Computer Practical ‘Stochastic models of infectious disease dynamics in small populations’

0. Introduction

In a stochastic model the number of individuals in each event (like infection or recovery) can only be an integer number in each small time step, based on a chance process. Every run with the model yields one particular outbreak in a population, and the model needs to be run many times to get information on averages, or on variation between runs. To simulate the number of infection events, we use the probability for each susceptible to make at least one contact with an infectious individual in each time step. For each susceptible, the mean number of contacts with infectious individuals is equal to $(\beta Y/N) * \Delta t$. By assuming that contacts occur by a Poisson point process (i.e. “random”), this means that the probability that no contacts are made is equal to exp($(-\beta Y/N) * \Delta t$). Then, in each time step, the number of new infections is binomially distributed as:

new infections $bin(n = X, p = 1\breve{}exp(-(\beta Y/N) * \Delta t))$

Similarly, the number of new recoveries is binomially distributed as:

new recoveries $bin(n = Y, p = 1\breve{}exp(-\gamma * \Delta t))$

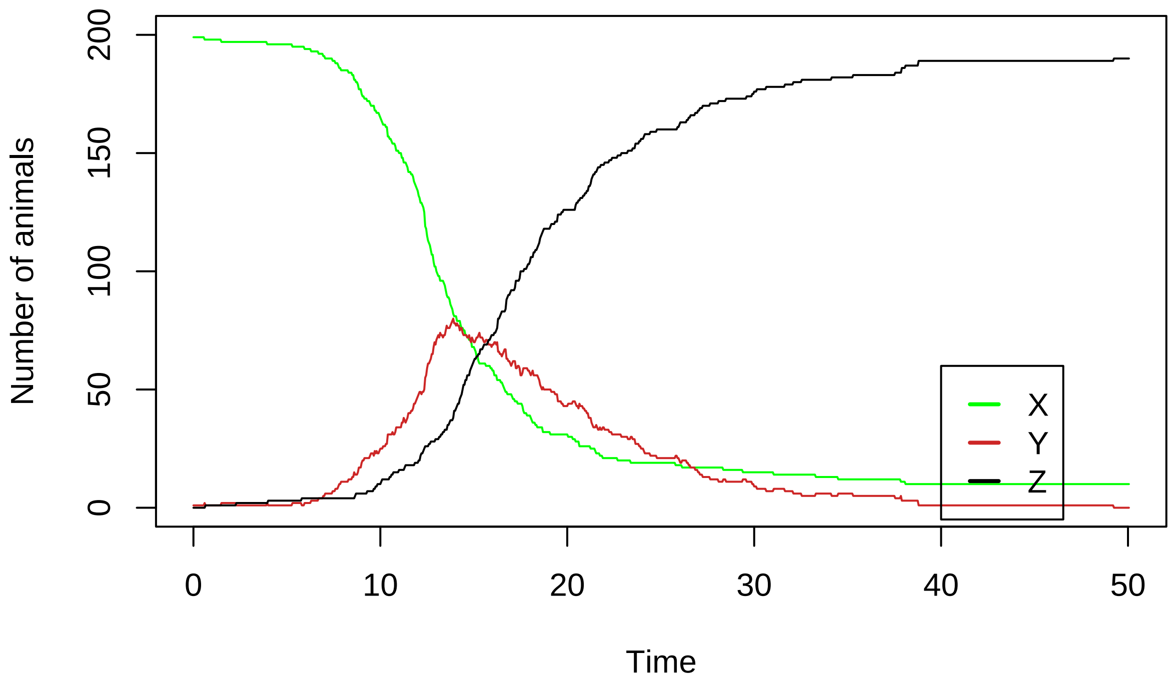
1. Stochastic SIR model.

```
#####simulation model#######
sir.sim <- function(N0, Y0, Z0, beta, gamma, deltat, duration){
  # start at time = 0
  time = 0

  #set initial population variables
  n = n0
  Y = Y0
  Z = Z0
  X  = N0 - Y0 - Z0

  #Create a data.frame to hold the simulated data
  res = data.frame(Time = c(0),X = c(X),Y = c(Y), Z = c(Z))

  #loop as long as the length of the simulation has not been filled
  while(time < duration)
  {
    ##simulate the number of new cases, recoveries and deaths during this time step
    #Each time check if the variabel is not 0 because rbinom cannot draw from a population of size 0
    ifelse (X * Y > 0, cases <- rbinom(1, X,  1 -  exp(- beta * (Y / n) * deltat)),cases <- 0)
    ifelse (Y > 0 , recoveries <- rbinom(1, Y, 1 -  exp(- gamma * deltat)),recoveries <- 0)

    #subtract/add from current values
    X =  X - cases
    Y = Y + cases - recoveries
    Z = Z + recoveries

    #advance one time step
    time <- time + deltat

    #store every thing
    res = rbind(res, data.frame(Time  = c(time),X = c(X),Y = c(Y), Z= c(Z)))
  }
  return(res)
}
```

