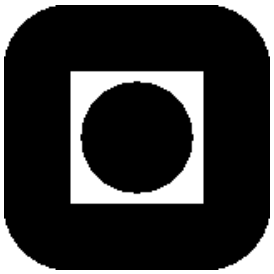
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Implementing Approximate Bayesian Inference for Survival Analysis using Integrated Nested Laplace Approximations

by

Rupali Akerkar, Sara Martino and Håvard Rue

PREPRINT
STATISTICS NO. 1/2010



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2010/S1-2010.pdf
Rupali Akerkar has homepage: http://www.math.ntnu.no/∼akerkar
E-mail: akerkar@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Implementing Approximate Bayesian Inference for Survival Analysis using Integrated Nested Laplace Approximations

Rupali Akerkar, Sara Martino and Håvard Rue
Department of Mathematical Sciences
NTNU, Norway

January 13, 2010

Abstract

In this report, we investigate the use of INLA, (Martino and Rue, 2008) to solve Bayesian inferential problems in Bayesisan Survival analsysis. In particular we consider the Exponential and Weibull-distributed lifetimes with and without censoring and frailty, and Cox-models with piecewise constant and piecewise linear baseline hazard. We demonstrate that all these models can (in most cases) be expressed as a latent Gaussian model (LGM) so that integrated nested Laplace approximations proposed by Rue et al. (2009) can be applied. We show comparison with the results obtained with INLA and those obtained with extensive runs with Markov chain Monte Carlo methods. The results obtained are again "practically exact" and support the general experience of Rue et al. (2009).

1 Background

Modern statistics has seen a great development over the last years, mainly within the Bayesian context and MCMC based inference; see for example Gamerman and Lopes (2006), Harrison and Stevens (1976), West et al. (1985), Gamerman (1998), etc. The Bayesian approach gives a unified approach to statistical modelling and inference, whereas MCMC sampling makes the magic possible; Parameters can be estimated and their uncertainties quantified, predictions can account for parameter and/or model uncertainty, and so on.

Despite the fact that simulation based inference *is* possible, it is not without serious problems in practice; Simple updating-schemes may produce slow converging/mixing samplers, which could seriously underestimate the variability of the random variables involved. Moreover, due to the large dimensions of the latent spatial field good block-updating schemes are often (relative) computationally expensive, whereas easier blocking schemes will still suffer from slow convergence/mixing. The computational expenses are also due to the Monte Carlo error itself, as accurate estimates require a long Markov chain run. In the context of survival analysis, there is a great body of work using latent Gaussian models; see for example Banerjee et al. (2004); Brezger et al. (2003); Fahrmeir (1994); Banerjee et al. (2003); Berzuini and Clayton (1994); Carlin and Banerjee (2003); Gamerman (1991). The important point, is that these models reduces to general latent Gaussian models although the structure of the likelihood depend on the (survival analysis) application.

Latent Gaussian models are subset of all Bayesian additive models with a structured additive predictor say η_i. In these models, the observation variable y_i is assumed to belong to an exponential family, where the mean μ_i is linked to this structured additive predictor η_i through a link function $g(.)$, so that $g(\mu_i) = \eta_i$. The

structured additive predictor η_i accounts for effects of various covariates in an additive way:

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i$$

Here, the $\{f^j(.)\}$s are unknown functions of the covariates **u**, the $\{\beta_k\}$s represent the linear effect of covariates **z** and the ϵ_is are unstructured terms. A Gaussian prior is assigned to α, $\{f^j(.)\}$, $\{\beta_k\}$ and $\{\epsilon_i\}$. We denote $\pi(.|.)$ as the conditional density of its arguments, and let **x** denote the vector of all the n Gaussian variables $\{\eta_i\}$, α, $\{f^j(.)\}$, and $\{\beta_k\}$, and $\boldsymbol{\theta}$ denotes the vector of hyperparameters, which are not necessarily Gaussian. The density $\pi(x|\boldsymbol{\theta_1})$ is Gaussian with (assumed) zero mean and precision matrix $Q(\boldsymbol{\theta_1})$ with hyperparameters $\boldsymbol{\theta_1}$.

The distribution for the n_d observational variables $\mathbf{y} = \{y_i : i \in I\}$ is denoted by $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_2)$ and we assume that $\{y_i : i \in I\}$ are conditionally independent given x and $\boldsymbol{\theta}_2$. For simplicity, denote by $\boldsymbol{\theta}$ = ($\boldsymbol{\theta}_1^T$, $\boldsymbol{\theta}_2^T$)^T with dim($\boldsymbol{\theta}$) = m. The posterior then reads (for nonsingular $\mathbf{Q}(\boldsymbol{\theta})$)

$$\begin{aligned}\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in I} \pi(y_i|x_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \mid \mathbf{Q}(\boldsymbol{\theta}) \mid^{n/2} \exp\Big(-\frac{1}{2}\mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{i \in I} \log \pi(y_i|x_i, \boldsymbol{\theta})\Big)\end{aligned}$$

The imposed linear constraints(if any) are denoted by $\mathbf{Ax} = \mathbf{e}$ for a $k \times n$ matrix $\mathbf{A}$ of rank k. The main aim is to approximate the posterior marginals of the latent field, $\pi(x_i|\mathbf{y})$ and the posterior marginals of the hyperparameters ,$\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$.

The latent Gaussian models we discuss in this paper satisfy two basic properties. The first is that the latent field **x** admit conditional independence properties. Hence, the latent field is a Gaussian Markov random field (GMRF) with a sparse precision matrix $\mathbf{Q}(\boldsymbol{\theta})$; see Rue and Held (2005). The second property is that the number of hyperparameters m is small, say $m \leq 6$. Both properties are usually required to produce fast inference.

In section 2, we have given short introduction to survival analysis, standard approach of doing MCMC in survival analysis: WinBUGS and new approach to perform approximate Bayesian inference using integrated nested Laplace approximations (INLA) introduced by Martino and Rue (2008). In further sections, we present worked out examples comparing the results for posterior marginals of the parameters for some survival models with (i) Gibbs sampler algorithm using the WinBUGS software and (ii) INLA (Integrated nested Laplace approximation). We demonstrate that all these models can (in most cases) be expressed as a latent Gaussian model (LGM) so that integrated nested Laplace approximations (Rue et al., 2009) can be applied. The purpose of these examples, is to compare the results to verify the quality of the determinstic approximations.

2 Survival Analysis

2.0.1 Introduction

The statistical analysis of survival data has been extensively studied in past few decades, particularly following the paper by (Cox, 1972) proposing a nonparamteric approach of regression models. In survival data analysis, models are generally specified through the hazard function. For any distribution function F(t) with density

function f(t), the survival function is $S(t) = P(T > t)$ and the corresponding hazard function $\lambda(t)$ is defined by means of a conditional probability

$$\lambda(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} P(t < T < t + \delta t \mid \ T > t)$$

thus we can write

$$\lambda(t) = \lim_{\delta t \to 0} \frac{S(t) - S(t + \delta t)}{S(t)} = \frac{f(t)}{S(t)}$$

Model proposed by Cox (1972) is

$$\lambda(t \mid \ z_1, \cdots, z_p) = \lambda_0(t) \ \exp(\ z_1\beta_1 + \cdots + \ z_p\beta_p)$$

Where, λ_0 is the baseline hazard function, z_i's are the covariates and β_i's are the regression parameters. For this model, the covariates are assumed to have fixed effects on failure pattern. A more comprehensive model is achieved by assuming

$$\begin{aligned} \lambda(t \mid \ z_1, \cdots, z_p) &= \lambda_0(t) \ \exp(\ z_1\beta_1(t) + \cdots + \ z_p\beta_p(t)) \\ &= \exp(z_0\beta_0(t) + \cdots + z_p\beta_p(t)) \end{aligned}$$

where $\beta_0(t) = \log(\lambda_0(t))$ and $z_0 = 1$. This model allows for effects that vary over time.
Different kinds of propotional hazard models are obtained by making assumptions about the baseline hazard function. For example, if the baseline hazard is constant over time, i.e., $\lambda_0(t)$ is simply λ_0 , we obtain exponential regression model, where

$$\begin{aligned} \lambda(t) &= \lambda_0 \exp(z_i^T \beta_i) \\ &= \exp(\log \lambda_0 + \ z_1\beta_1 + \cdots + \ z_p\beta_p) \end{aligned}$$

Another features of survival analysis are truncation and censoring. Truncation is about entering the study. It occurs when only those subjects whose event times lies within a certain observational time period $[T_l, T_r]$ are observed. No information is available about subjects whose event time is outside this frame.

- The observation is left truncated when T_r is infinite so the observation frame is $[T_l, \infty)$ and only those subjects are observed whose event time is greater than T_l, it is also called delayed entry.
- The observation is right truncated when $T_l = 0$ and so the observational frame is $(0, T_r]$, thus the subjects who have experienced the event are only observable.

The main variable of interest in survival analysis is *time-to-event*, a positive random variable. But the times-to-event are not always completely observed. These events are subject to censoring. A censored observation is defined as an observation with incomplete information. There are various types of censoring, such as right censoring, interval censoring and left censoring.

- The observation is uncensored if its failure time, say T is recorded.
- The observation is right censored if the censoring time (say C_r) is less than actual failure time, $C_r < T$ and this censoring time is being recorded.
- The observation is interval censored if the actual failure time is not observed exactly but it is known that $C_l < T < C_r$ and these C_l and C_r are known.

