

Bayesian Inference Project

Juntao Zhang

5/16/2022

Introduction

The reproduction number R_t is a common measure of transmissibility of an epidemic disease. By definition, R_t is the average number of secondary cases caused by an infected individual. R_t falling below 1 indicates the disease is unlikely to sustain. When it approaches 0, it indicates the disease is under control. The change of R_t across time can be used as a proxy for epidemic trajectories. Following the method proposed by Cori et al., R_t can be calculated by using Bayesian parametric estimation.

In this project, I use the same method to compute the R_t of COVID-19 in New York before the presence of effective vaccination (Mar 2020 to Dec 2020). The case data of New York is extracted from CDC.

```
library(dplyr)
library(lubridate)
library(rstan)
library(ggplot2)
```

```
df <-
  read.csv("NY_cases.csv")
df <- df %>% select(state, submission_date, new_case, tot_cases)
```

Method with Poisson Likelihood

Assuming the number of reported incident cases follows a poisson process, we have:

$$P(I_t | I_0, I_1, \dots, I_{t-1}, w_s, R_t) = \frac{(R_t \Lambda_t(w_s))_{I_t}^t \exp(-R_t \Lambda_t(w_s))}{I_t!},$$

where I_t is the number of incident cases arising at time t and $\Lambda_t(w_s)$ is the overall infectivity and can be computed as:

$$\Lambda_t(w_s) = \sum_{s=1}^t I_{t-s} w_s.$$

w_s is the serial interval distribution, which is the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases. It serves as the weight for reported new cases at each time step before time t .

I choose the distribution of the serial period of COVID-19 follows a gamma distribution with mean of 5.9 and standard deviation of 3.9, according to the paper of Liu et al.. I also choose the prior of R_t to follow a gamma distribution with mean = 1.5 and standard deviation of 2.

With the above assumptions and procedure, I use Stan to yield the posterior distribution of R_t on the last day of 2020 (Dec 31st, 2020) with all previous reported cases.

```
#get the weight of previous days using a gamma distribution
lastday <- df %>%
  mutate(days_away=rev(row_number())-1,
         weight=dgamma(days_away,shape=(5.9/3.9)^2,rate=5.9/3.9^2))
```

Stan Code

```
data { /* these are known and passed as a named list from R */
  int<lower = 0> I;           // number of cases in day t
  real<lower = 0> infect;     // weighted sum of previous infectivity
  real<lower = 0> alpha;     //shape parameter of gamma prior
  real<lower = 0> beta;      // rate parameter of gamma prior
  int<lower = 0, upper = 1> prior_only;
}
```

```
parameters {
  real<lower=0> Rt; // Reproduction number
}
```

```
model {
  if (!prior_only) {
    target +=poisson_lpmf(I | Rt*infect); // log-likelihood
  }
  target += gamma_lpdf(Rt | alpha, beta); //prior of Rt
}
```

```
#calculating overall infectivity
overallinfectivity=sum(lastday$new_case*lastday$weight)
#indicate new case on Dec 31,2020
It=lastday$new_case[nrow(lastday)]
#choose prior parameters (shape and rate)
a=9/16
b=3/8
```

```
post <- stan("reproduction_num.stan",
            data = list(infect=overallinfectivity,I=It, prior_only = 0, alpha = a, beta=b))
```

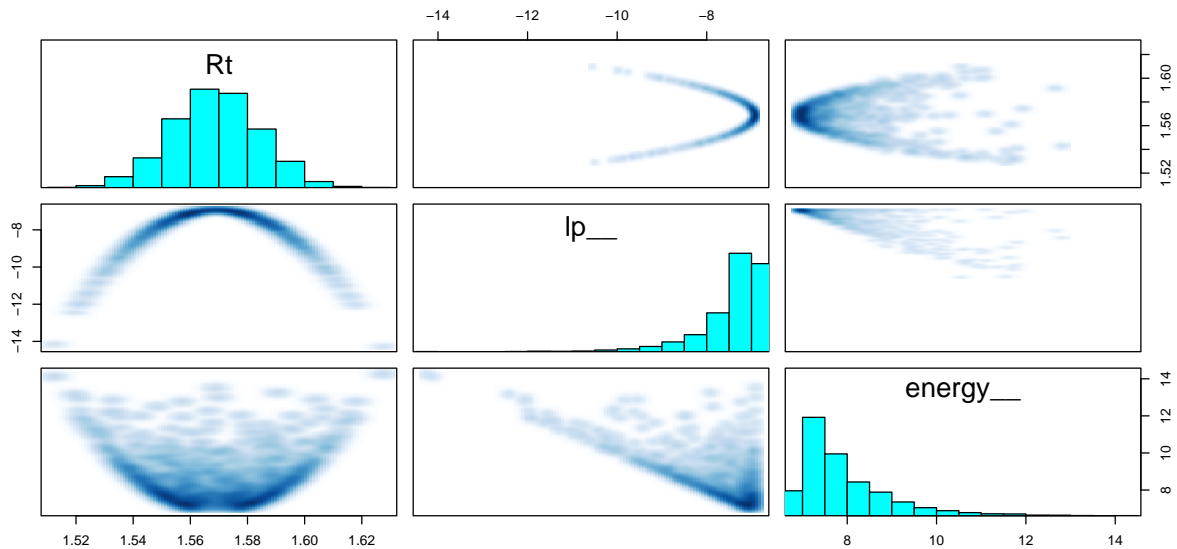
```
post
```

```
## Inference for Stan model: reproduction_num.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##      mean se_mean   sd  2.5%   25%   50%   75%  97.5% n_eff Rhat
## Rt      1.57     0.00 0.02   1.54   1.56   1.57   1.58   1.60  1383    1
## lp__ -7.42     0.02 0.73  -9.48 -7.61 -7.14 -6.96 -6.91  1484    1
##
## Samples were drawn using NUTS(diag_e) at Mon May 16 18:59:51 2022.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

The mean of the posterior distribution of R_t on Dec.31st, 2020 in New York is 1.57 with a 0.95 credible interval of (1.54,1.60).

Posterior Planes

