Approximate Bayesian Inference for Survival Models

Rupali Akerkar[1]
Department of Mathematical Sciences
NTNU,Norway

15 May 2009

[1]With H.Rue/S.Martino

Basic idea

- Survival model
- Present survival model as a latent Gaussian model
- Apply INLA
- Verify results with MCMC results

Basic idea

- Survival model
- Present survival model as a latent Gaussian model
- Apply INLA
- Verify results with MCMC results

Basic idea

- Survival model
- Present survival model as a latent Gaussian model
- Apply INLA
- Verify results with MCMC results

Basic idea

- Survival model
- Present survival model as a latent Gaussian model
- Apply INLA
- Verify results with MCMC results

Outline

- Survival Analysis
  - Some definitions
  - Censoring
  - Likelihood
- Survival Models
  - Parametric models
  - Semiparametric models for hazard
    - piecewise constant models
    - piecewise linear models
  - Spatial model

Let T be a random survival time, the following functions are defined:

- Density function:

$$T \sim f(t)$$

- Survival function:

$$S(t) = 1 - F(t) = \int_t^{\infty} f(u)du$$

- Hazard function:

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

thus

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

Censoring

- **Uncensored observation**: The failure time is recorded
- **Right censored observation**: The censoring time $C < T$ is recorded
- **Interval censored observation**: The failure time is not observed exactly but it is known that $T_{lo} < T < T_{up}$

Each observation is described by a triple(T_{lo}, T_{up}, δ) with

$$T_{lo} = T_{up} = T, \delta = 1 \; if \; the \; obs. \; is \; uncensored$$

$$T_{lo} = T_{up} = C, \delta = 0 \; if \; the \; obs. \; is \; right \; censored$$

$$T_{lo} < T_{up}, \delta = 0 \; if \; the \; obs. \; is \; interval \; censored$$

Likelihood

The likelihood function is:

$$L = \prod L_i$$

where:

- if i is uncensored

$$L_i = h(T)S(T) = h(T)exp\{-\int_0^T h(u)du\}$$

- if i is censored

$$L_i = S(C) = exp\{-\int_0^C h(u)du\}$$

- if i is interval censored

$$\begin{aligned} L_i &= S(T_{lo}) - S(T_{up}) \\ &= exp\{-\int_0^{T_{l0}} h(u)du\}\{1 - exp(-\int_{T_{lo}}^{T_{up}} h(u)du)\} \end{aligned}$$

The contribution to the log-likelihood of data i, $(T_{lo}, T_{up}, \delta_i)$, is in general

$$l_i = \delta_i h(T_{up,i}) - \int_0^{T_{up,i}} h(u)du + log\{1 - exp(-\int_{T_{lo,i}}^{T_{up,i}} h(u)du)\}$$

The contribution to the log-likelihood of data i, $(T_{lo}, T_{up}, \delta_i)$, is in general

$$l_i = \delta_i h(T_{up,i}) - \int_0^{T_{up,i}} h(u)du + log\{1 - exp(-\int_{T_{lo,i}}^{T_{up,i}} h(u)du)\}$$

Only included for uncensored data

The contribution to the log-likelihood of data i, $(T_{lo}, T_{up}, \delta_i)$, is in general

$$l_i = \delta_i h(T_{up,i}) - \int_0^{T_{up,i}} h(u)du + log\{1 - exp(-\int_{T_{lo,i}}^{T_{up,i}} h(u)du)\}$$

Included for all data

The contribution to the log-likelihood of data i, $(T_{lo}, T_{up}, \delta_i)$, is in general

$$l_i = \delta_i h(T_{up,i}) - \int_0^{T_{up,i}} h(u)du + log\{1 - exp(-\int_{T_{lo,i}}^{T_{up,i}} h(u)du)\}$$

Only included for interval censored data

Cox Model

The model proposed by Cox in 1972 is

$$h(t|z_1, ..., z_p) = h_0(t)\ \exp(z_1\beta_1 + ... + z_p\beta_p)$$

- h_0 = baseline hazard
- z_i = covariates
- β_i = regression parameters

For this model, the covariates are assumed to have fixed effects on failure pattern.
But the covariates effects do change with time.

