Introduction to spBayesSurv

Haiming Zhou
Northern Illinois University

Abstract

This tutorial provides an introduction to a set of programs for implementing Bayesian spatial survival models in R using the package **spBayesSurv**. The function `survregbayes` includes three most commonly-used semiparametric models: proportional hazards, proportional odds, and accelerated failure time. All manner of censored survival times are simultaneously accommodated including uncensored, interval censored, current-status, left and right censored, and mixtures of these. Left-truncated data are also accommodated leading to models for time-dependent covariates. Both georeferenced and areally observed spatial locations are handled via frailties. Model fit is assessed with conditional Cox-Snell residual plots, and model choice is carried out via LPML and DIC. The function `frailtyGAFT` extends the accelerated failure time frailty model to allow covariates-dependent baseline. The package can also fit two marginal survival models: proportional hazards (`spCopulaCoxph`) and linear dependent Dirichlet process mixture (`spCopulaDDP`), where the spatial dependence is modeled via spatial copulas. Note that all these models can also accommodate non-spatial data.

Keywords: Bayesian survival analysis, spatial dependence, semiparametric models, parametric models, interval-censored data.

1. Introduction

Due to the development of geographical information systems, many survival (time-to-event) data are spatially referenced. Spatial location plays a key role in survival prediction, serving as a proxy for unmeasured regional characteristics such as socioeconomic status, access to health care, pollution, etc. Literature on the spatial analysis of survival data has flourished over the last decade, including the study of leukemia survival (Henderson, Shimakura, and Gorst 2002), childhood mortality (Kneib 2006), asthma (Li and Lin 2006), breast cancer (Banerjee and Dey 2005; Zhou, Hanson, Jara, and Zhang 2015a), political event processes (Darmofal 2009), prostate cancer (Wang, Zhang, and Lawson 2012; Zhou, Hanson, and Zhang 2016), pine trees (Li, Hong, Thapa, and Burkhart 2015), threatened frogs (Zhou, Hanson, and Knapp 2015b), health and pharmaceutical firms (Arbia, Espa, Giuliani, and Micciolo 2016), and many others. In this tutorial we introduce a set of programs for implementing Bayesian spatial survival models in R using the package **spBayesSurv**, version 1.1.0. Note that the function syntaxes for `spCopulaCoxph`, `indeptCoxph`, `spCopulaDDP` and `anovaDDP` have changed in comparison to previous versions, and old syntaxes are no longer supported; see Section 4.

Suppose subjects are observed at m distinct spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_m$. Let t_{ij} be a random event time associated with the jth subject in $\mathbf{s}_i$ and $\mathbf{x}_{ij}$ be a related p-dimensional vector of covariates, $i = 1, \ldots, m, j = 1, \ldots, n_i$. Then $n = \sum_{i=1}^{m} n_i$ is the total number of subjects under

consideration. Assume the survival time t_{ij} lies in the interval (a_{ij}, b_{ij}), $0 \leq a_{ij} \leq b_{ij} \leq \infty$, and t_{ij} is independent with (a_{ij}, b_{ij}). Here left censored data are of the form $(0, b_{ij})$, right censored (a_{ij}, ∞), interval censored (a_{ij}, b_{ij}) and uncensored values simply have $a_{ij} = b_{ij}$, i.e., we define $(x, x) = \{x\}$. Therefore, the observed data will be $\mathcal{D} = \{(a_{ij}, b_{ij}, \mathbf{x}_{ij}, \mathbf{s}_i); i = 1, \ldots, m, j = 1, \ldots, n_i\}$. For areally-observed outcomes, e.g. county-level, there is typically replication (i.e. $n_i > 1$); for georeferenced, there may or may not be replication. Note although the models below are developed for spatial survival data, non-spatial data are also accommodated.

The tutorial is organized as follows. Section 2 introduces three commonly-used semiparametric frailty models, a test for parametric baseline, variable selection, left-truncation and time-dependent covariates. Section 3 describes a generalized accelerated failure time frailty model which allows covariates-dependent baseline function. Section 4 provides two spatial copula survival models for georeferenced, right-censored data.

2. Semiparametric Frailty Models

2.1. The Model

The function `survregbayes` supports three commonly-used semiparametric spatial frailty models: accelerated failure time (AFT), proportional hazards (PH), and proportional odds (PO). The AFT model has survival and density functions

$$S_{\mathbf{x}_{ij}}(t) = S_0(e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}t),\ f_{\mathbf{x}_{ij}}(t) = e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}f_0(e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}t), \tag{1}$$

while the PH model has survival and density functions

$$S_{\mathbf{x}_{ij}}(t) = S_0(t)^{e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}},\ f_{\mathbf{x}_{ij}}(t) = e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}S_0(t)^{e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}-1}f_0(t), \tag{2}$$

and the PO model has survival and density functions

$$S_{\mathbf{x}_{ij}}(t) = \frac{e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}S_0(t)}{1+(e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}-1)S_0(t)},\ f_{\mathbf{x}_{ij}}(t) = \frac{e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}f_0(t)}{[1+(e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}-1)S_0(t)]^2}, \tag{3}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of regression coefficients, v_i is an unobserved frailty associated with $\mathbf{s}_i$, and $S_0(t)$ is the baseline survival with density $f_0(t)$ corresponding to $\mathbf{x}_{ij} = \mathbf{0}$ and $v_i = 0$. The `survregbayes` function considers the following prior distributions:

$$\begin{aligned}
\boldsymbol{\beta} &\sim N_p(\boldsymbol{\beta}_0, \mathbf{S}_0),\\
S_0|\alpha, \boldsymbol{\theta} &\sim \text{TBP}_L(\alpha, S_{\boldsymbol{\theta}}),\ \alpha \sim \Gamma(a_0, b_0),\ \boldsymbol{\theta} \sim N_2(\boldsymbol{\theta}_0, \mathbf{V}_0),\\
(v_1, \ldots, v_m)'|\tau &\sim \text{ICAR}(\tau^2),\ \tau^{-2} \sim \Gamma(a_\tau, b_\tau),\ \text{or}\\
(v_1, \ldots, v_m)'|\tau, \phi &\sim \text{GRF}(\tau^2, \phi),\ \tau^{-2} \sim \Gamma(a_\tau, b_\tau),\ \phi \sim \Gamma(a_\phi, b_\phi),\ \text{or}\\
(v_1, \ldots, v_m)'|\tau &\sim \text{IID}(\tau^2),\ \tau^{-2} \sim \Gamma(a_\tau, b_\tau)
\end{aligned} \tag{4}$$

where TBP_L, ICAR, GRF and IID refer to the transformed Bernstein polynomial (TBP) (Chen, Hanson, and Zhang 2014; Zhou and Hanson 2017), intrinsic conditionally autoregressive (ICAR) (Besag 1974), Gaussian random field (GRF) prior, and independent Gaussian

(IID) prior distributions, respectively. We briefly introduce these priors but leave details to Zhou and Hanson (2017).

TBP prior

In semiparametric survival analysis, a wide variety of Bayesian nonparametric priors can be used to model S_0; see Müller, Quintana, Jara, and Hanson (2015) and Zhou and Hanson (2015) for reviews. The TBP prior is attractive in that it is centered at a given parametric family and it selects smooth densities. For a fixed positive integer L, the prior $\text{TBP}_L(\alpha, S_{\boldsymbol{\theta}})$ is defined as

$$S_0(t) = \sum_{j=1}^{L} w_j I(S_{\boldsymbol{\theta}}(t)|j, L-j+1), \ \mathbf{w}_L \sim \text{Dirichlet}(\alpha, \ldots, \alpha), \tag{5}$$

where $\mathbf{w}_L = (w_1, \ldots, w_L)'$ is a vector of positive weights, $I(\cdot|a, b)$ denotes a beta cumulative distribution function (cdf) with parameters (a, b), $\{S_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is a parametric family of survival functions with support on positive reals $\mathbb{R}^+$. The log-logistic $S_{\boldsymbol{\theta}}(t) = \{1 + (e^{\theta_1} t)^{\exp(\theta_2)}\}^{-1}$, the log-normal $S_{\boldsymbol{\theta}}(t) = 1 - \Phi\{(\log t + \theta_1)\exp(\theta_2)\}$, and the Weibull $S_{\boldsymbol{\theta}}(t) = 1 - \exp\left\{-(e^{\theta_1} t)^{\exp(\theta_2)}\right\}$ families are considered, where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$. In our experience, the three centering distributions yield almost identical posterior inferences. Clearly, the random distribution S_0 is centered at $S_{\boldsymbol{\theta}}$, that is, $E[S_0(t)|\alpha, \boldsymbol{\theta}] = S_{\boldsymbol{\theta}}(t)$. The parameter α controls the weights $\mathbf{w}_L$ to "adjust" the shape of the baseline survival S_0 relative to the prior guess $S_{\boldsymbol{\theta}}$. Large values of α indicate a strong belief that S_0 is close to $S_{\boldsymbol{\theta}}$; as $\alpha \to \infty$, $S_0 \to S_{\boldsymbol{\theta}}$ with probability 1. Smaller values of α allow more pronounced deviations of S_0 from $S_{\boldsymbol{\theta}}$. This adaptability makes the TBP prior attractive in its flexibility, but also anchors the random S_0 firmly about $S_{\boldsymbol{\theta}}$: $w_j = 1/L$ for $j = 1, \ldots, L$ implies $S_0(t) = S_{\boldsymbol{\theta}}(t)$ for $t \geq 0$. Moreover, unlike the mixture of Polya trees (Lavine 1992) or mixture of Dirichlet process (Antoniak 1974) priors, the TBP prior selects smooth densities, leading to efficient posterior sampling.

ICAR and IID priors

For areal data, the ICAR prior can be assumed on $\mathbf{v} = (v_1, \ldots, v_m)'$. Let e_{ij} be 1 if regions i and j share a common boundary and 0 otherwise; set $e_{ii} = 0$. Then the $m \times m$ matrix $\mathbf{E} = [e_{ij}]$ is called the adjacency matrix for the m regions. The prior $\text{ICAR}(\tau^2)$ on $\mathbf{v}$ is defined through the set of the conditional distributions

$$v_i|\{v_j\}_{j \neq i} \sim N\left(\sum_{j=1}^{m} e_{ij} v_j / e_{i+}, \ \tau^2 / e_{i+}\right), \ i = 1, \ldots, m, \tag{6}$$

where $e_{i+} = \sum_{j=1}^{m} e_{ij}$ is the number of neighbors of area $\mathbf{s}_i$. The induced prior on $\mathbf{v}$ under ICAR is improper; the constraint $\sum_{j=1}^{m} v_j = 0$ is used for identifiability (Banerjee, Carlin, and Gelfand 2014).

For non-spatial data, we consider the independent Gaussian prior $\text{IID}(\tau^2)$, defined as

$$v_1, v_2, \ldots, v_m \overset{iid}{\sim} N(0, \tau^2). \tag{7}$$

GRF priors

For georeferenced data, it is commonly assumed that $v_i = v(\mathbf{s}_i)$ arises from a Gaussian ran-

dom field (GRF) $\{v(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$ such that $\mathbf{v} = (v_1, \ldots, v_m)$ follows a multivariate Gaussian distribution as $\mathbf{v} \sim N_m(\mathbf{0}, \tau^2\mathbf{R})$, where τ^2 measures the amount of spatial variation across locations and the (i, j) element of $\mathbf{R}$ is modeled as $\mathbf{R}[i, j] = \rho(\mathbf{s}_i, \mathbf{s}_j)$. Here $\rho(\cdot, \cdot)$ is a correlation function controlling the spatial dependence of $v(\mathbf{s})$. In `survregbayes` the powered exponential correlation function $\rho(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s}, \mathbf{s}'; \phi) = \exp\{-(\phi\|\mathbf{s} - \mathbf{s}'\|)^\nu\}$ is used, where $\phi > 0$ is a range parameter controlling the spatial decay over distance, $\nu \in (0, 2]$ is a pre-specified shape parameter, and $\|\mathbf{s} - \mathbf{s}'\|$ refers to the distance (e.g., Euclidean, great-circle) between $\mathbf{s}$ and $\mathbf{s}'$. Therefore, the prior GRF(τ^2, ϕ) is defined as

$$v_i|\{v_j\}_{j\neq i} \sim N\left(-\sum_{\{j:j\neq i\}} p_{ij}v_j/p_{ii},\ \tau^2/p_{ii}\right),\ i = 1, \ldots, m, \tag{8}$$

where p_{ij} is the (i, j) element of $\mathbf{R}^{-1}$.

