Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2012 *Paper* 92

On penalized likelihood estimation for a non-proportional hazards regression model

Karthik Devarajan* Nader Ebrahimi†

*Fox Chase Cancer Center, karthik.devarajan@fccc.edu

†Northern Illinois University, nader@math.niu.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/cobra/art92

Copyright ©2012 by the authors.

On penalized likelihood estimation for a non-proportional hazards regression model

Karthik Devarajan and Nader Ebrahimi

Abstract

The fundamental assumption of proportionality of hazards in the Cox model sometimes does not hold in practice. In this paper, a semi-parametric generalization of the Cox model that permits crossing hazard curves is described. This model allows the interaction between covariates and the baseline hazard, and has been the subject of recent investigation. It includes, for the two sample problem, the case of two Weibull distributions and two extreme value distributions differing in both scale and shape parameters. The partial likelihood approach cannot be applied here to estimate the model parameters, and flexible methods based on splines and sieves for approximating the baseline hazard have been suggested. A theoretical framework for estimation in this generalized model is developed based on penalized likelihood methods. It is shown that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are exponential splines with knots at the unique failure times. Its relationship to prior computational approaches for this model is outlined.

On penalized likelihood estimation for a non-proportional hazards regression model

Karthik Devarajan [a,*], Nader Ebrahimi [b]

[a] *Biostatistics & Bioinformatics Department, Fox Chase Cancer Center, Philadelphia, PA 19111*

[b] *Division of Statistics, Northern Illinois University, DeKalb, IL 60115*

Abstract

The fundamental assumption of proportionality of hazards in the Cox model sometimes does not hold in practice. In this paper, a semi-parametric generalization of the Cox model that permits crossing hazard curves is described. This model allows the interaction between covariates and the baseline hazard, and has been the subject of recent investigation. It includes, for the two sample problem, the case of two Weibull distributions and two extreme value distributions differing in both scale and shape parameters. The partial likelihood approach cannot be applied here to estimate the model parameters, and flexible methods based on splines and sieves for approximating the baseline hazard have been suggested. A theoretical framework for estimation in this generalized model is developed based on penalized likelihood methods. It is shown that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are exponential splines with knots at the unique failure times. Its relationship to prior computational approaches for this model is outlined.

Key words: censored survival data analysis, crossing hazards, proportional hazards, penalty function, exponential spline, spline function, B-spline, knot, penalized likelihood, maximum likelihood, Weibull distribution, extreme value distribution.

* Corresponding author.
Email addresses: `karthik.devarajan@fccc.edu` (Karthik Devarajan), `nader@math.niu.edu` (Nader Ebrahimi).

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

1 Introduction

The primary goal in analyzing censored survival data is to assess the dependence of survival time on one or more explanatory variables or covariates. The Cox Proportional Hazards (PH) model (Cox, 1972) has become the standard tool for exploring this association. Given a vector of possibly time-dependent covariates $\mathbf{z}$, the hazard function at time t is assumed to be of the form $\lambda(t|\mathbf{z}) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{z}}$ where $\lambda_0(t)$ is the baseline hazard function, denoting the hazard under no covariate effect and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a p vector of regression coefficients. The focus is on inference for $\boldsymbol{\beta}$, with the baseline hazard function $\lambda_0(t)$, the non-parametric part, left completely unspecified. In spite of its attractive semi-parametric feature, the Cox PH model implicitly assumes that the hazard curves corresponding to two different values of the covariates do not cross. Although this assumption may be valid in many experimental settings, it has been found to be suspect in others. For example, if the treatment effect decreases with time, then one might expect the hazard curves corresponding to the treatment and control groups to converge. Other examples that indicate the presence of non-proportional hazards are detailed in Devarajan & Ebrahimi (2002, 2011), Wu & Hsieh (2009) and references therein.

In order to provide modeling flexibility that accounts for non-proportional hazards, several models have been proposed and investigated in the literature. In this paper, we describe one such semi-parametric generalization of the Cox PH model in which the hazard functions corresponding to different values of covariates can cross (see Devarajan & Ebrahimi, 2011 and references therein). The survival function corresponding to a covariate vector $\mathbf{z}$ is assumed to be of the form

$$S(t|\mathbf{z}) = \exp(\boldsymbol{\beta}'\mathbf{z})\{S_0(t)\}^{\exp(\boldsymbol{\gamma}'\mathbf{z})}, \tag{1.1}$$

where $S_0(t)$ is an arbitrary baseline survival function, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown p vectors of parameters. In terms of cumulative hazard functions, this model takes the form

$$\Lambda(t|\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}}\{\Lambda_0(t)\}^{\exp(\boldsymbol{\gamma}'\mathbf{z})}. \tag{1.2}$$

The Cox PH model is obtained as a special case of model (1.2) by setting $\boldsymbol{\gamma} = 0$. The conditional hazard function is

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\exp[(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z} + \{e^{\boldsymbol{\gamma}'\mathbf{z}} - 1\}\log\{\Lambda_0(t)\}]. \tag{1.3}$$

The hazards ratio corresponding to two different covariate vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ is

$$\frac{\lambda(t|\mathbf{z}_1)}{\lambda(t|\mathbf{z}_2)} = \exp\{(\boldsymbol{\beta}+\boldsymbol{\gamma})'(\mathbf{z}_1 - \mathbf{z}_2) + [e^{\boldsymbol{\gamma}'\mathbf{z}_1} - e^{\boldsymbol{\gamma}'\mathbf{z}_2}]\log[\Lambda_0(t)]\}. \tag{1.4}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

