Bayesian estimation for NHPP using `rstan`

Miao Cai miao.cai@slu.edu

2019-05-27

Contents

1 Theories of inference for the Power Law Process

1.1 Concepts

Intensity function The intensity function of a point process is:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t, t+\Delta t) \geq 1)}{\Delta t}$$

Nonhomogeneous Poisson Process The Nonhomogeneous Poisson Process (NHPP) is a Poisson process whose intensity function is non-constant.

When the intensity function of a NHPP has the form $\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1}$, where $\beta > 0$ and $\theta > 0$, the process is called **power law process** (PLP).

1. **Failure truncation**: When testing stops after a predetermined number of failures, the data are said to be failure truncated.
2. **Time truncation**: Data are said to be time truncated when testing stops at a predetermined time t.

Conditional probability

$$P(A \cap B) = P(A)P(B|A)$$
$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

1.2 The inference for the first two events

The first event

The cumulative density function (cdf) of time to the first event is $F(t_1)$: $F_1(t_1) = P(T_1 \leq t_1) = 1 - S(t_1)$.

The survival function for the first event $S_1(t_1)$ is:

$$\begin{aligned} S_1(t_1) &= P(T_1 > t_1) \\ &= P(N(0, t_1) = 0) \quad N\text{is the number of events} \\ &= e^{-\int_0^{t_1} \lambda_u du}(e^{-\int_0^{t_1} \lambda_u du})^0/0! \\ &= e^{-\int_0^{t_1} \lambda_u du} \end{aligned}$$

The probability density function (pdf) of time to the first event can be calculated by taking the first order derivative of the cdf $F_1(t_1)$:

$$
\begin{aligned}
f_1(t_1) &= \frac{d}{dt_1}F_1(t_1) \\
&= \frac{d}{dt_1}[1 - S_1(t_1)] \\
&= -\frac{d}{dt_1}S_1(t_1) \\
&= -\frac{d}{dt_1}e^{-\int_0^{t_1}\lambda(u)du} \\
&= -(-\lambda_{t_1})e^{-\int_0^{t_1}\lambda(u)du} \\
&= \lambda(t_1)e^{-\int_0^{t_1}\lambda(u)du}
\end{aligned}
$$

If this NHPP is a PLL, we plug in the intensity function $\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1}$, then we have:

$$
f_1(t_1) = \frac{\beta}{\theta}(\frac{t_1}{\theta})^{\beta-1}e^{-(\frac{t_1}{\theta})^{\beta}}, \quad t_1 > 0
$$

This pdf is identical with the pdf of Weibull distribution, so we have:

$$
T_1 \sim \text{Weibull}(\beta, \theta)
$$

The second event

The Survival function of the second event given the first event occurred at t_2 is:

$$
\begin{aligned}
S_2(t_2|t_1) &= P(T_2 > t_2|T_1 = t) \\
&= P(N(t_1, t_2) = 0|T_1 = t_1) \\
&= e^{-\int_{t_1}^{t_2}\lambda_u du}[\int_{t_1}^{t_2}\lambda_u du]^0/0! \\
&= e^{-\int_{t_1}^{t_2}\lambda_u du}
\end{aligned}
$$

The we can derive the pdf of t_2 conditioned on t_1

$$
\begin{aligned}
f(t_2|t_1) &= -\frac{d}{dt_2}S_2(t_2) \\
&= -\frac{d}{dt_2}e^{-\int_{t_1}^{t_2}\lambda(u)du} \\
&= \lambda(t_2)e^{-\int_{t_1}^{t_2}\lambda(u)du} \\
&= \frac{\beta}{\theta}(\frac{t_2}{\theta})^{\beta-1}e^{-[(\frac{t_2}{\theta})^{\beta}-(\frac{t_1}{\theta})^{\beta}]} \\
&= \frac{\frac{\beta}{\theta}(\frac{t_2}{\theta})^{\beta-1}e^{-(t_2/\theta)^{\beta}}}{e^{-(t_1/\theta)^{\beta}}}, \quad t_2 > t_1
\end{aligned}
\tag{1}
$$

1.3 Failure truncated case

In the failure truncated case, we know the total number of events n before the experiment starts. We can get the joint likelihood function for $t_1 < t_2 < \cdots < t_n$ in the failure truncated case based on Equation 1.