Full-scale approximation

As m increases evaluating $\mathbf{R}^{-1}$ from $\mathbf{R}$ becomes computationally impractical. To overcome this computational issue, we consider the full-scale approximation (Sang and Huang 2012) (FSA) due to its capability of capturing both large- and small-scale spatial dependence. Consider a fixed set of "knots" $\mathcal{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_K^*\}$ chosen from the study region. These knots are chosen using the function `cover.design` within the R package `fields`, which computes space-filling coverage designs using the swapping algorithm (Johnson, Moore, and Ylvisaker 1990). Let $\rho(\mathbf{s}, \mathbf{s}')$ be the correlation between locations $\mathbf{s}$ and $\mathbf{s}'$. The FSA approach approximates the correlation function $\rho(\mathbf{s}, \mathbf{s}')$ with

$$\rho^\dagger(\mathbf{s}, \mathbf{s}') = \rho_l(\mathbf{s}, \mathbf{s}') + \rho_s(\mathbf{s}, \mathbf{s}'). \tag{9}$$

The $\rho_l(\mathbf{s}, \mathbf{s}')$ in (9) is the reduced-rank part capturing the long-scale spatial dependence, defined as $\rho_l(\mathbf{s}, \mathbf{s}') = \rho'(\mathbf{s}, \mathcal{S}^*)\rho_{KK}^{-1}(\mathcal{S}^*, \mathcal{S}^*)\rho(\mathbf{s}', \mathcal{S}^*)$, where $\rho(\mathbf{s}, \mathcal{S}^*) = [\rho(\mathbf{s}, \mathbf{s}_i^*)]_{i=1}^K$ is an $K \times 1$ vector, and $\rho_{KK}(\mathcal{S}^*, \mathcal{S}^*) = [\rho(\mathbf{s}_i^*, \mathbf{s}_j^*)]_{i,j=1}^K$ is an $K \times K$ correlation matrix at knots $\mathcal{S}^*$. However, $\rho_l(\mathbf{s}, \mathbf{s}')$ cannot well capture the short-scale dependence due to the fact that it discards entirely the residual part $\rho(\mathbf{s}, \mathbf{s}') - \rho_l(\mathbf{s}, \mathbf{s}')$. The idea of FSA is to add a small-scale part $\rho_s(\mathbf{s}, \mathbf{s}')$ as a sparse approximate of the residual part, defined by $\rho_s(\mathbf{s}, \mathbf{s}') = \{\rho(\mathbf{s}, \mathbf{s}') - \rho_l(\mathbf{s}, \mathbf{s}')\}\,\Delta(\mathbf{s}, \mathbf{s}')$, where $\Delta(\mathbf{s}, \mathbf{s}')$ is a modulating function, which is specified so that the $\rho_s(\mathbf{s}, \mathbf{s}')$ can well capture the local residual spatial dependence while still permits efficient computations. Motivated by Konomi, Sang, and Mallick (2014), we first partition the total input space into B disjoint blocks, and then specify $\Delta(\mathbf{s}, \mathbf{s}')$ in a way such that the residuals are independent across input blocks, but the original residual dependence structure within each block is retained. Specifically, the function $\Delta(\mathbf{s}, \mathbf{s}')$ is taken to be 1 if $\mathbf{s}$ and $\mathbf{s}'$ belong to the same block and 0 otherwise. The approximated correlation function $\rho^\dagger(\mathbf{s}, \mathbf{s}')$ in (9) provides an exact recovery of the true correlation within each block, and the approximation errors are $\rho(\mathbf{s}, \mathbf{s}') - \rho_l(\mathbf{s}, \mathbf{s}')$ for locations $\mathbf{s}$ and $\mathbf{s}'$ in different blocks. Those errors are expected to be small for most entries because most of these location pairs are farther apart. To determine the blocks, we first use the R function `cover.design` to choose $B \leq m$ locations among the m locations forming B blocks, then assign each $\mathbf{s}_i$ to the block that is closest to $\mathbf{s}_i$. Here B does not need to be equal to K. When $B = 1$, no approximation is applied to the correlation ρ. When $B = m$, it reduces to the approach of Finley, Sang, Banerjee, and Gelfand (2009), so the local residual spatial dependence may not be well captured.

Applying the above FSA approach to approximate the correlation function $\rho(\mathbf{s}, \mathbf{s}')$, we can approximate the correlation matrix $\mathbf{R}$ with

$$\boldsymbol{\rho}^{\dagger}_{mm} = \boldsymbol{\rho}_l + \boldsymbol{\rho}_s = \boldsymbol{\rho}_{mK}\boldsymbol{\rho}^{-1}_{KK}\boldsymbol{\rho}'_{mK} + \left(\boldsymbol{\rho}_{mm} - \boldsymbol{\rho}_{mK}\boldsymbol{\rho}^{-1}_{KK}\boldsymbol{\rho}'_{mK}\right) \circ \boldsymbol{\Delta}, \tag{10}$$

where $\boldsymbol{\rho}_{mK} = [\rho(\mathbf{s}_i, \mathbf{s}^*_j)]_{i=1:m, j=1:K}$, $\boldsymbol{\rho}_{KK} = [\rho(\mathbf{s}^*_i, \mathbf{s}^*_j)]^K_{i,j=1}$, and $\boldsymbol{\Delta} = [\Delta(\mathbf{s}_i, \mathbf{s}_j)]^m_{i,j=1}$. Here, the notation "$\circ$" represents the element-wise matrix multiplication. To avoid numerical instability, we add a small nugget effect $\epsilon = 0.001$ when define $\mathbf{R}$, that is, $\mathbf{R} = (1-\epsilon)\boldsymbol{\rho}_{mm} + \epsilon\mathbf{I}_m$. It follows from equation (10) that $\mathbf{R}$ can be approximated by

$$\mathbf{R}^{\dagger} = (1-\epsilon)\boldsymbol{\rho}^{\dagger}_{mm} + \epsilon\mathbf{I}_m = (1-\epsilon)\boldsymbol{\rho}_{mK}\boldsymbol{\rho}^{-1}_{KK}\boldsymbol{\rho}'_{mK} + \mathbf{R}_s,$$

where $\mathbf{R}_s = (1-\epsilon)\left(\boldsymbol{\rho}_{mm} - \boldsymbol{\rho}_{mK}\boldsymbol{\rho}^{-1}_{KK}\boldsymbol{\rho}'_{mK}\right) \circ \boldsymbol{\Delta} + \epsilon\mathbf{I}_m$. Applying the Sherman-Woodbury-Morrison formula for inverse matrices, we can approximate $\mathbf{R}^{-1}$ by

$$\left(\mathbf{R}^{\dagger}\right)^{-1} = \mathbf{R}^{-1}_s - (1-\epsilon)\mathbf{R}^{-1}_s\boldsymbol{\rho}_{mK}\left[\boldsymbol{\rho}_{KK} + (1-\epsilon)\boldsymbol{\rho}'_{mK}\mathbf{R}^{-1}_s\boldsymbol{\rho}_{mK}\right]^{-1}\boldsymbol{\rho}'_{mK}\mathbf{R}^{-1}_s. \tag{11}$$

In addition, the determinant of $\mathbf{R}$ can be approximated by

$$\det\left(\mathbf{R}^{\dagger}\right) = \det\left\{\boldsymbol{\rho}_{KK} + (1-\epsilon)\boldsymbol{\rho}'_{mK}\mathbf{R}^{-1}_s\boldsymbol{\rho}_{mK}\right\}\det(\boldsymbol{\rho}_{KK})^{-1}\det(\mathbf{R}_s). \tag{12}$$

Since the $m \times m$ matrix $\mathbf{R}_s$ is a block matrix, the right-hand sides of equations (11) and (12) involve only inverses and determinants of $K \times K$ low-rank matrices and $m \times m$ block diagonal matrices. Thus the computational complexity can be greatly reduced relative to the expensive computational cost of using original correlation function for large value of m.

2.2. MCMC

The likelihood function for $(\mathbf{w}_L, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v})$ is given by

$$\mathcal{L}(\mathbf{w}_L, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = \prod_{i=1}^{m}\prod_{j=1}^{n_i}\left[S_{\mathbf{x}_{ij}}(a_{ij}) - S_{\mathbf{x}_{ij}}(b_{ij})\right]^{I\{a_{ij}<b_{ij}\}} f_{\mathbf{x}_{ij}}(a_{ij})^{I\{a_{ij}=b_{ij}\}}. \tag{13}$$

MCMC is carried out through an empirical Bayes approach coupled with adaptive Metropolis samplers (Haario, Saksman, and Tamminen 2001). Recall that $w_j = 1/L$ implies the underlying parametric model with $S_0(t) = S_{\boldsymbol{\theta}}(t)$. Thus, the parametric model provides good starting values for the TBP survival model. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ denote the parametric estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and let $\hat{\mathbf{V}}$ and $\hat{\mathbf{S}}$ denote their estimated covariance matrices, respectively. Set $\mathbf{z}_{L-1} = (z_1, \ldots, z_{L-1})'$ with $z_j = \log(w_j) - \log(w_L)$. The $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\mathbf{z}_{L-1}$, α and ϕ are all updated using adaptive Metropolis samplers, where the initial proposal variance is $\hat{\mathbf{S}}$ for $\boldsymbol{\beta}$, $\hat{\mathbf{V}}$ for $\boldsymbol{\theta}$, $0.16\mathbf{I}_{L-1}$ for $\mathbf{z}_{L-1}$ and 0.16 for α and ϕ. Each frailty term v_i is updated via Metropolis-Hastings, with proposal variance as the conditional prior variance of $v_i|\{v_j\}_{j\neq i}$; τ^{-2} is updated via a Gibbs step from its full conditional. A complete description and derivation of the updating steps are available in Zhou and Hanson (2017).

The function `survregbayes` sets the following hyperparameters as defaults: $\boldsymbol{\beta}_0 = \mathbf{0}$, $\mathbf{S}_0 = 10^{10}\mathbf{I}_p$, $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}$, $\mathbf{V}_0 = 10\hat{\mathbf{V}}$, $a_0 = b_0 = 1$, and $a_\tau = b_\tau = .001$. Note here we assume a somewhat informative prior on $\boldsymbol{\theta}$ to obviate confounding between $\boldsymbol{\theta}$ and $\mathbf{w}_L$. For georeferenced data, we set and $a_\phi = \phi_0 b_\phi + 1$ and $b_\phi = 1$ so that the prior of ϕ has mode at ϕ_0. Here ϕ_0 satisfies

$\rho(\mathbf{s}', \mathbf{s}''; \phi_0) = 0.001$, where $\|\mathbf{s}' - \mathbf{s}''\| = \max_{ij} \|\mathbf{s}_i - \mathbf{s}_j\|$. Note that Kneib and Fahrmeir (2007) simply fix ϕ at ϕ_0, while we allow ϕ to be random around ϕ_0.

2.3. Model Diagnostics and Comparison

For model diagnostics, we consider a general residual of Cox and Snell (1968), defined as $r(t_{ij}) = -\log\{S_{\mathbf{x}_{ij}}(t_{ij})|\mathcal{D}\}$, where the residual depends on the posterior $[\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{w}_L, v_i|\mathcal{D}]$. Given $S_{\mathbf{x}_{ij}}(\cdot)$, $-\log S_{\mathbf{x}_{ij}}(t_{ij})$ has a standard exponential distribution. If the model is "correct," and under the arbitrary censoring, the pairs $\{r(a_{ij}), r(b_{ij})\}$ are approximately a random arbitrarily censored sample from an Exp(1) distribution, and the estimated (Turnbull 1974) integrated hazard plot should be approximately straight with slope 1. Uncertainty in the plot is assessed through several cumulative hazards based on a random sample from $[\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{w}_J, v_i|\mathcal{D}]$. This is in contrast to typical Cox-Snell plots which only use point estimates.

For model comparison, we consider two popular model choice criteria: the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, and Van Der Linde 2002) and the log pseudo marginal likelihood (LPML) (Geisser and Eddy 1979), where the former places emphasis on the relative quality of model fitting and the latter focuses on the predictive performance. Both criteria are readily computed from the MCMC output; see Zhou and Hanson (2017) for more details.

2.4. Leukemia Survival Data

A dataset on the survival of acute myeloid leukemia in $n = 1,043$ patients (Henderson *et al.* 2002) is considered, named as `LeukSurv` in the package. It is of interest to investigate possible spatial variation in survival after accounting for known subject-specific prognostic factors, which include `age`, `sex`, white blood cell count (`wbc`) at diagnosis, and the Townsend score (`tpi`) for which higher values indicates less affluent areas. Both exact residential locations of all patients and their administrative districts (the boundary file is named as `nwengland.bnd` in the package) are available, so we can fit both geostatistical and areal models.

