Approximate Bayesian Inference for Semiparametric Proportional Harzard Models

Ziang Zhang

11/09/2019

1   Survival Analysis Model:

1.1   Introduction to Survival Analysis:

In survival analysis, we are analyzing datasets in which the response variable of interest is the time until the occurrence of a particular event, such as the lifetimes of patients with a specific kind of disease, the durability of a bunch of lightbulbs etc. More specifically, we may want to study the relationship between the lifetimes of patients with the types of medicine they are using, to conclude whether a certain type of medicine does improve the overall survival times of patients. To put it in a more general setting, the response variable T should be a non-negative random variable, and in most cases, it should be a continuous random variable, defined over the interval $[0, \infty)$. Let the probability density function of T be denoted as $f(t)$ and its cumulative distribution function be $F(t)$, then the survivor function $S(t)$ of T can be defined as:

$$S(t) = P(T > t) = \int_t^\infty f(x)dx \tag{1}$$

Notice that $S(t)$ is the probability of an observation to survive to time t, and therefore it should be a monotone decreasing function with $S(0) = 1$ and $S(\infty) = \lim_{t \to \infty} S(x) = 0$. In survival analysis study, we are mostly interested in the instantaneous rate of occurence at a specific time t given that the event does not occur before t. That instantaneous rate will be measured using a hazard function $h(t)$ which is defined as:

$$h(t) = \lim_{s \to 0} \frac{P(t \le T \le t + s | T \ge t)}{s} = \frac{f(t)}{S(t)} = -\frac{\partial}{\partial t} \log[S(t)] \tag{2}$$

And the realted cumulative hazard function $H(t)$ will be defined by:

$$H(t) = \int_0^t h(u)du = -\log[S(t)] \tag{3}$$

Often time, there will be *some* survival times in the dataset that we cannot observe their exact values due to censoring or truncations, which causes a great difficulty for us to carry out our analysis. I will present the details of these problems at the section below.

1.2   Types of Censoring and Truncations:

In survival analysis, we are mainly dealing with the problem of right-censoring, interval-censoring and left truncation. Right-censoring is when an indiviual's lifetime T_i is not exactly known because the individual is still alive when the study terminates at C_i, so we are only sure about that $T_i > C_i$ but not sure what exactly T_i is. Interval-censoring on the other hand, rises when the survival time is only known to be in an interval (L_i, U_i), and the left truncation problem happens when some survival times are not recorded unless they are bigger than a specified start time t^{tr}, so all the data with survival times less than t^{tr} are missed.

In general, we use the term "censoring" to refer to the scenarios where some lifetimes are only known to exceed their cutting times C_i, but we do not know how long do they last exactly. On the other hand, the

term "truncation" mostly refer to a data collected problem where only lifetimes greater than the start time t^{tr} are collected and observed. So these terms should be used in different situations depending on what kind of survival data are we dealing with.

However, Cox's semiparametric proportional hazard model cannot be handled using INLA-like algorithm when the problem of interval-censoring is present, since the trick that INLA uses to transform log-likelihood will no longer work in this case. Therefore, we will focus on the right-censoring problem right now, and come back to solve the cases of interval-censoring and left trucation later. Next, the three types of common right-censoring will be present.

Type-I right-censoring occurs when each indiviual's censoring time C_i is fixed and known beforehand. That means when we collect a bunch of survival times, we know whether each survival time is right-censored and when is it censored exactly. In this case, we will be able to write our original dataset $\{T_i, C_i : i = 1, ..., n\}$ as $\{t_i, \delta_i : i = 1, ..., n\}$ where:

$$t_i = \min\{T_i, C_i\}, \qquad \delta_i = I(T_i \geq C_i) \tag{4}$$

This is the most common type of right-censoring, and we will focus on this type of censoring for the rest of the passage.

