Bayesian hierarchical models for NHPP using `rstan`

Miao Cai miao.cai@slu.edu

2019-07-01

Contents

1 Model setting

Let $T_{d,s,i}$ denote the time to the d-th driver's s-th shift's i-th critical event. The total number critical events of d-th driver's s-th shift is $n_{d,s}$. The ranges of these notations are:

- $i = 1, 2, \cdots, n_{d,S_d}$,
- $s = 1, 2, \cdots, S_d$,
- $d = 1, 2, \cdots, D$.

We assume the times of critical events within the d-th driver's s-th shift were generated from a non-homogeneous Poisson process (NHPP) with a power law process (PLP), with a fix rate parameter β and varying scale parameters $\theta_{d,s}$ across drivers. The data generating process is then:

$$
\begin{aligned}
T_{d,s,1}, T_{d,s,2}, \cdots, T_{d,s,n_{d,s}} &\sim \text{PLP}(\beta, \theta_{d,s}) \\
\beta &\sim \text{Gamma}(1, 1) \\
\log \theta_{d,s} &= \gamma_{0d} + \gamma_1 x_{d,s,1} + \gamma_2 x_{d,s,2} + \cdots + \gamma_k x_{d,s,k} \\
\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0D} &\sim \text{i.i.d. } N(\mu_0, \sigma_0^2) \\
\gamma_1, \gamma_2, \cdots, \gamma_k &\sim \text{i.i.d. } N(0, 10^2) \\
\mu_0 &\sim N(0, 10^2) \\
\sigma_0 &\sim \text{Gamma}(1, 1)
\end{aligned}
$$

2 Simulating data

1. Random intercepts $\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0D}$. The standard deviation of μ_0 was intentionally set to small number 2 to make $\theta_{d,s}$ fall into a reasonably small range. If I otherwise set it as 10, $\theta_{d,s}$ may be more than 10^5 due to the exponentiation, which may not be realistic in real-life data.

$$\mu_0 \sim N(0, 2^2)$$
$$\sigma_0 \sim \text{Gamma}(1, 1)$$
$$\gamma_{01}, \gamma_{02}, \cdots, \gamma_{0D} \sim \text{i.i.d. } N(\mu_0, \sigma_0^2)$$

2. fixed parameters: 3 fixed parameters $\gamma_1, \gamma_2, \gamma_3$.

$$\gamma_1, \gamma_2, \gamma_3 \sim \text{i.i.d. } N(0, 10^2)$$

3. The number of observations in the d-th driver: N_d.

$$N_d \sim \text{Poisson}(100)$$

4. Data: 3 predictor variables $x_{d,s,1}, x_{d,s,2}, x_{d,s,3}$.

$$x_{d,s,1} \sim \text{N}(0, 10)$$
$$x_{d,s,2} \sim \text{Gamma}(10, 2)$$
$$x_{d,s,3} \sim \text{Poisson}(3.5)$$

5. Scale parameters of a NHPP (random effects): $\theta_{d,s}$.

$$\theta_{d,s} = \text{EXP}(\gamma_{0d} + \gamma_1 x_{d,s,1} + \gamma_2 x_{d,s,2} + \gamma_k x_{d,s,3})$$

6. Shape parameter of a NHPP (fixed effect): β.

$$\beta \sim \text{Gamma}(1, 1)$$

7. Simulate a NHPP based on β and $\theta_{d,s}$.

$$T_{d,s,1}, T_{d,s,2}, \cdots, T_{d,s,n_{d,s}} \sim \text{PLP}(\beta, \theta_{d,s})$$

```
set.seed(123)
D = 10 # the number of drivers
K = 3 # the number of predictor variables

# 1. Random-effect intercepts
# hyperparameters
mu0 = rnorm(1, mean = 0, sd = 2)
sigma0 = rgamma(1, 1, 1)
r_0D = rnorm(D, mean = mu0, sd = sigma0)

# 2. Fixed-effects parameters
R_K = rnorm(K, mean = 0, sd = 1)

# 3. The number of observations in the $d$-th driver: $N_{d}$
N_K = rpois(D, 100)

# 4. Generate data: x_1, x_2, .. x_K
sim1 = function(n = 10){
  x1 = rnorm(n, 0, 10)
  x2 = rgamma(n, 10, 2)
  x3 = rpois(n, 3.5)
  return(data.frame(x1, x2, x3))
}
simXD = function(ndrivers = D){
  XD = rep(list(data.frame()), ndrivers)
  for (i in 1:D) {
    XD[[i]] = sim1(N_K[i])
  }
  return(data.table::rbindlist(XD, idcol = "driver"))
```

```r
}
X = simXD()


# 5. Scale parameters of a NHPP
# 5a. parameter matrix: P
N_D = df[,.N,driver][["N"]]# N by driver
N_all = sum(N_D) # total N
P = cbind(r0 = rep(r_0D, N_D),
          t(replicate(N_all, R_K)))
theta = exp(rowSums(P))
hist(theta)
```

```r
y = c(0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1)
x = c(1, 3, 2, 0, -1, 5, 2, 6, 2, 5, 3, 3, 1)
fit = glm(y ~ x, family = "binomial")
yhat = predict(fit, type = "response")


sum(y)
sum(yhat)
```