PO model with ICAR frailties

We first need to sort the dataset by `district`, then obtain the adjacency matrix $\mathbf{E}$.

```
> library(coda)
> library(survival)
> library(spBayesSurv)
> library(fields)
> library(BayesX)
> library(R2BayesX)
> data(LeukSurv);
> attach(LeukSurv);
> d = LeukSurv[order(district),]; n = nrow(d); detach(LeukSurv);
> head(d);
    time cens    xcoord    ycoord age sex   wbc  tpi district
24     1    1 0.4123484 0.4233738  44   1 281.0 4.87        1
62     3    1 0.3925028 0.4531422  72   1   0.0 7.10        1
68     4    1 0.4167585 0.4520397  68   0   0.0 5.12        1
```

```
128    9    1 0.4244763 0.4123484  61   1   0.0 2.90        1
129    9    1 0.4145535 0.4520397  26   1   0.0 6.72        1
163   15    1 0.4013230 0.4785006  67   1  27.9 1.50        1
> nwengland=read.bnd(system.file("otherdata/nwengland.bnd",
+                                package="spBayesSurv"));
> adj.mat=bnd2gra(nwengland)
> E = diag(diag(adj.mat)) - as.matrix(adj.mat);
```

The following code is used to fit the PO model with ICAR frailties using the TBP prior with $L = 15$ and default settings for other priors. A burn-in period of 5,000 iterates was considered and the Markov chain was subsampled every 5 iterates to get a final chain size of 2,000. The argument `ndisplay=1000` will display the number of saved scans after every 1,000 saved iterates. If the argument `InitParamMCMC=TRUE` (not used here as it is the default setting), then an initial chain with `nburn=5000`, `nsave=5000`, `nkip=0` and `ndisplay=1000` will be run; otherwise, the initial values are obtained from fitting parametric non-frailty models via `survreg`. The total running time is 172 seconds.

```
> mcmc=list(nburn=5000, nsave=2000, nskip=4, ndisplay=1000);
> prior=list(maxL=15);
> ptm<-proc.time()
> res1 = survregbayes(formula=Surv(time,cens)~age+sex+wbc+tpi
+                     +frailtyprior("car",district),data=d,survmodel="PO",
+                     dist="loglogistic",mcmc=mcmc,prior=prior,Proximity=E);
Starting initial MCMC based on parametric model:
scan = 1000
scan = 2000
scan = 3000
scan = 4000
scan = 5000
Starting the MCMC for the semiparametric model:
scan = 1000
scan = 2000
> systime1=proc.time()-ptm; systime1;
   user  system elapsed
168.930   0.950 171.507
```

The term `frailtyprior("car",district)` indicates that the ICAR prior in (6) is used. One can also incorporate the IID prior in (7) via `frailtyprior("iid",district)`. The non-frailty model can be fit by removing the `frailtyprior` term. The argument `survmodel` is used to indicate which model will be fit; choices include `"PH"`, `"PO"`, and `"AFT"`. The argument `dist` is used to specify the distribution family of $S_{\boldsymbol{\theta}}$ defined in Section 2.1, and the choices include `"loglogistic"`, `"lognormal"`, and `"weibull"`. The argument `prior` is used to specify user-defined hyperparameters, e.g., for $p = 3$, $L = 15$, $\boldsymbol{\beta}_0 = \mathbf{0}$, $\mathbf{S}_0 = 10\mathbf{I}_p$, $\boldsymbol{\theta}_0 = \mathbf{0}$, $\mathbf{V}_0 = 10\mathbf{I}_2$, $a_0 = b_0 = 1$, and $a_\tau = b_\tau = 1$, the prior can be specified as below.

```
> prior=list(maxL=15,beta0=rep(0,3),S0=diag(10,3),theta0=rep(0,2),
+            V0=diag(10,2),a0=1,b0=1,taua0=1,taub0=1)
```

If `prior=NULL`, then the default hyperparameters given in Section 2.2 would be used. Note by default `survregbayes` standardizes each covariate by subtracting the sample mean and dividing the sample standard deviation. Therefore, the user-specified hyperparameters should be based on the model with scaled covariates unless the argument `scale.designX=FALSE` is added.

The output from applying the `summary` function to the returned object `res1` is given below.

```
> sfit1=summary(res1); sfit1
Proportional Odds model:
Call:
survregbayes(formula = Surv(time, cens) ~ age + sex + wbc + tpi +
    frailtyprior("car", district), data = d, survmodel = "PO",
    dist = "loglogistic", mcmc = mcmc, prior = prior, Proximity = E)

Posterior inference of regression coefficients
(Adaptive M-H acceptance rate: 0.2806):
      Mean        Median      Std. Dev.   95%CI-Low   95%CI-Upp
age   0.0515312   0.0513855   0.0034000   0.0448747   0.0582174
sex   0.1238992   0.1227971   0.1129346  -0.0996123   0.3539006
wbc   0.0059400   0.0059837   0.0007868   0.0043236   0.0074710
tpi   0.0616777   0.0614836   0.0165983   0.0310655   0.0945423

Posterior inference of precision parameter
(Adaptive M-H acceptance rate: 0.1856):
       Mean    Median  Std. Dev.  95%CI-Low  95%CI-Upp
alpha  0.9596  0.8766  0.4511     0.3387     2.0669

Posterior inference of conditional CAR frailty variance
          Mean      Median    Std. Dev.  95%CI-Low  95%CI-Upp
variance  0.074520  0.050776  0.085125   0.001517   0.283800

Log pseudo marginal likelihood: LPML=-5923.603
Deviance Information Criterion: DIC=11847.01
Number of subjects: n=1043
```

We can see that `age`, `wbc` and `tpi` are significant risk factors for leukemia survival. For example, lower `age` decreases the odds of a patient dying by any time; holding other predictors constant, a 10-year decrease in age cuts the odds of dying by $\exp(-10 \times 0.0515312) \approx 60\%$. The posterior mean for τ^2 is 0.074520, and is 0.9596 for precision parameter α. The LPML and DIC are -5924 and 11847, respectively.

The following code is used to produce trace plots (Figure 1) for $\boldsymbol{\beta}$, τ^2 and α. Note that the mixing for τ^2 is not very satisfactory. This is not surprising, since we are using very vague gamma prior $\Gamma(0.001, 0.001)$ and the total number of districts is only 24.

```
> par(mfrow=c(3,2));
> par(cex=1,mar=c(2.5,4.1,1,1))
> traceplot(mcmc(res1$beta[1,]), xlab="", main="age")
```

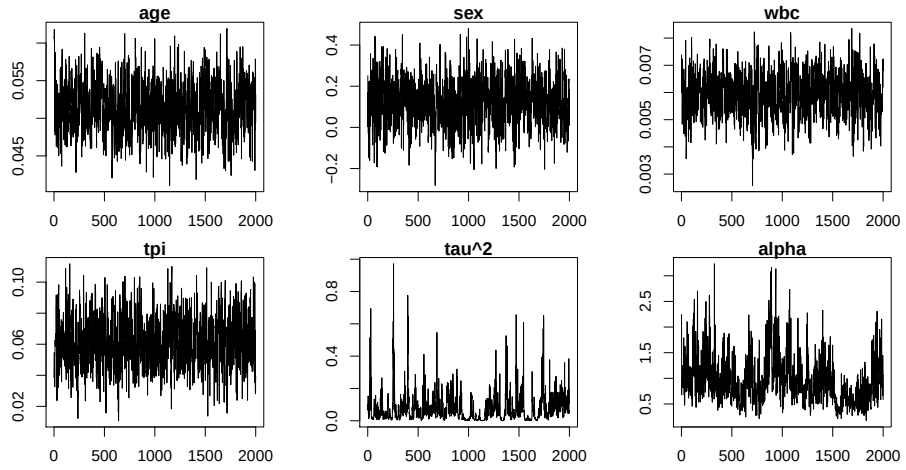


Figure 1: Leukemia survival data. Trace plots for $\boldsymbol{\beta}$, τ^2 and α under the PO model with ICAR frailties.

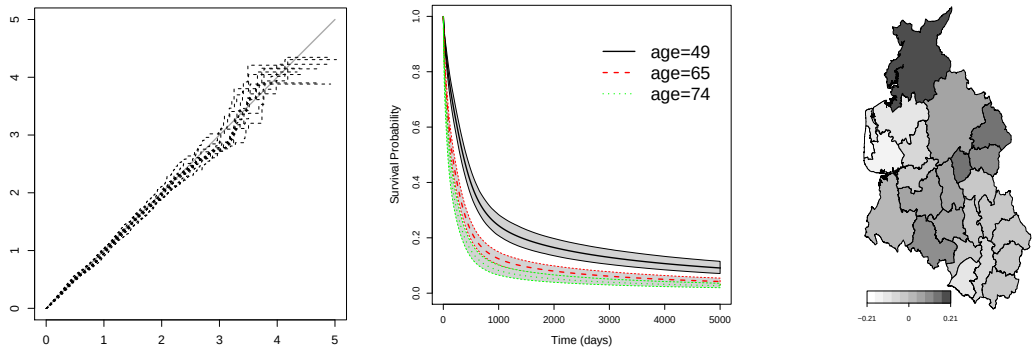
```
> traceplot(mcmc(res1$beta[2,]), xlab="", main="sex")
> traceplot(mcmc(res1$beta[3,]), xlab="", main="wbc")
> traceplot(mcmc(res1$beta[4,]), xlab="", main="tpi")
> traceplot(mcmc(res1$tau2), xlab="", main="tau^2")
> traceplot(mcmc(res1$alpha), xlab="", main="alpha")
```

The code below is used to generate the Cox-Snell plots with 10 posterior residuals (Figure 2, panel a).

```
> nrand = 10;
> Resid = cox.snell.survregbayes(res1, ncurves=nrand);
> r.max = ceiling(quantile(res1$Surv.cox.snell[,1], .99))+1
> xlim=c(0, r.max); ylim=c(0, r.max); width=8; height=8;
> xx = seq(0, r.max, 0.01);
> fit = survfit(Resid$resid1~1);
> par(cex=1.5,mar=c(2.1,2.1,1,1),cex.lab=1.4,cex.axis=1.1)
> plot(fit, fun="cumhaz", conf.int=F, mark.time=FALSE, xlim=xlim,
+      ylim=ylim, lwd=2, lty=2)
> lines(xx, xx, lty=1, lwd=3, col="darkgrey")
> for(i in 2:nrand){
+   fit = survfit(Resid[[i+1]]~1);
+   lines(fit, fun="cumhaz", conf.int=F, mark.time=FALSE, xlim=xlim,
+         ylim=ylim, lwd=2, lty=2)
+ }
```

The code below is used to generate survival curves for female patients with `wbc`=38.59 and `tpi`=0.3398 at different ages (Figure 2, panel b).

```
> tgrid = seq(0.1,5000,length.out=300);
> xpred = rbind(c(49, 0, 38.59, 0.3398),
```



(a) (b) (c)

Figure 2: Leukemia survival data. PO model with ICAR frailties. (a) Cox-Snell plot. (b) Survival curves with 95% credit interval bands for female patients with `wbc`=38.59 and `tpi`=0.3398 at different ages. (c) Map for the posterior mean frailties; larger frailties mean higher mortality rate overall.

```
+               c(65, 0, 38.59, 0.3398),
+               c(74, 0, 38.59, 0.3398));
> estimates=plot(res1, xpred=xpred, tgrid=tgrid);
> par(mfrow=c(1,1));
> par(cex=1.2,mar=c(4.1,4.1,1,1),cex.lab=1.3,cex.axis=1.1)
> plot(estimates$tgrid, estimates$Shat[,1], "l", lwd=3, ylim = c(0, 1),
+      xlab = "Time (days)", ylab="Survival Probability")
> polygon(x=c(rev(tgrid),tgrid),
+         y=c(rev(estimates$Shatlow[,1]),estimates$Shatup[,1]),
+         border=c("black"),col="lightgray");
> lines(estimates$tgrid, estimates$Shat[,1], lty=1, lwd=3)
> polygon(x=c(rev(tgrid),tgrid),
+         y=c(rev(estimates$Shatlow[,2]),estimates$Shatup[,2]),
+         border=c("red"),lty=2, col="lightgray");
> lines(estimates$tgrid, estimates$Shat[,2], lty=2, lwd=3, col="red")
> polygon(x=c(rev(tgrid),tgrid),
+         y=c(rev(estimates$Shatlow[,3]),estimates$Shatup[,3]),
+         border=c("green"),lty=2, col="lightgray");
> lines(estimates$tgrid, estimates$Shat[,3], lty=3, lwd=3, col="green")
> legend(2600,0.95, legend=c("age=49","age=65","age=74"),
+        lty=c(1,2,3),lwd=c(3,3,3),col=c("black","red","green"),bty="n",cex=2)
```

The code below is used to generate the map of posterior means of frailties for each district (Figure 2, panel c).

```
> frail0=(rowMeans(res1$v)); #$
> frail = frail0[as.integer(names(nwengland))];
> values = cbind(as.integer(names(nwengland)), frail)
> op <- par(no.readonly = TRUE)
> par(mar=c(3,0,0,0))
```

```
> plotmap(nwengland, x=values, col=(gray.colors(10,0.3,1))[10:1],
+         pos = "bottomleft",width = 0.5, height = 0.04)
```