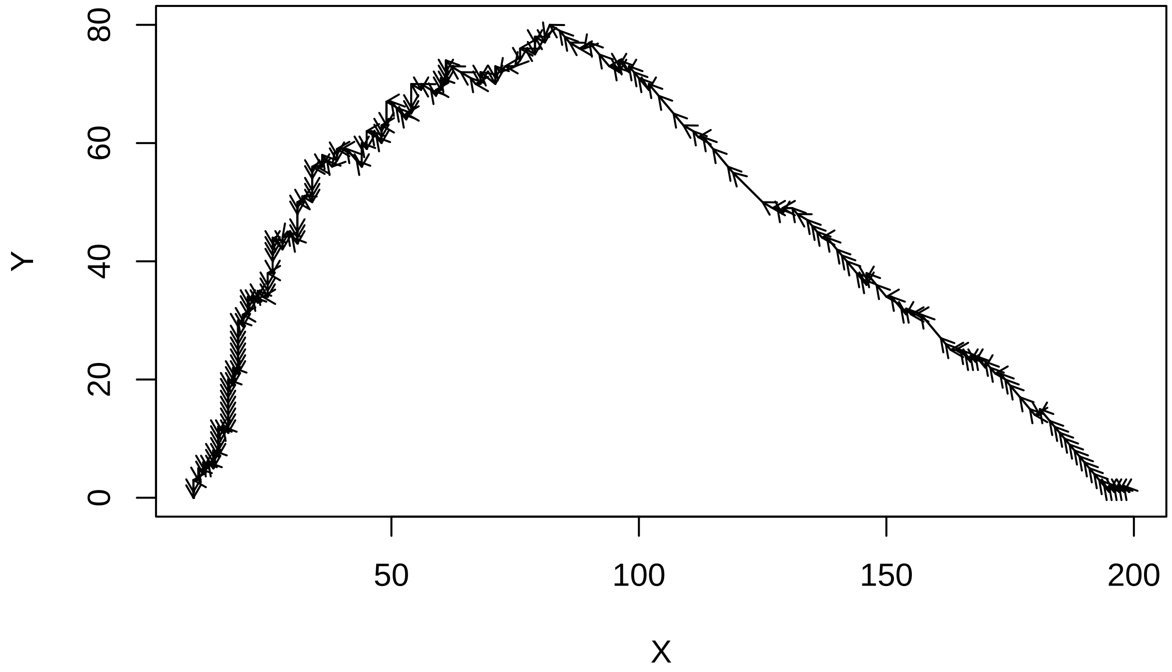
- Do you understand this code?