More comprehensive model

A more comprehensive model is achieved by assuming

$$h(t) = \exp(z_0\beta_0(t) + ... + z_1\beta_p(t))$$

where $\beta_0 = \log(h_0)$.

- Parametric models
  - Exponential
  - Weibull
- Parametric models with frailty
- Semiparametric models
  - Piecewise-constant baseline hazard
  - Piecewise-linear baseline hazard

The Weibull model

- The data : $(T_{lo,i},\ T_{up,i},\ \delta_i),\ i = 1,...,n_d$
- The hazards rate :

$$h(u; z, \alpha) = \alpha u^{\alpha-1} exp(\eta)$$

  with $\eta = z^T \beta$
- the log-likelihood :

$$l = \delta h(T_{up}) - \int_0^{T_{up}} h(u)du + log\{1 - exp(-\int_{T_{lo}}^{T_{up}} h(u)du)\}$$
$$= \delta[log\alpha + (\alpha - 1)logT_{up} + \eta] - e^{\eta} T_{up}^{\alpha} + log\{1 - e^{-e^{\eta}(T_{up}^{\alpha} - T_{lo}^{\alpha})}\}$$

- Priors for parameters :

$$\beta \sim N(0, \tau_\beta I)$$
$$\alpha \sim \pi(\alpha)$$

The Weibull model

- The data : $(T_{lo,i},\ T_{up,i},\ \delta_i),\ i = 1,\ldots,n_d$
- The hazards rate :

$$h(u; z, \alpha) = \alpha u^{\alpha-1} exp(\eta)$$

with $\eta = z^T\beta$

- the log-likelihood :

$$l = \delta h(T_{up}) - \int_0^{T_{up}} h(u)du + log\{1 - exp(-\int_{T_{lo}}^{T_{up}} h(u)du)\}$$
$$= \delta[log\alpha + (\alpha - 1)logT_{up} + \eta] - e^{\eta} T_{up}^{\alpha} + log\{1 - e^{-e^{\eta}(T_{up}^{\alpha} - T_{lo}^{\alpha})}\}$$

- Priors for parameters :

$$\beta \sim N(0, \tau_\beta I)$$

$$\alpha \sim \pi(\alpha)$$

The Weibull model

- The data : $(T_{lo,i},\ T_{up,i},\ \delta_i),\ i = 1,\ldots,n_d$
- The hazards rate :

$$h(u; z, \alpha) = \alpha u^{\alpha-1} exp(\eta)$$

with $\eta = z^T\beta$

- the log-likelihood :

$$l = \delta h(T_{up}) - \int_0^{T_{up}} h(u)du + log\{1 - exp(-\int_{T_{lo}}^{T_{up}} h(u)du)\}$$

$$= \delta[log\alpha + (\alpha - 1)logT_{up} + \eta] - e^{\eta}T_{up}^{\alpha} + log\{1 - e^{-e^{\eta}(T_{up}^{\alpha} - T_{lo}^{\alpha})}\}$$

- Priors for parameters :

$$\beta \sim N(0, \tau_\beta I)$$

$$\alpha \sim \pi(\alpha)$$

The Weibull model

- The data : $(T_{lo,i},\ T_{up,i},\ \delta_i),\ i = 1,...,n_d$
- The hazards rate :

$$h(u; z, \alpha) = \alpha u^{\alpha-1} exp(\eta)$$

with $\eta = z^T \beta$

- the log-likelihood :

$$l = \delta h(T_{up}) - \int_0^{T_{up}} h(u)du + log\{1 - exp(-\int_{T_{lo}}^{T_{up}} h(u)du)\}$$

$$= \delta[log\alpha + (\alpha - 1)log T_{up} + \eta] - e^{\eta} T_{up}^{\alpha} + log\{1 - e^{-e^{\eta}(T_{up}^{\alpha} - T_{lo}^{\alpha})}\}$$

- Priors for parameters :

$$\beta \sim N(0, \tau_\beta I)$$

$$\alpha \sim \pi(\alpha)$$

The Weibull model as a latent Gaussian model

- The latent Gaussian field

$$\mathbf{x} = \{\eta_1, ..., \eta_{n_d}, \beta\} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$$

- The hyperparameters

$$\boldsymbol{\theta} = \alpha$$

- The likelihood

$$\pi(data|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d} \pi(data_i|\eta_i, \boldsymbol{\theta})$$

We can apply INLA to this model without problems!!