PO model with GRF frailties

Note that all coordinates are distinct, so we have $m = 1043$ and $n_i = 1$ in terms of our notations. To use `frailtyprior` specify the prior, we need to create an `ID` variables consisting of 1043 distinct values. The powered exponential correlation function with $\nu = 1$ is used. To specify the number of knots and blocks for the FSA of $\mathbf{R}$, we consider $K = 100$ and $B = 1043$. The code below is used to fit a PO model with GRF frailties under above settings. The running time is 10478 seconds.

```
> mcmc=list(nburn=5000, nsave=2000, nskip=4, ndisplay=1000);
> prior=list(maxL=15, nu=1, nknots=100, nblock=1043);
> d$ID=1:nrow(d); #$
> locations=cbind(d$xcoord,d$ycoord);
> ptm<-proc.time()
> res2 = survregbayes(formula=Surv(time,cens)~age+sex+wbc+tpi
+                     +frailtyprior("grf",ID),data=d,survmodel="PO",
+                     dist="loglogistic",mcmc=mcmc,prior=prior,
+                     Coordinates=locations);
> sfit2=summary(res2); sfit2

Posterior inference of regression coefficients
(Adaptive M-H acceptance rate: 0.2838):
      Mean        Median      Std. Dev.   95%CI-Low   95%CI-Upp
age   0.0529726   0.0530458   0.0034661   0.0461137   0.0595578
sex   0.1068712   0.1057764   0.1114583  -0.1039998   0.3221860
wbc   0.0060615   0.0060646   0.0007782   0.0045370   0.0075862
tpi   0.0592036   0.0599728   0.0157041   0.0276731   0.0889597

Posterior inference of precision parameter
(Adaptive M-H acceptance rate: 0.2431):
       Mean    Median  Std. Dev.  95%CI-Low  95%CI-Upp
alpha  1.0894  0.9871  0.4984     0.4190     2.3426

Posterior inference of frailty variance
          Mean    Median  Std. Dev.  95%CI-Low  95%CI-Upp
variance  0.1468  0.1302  0.0604     0.0754     0.3084

Posterior inference of correlation function range phi
        Mean    Median  Std. Dev.  95%CI-Low  95%CI-Upp
range   7.924   7.705   2.157      4.229      12.402

Log pseudo marginal likelihood: LPML=-5921.707
Deviance Information Criterion: DIC=11842.45
Number of subjects: n=1043
```

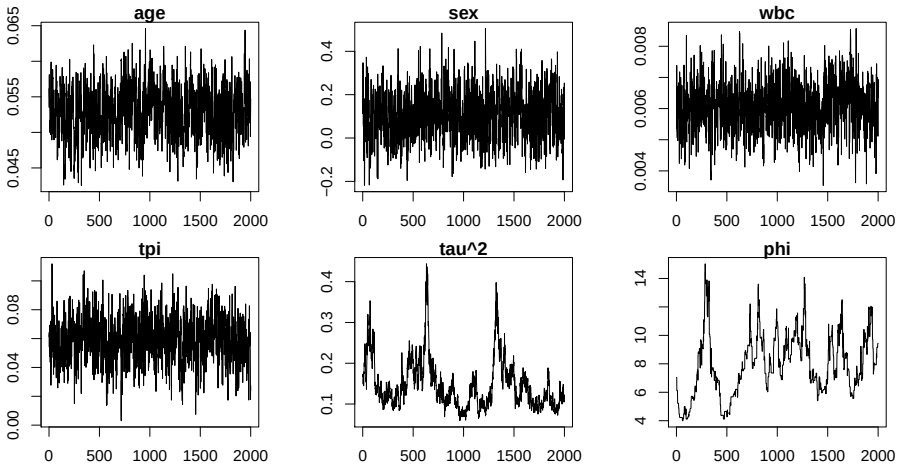


Figure 3: Leukemia survival data. Trace plots for $\boldsymbol{\beta}$, τ^2 and α under the PO model with GRF frailties.

```
> systime2=proc.time()-ptm; systime2;
     user    system   elapsed
10393.707    79.944 10478.230
```

The trace plots for $\boldsymbol{\beta}$, τ^2 and ϕ (Figure 3), Cox-Snell residuals and survival curves (Figure 4) can be obtained using the same code used for the PO model with ICAR frailties. The code below is used to generate the map of posterior means of frailties for each location (Figure 4).

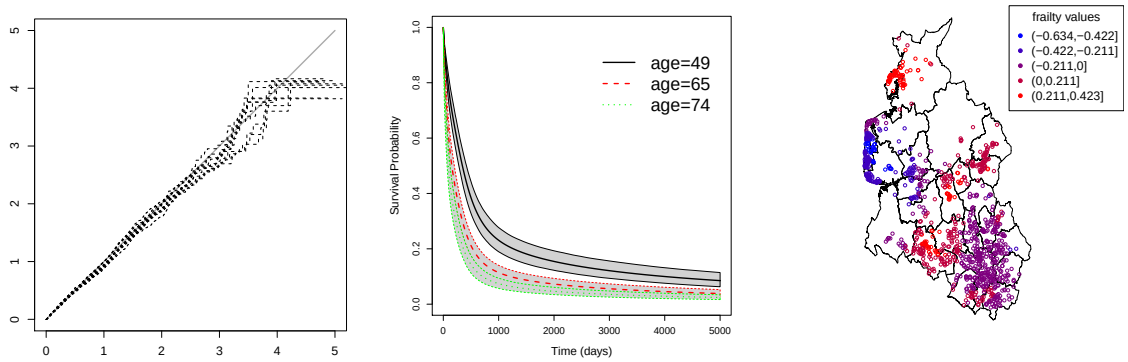
```
> frail= round((rowMeans(res2$v)),3); nclust=5; #$
> frail.cluster = cut(frail, breaks = nclust);
> frail.names = names(table(frail.cluster))
> rbPal <- colorRampPalette(c('blue','red'))
> frail.colors=rbPal(nclust)[as.numeric(frail.cluster)]
> par(mar=c(3,0,0,0))
> plot(nwengland)
> points(cbind(d$xcoord,d$ycoord), col=frail.colors)
> legend("topright",title="frailty values",legend=frail.names,
+        col=rbPal(nclust),pch=20,cex=1.7)
```

2.5. Variable Selection

The most direct approach is to multiply β_k by a latent Bernoulli variable γ_k, where $\gamma_k = 1$ indicates presence of covariate x_k in the model, and then assume an appropriate prior on $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. Following Kuo and Mallick (1998) and Hanson, Branscum, Johnson *et al.* (2014), we consider below independent priors

$$\gamma_1, \ldots, \gamma_p \overset{iid}{\sim} \text{Bern}(0.5) \text{ and } \boldsymbol{\beta} \sim N_p(\mathbf{0}, gn(\mathbf{X}'\mathbf{X})^{-1}), \tag{14}$$

where $\mathbf{X}$ is the usual design matrix, but with mean-centered covariates, i.e. $\mathbf{1}_n'\mathbf{X} = \mathbf{0}_p'$, and g is chosen by picking a number M such that a random $e^{\mathbf{x}'\boldsymbol{\beta}}$ is less than M with probability



(a) (b) (c)

Figure 4: Leukemia survival data. PO model with GRF frailties. (a) Cox-Snell plot. (b) Survival curves with 95% credit interval bands for female patients with `wbc`=38.59 and `tpi`=0.3398 at different ages. (c) Map for the posterior mean frailties; larger frailties mean higher mortality rate overall.

q, i.e. approximately $g = \left[\log M/\Phi^{-1}(q)\right]^2/p$. The function `survregbayes` sets $M = 10$ and $q = 0.9$ as the defaults. The MCMC procedure is described in Zhou and Hanson (2017).

To perform variable selection for the leukemia survival data, we simply need to add the argument `selection=TRUE` to the function `survregbayes`. A part of the output from `summary` is also shown. The model with `age`, `wbc` and `tpi` has the highest proportion (85.9%), and thus can be served as the final model.

```
> res3 = survregbayes(formula=Surv(time,cens)~age+sex+wbc+tpi
+                     +frailtyprior("car",district),data=d,survmodel="PO",
+                     dist="loglogistic",mcmc=mcmc,prior=prior,Proximity=E,
+                     selection=TRUE);
> systime3=proc.time()-ptm; systime3;
   user  system elapsed
312.111   1.632 316.392
> sfit3=summary(res3); sfit3
Variable selection:
age,wbc,tpi  age,sex,wbc,tpi  age,wbc  age,sex,wbc
prop.  0.8590       0.0985          0.0375   0.0050

Log pseudo marginal likelihood: LPML=-5925.228
Deviance Information Criterion: DIC=11849.45
Number of subjects: n=1043
```

2.6. Parametric vs. Semiparametric

Many authors have found parametric models to fit as well or better than competing semiparametric models (Cox and Oakes 1984, p. 123; Nardi and Schemper 2003). The semiparametric TBP models have their baseline survival functions centered at a parametric family $S_{\boldsymbol{\theta}}(t)$. Note that $\mathbf{z}_{J-1} = \mathbf{0}$ implies $S_0(t) = S_{\boldsymbol{\theta}}(t)$. Therefore, testing $H_0 : \mathbf{z}_{J-1} = \mathbf{0}$ versus $H_1 : \mathbf{z}_{J-1} \neq \mathbf{0}$

leads to the comparison of the semiparametric model with the underlying parametric model. Let BF_{10} be the Bayes factor between H_1 and H_0. Zhou *et al.* (2016) proposed to estimate BF_{10} by a large-sample approximation to the generalized Savage-Dickey density ratio (Verdinelli and Wasserman 1995). Adapting their approach BF_{10} is estimated

$$\widehat{BF}_{10} = \frac{p(\mathbf{0}|\hat{\alpha})}{N_{J-1}(\mathbf{0}; \hat{\mathbf{m}}, \hat{\mathbf{\Sigma}})}, \tag{15}$$

where $p(\mathbf{0}|\alpha) = \Gamma(\alpha J)/[J^{\alpha}\Gamma(\alpha)]^J$ is the prior density of $\mathbf{z}_{J-1}$ evaluated at $\mathbf{z}_{J-1} = \mathbf{0}$, $\hat{\alpha}$ is the posterior mean of α, $N_p(\cdot; \mathbf{m}, \mathbf{\Sigma})$ denotes a p-variable normal density with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$, and $\hat{\mathbf{m}}$ and $\hat{\mathbf{\Sigma}}$ are posterior mean and covariance of $\mathbf{z}_{J-1}$.

The Bayes factor BF_{10} under the semiparametric PO model with ICAR frailties can be obtained using the code below (here the object `res1` is obtained in Section 2.4).

```
> BF.survregbayes(res1)
[1] 2330.477
```

The $BF_{10} = 2330 \gg 1$ indicates that the semiparametric model significantly outperforms the loglogistic parametric model.

The function `survregbayes` also supports efficient parametric frailty models with loglogistic, lognormal or Weibull baseline function. For example, the following code fits a parametric loglogistic PO model with ICAR frailties for the leukemia survival data.

```
> prior=list(maxL=15, a0=-1,thete0=rep(0,2),V0=diag(1e10,2));
> state=list(alpha=Inf);
> ptm<-proc.time()
> res11 = survregbayes(formula=Surv(time,cens)~age+sex+wbc+tpi
+                      +frailtyprior("car",district),data=d,survmodel="PO",
+                      dist="loglogistic",mcmc=mcmc,prior=prior,state=state,
+                      Proximity=E,InitParamMCMC=FALSE);
scan = 1000
scan = 2000
> sfit11=summary(res11); sfit11
Proportional Odds model:
Call:
survregbayes(formula = Surv(time, cens) ~ age + sex + wbc + tpi +
frailtyprior("car", district), data = d, survmodel = "PO",
dist = "loglogistic", mcmc = mcmc, prior = prior, state = state,
Proximity = E, InitParamMCMC = FALSE)

Posterior inference of regression coefficients
(Adaptive M-H acceptance rate: 0.2903):
      Mean        Median      Std. Dev.   95%CI-Low   95%CI-Upp
age   0.0504202   0.0505329   0.0034000   0.0437902   0.0570289
sex   0.1146655   0.1110484   0.1136106  -0.0962541   0.3412315
wbc   0.0062118   0.0062088   0.0007854   0.0046620   0.0077172
tpi   0.0596343   0.0592660   0.0155276   0.0303326   0.0912814
```

```
Posterior inference of baseline parameters
Note: the baseline estimates are based on scaled covariates
(Adaptive M-H acceptance rate: 0.2811):
         Mean      Median    Std. Dev.  95%CI-Low  95%CI-Upp
theta1  -5.12357  -5.12354   0.06320   -5.24813   -5.00353
theta2  -0.10715  -0.10640   0.02826   -0.16283   -0.05271

Posterior inference of conditional CAR frailty variance
          Mean     Median   Std. Dev.  95%CI-Low  95%CI-Upp
variance  0.07614  0.06035  0.07411    0.00150    0.26608

Log pseudo marginal likelihood: LPML=-5950.056
Deviance Information Criterion: DIC=11899.71
Number of subjects: n=1043
> systime11=proc.time()-ptm; systime11;
  user  system elapsed
25.123   0.112  25.369
```

In parametric models, the prior for $\boldsymbol{\theta}$ can be set to be relatively vague. The LPML is -5950, much worse than the value under the semiparametric PO model. Note that setting a_0 at any negative value will force the α to be fixed at the value specified in the argument `state`. For example, setting `prior=list(a0=-1)` and `state=list(alpha=1)` will fix $\alpha = 1$ throughout the MCMC; setting `prior=list(a0=-1)` and `state=list(alpha=Inf)` will fit a parametric model.