Since this ratio is a monotone function of t, the model allows the hazards ratio to invert over time. In other words, it allows crossing of hazard curves. For example, when treatment effect decreases or increases over time, model (1.2) can be applied. This model includes, for the two sample problem, the case of two Weibull and two extreme value distributions differing in both scale and shape parameters. An interesting property of this model is that the ratios of hazard to cumulative hazard functions corresponding to two covariate vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are proportional. Using (1.2) and (1.3), it can be shown that

$$\frac{\lambda(t|\mathbf{z})}{\Lambda(t|\mathbf{z})} = \frac{\lambda_0(t)}{\Lambda_0(t)} \exp(\boldsymbol{\gamma}'\mathbf{z}). \tag{1.5}$$

Devarajan & Ebrahimi (2011) discussed the unique properties of this model and provided an interpretation of it. In addition, they demonstrated its relationship to the time-dependent coefficient Cox PH model. For small $\boldsymbol{\gamma}$, equation (1.2) reduces to

$$\Lambda(t|\mathbf{z}) = e^{\boldsymbol{\beta}'\mathbf{z}}\{\Lambda_0(t)\}^{1+\boldsymbol{\gamma}'\mathbf{z}}. \tag{1.6}$$

For two different covariate vectors $\mathbf{z_1}$ and $\mathbf{z_2}$, $\log\{\frac{\Lambda(t|\mathbf{z_1})}{\Lambda(t|\mathbf{z_2})}\} = \boldsymbol{\eta}(t)(\mathbf{z_1} - \mathbf{z_2})$, where $\boldsymbol{\eta}(t) = [\boldsymbol{\beta} + \log\{\Lambda_0(t)\} \cdot \boldsymbol{\gamma}]'$. This is the Cox PH model with time-dependent coefficients $\boldsymbol{\eta}(t)$ that allows crossing hazard curves. Quantin et al. (1996) and Devarajan & Ebrahimi (2002) used model (1.2) for goodness of fit testing of the Cox PH model. Inference for model (1.2) has been discussed in Devarajan (2000), Devarajan & Ebrahimi (2011), Hsieh (2001) and Wu & Hsieh (2009).

In the following section, we develop a theoretical framework for estimation in this non-proportional hazards model based on penalized likelihood methods. We show that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are exponential splines with knots at the unique failure times. We conclude by briefly discussing the relationship of this approach to prior computational approaches for this model.

2 Maximum Penalized Likelihood Estimation for the Non-Proportional Hazards Regression Model

The partial likelihood approach of Cox (1972) is the standard inferential method for the Cox PH model. Due to the PH assumption, the baseline hazard $\lambda_0(t)$, drops out of this partial likelihood and can be left completely unspecified. On the other hand, it is evident from equations (1.3) and (1.4) that the generalized model (1.2) allows crossing hazard curves over time. Due to the non-constant hazards ratio, a factorization of the likelihood into a "partial" likelihood and a nuisance part is not possible. The baseline hazard must be specified in the likelihood or explicitly estimated. For example, Bordes & Breuils (2006) discussed sequential estimation for semi-parametric models with application to the PH model while Wu & Hsieh

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

(2009) adopted the method of sieves and used piecewise constants for estimating the baseline hazard in model (1.2). Devarajan (2000) and Devarajan & Ebrahimi (2011) harnessed the attractive computational properties of B-splines and used a set of cubic B-spline basis functions to model the baseline hazard. This spline-based approach can be viewed as a generalization of the piecewise constant approach. Here we develop a theoretical framework for estimating the baseline hazard using penalized likelihood methods. We show that the maximum penalized likelihood estimate of the baseline hazard and baseline cumulative hazard in model (1.2) are exponential splines with knots at the unique failure times. We briefly discuss the relationship of this approach to prior computational approaches for approximating the baseline hazard in this model.

The maximum penalized likelihood (MPL) approach applies the method of maximum likelihood to curve estimation. Often in survival analysis, one is interested in estimating the hazard function non-parametrically for a group of individuals in a study in the presence of censoring. In many situations, several other concomitant variables are also included in the study, thus giving rise to regression parameters associated with those variables that need to be estimated as well. First, we give a brief introduction to the concept of maximum penalized likelihood estimation in the context of hazard estimation.

The observed data consist of independent observations on the triple $(X, \delta, \mathbf{z})$, where X is the minimum of a failure and censoring time pair $(T, C), \delta = I(T \leq C)$ is the indicator of the event that a failure has been observed and $\mathbf{z} = (z_1, \ldots, z_p)'$ is a p vector of covariates. The random variables T and C denote the survival and censoring times respectively which are assumed to be independent. We are interested in estimating the hazard function in the presence of censoring. For the sake of simplicity, let us assume that there are no covariates. The inclusion of covariates is discussed later. The log-likelihood of a random sample of size n of censored survival data can be expressed in terms of the hazard function as

$$\ell[\lambda(.)] = \sum_{i=1}^{n} \left\{ \delta_i \log \lambda(t_i) - \int_0^{t_i} \lambda(u) du \right\} \tag{2.1}$$

