

ST433/533 Applied Spatial Statistics

Lab activity for 11/11/2020

A. CLARIFICATION QUESTIONS

(1) I am really interested in applications of the Strauss process. Are we able to use this process to model any system with some degree of repulsion between their interacting particles? Would it be possible (or useful) to model a non-stationary process as a mixture of a Strauss process and a Poisson cluster process since one would determine inhibition and one would determine attraction? (or am I mixing up definitions) If so, what would the applications of this be?

I'm sure there are other approaches, but one idea might be a cluster process with a Strauss process for the cluster centers. This model would have both the cluster and Strauss processes as special cases, and in general would be a mix between the two.

(2) For IPP, I am confused about the method of using the log Gaussian Cox process. Since IPP is a Poisson process, how do we assume $\log(\lambda)$ to be a Gaussian process?

An IPP is usually defined for a fixed intensity function, $\lambda(s)$. However, in reality, you don't know the intensity function and you need some assumptions or a model for it, and the Cox process picks a Gaussian process for the log intensity.

(3) I would like to see more about point processes and their real utility/applications, is any good resources you recommend?

The class textbook is great. Also, the "Handbook of Spatial Statistics" by Gelfand et al has a section on point patterns (Part IV) and this book is now available online.

(4) Are these methods only to create points and test their spatial distribution? or do they have other purposes?

Well, the methods are designed to be used on real data. Common objectives are to test for covariate effects, hotspots or clustering, and to make predictions about the locations of future events.

(5) Generally speaking, is it easier to determine if a spatial data is clustering than under repulsion? Is inhibition spatial data more difficult to determine?

It's definitely easier to build a model for clustering than inhibition. It's probably easier to visually detect clustering, but I'm not sure about the formal tests.

(6) For the Poisson process model, the number of samples in subdomain B has a Poisson distribution with λ equal to the volume under the intensity function within subdomain B. Does the property of the Poisson distribution that the expected value and the variance are equal impose any additional assumptions to the Poisson process model?

Yes, the overall number of points is $\text{Poisson}(\lambda_{\text{integrated}})$, so the mean and variance are both $\lambda_{\text{integrated}}$.

B. BRIAN'S DISCUSSION QUESTIONS

In today's lab, we will discuss a spatiotemporal extension of the SIR model for the spread of an infectious disease. The non-spatial model is given here

https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

Recall, the SIR model classifies people as either susceptible (S) to the disease, infected (I) by the disease (and thus contagious) and recovered (R) from the disease (and thus not contagious or susceptible). In a closed population of N people, the spread of the disease is governed by the equations

$$S_t = S_{t-1} - \beta S_{t-1} I_{t-1} / N$$

$$I_t = I_{t-1} + \beta S_{t-1} I_{t-1} / N - \gamma I_{t-1}$$

$$R_t = R_{t-1} + \gamma I_{t-1}$$

In this model, $S_{t-1}I_{t-1}$ is the number of potential interactions between a susceptible and an infected people, $\beta > 0$ controls the spread of infection due to these interactions, and $\gamma > 0$ controls the rate by which infected people recover. You can plot the SIR curve at <https://shiny.stat.ncsu.edu/bjreich/SIR/>.

Today we will consider a spatial extension. We will use a fairly simple model; here is a more complicated example: <https://www.nature.com/articles/s41586-020-2923-3> (NYT article). Let S_{jt} , I_{jt} and R_{jt} be the counts for region j (say county j) and time t (say t days since the outbreak). For simplicity, we'll assume all regions have population N. This extension allows for both interactions in the same region, but also in different regions via the additional term in blue below,

$$S_{jt} = S_{jt-1} - \beta_j S_{jt-1} I_{jt-1} / N - \phi \sum_{k=1}^n w_{jk} S_{jt-1} I_{kt-1} / N$$

$$I_{jt} = I_{jt-1} + \beta_j S_{jt-1} I_{jt-1} / N + \phi \sum_{k=1}^n w_{jk} S_{jt-1} I_{kt-1} / N - \gamma I_{jt-1}$$

$$R_{jt} = R_{jt-1} + \gamma I_{jt-1}$$

In this expression, $S_{jt-1}I_{kt-1}$ is the number of potential contacts between susceptible people in region j and infected people in region k, and the number of infections that results from these potential encounters is determined by the parameter $\phi > 0$ and the weights w_{jk} . For simplicity, below we take $w_{jk}=1$ if regions j and k are adjacent and $w_{jk}=0$ otherwise (with $w_{jj}=0$) so that for a given time point the disease can only spread to neighboring regions. In this example, the local infection rate β_j is allowed to vary by region, but the γ and ϕ are constant for all regions.