(a) Do one simulation‘below. The code will produce 2 figures: epidemic curve and X vs Y:

```
####parameters#####
beta = 0.6
gamma = 0.2


####simulation settings####
deltat = 0.05
simlength = 50

#####initial values####
n0 = 200
Y0 = 1
Z0 = 0
X0 = n0 - Y0 - Z0

sim <- sir.sim(n0, Y0,Z0, beta, gamma, deltat, simlength)
plot(sim$Time, sim$X, type = "l", xlab = "Time", ylab = "Number of animals",col = "green",ylim = c(0,n0)
lines(sim$Time, sim$Y, type = "l", xlab = "Time", ylab = "Number of animals",col = "firebrick3")
lines(sim$Time, sim$Z, type = "l", xlab = "Time", ylab = "Number of animals",col = "black")
legend(40,60,c("X","Y","Z"),col = c("green","firebrick3", "black"),lwd = 2,seg.len = 1)
```



```
plot(sim$X, sim$Y, type = "l", xlab = "X", ylab = "Y")
s <- seq(length(sim$X)-1)
suppressWarnings(arrows(x0 = sim$X[s], y0 = sim$Y[s], x1 = sim$X[s + 1], y1 = sim$Y[s + 1],length = .075
```



(b) Click Run the simulation a couple of times. Describe the differences between the different runs. How long do the outbreaks take? What is the peak number of infected people? How many are still susceptible at the end? How does this relate to what we saw in the deterministic model?

Because in the stochastic model every simulation is different, there is variation in characteristics such as output duration, peak prevalence, or final size. It can therefore be informative to do many re-runs of the model, and store the output parameter of interest.

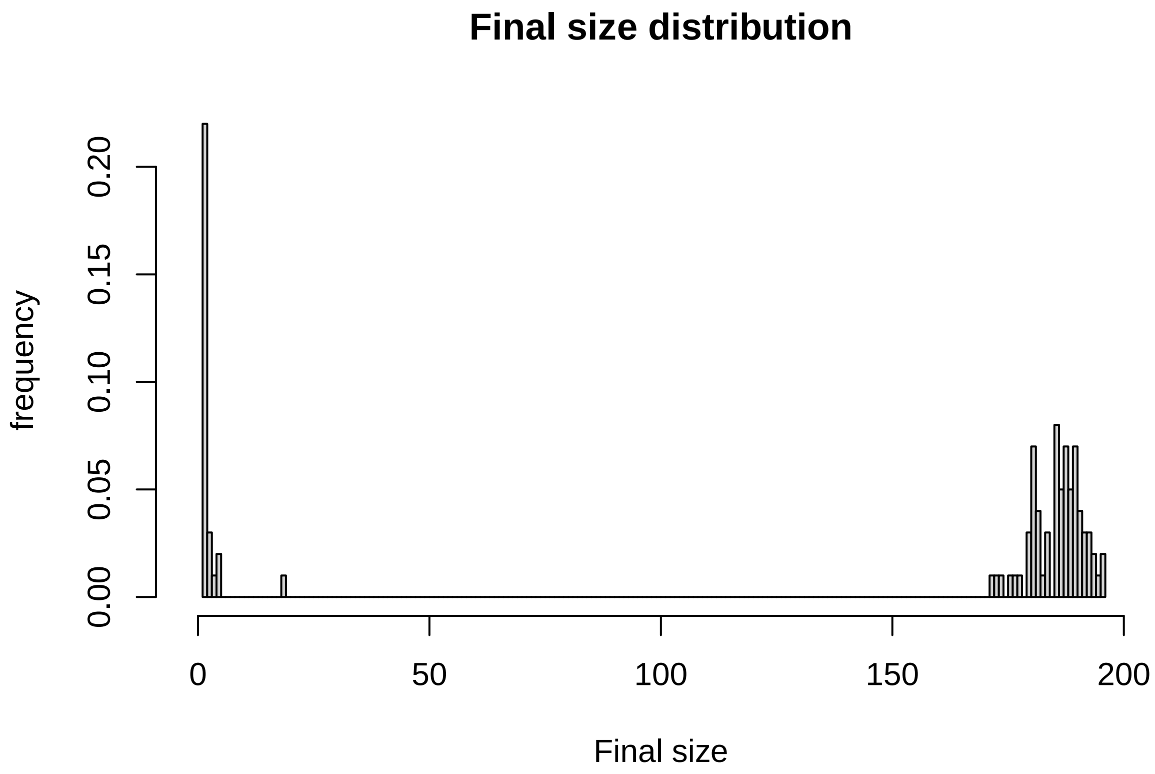
```
#####function for multiple runs#####
sir.sim.multiple <- function(n0, i0, r0, beta, gamma, deltat, duration, repeats)
{
  res <- data.frame(Time = c(),X =c(), Y =c(), Z=c())
  r <-  1
  while(r <= repeats)
  {
    sim <- sir.sim(n0, i0, r0, beta, gamma, deltat, simlength)
    res <- rbind(res, cbind(sim, sim = rep(r, length(sim$Time))))
    r = r + 1
  }
  return(res)
}
```