where $\lambda(t)$ is the hazard function and δ is as defined above. The subscript refers to the ith individual in the sample for $i = 1, ..., n$. An application of the maximum likelihood method to the log-likelihood function given above results in a sum of Dirac delta functions with spikes at the uncensored observations as the hazard estimate. This is due to the fact that the log-likelihood $\ell[.]$ as a functional of $\lambda(.)$ is unbounded from above and hence cannot be maximized with respect to $\lambda(t)$. Good and Gaskins (1971) introduced the idea of non-parametric roughness penalties and showed how to impose a smoothness constraint on $\ell[.]$ by maximizing a penalized version of the log-likelihood to obtain a bounded estimate. The penalized log-likelihood is defined as

$$J[\lambda(t)] = \ell[\lambda(t)] - \alpha R[\lambda(t)]$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

$$= \sum_{i=1}^{n} \left\{ \delta_i \log \lambda(t_i) - \int_0^t \lambda(u)du \right\} - \alpha R[\lambda(t)] \tag{2.2}$$

where $\alpha > 0$ is the *smooothing parameter* that controls the amount of smoothing and $R[\lambda(t)]$ is the *roughness penalty*. If S denotes the class of all sufficiently smooth functions $\lambda(t)$ such that $R[\lambda(t)]$ is defined and for which $\lambda(t) \geq 0, \ \ \forall \ \ t > 0$, then the hazard estimate $\widehat{\lambda(t)}$ is a *maximum penalized likelihood estimate* if $\widehat{\lambda(t)} \in S$ and $J[\widehat{\lambda(t)}] \geq J[\lambda(t)], \ \ \forall \ \ \lambda(t) \in S$. That is, $\widehat{\lambda(t)}$ maximizes the penalized log-likelihood $J[\lambda(t)]$ over the class of all $\lambda(t)$ satisfying the conditions given above.

An excellent exposition of penalized likelihood methods is given in Tapia & Thompson (1978) and Silverman (1986). Cox & O'Sullivan (1990) discuss the asymptotic properties of penalized likelihood estimates. Penalized likelihood methods have been applied for hazard estimation by Gu (1996), O'Sullivan (1988), Anderson & Senthilselvan (1980), Bartoszynski et al (1981), among others. The most general form of the penalty functional $R[\lambda(t)]$ is

$$R[\lambda(t)] = \int_0^\infty \left[\lambda^{(m)}(x)\right]^2 dx, \quad m > 0$$

where $\lambda^{(m)}(t)$ is the mth derivative of $\lambda(t)$. The commonly used forms for $R[\lambda(t)]$ are $\int_0^\infty [\lambda^{(1)}(x)]^2 dx$ which penalizes the slope of the curve and $\int_0^\infty [\lambda^{(2)}(x)]^2 dx$ which penalizes the curvature. Sometimes, it is convenient to work with the logarithm or the square root of the hazard to avoid explicit positivity constraints on the hazard. The functional (2.2) is maximized with respect to $\lambda(t)$ over the class of absolutely continuous functions on $(0, \infty)$ whose m^{th} derivative is square integrable. As noted earlier, the solution corresponding to $\alpha = 0$ is the sum of Dirac delta function spikes at the uncensored observations. As $\alpha \to \infty$, the solution tends to an $(m-1)^{th}$ degree polynomial. O'Sullivan (1988) notes that for $m = 2$, the log-hazard estimator is linear in t and has a Weibull form on the logarithmic scale.

Next, we describe estimation of the underlying baseline quantities of interest (such as the hazard, cumulative hazard and their ratio) in models (1.2) and (1.6) using MPL methods.

2.1 MPL Estimation of the Baseline Hazard

We obtain the MPL estimate of the baseline hazard function, $\lambda_0(t)$, in model (1.2) assuming that the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are known. In order to avoid explicit positivity constraints, we estimate the square root of the hazard function. That is, *given* $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, we obtain the maximum penalized likelihood estimate of the function $g(t) = \sqrt{\lambda_0(t)}$ using calculus of variations (Reinsch, 1967).

Using (2.1), (1.2) and (1.3), the log-likelihood for this model is given by

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$$\ell[\lambda(.)] = \sum_{i=1}^{n} \delta_i \log \lambda(t_i|\mathbf{z}_i) - \int_0^{t_i} \lambda(y|\mathbf{z}_i)dy$$
$$= \sum_{i=1}^{n} \left\{ \delta_i \left[\log \lambda_0(t_i) + (\boldsymbol{\beta} + \boldsymbol{\gamma})'\mathbf{z}_i + (e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} - 1) \cdot \log \Lambda_0(t_i) \right] \right\} -$$
$$\sum_{i=1}^{n} \left\{ [\Lambda_0(t_i)]^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i)} e^{\boldsymbol{\beta}'\boldsymbol{z}_i} \right\} \tag{2.3}$$

Using the log-likelihood given by (2.3) and the penalty functional

$$R[\lambda(t)] = \alpha \int_0^{\infty} [g^{(1)}(u)]^2$$

where $\alpha > 0$ is the smoothing parameter, the penalized log-likelihood can be written in terms of $g(t) = \sqrt{\lambda_0(t)}$ as

$$\ell[g(t)] = \sum_{i=1}^{n} \left\{ 2\delta_i \log g(t_i) + (\boldsymbol{\beta} + \boldsymbol{\gamma})'\mathbf{z}_i - e^{\boldsymbol{\beta}'\boldsymbol{z}_i} \left(\int_0^{t_i} g^2(u)du \right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i)} \right.$$
$$\left. + (e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} - 1) \log \left(\int_0^{t_i} g^2(u)du \right) \right\} - \left\{ \alpha \int_0^{\infty} [g'(u)]^2 du \right\}. \tag{2.4}$$

We need to maximize this penalized log-likelihood with respect to the function $g(t)$. The following theorem guarantees the existence of the maximum penalized likelihood estimator of the functional $g(t) = \sqrt{\lambda_0(t)}$ in (2.4) (Cox & O'Sullivan, 1990).

