Test of id_spatial_sim

Steven Riley

October 8, 2014

1 Playing around with the ebola library

When this is not being run in bacth mode, we need to set the working directoy. This appears here as a comment, so it is not run when in batch mode.

```
# setwd("~/Dropbox/git/id_spatial_sim/testsuite")
```

First we declare some packages that we will use below.

```
require("raster")

## Loading required package:  raster
## Loading required package:  methods
## Loading required package:  sp
```

The first line of the batch file builds a synthetic population with density proportional the ebola affected region in west Africa, but much smaller. With a total population of only 100,000. Each person has, one average, 10 network links but these links are distributed entirely randomly in space. This takes a while because the average population is very low and there is high variability. Hence the accept-reject method for assinging nodes has many rejection steps. We assume that only one individual lives in a household for this population.

The second line of the batch file runs an outbreak of only two generations 20 times. The outbreak is seeded in the same area as the reported patient zero for the 2014 Ebola outbreak. There are 4 iniitally infectious individuals at time $t = 0$. Transmission is only via the spatial kernel and thus allows us to test that the basic reproductive number is parameterized correctly. We can also report the serial interval.

We first load the linelist of events from all the realizations. And check the dimensions of the output. The output was designed before csvs became so dominant!

```
dat0 <- read.table(file="./output/ft_sp_pset_0_Events.out",header=TRUE)
dimDat0 <- dim(dat0)
noevents <- dimDat0[1]
nocols <- dimDat0[2]
```

The column headings describe the information captured in the event file

```
names(dat0)

##  [1] "Run"        "Day"        "Event"      "Index"      "X"
##  [6] "Y"          "Generation" "infector"   "infect_x"   "infect_y"
```

We can subset these 'data' to look at only infection. Then examine the number of infections by generation for each realization.

```
tabInfs0 <- dat0[dat0$Event==0,]
table(tabInfs0$Run,tabInfs0$Generation)

##
##      1  2
##   0  4  5
##   1  4  0
##   2  4  5
##   3  4 12
##   4  4  8
##   5  4  8
##   6  4 11
```

```
##   7  4 14
##   8  4 11
##   9  4  2
```

And look at the average number in the second generation divided by the number of seeds as a test of the R_0 parameterization.

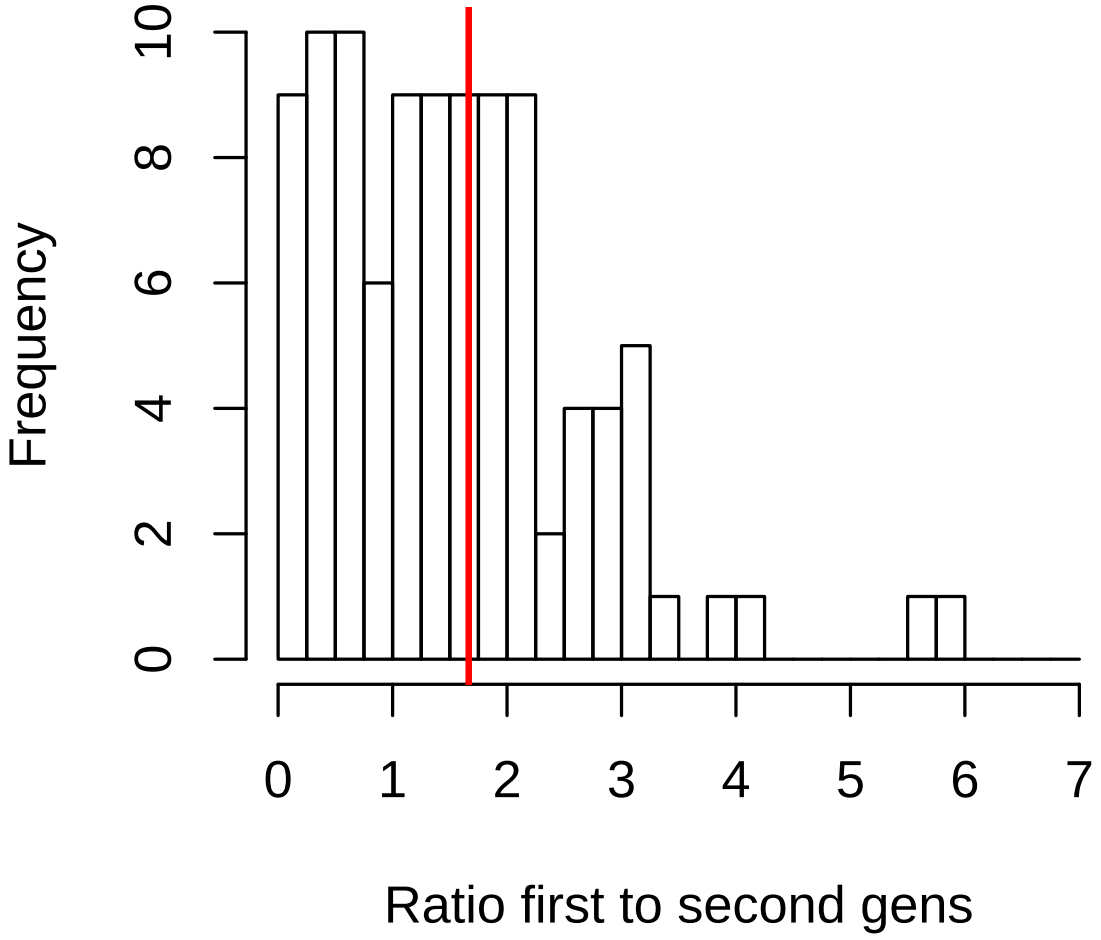
```
table(tabInfs0$Generation)[2]/table(tabInfs0$Generation)[1]

##   2
## 1.9
```

