HW07

Chongdan Pan

2022-03-25

Question 7.1

a. The process is very smooth, I just need to follow the tutorial and run test.sbat

b. Yes, I’ve used the jupyter notebook on the greatlake cluster

c. The time 0 is using one core to generate normal distributioned random variables while time 1-4 are using multiple cores. My laptop also has multiple cores so that it’s not too slower than the cluster, but the cluster’s advantage will be more obvious as the number of variables becomes larger.

	user.self	sys.self	elapsed	user.child	sys.child
My Time 0	1.955	0.221	2.176	0	0
My Time 1	0.234	0.504	1.24	0.27	0.145
My Time 2	0.144	0.524	1.118	3.75	0.46
My Time 3	0.193	0.619	1.257	5.764	1.427
My Time 4	0.632	1.245	2.318	5.52	7.384
Greatlake Time 0	6.409	0.235	6.661	0	0
Greatlake Time 1	0.215	0.88	2.079	0	0
Greatlake Time 2	0.184	0.79	1.385	1.085	0.264
Greatlake Time 3	0.342	0.6	1.386	6.79	2.007
Greatlake Time 4	1.054	0.931	2.244	6.974	2.818

7.2

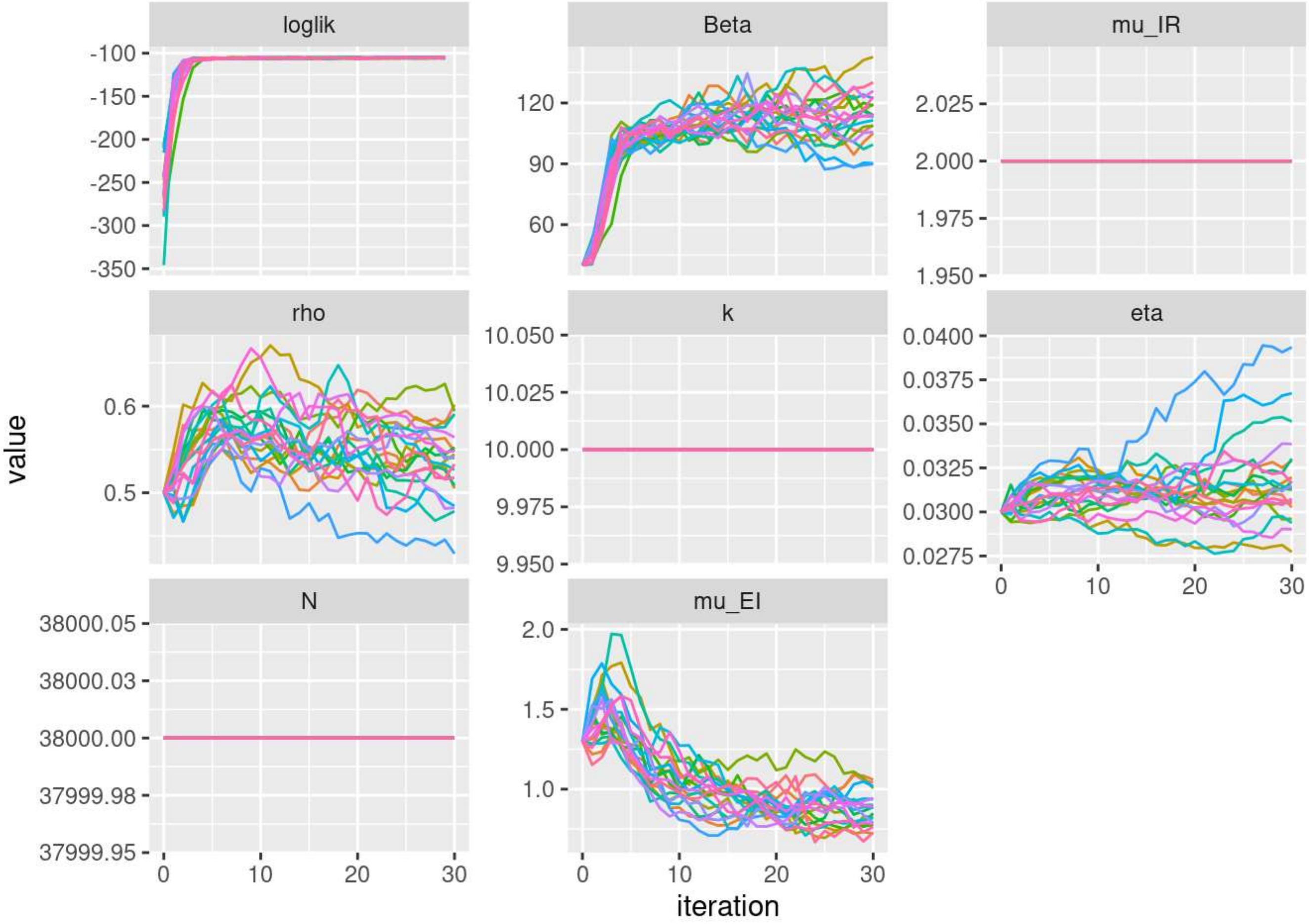
The following table shows the time required for each step for my laptop and great lake cluster. It turns out that the cluster’s supremacy is more obvious as we’re processing more data or simulation.