$$
\begin{aligned}
f(t_1, t_2, \cdots, t_n) &= f(t_1)f(t_2|t_1)f(t_3|t_1,t_2)\cdots f(t_n|t_1,t_2,\ldots,t_{n-1}) \\
&= \lambda(t_1)e^{-\int_0^{t_1}\dot{\lambda}(u)du}\lambda(t_2)e^{-\int_{t_1}^{t_2}\dot{\lambda}(u)du}\cdots\lambda(t_n)e^{-\int_{t_{n-1}}^{t_n}\lambda(u)du} \\
&= \Big(\prod_{i=1}^{n}\lambda(t_i)\Big)e^{-\int_0^{t}\lambda(u)du} \\
&= \Big(\prod_{i=1}^{n}\frac{\beta}{\theta}(\frac{t_i}{\theta})^{\beta-1}\Big)e^{-(t_n/\theta)^{\beta}}, \quad t_1 < t_2 < \cdots < t_n
\end{aligned}
\tag{2}
$$

The log-likelihood function in the failure truncated case is therefore:

$$
\log\ell = n\log\beta - n\beta\log\theta + (\beta-1)\Bigg(\sum_{i=1}^{n}\log t_i\Bigg) - \Big(\frac{t_n}{\theta}\Big)^{\beta}
$$

1.4 Time Truncated Case

We assume that the truncated time is τ. The derivation of $f(t_1, t_2, \cdots, t_n|n)$ is messy in math, we directly give the conclusion here:

$$
f(t_1, t_2, \cdots, t_n|n) = n!\prod_{i=1}^{n}\frac{\lambda(t_i)}{\Lambda(\tau)}
$$

Therefore, the joint likelihood function for $f(n, t_1, t_2, \cdots, t_n)$ is:

$$
\begin{aligned}
f(n, t_1, t_2, \cdots, t_n) &= f(n) f(t_1, t_2, \cdots, t_n | n) \\
&= \frac{e^{-\int_0^\tau \lambda(u)du} [\int_0^\tau \lambda(u)du]^n}{n!} n! \frac{\prod_{i=1}^n \lambda(t_i)}{[\Lambda(\tau)]^n} \\
&= \Big(\prod_{i=1}^n \lambda(t_i) \Big) e^{-\int_0^\tau \lambda(u)du} \\
&= \Big(\prod_{i=1}^n \frac{\beta}{\theta} (\frac{t_i}{\theta})^{\beta - 1} \Big) e^{-(\tau/\theta)^\beta}, \\
n = 0, 1, 2, \cdots, &\quad 0 < t_1 < t_2 < \cdots < t_n
\end{aligned}
\tag{3}
$$

The log likelihood function l is then:

$$
\begin{aligned}
l &= \log \Bigg(\Big(\prod_{i=1}^n \frac{\beta}{\theta} (\frac{t_i}{\theta})^{\beta - 1} \Big) e^{-(\tau/\theta)^\beta} \Bigg) \\
&= \sum_{i=1}^n \log \Big(\frac{\beta}{\theta} (\frac{t_i}{\theta})^{\beta - 1} \Big) - (\frac{\tau}{\theta})^\beta \\
&= n \log \beta - n\beta \log \theta + (\beta - 1) \Bigg(\sum_{i=1}^n \log t_i \Bigg) - \Big(\frac{\tau}{\theta} \Big)^\beta
\end{aligned}
\tag{4}
$$

1.5 Bayesian inference

After having the joint likelihood function in both the failure and time truncated case, it is straightforward to conduct Bayesian inference according to the Bayes theorem:

Theorem 1.1 (The Bayes Theorem)**.**

$$P(\theta | D) = \frac{P(\theta) \times P(D|\theta)}{P(D)}$$

Where θ is the parameter to be estimated, D is the observed data, $P(\theta)$ is the prior belief about the parameter θ, $P(D|\theta)$ is the likelihood function, and $P(D)$ is the normalizing constant to make the posterior density function integrates to 1.

The 1.1 can be written in a proportional format:

$$P(\theta | D) \propto P(\theta) \times P(D|\theta)$$

which means that the posterior density of a parameter is proportional to the product of the prior and the likelihood function, which is the key of Bayesian inference.