```
pairs(post, pars = "p", include = FALSE)
```



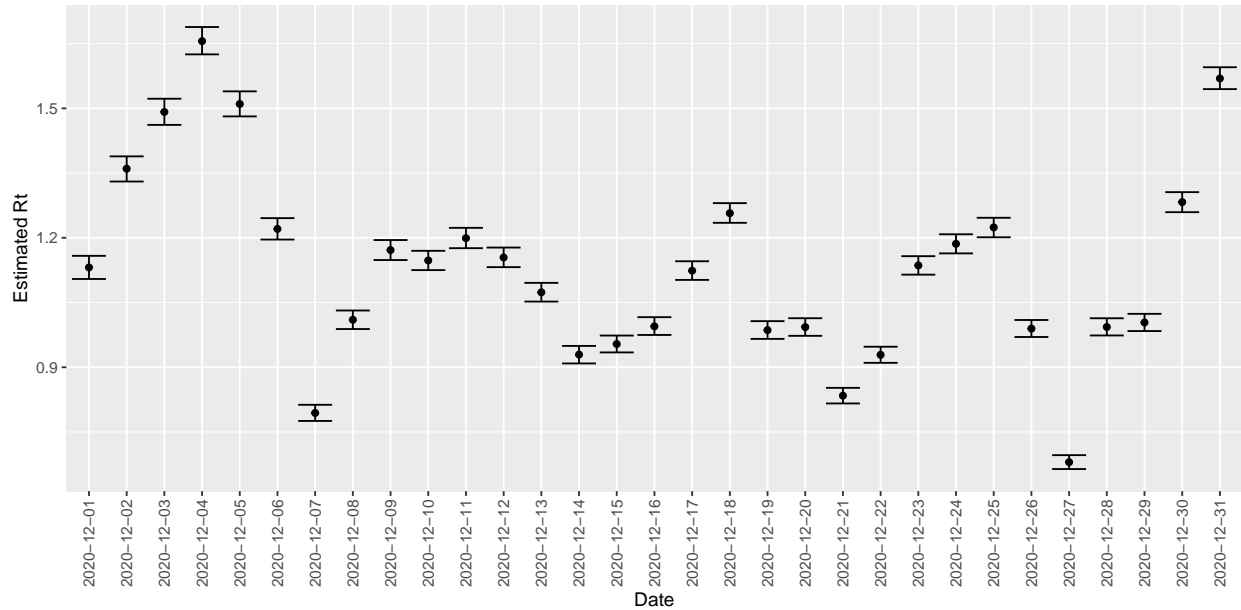
Compute For All Time Points

Essentially, we can compute the posterior distribution of R_t for all time points in the data set, assuming the same serial interval distribution over time.

```
df["Rt"]<-NA
df["lower"]<-NA
df["upper"]<-NA
for (i in 2:nrow(df)){
  dt<-df[1:i,] %>%
  mutate(days_away=rev(row_number())-1,
          weight=dgamma(days_away,shape=(5.9/3.9)^2,rate=5.9/3.9^2))
  overall=sum(dt$new_case*dt$weight)
  I_t=dt$new_case[nrow(dt)]
  a=9/16
  b=3/8
  posterior <- stan("reproduction_num.stan",
                    data = list(infect=overall,I=I_t, prior_only = 0, alpha = a, beta=b))
  draws <- as.data.frame(posterior) %>% select(-starts_with("p"))
  df$Rt[i]<-mean(draws$Rt)
  df$lower[i]<-quantile(draws$Rt, probs = c(.05))
  df$upper[i]<-quantile(draws$Rt, probs = c(.95))
}
```

Plot for estimated R_t with 95% credible interval in New York from 2020-12-01 to 2020-12-31

```
ggplot(df[271:301,], aes(submission_date,Rt)) +
  geom_point() +
  geom_errorbar(aes(ymin = lower, ymax = upper))+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  xlab("Date") + ylab("Estimated Rt")
```



From the above plot we can see how the R_t changes through December 2020. The small fluctuations can be reduced by using a sliding window or assuming that R_t remains the same for a period of time to have a clear look at the overall trend.

Method with Negative Binomial Likelihood

The method can be modified by using a negative binomial likelihood instead of a poisson likelihood since negative binomial distribution allows for overdispersion. The negative binomial distribution has an extra dispersion parameter ϕ and it is assumed to follow a half cauchy distribution with location parameter =30 and scale = 3.

Stan Code