- The observation is left censored if the time to event is less than left censoring time, in other words if the event has already occurred before start of the study, i.e. $T < C_l$

Each observation can be represented by a triplet (C_l, C_r, δ) when censoring mechanism is used with

$$\begin{aligned} C_l = C_r = T, \delta = 1, &\quad \text{if the obs. is uncensored} \\ C_l = C_r = T, \delta = 0, &\quad \text{if the obs. is right censored} \\ C_l = C_r = T, \delta = 0, &\quad \text{if the obs. is left censored} \\ C_l < C_r, \delta = 0, &\quad \text{if the obs. is interval censored} \end{aligned}$$

Here we assumed that the lifetimes and censoring times are independent. The contribution to the likelihood function by an uncensored observation (for which exact event time is known) is the density at that duration and by a censored observation it is the survival function evaluated at that time point.The likelihood function is $L = \prod_{i=1}^{n} L_i$, where

$$L_i = \begin{cases} S(t_i)\lambda(t_i) = \lambda(t_i)\exp\{-\int_0^{t_i}\lambda(x)dx\} & \text{if obs. } i \text{ is uncensored} \\ S(C_i) = \exp\{-\int_0^{C_i}\lambda(x)dx\} & \text{if obs. } i \text{ is right censored} \\ S(C_i) = 1 - \exp\{-\int_0^{C_i}\lambda(x)dx\} & \text{if obs. } i \text{ is left censored} \\ S(C_l) - S(C_r) = \exp\{-\int_0^{C_l}\lambda(x)dx\}\{1 - \exp\big(-\int_{C_l}^{C_r}\lambda(x)dx\big)\} & \text{if obs. } i \text{ is interval censored} \end{cases}$$

For truncated data we are restricted to use conditional distribution in constructing the likelihood. The likelihood function for truncated data, say i with truncation interval $(T_L, T_R)$is given by

$$L_i = \begin{cases} f(t_i)/[S(T_l) - S(T_r)] & \text{if } i \text{ is interval-truncated and uncensored} \\ S(C_i)/[S(T_l) - S(T_r)] & \text{if } i \text{ is interval-truncated and censored} \\ f(t_i)/S(T_l) & \text{if } i \text{ is left truncated and uncensored} \\ S(C_i)/S(T_l) & \text{if } i \text{ is left truncated and censored} \\ f(t_i)/[1 - S(T_r)] & \text{if } i \text{ is right truncated} \end{cases}$$

An example:
Suppose we have a sample of n observations from an exponential distribution with parameter λ. Let t_i be the observation times, δ_i be the censoring indicator for unit i, let C_l and C_r be the left and right censoring time and let for truncation the observation time window is (T_l, T_r) . In exponential distribution, the hazard function is a constant and is equal to λ , the density function and the survival function for $T > 0$ are

$$f(t) = \lambda\exp(-\lambda t) \qquad S(t) = \exp(-\lambda t)$$

The corresponding components of the likelihoods for various censoring and truncation schemes may all be written as:

$f(t_i) = \lambda \exp(-\lambda t_i)$	exact lifetimes
$S(C_r) = \exp(-\lambda C_r)$	right censored data
$1 - S(C_l) = 1 - \exp(\lambda C_l)$	left censored data
$S(C_l) - S(C_r) = \exp(\lambda C_l) - \exp(-\lambda C_r)$	interval censored data
$f(t_i)/S(T_l) = \lambda \exp\{-\lambda(t_i - T_l)\}$	left truncated & uncensored data
$S(t_i)/S(T_l) = \exp\{-\lambda(t_i - T_l)\}$	left truncated & censored data
$f(t_i)/(1 - S(T_r)) = \lambda \exp(-\lambda t_i)/[1 - \exp(-\lambda T_r)]$	right truncated data
$f(t_i)/[S(T_l) - S(T_r)] = \lambda \exp(-\lambda t_i)/[\exp(-\lambda T_l) - \exp(-\lambda T_r)]$	interval truncated data

The likelihood function may be constructed by putting together the component parts as per the censoring/ truncation scheme.

2.1 Standard MCMC approach: WinBUGS

There are many software packages available for fitting either marginal or random effects models to correlated survival data, but the highly specialized standard tool for doing Bayesian analyis of survival models, is *WinBUGS*. WinBUGS is a software package for Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods. WinBUGS is composed of two seperate components: a component for specifying the model to be fitted and another component that runs the simulations to obtain the sample distributions of the parameters in the model Spiegelhalter et al. (2003). There are two options for specifying a model in WinBUGS. A model can be specified by either text statements, as in classic BUGS, or can be specified graphically by using graphical interface called doodleBUGS. Currently all versions of BUGS and WinBUGS are free and available from (http:// www.mrc-bsu.cam.ac.uk/bugs).

WinBUGS can fit a survival model with a very complex model structure. Once the data and model are specified, the package automatically sets up the sampling distributions and automatically select and implements the appropriate sampling algorithm for the unknown quantities in the model. BUGS generally uses Gibbs sampling, although it can use simple Metropolis-within-Gibbs schemes when needed.

It must be noted that of'course any Markov Monte Carlo scheme can be used instead of the WinBUGS program; the results are the same.

2.2 The new approach: INLA

A new instrument which allows the user to easily perform approximate Bayesian inference using integrated nested Laplace approximation is introduced by Rue et al. (2009). It provides fast deterministic alternative to Markov chain Monte Carlo (MCMC), which at moment is the standard tool for inference in such models. INLA compute posterior marginals for each component in model, from which posterior expectation and standard deviations can easily be found. The main aim of the latent Gaussian models we discussed earlier is to approximate the posterior marginals of $\pi(x_i|\mathbf{y})$, $\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$. The posterior marginal of interest can be written as

$$\pi(x_i \mid \mathbf{y}) = \int \pi(x_i \mid \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} \mid \mathbf{y}) \mathbf{d}\boldsymbol{\theta},$$

$$\pi(\theta_j \mid \mathbf{y}) = \int \pi(\boldsymbol{\theta} \mid \mathbf{y}) \mathbf{d}\boldsymbol{\theta}_{-j}$$

The key feature of the new approach is to use the above form to construct nested approximations, as this approach makes Laplace approximations very accurate when applied to latent Gaussian models

$$\tilde{\pi}(x_i \mid \mathbf{y}) = \int \tilde{\pi}(x_i \mid \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \mathbf{d}\boldsymbol{\theta},$$

$$\tilde{\pi}(\theta_j \mid \mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \mathbf{d}\boldsymbol{\theta}_{-j}$$

Here $\tilde{\pi}(\cdot \mid \cdot)$ is an approximated (conditional) density of its arguments. Approximations to $\pi(x_i|\mathbf{y})$ are computed by approximating $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ and $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ and using numerical integration to integrate out $\boldsymbol{\theta}$. The approximation of $\pi(\theta_j|\mathbf{y})$ is computed by integrating out $\boldsymbol{\theta}_{-j}$ from $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$. The posterior marginal $\pi(\boldsymbol{\theta}|\mathbf{y})$ of the hyperparameters $\boldsymbol{\theta}$ is approximated using a Laplace approximation

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi_G}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}$$

(Tierney and Kadane, 1986) where $\tilde{\pi_G}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to the full conditional of $\mathbf{x}$, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional for $\mathbf{x}$, for a given $\boldsymbol{\theta}$ (Rue and Held, 2005). The proportionality sign comes from the fact that the normalizing constant for $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$ is unknown. The approximations of the posterior marginals of the latent field are obtained using the finite sum

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y}) \Delta_k$$

The sum is evaluated at support points θ_k using appropriate weights Δ_k. Rue and Martino (2007) have discussed three different approaches with their features to approximate $\tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})$, namely a Gaussian, a full Laplace and a simplified Laplace approximation.

The main advantge of the INLA approach over MCMC is the great improvement in speed but also accurancy. All procedures necessary to perform INLA are efficiently implemented in the GMRFLib library. This is an open source library written in (ANSI)C and Fortran, which is freely available on the web page (http://www.math.ntnu.no/∼hrue/GMRFLib/). The components of the model and the options for the INLA procedures are specified through a ini file. The INLA program reads the ini file, then it builds and solves the model returning the required approximate posterior marginal densities and summary statistics. A R package called R-INLA is also available and it works as an interface to INLA program and its usage is similar to all other R functions. Both the INLA program and the R package are available for Unix, Windows and Mac and can also be freely downloaded from http://www..r-inla.org.

2.3 Working of the inla-R package for survival analysis

Time-to-event phenomena are presented in a very special way, as very often the data are censored due to some reason. Also some cases include a special feature called truncation. These aspects of survival data are very well included in R-INLA package.
To load R-INLA, use

```
>library(INLA)
```

The routines in R-INLA work with objects of class "inla.surv", which is a data structure that combines times, censoring and truncation information. Such objects are constructed using the 'inla.surv' function, which takes four arguments: an observation time, an event indicator(censoring indicator), an ending time for interval censored data and left truncation time.This can be done as follows:

```
>inla.surv(time, event, time2, truncation)
```

The observation time is the follow up time for right censored data and starting time for interval censored data. The event indicator can be coded as 1 for uncensored observation, 0 for right censored observation, 2 for left censored data and 3 for interval censored data. The left truncation time if missing is considered 0.
The out put of 'inla.surv' is a data frame consisting of 5 columns, the names of which can be seen using the command,

```
>names(inla.surv(time, event, time2, truncation))
```

resulting in:

```
[1] "time"       "lower"      "upper"      "event"      "truncation"
```

A print method is associated with the 'inla.surv'object that displays the objects in special format, with a '+' marking censoring observations. After the data object next stage is to give symbolic description of the model to be fit through an object "formula".

```
> formula = inla.surv(time, event, time2, truncation) ~ covariate
```

The main function of the package is "inla". It is similar to that of the *glm*() R function for solving generalized linear models. Once the formula is defined, we only have to call the inla() function specifying the likelihood family and some additional parameters. For example,

```
>model=inla(formula,family="exponential", data= d, verbose=TRUE,keep=TRUE)
```

The summary and some plots of the most relevant features of the fitted model is provided by:

```
>summary(model)
>plot(model)
```

Also, it is further possible to correct the hyperparameters, using

```
>h=inla.hyperpar(model)
>plot(h$marginals$'NAME OF HYPERPARAMETER')
```

3 Example1: Exponential model

Parametric models play an important role in Bayesian survival analysis. These models offer straight forward modeling and analysis techniques. In this section we will discuss exponential model for univariate right censored survival data.

3.1 The model

Suppose $\mathbf{t} = (t_1, t_2, ...t_n)^{'}$ are n i.i.d. survival times, each having an exponential distribution with parameter λ. Denote the censoring indicators by $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta n)^{'}$, where $\delta_i = 0$ if t_i is right censored and $\delta_i = 1$ if t_i is failure time. Let $f(t_i|\lambda) = \lambda \exp(-\lambda t_i)$ denote the density for t_i and $S(t_i|\lambda) = \exp(-\lambda t_i)$ denote the survival function for t_i.
To build a regression model, we introduce covariates through λ, and write $\lambda_i = \exp(\mathbf{z}_i^{'}\boldsymbol{\beta})$, where $\mathbf{z}_i$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters and $D = (n, \mathbf{t}, \mathbf{Z}, \boldsymbol{\delta})$ denote the observed data for regression model, where $\mathbf{Z}$ is the $n \times p$ matrix of covariates with i-th row as $\mathbf{z}_i$. Using all these we get the likelihood function as

$$\begin{aligned} L(\boldsymbol{\beta}|D) &= \prod_{i=1}^{n} f(t_i|\lambda_i)^{\delta_i} S(t_i|\lambda_i)^{(1-\delta_i)} \\ &= \prod_{i=1}^{n} \exp(\mathbf{z}_i^{'}\boldsymbol{\beta})^{\delta_i} \exp(-\exp(\mathbf{z}_i^{'}\boldsymbol{\beta})t_i)^{\delta_i} \exp(-\exp(\mathbf{z}_i^{'}\boldsymbol{\beta})t_i)^{1-\delta_i} \\ &= \exp\Big\{\sum_{i=1}^{n} \delta_i \mathbf{z}_i^{'}\boldsymbol{\beta}\Big\} \exp\Big\{-\sum_{i=1}^{n} t_i \exp(\mathbf{z}_i^{'}\boldsymbol{\beta})\Big\} \end{aligned}$$

If we assume a p dimensional normal prior for $\boldsymbol{\beta}$ with mean as $\boldsymbol{\mu}_0$ and covariance matrix as Σ_0.Then the posterior distribution of $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}|D) \propto L(\boldsymbol{\beta}|D)\pi(\boldsymbol{\beta}|\boldsymbol{\mu}_0, \Sigma_0)$$