Note

In more general term we can have

$$\eta = \sum_j f_j(z_j) + \sum_k \beta_j \tilde{z}_j + \epsilon$$

where $f_j(z_j)$ can represent

- smooth effect of covariate
- time varying effect of covarite
- space effect

As long as the prior for $f()$ is Gaussian we are still in the latent Gaussian model framework!

Example1 - Kidney data

(Nahman et al.,1992) The time to the first infection for kidney dialysis patients is analysed. The data are right censored. One binary covariate z is catheter placement.
The model is

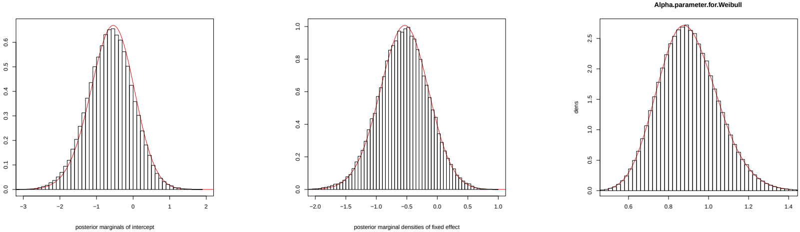
$$h(t; z) = \alpha t^{\alpha-1} exp\{\beta_0 + \beta_1 z\}$$

we assume

$$\beta_0, \beta_1 \sim N(0, 0.001)$$

$$\alpha \sim \Gamma(1, 0.001)$$

Example



Solid line from INLA (0.165sec and 0.0235 for inla.hyperpar), histogram from 3×10^6 samples from Winbugs(1900 sec)

Parametric models with frailty - Hazard function

Let t_{ij} be the survival times for the j^{th} subject in the i^{th} cluster, i=1,...,n,j=1,...,m_i, the hazard function is given as:

$$\begin{aligned} h(t_{ij}; z_{ij}, w_i, \alpha) &= \alpha t_{ij}^{\alpha-1} w_i exp(z_{ij}^T \beta) \\ &= \alpha t_{ij}^{\alpha-1} exp(\eta_{ij}) \end{aligned}$$

with

$$\eta_{ij} = z_{ij}^T \beta + log(w_i)$$

Parametric models with frailty - Log-likekihood function

The likelihood function for the generic data (T_{lo}, T_{up}, δ) is then

$$
\begin{aligned}
l &= \delta h(T_{up}) - \int_0^{T_{up}} h(u)du + log\{1 - exp(-\int_{T_{lo}}^{T_{up}} h(u)du)\} \\
&= \delta[log\alpha + (\alpha - 1)logT_{up} + \eta] - e^{\eta} T_{up}^{\alpha} + log\{1 - e^{-e^{\eta}(T_{up}^{\alpha} - T_{lo}^{\alpha})}\}
\end{aligned}
$$

Log-normal model for frailty

If we assume $log(w_i) = \epsilon_i$ to have a Gaussian prior $N(0, \tau_w)$, the parametric frailty model falls into the latent Gaussian family.

- The latent Gaussian field

$$\mathbf{x} = \{\eta_{11}, ..., \eta_{nm_n}, \epsilon_1, ...\epsilon_n, \beta\} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$$

- the hyperparameters

$$\boldsymbol{\theta} = (\alpha, \tau_w)$$

- The likelihood

$$\pi(data|\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n_d} \pi(data|\eta_i, \boldsymbol{\theta})$$

...and so no problem to apply INLA!!

