Bayesian View of COVID-19 Estimation

Matthew Farant

5/11/2020

COVID-19 Test on Bogor's Commuterline

Commuterline, locally known as *Kereta Rel Listrik* (KRL), is one of the busiest transportation mode in Jabodetabek. It carries almost 141,330 passengers everyday, even in the midst of physical distancing "PSBB". On Monday (May 11, 2020), the district government of Bogor conducts a rapid and swab test to a sample of **200 KRL passengers**. According to the test, they found **1 person** to be tested positive of COVID-19. Roughly, we can estimate that for each 200 passengers, there will be 1 person tested positive (0.5% infection rate). However, estimating the infection rate by only dividing those two numbers might be impractical.

Bayes Theorem for Estimating the Infection Rate

The Bayes' theorem is stated mathematically as the following equation:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

With "H" as the hypothesis and "E" as the evidence, we can utilize the equation above to estimate the rate of infection in Bogor's KRL, given the test result on Monday. Our hypothesis is: The rate of infection will be the proportion of positively tested people and the total number of people available, denoted by r. The evidence shows that from 200 passengers tested, only 1 person tested positive for COVID-19. The likelihood of this evidence occured given the hypothesis is true will be expressed like a Binomial distribution PDF as below

$$P(E|H) = \binom{200}{1} r(1-r)^{199}$$

As we know, the conjugate prior of Binomial distribution is the Beta distribution. Hence, the posterior distribution will be Beta(2,200) which is representated as the following expression

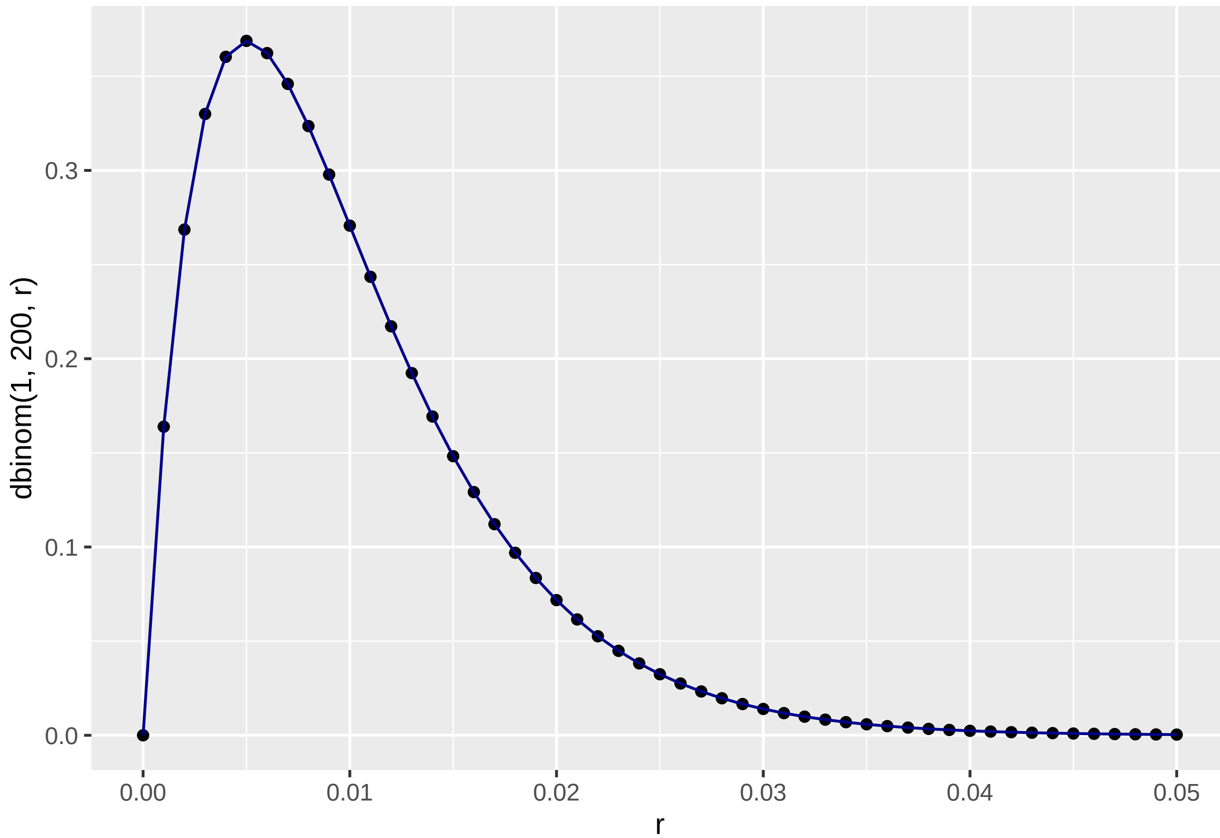
$$\frac{\Gamma(202)}{\Gamma(2)\Gamma(200)} r(1-r)^{199}$$

We can generate the plot of the Beta distribution by iterating each possible value of r using for loop or sapply() function.

```
#define function for P(E|H)
beta<-function(r){
  return(200*r*(1-r)^(199))
}
#possible values of r
r<-seq(from=0,to=5/100,by=0.1/100)
#for loop
p<-c()
for(i in 1:length(r)){
  p[i]<-print(beta(r[i]))
}
```

Alternatively, we can also use dbinom() function to generate the plot

```
library(ggplot2)
r<-seq(from=0,to=5/100,by=0.1/100)
#density plot
dat<-data.frame(r,dbinom(1,200,r))
plot<-ggplot(dat,aes(r,dbinom(1,200,r)))+
  geom_point()+
  geom_line(col='darkblue')
plot
```



With Beta distribution as a posterior PDF and 95% Confidence Interval, we can say that the interval of the rate of infection will be 0.00121 until 0.02741 (**0.121% until 2.741%**) by finding the upper and lower bounds that generates an area of 0.95 under the Beta distribution curve. Therefore, suppose there are 1000 people inside the KRL trains, there will be approximately **1 until 27 people** tested positive with COVID-19.

Conclusion

Besides the SIR model, which is a deterministic approach of estimating COVID-19 behaviour, we can also use Bayes' Theorem as a different approach to estimate the number of infected people. The main idea of using this method is taking account the randomness and uncertainty of a virus behaviour.