3.2 Data: Times to Infection of Kidney Dialysis Patients

In a study Nahman et al. (1992), (given in the book by Klein and Moeschberger (2003)) designed to assess the time to first exitsite infection (in months) in patients with renal insufficiency, 43 patients utilized a surgically placed catheter (Group 1), and 76 patients utilized a percutaneous placement of their catheter (Group 2), a total of 119 patients. Where the variables represented in the data set are time to infection in months/10 denoted by t, Infection indicator (0=no, 1=yes) or event denoted by δ and catheter placement (1=surgically, 2=percutaneously) denoted by trt. First few lines of data are given below.

time	event	placement
0.15	1	1
0.35	1	1
0.45	1	1
0.45	1	1
0.55	1	1

The model for this example can be specified as:

$$t_i \sim \exp(\lambda_i)$$

Where i is from 1 to 119. t_i is the time to infection. The corresponding density function and survival function are as follows:

$$f(t_i \mid \lambda) = \lambda \exp(-\lambda t_i) \qquad S(t_i \mid \lambda) = \exp(-\lambda t_i)$$

For this example we have single covariate, catheter placement (trt) and therefore $\boldsymbol{\beta} = (\beta_0, \beta_1)^{'}$, where β_0 denotes the intercept term and β_1 is the coefficient for the catheter placement covariate (trt). Here, $\lambda_i = \exp(\beta_0 + trt_i\beta_1)$. Thus, the likelihood function is

$$L(\beta \mid D) = \exp\big\{\sum_{i=1}^{119} \delta_i(\beta_0 + trt_i^{'}\beta_1)\big\} \exp\big\{-\sum_{i=1}^{119} t_i \exp(\beta_0 + trt_i^{'}\beta_1)\big\}$$

3.3 Model specification in WinBUGS

In WinBUGS the model code for fitting the above mentioned model is as follows:

```
model {
   for(i  in  1 : n)
   {
       t[i] ~ dexp(lambda[i])
       lambda[i] <- exp(beta0 + trt[i] * beta1)
   }
   ## the  priors  are
   beta1 ~ dnorm(0.0,0.001)
   beta0 ~ dnorm(0.0,0.001)
}
```

The data is formated in the following way:
S-Plus format:
list($n = 119$)
Rectangular format:

```
t[]    t.cen[]     trt[]
0.15    0           1
0.35    0           1
0.45    0           1
0.45    0           1
0.55    0           1
```

3.4 Implementing using the $INLA$ package for R

```
 >library(INLA)
 >data=read.table("data.expo.txt", header = T)
 >inla.surv(time, event, time2, truncation)
# Here we have right censored data and thus,
  the time is represented in this way
 >inla.surv(data$time, data$event)
 >formula = inla.surv(data$time, data$event) ~ placement
 >model=inla(formula,family="exponential", data= data, verbose=TRUE,
  keep=TRUE )
 The summary of the fitted model is provided by:
 >summary(model)

Call:
c("inla.surv(formula = formula, family = "exponential", data = d,
 ", " ,
 verbose = TRUE, keep = TRUE)")

Fixed effects:
               mean        sd     0.025quant  0.975quant    kld dist.
placement -0.5335606  0.3968638    -1.312941   0.2447266         0
intercept  -0.6409269  0.5978231   -1.814960   0.5314603         0

The model has no random effects

The model has no hyperparameters
```

Also there is possibility to plot the most relevant features of the fitted model with

```
>plot(model)
```

3.5 The Result

The processing time using WinBUGS was 1930 seconds for 10^7 iterations after a 10^3 iteration burn-in. The processing time for INLA was 0.366 seconds. We see already in this example that the processing time is greatly improved which is of vital importance for the end-user. Posterior Summaries of intercept and fixed effect for Exponential model using WinBUGS and INLA are given below.

1. Intercept

	WinBUGS	INLA
mean	-0.6448	-0.6409
standard deviation	0.6076	0.5978

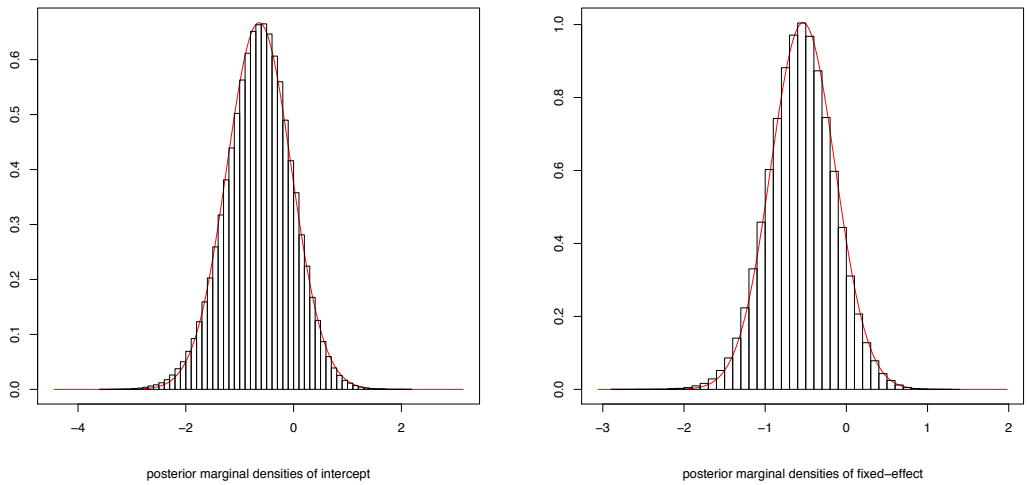


Figure 1: The graphs showing posterior marginals of intercept and fixed effect. The density shown by straight line is obtained using INLA and with histogram is using Winbugs.

2. Fixed effect

	WinBUGS	INLA
mean	-0.5586	-0.5336
standard deviation	0.405	0.3969

However, the improved processing time, might not be worth while, unless the accurancy of the posterior marginals are satisfactory. In Fig.1, we have compared the posterior marginals for the intercept and the fixed effect obtained using INLA with the one obtained using WinBUGS. As its clear from the figures, the approximations using the INLA approach are very accurate.

If we want to run WinBUGS until we get comparable results, then the improvement in processing time gets very very large. In Fig.1,there are two graphs for comparing posterior marginals of intercept and fixed effect. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

4 Example2: Weibull Model

4.1 The model

The Weibull model is perhaps the most widely used parametric survival model. Suppose we have independent identically distributed survival times $\mathbf{t}= (t_1, t_2, ..., t_n)^{'}$, each having a Weibull distribution, with parameters α, where $\alpha > 0$ and λ. The density for t_i is given by

$$f(t_i|\alpha, \lambda) = \alpha t_i^{\alpha-1} \exp(\lambda - \exp(\lambda) t_i^{\alpha})$$

The survival function is given by $S(t_i|\alpha, \lambda) = \exp(-\exp(\lambda) t_i^{\alpha})$. We can write the likelihood function of (α, λ) as

$$\begin{aligned} L(\alpha, \lambda \mid D) &= \prod_{i=1}^{n} f(t_i|\alpha, \lambda)^{\delta_i} S(t_i|\alpha, \lambda)^{(1-\delta_i)} \\ &= \alpha^{\sum \delta_i} \exp\left\{\lambda \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} (\delta_i(\alpha - 1)\log(t_i) - \exp(\lambda) t_i^{\alpha})\right\} \end{aligned}$$

Where δ_i is the indicator variable taking value 1 if t_i is failure time and 0 if t_i is right censored. To build the regression model, we introduce covariates through λ and write $\lambda_i = \mathbf{z}_i^{'}\boldsymbol{\beta}$. Where $\mathbf{z}_i$ is a $p \times 1$ vector of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. Assuming gamma prior with parameters (α_0, κ_0) for α and normal prior with parameters (μ_0, σ_0^2) for λ, the joint posterior distribution of (α, λ) is given by

$$\pi(\alpha, \lambda \mid D) \propto L(\alpha, \lambda \mid D)\pi(\alpha \mid \alpha_0, \kappa_0)\pi(\lambda \mid \mu_0, \sigma_0^2)$$

To build the Weibull regression model, we introduce covariates through λ and write $\lambda_i = \mathbf{z}_i^{'}\boldsymbol{\beta}$. The corresponding hazard function is then given by $h(t_i|\alpha, \lambda_i) = \alpha t_i^{\alpha-1} \exp(\lambda_i)$. If we assume $N_p(\boldsymbol{\mu_0}, \Sigma_0)$ prior for $\boldsymbol{\beta}$ and a gamma prior for α. Then the joint posterior is given by

$$\pi(\boldsymbol{\beta}, \alpha \mid D) \propto \alpha^{\alpha_0+d-1} \exp\Big\{ \sum_{i=1}^{n} (\delta_i + \mathbf{z}_i^{'}\boldsymbol{\beta} + \delta_i(\alpha - 1)\log(t_i)) - t_i^{\alpha} \exp(\mathbf{z}_i^{'}\boldsymbol{\beta})) - \kappa_0\alpha - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\Sigma_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\Big\}$$

where $D = (n, \mathbf{t}, \mathbf{Z}, \boldsymbol{\delta})$ denote the observed data for regression model, where $\mathbf{Z}$ is the $n \times p$ matrix of covariates with i-th row as $\mathbf{z}_i$ and $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_n)^{'}$.

4.2 Data-Times to Infection of Kidney Dialysis Patients

Like in the earlier example we use the same data used by Nahman et al. (1992), the time to first exitsite infection (in months) in patients with renal insufficiency, 43 patients utilized a surgically placed catheter (Group 1), and 76 patients utilized a percutaneous placement of their catheter (Group 2), a total of 119 patients.

Again for this example we used single covariate, catheter placement(trt) and therefore $\boldsymbol{\beta} = (\beta_0, \beta_1)^{'}$, where β_0 denotes the intercept term and $\beta_1$is the coefficient for the catheter placement covariate.

Further we assume normal priors for both β_0 and β_1 with $mean = 0$ and $precision = 0.001$ and for α the shape parameter of Weibull we assume gamma prior with parametres 1 and 0.001. The model for this example can be specified as:

$$t_i \sim \text{Weibull}(\alpha, \lambda_i)$$

Where i is from 1 to 119. t_i is the time to infection. The corresponding density function and survival function are as follows:

$$f(t_i \mid \alpha, \lambda) = \alpha t^{\alpha-1} \exp(\lambda - \exp(\lambda) t^{\alpha}), \qquad S(t_i \mid \alpha, \lambda) = \exp(-\exp(\lambda) t^{\alpha})$$

For this example we have single covariate, catheter placement (trt) and therefore $\beta = (\beta_0, \beta_1)^{'}$, where β_0 denotes the intercept term and β_1 is the coefficient for the catheter placement covariate (trt). Thus, the likelihood function with $\lambda_i = \exp(\beta_0 + trt_i\beta_1)$ is

$$L(\alpha, \beta \mid D) = \alpha^{\sum d_i} \exp\left\{\lambda \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} (\delta_i(\alpha - 1)\log(t_i) - \exp(\lambda) t_i^{\alpha})\right\}$$

4.3 Model specification in WinBUGS

In WinBUGS the model code for fitting the above mentioned model is as follows:

```
 model {
     for(i  in  1 : n)
     {
        t[i] ~ dweib(alpha,  lambda[i]) I(t.cen[i],)
        lambda[i] <- exp(beta0 + trt[i] * beta1 )
     }
#the  priors  are
   beta1 ~ dnorm(0.0,0.001)
   beta0 ~ dnorm(0.0,0.001)
   alpha ~ dgamma(1, 0.001)
}
```

The data are same as earlier example.

