

Computational Biology, Problem Set 3

Problem 1, Stochastic dynamics in large but finite populations

Audrey Ely, Konstantinos Zakkas

March 13, 2021

a) Deterministic model

The infection sustains ad infinitum if the population size of infectives I reaches a stable steady state. First, let us find the steady states of the system by introducing $N = S + I$. We assume that N is constant.

The first steady state is for $I = 0$ and $S = N$, where there is no infection. Now, we assume $I \neq 0$ to find the second steady state :

$$\begin{aligned}\frac{dI}{dt} &= 0 \\ \Leftrightarrow \frac{\alpha}{S+I}SI - \beta I &= 0 \\ \Leftrightarrow \frac{\alpha}{N}(N-I)I - \beta I &= 0 \\ \Leftrightarrow \alpha(1 - \frac{I}{N}) - \beta &= 0 \\ \Leftrightarrow I = N(1 - \frac{1}{r_0})\end{aligned}$$

where r_0 is the reproductive rate of the disease : $r_0 = \frac{\alpha}{\beta}$. Thus, the steady state is reached for $I^* = N(1 - \frac{1}{r_0})$ and in that case $S = \frac{N}{r_0}$. Note : $I^* > 0 \Leftrightarrow r_0 > 1$: there exists a non-negative and non-null steady state for $\alpha > \beta$.

To know under which conditions the steady state is stable, we introduce a small perturbation η around the fixed point I^* . Near this point, we approximate the dynamics with $I = I^* + \eta$. Let f be the function defined such as $f(I) = \frac{dI}{dt}$. We

can then approximate :

$$\begin{aligned}
f(I^* + \eta) &\approx \eta f'(I^*) \\
f'(I) &= \frac{-2\alpha}{N}I + \alpha - \beta \\
\Rightarrow f'(I^*) &= \frac{-2\alpha}{N}N(1 - \frac{\beta}{\alpha}) + \alpha - \beta \\
\Rightarrow f'(I^*) &= \beta - \alpha
\end{aligned}$$

Then, the steady state I^* is stable if $\beta < \alpha$, and unstable otherwise.

b) Stochastic model for finite population size

In this part, we consider the assumptions given in the instructions. $\rho_n(t)$ is the probability to observe n infected individuals at time t in a finite population of N individuals. In a small time interval δt , changes in ρ_n happen :

- Due to infection : $(b_{n-1}\rho_{n-1} - b_n\rho_n)\delta t$
- Due to recovery : $(d_{n+1}\rho_{n+1} - d_n\rho_n)\delta t$

From there, we can deduce :

$$\begin{aligned}
\rho_n(t + \delta t) &= \rho_n(t) + \delta t(b_{n-1}\rho_{n-1}(t) + d_{n+1}\rho_{n+1}(t) - (b_n + d_n)\rho_n(t)) \\
\Rightarrow \frac{d\rho_n(t)}{dt} &= b_{n-1}\rho_{n-1}(t) + d_{n+1}\rho_{n+1}(t) - (b_n + d_n)\rho_n(t)
\end{aligned}$$

The last line is the Master equation.

In the deterministic model, it is possible that the infection sustain ad infinitum, whereas in the stochastic model, it is not possible. Indeed, there is a quasi-steady state, that disappear in the long run due to stochastic fluctuations.

c) Towards efficient simulation of the stochastic model

In order to find the probability distributions of t_b and t_d , we will take small steps dt . At each time step, we consider a random number between 0 and 1 : if this number is under the threshold of b_ndt (resp. d_ndt), this means that an event of infection (resp. recovery) occurs, and then we can register the value of t_b (resp. t_d). This simulation is run many times to get good statistics. The corresponding R code is at the end of the document.