Theorem 1 *If there exists a unique $(m-1)^{th}$ degree polynomial minimizing the likelihood part of (2.4), then the penalized likelihood in (2.4) has a unique maxima over the Sobolev space $W_2{}^m[a,b]$ of absolutely continuous functions whose m^{th} derivative is square integrable.*

First, we define the exponential spline and then state a key result from Hewitt & Stromberg (1975) used in obtaining the maximum penalized likelihood estimate.

Definition: An *exponential spline* with knots at $t_1 < t_2 < ... < t_n$ is any function f of the form

$$f(t) = \sum_{i=1}^{n} \alpha_i e^{t-t_i}$$

for some real spline coefficients $\alpha_i, i = 1, 2, ..., n$.

Lemma: For continuously differentiable functions f and g, we have,

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

$$\int_a^b f(u)g(u)du = f(b-)g(b-) - f(a+)g(a+) - \int_a^b g(u)f'(u)du \tag{2.5}$$

Theorem 1 guarantees the existence of the maximum penalized likelihood estimator of the functional $g(t) = \sqrt{\lambda_0(t)}$ in (2.4). (Cox & O'Sullivan, 1990). We now state our result in the following theorem.

Theorem 2 *Given the penalized log-likelihood (2.4) and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the maximum penalized likelihood of the function $g(t) = \sqrt{\lambda_0(t)}$ is an exponential spline with knots at the unique failure times if $\int_0^x g^2(u)du$ is finite for every x.*

Proof: Consider a small increment $\epsilon \cdot \eta(t), \quad \epsilon > 0$ to the function $g(t)$ in the class $C_2(I)$ of functions where $I = [t_0, t_n]$ with $t_0 \equiv 0$. A necessary condition for the functional $\ell[\cdot]$ to have a maximum is that as $\epsilon \to 0$,

$$\frac{\partial}{\partial \epsilon} \ell[g(u) + \epsilon \cdot \eta(u)] = 0$$

Now, $\frac{\partial}{\partial \epsilon} \ell[g(u) + \epsilon \cdot \eta(u)] = 0$ gives

$$\frac{\partial}{\partial \epsilon} \left[\sum_{i=1}^n \left\{ 2\delta_i \log\{g(t_i) + \epsilon \cdot \eta(t_i)\} + (\boldsymbol{\beta} + \boldsymbol{\gamma})'\mathbf{z}_i - e^{\boldsymbol{\beta}'\boldsymbol{z}_i} \left(\int_0^{t_i} [g(u) + \epsilon \cdot \eta(u)]^2 du \right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i)} \right.\right.$$

$$\left.\left. + (e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} - 1) \log \left(\int_0^{t_i} [g(u) + \epsilon \cdot \eta(u)]^2 du \right) \right\} - \left\{ \alpha \int_0^\infty [g'(u) + \epsilon \cdot \eta'(u)]^2 du \right\} \right] = 0$$

which implies that

$$0 = \sum_{i=1}^n \left\{ 2\delta_i \frac{\eta(t_i)}{g(t_i)} - \left\{ e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_i} \right\} \left(\int_0^{t_i} g^2(u)du \right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i)-1} \left(\int_0^{t_i} 2g(u) \cdot \eta(u)du \right) \right.$$

$$\left. + (e^{\boldsymbol{\gamma}'\mathbf{z}_i} - 1) \left(\frac{\int_0^{t_i} 2g(u) \cdot \eta(u)du}{\int_0^{t_i} g^2(u)du} \right) \right\} - \left\{ 2\alpha \int_0^\infty g'(u) \cdot \eta'(u)du \right\} \tag{2.6}$$

The last term of equation (2.6) can be rewritten using the Lemma (2.5) and integration by parts as follows.

$$\int_0^\infty g'(u) \cdot \eta'(u)du = \int_{t_0}^{t_n} g'(u) \cdot \eta'(u)du + \int_{[0,\infty)/[t_0,t_n]} g'(u) \cdot \eta'(u)du$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$$= \sum_{i=1}^{n} \left\{ \int_{t_{i-1}}^{t_i} g'(u) \cdot \eta'(u) du \right\} + \int_{[0,\infty)/[t_0,t_n]} g'(u) \cdot \eta'(u) du \tag{2.7}$$

Integrating the first term on the right hand side of equation (2.7) by parts and using equation (2.5), we get,

$$\int_0^\infty g'(u) \cdot \eta'(u) du = \sum_{i=1}^{n} [g'(t_i^-) - g'(t_i^+)]\eta(t_i) - g'(t_0^+)\eta(t_0) + g'(t_m^+)\eta(t_n) - \sum_{i=1}^{n} \left\{ \int_{t_{i-1}}^{t_i} g''(u) \cdot \eta(u) du \right\} + \int_{[0,\infty)/[t_0,t_n]} g'(u) \cdot \eta'(u) du \tag{2.8}$$