4.4 Implementing using the $INLA$ package for R

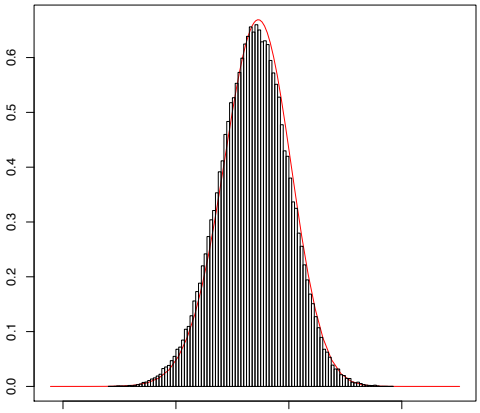
```
 >library(INLA)
 >data=read.table("data.txt", header = T)
 >inla.surv(data$time, data$event)
 >formula = inla.surv(data$time, data$event)~ placement
 > model=inla(formula, family="weibull", data= data, verbose=TRUE,
   keep=TRUE )
 >summary(model)
Call:
c("inla.surv(formula = formula, family = "weibull", data = d, ", "
    , verbose = TRUE, keep = TRUE)")

Fixed effects:
               mean     sd      0.025quant 0.975quant    kld dist.
placement -0.5474926  0.3972345  -1.327611  0.2315099         0
intercept -0.6009739  0.6008539  -1.781071  0.5772280         0

The model has no random effects

Model hyperparameters:
                            mean   sd     0.025quant 0.975quant
Alpha.parameter.for.Weibull 0.916  0.149  0.658      1.241
```

4.5 The Result

The processing time using WinBUGS was approx. 1900 seconds for 3 x 10^6 iterations after a 10^3 iteration burn-in. The processing time for INLA was 0.165 seconds. The time required for improving hyperparameters using inla.hyperpar was 0.235 seconds.

Posterior estimates of beta0 and beta1 and alpha for Weibull model using WinBUGS and INLA are given below.

1. Intercept



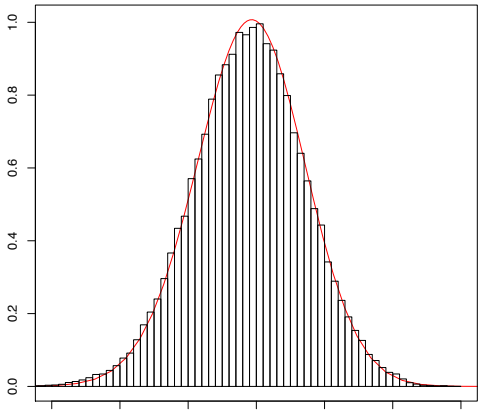


Figure 2: The graphs showing posterior marginals of intercept and fixed effect. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

	WinBUGS	INLA
mean	-0.5981	-0.5841
standard deviation	0.6082	0.0.6008

2. Fixed effect

	WinBUGS	INLA
mean	-0.577	-0.5475
standard deviation	0.4051	0.3972

3. alpha

	WinBUGS	INLA
mean	0.8991	0.9040
standard deviation	0.1478	0.146

The graphical comparison of results from two methods is given in Fig.2 and Fig.3. There are three graphs for comparing posterior marginals of intercept effect, fixed effect and weibull parameter. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS. As it is clear from the figures, the approximations using the INLA approach is quite accurate. If we want to run WinBUGS until we get exact results, then the improvement in processing time gets very very large.

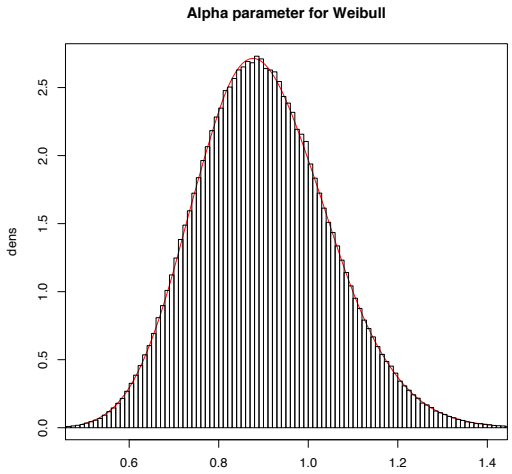


Figure 3: The graphs showing posterior marginals of weibull parameter.The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

5 Example3 - Semiparametric Models- Piecewise constant Hazard model

In the earlier sections we discussed the parametric models for univariate right censored survival data. In this section we will discuss the semiparametric models. The popularity of semiparametric approaches for analyzing univariate data begins with the seminal paper of (Cox, 1972) on the proportional hazard model. In parametric methods, the baseline hazard function is assigned a parametric form whereas in semiparametric approach, the baseline hazard function is assigned a nonparametric prior. One of the most convenient and popular models for semiparametric models is the piecewise constant hazard model. A piecewise constant baseline hazard model is an exponential hazard rate model where the costant rate is allowed to vary within pre-defined time intervals or segments.
In this section we will consider fitting of a proprtional hazards model of the usual form,

$$h_i(t \mid \mathbf{z}_i) = h_0(t) \ \exp(\mathbf{z}_i \boldsymbol{\beta})$$

where $h_i(t \mid \mathbf{z}_i)$ is the hazard at time t for an individual i with covariates $\mathbf{z}_i$, the corresponding regression parameters $\boldsymbol{\beta}$ and $h_0(t)$ is a baseline hazard function.

5.1 Model

To construct piecewise constant baseline hazard model, we first partition the time axis into J intervals with cutpoints $0 = s_1 < s_2 < ...s_J$, define the k-th interval as $(s_k, s_{k+1}]$. We then assume that the baseline hazard is constant within each interval, so that,

$$h_0(t) = \lambda_k \quad \text{for} \quad t \in I_k = (s_k, s_{k+1}]$$

where t is the event time.
Breslow (1972) and Breslow (1974) proposed the use of Piecewise exponential distributions in survival data analysis with successive death times to replace the base-line in (Cox, 1972). Holford (1980) and Laird and Oliver (1981) in their papers independently noted that the piecewise hazard model was equivalent to a certain Poisson regression model. Kalbfleisch and Prentice (1973) suggested that the cutpoints should be selected independently of the data and we have also assumed the same.

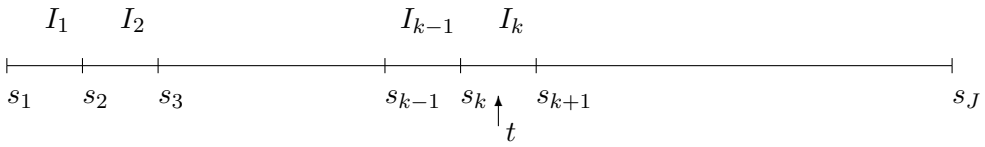


Figure 4: The time line is partitioned into J non overlapping intervals.

The observed survival times may be terminated either by failure or by censoring. It is assumed that the times of failure are independent of the times of censoring. If the individual lived beyond the end of the interval then the time lived in the interval equals the width of the interval. Let t_i denote the time lived by the i-th individual, for i is $1, 2, \ldots, n$. The baseline survival is then

$$\begin{aligned} S_{0i}(t) &= \exp\{-\int_0^t h_0(u)du\} \\ &= \exp\{-\sum_{j=1}^{k-1}(s_{j+1} - s_j)\lambda_j - (t - s_k)\lambda_k\} \end{aligned}$$

for $t \in I_k$, i.e., the i-th individual lived until or censored in k-th interval. In general the hazard rate for i-th individual in j-th interval is given by

$$\begin{aligned} h_i^{(j)}(t) &= h_0(t)\exp(\mathbf{z}_i^T\boldsymbol{\beta}) \\ &= \exp\{\mathbf{z}_i^T\boldsymbol{\beta} + \log h_0(t)\} \\ &= \exp\{\mathbf{z}_i^T\boldsymbol{\beta} + \log \lambda_j\} \end{aligned}$$

Let $\eta_i^{(j)} = \mathbf{z}_i^T\boldsymbol{\beta} + \log\lambda_j$ and let $\delta_i^{(j)}$ is the indicator of death/failure of ith individual in jth interval(taking the value 1 if individual died and 0 if censored). The log-likelihood function for i-th observation with time t, $t \in I_k$ can be written as

$$\begin{aligned} \log[h_i(t)^{\delta_i}S_i(t)] &= \delta_i^{(k)}\eta_i^{(k)} - \{\sum_{j=1}^{k-1}(s_{j+1}-s_j)\exp(\eta_i^{(j)}) + (t-s_k)\exp(\eta_i^{(j)})\} \\ &= \delta_i^{(j)}\eta_i^{(k)} - (t-s_k)\exp(\eta_k) - \sum_{j=1}^{k-1}(s_{j+1}-s_j)\exp(\eta_i^{(j)}) \end{aligned}$$

- $\delta_i^{(j)}\eta_i^{(k)} - (t-s_k)\exp(\eta_i^{(k)})$: can be seen as the log-likelihood from Poisson with mean $(t-s_k)\exp(\eta_k)$ observed to be 0 or 1 according to $\delta_i^{(j)}$.
- $-\sum_{j=1}^{k-1}(s_{j+1}-s_j)\exp(\eta_i^{(j)})$: can be seen as the likelihood from $k-1$ Poisson with mean $(s_{j+1}-s_j)\exp(\eta_i^{(j)})$ observed to be zero.

Each data point t_i is written as k "augmented data points" y_{i1},....,y_{ik} coming from Poisson distribution.