Example2 - Rat data

(Mantel et al.,1977) study time till tumor development in rats from 50 distinct litters, the covarite z is a treatment(drug or placebo), w is the frailty variable (litter/cluster). The model is

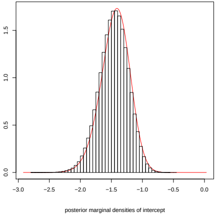
$$h(t_{ij}; z_{ij}, w_i, \alpha) = \alpha t^{\alpha-1} exp\{\beta_0 + \beta_1 z + log(w_i)\}$$

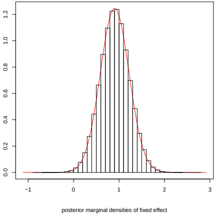
We assume

$$\beta_0, \beta_1 \sim N(0, 0.001)$$

$$\alpha, \tau_w \sim \Gamma(1, 0.001)$$

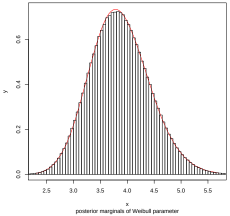
Example2 - Rat data

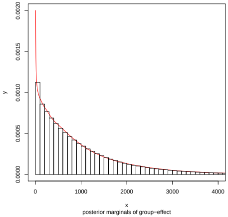




Solid line from INLA(1.422 sec) and histogram from 10^6 samples from Winbugs(1687 sec)

Example - Rat data





Solid line from INLA(1.422 sec) and histogram from 10^6 samples from Winbugs(1687 sec)

Piecewise constant model for $h_0(t)$

Divide the time line into J predefined intervals, $I_k = (S_k, S_{k+1}]$ for $k = 1, ..., J$ with $0 = s_1 < ... < s_J < \infty$, we define the baseline hazard as:

$$h_0(t) = \lambda_k \, if \; t \in I_k = (s_k, s_{k+1}]$$

and the baseline survival is then:

$$S_0(t) = exp\{-\int_0^t h_0(u)du\} = exp\{\sum_{j=1}^{k-1}(s_{j+1} - s_j)\lambda_j - (t - s_k)\lambda_k\}$$

Piecewise constant model for $h_0(t)$

In general, let the hazard rate be

$$\begin{aligned} h(t;.) &= h_0(t) exp(z^T\beta) = exp\{z^T\beta + log h_0(t)\} \\ &= exp\{z^T\beta + log\lambda_k\}; t \in I_k \end{aligned}$$

and assume a RW prior for the piecewise baseline hazard

$$log\lambda_1, ..., log\lambda_J \mid \tau_\lambda \sim RW(\tau_\lambda)$$

then

$$\eta_k = z^T\beta + log\lambda_k \mid ... \sim Gaussian$$

Log-likelihood for right censored data

The log-likelihood contribution for a (possibly) right censored observation $t \in I_k$ is :

$$log[h(t;.)^{\delta} S(t;.)] = \delta\eta_k - \{\sum_{j=1}^{k-1}(s_{j+1} - s_j)e^{\eta_j} + (t - s_k)e^{\eta_k}\}$$

$$= \delta\eta_k - (t - s_k)e^{\eta_k} - \sum_{j=1}^{k-1}(s_{j+1} - s_j)e^{\eta_j}$$

- This can be seen as the log likelihood from a Poisson with mean $(t - s_k)e^{\eta_k}$ observed to be 0 or 1 according to δ
- This can be seen as the likelihood from $k - 1$ Poisson with mean $(s_{j+1} - s_j)e^{\eta_j}$ observed to be 0

Log-likelihood for right censored data

The log-likelihood contribution for a (possibly) right censored observation $t \in I_k$ is :

$$log[h(t;.)^{\delta} S(t;.)] = \delta\eta_k - \{\sum_{j=1}^{k-1}(s_{j+1} - s_j)e^{\eta_j} + (t - s_k)e^{\eta_k}\}$$

$$= \delta\eta_k - (t - s_k)e^{\eta_k} - \sum_{j=1}^{k-1}(s_{j+1} - s_j)e^{\eta_j}$$

- This can be seen as the log likelihood from a Poisson with mean $(t - s_k)e^{\eta_k}$ observed to be 0 or 1 according to δ
- This can be seen as the likelihood from $k - 1$ Poisson with mean $(s_{j+1} - s_j)e^{\eta_j}$ observed to be 0

Log-likelihood for right censored data

The log-likelihood contribution for a (possibly) right censored observation $t \in I_k$ is :

$$
\begin{aligned}
log[h(t;.)^{\delta} S(t;.)] &= \delta\eta_k - \{\sum_{j=1}^{k-1}(s_{j+1} - s_j)e^{\eta_j} + (t - s_k)e^{\eta_k}\} \\
&= \delta\eta_k - (t - s_k)e^{\eta_k} - \sum_{j=1}^{k-1}(s_{j+1} - s_j)e^{\eta_j}
\end{aligned}
$$

