Bayesian estimation for NHPP using `rstan`

Miao Cai miao.cai@slu.edu

2019-05-24

1 One estimation result each at $N = 5, 10, 50, 500, 1000$

I set the theoretical values as: $\beta = 2, \theta = 5$

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

Table 1: Parameter estimates when N = 5

	mean	sd	2.5%	50%	97.5%
beta	2.778	0.649	1.724	2.720	4.236
theta	4.198	0.594	2.962	4.217	5.282
lp___	-18.571	1.033	-21.450	-18.263	-17.546

Table 2: Parameter estimates when N = 10

	mean	sd	2.5%	50%	97.5%
beta	1.991	0.219	1.607	1.983	2.398
theta	5.495	0.667	4.240	5.474	6.846
lp___	-92.420	0.980	-94.860	-92.173	-91.422

Table 3: Parameter estimates when N = 50

	mean	sd	2.5%	50%	97.5%
beta	2.103	0.106	1.906	2.099	2.319
theta	5.296	0.286	4.798	5.316	5.896
lp___	-409.260	1.109	-412.175	-408.970	-408.196

Table 4: Parameter estimates when N = 100

	mean	sd	2.5%	50%	97.5%
beta	2.074	0.083	1.916	2.080	2.225
theta	5.076	0.210	4.669	5.078	5.450
lp___	-756.363	1.011	-759.102	-756.064	-755.341

Table 5: Parameter estimates when N = 500

	mean	sd	2.5%	50%	97.5%
beta	1.978	0.037	1.909	1.977	2.055
theta	4.948	0.101	4.754	4.949	5.154
lp___	-3774.525	1.066	-3777.162	-3774.204	-3773.444

Table 6: Parameter estimates when N = 1000

	mean	sd	2.5%	50%	97.5%
beta	2.032	0.024	1.984	2.032	2.080
theta	5.066	0.063	4.936	5.066	5.179
lp___	-7357.631	0.896	-7359.840	-7357.375	-7356.686

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 30 simulations and estimations per each $N = 5, 10, 50, 100, 500$

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

Table 7: 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	2.068	0.294	0.398
10	2.011	0.242	0.255
50	2.049	0.121	0.117
100	2.020	0.066	0.082
500	2.007	0.042	0.036

Table 8: 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	4.912	0.855	0.909
10	5.116	0.699	0.695
50	5.148	0.265	0.308
100	5.007	0.154	0.216
500	5.031	0.110	0.097

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.

Appendix - stan code

```
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);
}
```