Section 3 reveal that $\hat{\boldsymbol{\beta}}$ performs as well as or better than $\hat{\boldsymbol{\beta}}_{\mathrm{LW}}$.

The article is organized as follows. In Section 2 we present our proposed estimators and provide theoretical comparisons with standard Cox regression estimators. We then present our proposed extension when there are MCIs. In Section 3, we present the results of simulation studies comparing the proposed and other approaches under discussion. We also provide illustrations using data from a heart transplant study and a study on recidivism, and another illustration using pseudo-real data. Technical details are given in the Appendix.

2. Proposed estimators and large sample results

Let $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{\Sigma}$. When there are no MCIs, we observe n independent and identically distributed triplets $(X_i, \delta_i, \boldsymbol{Z}_i)$, $i = 1, \ldots, n$, where $X = \min(T, C)$ is the minimum of the failure and censoring times, δ is the censoring indicator (1 when uncensored and 0 when censored), and $\boldsymbol{Z}$ denotes a $p \times 1$ vector of covariates. The conditional hazard function of the failure time given $\boldsymbol{Z}$ takes the form $\lambda(t|\boldsymbol{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}}$, where $\boldsymbol{\beta}$ is the $p \times 1$ regression parameter and $\lambda_0(t)$ is a baseline hazard function. Writing $N_i(t) = I(X_i \le t)$ and $Y_i(t) = I(X_i \ge t)$, $i = 1, \ldots, n$, the Cox partial likelihood estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{C}}$, solves $S_{\mathrm{C}}(\boldsymbol{\beta}) = 0$, where

$$S_{\mathrm{C}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} \delta_i \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j} \boldsymbol{Z}_j}{\sum_{j=1}^{n} Y_j(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_j}} \right] dN_i(t). \tag{2.1}$$

Breslow's [4] estimator of the baseline cumulative hazard function is given by

$$\hat{\Lambda}_{0\mathrm{C}}(t) \equiv \hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\mathrm{C}}) = \sum_{i=1}^{n} \int_0^{t} \frac{\delta_i}{\sum_{j=1}^{n} Y_j(s) e^{\hat{\boldsymbol{\beta}}_{\mathrm{C}}^T \boldsymbol{Z}_j}} dN_i(s). \tag{2.2}$$

Andersen and Gill [1] proved that $\hat{\boldsymbol{\beta}}_{\mathrm{C}} \xrightarrow{P} \boldsymbol{\beta}_0$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{C}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathrm{C}}^{-1})$, where Eq. (A.5) defines $\boldsymbol{\Sigma}_{\mathrm{C}}$. They also derived the weak convergence of $n^{1/2}(\hat{\Lambda}_{0\mathrm{C}}(t) - \Lambda_0(t))$, with the limiting variance function given by the first two terms of Eq. (A.45); see also [24].

2.1. Censoring indicators always observed

To tie the SRC framework to Cox regression, we write $m(X, \boldsymbol{Z}, \boldsymbol{\theta}_0) = P(\delta = 1|X, \boldsymbol{Z})$. The unknown $\boldsymbol{\theta} \in \mathbb{R}^k$, whose true value is $\boldsymbol{\theta}_0$, can be estimated by maximizing the quantity

$$\prod_{i=1}^{n} \{m(X_i, \boldsymbol{Z}_i, \boldsymbol{\theta})\}^{\delta_i} \{1 - m(X_i, \boldsymbol{Z}_i, \boldsymbol{\theta})\}^{1-\delta_i}. \tag{2.3}$$