2.7. Left-Truncation and Time-Dependent Covariates

The survival time t_{ij} is left-truncated at $u_{ij} \geq 0$ when u_{ij} is the time when the ijth subject is first observed. Left-truncation often occurs when age is used as the time scale. Given the observed left-truncated data $\mathcal{D} = \{(u_{ij}, a_{ij}, b_{ij}, \mathbf{x}_{ij}, \mathbf{s}_i)\}$, the likelihood function (13) becomes

$$L(\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[S_{\mathbf{x}_{ij}}(a_{ij}) - S_{\mathbf{x}_{ij}}(b_{ij})\right]^{I\{a_{ij}<b_{ij}\}} f_{\mathbf{x}_{ij}}(a_{ij})^{I\{a_{ij}=b_{ij}\}} / S_{\mathbf{x}_{ij}}(u_{ij}).$$

Allowing for left-truncation allows the semiparametric AFT, PH and PO models to be easily extended to handle time-dependent covariates. Following Kneib (2006) and Hanson, Johnson, and Laud (2009), assume the covariate vector $\mathbf{x}_{ij}(t)$ is a step function that changes at o_{ij} ordered times $t_{ij,1} < \ldots < t_{ij,o_{ij}} \leq a_{ij}$, i.e.,

$$\mathbf{x}_{ij}(t) = \sum_{k=1}^{o_{ij}} \mathbf{x}_{ij,k} I(t_{ij,k} \leq t < t_{ij,k+1}),$$

where $t_{ij,1} = u_{ij}$ and $t_{ij,o_{ij}+1} = \infty$. Assuming one of PH, PO, or AFT holds conditionally on

each interval, the survival function for the ijth individual at time a_{ij} is

$$P(t_{ij} > a_{ij}) = P(t_{ij} > a_{ij} | t_{ij} > t_{ij,o_{ij}}) \prod_{k=1}^{o_{ij}} P(t_{ij} > t_{ij,k} | t_{ij} > t_{ij,k-1})$$
$$= \frac{S_{\mathbf{x}_{ij,o_{ij}}}(a_{ij})}{S_{\mathbf{x}_{ij,o_{ij}}}(t_{ij,o_{ij}})} \prod_{k=1}^{o_{ij}} \frac{S_{\mathbf{x}_{ij,k}}(t_{ij,k})}{S_{\mathbf{x}_{ij,k}}(t_{ij,k-1})},$$

where $t_{ij,0} = 0$. This leads to the usual PH model for time-dependent covariates (Cox 1972), the AFT model first proposed by Prentice and Kalbfleisch (1979) and developed by Hanson *et al.* (2009), and particular piecewise PO model. Thus one can replace the observation $(u_{ij}, a_{ij}, b_{ij}, \mathbf{x}_{ij}(t), \mathbf{s}_i)$ by a set of new o_{ij} observations $(t_{ij,1}, t_{ij,2}, \infty, \mathbf{x}_{ij,1}, \mathbf{s}_i)$, $(t_{ij,2}, t_{ij,3}, \infty, \mathbf{x}_{ij,2}, \mathbf{s}_i)$, ..., $(t_{ij,o_{ij}}, a_{ij}, b_{ij}, \mathbf{x}_{ij,o_{ij}}, \mathbf{s}_i)$. This way we get a new left-truncated data set of size $N = \sum_{i=1}^{m} \sum_{j=1}^{n_i} o_{ij}$. Then the likelihood function becomes

$$L(\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left\{ \left[S_{\mathbf{x}_{ij,o_{ij}}}(a_{ij}) - S_{\mathbf{x}_{ij,o_{ij}}}(b_{ij}) \right]^{I\{a_{ij} < b_{ij}\}} f_{\mathbf{x}_{ij,o_{ij}}}(a_{ij})^{I\{a_{ij} = b_{ij}\}} \right.$$
$$\left. \times \prod_{k=1}^{o_{ij}} \frac{S_{\mathbf{x}_{ij,k}}(t_{ij,k})}{S_{\mathbf{x}_{ij,k}}(t_{ij,k-1})} \right\}.$$

PBC data

We use the primary biliary cirrhosis (PBC) dataset (available in the package `survival` as `pbc`) as an example to show how to incorporate time-dependent covariates in the function `survregbayes`. Although this is not a spatial dataset, spatial frailties can be added similarly as in Section 2.4. The following code is copied from Therneau, Crowson, and Atkinson (2016) to create the data frame with time-dependent covariates.

```
> temp <- subset(pbc, id <= 312, select=c(id:sex, stage)) # baseline data
> pbc2 <- tmerge(temp, temp, id=id, endpt = event(time, status))
> pbc2 <- tmerge(pbc2, pbcseq, id=id, ascites = tdc(day, ascites),
+                bili = tdc(day, bili), albumin = tdc(day, albumin),
+                protime = tdc(day, protime), alk.phos = tdc(day, alk.phos))
> pbc2 = pbc2[,c("id","tstart","tstop","endpt","bili","protime")];
> head(pbc2);
  id tstart tstop endpt bili protime
1  1      0   192     0 14.5    12.2
2  1    192   400     2 21.3    11.2
3  2      0   182     0  1.1    10.6
4  2    182   365     0  0.8    11.0
5  2    365   768     0  1.0    11.6
6  2    768  1790     0  1.9    10.6
> coxph(Surv(tstart,tstop,endpt==2)~log(bili)+log(protime), data=pbc2)
Call:
coxph(formula = Surv(tstart, tstop, endpt == 2) ~ log(bili) +
log(protime), data = pbc2)
```

```
coef exp(coef) se(coef)     z      p
log(bili)     1.241     3.460    0.097 12.80 <2e-16
log(protime)  3.983    53.699    0.436  9.14 <2e-16

Likelihood ratio test=332  on 2 df, p=0
n= 1807, number of events= 125
```

We can fit the Bayesian PH model with TBP baseline as follows. The output for regression coefficients is partial.

```
> mcmc=list(nburn=5000, nsave=2000, nskip=4, ndisplay=1000);
> ptm<-proc.time()
> fit1 = survregbayes(Surv(tstart,tstop,endpt==2)~log(bili)+log(protime),
+                     data=pbc2, survmodel="PH", dist="loglogistic",
+                     mcmc=mcmc, subject.num=id);
> fit1
Proportional hazards model:
Call:
survregbayes(formula = Surv(tstart, tstop, endpt == 2) ~ log(bili) +
        log(protime), data = pbc2, survmodel = "PH", dist = "loglogistic",
        mcmc = mcmc, subject.num = id)

Posterior means for regression coefficients:
   log(bili)  log(protime)
       1.315         4.217

LPML: -1018.964
DIC: 2033.276
n=1807
> systime1=proc.time()-ptm; systime1;
   user  system elapsed
234.975   1.108 237.089
```

Equivalently, one can also run the following code to obtain the same analysis. The argument `truncation_time` is used to specify the start time point for each time interval, i.e. `tstart`. The end time point `tstop` together with `endpt` are formulated as interval censored data using `type="interval2"` of `Surv`. This format is more general than the former one, as one can easily incorporate interval censored data.

```
> pbc2$tleft=pbc2$tstop; pbc2$tright=pbc2$tstop;
> pbc2$tright[which(pbc2$endpt!=2)]=NA;
> fit11 = survregbayes(Surv(tleft,tright,type="interval2")~log(bili)
+                     +log(protime), data=pbc2, survmodel="PH",
+                     dist="loglogistic", mcmc=mcmc,
+                     truncation_time=tstart, subject.num=id);
```

3. GAFT Frailty Models

3.1. The Model

The generalized accelerated failure time (GAFT) frailty model (Zhou *et al.* 2016) generalizes the AFT model (1) to allow the baseline survival function $S_0(t)$ to depend on certain covariates, say a q-dimensional vector $\mathbf{z}_{ij}$ which is usually a subset of $\mathbf{x}_{ij}$. Specifically, the GAFT frailty model is given by

$$S_{\mathbf{x}_{ij}}(t) = S_{0,\mathbf{z}_{ij}}\left(e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}t\right), \tag{16}$$

or equivalently,

$$y_{ij} = \log(t_{ij}) = \tilde{\mathbf{x}}_{ij}'\tilde{\boldsymbol{\beta}} + v_i + \epsilon_{ij}, \tag{17}$$

where $\tilde{\mathbf{x}}_{ij} = (1, \mathbf{x}_{ij}')'$ includes an intercept, $\tilde{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}')'$ is a vector of corresponding coefficients, ϵ_{ij} is a heteroscedastic error term independent of v_i, and $P(e^{\beta_0+\epsilon_{ij}} > t|\mathbf{z}_{ij}) = S_{0,\mathbf{z}_{ij}}(t)$. Note the regression coefficients $\boldsymbol{\beta}$ here are defined differently with those in model (1). Here we assume

$$\epsilon_{ij}|G_{\mathbf{z}_{ij}} \overset{ind.}{\sim} G_{\mathbf{z}_{ij}},$$

where $G_{\mathbf{z}}$ is a probability measure defined on $\mathbb{R}$ for every $\mathbf{z} \in \mathcal{X}$; this defines a model for the entire collection of probability measures $\mathcal{G}_{\mathcal{X}} = \{G_{\mathbf{z}} : \mathbf{z} \in \mathcal{X}\}$ so that each element is allowed to smoothly change with the covariates $\mathbf{z}$. The `frailtyGAFT` function considers the following prior distributions:

$$\begin{aligned}
\tilde{\boldsymbol{\beta}} &\sim N_{p+1}(\mathbf{m}_0, \mathbf{S}_0) \\
G_{\mathbf{z}}|\alpha, \sigma^2 &\sim \text{LDTFP}_L(\alpha, \sigma^2),\ \alpha \sim \Gamma(a_0, b_0),\ \sigma^{-2} \sim \Gamma(a_\sigma, b_\sigma), \\
(v_1, \ldots, v_m)'|\tau &\sim \text{ICAR}(\tau^2),\ \tau^{-2} \sim \Gamma(a_\tau, b_\tau),\ \text{or} \\
(v_1, \ldots, v_m)'|\tau, \phi &\sim \text{GRF}(\tau^2, \phi),\ \tau^{-2} \sim \Gamma(a_\tau, b_\tau),\ \phi \sim \Gamma(a_\phi, b_\phi),
\end{aligned} \tag{18}$$

where LDTFP_L refers to the linear dependent tailfree process prior (LDTFP) prior as described in (Zhou *et al.* 2016).