- Run the model multiple times:

```
repeats <- 100
multi.sim <- sir.sim.multiple(n0, Y0, Z0, beta, gamma, deltat, duration, repeats)
```

- Plot a histogram of the final size of the simulations by selecting the number of recovered individuals for Y = 0.

```
##plot the end of the epidemic in a histogram
hist(multi.sim$Z[floor(multi.sim$Time) == simlength],breaks = n0, xlim = c(0-1,n0+1), xlab = "Final siz
```



(c) What do you see? Calculate the mean final size. Why is the mean result of 100 runs with the stochastic model not equal to the result of the deterministic model? Estimate the probability of a major outbreak from the 100 model runs.

(d) Change the population size in BM to N = 100, and repeat the analyses of the final sizes. Does the probability of a major outbreak change if the population size is smaller? Now change R0 and repeat it again. How does that affect the final size distribution?

4. Stochastic model with birth/immigration and death/emigration

Add birth and death to the model by changing this code. Also add three extra variable DX, DY and DZ to count the number of dead susceptibles, infected and recovered individuals. To keep the population constant the number of births should equal DX + DY +DZ.

```
sir.sim.demo <- function(N0, Y0, Z0, beta, gamma, mu, deltat, duration){
  # start at time = 0
  time = 0

  #set initial population variables
  n = n0
  Y = Y0
  Z = Z0
  X  = N0 - Y0 - Z0
  #Create a data.frame to hold the simulated data
  res = data.frame(Time = c(0),X = c(X),Y = c(Y), Z = c(Z))

  #loop as long as the length of the simulation has not been filled
  while(time < duration)
  {
    ##simulate the number of new cases, recoveries and deaths during this time step
    #Each time check if the variabel is not 0 because rbinom cannot draw from a population of size 0
    ifelse (X * Y > 0, cases <- rbinom(1, X,  1 -  exp(- beta * (Y / n) * deltat)),cases <- 0)
    ifelse (Y > 0 , recoveries <- rbinom(1, Y, 1 -  exp(- gamma * deltat)),recoveries <- 0)

    #subtract/add from current values
    X =  X - cases
    Y = Y + cases - recoveries
    Z = Z + recoveries

    #advance one time step
    time <- time + deltat

    #store every thing
    res = rbind(res, data.frame(Time  = c(time),X = c(X),Y = c(Y), Z= c(Z)))
  }
  return(res)
}
```