	user	system	elapsed
Laptop Local Search	102.49	11.15	6.37
Laptop Local Likelihood Estimate	11.8	10.55	6.96
Laptop Global Search	3319.62	11.28	177.61
Laptop Profile Likelihood Interval	19025.7	61.35	1006.76
Cluster Local Search	86.54	11.57	5.69
Cluster Local Likelihood Estimate	113.08	18.19	7.82
Cluster Global Search	2139.335	43.533	73.755
Cluster Profile Likelihood Interval	4714	69.47	182.2

a.

1. A local search

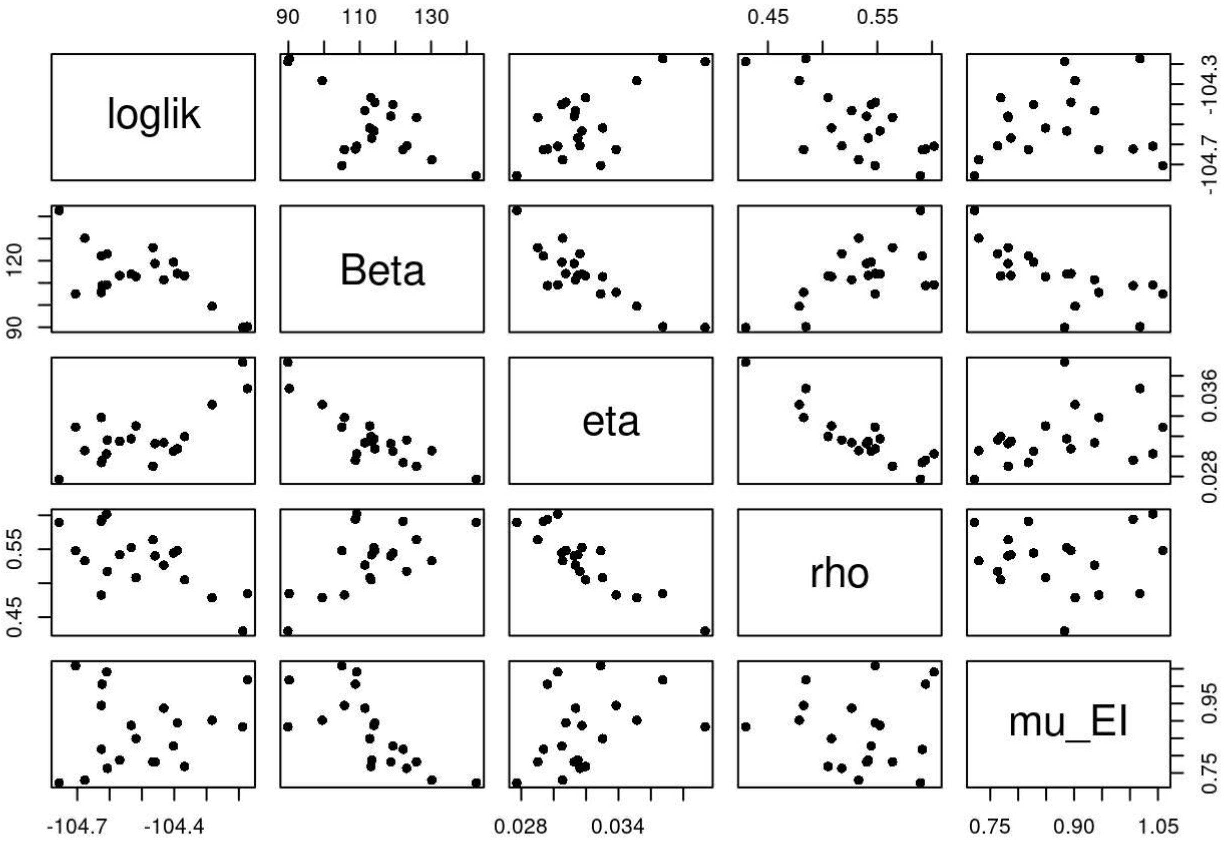
```
library(tidyverse)
library(pomp)
library(foreach)
library(doParallel)
library(doRNG)
cores <-  as.numeric(Sys.getenv('SLURM_NTASKS_PER_NODE', unset=NA))
if(is.na(cores)) cores <- detectCores()
registerDoParallel(cores)
registerDoRNG(1)
source("SEIR.R")

fixed_params <- c(N=38000, mu_IR=2, k=10)
coef(measSEIR, names(fixed_params)) <- fixed_params

# Local Particle Filtering
system.time(
 foreach(i=1:20,.combine=c) %dopar% {
    measSEIR %>%
      mif2(
          Np=1000, Nmif=30,
          cooling.fraction.50=0.5,
          rw.sd=rw.sd(Beta=0.02, rho=0.02, eta=ivp(0.02), mu_EI=0.02),
        )
  } -> mifs_local
) -> time0
mifs_local %>%
traces() %>%
melt() %>%
ggplot(aes(x=iteration,y=value,group=L1,color=factor(L1)))+
geom_line()+
guides(color="none")+
facet_wrap(~variable,scales="free_y")
```