The LDTFP prior considered in Zhou *et al.* (2016) is centered at a normal distribution Φ_σ with mean 0 and variance σ^2, that is, $E(G_{\mathbf{z}}) = N(0, \sigma^2)$ for every $\mathbf{z} \in \mathcal{X}$. Define the function $k_\sigma(x) = \lceil 2^L \Phi_\sigma(x) \rceil$, where $\lceil x \rceil$ is the ceiling function, the smallest integer greater than or equal to x. Further define probability $p_{\mathbf{z}}(k)$ for $k = 1, \ldots, 2^L$ as

$$p_{\mathbf{z}}(k) = \prod_{l=1}^{L} Y_{l,\lceil k2^{l-L} \rceil}(\mathbf{z}),$$

where $Y_{j+1,2k-1}(\mathbf{z}) = \left(1 + \exp\{-\tilde{\mathbf{z}}'\boldsymbol{\gamma}_{j,k}\}\right)^{-1}$ and $Y_{j+1,2k}(\mathbf{z}) = 1 - Y_{j+1,2k-1}(\mathbf{z})$ for $j = 0, \ldots, L-1$, $k = 1, \ldots, 2^j$, where $\tilde{\mathbf{z}} = (1, \mathbf{z}')'$ includes an intercept, and $\boldsymbol{\gamma}_{j,k} = (\gamma_{j,k,0}, \ldots, \gamma_{j,k,q})'$ is a vector of coefficients. Note there are $2^L - 1$ regression coefficient vectors $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_{j,k}\}$, e.g. for $L = 3$, $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_{0,1}, \boldsymbol{\gamma}_{1,1}, \boldsymbol{\gamma}_{1,2}, \boldsymbol{\gamma}_{2,1}, \boldsymbol{\gamma}_{2,2}, \boldsymbol{\gamma}_{2,3}, \boldsymbol{\gamma}_{2,4}\}$. For a fixed integer $L > 0$, the random density associated with $\text{LDTFP}_L(\alpha, \sigma^2)$ is defined as

$$f_{\mathbf{z}}(e) = 2^L \phi_\sigma(e) p_{\mathbf{z}}\{k_\sigma(e)\},\ \boldsymbol{\gamma}_{j,k} \overset{ind.}{\sim} N_{q+1}\left(\mathbf{0}, \frac{2n}{\alpha(j+1)^2}(\mathbf{Z}'\mathbf{Z})^{-1}\right) \tag{19}$$

with cdf

$$G_{\mathbf{z}}(e) = p_{\mathbf{z}}\{k_\sigma(e)\}\left\{2^L\Phi_\sigma(e) - k_\sigma(e)\right\} + \sum_{k=1}^{k_\sigma(e)} p_{\mathbf{z}}(k), \tag{20}$$

where $\mathbf{Z}$ is the $n \times (q+1)$ design matrix with mean-centered covariates $\tilde{\mathbf{z}}_{ij}$s. Furthermore, the LDTFP is specified by setting $\boldsymbol{\gamma}_{0,1} \equiv \mathbf{0}$, such that for every $\mathbf{z} \in \mathcal{X}$, $G_{\mathbf{z}}$ is almost surely a median-zero probability measure. This is important to avoid identifiability issues. As shown by Jara and Hanson (2011), the LDTFP has appealing theoretical properties such as continuity as a function of the covariates, large support on the space of conditional density functions, straightforward posterior computation relying on algorithms for fitting generalized linear models, and the process closely matches conventional Polya tree priors (see, e.g., Hanson 2006) at each value of the covariate, which justify its choice here.

3.2. MCMC

Let $\Omega = (\mathbf{y}_c, \tilde{\boldsymbol{\beta}}, \mathbf{v}, \tau^2, \sigma^2, \boldsymbol{\gamma}, \alpha)$ denote collectively the model parameters to be updated, where $\mathbf{y}_c = \{y_{ij} : a_{ij} < b_{ij}\}$ are censored log-survival times. The $y_{ij} \in \mathbf{y}_c$, each component of $\tilde{\boldsymbol{\beta}}$, v_i and σ are all sampled using the single-variable slice sampling method (Neal 2003). For the LDTFP regression parameters $\gamma_{j,k}$, we utilize Metropolis-Hastings steps with Gaussian proposals based on iterative weighted least squares (Gamerman 1997), recognizing that the $\gamma_{j,k}$ full conditionals are proportional to logistic regression likelihoods. The hyperparameter τ^2 and α are sampled according to their conjugate full conditional distributions. A complete description of updating steps is available in Zhou *et al.* (2016).

The function `frailtyGAFT` sets the following hyperparameters as defaults: $\mathbf{m}_0 = \mathbf{0}$, $\mathbf{S}_0 = 10^5\mathbf{I}_p$, $a_0 = b_0 = 1$, $a_\tau = b_\tau = .001$, and $a_\sigma = 2 + \hat{\sigma}_0^4/(100\hat{v}_0)$, $b_\sigma = \hat{\sigma}_0^2(a_\sigma - 1)$, where $\hat{\sigma}_0^2$ and $\hat{v}_0$ are the estimates of σ^2 and its asymptotic variance from fitting the parametric lognormal AFT model, respectively. Note here we assume a somewhat informative prior on σ^2 so that its mean is $\hat{\sigma}_0^2$ and variance is $100\hat{v}_0$. For georeferenced data, we set and $a_\phi = \phi_0 b_\phi + 1$ and $b_\phi = 1$ so that the prior of ϕ has mode at ϕ_0. Here ϕ_0 satisfies $\rho(\mathbf{s}', \mathbf{s}''; \phi_0) = 0.001$, where $\|\mathbf{s}' - \mathbf{s}''\| = \max_{ij} \|\mathbf{s}_i - \mathbf{s}_j\|$. Note that Kneib and Fahrmeir (2007) simply fix ϕ at ϕ_0, while we allow ϕ to be random around ϕ_0. Again, the user-defined hyperparameters can be specified via the argument `prior`, e.g., for $p = 3$, $L = 5$, $\mathbf{m}_0 = \mathbf{0}$, $\mathbf{S}_0 = 10\mathbf{I}_{p+1}$, $a_0 = b_0 = 1$, and $a_\sigma = b_\sigma = 2$, the prior can be specified as below.

```
> prior=list(maxL=5,m0=rep(0,4),S0=diag(10,4),siga0=1,sigb0=1,a0=1,b0=1)
```

Given a set of posterior samples $\{\boldsymbol{\Omega}^{(s)}, s = 1, \ldots, S\}$, all the inference targets can be easily estimated. For example, the baseline survival function $S_{0,\mathbf{z}}(t) = P(e^{\beta_0 + \epsilon} > t|\mathbf{z})$ given the covariate $\mathbf{z}$ is estimated by

$$S_{0,\mathbf{z}}(t) = \frac{1}{S}\sum_{s=1}^{S}\left\{1 - G_{\mathbf{z}}^{(s)}\left(\log t - \beta_0^{(s)}\right)\right\}, \tag{21}$$

where $G_{\mathbf{z}}^{(s)}(\cdot)$ is given in (20) with all unknown parameters replaced by corresponding posterior values in the sth iterate.

3.3. Bayesian Hypothesis Testing

The GAFT frailty model includes the following as important special cases: an AFT frailty model with nonparametric baseline where $G_{\mathbf{z}} = G_{\mathbf{z}'}$ for all $\mathbf{z} = \mathbf{z}'$ and parametric baseline model $G_{\mathbf{z}} = N(0, \sigma^2)$ for all $\mathbf{z} \in \mathcal{X}$. Hypothesis tests can be constructed based on the LDTFP coefficients $\{\boldsymbol{\gamma}_{l,k} : k = 1, \ldots, 2^l, l = 1, \ldots, L-1\}$, where $\boldsymbol{\gamma}_{l,k} = (\gamma_{l,k,0}, \ldots, \gamma_{l,k,q})'$. Let $\boldsymbol{\gamma}_{l,k,-j}$ denote the subvector of $\boldsymbol{\gamma}_{l,k}$ without element $\gamma_{l,k,j}$ for $j = 0, \ldots, q$. Set $\boldsymbol{\Upsilon}_j = (\gamma_{l,k,j}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$, $\boldsymbol{\Upsilon}_{-j} = (\boldsymbol{\gamma}'_{l,k,-j}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$ and $\boldsymbol{\Upsilon} = (\boldsymbol{\gamma}'_{l,k}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$. Testing the hypotheses $H_0 : \boldsymbol{\Upsilon}_{-0} = \mathbf{0}$ and $H_0 : \boldsymbol{\Upsilon} = \mathbf{0}$ leads to global comparisons of the proposed model with the above two special cases respectively. Similarly, we may also test the null hypothesis $H_0 : \boldsymbol{\Upsilon}_j = \mathbf{0}$ for the jth covariate effect of $\mathbf{z}$ on the baseline survival, $j = 1, \ldots, q$.

Suppose we wish to test $H_0 : \boldsymbol{\Upsilon}_j = \mathbf{0}$ versus $H_1 : \boldsymbol{\Upsilon}_j \neq \mathbf{0}$, for fixed $j \in \{1, \ldots, q\}$. Following Zhou *et al.* (2016), the Bayes factor between hypotheses H_1 and H_0 can be approximated by

$$\hat{BF}_{10} = \frac{\prod_{l=1}^{L-1} \prod_{k=1}^{2^l} N\left(0 \middle| 0, \frac{2n}{\hat{\alpha}(l+1)^2} (\mathbf{Z}'\mathbf{Z})^{-1}_{jj}\right)}{N_{2^L-2}(\boldsymbol{\Upsilon}_j = \mathbf{0}; \hat{\mathbf{m}}_j, \hat{\mathbf{S}}_j)}, \tag{22}$$

where $N_p(\cdot; \mathbf{m}, \mathbf{S})$ denotes a p-variate normal density with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$, and $\hat{\mathbf{m}}_j$ and $\hat{\mathbf{S}}_j$ are the sample mean and covariance for $\boldsymbol{\Upsilon}_j$.

3.4. Leukemia Survival Data

The code below is used to fit the GAFT model with ICAR frailties for the leukemia survival data. As suggested by Zhou *et al.* (2016), the gamma prior $\Gamma(a_0 = 5, b_0 = 1)$ is used for α. We include all four covariates affect the baseline survival.

```
> mcmc=list(nburn=5000, nsave=2000, nskip=4, ndisplay=1000);
> prior=list(maxL=4, a0=5, b0=1);
> ptm<-proc.time()
> res1 = frailtyGAFT(formula=Surv(time,cens)~age+sex+wbc+tpi
+                    +baseline(age,sex,wbc,tpi)+frailtyprior("car",district),
+                    data=d,mcmc=mcmc,prior=prior,Proximity=E);
scan = 1000
scan = 2000
> sfit1=summary(res1); sfit1
Generalized accelerated failure time frailty model:
Call:
frailtyGAFT(formula = Surv(time, cens) ~ age + sex + wbc + tpi +
    baseline(age, sex, wbc, tpi) + frailtyprior("car", district),
    data = d, mcmc = mcmc, prior = prior, Proximity = E)

Posterior inference of regression coefficients
            Mean       Median     Std. Dev.  95%HPD-Low  95%HPD-Upp
intercept   8.561833   8.559441   0.139238   8.297258    8.843160
age        -0.050588  -0.050576   0.001988  -0.054635   -0.046877
sex        -0.263716  -0.275361   0.151775  -0.545586    0.009187
```

```
wbc         -0.004093  -0.004278   0.001013  -0.005650   -0.001764
tpi         -0.063257  -0.065633   0.022226  -0.099564   -0.019121

Posterior inference of scale parameter
       Mean    Median  Std. Dev.  95%HPD-Low  95%HPD-Upp
scale  2.1299  2.1250  0.0985     1.9640      2.3529

Posterior inference of precision parameter of LDTFP
        Mean    Median  Std. Dev.  95%HPD-Low  95%HPD-Upp
alpha   6.618   6.350   2.007      3.214       10.738

Posterior inference of conditional CAR frailty variance
          Mean    Median  Std. Dev.  95%HPD-Low  95%HPD-Upp
variance  0.3167  0.2880  0.1394     0.1093      0.5818

Bayes factors for LDTFP covariate effects:
intercept        age        sex        wbc        tpi    overall  normality
   1.9716    44.0512     1.1188    44.0041     0.5366    44.0503  6372.6117

Log pseudo marginal likelihood: LPML=-5936.359
Number of subjects:=1043
> systime1=proc.time()-ptm; systime1;
user  system elapsed
508.686   0.518 509.806
```

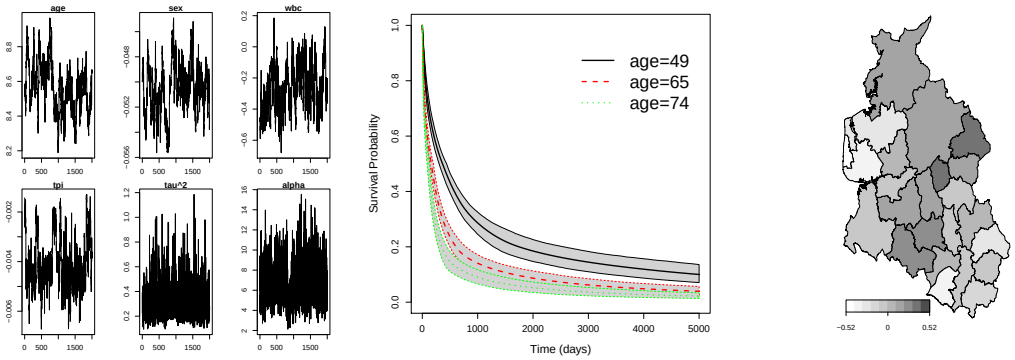
The Bayes factors for testing `age` and `wbc` effects on LDTFP are both 44, indicating that the baseline survival function under the AFT model significantly depends on `age` and `wbc`, and thus GAFT should be considered. The trace plots, survival curves and frailty map (Figure 5) can be obtained using the code similarly to Section 2.4. The only difference is that we need to specify the baseline covariates for plotting survival curves by including the arguement `xtfpred=xpred` into the `plot` function.

```
> estimates=plot(res1, xpred=xpred, xtfpred=xpred, tgrid=tgrid);
```

4. Survival Models via Spatial Copulas

Currently the package only supports spatial copula models for georeferenced (without replication, i.e. $n_i = 1$), right-censored spatial data. Suppose subjects are observed at n distinct spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Let t_i be a random event time associated with the subject at $\mathbf{s}_i$ and $\mathbf{x}_i$ be a related p-dimensional vector of covariates, $i = 1, \ldots, n$. For right-censored data, we only observe t_i^o and a censoring indicator δ_i for each subject, where δ_i equals 1 if $t_i^o = t_i$ and equals 0 if t_i is censored at t_i^o. Therefore, the observed data will be $\mathcal{D} = \{(t_i^o, \delta_i, \mathbf{x}_i, \mathbf{s}_i); i = 1, \ldots, n\}$. Note although the models below are developed for spatial survival data, non-spatial data are also accommodated.

