Bayesian Survival Analysis using Stan - Notes on Work in Progress

Sam Brilleman

8 October 2018

Proportional-hazards model formulation

A flexible proportional hazards model can be specified on the hazard scale, with M-splines used to model the baseline hazard. Let the hazard for individual i at time t be formulated as

$$\begin{aligned} h_i(t) &= h_0(t) \exp(\boldsymbol{X_i^T} \boldsymbol{\beta}) \\ &= m(t; \boldsymbol{\gamma}, \boldsymbol{k_0}) \exp(\boldsymbol{X_i^T} \boldsymbol{\beta}) \end{aligned} \tag{1}$$

where $h_0(t)$ is the baseline hazard at time t, $\boldsymbol{X_i}$ is a vector of baseline covariates for individual i with associated parameters (i.e. hazard ratios) $\boldsymbol{\beta}$, and $m(t; \boldsymbol{\gamma}, \boldsymbol{k_0})$ is an M-spline function with basis evaluated at a vector of knot locations $\boldsymbol{k_0}$ and parameter vector $\boldsymbol{\gamma}$.

This leads to a closed form expression for the cumulative hazard

$$\begin{aligned} H_i(t) &= H_0(t) \exp(\boldsymbol{X_i^T} \boldsymbol{\beta}) \\ &= \int_{s=0}^{t} m(s; \boldsymbol{\gamma}, \boldsymbol{k_0}) ds \exp(\boldsymbol{X_i^T} \boldsymbol{\beta}) \\ &= i(t; \boldsymbol{\gamma}, \boldsymbol{k_0}) \exp(\boldsymbol{X_i^T} \boldsymbol{\beta}) \end{aligned} \tag{2}$$

where $i(t; \boldsymbol{\gamma}, \boldsymbol{k_0})$ is an I-spline function (i.e. the integral of an M-spline).

Assuming that we constrain the baseline hazard coefficients to be positive, i.e. $\boldsymbol{\gamma} > 0$, this model formulation satisfies four desirable criteria:

- the M-spline basis (i.e. baseline hazard) is strictly positive
- the I-spline basis (i.e. baseline cumulative hazard) is strictly positive and monotonically increasing
- the likelihood is strictly positive
- there is a closed-form expression for the (log-)likelihood, which leads to faster computation and improved accuracy compared with numerical approximations

This model can therefore be implemented in Stan **and** estimation should be relatively fast.

Extension to time-dependent coeffients (i.e. non-proportional hazards)

We can extend the previous model formulation to allow for time-dependent coefficients (i.e. non-proportional hazards).

We define the hazard for individual i at time t as

$$
\begin{aligned}
h_i(t) &= h_0(t)\exp(\boldsymbol{X_i^T}\boldsymbol{\beta(t)}) \\
&= m(t;\boldsymbol{\gamma},\boldsymbol{k_0})\exp(\boldsymbol{X_i^T}\boldsymbol{\beta(t)})
\end{aligned}
\tag{3}
$$

with parameter vector $\boldsymbol{\beta(t)} = \{\beta_p(t); p = 1, ..., P\}$. Each element of the parameter vector corresponds to a, possibly time-dependent, regression coefficient (i.e. hazard ratio). We formulate each regression coefficient as

$$
\beta_p(t) = \begin{cases} \theta_{p0} & \text{for proportional hazards} \\ b(t;\boldsymbol{\theta_p},\boldsymbol{k_p}) & \text{for non-proportional hazards} \end{cases}
\tag{4}
$$

where θ_{p0} is a time-fixed hazard ratio, or alternatively, $b(t;\boldsymbol{\theta_p},\boldsymbol{k_p})$ is a B-spline function with basis evaluated at a vector of knot locations $\boldsymbol{k_p}$ and parameter vector $\boldsymbol{\theta_p}$.

To avoid overfitting in the B-spline function for the estimated time-dependent coefficient, we can penalise the B-spline coefficients [1]. This can be implemented in practice by specifying

$$
p(\boldsymbol{\theta_p}|\tau_p) \propto \tau^{\rho(K_p)/2}\exp\left(\frac{-\tau_p}{2}\boldsymbol{\theta_p^T}\boldsymbol{K_p}\boldsymbol{\theta_p}\right)
\tag{5}
$$

as the prior distribution for the B-spline coefficients $\boldsymbol{\theta_p}$, where

- $\boldsymbol{K_p} = \Delta_r^T\Delta_r + 10^{-6}I$ and Δ_r denotes the r^{th}-difference penalty matrix
- $\rho(\boldsymbol{K_p})$ denotes the rank of $\boldsymbol{K_p}$, and
- τ_p controls the smoothness with lower values resulting in a less flexible (i.e. smoother) function

For the hyperparameter τ_p we can allow the user to choose between an exponential, half-normal, half-t, or half-Cauchy prior distribution.

Unfortunately, the formulation on the previous slide **does not lead to a closed form expression for the cumulative hazard**, hence there is no closed form expression for the (log-)likelihood. Therefore, quadrature is required to approximate the cumulative hazard at each MCMC iteration, so estimation is much slower than the situation without time-dependent coefficients.

References

[1] Lang S, Brezger A. Bayesian p-splines. *Journal of Computational and Graphical Statistics* 2004; 13: 183–212.