Just as we get from SIR model from slides, the log likelihood increase without too much iterations. The is an obvious growing trend for beta, whose value should lie in [90,100]. It also looks like that the value of mu_EI should be in [0.8, 1], but it’s not so clear for other parameters.

2. Estimate the likelihood

```
# Calculate the likelihood
system.time(
 foreach(mf=mifs_local,.combine=rbind) %dopar% {
    evals <- replicate(10, logLik(pfilter(mf,Np=5000)))
    ll <- logmeanexp(evals,se=TRUE)
    mf %>% coef() %>% bind_rows() %>%
    bind_cols(loglik=ll[1],loglik.se=ll[2])
  } -> local_logliks
) -> time1
pairs(~loglik+Beta+eta+rho+mu_EI,data=local_logliks,pch=16)
```

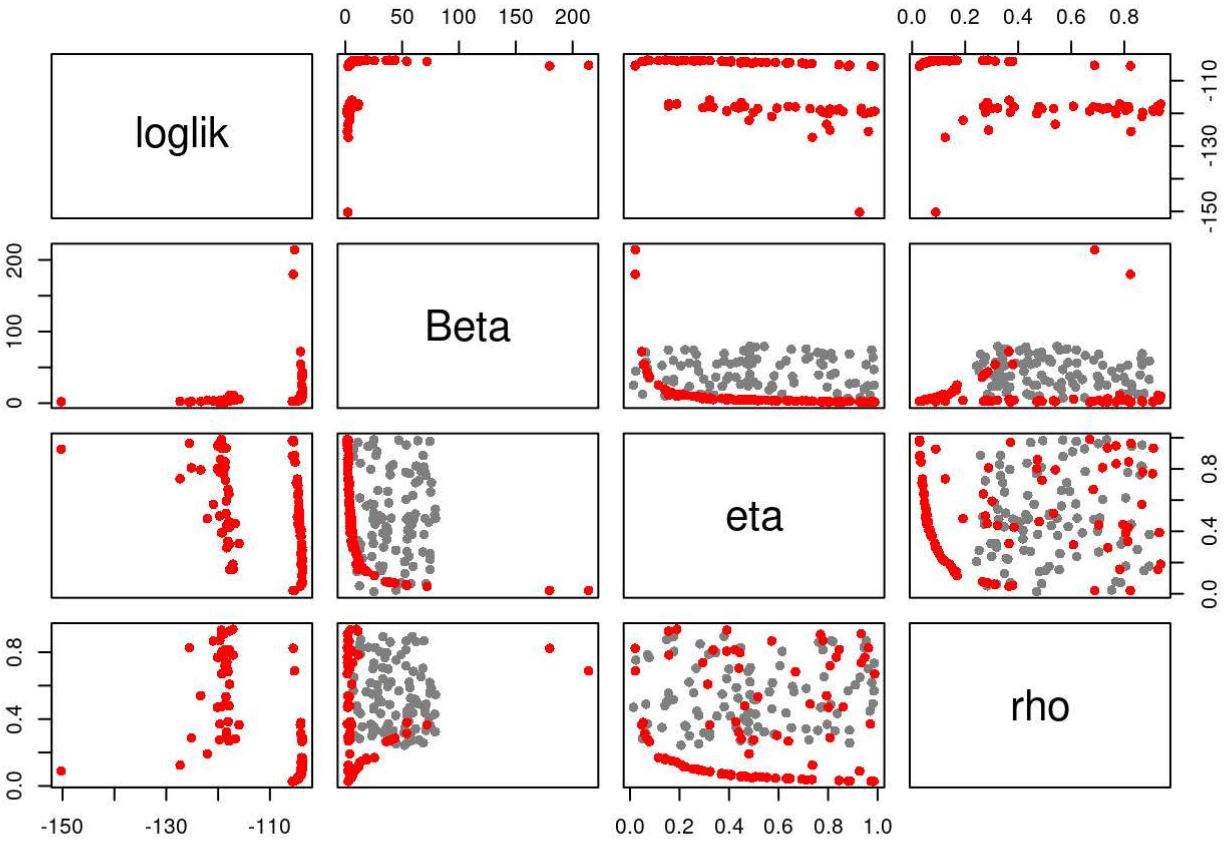


```
bind_rows(local_logliks) %>% filter(is.finite(loglik)) %>%
arrange(-loglik) %>% write_csv("measles_params.csv")
```