Using (2.8), (2.6) becomes

$$\sum_{i=1}^{n} \left\{ 2\delta_i \frac{\eta(t_i)}{g(t_i)} - \left\{ e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_i} \right\} \left(\int_0^{t_i} g^2(u) du \right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_i)-1} \left(\int_0^{t_i} 2g(u) \cdot \eta(u) du \right) \right.$$

$$\left. + (e^{\boldsymbol{\gamma}'\mathbf{z}_i} - 1) \left(\frac{\int_0^{t_i} 2g(u) \cdot \eta(u) du}{\int_0^{t_i} g^2(u) du} \right) \right\} - 2\alpha \left\{ \sum_{i=1}^{n} [g'(t_i^-) - g'(t_i^+)]\eta(t_i) - g'(t_0^+)\eta(t_0) \right.$$

$$\left. + g'(t_m^+)\eta(t_n) - \sum_{i=1}^{n} \left\{ \int_{t_{i-1}}^{t_i} g''(u) \cdot \eta(u) du \right\} + \int_{[0,\infty)/[t_0,t_n]} g'(u) \cdot \eta'(u) du \right\} = 0$$

Now, collecting terms with $\int a(u)\eta(u)du$, $\eta(u)$ etc. and simplifying, we have,

$$\sum_{i=1}^{n} \left\{ \int_{t_{i-1}}^{t_i} \left(\sum_{l=i}^{m} \left(\frac{-e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_l} \left(\int_0^{t_l} g^2(u) du \right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_l)} + (e^{\boldsymbol{\gamma}'\boldsymbol{z}_l} - 1)}{\int_0^{t_l} g^2(u) du} \right) 2g(u) + 2\alpha \cdot g''(u) \right) \eta(u) du \right\}$$

$$+ \sum_{i=1}^{n} \left\{ \frac{2\delta_i}{g(t_i)} - 2\alpha[g'(t_i^-) - g'(t_i^+)] \right\} \eta(t_i) - 2\alpha g'(t_m^+)\eta(t_m) + 2\alpha g'(t_0^+)\eta(t_0)$$

$$-2\alpha \int_{[0,\infty)/[t_0,t_n]} g'(u) \cdot \eta'(u) du = 0$$

Since, $\eta(.)$ is an arbitrary function in $C_2(I)$, by suitable choice of $\eta(.)$, we can show that each

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

term in the above equation which is a coefficient of $\eta(.)$ or $\eta'(.)$ is identically zero. Hence, we obtain a second order ordinary linear differential equation given by

$$0 = \alpha g''(u) + g(u)\left\{\sum_{l=i}^{m}\left(\frac{-e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_l}\left(\int_0^{t_l} g^2(u)du\right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_l)} + (e^{\boldsymbol{\gamma}'\boldsymbol{z}_l} - 1)}{\int_0^{t_l} g^2(u)du}\right)\right\} \tag{2.9}$$

and the following side conditions

$$g'(t_i^-) - g'(t_i^+) = \frac{\delta_i}{\alpha g(t_i)}, \quad i = 1, ..., n$$

$$g'(t_m^+) = g'(t_0^+) = 0$$

and

$$g'(u) = 0, \quad u \in [0, \infty)/[t_0, t_n]$$

Since the function g exists, assuming that the definite integral $\int_0^{t_l} g^2(u)du$ in (2.9) is finite, the solution to the differential equation

$$g''(u) - \frac{1}{\alpha}\left\{\sum_{l=i}^{m}\left(\frac{e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_l}\left(\int_0^{t_l} g^2(u)du\right)^{\exp(\boldsymbol{\gamma}'\boldsymbol{z}_l)} - (e^{\boldsymbol{\gamma}'\boldsymbol{z}_l} - 1)}{\int_0^{t_l} g^2(u)du}\right)\right\} g(u) = 0$$

is given by

$$g(t) = a_1 \cdot \exp(\sqrt{\alpha} \cdot t) + a_2 \cdot \exp(-\sqrt{\alpha} \cdot t), \quad t_{i-1} < t < t_i, \quad i = 1, ..., n \tag{2.10}$$

where a_1 and a_2 are arbitrary constants (Boyce & DiPrima, 1969). Hence $g(t) = \sqrt{\lambda_0(t)}$ is an exponential spline with knots at the unique failure times. From (2.10), we see that $g^2(t) = \lambda_0(t)$ and $\Lambda_0(t) = \int_0^t \lambda_0(x)dx$ are also exponential splines with knots at the unique failure times.

2.2 MPL Estimation of the Ratio $\frac{\lambda_0(t)}{\Lambda_0(t)}$

For two covariate vectors $\mathbf{z_1}$ and $\mathbf{z_2}$, the ratios of the hazard to cumulative hazard functions are proportional not only in model (1.2) but also in the time-dependent coefficient Cox model (1.6). We utilize this property by working with the baseline ratio $\frac{\lambda_0(t)}{\Lambda_0(t)}$ and develop an approach to estimation in these models. The function $g(t) = \frac{\lambda_0(t)}{\Lambda_0(t)}$ is estimated non-parametrically assuming that the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are known. That is, *given* $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, we obtain the maximum penalized likelihood estimate of the functional $g(t)$ using similar methods as in §2.1.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