5.2 Data -Survival times of gastric cancer patients

The data used here consists of the survival times of 90 gastric cancer patients equally divided in two treatment groups, which are combined treatment (chemotherapy and radiation) and chemotherapy alone. It is taken from Stablein et al. (1981). Treatment effect is the only covariate used. To construct piecewise constant hazard model, we divide the time axis into two intervals of equal width, with cutpoints at $(0, 900, 1800)$. Thus the baseline survival for the i-th individual is

$$S_{0i}(t) = \begin{cases} = \exp\{-t\lambda_1\} & \text{if} \quad t \in (0, 900] \\ = \exp\{-900\lambda_1 - (t-900)\lambda_2\} & \text{if} \quad t \in (900, 1800) \end{cases}$$

The hazard rate is $h_i(t) = \exp\{\beta_0 + trt_i\beta_1 + \log\lambda_i^{(j)}\}$, where β_0 is the intercept term, β_1 is the coefficient for treatment effect, $\lambda_i^{(j)}$ is the baseline hazard for i-th individual in j-th interval, j is either 1 or 2 and i is from 1 to 90. The loglikelihood function for the i-th individual with observation time t_i is

$$l_i = \begin{cases} \delta_i^{(1)}\eta_i^{(1)} - t_i\exp(\eta_i^{(1)}) & \text{if} \quad 0 \le t_i \le 900 \\ \delta_i^{(2)}\eta_i^{(2)} - 900\exp(\eta_i^{(1)}) - (t_i-900)\exp(\eta_i^{(2)}) & \text{if} \quad 900 \le t_i \end{cases}$$

Where $\eta_i^{(j)} = \beta_0 + trt_i\beta_1 + \log\lambda_j$, $j = 1, 2$.
Further we assined flat prior for β_0 and normal prior for β_1 with mean $= 0$ and precision $= 0.001$ and for $\log\lambda_j$, we assigned random walk prior with precision as τ. The hyperparameter τ is assigned a gamma prior with parameters 1 and 0.001.

5.3 Model specification in WinBUGS

In WinBUGS the model code for fitting the above mentioned model is as follows:

```
model {
    for (j in 1 :T ) {
       for(i  in  1 : N) {
         y[j,i] ~ dpois(eta[j,i] * t[j,i])
        log(eta[j,i]) <- beta0 + trt[i] * beta1 + lambda[j]
       }
    }
   # the priors are
   beta0 ~ dflat()
   beta1 ~ dnorm(0.0,0.001)
   lambda[1:T] ~ car.normal(adj[], weights[], num[],tau0)
   tau0 ~ dgamma(1,0.001)
   logtau0 <- log(tau0)
}
```

Where,

- T = Number of time intervals or number of cutpoints. In this example the number of time intervals considered are 2 or in terms of cutpoints these are 0, 900, 1800.
- N = Number of individuals. In this example N=90.
- y[j,i] = δ_{ji} is the indicator of death of ith individual in j th interval.
- t[j,i] = time survived by i th individual in j th interval.
- adj[] = adjacency vector listing neighbouring time points.
- weights[] = weight vector.
- num[] = number of neighbours

5.4 Implementing using the $INLA$ package for R

```
>library(INLA)
>data=data.frame(time,event,trt)
>inla.surv(time,event)
>formula = inla.surv(time,event) ~ trt
>model = inla(formula,family="piecewise.constant",control.hazard=
list(model="rw1", cutpoints = c(0,900,1800)),data=data,keep=T)
>summary(model)
>plot(model)
# For further improving hyperparameters,
>h=inla.hyperpar(model)
>plot(h$marginals$'Precision for baseline.hazard')
```

5.5 Result

The processing time using WinBUGS was 685 seconds for 5 x 10^6 iterations after a 10^3 iteration burn-in. The processing time for INLA was 0.090 seconds.

Posterior estimates of intercept, fixed effect, random effect and log precision for Piecewise constant hazard model using WinBUGS and INLA are given below.

1. Intercept

	WinBUGS	INLA
mean	-6.426	-6.41689
standard deviation	0.1652	0.16490222

2. Fixed effect

	WinBUGS	INLA
mean	-0.1291	-0.12958986
standard deviation	0.226	0.22273412

3. Hazard

	WinBUGS	INLA
mean0	0.011	0.014707346
mean1	-0.011	-0.014745855
standard deviation0	0.04057	0.046269842
standard deviation1	0.04057	0.046397741

4. Log precision

	WinBUGS	INLA
mean	6.289	6.2954938
standard deviation	1.31	1.3054785

The graphical comparison of results from two methods are given in Fig.5 and Fig.6. There are four graphs for comparing posterior marginals of intercept effect, fixed effect, random effect (hazard) and log precision of the hyperparameter. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

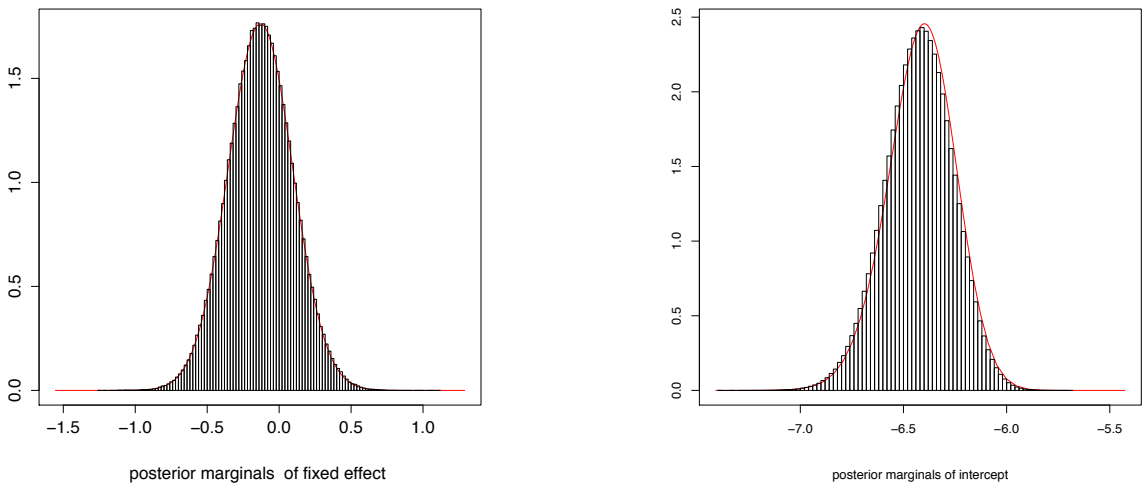


Figure 5: The graphs showing posterior marginals of intercept and fixed effect. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

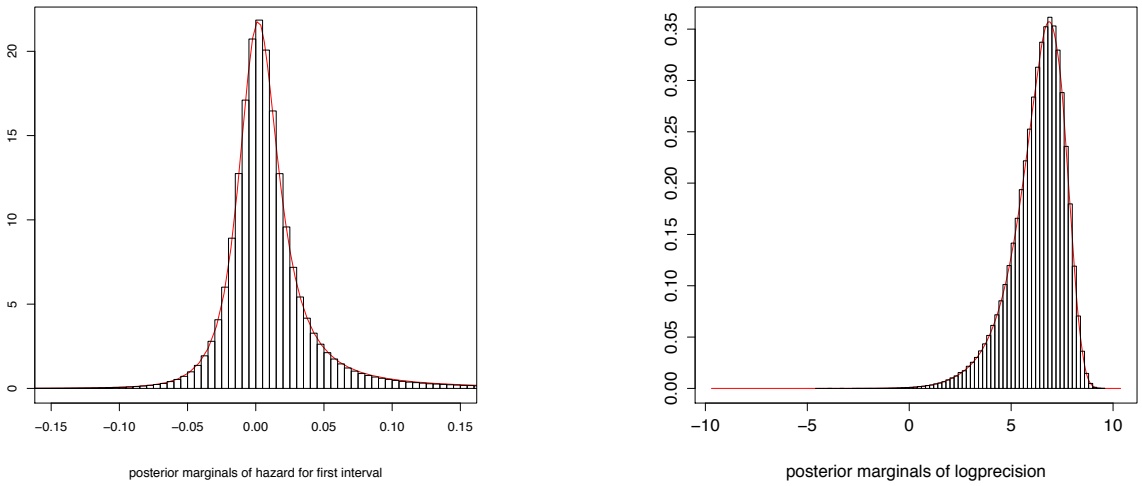


Figure 6: The graphs showing posterior marginals of log baseline hazard and log precision. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

6 Example4 - Frailty Model

In this section we will discuss the proportional hazard model with frailties especially the Weibull model with additive frailties.
The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data. The term frailty was introduced by Vaupel et al. (1979) in univariate survival models. Essentially these models extend the Cox proportional hazards model, discussed by Cox (1972), by adding random effects into the baseline hazard to model the intracluster correlation.

6.1 The model

Let t_{ij} be the survival time for the j-th individual in the i-th cluster , $i = 1, ..., n$, and $j = 1, ...m_i$. Here the m_i's represent the number of individual in the i-th cluster. We assumed that these t_{ij} follow i.i.d. Weibull distribution such that

$$t_{ij} \sim W(\alpha, \eta_{ij}), \quad \alpha > 0,$$

For the frailty models the conditional hazard function of t_{ij} given the unobserved frialty b_i for the i-th cluster, the fixed covariate vector $\boldsymbol{z}_{ij}$ and the weibull parameter α is given by

$$h(t_{ij}|z_{ij}, b_i, \alpha) = \alpha t_{ij}^{\alpha-1} \exp(\eta_{ij})$$

with

$$\eta_{ij} = \beta_0 + \boldsymbol{z}_{ij}^T \boldsymbol{\beta} + b_i$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and β_0 is denoting intercept and $\mathbf{z}_{ij}$ is a $p \times 1$ covariate vector. This is the modeling strategy used in BUGS (Spiegelhalter et al. 1995), is based on an additive frailty model. The complete data likelihood is given by

$$L(\beta, \alpha \mid D) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left(\alpha t_{ij}^{\alpha-1} \exp(\eta_{ij})\right)^{\delta_{ij}} \exp\left(-\exp(\eta_{ij}) t_{ij}^{\alpha}\right)$$

where, δ_{ij} is the censoring indicator, taking value 1 if the j-th individual of the i-th cluster dies and 0 otherwise and $D = (\mathbf{t}, \mathbf{Z}, \boldsymbol{\delta}, \mathbf{b})$ denotes the complete data set with $\mathbf{t} = (t_{11}, \cdots, t_{nm_n})'$, $\mathbf{Z} = (\mathbf{z}_{11}, \cdots, \mathbf{z}_{nm_n})$, $\boldsymbol{\delta} = (\delta_{11}, \cdots, \delta_{nm_n})'$ and $\mathbf{b} = (b_1, \cdots, b_n)$.

6.2 Data

Mantel et al. (1977) reports the results of a litter-matched study of the tumorigenesis of a drug. In this experiment, rats were taken from 50 distinct litters, and one rat of the litter was randomly selected and given the drug. For each litter, 2 rats were selected as controls and were given placebo. All mice were females. Possible associations between litter mates in their time to development of tumors may be due to common genetic backgrounds shared by siblings.
Here we have single covariate treatment denoting whether the rat was fed with drug(1) or placebo(0). Time to tumor development (in days), indicator of tumor development (1=yes, 0=no) and litters are given in data. A portion of data is given below.

time	indicator	treatment	litter
101	0	1	1
104	0	1	2
104	0	1	3
77	0	1	4
89	0	1	5
88	1	1	6
104	1	1	7
96	1	1	8

The model for this example with one covariate (treatment effect) is:

$$t_{ij} \sim W(\alpha, \eta_{ij})$$

Where i is 1 to 3, j is 1 to 50 and $\log(\eta_{ij}) = \beta_0 + trt_{ij}\beta_1 + b_i$, we assumed normal priors for β_0 and β_1 with parameters 0 and 0.001, assume $b_i \sim N(0, \tau)$ and to complete the model we assumed Gamma prior for τ and α with parameters 1 and 0.001.