Based on the pair plot, it turns out the parameters lies in the region shown in the previous part, but it also shows a relationship between parameters and log likelihood similar to SIR model.

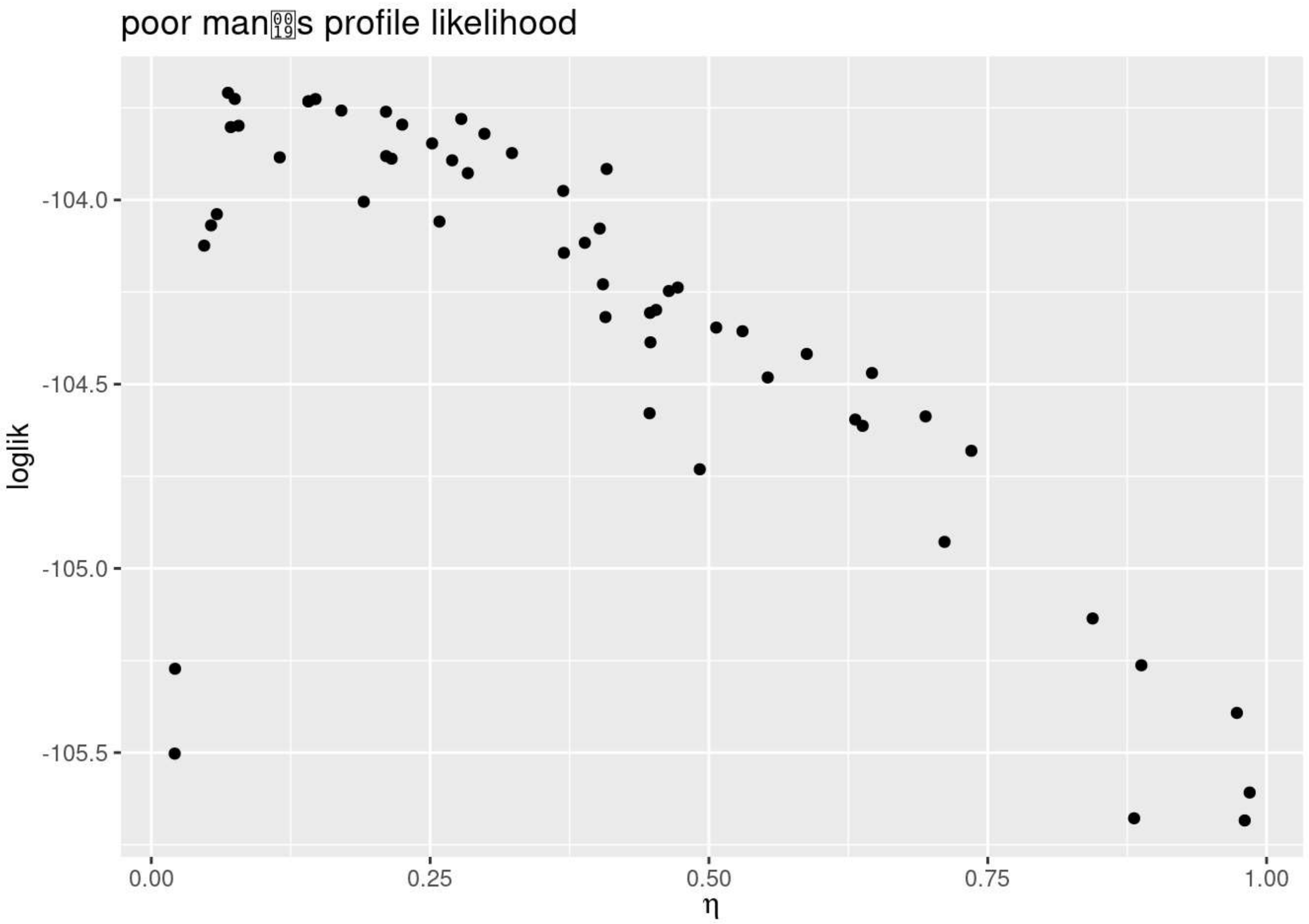
3. Global Search

```
set.seed(1)
# Set the box
runif_design(
  lower=c(Beta=5,rho=0.2,eta=0,mu_EI=0.5),
  upper=c(Beta=80,rho=0.9,eta=1,mu_EI=2),
  nseq=100
) -> guesses
mf1 <- mifs_local[[1]]

system.time(
  # A global search
  foreach(guess=iter(guesses,"row"), .combine=rbind) %dopar% {
    mf1 %>% mif2(params=c(guess,fixed_params)) %>% mif2(Nmif=100) -> mf
    replicate(
      10,
      mf %>% pfilter(Np=1000) %>% logLik()
    ) %>% logmeanexp(se=TRUE) -> ll

    mf %>% coef() %>% bind_rows() %>%
    bind_cols(loglik=ll[1],loglik.se=ll[2])
  } -> global_logliks
) -> time2

read_csv("measles_params.csv") %>%
bind_rows(global_logliks) %>% filter(is.finite(loglik)) %>%
arrange(-loglik) %>% write_csv("measles_params.csv")

global_logliks %>% filter(loglik>max(loglik)-50) %>% bind_rows(guesses) %>%
mutate(type=if_else(is.na(loglik),"guess","result")) %>% arrange(type) -> color_liks
pairs(~loglik+Beta+eta+rho, data=color_liks, pch=16, col=ifelse(color_liks$type=="guess",grey(
0.5),"red"))
```