Code to simulate the spread of the disease is given on the next page. This uses a 20x20 grid of 400 regions, $N=10,000$ and 100 time steps. It begins with one infected person in region 50 and sets $\beta_j = 0.5$, $\gamma = 0.5$ and $\phi = 0.1$. Refer to this as the baseline case. Using this code, answer the following questions.

(1) What are the most important assumptions being made by this model?

- Recovered people can't be infected again
- Disease can only spread to neighbors
- No deaths

- Square lattice

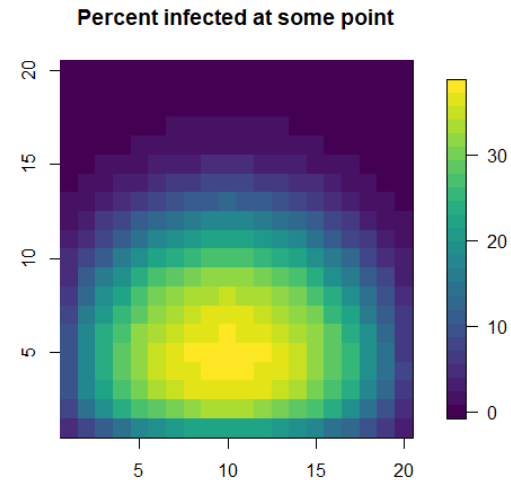
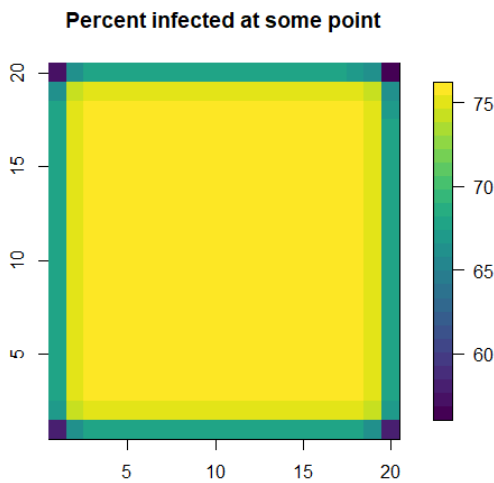
- Many others

(2) Starting with the baseline case, say the local spread is reduced by 50%, i.e., β_j is cut in half for all regions. What interventions might accomplish this, and what effect does this reduction have on the spread of the disease?

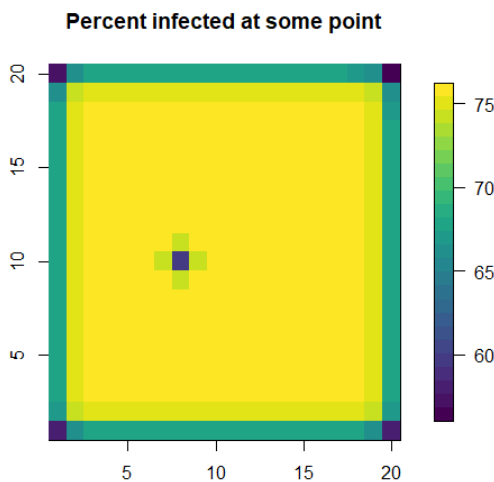
- Universal mask mandate. Reduced the proportion infected from 0.74 to 0.14.

Baseline

50% reduction in all beta



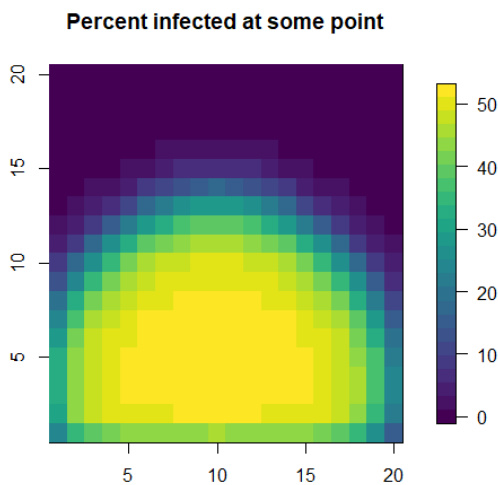
(3) Start with the baseline case except that region 188 is able to cut β_{188} in half. What effect does this have on the spread of the disease? How do these results compare to (2) and what conclusions can be drawn from this comparison?