6.3 Model specification in WinBUGS

In WinBUGS the model code for fitting the above mentioned model is as follows:

```
 model {
    for (i in 1 :N )
       for(j  in  1 : M) {
            t[i,j] ~ dweib(alpha, eta[i,j])I(t.cen[i, j])
            log(eta[i,j]) <- beta0 + beta1 * trt[i,j]  + b[i]
            }
          b[i] ~ dnorm(0.0, tau)
        }
   # the priors are
     beta0 ~ dnorm(0.0, 0.001)
     beta1 ~ dnorm(0.0,0.001)
     tau ~ dgamma(1,0.001)
     alpha ~ dgamma(1,0.001)
}
```

6.4 Implementing using the $INLA$ package for R

In this example we also used the $inla$ package for **R**. Using the INLA package all data files are automatically build by the package itself starting from a usual R data frame.

```
>library(INLA)
>Rats=read.table("data.txt",header =TRUE)
>time=Rats$time
>event=Rats$event
>trt=Rats$trt
>group=Rats$group
>inla.surv(time, event)
>formula = inla.surv(time,event) ~ trt + f(group,model="iid", param =
```

```
            c(1, 0.001), initial=0.01 )
>model = inla(formula, family="weibull", data = Rats,
         verbose=T,control.data=list(param=c(1,0.001),fixed=F,
         initial =  0.3), control.inla = list(int.strategy="grid",
         diff.logdens=20, dz=0.5), keep=T )
>summary(model)
>plot(model)
# To improve further the posterior marginals of hyperparameters,
>h=inla.hyperpar(model)
>summary(h)
# and for plotting,
>plot(h$marginals$'Alpha parameter for Weibull')
>plot(h$marginals'Precision for group')
```

6.5 Result

The processing time using WinBUGS was 1219 seconds for 10^6 iterations after a 10^3 iteration burn-in. The processing time for INLA was 4.972 seconds and for correction of hyperparametrs was 0.235sec. Posterior estimates of intercept, fixed effect, random effect and log precision for frailty model using WinBUGS and INLA are given below.

1. Intercept

	WinBUGS	INLA
mean	-1.464	-1.4405
standard deviation	0.2338	0.2310

2. Fixed effect

	WinBUGS	INLA
mean	0.9095	0.9064
standard deviation	0.3212	0.3154

3. Weibull parameter

	WinBUGS	INLA
mean	3.8630	3.8638
standard deviation	0.5505	0.5462

4. Precision of random effect(frailty)

	WinBUGS	INLA
mean	983.1	992.4655
standard deviation	998.9855	989.1115

The graphical comparison of results from two methods is given in Fig.7 and Fig.8. There are four graphs comparing posterior marginals of intercept , age effect, random effect (frailty) and log precision of the hyperparameter. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS. As we have seen in earlier examples, the approximations using the INLA approach are less time consuming and quite accurate.

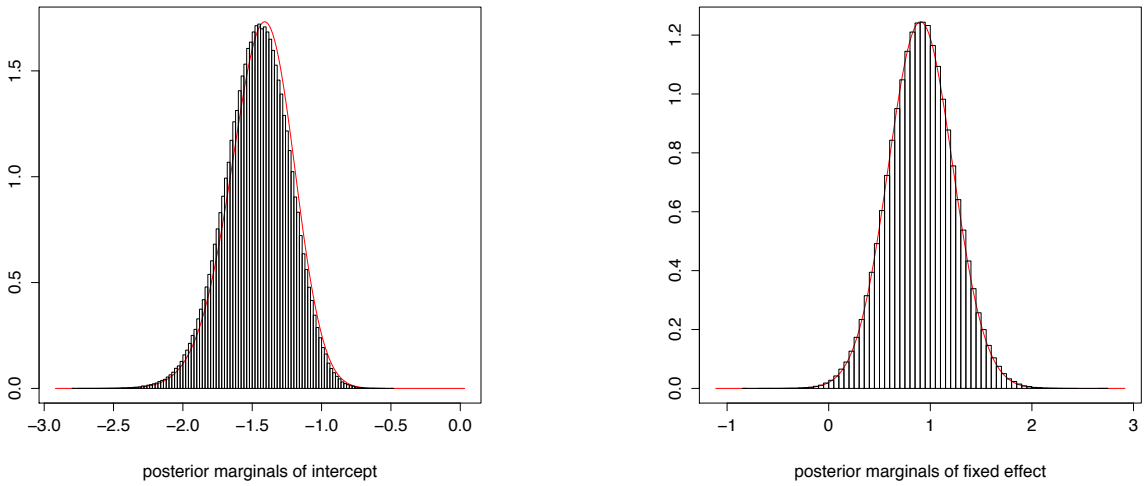


Figure 7: The graphs showing posterior marginals of intercept and fixed effect. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

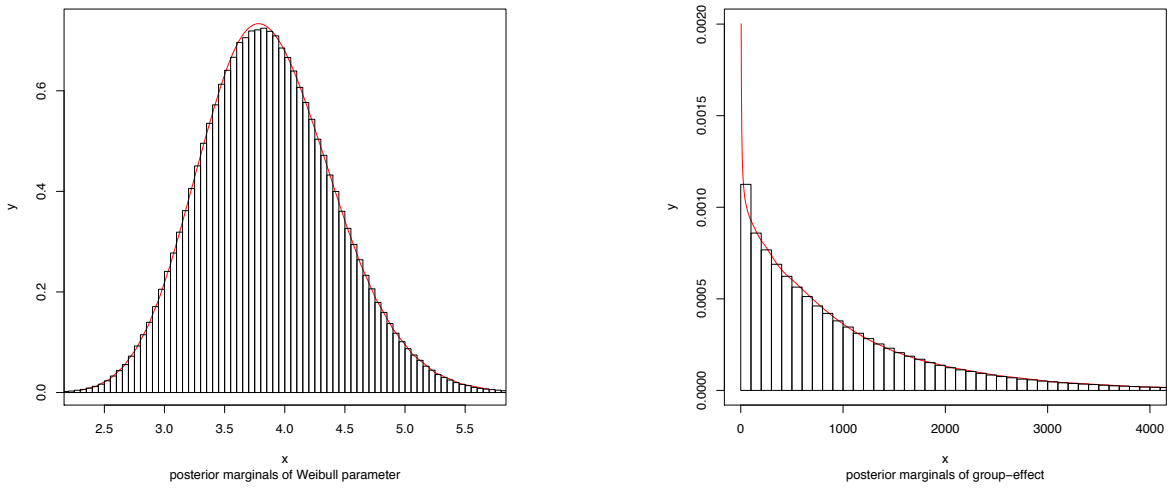


Figure 8: The graphs showing posterior marginals of weibull shape parameter and precision of the random effect. The density shown by straight line is obtained using INLA and with histogram is using WinBUGS.

7 Example5 - Spatial model

In this section we will discuss spatial models. We will follow the piecewise constant hazard model with additional covariates, including covariate with random effects corresponding to spatial effect.

7.1 The model

We will follow a model similar to the one we have discussed in earlier example - piecewise constant hazard model with some extensions.

7.2 Data

We are analysing a data set used by Henderson et al. (2002) on leukaemia survival times in Northwest England. The data set contains information on all 1043 cases of acute myeloid leukaemia in adults who have been diagnosed between 1982 and 1998 in Northwest England. Continuous covariates include the age of the patient, the white blood cell count (wbc) at diagnosis, and the Townsend deprivation index(tpi) which measures the deprivation for the enumeration district of residence.Higer values of this index indicate poorer regions while samller values correspond to wealthier regions. The sex of a patient is incuded using indicators 0 (for male) and 1 (for female). Spatial information is used on district on district level. A portion of data is given below.

time	cens	age	sex	wbc	tpi	district
1	1	61	0	13.3	-1.96	9
1	1	76	0	450.0	-3.39	7
1	1	74	0	154.0	-4.95	7
1	1	79	1	500.0	-1.40	24
1	1	83	1	160.0	-2.59	7
1	1	81	1	30.4	0.03	11
1	1	76	0	41.3	3.95	17
1	1	87	0	280.0	1.91	21

For this example, let y_{ji} be the indicator of death in j^{th} interval of i^{th} individual.

$$y_{ji} \sim \text{Pois}(\eta_{ji} t_{ji})$$

We introduce covariates through λ

$$\log(\eta_{ji}) = \alpha + \beta.haz_j + \beta.sex * sex_i + \beta.age_i + \beta.tpi_i + \beta.wbc_i + \beta.spatial_i$$

where, j = 1 to 10 , i = 1 to 1043

- t_{ji} is time survived by i th individual in j th interval.
- The fixed effect regression coefficent $\beta.sex$ is assume Gaussian prior with parameters 0 and 0.001 and the intercept term (α) is assume flat prior.
- The baseline hazard function $\beta.haz$ and the random effects regression coefficients $\beta.age$, $\beta.wbc$, $\beta.tpi$ and $\beta.spatial$ follows random walk one with precisions τ_0 , τ_1 , τ_2 , τ_3 and τ_4.
- The above model has 5 hyperparameters namely τ_0 , τ_1 , τ_2 , τ_3 and τ_4, each of them assume a Gamma prior with parameters 10 and 1.
- Since we use RW1 prior for baseline hazard and random effects. We need to discretise the continuous covariates in the data set. Covariate age is divided into 10 groups, covariates tpi and wbc are divided into 19 groups and covariate spatial effect is arranged according to district information of patients.

7.3 Model specification in WinBUGS

In WinBUGS the model code for fitting the above mentioned model is given below. For RW1 prior we need to specify weight and adjacency matrix, this could be given in the data file instead.

```
model{
      for (j in 1 :T ){
        for(i  in  1 : N){
            y[j,i] ~ dpois(eta[j,i])
            eta[j,i] <- lambda[j,i] * t[j,i]
            log(lambda[j,i]) <- alpha + beta.haz[j] + beta.sex * sex[i]+
                                beta.age[i] + beta.tpi[i] + beta.wbc[i] +
                                beta.spatial[i]
        }
      }
   # the priors are
      alpha ~ dflat()
      beta.sex ~ dnorm(0.0,0.001)
      beta.haz[1:T] ~ car.normal(adj[], weights[], num[],tau0)
      tau0 ~ dgamma(10, 1)
      beta.age[1:T] ~ car.normal(adj2[], weights2[], num2[],tau1)
      tau1  ~ dgamma(10 ,1)
      beta.wbc[1:T] ~ car.normal(adj3[], weights3[], num3[],tau2)
      tau2 ~ dgamma(10,1)
      beta.tpi[1:T] ~ car.normal(adj3[], weights3[], num3[],tau3)
      tau3 ~ dgamma(10,1)
      beta.spatial[1:T] ~ car.normal(adj[], weights[], num[],tau4)
      tau4 ~ dgamma(10,1)

  # here we are specifying weight matrix and adjacency matrix corresponding
  # to RW1 prior for covariate age, this could be given in the data file
  #   instead

       for(t in 1:1) {
weights2[t] <- 1;
adj2[t] <- t+1;
num2[t] <- 1
      }
   for(t in 2:(A-1)){
weights2[2+(t-2)*2] <- 1;
adj2[2+(t-2)*2] <- t-1;
weights2[3+(t-2)*2] <- 1;
adj2[3+(t-2)*2] <- t+1;
num2[t] <- 2
    }
   for(t in A:A) {
weights2[(A-2)*2 + 2] <- 1;
adj2[(A-2)*2 + 2] <- t-1;
num2[t] <- 1
```

```
    }
  }
```

where,

- T = Number of time intervals or number of cutpoints. In this example number of time intervals considered are 20.
- N = Number of individuals. In this example N=1043.
- y[j,i] = 1 if failure observed or 0 if censored.
- E[j,i] = time survived by ith individual in jth interval.
- adj[] = adjacency vector listing neighbouring time points.
- weights[] = weight vector.
- num[] = number of neighbours.