Type-II right-censoring occurs when we only observed the r smallest survival times in our sample. So the survial times that we can observed will be like $t_{(1)} < t_{(2)} < ... < t_{(r)}$, and the other survival times will be censored so we don't know the exact numbers. In this scenario, we have a censoring time $t_{(r)}$ that is itself random.

Lastly, independent random censoring happens when both the ith survival time T_i and the ith censoring time C_i are random variable that are independent.

1.3   Cox Proportional Hazard Model:

In most survival analysis study, we are interested in incorporating some covariates $\tilde{X} = \{X_1, X_2, ..., X_p\}$ into the distribution of survival time T, and studying their effects on the survival time T. Therefore, we often need to use different kinds of models to specify the dependence of T on $\tilde{X}$, and among those models, the proportional hazard model introduced by Cox(1972) is the most popular choice.

Let $h(t|\tilde{x})$ denote the hazard function of T at time t for a subject with covariates $\tilde{x} = (x_1, x_2, ..., x_p)$. The Cox Proportional Hazard Model can be specified as follows:

$$h(t|\tilde{x}) = h_0(t)\exp(\beta_1 x_1 + ... + \beta_p x_p) \tag{5}$$

where $h_0(t)$ is an arbitrary basedline hazard function that does only depend on time, and β_i's are the unknown parameters that we are interested in estimating. The reason that it is called a "proportional" hazard model is because for any two subjects, the ratio of their hazard function will be constant over time. This is a very strict assumption which should be checked before adopting this model.

Notice that the baseline hazard function is left to be arbitrary, which implies that the Cox Proportional Hazard Model will be a semiparametric model. There are different ways to define the baseline hazard functions, and the piecewise constant baseline hazard model will be the most convenient and popular choice. We will focus on this kind of model in the rest of this paper, and I will introduce it in details in the next section.

1.4   Proportional Hazard Model with Piecewise Constant Baseline Hazard:

Firstly, we break the time axis into K intervals with endpoints $0 = s_0 < s_1 < ... < s_K < \max\{t_i : i = 1, ..., n\}$, and assumes that the basedline hazard function is constant in each interval. i.e: $h_0(t) = \lambda_k$ for

$t \in (s_{k-1}, s_k), \ k = 1, 2, ..., K$ Let $\eta_{ik} = \log(\lambda_k) + \beta_1 x_{i1} + ...\beta_p x_{ip}$, the model that we will be focusing on will be the semiparametric proportional hazard model, specified at below:

$$\begin{aligned} h(t_i) &= h_0(t_i)\exp(\beta_1 x_{i1} + ...\beta_p x_{ip}) \\ &= \exp[\log(\lambda_k) + \beta_1 x_{i1} + ...\beta_p x_{ip}] \qquad t_i \in (s_{k-1}, s_k] \\ &= \exp(\eta_{ik}) \end{aligned} \tag{6}$$

Using this information, we can easily derive the likelihood for that single observation to be:

$$\begin{aligned} L &= f(t_i)^{\delta_i} S(t_i)^{(1-\delta_i)} \\ &= h(t_i)^{\delta_i} S(t_i) \\ &= \exp(\delta_i \eta_{ik}) \left\{ \exp\left[-\int_0^{t_i} h(u)du \right] \right\} \\ &= \exp(\delta_i \eta_{ik}) \left\{ \exp\left[-\sum_{j=1}^{k-1}(s_{j+1} - s_j)\exp(\eta_{ij}) - (t_i - s_k)\exp(\eta_{ik}) \right] \right\} \end{aligned} \tag{7}$$

Therefore, the full-likelihood of the dataset will be:

$$\begin{aligned} L &= \prod_{i=1}^{n} \exp(\delta_i \eta_{ik_{(i)}})\exp\left\{ -\sum_{j=1}^{k_{(i)}-1}(s_{j+1} - s_j)\exp(\eta_{ij}) - (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) \right\} \\ &= \prod_{i=1}^{n} \exp\left\{ \delta_i \eta_{ik_{(i)}} - \sum_{j=1}^{k_{(i)}-1}(s_{j+1} - s_j)\exp(\eta_{ij}) - (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) \right\} \end{aligned} \tag{8}$$

I emphasize the subscript for $k_{(i)}$ because each survival time will correspond to a different value of k, depending on which interval the survival time lies in.

