On Bayesian Analysis of the Proportional Hazards Model

Sull’Analisi Bayesiana del Modello a Rischi Proporzionali

Luca La Rocca
Università di Modena e Reggio Emilia
Dipartimento di Scienze Sociali, Cognitive e Quantitative
Via G. Giglioli Valle 9, 42100 Reggio Emilia
The proportional hazards model

It is a well known model for regression in survival analysis, introduced by Cox (1972), in which

\( T_1, \ldots, T_N \) are the survival times of interest

\( x_1, \ldots, x_N \) are the corresponding vectors of covariates

the unknown distribution of \( T_i \) is described through its hazard rate

\[
\rho_i(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}(t \leq T_i \leq t + h \mid T_i \geq t), \quad t \in \mathbb{R}_+, \quad i = 1 \ldots N
\]

and it is assumed that this can be factored as

\[
\rho_i = e^{\langle \beta, x_i \rangle} \rho_\star, \quad i = 1 \ldots N
\]

where \( e \) is the basis of the natural logarithms and \( \langle \cdot, \cdot \rangle \) denotes the ordinary scalar product, say in \( \mathbb{R}^p \).
Analysis of the proportional hazards model

Interest lies in estimating both

- the vector of regression parameters $\beta$
- and the so-called baseline hazard rate $\rho_*$

from possibly right censored observations, that is having observed an event of the form

$$\{T_1 = t_1, T_2 > t_2\}$$

where, for simplicity, the case $N = 2$ has been considered. It is assumed that the censoring mechanism be non-informative.
Bayesian analysis of the proportional hazards model

First, a joint prior distribution on $\beta$ and $\rho_\star$ needs to be elicited.

This can be done by building a stochastic process $\rho_\star$ such that

$$\rho_\star \geq 0, \quad \int_0^t \rho_\star(s)ds < \infty, \quad \int_0^\infty \rho_\star(s)ds = \infty$$

together with a random vector $\beta$ on a suitable probability space.

Then, the corresponding posterior distribution has to be computed.

The Bayes formula based on the standard likelihood

$$L(t \mid o, x; \rho_\star, \beta) = \prod_{i=1}^{N} \left[ e^{<\beta, x_i>} \rho_\star(t_i) \right]^{o_i} \exp \left\{ -e^{<\beta, x_i>} \int_0^{t_i} \rho_\star(s) ds \right\}$$

where $o_i = 1$, if $t_i$ is exact, and $o_i = 0$, if $t_i$ is right censored, can be approximated by means of ad hoc MCMC techniques.
Proportional hazards without the hazard rate

An alternative definition of the proportional hazards model is given by the formula

\[ \log \Sigma_i(t) = e^{\langle \beta, x_i \rangle} \log \Sigma_*(t), \quad t \in \mathbb{R}_+ \]

which relates the unknown survival function \( \Sigma_i \) of the \( i \)-th survival time to the baseline survival function \( \Sigma_* \).

Note that the above formula does not require the hazard rate to be defined.

- Kalbfleisch (1978) built \( -\log \Sigma_* \) as a gamma process and estimated \( \beta \) by maximizing its marginal likelihood

- Hjort (1990) built \( -\int_{[0, \cdot]} \Sigma_*^{-1}(t) \Sigma_*(dt) \) as a beta process and suggested simulation techniques alternatively to the empirical Bayes approach
Building the prior hazard rate

An infinitely smooth possibility (La Rocca, 2003) is to take

$$\rho_*(t) = q[1 - K(t)]\xi_0 + \sum_{j=1}^{\infty} \xi_j k(t - \sigma_j), \quad t \in \mathbb{R}_+$$

where

$$\xi_0, \xi_1, \xi_2, \ldots \overset{i.i.d.}{\sim} \mathcal{G}(a, b), \ a > 0, \ b > 0$$ independently of $$\sigma_1, \sigma_2, \ldots$$

$$\sigma_j = \tau_1 + \cdots + \tau_j, \ j \geq 1$$ with $$\tau_1, \tau_2, \ldots \overset{i.i.d.}{\sim} \mathcal{E}(q), \ q > 0$$

$$k$$ is a zero mean normal density on $$\mathbb{R}$$ with standard deviation $$q^{-1}$$

and finally $$K(y) = \int_{-\infty}^{y} k(x)dx, \ y \in \mathbb{R}.$$
The treatment/placebo scenario

It is a simple important case of the proportional hazards model, in which

\[ x_i \in \{0, 1\} \]

for all \( i = 1 \ldots N \). The main goal is determining whether the hazard ratio

\[ \zeta = e^\beta \]

is significantly different from one. In this case, the conjugate choice

\[ \rho_\star \perp \zeta \sim \mathcal{G}(c, d) \]

is possible, which helps the implementation of a Gibbs-type MCMC solution.

When no specific prior information is available, condition

\[ \mathbb{P}\{\zeta < 1\} = \mathbb{P}\{\zeta > 1\} \]

can be imposed in order to help fixing the values of \( c \) and \( d \).
Prior Comparison Plot

\[ \zeta \sim G(1.31425, 1) \]
\[ \zeta = e^\beta, \beta \sim N(0, 1) \]
\[ \zeta \sim G(0.1, 0.000593391) \]
\[ \zeta_{0.5} = 1 \]
The leukemia remission times

A well known dataset has been analyzed, in order to validate the suggested approach.

Data consist of 21 treatment/placebo pairs of leukemia remission times, with 12 right censored observations in the group of treated patients, clearly showing a shorter remission for patients receiving placebo.

The hyperparameters $a$, $b$ and $q$ have been chosen by setting

$$qt(n) = 10$$

$$\mathbb{E}[\rho_*(s)] = \frac{\sum_{i=1}^{n} \mathbb{I}_{\{o_i=1, x_i=0\}}}{\sum_{i=1}^{n} t_i \mathbb{I}_{\{x_i=0\}}}$$

$$s \in \mathbb{R}^+$$

$$\frac{\text{Std}[\rho_*(s)]}{\mathbb{E}[\rho_*(s)]} \to 1, \quad \text{as } s \to \infty$$

and an ad hoc MCMC solution has been implemented in R.
Estimation of the regression coefficient

The posterior expected value of $\beta$ is found to be

$$\hat{\beta} = 1.26$$

which compared with the available estimates

<table>
<thead>
<tr>
<th>Author</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox (1972)</td>
<td>1.65</td>
</tr>
<tr>
<td>Kalbfleish (1978)</td>
<td>1.46–1.61</td>
</tr>
<tr>
<td>Laud <em>et al.</em> (1998)</td>
<td>1.62–1.71</td>
</tr>
<tr>
<td>Ibrahim <em>et al.</em> (2001)</td>
<td>1.59</td>
</tr>
</tbody>
</table>

clearly shows that the suggested approach is conservative.