The use of copulas in the spatial context was first proposed by Bárdossy (2006), where the empirical variogram is replaced by empirical copulas to investigate the spatial dependence



(a) (b) (c)

Figure 5: Leukemia survival data. GAFT model with ICAR frailties. (a) Trace plots for $\boldsymbol{\beta}$, τ^2 and α. (b) Survival curves with 95% credit interval bands for female patients with `wbc`=38.59 and `tpi`=0.3398 at different ages. (c) Map for the posterior mean frailties; larger frailties mean higher mortality rate overall.

structure. Copulas completely describe association among random variables separately from their univariate distributions and thus capture joint dependence without the influence of the marginal distribution (Li 2010). In the context of survival models, the idea of spatial copula approach is to first assume that the survival time t_i at location $\mathbf{s}_i$ marginally follows a model $S_{\mathbf{x}_i}(t)$, then model the joint distribution of $(t_1, \ldots, t_n)'$ as

$$P(t_1 \leq a_1, \ldots, t_n \leq a_n) = C(F_{\mathbf{x}_1}(a_1), \ldots, F_{\mathbf{x}_n}(a_n)), \tag{23}$$

where $F_{\mathbf{x}_i}(t) = 1 - S_{\mathbf{x}_i}(t)$ is the cumulative distribution function and the function C is an n-copula used to capture spatial dependence.

The current package assumes a spatial version of the Gaussian copula (Li 2010), defined as

$$C(u_1, \ldots, u_n) = \Phi_n\left(\Phi^{-1}\{u_1\}, \ldots, \Phi^{-1}\{u_n\}; \mathbf{R}\right), \tag{24}$$

where $\Phi_n(\cdot, \ldots, \cdot; \mathbf{R})$ denotes the distribution function of $N_n(\mathbf{0}, \mathbf{R})$. To allow for a nugget effect, we consider $\mathbf{R}[i,j] = \theta_1 \rho(d_{ij}; \theta_2) + (1 - \theta_1) I(\mathbf{s}_i = \mathbf{s}_j)$, where $\rho(d_{ij}; \theta_2) = \exp\{-\theta_2 d_{ij}\}$. Here $\theta_1 \in [0, 1]$, also known as a "partial sill" in Waller and Gotway (2004), is a scale parameter measuring a local maximum correlation, and θ_2 controls the spatial decay over distance. Note that all the diagonal elements of $\mathbf{R}$ are ones, so it is also a correlation matrix. Under the above spatial Gaussian copula, the likelihood function based on upon the complete data $\{(t_i, \mathbf{x}_i, \mathbf{s}_i), i = 1, \ldots, n\}$ is

$$\mathcal{L} = |\mathbf{R}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{R}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \prod_{i=1}^{n} f_{\mathbf{x}_i}(t_i), \tag{25}$$

where $z_i = \Phi^{-1}\left\{F_{\mathbf{x}_i}(t_i)\right\}$ and $f_{\mathbf{x}_i}(t)$ is the density function corresponding to $S_{\mathbf{x}_i}(t)$. We next discuss two marginal spatial survival models for $S_{\mathbf{x}_i}(t)$ that are accommodated in the package. Note that for large n, the FSA introduced in Section 2.1 (with ϵ replaced by $1 - \theta_1$) can be applied.

4.1. Proportional Hazards Model via Spatial Copulas

Assume that $t_i|\mathbf{x}_i$ marginally follows the proportional hazards (PH) model with cdf

$$F_{\mathbf{x}_i}(t) = 1 - \exp\left\{-\Lambda_0(t)e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\} \tag{26}$$

and density

$$f_{\mathbf{x}_i}(t) = \exp\left\{-\Lambda_0(t)e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\}\lambda_0(t)e^{\mathbf{x}_i'\boldsymbol{\beta}},$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, $\lambda_0(t)$ is the baseline hazard function and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the cumulative baseline hazard function. The piecewise exponential model provides a flexible framework to deal with the baseline hazard (e.g. Walker and Mallick 1997). We partition the time period $\mathbb{R}^+$ into M intervals, say $I_k = (d_{k-1}, d_k], k = 1, \ldots, M$, where $d_0 = 0$ and $d_M = \infty$. Specifically, we set $d_k = F_h^{-1}(k/M), k = 0, \ldots, M$, where $F_h(\cdot)$ is the cdf of exponential distribution with rate parameter h. The baseline hazard is then assumed to be constant within each interval, i.e.

$$\lambda_0(t) = \sum_{k=1}^{M} h_k I\{t \in I_k\},$$

where h_ks are unknown hazard values. Consequently, the cumulative baseline hazard function can be written as

$$\Lambda_0(t) = \sum_{k=1}^{M(t)} h_k \Delta_k(t),$$

where $M(t) = \min\{k : d_k \geq t\}$ and $\Delta_k(t) = \min\{d_k, t\} - d_{k-1}$. After incorporating spatial dependence via the copula in (24), the `spCopulaCoxph` function considers the following prior distributions:

$$\begin{aligned} \boldsymbol{\beta} &\sim N_p(\boldsymbol{\beta}_0, \mathbf{S}_0), \\ h_k|h &\overset{iid}{\sim} \Gamma(r_0 h, r_0), k = 1, \ldots, M, \\ (\theta_1, \theta_2) &\sim \text{Beta}(\theta_{1a}, \theta_{1b}) \times \Gamma(\theta_{2a}, \theta_{2b}) \end{aligned} \tag{27}$$

The `spCopulaCoxph` function sets the following default hyperparameter values: $M = 10$, $r_0 = 1$, $h = \hat{h}$, $\boldsymbol{\beta}_0 = \mathbf{0}$, $\mathbf{S}_0 = 10^5\mathbf{I}_p$, $\theta_{1a} = \theta_{1b} = \theta_{2a} = \theta_{2b} = 1$, where $\hat{h}$ is the maximum likelihood estimate of the rate parameter from fitting an exponential PH model. A function `indeptCoxph` is also provided to fit the non-spatial standard PH model with above baseline and prior settings.

4.2. Bayesian Nonparametric Survival Model via Spatial Copulas

We assume that $y_i = \log t_i$ given $\mathbf{x}_i$ marginally follows a linear dependent Dirichlet process mixture (LDDPM) model (De Iorio, Johnson, Müller, and Rosner 2009) with cdf,

$$F_{\mathbf{x}_i}(t) = \int \Phi\left(\frac{\log t - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) dG\{\boldsymbol{\beta}, \sigma^2\}, \tag{28}$$

where $\Phi(\cdot)$ is the cdf of the standard normal, and G follows the Dirichlet Process (DP) prior. This Bayesian nonparametric model treats the conditional distribution $F_{\mathbf{x}}$ as a function-valued

parameter and allows its variance, skewness, modality and other features to flexibly vary with the $\mathbf{x}$ covariates. After incorporating spatial dependence via the copula in (24), the function `spCopulaDDP` assumes the following prior distributions:

$$\begin{aligned} G = \sum_{k=1}^{N} w_k \delta_{(\boldsymbol{\beta}_k, \sigma_k^2)},\ w_k = V_k \prod_{j=0}^{k-1}(1-V_j),\ V_0 = 0, V_N = 1 \\ V_k \overset{iid}{\sim} \text{Beta}(1, \alpha), k = 1, \ldots, N,\ \alpha \sim \Gamma(a_0, b_0) \\ \boldsymbol{\beta}_k | \boldsymbol{\mu} \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), k = 1, \ldots, N,\ \boldsymbol{\mu} \sim N_p(\mathbf{m}_0, \mathbf{S}_0) \\ \sigma_k^{-2} | \boldsymbol{\Sigma} \overset{iid}{\sim} \Gamma(\nu_a, \nu_b), k = 1, \ldots, N,\ \boldsymbol{\Sigma}^{-1} \sim W_p\left((\kappa_0 \boldsymbol{\Sigma}_0)^{-1}, \kappa_0\right) \\ (\theta_1, \theta_2) \sim \text{Beta}(\theta_{1a}, \theta_{1b}) \times \Gamma(\theta_{2a}, \theta_{2b}). \end{aligned} \tag{29}$$

The following default hyperpriors are considered in `spCopulaDDP`: $a_0 = b_0 = 2$, $\nu_a = 3$, $\nu_b = \hat{\sigma}^2$, $\theta_{1a} = \theta_{1b} = \theta_{2a} = \theta_{2b} = 1$, $\mathbf{m}_0 = \hat{\boldsymbol{\beta}}$, $\mathbf{S}_0 = \hat{\boldsymbol{\Sigma}}$, $\boldsymbol{\Sigma}_0 = 30\hat{\boldsymbol{\Sigma}}$, and $\kappa_0 = 7$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2 from fitting the log-normal accelerated failure time model $\log(t_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\epsilon_i, \epsilon_i \sim N(0,1)$, and $\hat{\boldsymbol{\Sigma}}$ is the asymptotic covariance estimate for $\hat{\boldsymbol{\beta}}$. A function `indeptDDP` is also provided to fit the non-spatial LDDPM model in (28) with above prior settings.

4.3. Leukemia Survival Data

PH model with spatial copula

The following code is used to fit the piecewise exponential PH model (26) with the Gaussian spatial copula (24) using $M = 20$ and default priors. We consider $K = 100$ and $B = 1043$ for the number of knots and blocks in the FSA of **R**. The total running time is 15075 seconds.

```
> mcmc=list(nburn=5000, nsave=2000, nskip=4, ndisplay=1000);
> prior=list(M=20, nknots=100, nblock=1043);
> ptm<-proc.time()
> res1 = spCopulaCoxph(formula=Surv(time,cens)~age+sex+wbc+tpi,data=d,
+                      mcmc=mcmc,prior=prior,
+                      Coordinates=cbind(d$xcoord,d$ycoord));
> sfit1=summary(res1); sfit1
Spatial Copula Cox PH model with piecewise constant baseline hazards
Call:
spCopulaCoxph(formula = Surv(time, cens) ~ age + sex + wbc +
     tpi, data = d, mcmc = mcmc, prior = prior, Coordinates = cbind(d$xcoord,
     d$ycoord))

Posterior inference of regression coefficients
(Adaptive M-H acceptance rate: 0.2438):
      Mean        Median      Std. Dev.   95%CI-Low   95%CI-Upp
age   0.0282502   0.0282043   0.0018892   0.0246622   0.0319507
sex   0.0568409   0.0558295   0.0612507  -0.0663804   0.1774189
wbc   0.0028199   0.0028274   0.0004191   0.0020218   0.0036204
```

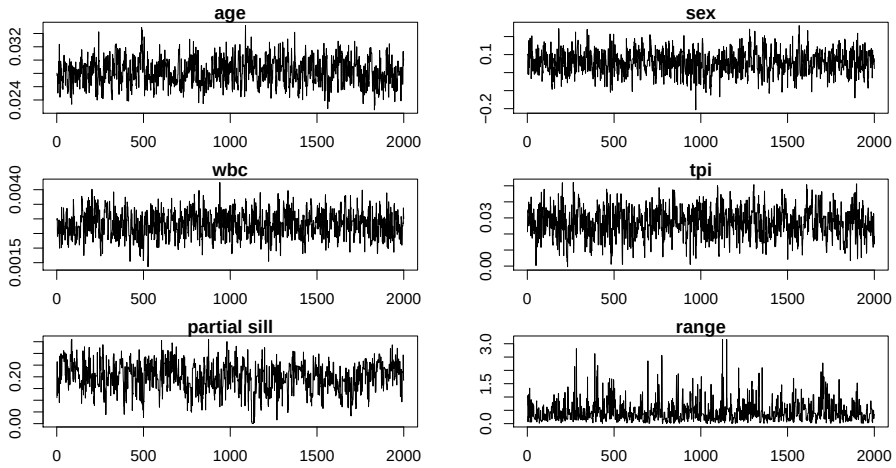


Figure 6: Leukemia survival data. Trace plots for $\boldsymbol{\beta}$, θ_1 and θ_2 under the PH model with spatial copula.