We penalize for both slope and curvature of the function $g(t)$ and use the penalty functional

$$R[g(x)] = \alpha_1 \int_0^\infty [g^{(1)}(x)]^2 + \alpha_2 \int_0^\infty [g^{(2)}(x)]^2 \tag{2.11}$$

where $\alpha_1 > 0$ and $\alpha_2 > 0$ are smoothing parameters. Using the log-likelihood (2.3) and (2.11), the penalized log-likelihood can be written in terms of $g(t) = \frac{\lambda_0(t)}{\Lambda_0(t)}$ as

$$\begin{aligned}\ell[g(t)] = &\sum_{i=1}^n \left\{ \delta_i \log g(t_i) + (\boldsymbol{\beta} + \boldsymbol{\gamma})'\mathbf{z}_i - e^{\boldsymbol{\beta}'\boldsymbol{z}_i} \exp\left[e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} G(t_i)\right] + e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} G(t_i) \right\} \\ &- \left\{ \alpha_1 \int_0^\infty [g'(u)]^2 du + \alpha_2 \int_0^\infty [g''(u)]^2 du \right\}\end{aligned} \tag{2.12}$$

where $G(t) = \int_0^t g(u)du$. Here and elsewhere, $g'(u) \equiv g^{(1)}(u)$ and $g''(u) \equiv g^{(2)}(u)$. Using the linear approximation $e^x \approx 1 + x$, we rewrite the third term in (2.12) as

$$e^{\boldsymbol{\beta}'\boldsymbol{z}_i} \exp\left[e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} G(t_i)\right] \approx e^{\boldsymbol{\beta}'\boldsymbol{z}_i} \left[1 + e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} G(t_i)\right],$$

and the penalized log-likelihood becomes

$$\begin{aligned}\ell[g(t)] \approx &\sum_{i=1}^n \left\{ \delta_i \log g(t_i) + (\boldsymbol{\beta} + \boldsymbol{\gamma})'\mathbf{z}_i - e^{\boldsymbol{\beta}'\boldsymbol{z}_i} - e^{\boldsymbol{\gamma}'\boldsymbol{z}_i} \left(e^{\boldsymbol{\beta}'\boldsymbol{z}_i} - 1\right) G(t_i) \right\} \\ &- \left\{ \alpha_1 \int_0^\infty [g'(u)]^2 du + \alpha_2 \int_0^\infty [g''(u)]^2 du \right\}\end{aligned} \tag{2.13}$$

This penalized log-likelihood is maximized with respect to the function $g(t)$. Theorem 1 guarantees the existence of the maximum penalized likelihood estimator of the functional $g(t) = \frac{\lambda_0(t)}{\Lambda_0(t)}$ in (2.13) (Cox & O'Sullivan, 1990). We now state our result in the following theorem.

Theorem 3 *Given the penalized log-likelihood (2.13) and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the maximum penalized likelihood estimate of the functional $g(t) = \frac{\lambda_0(t)}{\Lambda_0(t)}$ of $\lambda_0(t)$ is an exponential spline with knots at the distinct failure times.*

Proof: Consider a small increment $\epsilon \cdot \eta(t)$, $\epsilon > 0$ to the function $g(t)$ in the class $C_2(I)$ of functions where $I = [t_0, t_n]$ with $t_0 \equiv 0$. A necessary condition for the functional $\ell[\cdot]$ to have a maximum is that as $\epsilon \to 0$,

$$\frac{\partial}{\partial \epsilon} \ell[g(u) + \epsilon \cdot \eta(u)] = 0$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

Now, $\frac{\partial}{\partial\epsilon}\ell[g(u)+\epsilon\cdot\eta(u)]=0$ implies that

$$\frac{\partial}{\partial\epsilon}\left[\sum_{i=1}^{n}\left\{\delta_i\log\{g(t_i)+\epsilon\cdot\eta(t_i)\}+(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_i-e^{\boldsymbol{\beta}'\boldsymbol{z}_i}\right.\right.$$

$$\left.+\left\{-e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_i}+e^{\boldsymbol{\gamma}'\boldsymbol{z}_i}\right\}\int_0^{t_i}[g(u)+\epsilon\cdot\eta(u)]\,du\right\}$$

$$\left.-\left\{\alpha_1\int_0^{\infty}[g'(u)+\epsilon\cdot\eta'(u)]^2du+\alpha_2\int_0^{\infty}[g''(u)+\epsilon\cdot\eta''(u)]^2du\right\}\right]=0$$

which implies that

$$0=\sum_{i=1}^{n}\left\{\delta_i\frac{\eta(t_i)}{g(t_i)}+\left\{-e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_i}+e^{\boldsymbol{\gamma}'\boldsymbol{z}_i}\right\}\int_0^{t_i}\eta(u)du\right\}-$$

$$\left\{2\alpha_1\int_0^{\infty}g'(u)\cdot\eta'(u)du+2\alpha_2\int_0^{\infty}g''(u)\cdot\eta''(u)du\right\} \tag{2.14}$$

The last two terms of the above equation are rewritten using (2.5) and (2.8). Integrating the last term of equation (2.14), we have,

$$\int_0^{\infty}g''(u)\cdot\eta''(u)du=\int_{t_0}^{t_n}g''(u)\cdot\eta''(u)du+\int_{[0,\infty)/[t_0,t_n]}g''(u)\cdot\eta''(u)du$$

$$=\sum_{i=1}^{n}\left\{\int_{t_{i-1}}^{t_i}g''(u)\cdot\eta''(u)du\right\}+$$

$$\int_{[0,\infty)/[t_0,t_n]}g''(u)\cdot\eta''(u)du \tag{2.15}$$

Again, integrating the first term on the right hand side of equation (2.15) by parts and using equation (2.5), we get,

$$\int_0^{\infty}g''(u)\cdot\eta''(u)du=\sum_{i=1}^{n}[g''(t_i^-)-g''(t_i^+)]\eta(t_i)-g''(t_0^+)\eta(t_0)+g''(t_m^+)\eta(t_n)-$$

$$\sum_{i=1}^{n}[g'''(t_i^-)-g'''(t_i^+)]\eta(t_i)-g'''(t_0^+)\eta(t_0)+g'''(t_m^+)\eta(t_n)+$$

$$\sum_{i=1}^{n}\left\{\int_{t_{i-1}}^{t_i}g''''(u)\cdot\eta(u)du\right\}+ \tag{2.16}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$$\int\limits_{[0,\infty)/[t_0,t_n]} g''(u)\cdot\eta''(u)du$$