By taking the logarithm, the log-likelihood function for the ith observation $t_i \in (s_{k-1}, s_k]$ can be written as :

$$\begin{aligned} l &= \log[f(t_i)^{\delta_i} S(t_i)^{(1-\delta_i)}] \\ &= \log[h(t_i)^{\delta_i} S(t_i)] \\ &= \delta_i \eta_{ik} - (t_i - s_k)\exp(\eta_{ik}) - \sum_{j=1}^{k-1}[(s_{j+1} - s_j)\exp(\eta_{ij})] \end{aligned} \tag{9}$$

Similarly, the full log-likelihood can be derived as:

$$l = \sum_{i=1}^{n} \left\{ \delta_i \eta_{ik_{(i)}} - (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) - \sum_{j=1}^{k_{(i)}-1}[(s_{j+1} - s_j)\exp(\eta_{ij})] \right\} \tag{10}$$

It can see from the above expression that by considering a piece-wise constant baseline hazard, we make the corresponding log-likelihood much easier to work with, since the integral $\int_0^{t_i} h(u)du$ can be replaced by a sum.

2 Inference Methodology:

2.1 Data Augmentation Using Poisson Likelihood:

Here the INLA algorithm cannot directly be applied, because if we look at the log-likelihood of a single survival time $\{t_i, \delta_i\}$, we can find that it depends on more than one η. To use INLA, we required a conditional

independent latent field together with a sparse Hessian matrix for the log-likelihood. That means we need to make sure that for a single data point, the log-likelihood should be free of terms from latent field once we condition on one of the term from the latent field.

To solve this puzzle, we will utilize a data "augmentation" trick to transform the log-likelihood of a single data point into the form that INLA likes. Notice that if we are looking at a random variable X_i that follows a poisson distribution with mean $(t_i - s_k)\exp(\eta_{ik})$, then the log-likelihood corresponding to a single data point $\{X_i = 0\}$ will be:

$$
\begin{aligned}
l &= \log\left\{ P\big[X_i = 0 | \lambda = (t_i - s_k)\exp(\eta_{ik})\big] \right\} \\
&= 0 \times \ln[(t_i - s_k)\exp(\eta_{ik})] - (t_i - s_k)\exp(\eta_{ik}) - \ln(0!) \\
&= -(t_i - s_k)\exp(\eta_{ik})
\end{aligned}
\tag{11}
$$

Similarly, when $X_i = 1$, the log-likelihood of this single data point is:

$$
\begin{aligned}
l &= \log\left(P(X_i = 1 | \lambda = (t_i - s_k)\exp(\eta_{ik})) \right) \\
&= 1 \times \ln((t_i - s_k)\exp(\eta_{ik})) - (t_i - s_k)\exp(\eta_{ik}) - \ln(1!) \\
&= \ln(t_i - s_k) + \eta_{ik} - (t_i - s_k)\exp(\eta_{ik}) \\
&\propto \eta_{ik} - (t_i - s_k)\exp(\eta_{ik})
\end{aligned}
\tag{12}
$$

Here we can basically ignore the term $\ln(t_i - s_k)$ as it does not depend on any term from the latent field. So when we later take derivative, this term will just disappear which means it won't affect our C matrix.

We showed that the first two terms of the log-likelihood of a single data point $\{t_i, \delta_i\}$ can be viewd as the log-likelihood of a single data point $X_i \sim \text{Poisson}\big[\lambda = (t_i - s_k)\exp(\eta_{ik})\big]$ being 0 when $\delta_i = 0$ and being 1 when $\delta_i = 1$.