- This can be seen as the log likelihood from a Poisson with mean $(t - s_k)e^{\eta_k}$ observed to be 0 or 1 according to δ
- This can be seen as the likelihood from $k - 1$ Poisson with mean $(s_{j+1} - s_j)e^{\eta_j}$ observed to be 0

Log-likelihood for right censored data

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

Piecewise constant model for $h_0(t)$ as latent Gaussian field

- The latent Gaussian field

$$\mathbf{x} = \{log\lambda_1, ..., log\lambda_J, \beta, \eta_{11}...\}$$

- The hyperparameters

$$\boldsymbol{\theta}$$

- The "augmented" Poisson data

Example3 -Times to death for a Breast-cancer trial

(Sedmak et al., 1989) The time to death of 45 breast cancer patients is analysed. The data are right censored. One binary covariate z(immunohistochemical response) is also recorded. The model is:

$$h(t; z) = h_0(t) exp\{\beta_0 + \beta_1 z\}$$

Moreover we divide the time line into 2 equal intervals

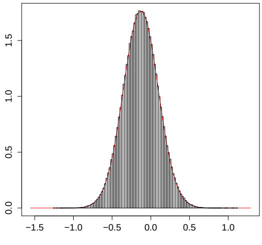
$$h_0(t) = \lambda_k \; t \in I_k$$

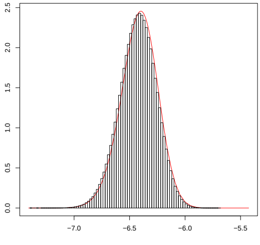
and assume

$$log\lambda_1, log\lambda_2 \sim RW1(\tau_\lambda)$$

$$\beta_0, \beta_1 \sim N(0, 0.001)$$

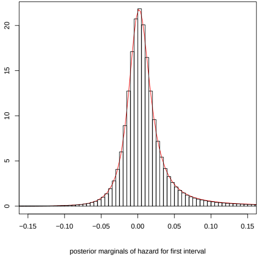
Example3 -Times to death for a Breast-cancer trial

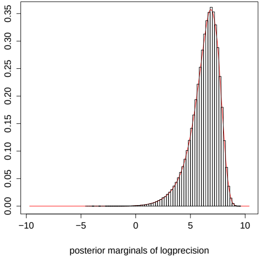




The solid line is the INLA approximations(0.090sec) and the histogram is from $5x10^6$ Winbug samples(685sec)

Example3 -Times to death for a Breast-cancer trial





The solid line is the INLA approximations(0.090sec) and the histogram is from $5x10^6$ Winbug samples(685sec)

Piecewise linear model for $h_0(t)$

Divide the time line into J predefined intervals, $I_k = (S_k, S_{k+1}]$ for $k = 1, ..., J$ with $0 = s_1 < ... < s_J < \infty$, we define the baseline hazard as:

$$h_0(t) = \lambda_j + \frac{\lambda_{j+1} - \lambda_j}{s_{j+1} - s_j} t \; if \; t \in I_j = (s_j, s_{j+1}]$$

Log-likelihood for right censored data

The log-likelihood contribution for a (possibly) right censored observation $t \in I_k$

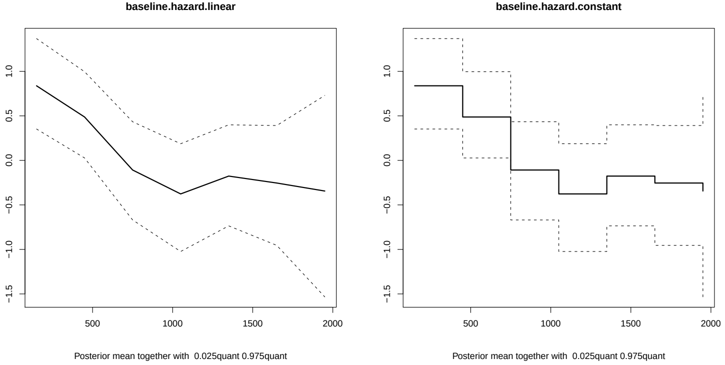
$$log[h(t;.)S(t;.)] = \delta\eta_k - w_1 e^{\eta_1} - \sum_{j=2}^{k-1} w_j e^{\eta_j} - w_k e^{\eta_k} - w_{k+1} e^{\eta_{k+1}}$$