2 Bayesian estimation in simulated multiple shifts from one driver

2.1 Parameter setup and priors

- parameters: $\beta = 2, \theta = 10$
- Priors:

$$
\begin{aligned}
\beta &\sim \text{Gamma}(1,1), & E(\beta) &= \alpha/\beta = 1, & V(\beta) &= \alpha/\beta^2 = 1 \\
\theta &\sim \text{Gamma}(1,0.01), & E(\beta) &= \alpha/\beta = 100, & V(\beta) &= \alpha/\beta^2 = 10000
\end{aligned}
$$

In `Stan`, the parameterization of a Gamma distribution is:

$$\text{Gamma}(y|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y)$$

In this parameterization, the mean of a Gamma distribution is α/β and the variance is α/β^2.

2.2 Time truncated case

2.2.1 One estimation result each at at different sample sizes

Table 1: Parameter estimates when N = 5

	mean	sd	2.5%	50%	97.5%
beta	1.918	0.399	1.220	1.927	2.675
theta	9.252	2.091	5.312	9.430	13.113
lp___	-56.122	1.094	-58.743	-55.843	-54.921

Table 2: Parameter estimates when N = 10

	mean	sd	2.5%	50%	97.5%
beta	2.283	0.361	1.650	2.277	3.049
theta	11.228	1.326	8.652	11.186	13.626
lp___	-76.542	1.045	-79.312	-76.189	-75.543

Table 3: Parameter estimates when N = 50

	mean	sd	2.5%	50%	97.5%
beta	1.862	0.124	1.629	1.857	2.111
theta	8.973	0.680	7.513	8.939	10.256
lp___	-545.990	1.207	-549.391	-545.636	-544.695

Table 4: Parameter estimates when N = 100

	mean	sd	2.5%	50%	97.5%
beta	1.988	0.098	1.823	1.985	2.205
theta	9.861	0.450	9.012	9.827	10.796
lp___	-1021.537	1.025	-1024.400	-1021.179	-1020.481

Table 5: Parameter estimates when N = 400

	mean	sd	2.5%	50%	97.5%
beta	2.008	0.044	1.924	2.005	2.096
theta	9.891	0.207	9.486	9.887	10.291
lp___	-4028.537	1.149	-4031.990	-4028.179	-4027.481

2.2.2 30 simulations and estimations at different sample sizes

Since the parameter estimates of one simulation at different sample size may subject to sampling error. Here are the results:

Table 6: Summary results for parameter β

sample size	mean of the posterior means	s.d. of the posterior means	mean of the posterior s.e.
5	1.993	0.130	0.179
10	2.041	0.136	0.131
50	2.012	0.052	0.057
100	2.006	0.041	0.040
250	1.999	0.022	0.026
500	2.001	0.016	0.018

Table 7: Summary results for parameter θ

sample size	mean of the posterior means	s.d. of the posterior means	mean of the posterior s.e.
5	9.973	0.980	1.512
10	10.411	1.107	1.099
50	10.053	0.463	0.480
100	10.048	0.385	0.339
250	9.993	0.181	0.219
500	10.016	0.129	0.154

2.3 Failure truncated case

2.3.1 One estimation result each at different sample sizes

Then I ran one simulation per each number of shifts. I tried the number of shifts (N) as $5, 10, 50, 500, 1000$.

Table 8: Parameter estimates when N = 5

	mean	sd	2.5%	50%	97.5%
beta	2.703	0.599	1.703	2.668	3.992
theta	8.325	1.228	5.975	8.357	10.642
lp___	-28.349	1.067	-31.129	-27.966	-27.294

Table 9: Parameter estimates when N = 10

	mean	sd	2.5%	50%	97.5%
beta	2.037	0.234	1.604	2.033	2.539
theta	11.220	1.413	8.366	11.205	13.885
lp___	-132.706	0.954	-135.032	-132.423	-131.669

Table 10: Parameter estimates when N = 50

	mean	sd	2.5%	50%	97.5%
beta	2.105	0.119	1.892	2.105	2.314
theta	10.616	0.617	9.428	10.650	11.696
lp___	-601.260	1.002	-603.863	-601.017	-600.242

Table 11: Parameter estimates when N = 100

	mean	sd	2.5%	50%	97.5%
beta	2.081	0.083	1.929	2.082	2.250
theta	10.184	0.412	9.359	10.186	10.969
lp___	-1108.571	1.091	-1111.728	-1108.224	-1107.516

Table 12: Parameter estimates when N = 500

	mean	sd	2.5%	50%	97.5%
beta	1.974	0.032	1.914	1.974	2.035
theta	9.872	0.185	9.518	9.862	10.207
lp___	-5483.755	0.970	-5486.301	-5483.504	-5482.794

Table 13: Parameter estimates when N = 1000

	mean	sd	2.5%	50%	97.5%
beta	2.029	0.027	1.979	2.028	2.083
theta	10.123	0.137	9.872	10.113	10.419
lp___	-10711.933	0.993	-10714.448	-10711.653	-10710.897

The parameter estimates at different sample sizes seem to be quite well: **the points estimates are getting closer to true parameter values as the number of shifts increases**.