Next step will be to figure out a similar way to deal with the last term in equation (3). Notice that for a Poisson random variable Y_j with mean $(s_{j+1} - s_j)\exp(\eta_{ij})$, the log-likelihood for observing it being 0 will be:

$$
\begin{aligned}
l &= \log\left\{ P\big[Y_j = 0 | \lambda = (s_{j+1} - s_j)\exp(\eta_{ij})\big] \right\} \\
&= -(s_{j+1} - s_j)\exp(\eta_{ij})
\end{aligned}
\tag{13}
$$

Similarly, if we gather a sample of $\{Y_{i_1} = 0, Y_{i_2} = 0, ..., Y_{i_k} = 0\}$ where each $Y_{i_j} \sim \text{Poisson}\big[\lambda = (s_{j+1} - s_j)\exp(\eta_{ij})\big]$ is independent of others, then the log-likelihood of this sample will simply be the sum of log-likelihood of each term due to independence, which sums to be $\sum_{j=1}^{k-1}(s_{j+1} - s_j)\exp(\eta_{ij})$, that is exactly what we want.

Putting these two pieces information together, which means if we have a sample being $\{X_i = \delta_i, Y_{i_1} = 0, Y_{i_2} = 0, ..., Y_{i_k} = 0\}$, and all the terms in this sample being mutually independent, then the log-likelihood of this sample will just be the log-likelihood of the single data point $\{t_i, \delta_i\}$. Doing this for all the data points $\{t_i, \delta_i | i = 1, ..., n\}$. We retrieve the original log-likelihood from the log-likelihood of a sample of $\sum_{i=1}^{n} k_{(i)}$ number of independent, but non-identical Poisson random variables. In other words, we augment our original dataset $\{t_i, \delta_i | i = 1, ..., n\}$ into a huge dataset $\{x_i, y_{i_1}, y_{i_2}, ..., y_{i_{k_{(i)}}} | i = 1, 2, ...n\}$, where all the terms in this new dataset are mutually independent. This is the cure for our problem since the log-likelihood of each term from this new "augmented" dataset, will only depend on the latent field through one η.

2.2 Methodology for Approximation Using INLA:

Here I will present how the Bayesian approximation can be carried out using an INLA-type of algorithm. Firstly, to make the covariance matrix of the joint gaussian latent field non-singular, and to simplify the

Hessian matrix that we are going to derive later, we will assume that for each η_{ij}, a normal random noise ϵ_{ij} is added. We assume that $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ being mutually independent across different i and j. In other words, we will write $\eta_{ij} = \log(\lambda_k) + \beta_1 x_{i1} + ...\beta_p x_{ip} + \Gamma_i + \eta_{ij}$, where Γ_i is any random effect that we believe exists in the context of the study.

Then, the latent field can be denoted as:

$$\tilde{W} = \left[\eta_{11}, \eta_{12}, ..., \eta_{1k}, \eta_{2k}, ..., \eta_{nk}, \Gamma_1, ..., \Gamma_q, \beta_1, ..., \beta_p, \log(\lambda_1), ..., \log(\lambda_k) \right]^T \tag{14}$$

Besides assume that $\tilde{W}$ is a GMRF, we also assume that $\log(\lambda_{k+1}) - \log(\lambda_k)$ follows $N(0, \tau^{-1})$, a RW1 model. So we will just use $\tilde{\theta}$ to denote the hyperparameter vector that determines the precision matrix of our latent field.