7.4 Implementing using the $INLA$ package for R

to read the data files

```
>data(Leuk)
```

to see the various components of data

```
>Leuk$
Leuk$age        Leuk$district  Leuk$time      Leuk$wbc       Leuk$ycoord
Leuk$cens       Leuk$sex       Leuk$tpi       Leuk$xcoord
\small}
```

to make a discretised version of the continuous covariate in the data set.

```
>age.group = inla.group(Leuk$age, n=11)
>wbc.group = inla.group(Leuk$wbc, n=20)
>tpi.group = inla.group(Leuk$tpi, n=20)
```

here each value is replaced by the median value of the interval to which it belongs. To ensure that the descretised values will be the same in INLA and winBUGS we have some different settings.

```
>Leuk$wbc1=wbc
>Leuk$tpi1=tpi
>Leuk$age1=age
>Leuk$time=time
>Leuk$cens=event
```

to sort the data set per district

```
>oo=order(Leuk$district)
>Leuk=Leuk[oo,]
```

the function to plot the regions is as follows:

```
>source("draw-map-newengland.R")
```

the formula and the model are specified in the following ways:

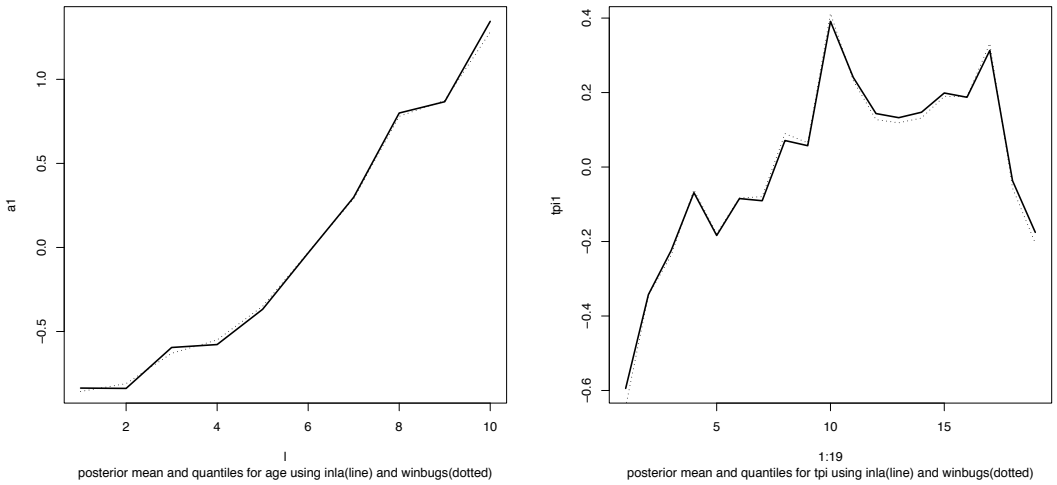


Figure 9: The graphs showing posetior means for age and town deprivation index using INLA(straight line) and WinBUGS(dotted line).

```
>inla.surv(time,event)>formula= inla.surv(time,event)~sex+
              f(age1,model="rw1", values=lab.age)
            + f(wbc1,model="rw1",values=lab.wbc)
            + f(tpi1,model="rw1",values=lab.tpi)
            + f(district,model="besag",graph.file="new.graph")
>model=inla(formula,family="piecewise.constant", data=Leuk,
          verbose=TRUE, control.hazard=list(model="rw1",n.intervals=20),
          keep=T)
```

to plot the mean of the spatial field,

```
> newengland.map(model1$summary.random$district$mean)
```

7.5 Result

Like other survival models we compared the results obtained from INLA and Winbugs for this model. The posterior means of covariates are quite comparable but as WinBUGS is taking a lot more time for execution than expected, we were unable to get reliable results for hyperparameters.
In Fig.9 and Fig.10, the graph of posterior means for age, town deprivation index(tpi) and spatial effect are given, which show that the posterior means are very close.

Since for this particular example WinBUGS did not perform well. We compared our results, which we got from INLA approximations, with those given in Kneib and Fahrmeir (2007). In the article, same data set is studied using mixed model approach for geoadditive hazard regression. They have considered link function as,

$$\eta_i(t) = \gamma_0 + \gamma_1 sex_i + g_0(t) + f_1(age_i) + f_2(wbc_i) + f_3(tpi_i) + f_{spat}(s_i)$$

where g_0 is the log-baseline,f_1,f_2,f_3 are smooth functions of the continuous covariates and f_{spat} is a spatial effect. Both g_0 and f_j were modelled as cubic P-splineswith second-order difference penalty and 20 knots.
While checking the estimated spatial effect presented in Fig.11, we found that these are similar to those are mentioned in article by Kneib and Fahrmeir (2007) on page218. In Fig.12, we presented the posterior means with together with 95% credible intervals(dashed lines) for log-baseline, the effects of age, white blood count, the town deprivation index and district. These are quite comparative with the figures given in figure2 by Kneib and Fahrmeir (2007).

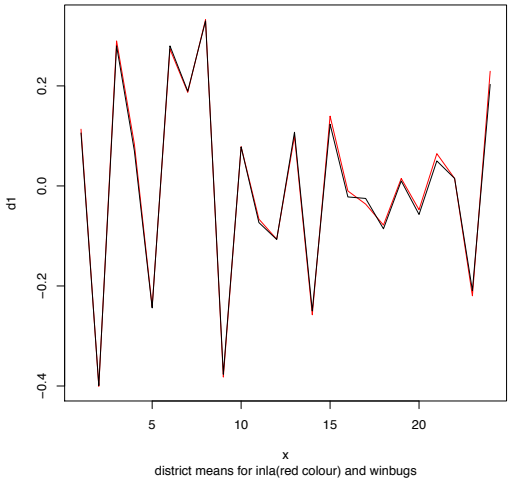


Figure 10: The graphs showing posetior means for districts using INLA(red colour) and WinBUGS(black colour).

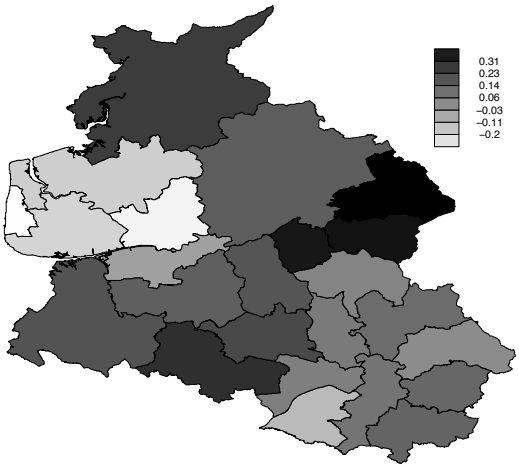


Figure 11: The graphs showing estimates of spatial effect based on a district level analysis.

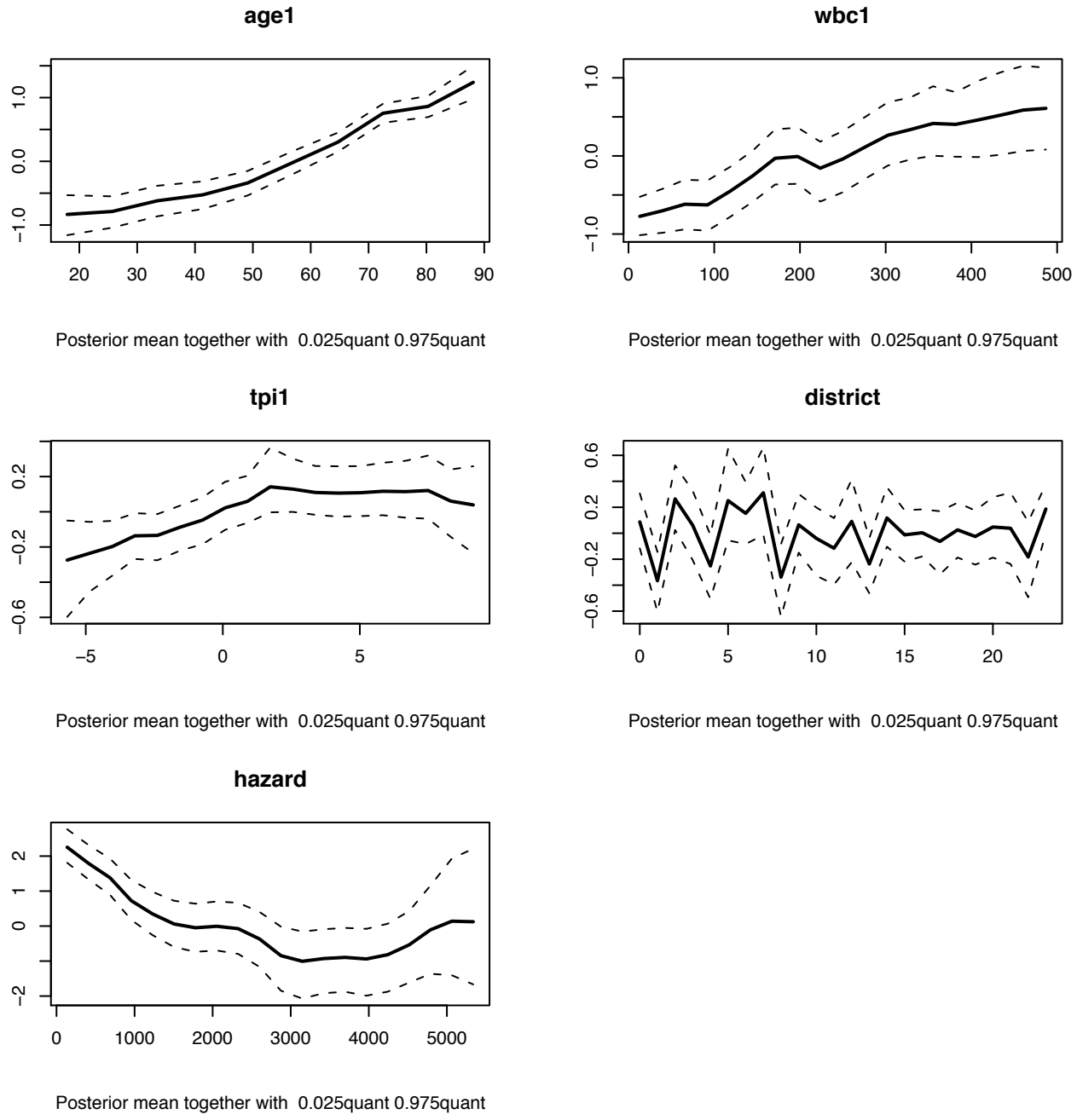


Figure 12: The graphs showing estimates of posterior means together with 95% credible intervals(dashed lines) for log-baseline, the effects of age, white blood count, the town deprivation index and district.