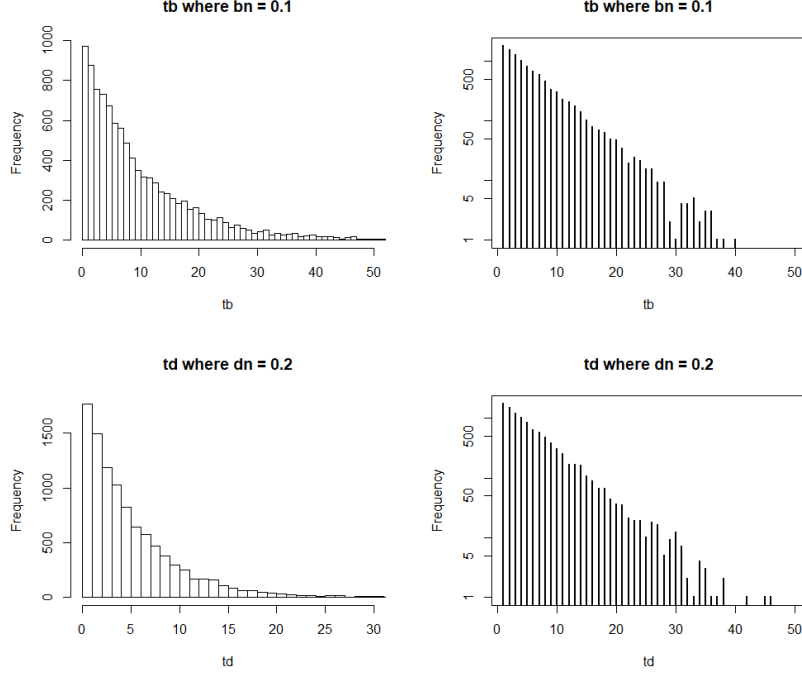


Figure 1: Histograms of t_b and t_d for the first set of values.

- For the set of values $b_n = 0.1, d_n = 0.2$, we get the following means : $\langle t_b \rangle \approx 9.887$ and $\langle t_d \rangle \approx 4.980$. The histograms we get for t_b and t_d , with a log-y-axis or not, are shown in figure 1. This figure reveals that the logarithm of the frequencies of t_b (resp. t_d) is inversely proportional to the value of t_b (resp. t_d) in time unit.
- For the set of values $b_n = 1, d_n = 2$, we get the following means : $\langle t_b \rangle \approx 0.994$ and $\langle t_d \rangle \approx 0.504$. The histograms show a trend similar to the first case.
- For the set of values $b_n = 10, d_n = 5$, we get the following means : $\langle t_b \rangle \approx 0.101$ and $\langle t_d \rangle \approx 0.200$. The histograms show a trend similar to the first case.

For the three sets of values, we can observe that the mean of time t_b (resp. t_d) is approximately equal to the inverse of b_n (resp. d_n).

d) Population distribution at different times in the stochastic model

R Code

```
# Here is the code for the first set of values for 1c).
# The other sets only require to change the values of bn and dn.

bn = 0.1
dn = 0.2

dt = 0.1

data.tb=c()
data.td=c()

for (i in 1:10^4){
  tb = 0
  td = 0
  t = 0
  while (tb==0 | td==0) {
    t = t+dt      # Recording of the time step
    nb = runif(1) # Random number between 0 and 1
    if (tb==0 & nb < bn*dt){
      tb = t
      data.tb = c(data.tb, c(tb))
    }
    if (td==0 & nb < dn*dt){
      td = t
      data.td = c(data.td, c(td))
    }
  }
}

mean(data.tb)
mean(data.td)

# Histogram with regular axis
hist(data.td, xlim=c(0,30), breaks=seq(0,100,by=1), main = "td where dn = 0.2",
      xlab = "td")

# Histogram with log axis
hist.data = hist(data.td, plot=F, xlim=c(0,30), breaks=seq(0,50,by=1))
plot(hist.data$count, log="y", type='h', lwd=2, lend=2, main = "td where dn = 0.2",
      xlab = "td", ylab = "Frequency")
```

```
# Histogram with regular axis
hist(data.tb, xlim=c(0,50), breaks=seq(0,100,by=1), main = "tb where bn = 0.1",
      xlab = "tb")

# Histogram with log axis
hist.data = hist(data.tb, plot=F, xlim=c(0,50), breaks=seq(0,60,by=1))
plot(hist.data$count, log="y", type='h', lwd=2, lend=2, main = "tb where bn = 0.1",
      xlab = "tb", ylab = "Frequency")
```