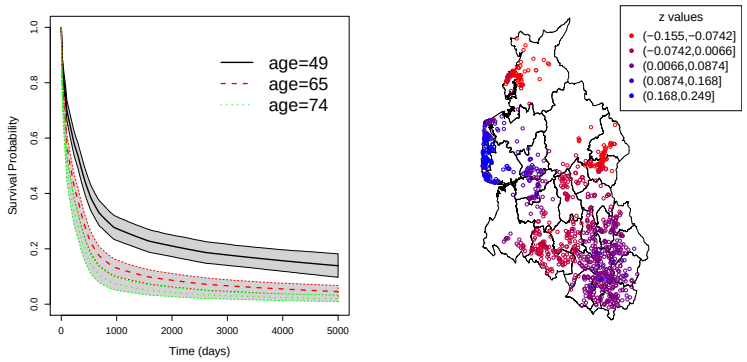
```
tpi    0.0265696    0.0265197    0.0089057    0.0084522    0.0442770

Posterior inference of spatial sill and range parameters
(Adaptive M-H acceptance rate: 0.1932):
       Mean     Median   Std. Dev.  95%CI-Low  95%CI-Upp
sill   0.19922  0.20228  0.05963    0.07301    0.29884
range  0.43634  0.33631  0.37704    0.02665    1.48672

Log pseudo marginal likelihood: LPML=-5931.167
Number of subjects: n=1043
> systime1=proc.time()-ptm; systime1;
     user    system   elapsed
14875.147   180.753 15075.110
```

The trace plots (Figure 6) and survival curves (Figure 7, panel a) can be obtained using the code similarly to Section 2.4, where the only difference is that we also present the trace plots for partial sill θ_1 and range θ_2.

```
> traceplot(mcmc(res1$theta[1,]), xlab="", main="partial sill")
> traceplot(mcmc(res1$theta[2,]), xlab="", main="range")
```

Note that the higher the value of $z_i = \Phi^{-1}\{F_{\mathbf{x}_i}(t_i)\}$ is, the longer the survival time t_i (i.e. lower mortality rate) would be. The posterior sample of z_is is saved in `res1$Zpred`. The following code is used to show the posterior mean of z_i values on the map (Figure 7, panel b).

```
> frail= round((rowMeans(res1$Zpred)),3); nclust=5;
> frail.cluster = cut(frail, breaks = nclust);
> frail.names = names(table(frail.cluster))
> rbPal <- colorRampPalette(c('red','blue'))
> frail.colors=rbPal(nclust)[as.numeric(frail.cluster)]
```



(a) (b)

Figure 7: Leukemia survival data. PH model with spatial copula. (a) Survival curves with 95% credit interval bands for female patients with `wbc`=38.59 and `tpi`=0.3398 at different ages. (b) Map for the posterior mean of z_i values; smaller z values mean higher mortality rate overall.

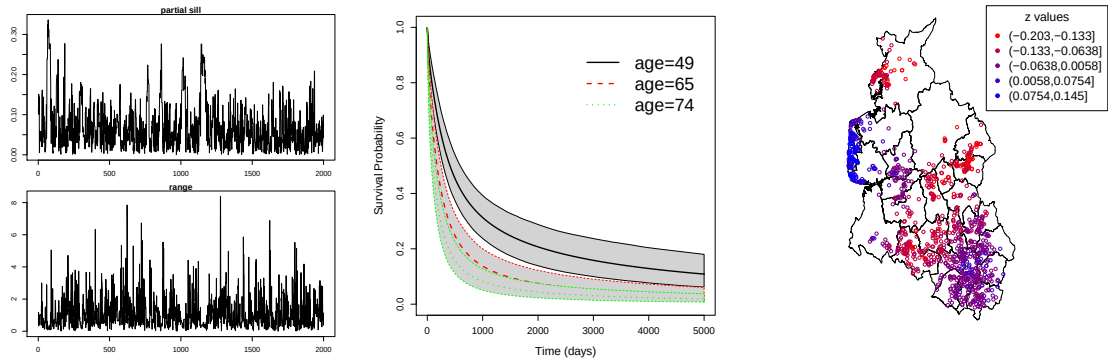
```
> par(mar=c(3,0,0,0))
> plot(nwengland)
> points(cbind(d$xcoord,d$ycoord), col=frail.colors)
> legend("topright",title="z values",legend=frail.names,
+        col=rbPal(nclust),pch=20, cex=1.7)
```

LDDPM model with spatial copula

The following code is used to fit the LDDPM model (28) with the Gaussian spatial copula (24) using $N = 10$ and default priors. For the FSA, $K = 100$ and $B = 1043$ are used. The total running time is 19491 seconds. Note this is no `summary` output as before, as we are fitting a nonparametric model.

```
> mcmc=list(nburn=5000, nsave=2000, nskip=4, ndisplay=1000);
> prior=list(N=10, nknots=100, nblock=1043);
> ptm<-proc.time()
> res1 = spCopulaDDP(formula=Surv(time,cens)~age+sex+wbc+tpi,data=d,
+                    mcmc=mcmc,prior=prior,
+                    Coordinates=cbind(d$xcoord,d$ycoord));
> systime1=proc.time()-ptm; systime1;
     user    system   elapsed
19310.243   177.188 19490.923
> sum(log(res1$cpo)); ## LPML $
[1] -5931.866
```

The trace plots, survival curves, and map of z_is (Figure 8) can be obtained using the same code used for the PH copula model, where the only difference is that we only present the trace plots for partial sill θ_1 and range θ_2.



(a) (b) (c)

Figure 8: Leukemia survival data. LDDPM model with spatial copula. (a) Trace plots for partial sill θ_1 and range θ_2. (b) Survival curves with 95% credit interval bands for female patients with `wbc`=38.59 and `tpi`=0.3398 at different ages. (c) Map for the posterior mean of z_i values; smaller z values mean higher mortality rate overall.

References

Antoniak CE (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems." *Annals of Statistics*, **2**, 1152–1174.

Arbia G, Espa G, Giuliani D, Micciolo (2016). "A spatial analysis of health and pharmaceutical firm survival." *Journal of Applied Statistics*, p. in press.

Banerjee S, Carlin BP, Gelfand AE (2014). *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. Chapman and Hall/CRC Press.

Banerjee S, Dey DK (2005). "Semiparametric proportional odds models for spatially correlated survival data." *Lifetime Data Analysis*, **11**(2), 175–191.

Bárdossy A (2006). "Copula-based geostatistical models for groundwater quality parameters." *Water Resources Research*, **42**(11), 1–12.

Besag J (1974). "Spatial interaction and the statistical analysis of lattice systems." *Journal of the Royal Statistical Society: Series B*, **36**(2), 192–236.

Chen Y, Hanson T, Zhang J (2014). "Accelerated hazards model based on parametric families generalized with Bernstein polynomials." *Biometrics*, **70**(1), 192–201.

Cox DR (1972). "Regression models and life-tables (with discussion)." *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 187–220.

Cox DR, Oakes D (1984). *Analysis of Survival Data*. Chapman & Hall: London.

Cox DR, Snell EJ (1968). "A general definition of residuals." *Journal of the Royal Statistical Society. Series B (Methodological)*, **30**(2), 248–275.

Darmofal D (2009). "Bayesian Spatial Survival Models for Political Event Processes." *American Journal of Political Science*, **53**(1), 241–257. ISSN 1540-5907. doi:10.1111/j.1540-5907.2008.00368.x. URL http://dx.doi.org/10.1111/j.1540-5907.2008.00368.x.

De Iorio M, Johnson WO, Müller P, Rosner GL (2009). "Bayesian nonparametric nonproportional hazards survival modeling." *Biometrics*, **65**(3), 762–771.

Finley AO, Sang H, Banerjee S, Gelfand AE (2009). "Improving the performance of predictive process modeling for large datasets." *Computational statistics & data analysis*, **53**(8), 2873–2884.

Gamerman D (1997). "Sampling from the posterior distribution in generalized linear mixed models." *Statistics and Computing*, **7**(1), 57–68.

Geisser S, Eddy WF (1979). "A Predictive approach to model selection." *Journal of the American Statistical Association*, **74**(365), 153–160.

Haario H, Saksman E, Tamminen J (2001). "An adaptive Metropolis algorithm." *Bernoulli*, **7**(2), 223–242.

Hanson T, Johnson W, Laud P (2009). "Semiparametric inference for survival models with step process covariates." *Canadian Journal of Statistics*, **37**(1), 60–79.

Hanson TE (2006). "Inference for Mixtures of Finite Polya Tree Models." *Journal of the American Statistical Association*, **101**(476), 1548–1565.

Hanson TE, Branscum AJ, Johnson WO, *et al.* (2014). "Informative g-Priors for Logistic Regression." *Bayesian Analysis*, **9**(3), 597–612.

Henderson R, Shimakura S, Gorst D (2002). "Modeling spatial variation in leukemia survival data." *Journal of the American Statistical Association*, **97**(460), 965–972.

Jara A, Hanson TE (2011). "A class of mixtures of dependent tailfree processes." *Biometrika*, **98**(3), 553–566.

Johnson ME, Moore LM, Ylvisaker D (1990). "Minimax and maximin distance designs." *Journal of statistical planning and inference*, **26**(2), 131–148.

Kneib T (2006). "Mixed model-based inference in geoadditive hazard regression for interval-censored survival times." *Computational Statistics & Data Analysis*, **51**(2), 777–792.

Kneib T, Fahrmeir L (2007). "A Mixed Model Approach for Geoadditive Hazard Regression." *Scandinavian Journal of Statistics*, **34**(1), 207–228.

Konomi BA, Sang H, Mallick BK (2014). "Adaptive Bayesian nonstationary modeling for large spatial datasets using covariance approximations." *Journal of Computational and Graphical Statistics*, **23**, 802–929.

Kuo L, Mallick B (1998). "Variable selection for regression models." *Sankhyā: The Indian Journal of Statistics, Series B*, **60**, 65–81.

Lavine M (1992). "Some Aspects of Polya Tree Distributions for Statistical Modelling." *The Annals of Statistics*, **20**, 1222–1235.

Li J (2010). "Application of Copulas as a New Geostatistical Tool." *Dissertation*.

Li J, Hong Y, Thapa R, Burkhart HE (2015). "Survival Analysis of Loblolly Pine Trees with Spatially Correlated Random Effects." *Journal of the American Statistical Association*, **110**(510), 486–502.

Li Y, Lin X (2006). "Semiparametric normal transformation models for spatially correlated survival data." *Journal of the American Statistical Association*, **101**(474), 591–603.

Müller P, Quintana F, Jara A, Hanson T (2015). *Bayesian Nonparametric Data Analysis*. Springer-Verlag: New York.

Nardi A, Schemper M (2003). "Comparing Cox and parametric models in clinical studies." *Statistics in Medicine*, **22**(23), 3597–3610.

Neal RM (2003). "Slice sampling." *Annals of Statistics*, **31**(3), 705–767.

Prentice RL, Kalbfleisch JD (1979). "Hazard rate models with covariates." *Biometrics*, **35**, 25–39.

Sang H, Huang JZ (2012). "A full scale approximation of covariance functions for large spatial data sets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74**(1), 111–132.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society, Series B*, **64**(4), 583–639.

Therneau T, Crowson C, Atkinson E (2016). "Using time dependent covariates and time dependent coefficients in the Cox model."

Turnbull BW (1974). "Nonparametric estimation of a survivorship function with doubly censored data." *Journal of the American Statistical Association*, **69**(345), 169–173.

Verdinelli I, Wasserman L (1995). "Computing Bayes factors using a generalization of the Savage-Dickey density ratio." *Journal of the American Statistical Association*, **90**(430), 614–618.

Walker SG, Mallick BK (1997). "Hierarchical Generalized Linear Models and Frailty Models with Bayesian Nonparametric Mixing." *Journal of the Royal Statistical Society, Series B: Methodological*, **59**, 845–860.

Waller LA, Gotway CA (2004). *Applied spatial statistics for public health data*. John Wiley & Sons.

Wang S, Zhang J, Lawson AB (2012). "A Bayesian normal mixture accelerated failure time spatial model and its application to prostate cancer." *Statistical Methods in Medical Research*, **http://dx.doi.org/10.1177/0962280212466189**.

Zhou H, Hanson T (2015). "Bayesian spatial survival models." In *Nonparametric Bayesian Inference in Biostatistics*, pp. 215–246. Springer.

Zhou H, Hanson T (2017). "A unified framework for fitting Bayesian semiparametric models to arbitrarily censored spatial survival data." *arXiv preprint arXiv:1701.06976*.

Zhou H, Hanson T, Jara A, Zhang J (2015a). "Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model." *The Annals of Applied Statistics*, **9**(1), 43–68.

Zhou H, Hanson T, Knapp R (2015b). "Marginal Bayesian Nonparametric Model for Time to Disease Arrival of Threatened Amphibian Populations." *Biometrics*, **71**(4), 1101–1110.

Zhou H, Hanson T, Zhang J (2016). "Generalized accelerated failure time spatial frailty model for arbitrarily censored data." *Lifetime Data Analysis*, **in press**.

Affiliation:

Haiming Zhou
Division of Statistics
Northern Illinois University
E-mail: zhouh@niu.edu
URL: http://niu.edu/stat/people/faculty/Zhou.shtml