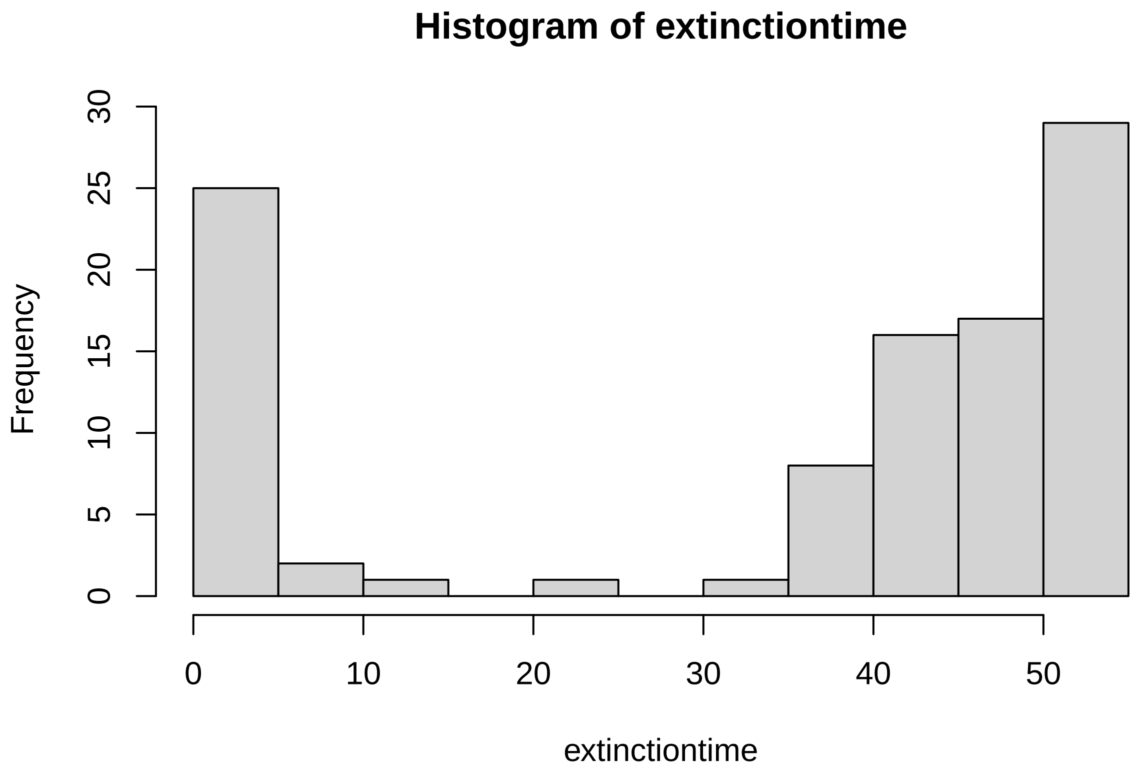
(a) Run the model, customize the graph, make a graph of Y against X, as in exercise 3a.

(b) Look at the graph of X, Y, and Z in time, and the graph of Y against X (bottom). Run the model a couple of times. How do the graphs differ from the deterministic model? We will now compare the long-term dynamics related to the initial conditions. Two initial conditions will be considered: (1) starting with a susceptible population and one infected individual, and (2) starting in the endemic steady state which will be determined from the deterministic model.

(c) For the second initial condition we need the steady state of the deterministic model.Calculate values of the steady state of the deterministic model.

We will compare the long-term dynamics by two output parameters: the number of outbreaks going extinct, and the distribution of extinction times over many simulations. Therefore, we need the extinction time as output from the simulation.

```
##
sir.sim.demo.multiple <- function(n0, i0, r0, beta, gamma,mu, deltat, duration, repeats)
{
  res <- data.frame(Time = c(),X =c(), Y =c(), Z=c())
  r <-  1
  while(r <= repeats)
  {
    sim <- sir.sim.demo(n0, i0, r0, beta, gamma, mu, deltat, simlength)
    res <- rbind(res, cbind(sim, sim = rep(r, length(sim$Time))))
    r = r + 1
  }
  return(res)
}
##plot the end of the epidemic (Y == 0)
select<- function(r){
  #no extinction
  if(length(multi.sim$Time[multi.sim$sim==r & multi.sim$Y ==0])==0) return(simlength+1)
  return(min(multi.sim$Time[multi.sim$sim==r & multi.sim$Y ==0]))

  }
extinctiontime= sapply(FUN = select, c(1:repeats))

hist(extinctiontime)
```



(d) Start simulations with a susceptible population and one infected. Make a histogram of the extinction times. Calculate in what proportion of simulations the outbreak got extinct.

(e) Repeat (c) and (d) with simulations starting in the endemic steady state. Describe the differences in extinction behaviour between the two initial condition. How do explain the differences?

ADDITIONAL QUESTIONS

(g) Make the population size smaller (e.g. N = 20) and larger (e.g. N = 1000) and repeat (c) and (d) for the endemic equilibrium as starting condition. How do the results differ? How do you explain this? What will you expect for very large populations?