1 Background

Survival analysis concerns the analysis of time to failure events. Let $\boldsymbol{T}$ denote a time to failure event with density $f(t)$. Then the survivor function $S(t) = 1 - F(t)$ where $F(t)$ is the cumulative density function. And, $h(t)$, the hazard rate is given by $h(t) = f(t)/S(t)$.

1.1 Proportional hazards model

Let $\mathcal{D} = (t_i, \delta_i, x_i)_{i=1}^n$ denote our observed data for $t_i \in \mathbb{R}^+, \delta_i \in \{0, 1\}, x_i \in R^p$, where t_i is an observed time, δ_i is a censoring indicator with $\delta_i = 0$ denoting right censoring, and x_i are observed covariates. A popular approach for analysing failure data when there are covariates is through the proportional hazards models (PHM). In the PHM, it is assumed the hazard rate takes the form

$$h(t; x, \beta) = h_0(t) \exp(\beta^\top x) \tag{1}$$

where $\beta \in \mathbb{R}^p$. Under the PHM assumption, the density is given by

$$f(t; x, \beta, \delta) = h_0(t) \exp(\beta^\top x) \exp(-H(t)) \tag{2}$$

where $H(t) = \int_0^t h_(s) ds$, the cumulative hazard rate.

1.2 Bayesian survival analysis

Fitting PHMs in a Frequentist setting involves finding the coefficients that maximise the partial likelihood, given as

$$L_p(\mathcal{D}; \beta) = \prod_{\{i:\delta_i=1\}} \frac{\exp(\beta^\top x_i}{\sum_{\{r \in \mathbb{R}_i\}} \exp(\beta^\top x_r} \tag{3}$$

where $R_i := \{r : t_r \geq t_i\}$, known as the risk set. In a Bayesian setting we can approach posterior inference by the evaluating the posterior

$$p(\beta; \mathcal{D}) \propto L_p(\mathcal{D}; \beta) \pi(\beta) \tag{4}$$

where $\pi(\beta)$ is our prior distribution on β. Use of the partial likelihood in Equation (4) can be thought of as using a profile likelihood, where $H_0(t)$ a nuisance parameter is given by the MLE of $H_0(t)$.

1.3 Spike and slab prior

Spike and slab priors are a class of prior that impose shrinkage. Generally, the form of spike and slab priors is given by

$$z_j \phi_{\text{slab}}(\beta_j) + (1 - z_j) \phi_{\text{spike}}(\beta_j) \tag{5}$$

where $z_j \in \{0, 1\}$, ϕ_{slab} is a density that spreads mass in the domain of β_j and ϕ_{spike} concentrates mass around some point (typically 0). We consider a spike and slab prior where the slab is given by the Laplace density (centered at 0) and the spike by a Dirac mass on zero, formally,

$$\pi(\beta_j|z_j) \sim z_j \text{Laplace}(\lambda) + (1-z_j)\delta(\beta_j) \tag{6}$$

$$\pi(z_j|w) \sim \text{Bernoulli}(w) \tag{7}$$

$$\pi(w) \sim \text{Beta}(a_0, b_0) \tag{8}$$

for $a_0, b_0, \lambda > 0$.

2 Posterior Inference

We can perform posterior inference by constructing the Gibbs sampler presented in Algorithm 1.

Algorithm 1 Gibbs sampler for Equation (4)

Initialise $z^{(0)}, \beta^{(0)}, w^{(0)}$
for $i = 1, \dots N$ **do**
  $w^{(i)} \sim p(w|\mathcal{D}, \beta^{(i-1)}, z^{(i-1)})$
  for $j = 1, \dots p$ **do**
    $z_j^{(i)} \sim p(z_j|\mathcal{D}, \beta^{(i-1)}, z_{1:(j-1)}^{(i)}, z_{(j+1):p}^{(i-1)}, w^{(i)})$
  for $j = 1, \dots p$ **do**
    $\beta_j^{(i)} \sim p(\beta_j|\mathcal{D}, \beta_{1:(j-1)}^{(i-1)}, \beta_{(j+1):p}^{(i-1)}, z^{(i)}, w^{(i)})$

where $x_{k:l} = (x_i)_{i=k}^{l}$, for $1 \leq l < k \leq p,\ x \in \mathbb{R}^p$ and N are the number of iterations we run our sampler for.

