ensemble-IC3

Saier Gong + Zhijie Huang

7/13/2020

IMPORT DATA

```
data<-read.csv(file="US-Coronavirus-data.csv",header = TRUE) %>%
#data<-data %>%
  #mutate(date=as.character(date)) %>%
  mutate(date=as.Date(date)) %>%
  #mutate(date=paste(month(date),"/",day(date))) %>%
  #mutate(date=as.Date(date,"%m-%d")) %>%
  #mutate(S=N-C) %>%
  mutate(daily.D=D-c(0,D[-length(D)])) %>%
  filter(!is.na(C)) %>%
  filter(date>=as.Date('2020-02-29'))

DAILY<-data %>%
  mutate(add.D=D-c(0,D[1:(length(D)-1)])) %>%
  mutate(add.C=C-c(0,C[1:(length(C)-1)])) %>%
  select(date,C,add.C,D,add.D,X) %>%
  filter(date<=as.Date("2020-07-31"))

DAILY.train<-head(DAILY,nrow(DAILY)-7)
```

MODEL & MSE & RMSE

We estimate new reporting cases in the coming week by sampling from a Gamma distribution with mean and standard deviation estimated from the number of observed cases in the last week.

```
####
mse<-c()
exam<-c()
for (i in 1:5000) {


mu<-mean(tail(DAILY.train$add.C,7))
sd<-sd(tail(DAILY.train$add.C,7))

##gamma(shape,rate)
rate<-mu/sd^2
shape<-mu*rate
```

```r
## estimated reported daily cases
new.daily.C.train<-rgamma(7,shape=shape,rate=rate)
#(sum(new.daily.C.report))

daily.C.train<-append(DAILY.train$add.C,new.daily.C.train)


library(purrr)
### daily death ~ Binom(n,p)
### assume gamma(mean=10,sd=2)----> shape=25;rate=2.5

##n=?
#t=nrow(DAILY)
f1<-function(t){

  n=sum(map_dbl(0:(t-1),function(x) dgamma(x,shape=25,rate=2.5) * daily.C.train[(t-x)]))


  return(n)
}

n<-map_dbl((1+nrow(DAILY.train)):(7+nrow(DAILY.train)),function(t) f1(t))

##p=?
p=mean(tail(DAILY.train$add.D,7)/tail(DAILY.train$add.C,7))

new.daily.D<-map_dbl(n,function(x) rbinom(1,size=round(x),prob=p))

D.total.exam<-map_dbl(1:(nrow(DAILY.train)+7),function(x) sum(head(append(DAILY.train$add.D,new.daily.D
D.exam<-tail(D.total.exam,7)

error<-mean((D.exam-tail(DAILY$D,7))^2)
mse<-append(mse,error)
exam<-rbind(exam,D.exam)

}
mean(mse)
```

```
## [1] 577800.5
```

```r
(rmse<-sqrt(mean(mse)))
```

```
## [1] 760.1319
```

```r
colMeans(exam)
```

```
## [1] 146994.9 147920.6 148834.4 149730.5 150615.8 151503.9 152406.0
```

FORECAST

```r
####We estimate new reporting cases in the coming week by sampling from a Gamma distribution with mean
tt<-7

D.fore<-c()
for (i in 1:5000) {

mu<-mean(tail(DAILY$add.C,7))
sd<-sd(tail(DAILY$add.C,7))

##gamma(shape,rate)
rate<-mu/sd^2
shape<-mu*rate

## estimated reported daily cases
new.daily.C<-rgamma(tt,shape=shape,rate=rate)
#(sum(new.daily.C.report))

daily.C<-append(DAILY$add.C,new.daily.C)



f1<-function(t){

  n=sum(map_dbl(0:(t-1),function(x) dgamma(x,shape=25,rate=2.5) * daily.C[(t-x)]))


  return(n)
}

n<-map_dbl((1+nrow(DAILY)):(tt+nrow(DAILY)),function(t) f1(t))

##p=?
p=mean(tail(DAILY$add.D,7)/tail(DAILY$add.C,7))

new.daily.D.fore<-map_dbl(n,function(x) rbinom(1,size=round(x),prob=p))

D.total.fore<-map_dbl(1:(nrow(DAILY)+tt),function(x) sum(head(append(DAILY$add.D,new.daily.D.fore),x)) )
fore.tt<-tail(D.total.fore,tt)
D.fore<-rbind(D.fore,fore.tt)

}

answer=colMeans(tail(D.fore,1250))
temp<-seq(from=as.Date('2020-08-01'),by='day',length.out = tt)
pred=tibble(date=temp,`prediction in total deaths`=answer)
pred
```

```
## # A tibble: 7 x 2
##   date       `prediction in total deaths`
##   <date>                            <dbl>
## 1 2020-08-01                      154964.
## 2 2020-08-02                      156110.
## 3 2020-08-03                      157236.
```

```
## 4 2020-08-04                    158332.
## 5 2020-08-05                    159412.
## 6 2020-08-06                    160498.
## 7 2020-08-07                    161601.
```

```
#(pred[1:7,]);(pred[8:14,]);(pred[15:21,]);(pred[22:28,])
```