Now, let's derive the negated Hessian matrix of the log-likelihood with respect to the latent field. To do that, let's first consider the log-likelihood consider only one survival time $\{t_i, \delta_i\}$ where $t_i \in (s_{k_{(i)}-1}, s_{k_{(i)}}]$. In this case, the log-likelihood for this data point will be:

$$l = \delta_i \eta_{ik_{(i)}} - (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) - \sum_{j=1}^{k_{(i)}-1} \left[(s_{j+1} - s_j)\exp(\eta_{ij}) \right] \tag{15}$$

The derivative with respect to $\eta_{ik_{(i)}}$ will be

$$\frac{\partial l}{\partial \eta_{ik_{(i)}}} = \delta_i - (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) \tag{16}$$

That means the negated second derivative will be:

$$-\frac{\partial^2 l}{\partial \eta_{ik_{(i)}}^2} = (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) \tag{17}$$

For first and negated second derivatives with η_{ij} where $j < k_{(i)}$, we have:

$$\frac{\partial l}{\partial \eta_{ij}} = -(s_{j+1} - s_j)\exp(\eta_{ij})$$
$$-\frac{\partial^2 l}{\partial \eta_{ij}^2} = (s_{j+1} - s_j)\exp(\eta_{ij}) \tag{18}$$

Apparenly, for η_{ij} where $j > k_{(i)}$, we have the second derivatives of log-likelihood being 0's. Combine them together, we know that the negated Hessian matrix for the log-likelihood of $\{t_i, \delta_i\}$, H_i will be:

$$\begin{bmatrix} (s_2 - s_1)\exp(\eta_{i1}) & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & (s_3 - s_2)\exp(\eta_{i2}) & 0 & \ddots & & & \\ 0 & 0 & \ddots & & & \ddots & \\ \vdots & \cdots & \cdots & (s_{k_{(i)}} - s_{k_{(i)}-1})\exp(\eta_{i(k_{(i)}-1)}) & \ddots & & \vdots \\ \vdots & & & 0 & (t_i - s_{k_{(i)}})\exp(\eta_{ik_{(i)}}) & \cdots & \vdots \\ \vdots & & & & \ddots & & \ddots & \vdots \\ \vdots & & & & & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & \cdots & & \cdots & \cdots & 0 \end{bmatrix} \tag{19}$$

This is a very sparse matrix with only diagonal terms.

Repeating this procedure for the rest data points, using the property of independence, we can get the negated Hessian matrix H for the full log-likelihood will be:

$$
H = \begin{pmatrix}
H_1 & 0 & 0 & \cdots & & & \\
0 & H_2 & 0 & \cdots & & & \\
& \cdots & & \ddots & & & \\
& & & & H_n & \cdots & \vdots \\
& & & \ddots & & & \vdots \\
& & & & & & 0
\end{pmatrix}
\tag{20}
$$

Here we bulid a block diagonal matrix H using each block H_i obtained from above procedures. The negated Hessian matrix is very sparse, which is exactly what we want it to be. Then, we will try to derive the precision matrix Q of the latent field (To be continued).

3 Proposed Methodology for Approximation:

In the paper "Approximate Bayesian Inference for Case-Crossover Models", the author suggested a new type of algorithm to do the approximation while allowing the log-likelihood of each observation to be dependent on more than one element from the latent field, which means the ad-hoc method using "data augmentation" is no longer needed (Stringer,2019). Here we will demonstrate how that algorithm can be used to estimate the parameters in Cox Proportional Hazard Model. For simplicity, let's assume that our main interest is the β_i's in the model but not the baseline hazard $h_0(t)$.

3.1 Approximation using Partial Likelihood:

Assume that $\{t_i : i = 1, ..., k\}$ is a set of k distinct lifetimes that we actually *observed*, such that $t_{(1)} < t_{(2)} < ... < t_{(k)}$, and the result n-k lifetimes are the censored lifetimes that are not observed. Let $R_i = R(t_{(i)})$ be the set of individuals who are alived and uncensored prior to time $t_{(i)}$ (including the i-th individual who dies at $t_{(i)}$).