- The percent infected in region 188, and to some extent its neighbors, is reduced.
- An individual region can somewhat control its own spread, but not much of a reduction as when all regions had the mandate.

(4) Start with the baseline case, but you are able to cut the spatial parameter ϕ in half. What interventions might do this, and what would be the effect on the spread of the disease? Is it better to focus on cutting all the β_j by 50% or cutting ϕ by 50%?

- Social distancing, stay at home orders, airline/bus restrictions, build the wall 😊
- The overall reduction was more for the scenario (2) than scenario (4) at least for 100 iterations.



(5) Say at time $t=1$, all regions are 10% infected and 0% recovered, i.e., $I_{j1}=N/10$ and γ and ϕ are as in the baseline case. We have a vaccine that can be administered at time $t=1$ and reduce β_j , how effective must it be to control the spread so no region has more than 30% of its population infected?

We must have $\beta_j = 0.04$.

(6) Starting with the baseline case, how would you estimate the different effects of starting a mask mandate that reduces β_j by half for all regions at day 10 versus day 50? Code it up and see how this affects the spread of the disease.

BR: You would need to allow beta to change with both space and time. In the code you could multiple beta by 0.5 when you get to time step 10 or 50.

(7) How would you estimate the model parameters? What data would you need/desire?

BR: You can fit the true curve with observations as in the Bayesian lecture (GUI). You might do a separate analysis by region to estimate the beta's and then another analysis to estimate the other parameters.

(8) How would you test that the model fits well?

BR: You could fit the model for time 1 to T and then predict for time T to 100.

(9) The spatial model with binary w_{jk} is probably too simple, what might be a more realistic model?

BR: Commuting/mobility data, or maybe an exponential decay with distance between centroids.

(10) The model is deterministic, how would you make it stochastic?

BR: You could assume the observed I was Poisson with mean given by I in the SIR equation.

```

##### R code to simulate from the spatial SIR model #####
rm(list=ls())

library(fields)
library(viridis)

# The main simulation function
Spatial_SIR <- function(S1,I1,N,beta,gamma,phi,W,nt){
  ns <- length(S1)
  S <- I <- R <- matrix(0,ns,nt)
  S[,1] <- S1
  I[,1] <- I1
  R[,1] <- N-S1-I1
  for(t in 2:nt){
    newI <- beta*S[,t-1]*I[,t-1]/N + phi*S[,t-1]*(W%*%I[,t-1])/N
    newR <- gamma*I[,t-1]
    S[,t] <- S[,t-1] - newI
    I[,t] <- I[,t-1] + newI - newR
    R[,t] <- R[,t-1] + newR
  }
  output <- list(S=S,I=I,R=R)
  return(output)}

# A plotting function
plot_grid <- function(x,main=""){
  m <- sqrt(length(x))
  image.plot(1:m,1:m,matrix(x,m,m),col=viridis(25),
    xlab="",ylab="",main=main)
}

m <- 20      # Generate data on an m x m grid
ns <- m^2     # Number of regions
nt <- 100     # Number of time steps
N <- 10000    # Population of each region

s <- expand.grid(1:m,1:m)  # Neighbors get w[j,k]=1
W <- as.matrix(dist(s))==1

beta <- rep(0.5,ns) # The SIR infection parameter
gamma <- 0.5        # The SIR recovery parameter
phi <- 0.1          # The spatial parameter
I1 <- rep(0,ns)     # The number of infected in each region at time 1
I1[50] <- 1
S1 <- N-I1

plot_grid(I1,"Number of infected in at time 1")
plot_grid(W[50,],"Neighbors of region 50")

# Simulate the process

sim <- Spatial_SIR(S1,I1,N,beta,gamma,phi,W,nt)

# Plot the infection curve for each region
matplot(t(sim$I),type="l",lty=1,
  xlab="Time since outbreak",ylab="Number of infected")
legend("topleft",c("Region 1","Region 2", "Etc"),lty=1,col=1:3,bty="n",cex=0.7)

# Plot the number of infected in each time point in each region
for(t in 1:nt){
  plot_grid(sim$I[,t],main=paste("Number of infected at time",t,"of",nt))
  Sys.sleep(0.1) # Pause
}

# Plot the percentage of people that were infected at some point
plot_grid(100-100*sim$S[,nt]/N,main=paste("Percent infected at some point"))

```