b.

```
color_liks %>%
filter(type=="result") %>%
filter(loglik>max(loglik)-10) %>%
ggplot(aes(x=eta,y=loglik))+
geom_point()+
labs(
  x=expression(eta),
  title="poor man  s profile likelihood"
)
```



The max likelihood we get is -103.7, which is higher than the value we got from SIR model, implying that the SEIR has a high accuracy for fitting.

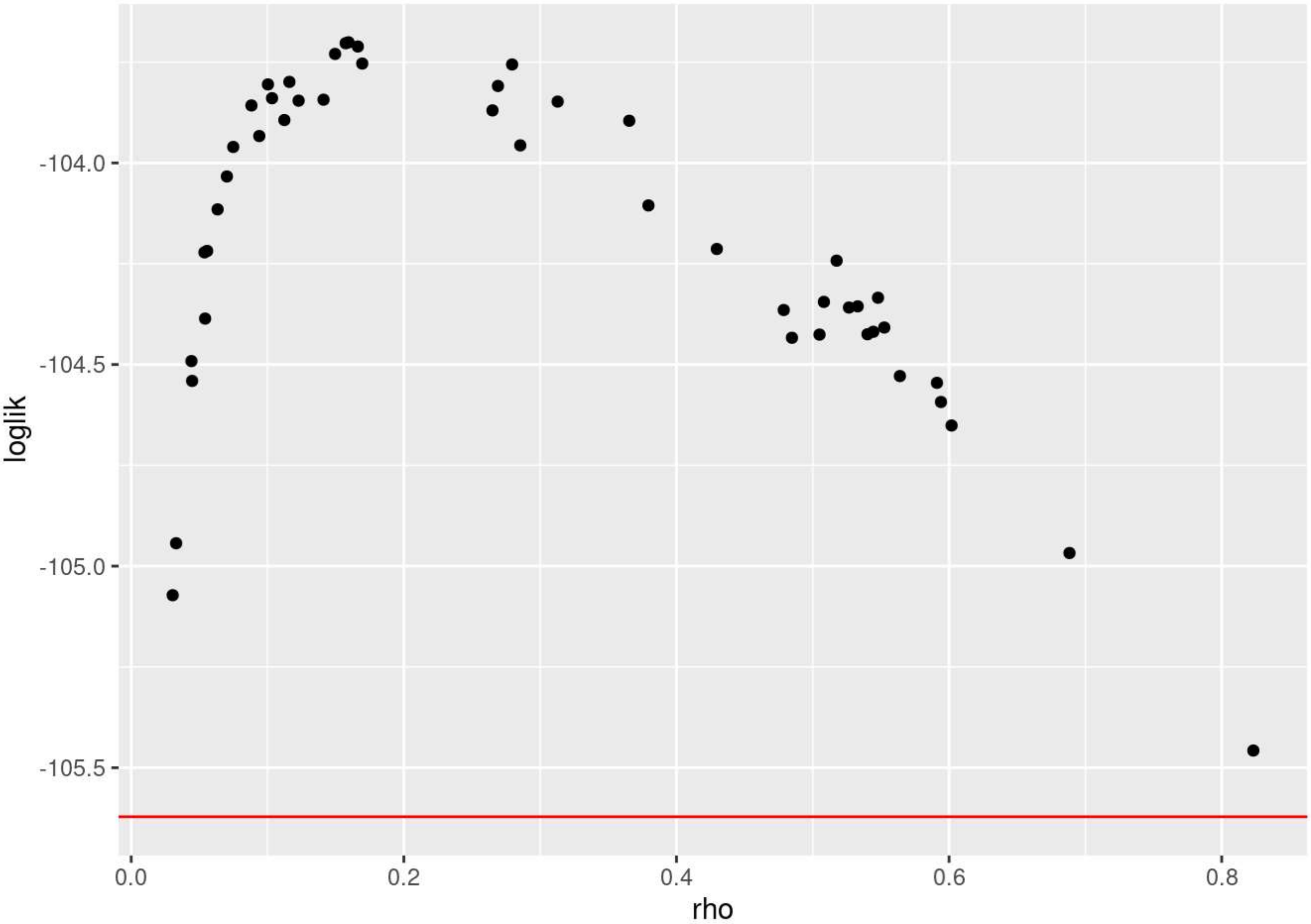
c.

```
global_logliks %>% filter(loglik>max(loglik)-20,loglik.se<2) %>%
sapply(range) -> box
```

Compared to the estimates from SIR models, SEIR has a larger Beta,but similar values for eta and rho’s estimates. Note that the difference may be caused by a different mu_IR of our model.

d.

```
read_csv("measles_params.csv") %>%
group_by(cut=round(rho,2)) %>%
filter(rank(-loglik)<=10) %>%
ungroup() %>%
arrange(-loglik) %>%
select(-cut,-loglik,-loglik.se) -> guesses

system.time(
  foreach(guess=iter(guesses,"row"), .combine=rbind) %dopar% {
    mf1 %>% mif2(params=guess,
    rw.sd=rw.sd(Beta=0.02,eta=ivp(0.02),mu_EI=0.02)) %>%
    mif2(Nmif=100,cooling.fraction.50=0.3) %>%
    mif2() -> mf
    replicate(
    10,
    mf %>% pfilter(Np=5000) %>% logLik()) %>%
    logmeanexp(se=TRUE) -> ll