```
data { /* these are known and passed as a named list from R */
  int<lower = 0> I;          // number of cases in day t
  real<lower = 0> infect;    // weighted sum of previous infectivity
  real<lower = 0> alpha;     //shape parameter of gamma prior
  real<lower = 0> beta;      // rate parameter of gamma prior
  real<lower = 0> mu;        //location parameter of cauchy prior
  real<lower = 0> sigma;     //scale parameter of cauchy prior
  int<lower = 0, upper = 1> prior_only;
}

parameters {
```

```

    real<lower=0> Rt; // Reproduction number
    real<lower=0> phi; // dispersion parameter
  }

  model {
    if (!prior_only) {
      target += neg_binomial_2_lpmf(I | Rt*infect, phi); // log-likelihood
    }
    target += gamma_lpdf(Rt | alpha, beta); //prior of Rt
    target += cauchy_lpdf(phi | mu, sigma); //prior of phi
  }

  post_neg <- stan("reproduction_num2.stan",
    data = list(infect=overallinfectivity,I=It, prior_only = 0,
      alpha = a, beta=b, mu=30,sigma=5))

```

```
post_neg
```

```

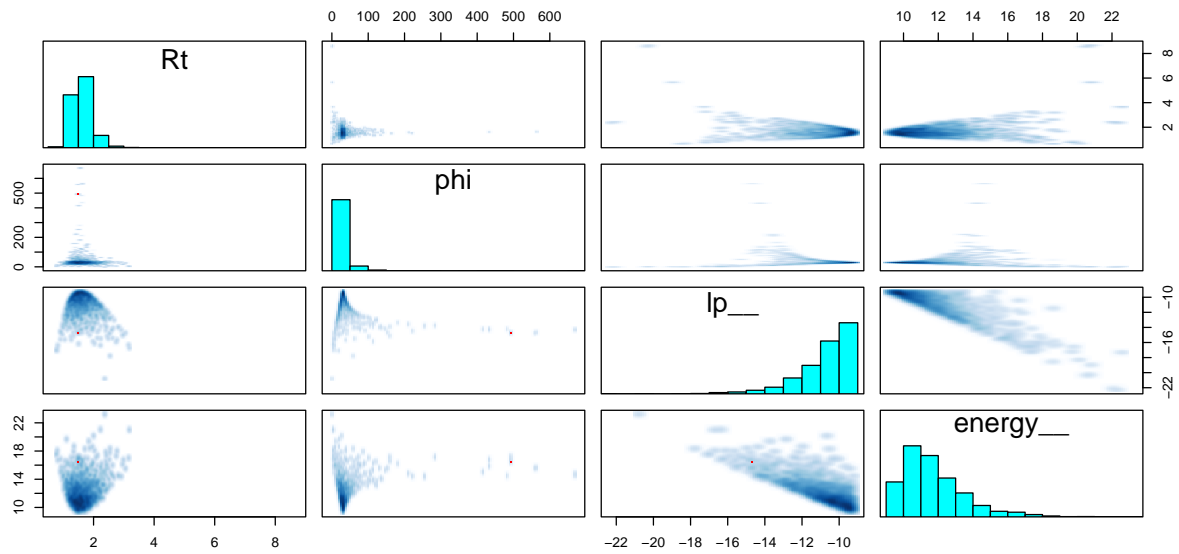
## Inference for Stan model: reproduction_num2.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##      mean se_mean    sd  2.5%   25%   50%   75% 97.5% n_eff Rhat
## Rt      1.62    0.01 0.36   1.11   1.40   1.58   1.80  2.37 1143 1.00
## phi    34.51    1.91 32.87  10.27  26.21  30.20  35.10 79.91   295 1.01
## lp__ -10.74    0.06  1.48 -14.68 -11.37 -10.32  -9.65  -9.24   667 1.01
##
## Samples were drawn using NUTS(diag_e) at Mon May 16 19:01:13 2022.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).

```

After modification, the negative binomial likelihood has higher standard deviation for the estimated R_t and gives a wider credible interval which implies higher uncertainty.

Posterior Planes

```
pairs(post_neg, pars = "p", include = FALSE)
```



Reference

Anne Cori, Neil M. Ferguson, Christophe Fraser, Simon Cauchemez, A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, *American Journal of Epidemiology*, Volume 178, Issue 9, 1 November 2013, Pages 1505–1512, <https://doi.org/10.1093/aje/kwt133>

Liu, X., Xu, X., Li, G., Xu, X., Sun, Y., Wang, F., Shi, X., Li, X., Xie, G., & Zhang, L. (2021). Differential impact of non-pharmaceutical public health interventions on COVID-19 epidemics in the United States. *BMC Public Health*, 21(1), 965. <https://doi.org/10.1186/s12889-021-10950-2>