8 Remarks

8.1 Piecewise linear Hazard model

In this section we examine piecewise linear hazard model.
To construct this model, we first construct a finite partition of the time axis, $s_1 < s_2 < ... < s_j$, with $s_J > t_i$ for all $i = 1, 2, ..., n$ and t_i are the n survival times. Thus, we have J intervals $(0, s_1], (s_1, s_2], ..., (s_{J-1}, s_J]$. The observed survival times may be terminated either by failure or by censoring. It is assumed that conditionally the times of failure are independent of the times of censoring. Let t_{ij} denote the time lived by the i-th individual in the j-th interval. If the individual lived beyond the end of the interval then the time lived in the interval equals the width of the interval. The base line hazard is defined as

$$\begin{aligned} h_0(t) &= \lambda_j + \frac{\lambda_{j+1} - \lambda_j}{s_{j+1} - s_j}(t - s_j),\ if\ t \in I_j = (s_j, s_{j+1}] \\ &= \lambda_j\big(1 - \frac{t - s_j}{s_{j+1} - s_j}\big) + \lambda_{j+1}\big(\frac{t - s_j}{s_{j+1} - s_j}\big) \end{aligned}$$

The base line survival if $t \in I_k = (s_k, s_{k+1}]$ is

$$S_0(t) = \exp\Big\{ - \lambda_1\big(\frac{s_2 - s_1}{2}\big) + \lambda_2\big(\frac{s_3 - s_1}{2}\big) + \cdots + \lambda_k\big(\frac{t}{2} - \frac{s_{k-1}}{2} - \frac{(t - s_k)^2}{2(s_{k+1} - s_k)}\big) + \lambda_{k+1}\big(\frac{(t - s_k)^2}{2(s_{k+1} - s_k)}\big)\Big\}$$

and the hazard rate is given by

$$\begin{aligned} h(t) &= h_0(t)\exp(z^T\beta) \\ &= \exp\{z^T\beta + \log h_0(t)\} \\ &= \exp\big\{z^T\beta + \log\big[\lambda_j\big(1 - \frac{t - s_j}{s_{j+1} - s_j}\big) + \lambda_{j+1}\big(\frac{t - s_j}{s_{j+1} - s_j}\big)\big]\big\} \end{aligned}$$

where $t \in I_k$.
Here the hazard function depends on the $\log$ of the sum of two Gaussian variables and thus resulting in a non Gaussian variable. We could not use INLA for this setting. If we want to remove the dependency of the hazard rate on this extra λ_{j+1}. Then we need to create extra n new knots on the time axis at points λ_j's, meaning increasing time intervals from J to $J+n$. Which consequently increase the computing time from $o(n)$ to $o(n^2)$. Instead it is advicable to use piecewise constant baseline hazard model with more time intervals, i.e., finer partition of time axis.

prior distribution	parameters	intervals on original scale	intervals on new scale	parameters of new prior
RW1	a and b	n	$n \times k$	a and b/k
RW2	a and b	n	$n \times k$	a and b/k^3
CRW2	a and b	n	$n \times k$	a and b/k^3

Table 1: scaling of priors - when values on new scale are used

8.2 Scaling of Random walk priors

In modeling, sometimes priors are RW models. In such cases if we want to compare such models we need to be careful with the number of nodes used for such priors. If we change the number of nodes then scaling of prior is necessary in order to obtain similar variance as earlier. To understand the need of scaling, we focus our attention on one specific case. We assume that the location of the nodes at the time axis are all positive integers, such that $x_1 < x_2 < x_3 < \cdots$ and the distance between consecutive nodes is constant, i.e., $x_2 - x_1 = x_3 - x_2 = x_4 - x_3$ and so on.
For RW1 models,

$$x_{i+1} - x_i \sim N(0, {\tau_1}^{-1}), \qquad i = 1, \ldots, n-1$$

Hence, var$(x_2 - x_1) = \frac{1}{\tau_1}$. We further assume, $\tau_1 \sim \Gamma(a, b)$.
Now, we divide each interval into two equal intervals, such that $x_1^{'} < x_2^{'} < x_3^{'} < \cdots$, is shown in Fig.13.
Thus $x_2 - x_1 = 2(x_2^{'} - x_1^{'}) = x_3^{'} - x_1^{'}$.

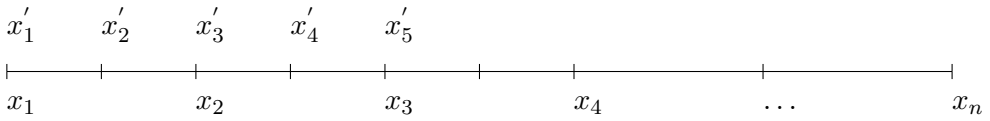


Figure 13: The graphs showing the division of time axis into intervals.

Let the var$(x_2^{'} - x_1^{'}) = \frac{1}{\tau_2}$.
It is clear that var$(x_2 - x_1) =$ var$(x_3^{'} - x_1^{'})$. Thus, $2\tau_1 = \tau_2$.
Finally, after comparing the desity functions for τ_1 and τ_2, we see that $\tau_2 \sim \Gamma(a, b/2)$.
Thus, for a random walk prior, if we change the number of time intervals then we need to scale the parameters of the new prior according to new number of intervals to get the same reuslts as earlier. In general, if we divide each interval into k equal intervals, then new variance, say τ_k will be equal to b/k. The various priors with their original parameters and their changed parameters are given in table 1.

Now we will compare the results of posterior means of original prior and the posterior means of the rescaled prior considering the case when original prior is RW2.

Example: When prior is RW2 model

Here we are assuming that the random variable y follows second order random walk (RW2). The distance between consecutive nodes is constant and is equal to 1. Initially there are n nodes but then we will increase them to $n \times k$ nodes, all equally spaced.

```
n = 100
x = 1:n
s = 1
y = sin(0.3*x) + rnorm(n,sd=s)
##prior
a = 1
b= 0.1
ind = 1:n
data = list(ind=ind, y=y)
values = seq(1,n,length=n)
formula = y ~ f(ind, values=values, model="rw2", param=c(a,b)) + 1
m = inla(formula, data=data, control.data=list(initial = log(1/s^2),
    fixed=T))
plot(values, m$summary.random$ind$mean, type="l")
points(values, y)

k =10
values = seq(1,k*n,length=k*n)
ind = values[seq(1,k*n, length=n)]
data = list(ind=ind, y=y)
formula = y ~ f(ind, values=values, model="rw2", param=c(a,b/k^3)) + 1
mm = inla(formula, data=data, control.data=list(initial = log(1/s^2),
     fixed=T), keep=T)
lines(seq(1,n,length=k*n), mm$summary.random$ind$mean, type="l",col="red")
```

The graph for the above example is given in Fig.14.

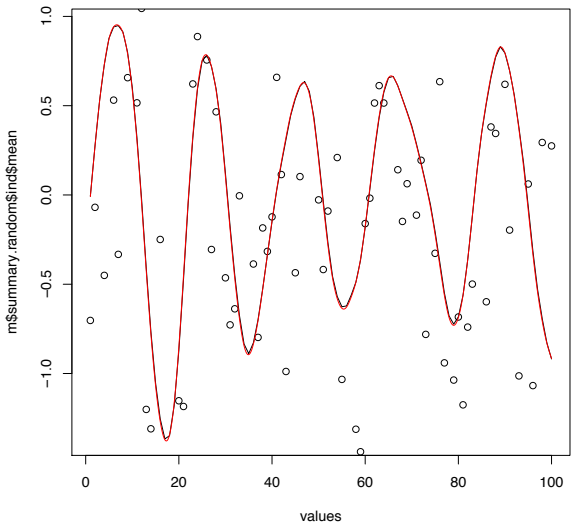


Figure 14: The posterior means of RW2 model with n intervals and with $k \times n$ intervals with proper scaling.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, volume 101 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(123–142).

Berzuini, C. and Clayton, C. (1994). Bayesian survival analysis on multiple time scales. *Statistics in Medicine*, 13:823–838.

Breslow, N. (1972). Discussion on regression models and life-tables (by d. r. cox). *Journal of the Royal Statistical Society, Series B*, 34:216–217.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30:89–99.

Brezger, A., Kneib, T., and Lang, S. (2003). *BayesX: Software for Bayesian inference*. Department of statistics, University of Munich, version 1.1 edition. http://www.stat.uni-muenchen.de/∼lang/bayesx.

Carlin, B. P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In *Bayesian Statistics, 7*, pages 45–63. Oxford Univ. Press, New York.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220.

Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika*, 81(2):317–330.

Gamerman, D. (1991). Dynamic bayesian models for survival data. *Journal of the Royal Statistical Society, Series B*, 40(1):63– 69.

Gamerman, D. (1998). Markov chain monte carlo for dynamic generalised linear models. *Biometrika*, 85(1):215–227.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*. Chapman and Hall, Boca Raton, 2nd edition.

Harrison, P. and Stevens, C. (1976). Bayesian forecasting(with discussion). *Journal of the Royal Statistical Society, Series B*, 38:205–247.

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

Holford, T. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, 36:299–305.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihood based on cox's regression and life model. *Biometrika*, 60:267–278.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 2nd edition.

Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34:207–228.

Laird, N. and Oliver, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76:231–240.

Mantel, N., Bohidar, N. R., and Ciminera, J. L. (1977). Mantel - haenszel analysis of litter -matched time -to -response data, with modifications for recovery of interlitter information. *Cancer Research*, 37:3863–3868.

Martino, S. and Rue, H. (2008). Implementing approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations: A manual for the inla-program. Technical Report no 2, Department of mathematical sciences, Norwegian University of Science and Technology.

Nahman, N. S., F.Middendorf, D., Bay, W. H., McElligott, R., Powell, S., and Anderson, J. (1992). Modification of the percutaneous approach to peritoneal dialysis catheter placement under peritoneoscopic visualization: Clinical results in 78 patients. *Journal of The American Society of Nephrology*, 3:103–107.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Rue, H. and Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192. Special Issue: Bayesian Inference for Stochastic Processes.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Spiegelhalter, D. J., Thomas, A., Best, N., and Lunn, D. (2003). *WinBugs User Manual Version 1.4*. Cambridge: MRC Biostatistics Unit.

Stablein, D. M., Carter, W. H., and Novak, J. W. (1981). Analysis of survival data with non-proportional hazard functions. *Controlled Clinical Trials*, 2:149–159.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Vaupel, J. M., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80:73–97.