Using (2.8) and (2.16), equation (2.14) becomes,

$$\sum_{i=1}^{n}\left\{\delta_i\frac{\eta(t_i)}{g(t_i)}-\left\{e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_i}-e^{\boldsymbol{\gamma}'\boldsymbol{z}_i}\right\}\int\limits_0^{t_i}\eta(u)du\right\}-2\alpha_1\sum_{i=1}^{n}\left[g'(t_i^-)-g'(t_i^+)\right]\eta(t_i)$$

$$+2\alpha_1 g'(t_0^+)\eta(t_0)-\ 2\alpha_1 g'(t_m^+)\eta(t_n)\ -\ 2\alpha_1\int\limits_{[0,\infty)/[t_0,t_n]} g'(u)\cdot\eta'(u)du$$

$$+2\alpha_1\sum_{i=1}^{n}\int\limits_{t_{i-1}}^{t_i} g''(u)\cdot\eta(u)du-\ 2\alpha_2\sum_{i=1}^{n}[g''(t_i^-)-g''(t_i^+)]\eta(t_i)+\ 2\alpha_2 g''(t_0^+)\eta(t_0)$$

$$-2\alpha_2 g''(t_m^+)\eta(t_n)\ +2\alpha_2\sum_{i=1}^{n}[g'''(t_i^-)-g'''(t_i^+)]\eta(t_i)\ +\ 2\alpha_2 g'''(t_0^+)\eta(t_0)$$

$$-2\alpha_2 g'''(t_m^+)\eta(t_n)-\ 2\alpha_2\int\limits_{[0,\infty)/[t_0,t_n]} g''(u)\cdot\eta''(u)du\ -\ 2\alpha_2\sum_{i=1}^{n}\int\limits_{t_{i-1}}^{t_i} g''''(u)\cdot\eta(u)du=0$$

Now, collecting terms with $\int a(u)\eta(u)du$, $\eta(u)$ etc. and simplifying, we have,

$$\sum_{i=1}^{n}\left\{\int\limits_{t_{i-1}}^{t_i}\left(\sum_{l=i}^{m}\left\{-e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_l}+e^{\boldsymbol{\gamma}'\boldsymbol{z}_l}\right\}+2\alpha_1\cdot g''(u)-2\alpha_2\cdot g''''(u)\right)\eta(u)du\right\}$$

$$+\sum_{i=1}^{n}\left\{\frac{\delta_i}{g(t_i)}+2\alpha_2[g'''(t_i^-)-g'''(t_i^+)]-2\alpha_2[g''(t_i^-)-g''(t_i^+)]-2\alpha_1[g'(t_i^-)-g'(t_i^+)]\right\}\eta(t_i)$$

$$+\left\{2\alpha_2 g''(t_0^+)+2\alpha_2 g'''(t_0^+)+2\alpha_1 g'(t_0^+)\right\}\eta(t_0)$$

$$-\left\{2\alpha_1 g'(t_m^+)+\ 2\alpha_1 g''(t_m^+)+\ 2\alpha_2 g'''(t_m^+)\right\}\eta(t_n)$$

$$-2\alpha_1\int\limits_{[0,\infty)/[t_0,t_n]} g'(u)\cdot\eta'(u)du-\ 2\alpha_2\int\limits_{[0,\infty)/[t_0,t_n]} g''(u)\cdot\eta''(u)du=0$$

Since, $\eta(.)$ is an arbitrary function in $C_2(I)$, by suitable choice of $\eta(.)$, we can show that each term in the above equation which is a coefficient of $\eta(.)$ or $\eta'(.)$ or $\eta''(.)$ is identically zero. Hence, we obtain a fourth order ordinary linear differential equation given by

$$2\alpha_2\cdot g''''(u)-2\alpha_1\cdot g''(u)=\sum_{l=i}^{m}\left\{-e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_l}+e^{\boldsymbol{\gamma}'\boldsymbol{z}_l}\right\}$$

and the following side conditions

$$\alpha_1[g'(t_i^-)-g'(t_i^+)]+\alpha_2[g''(t_i^-)-g''(t_i^+)-g'''(t_i^-)+g'''(t_i^+)]=\frac{\delta_i}{2g(t_i)},\quad i=1,...,n$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

$$g'(t_0^+) = g'(t_m^+) = g''(t_0^+) = g''(t_m^+) = g'''(t_0^+) = g''(t_m^+) = 0$$

$$g'(u) = 0, \quad u \in [0, \infty)/[t_0, t_n]$$

and

$$g''(u) = 0, \quad u \in [0, \infty)/[t_0, t_n]$$

The solution to the differential equation

$$g''''(u) - \frac{\alpha_1}{\alpha_2} g''(u) = \frac{1}{2\alpha_2} \sum_{l=i}^{m} \left\{ -e^{(\boldsymbol{\beta}+\boldsymbol{\gamma})'\mathbf{z}_l} + e^{\boldsymbol{\gamma}'\boldsymbol{z}_l} \right\}$$

is given by

$$g(t) = c_1 \cdot \exp\left(\sqrt{\frac{\alpha_1}{\alpha_2}} \cdot t\right) + c_2 \cdot \exp\left(-\sqrt{\frac{\alpha_1}{\alpha_2}} \cdot t\right), \quad t_{i-1} < t < t_i, \quad i = 1, 2, ..., n$$

where c_1 and c_2 are arbitrary constants (Boyce & DiPrima, 1969). Hence $g(t) = \frac{\lambda_0(t)}{\Lambda_0(t)}$ is an exponential spline with knots at the unique failure times.