Define the hazard function for the i-th individual to be $h_0(t)\exp(\eta_i)$, and let $\Delta_{ij} = \log \lambda_i - \log\lambda_j$, then the partial likelihood for Cox Proportional Hazard Model can be written as:

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{k} \left\{ \frac{exp[\eta_{(i)}]}{\sum_{l \in R_i} exp[\eta_{(l)}]} \right\} \\
&= \prod_{i=1}^{k} \left\{ \frac{1}{\sum_{l \in R_i} exp[\eta_{(l)} - \eta_{(i)}]} \right\} \\
&= \prod_{i=1}^{k} \left\{ \frac{1}{\sum_{l \in R_i} exp[-\Delta_{il}]} \right\} \\
&= \prod_{i=1}^{k} \left\{ \frac{1}{1 + \sum_{l \in R_i, l \neq i} exp[-\Delta_{il}]} \right\}
\end{aligned}
\tag{21}
$$

Notice that this partial likelihood does not include any information on the baseline hazard function $h_0(t)$, meaning that all of the information are used to estimate the regression parameters in the model, which hopefully should result in a more precise estimation for them. Here it is obvious that the partial likelihood only depend on those "differenced linear predictors" Δ_{ij}, so our latent field in this case will be $\{\Delta, \beta, \Gamma\}$.

In next section, I will derive the corresponding Hessian matrix and Precision matrix under this setup.

6

3.2 Derivation of Hessian matrix and Precision matrix:

4 Example from Diabetics Dataset

Firstly, I will use the dataset "diabetics" to demonstrate the equivalence between "coxph" approach in survival package and inla approach. This dataset contains the results from a trail of laser coagulation for the treatment of diabetic retinopathy from 197 patients. Each patient had one eye randomized to laser treatment and the other eye received no treatment.

The variable "id" specifies the subject's ID.

The variable "laser" is a categorical variable with levels xenon or argon.

The varaibale "age" is the age of the subject at diagnosis.

The variable "eye" is a categorical variable with levels left or right.

The variable "trt" is a categorical variable with levels 0 for no treatment and 1 for treatment using laser.

The variable "risk" classifies the risk levels of the patients.

The response variable in this dataset will be "time", which are the actual time to blindness in months, minus the minimum possible time to event (6.5 months), and "status" indicates whether the time is censored with 1 for visual loss and 0 for censored. The censoring can be due to death, dropout, or end of the study.

Let use briefly view the structure of "diabetics":

4.1 Dataset:

```
head(as_tibble(diabetic))
```

```
## # A tibble: 6 x 8
##       id laser   age eye      trt  risk  time status
##    <int> <fct> <int> <fct>  <int> <int> <dbl>  <int>
## 1      5 argon    28 left       0     9  46.2      0
## 2      5 argon    28 right      1     9  46.2      0
## 3     14 xenon    12 left       1     8  42.5      0
## 4     14 xenon    12 right      0     6  31.3      1
## 5     16 xenon     9 left       1    11  42.3      0
## 6     16 xenon     9 right      0    11  42.3      0
```

We can see that those survival times are right-censored. We will fit a cox proportional hazard model with piece-wise constant baseline hazard, assuming that each individual will have the same baseline hazard function. The variable ID will be treated as a random effect(which can be added using code "frailty.gaussian(id)"). The variables "age", "eye", "trt" and "laser" will be included as fixed effects.

4.2 Survival: coxph

```
diabetic.CoxPh <- coxph(Surv(time, status)~age + eye + trt + laser + frailty.gaussian(id), data = diabet
summary(diabetic.CoxPh)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + eye + trt + laser +
##     frailty.gaussian(id), data = diabetic)
##
```

```
##    n= 394, number of events= 155
##
##                       coef      se(coef) se2       Chisq  DF      p
## age                   0.009548 0.01323   0.009879   0.52  1.00 4.7e-01
## eyeright              0.483005 0.17501   0.168693   7.62  1.00 5.8e-03
## trt                  -1.007507 0.17930   0.174315  31.57  1.00 1.9e-08
## laserargon           -0.182388 0.39471   0.293566   0.21  1.00 6.4e-01
## frailty.gaussian(id)                              131.35 79.63 2.4e-04
##
##            exp(coef) exp(-coef) lower .95 upper .95
## age           1.0096     0.9905    0.9837    1.0361
## eyeright      1.6209     0.6169    1.1503    2.2842
## trt           0.3651     2.7388    0.2569    0.5189
## laserargon    0.8333     1.2001    0.3844    1.8062
##
## Iterations: 6 outer, 24 Newton-Raphson
##      Variance of random effect= 0.9447455
## Degrees of freedom for terms=  0.6  0.9  0.9  0.6 79.6
## Concordance= 0.867  (se = 0.867 )
## Likelihood ratio test= 228  on 82.62 df,    p=2e-15
```