where, w's, the weights are

$$w_1 = \frac{s_2 - s_1}{2}, \; w_j = \frac{s_{j+1} - s_{j-1}}{2}$$

$$w_k = t - \frac{s_k + s_{k-1}}{2} - \frac{(t - s_k)^2}{2(s_{k+1} - s_k)} \; and \; w_{k+1} = \frac{(t - s_k)^2}{2(s_{k+} - s_k)}$$

Example4-Times to death for a Breast-cancer trial



Piecewise linear and constant baseline hazard

Example5- Leukemia survival data

(Henderson et al.,2002) We study time to death for 1043 leukaemia patients.
The covariates included are age, wbc, tpi, sex and spatial information on district level. Here η is

$$\eta_{ij} = \alpha + \beta.haz_j + \beta.sex * sex_i + \beta.age_i + \beta.tpi_i + \beta.wbc_i + \beta.spatial_i$$

Example5- Leukemia survival data

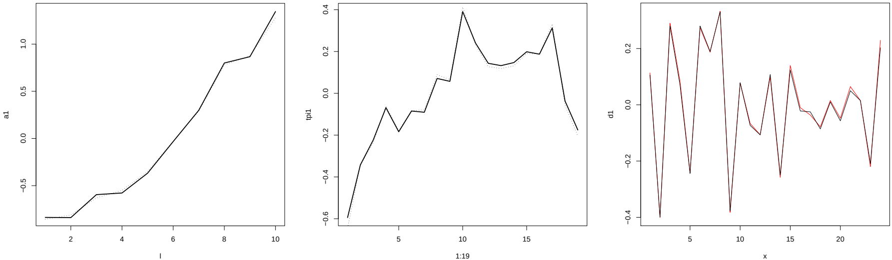
Assume

$$\alpha, \beta.sex \sim N(0, 0.001)$$

$$\beta.haz, \beta.age, \beta.tpi, \beta.wbc \sim RW1(\tau's)$$

$$\beta.spatial \sim besag$$

Example5- Leukemia survival data



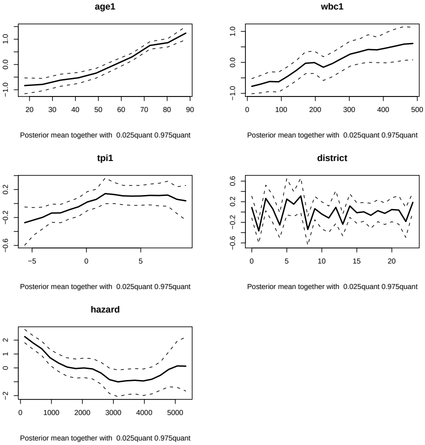
posterior mean and quantiles for age using inla(line) and winbugs(dotted)

posterior mean and quantiles for tpi using inla(line) and winbugs(dotted)

district means for inla(red colour) and winbugs

Posterior means of age,tpi and spatial effect, Solid line(red) is the INLA approximations and dotted line from Winbugs samples.

Example5-Leukemia survival data



Compared these with results by Kneib et al.,2007.

Example5-Leukemia survival data-Demo

demo(Leuk)

New inla input

data=read.table("data1.txt", header = T) this is some data
n=length(data$time)
surv.time = list(truncation=rep(0,n), event =
data$event,lower=data$time,
upper = rep(0,n), time=data$time)
d = c(as.list(data), surv.time = surv.time)
formula = surv.time ∼ *placement*
model=inla(formula,family="weibull", data= d, verbose=TRUE,
keep=TRUE)
h=inla.hyperpar(model)

Summary

- Many survival models fall in the latent Gaussian models family
- For such models INLA is a fast and reliable tool for estimate
- ... there is still lot to do to make INLA a more general tool for survival models