Its difficult to tell if this is accurate with such small numbers, so we can load up the similar run with 100 realisations.

```
dat1 <- read.table(file="./output/ft_sp_pset_1_Events.out",header=TRUE)
dim(dat1)

## [1] 4264   10

tabInfs1 <- dat1[dat1$Event==0,]
estR0 <- table(tabInfs1$Generation)[2] / table(tabInfs1$Generation)[1]
estR0

##     2
## 1.665
```

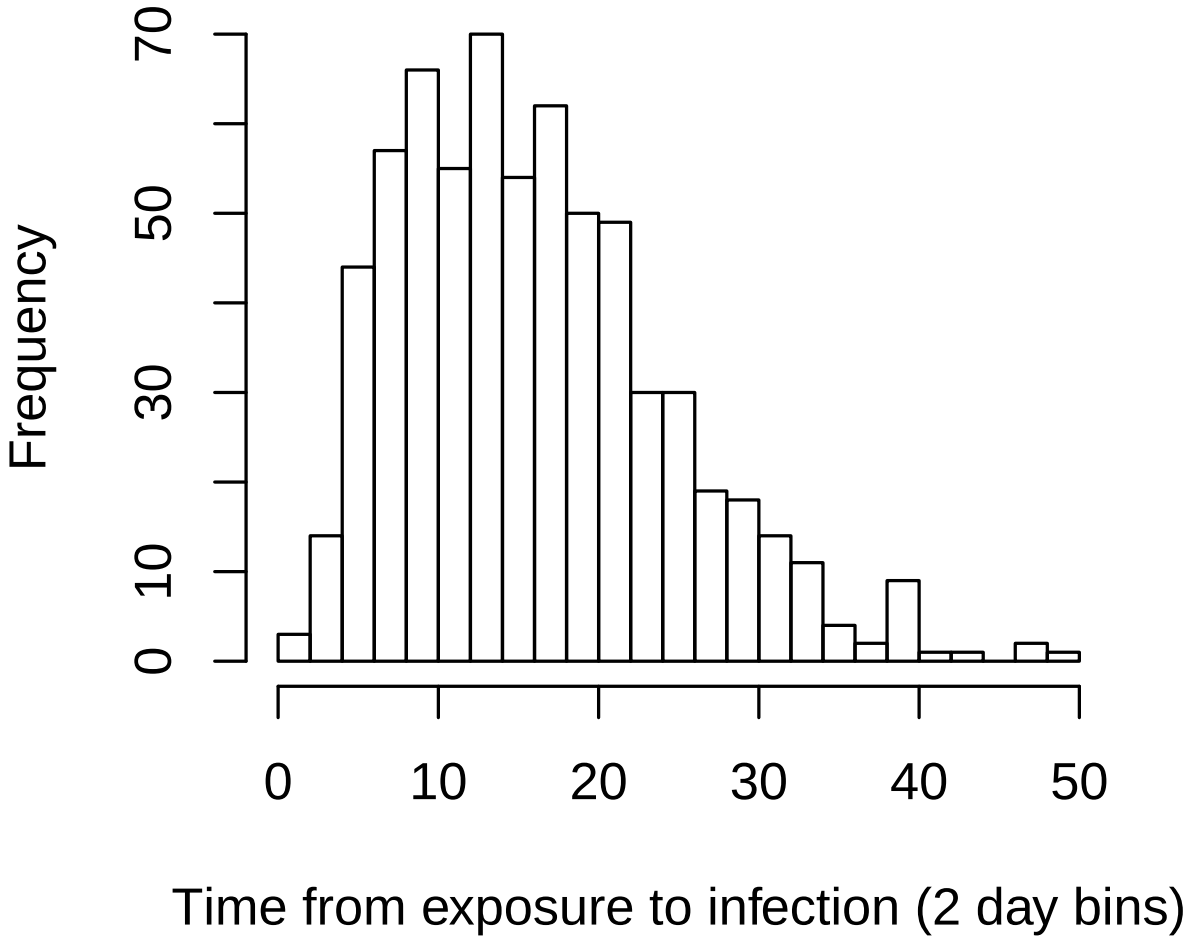
Might also be worth looking at the distribution of the ratio of secondary cases for each realization. So we make a table of generations by run and plot a histogram of the ratios for each realization. You can see quite a bit of variance in the size of the second generation of this model. Note that the offspring distribution will be even more highly over-dispersed because these results are based on a seed of 4.

```
tabRunGen1 <- table(tabInfs1$Run,tabInfs1$Generation)
hist( tabRunGen1[,2]/tabRunGen1[,1],breaks=seq(0,7,0.25),
      xlab="Ratio first to second gens",main="")
abline(v=estR0,lwd=2,col="red")
```



It is also straightforward to look at the distributions of different waiting times in the model because we 'observe' them directly in this idealized linelist. So the average serial interval is equal to the average time of the infection event in the second generation.

```
vecTimesG1 <- tabInfs1[tabInfs1$Generation==2,"Day"]
mean(vecTimesG1)

## [1] 16.51

hist(vecTimesG1,breaks=seq(0,50,2),main="",
    xlab="Time from exposure to infection (2 day bins)")
```



Its a highly over-dospersed distribution, suggesting that events such as the long time from exposure to infection for the non-African infection event in Spain are not entirely inconsistent with the NEJM estimated parameters.

1.1 Spatial analysis

Next we load up the population desity on which the model was based and look at the spatial distribution of these initial infections.