From the output above, it can be seen that variables "age" and "eyeright" have positive association with the rate of occurence of visual loss, and variables "trt" and whether using "argon" type of laser are negatively associated with the rate. All the fixed effects that we included in this study have significant effects for the risk of visual loss.

4.3   Bayesian: INLA

Now we fit the same model using INLA:

```
formula = inla.surv(time, status) ~ age + eye + trt + laser + f(id, model = "iid")
diabetic.INLA <- inla(formula, control.compute = list(dic = TRUE), family = "coxph",
                    data = diabetic, control.hazard=list(model="rw2", n.intervals=20))
diabetic.INLA$summary.fixed
```

```
##                      mean          sd   0.025quant      0.5quant    0.975quant
## (Intercept) -4.763690680 0.24992878 -5.27236824 -4.757670697 -4.28883926
## age          0.007159365 0.01085487 -0.01385225  0.007023859  0.02907121
## eyeright     0.385724694 0.17636318  0.04933510  0.382188223  0.74274047
## trt         -0.867197367 0.19238190 -1.26409308 -0.860338029 -0.50840693
## laserargon  -0.118806249 0.32291823 -0.77650975 -0.111808889  0.49856487
##                    mode           kld
## (Intercept) -4.746048087 3.767951e-06
## age          0.006834771 8.315113e-06
## eyeright     0.375234593 6.150518e-06
## trt         -0.846041739 1.474631e-05
## laserargon  -0.099033910 3.891890e-06
```

It seems like these two approaches are similar enough. Though in the classic frequentist approach, the effects of age and using argon-type laser are significant, but inla gives insignificant results(the 95% credible interval contains zero). There is an estimate for intercept in the INLA's method because we used a random walk prior in that.

5 Example from Bladder Dataset:

Next, we will study the two approaches on the dataset "bladder1". This is the full dataset that contains the result from a study on recurrences of bladder cancer from 118 subjects. In this dataset, the variables that we are interested in are "id", "number", "size", "recur", "times" and "censored".

The variable "id" is just the patient ID.

The variable "number" specifies initial number of tumours of each subject.

The variable "size" is the size of largest initial tumour.

The variable "recur" is the number of recurrence of bladder cancer for that subject.

The response variable will be "time" which is computed to be the duration of times until recurrence or death, censored by the variable "censored" with 0 means being censored.

5.1 Dataset:

```
data <- as_tibble(bladder1)
data <- select(data,-c(rsize,rtumor,enum))
data <- data %>% mutate(censored = status==0)
for (i in 1:length(data$censored)) {
  if(data$censored[i]) data$censored[i] <- 0
  else data$censored[i] <- 1
}
data <- data %>% mutate(times = stop-start)
head(data)
```

```
## # A tibble: 6 x 10
##      id treatment number  size recur start  stop status censored times
##   <int> <fct>      <int> <int> <int> <int> <int>  <dbl>    <dbl> <int>
## 1     1 placebo        1     1     0     0     0      3        1     0
## 2     2 placebo        1     3     0     0     1      3        1     1
## 3     3 placebo        2     1     0     0     4      0        0     4
## 4     4 placebo        1     1     0     0     7      0        0     7
## 5     5 placebo        5     1     0     0    10      3        1    10
## 6     6 placebo        4     1     1     0     6      1        1     6
```

```
tail(data)
```

```
## # A tibble: 6 x 10
##      id treatment number  size recur start  stop status censored times
##   <int> <fct>      <int> <int> <int> <int> <int>  <dbl>    <dbl> <int>
## 1   115 thiotepa       4     1     3    24    47      1        1    23
## 2   115 thiotepa       4     1     3    47    50      0        0     3
## 3   116 thiotepa       3     4     0     0    54      0        0    54
## 4   117 thiotepa       2     1     1     0    38      1        1    38
## 5   117 thiotepa       2     1     1    38    54      0        0    16
## 6   118 thiotepa       1     3     0     0    59      3        1    59
```