    mf %>% coef() %>% bind_rows() %>%
    bind_cols(loglik=ll[1],loglik.se=ll[2])
  } -> profile_logliks
) -> time3

profile_logliks %>%
  filter(loglik>max(loglik)-10,loglik.se<1) %>%
  group_by(round(rho,2)) %>%
  filter(rank(-loglik)<3) %>%
  ungroup() %>%
  ggplot(aes(x=rho,y=loglik))+
  geom_point()+
  geom_hline(
  color="red",
  yintercept=max(profile_logliks$loglik)-0.5*qchisq(df=1,p=0.95)
)
```



The SEIR model's confidence interval for ρ is about [0.01,0.8], which is much larger than the interval we get from page 75 of the slides. The interval is much larger, but it looks like to be consistent with our previous pair plot. The maximum likelihood estimates for ρ is about 0.15 which is larger too. In addition, the likelihood we got from SEIR is also greater than SIR model, so I think the different set up of the model lead to different confidence interval of ρ and lead to a better performance.

Refernce

- Slides about Chapter 12 (https://ionides.github.io/531w22/12/index.html)
- Slides about Chapter 13 (https://ionides.github.io/531w22/13/index.html)
- Slides about Chapter 14 (https://ionides.github.io/531w22/14/index.html)
- Previous Solution (https://ionides.github.io/531w21/hw07/sol07.html)