3 Summary and Discussion

In summary, we developed a theoretical framework for maximum penalized likelihood estimation for a semi-parametric generalization of the Cox PH model. We showed that the optimal solution to the baseline hazard, baseline cumulative hazard and their ratio are exponential splines with knots at the unique failure times. From a computational perspective, models based on splines are intermediate in structure between completely parametric models and non-parametric methods, and provide flexibility in incorporating a variety of shapes. It is well-known that any spline function of a given order and pre-specified knot sequence can be approximated by a linear combination of B-splines for that knot sequence. Moreover, an interesting feature of the B-spline basis function is that it has the properties of a probability density function. So a linear combination of B-spline basis functions is like a mixture of densities that allows for various distributions (de Boor, 2001). A full likelihood approach based on a linear combination of B-splines for the baseline hazard has already been outlined in our earlier work (Devarajan & Ebrahimi, 2011) and would aid in the implementation of the proposed method. Simultaneous estimation of the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the baseline hazard can be achieved via an iterative scheme. The testing and comparison of the performance of this approach to existing methods of estimation for this generalized model using real data are especially important and could form the core of a future, larger investigation.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

4 Acknowledgement

The work of the first author was supported in part by NIH grant P30 CA 06927.

References

[1] Anderson, J.A. and A. Senthilselvan *(1980). Smooth Estimates for the Hazard Function, Jorunal of the Royal Statistical Society-Series B, 42, 322-327.*

[2] Bartoszynski, R., Brown, B.W., McBride, C.M. and J.R. Thompson *(1981). Some Nonparametric Techniques for Estimating the Intensity Function of a Cancer Related Nonstationary Poisson Process, Annals of Statistics, 9, 1050-1060.*

[3] de Boor, C. *(2001). A Practical Guide to Splines, Springer-Verlag, New York.*

[4] Bordes, L. and Breuils, C. (2006). Sequential estimation for semiparametric models with application to the proportional hazards model, *Journal of Statistical Planning and Inference*, **136**, 3735-3759.

[5] Boyce, W. and R.C. Diprima *(2008). Elementary Differential Equations and Boundary Value Problems, Ninth Edition, John Wiley & Sons, Inc., New York*

[6] Cox, D.R. (1972). Regression Models and Life Tables (with Discussion), *Journal of the Royal Statistical Society-Series B*, **34**, 187-202.

[7] Cox, D.D. and F. O'Sullivan *(1990). Asymptotic Analysis of Penalized Likelihood and Related Estimators, Annals of Statistics, 18, 1676-1695.*

[8] Devarajan, K. (2000). *Inference for a non-proportional hazards regression model and applications.* Ph.D dissertation, Division of Statistics, Northern Illinois University.

[9] Devarajan, K. and N. Ebrahimi (2011). *A Semi-parametric Generalization of the Cox Proportional Hazards Regression Model: Inference and Applications, Computational Statistics & Data Analysis, 55(1):667-676.*

[10] Devarajan, K. and Ebrahimi, N. (2002). Goodness-of-Fit Testing for the Cox Proportional Hazards Model. In Huber, C., Balakrishnan, N., Nikulin, M.S., & Mesbah, M.(Eds.), *Goodness-of-Fit Tests and Model Validity* (237-254). Birkhauser, Boston.

[11] Good, I.J. and R.A. Gaskins *(1971). Non-parametric Roughness Penalties for Probability Densities, Biometrika, 58, 255-277.*

[12] Gu, C. *(1996). Penalized Likelihood Hazard Estimation: A General Procedure, Statistica Sinica, 6, 861-876.*

[13] Hewitt, A. and P. Stromberg *(1975). Real & Abstract Analysis, Springer, New York.*

[14] Hsieh, F. (2001). On heteroscedastic hazards regression models: theory and application, *Journal of the Royal Statistical Society-Series B*, **63(1)**, 63-80.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/cobra/art92

[15] O'Sullivan, F. (1988). Fast Computation of Fully Automated Log-Density and Log-Hazard Estimators, *SIAM Journal of Scientific Statistical Computing*, **9**, 363-379.

[16] Quantin, C., Moureau, T., Asselain, B., Maccario, J. and J. Lellouch (1996). A Regression Model for Testing the Proportional Hazards Hypotheses, *Biometrics*, **52**, 874-885.

[17] Reinsch, C.H. *(1967). Smoothing by Spline Functions, Numerische Mathematik, 10, 1977-1983.*

[18] Silverman, B.W. *(1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.*

[19] Tapia, R.A. and J.R. Thompson *(1978). Nonparametric Probability Density Estimation, Johns Hopkins University Press, Baltimore, MD, 1978.*

[20] Wu, H-D.I., Hsieh, F. (2009). Heterogeneity and varying effect in hazards regression, *Journal of Statistical Planning and Inference*, 139, 4213-4222.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press