However, this is only one simluation for each sample size, which may be subject to sampling error (but at least the estimates seem reasonably well). In the following section, I need to **scale up the simulation to see if we get consistently good results as we perform repeated simulations**.

2.3.2 30 simulations and estimations at each different sample sizes

I simulated NHPP for 30 times and accordingly performed 30 Bayesian estimation for $\beta = 2$ and $\theta = 10$ at each sample size ($N = 5, 10, 50, 100, 500$).

Table 14: Summary results for parameter β

sample size	mean of the posterior means	s.d. of the posterior means	mean of the posterior s.e.
5	2.060	0.353	0.354
10	2.033	0.151	0.240
50	1.996	0.124	0.106
100	1.984	0.078	0.071
250	2.004	0.052	0.046
500	1.995	0.027	0.033

Table 15: Summary results for parameter θ

sample size	mean of the posterior means	s.d. of the posterior means	mean of the posterior s.e.
5	10.094	2.021	1.813
10	10.470	1.039	1.369
50	9.977	0.588	0.598
100	9.889	0.457	0.404
250	10.043	0.288	0.263
500	10.005	0.180	0.188

It seems that Bayesian estimates of β and θ are getting closer to true parameter values as the number of shifts increase:

- The bias was getting smaller: $|\hat{\beta} - \beta|$ is getting closer to 0 as N increases (the 2nd column),
- The standard error of posterior mean was getting smaller, as we can tell from the 3rd column.

3 Hierarchical Bayesian model for PLP

3.1 Time truncated case

$$\beta_{d(i)} \sim N(2, 0.5^2)$$
$$\theta_{d(i)} \sim N(10, 2^1)$$
$$T_{i,d(i)} \sim PLP(\beta_{d(i)}, \theta_{d(i)})$$

how to incoporate covariates into this hierarchical model?

Appendix - stan code

Time truncated case

```
functions{
  real nhpp_log(vector t, real beta, real theta, real tau){
    vector[num_elements(t)] loglik_part;
    real loglikelihood;
    for (i in 1:num_elements(t)){
      loglik_part[i] = log(beta) - beta*log(theta) + (beta - 1)*log(t[i]);
    }
    loglikelihood = sum(loglik_part) - (tau/theta)^beta;
    return loglikelihood;
  }
}
data {
  int<lower=0> N; //total # of obs
  int<lower=0> K; //total # of groups
  vector<lower=0>[K] tau;//truncated time
  vector<lower=0>[N] event_time; //failure time
  int s[K]; //group sizes
}
parameters{
  real<lower=0> beta;
  real<lower=0> theta;
}
model{
  int position;
  position = 1;
  for (k in 1:K){
    segment(event_time, position, s[k]) ~ nhpp(beta, theta, tau[k]);
    position = position + s[k];
  }
```

```
//PRIORS
  beta ~ gamma(1, 1);
  theta ~ gamma(1, 0.01);
}
```

Failure truncated case

```
functions{
  real nhpp_log(vector t, real beta, real theta, real tn){
    vector[num_elements(t)] loglik_part;
    real loglikelihood;
    for (i in 1:num_elements(t)){
      loglik_part[i] = log(beta) - beta*log(theta) + (beta - 1)*log(t[i]);
    }
    loglikelihood = sum(loglik_part) - (tn/theta)^beta;
    return loglikelihood;
  }
}
data {
  int<lower=0> N; //total # of obs
  int<lower=0> K; //total # of groups
  vector<lower=0>[K] tn;//truncated time
  vector<lower=0>[N] event_time; //failure time
  int s[K]; //group sizes
}
parameters{
  real<lower=0> beta;
  real<lower=0> theta;
}
model{
  int position;
  position = 1;
  for (k in 1:K){
```

```
    segment(event_time, position, s[k]) ~ nhpp(beta, theta, tn[k]);
    position = position + s[k];
  }
//PRIORS
  beta ~ gamma(1, 1);
  theta ~ gamma(1, 0.01);
}
```