Here the variable "id" specifies different individuals, and should be treated as a random effect. The variable "time" is computed using the difference between variables "start" and "stop", which denote the start time and end time of each time interval. It seemes like a interval censoring problem but the start time is known before hand, so we can treat it as a regular type-I right censoring.

In this study, a nonzero value of "status" can be death from bladder disease, death from other reason or recurrence. Here we will just view all of these situations as "occurence" for simplicity. So the variable "censored" is created such that it is 1 if "status" is non-zero, otherwise 0. We will include "number", "size" and "recur" as fixed effects in this study.

5.2 Survival:coxph

```r
bladder.CoxPh <- coxph(Surv(times, censored)~ number + size + recur + frailty.gaussian(id), data = data
summary(bladder.CoxPh)
```

```
## Call:
## coxph(formula = Surv(times, censored) ~ number + size + recur +
##     frailty.gaussian(id), data = data)
##
##   n= 294, number of events= 218
##
##                      coef    se(coef) se2     Chisq DF   p
## number             0.064848 0.04246  0.04051  2.33 1.00 1.3e-01
## size               0.008336 0.04797  0.04631  0.03 1.00 8.6e-01
## recur              0.229848 0.02559  0.02431 80.66 1.00 2.7e-19
## frailty.gaussian(id)                          5.12 4.59 3.5e-01
##
##        exp(coef) exp(-coef) lower .95 upper .95
## number    1.067    0.9372     0.9818     1.160
## size      1.008    0.9917     0.9179     1.108
## recur     1.258    0.7947     1.1968     1.323
##
## Iterations: 8 outer, 43 Newton-Raphson
##      Variance of random effect= 0.0257231
## Degrees of freedom for terms= 0.9 0.9 0.9 4.6
## Concordance= 0.694   (se = 0.694 )
## Likelihood ratio test= 105.4   on 7.33 df,    p=<2e-16
```

Fitting this model using the traditional partial likelihood approach gives insignificant results for all the fixed effects except "recur", which has a strong positive effect. But we will still proceed to check what will happen if we fit it using a Bayesian approach.

5.3 Bayesian: INLA

```r
formula = inla.surv(times, censored) ~  -1 + number + size + recur + f(id, model = "iid")
bladder.INLA <- inla(formula, control.compute = list(dic = TRUE), family = "coxph",
                     data = data, control.hazard=list(model="rw1", n.intervals=20))
bladder.INLA$summary.fixed
```

```
##               mean         sd 0.025quant    0.5quant 0.975quant        mode
## number -0.4033400 0.11182124 -0.6273909 -0.4019900 -0.1869216 -0.3992871
## size   -0.7277314 0.11391556 -0.9608581 -0.7245318 -0.5126347 -0.7182870
## recur   0.3002257 0.09091173  0.1254837  0.2987694  0.4835155  0.2959388
##               kld
## number 3.919691e-05
## size   6.816036e-05
## recur  7.664976e-06
```

Indeed, the two results seem pretty similar in general. In both cases, we can see that there are no apparent relationships between all of the fixed effects and the rate of occurence of bladder cancer's recurrence, or death, except the variable "recur" with a postive effect.