Ignoring the superscript for clarity, the conditional distribution for $p(z_j|\mathcal{D}, \beta, z_{-j}, w)$ is given by

$$p(z_j|\mathcal{D}, \beta, z_{-j}, w) \propto L_p(\mathcal{D}; \beta, z, w)\pi(z_j|w) \tag{9}$$

we can evaluate the RHS for $z_j = 0$ and $z_j = 1$ to obtain the unnormalised probabilities. For $p(\beta_j|\mathcal{D}, \beta_{-j}, z, w)$, we turn to the Metropolis Hastings algorithm to sample from the conditional distribution noting

$$p(\beta_j|\mathcal{D}, \beta_{-j}, z, w) \propto L_p(\mathcal{D}; \beta, z, w)\pi(\beta_j). \tag{10}$$

To sample β_j we propose some kernel Q and let $\beta_j^{(i+1)} \sim Q$, letting

$$A := \min\left(1, \frac{L_p(\mathcal{D}; \beta_{-j}, \beta_j^{(i+1)}, z, w)\pi(\beta_j^{(i+1)})}{L_p(\mathcal{D}; \beta_{-j}, \beta_j^{(i)}, z, w)\pi(\beta_j^{(i)})} \frac{Q(\beta_j^{(i)}|\beta_j^{(i+1)})}{Q(\beta_j^{(i+1)}|\beta_j^{(i)})}\right) \tag{11}$$

we accept $\beta_j^{(i+1)}$ with probability A, and we reject the proposal with probability $1 - A$; letting $\beta_j^{(i+1)} \leftarrow \beta_j^{(i)}$ if we reject the proposal.

3 Example

`R` code implementing Algorithm 1 is available at `https://github.com/mkomod/mcmc_ss_surv`. Within our implementation we have taken $Q = \mathcal{N}(\beta_j; \beta_j^{(i)}; \sigma^2)$ where $\sigma = 0.3$. Running our algorithm for $p = 50$, letting $\beta = (4, 4, 0, \ldots, 0)$, $\lambda = 0.01$ and $N = 10,000$ (using a burn-in 1,000). We present the chains for $\beta_1, \ldots, \beta_4$ in Figure 1.

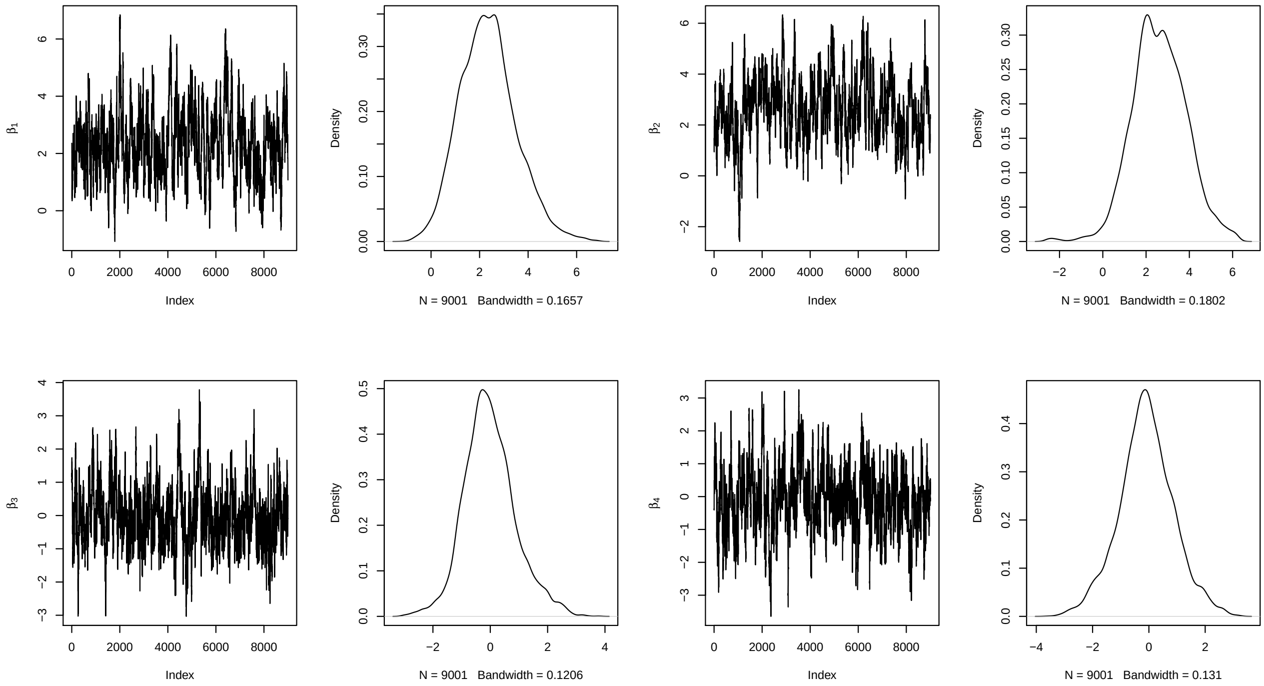

Figure 1: Mixing and marginal densities for $\beta_1, \ldots \beta_4$

References

W. Bruinsma. Spike and slab priors. `https://wesselb.github.io/assets/write-ups/Bruinsma,%20Spike%20and%20Slab%20Priors.pdf`, 2021.

D. R. Cox. Regression Models and Life-Tables. *J. R. Stat. Soc. Ser. B*, 34(2):187–220, feb 1972.

M. Stephens. The metropolis hastings algorithm. `https://stephens999.github.io/fiveMinuteStats/MH_intro.html`, 2018.