Homework 4

Stat 597a: Spatial Models

Claire Kelling

Due Nov 14, 2017

Problem 1a:

Load the data and create two ppp objects from it, one for affected and one for unaffected individuals. A key question is: what is the window? It is not available for this data. For now, take the window to be the same for each ppp object: use a rectangular region based on the range of x and y for both datasets. (Have a look at the help file for owin.)

I created these two ppp objects, please find the code in the appendix.

Problem 1b:

For each dataset separately, create Monte Carlo simulation envelopes for the F and G functions and plot them. Clearly label your plots and turn in a few sentences describing any choices you made in creating the envelopes. Is their evidence against CSR in this dataset? If so, what type of violation is suggested?

I have included my plots below for the F and G functions for both datasets.

Choices I made in creating the envelopes?

I don't see too much evidence against CSR in my dataset. For the most part, the observed curve lies within the simulation envelope for both that affected and the unaffected datasets, for both the F and G functions. However, for the unaffected dataset and the F curve, the observed curve is not very close to the mean simulated plot and lies outside the envelope on the right hand side of the plot.
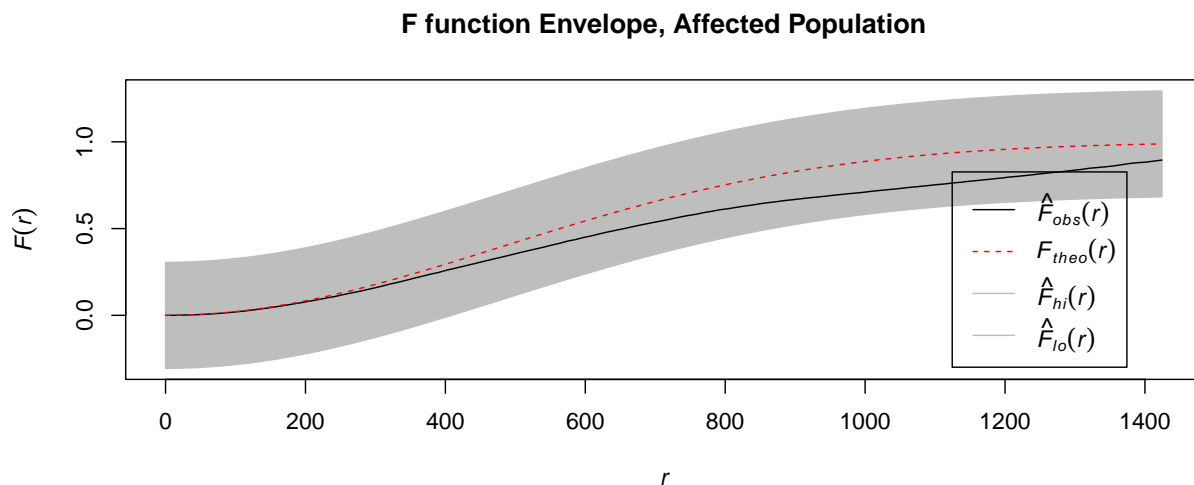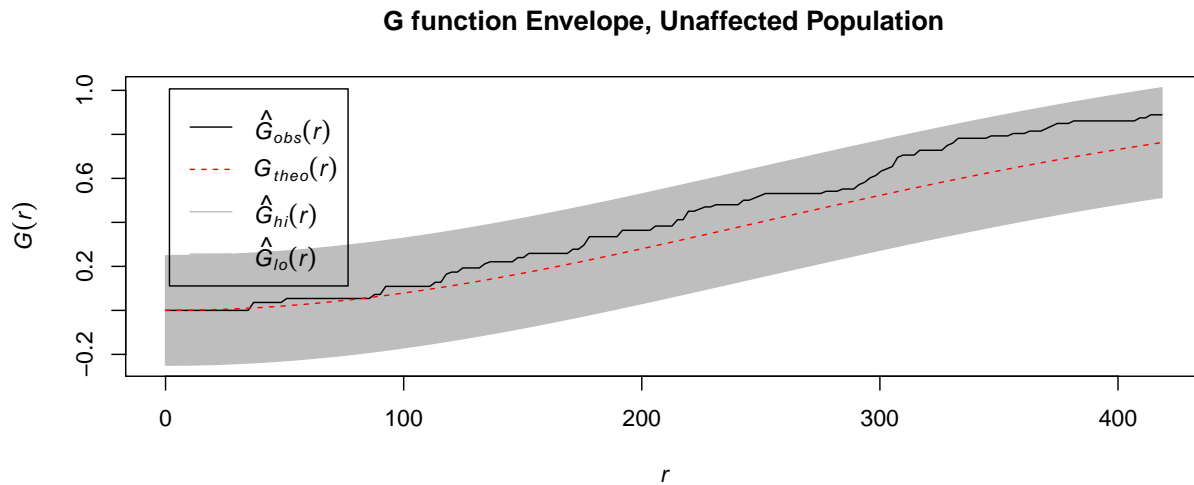
F function Envelope, Affected Population

G function Envelope, Affected Population

F function Envelope, Unaffected Population
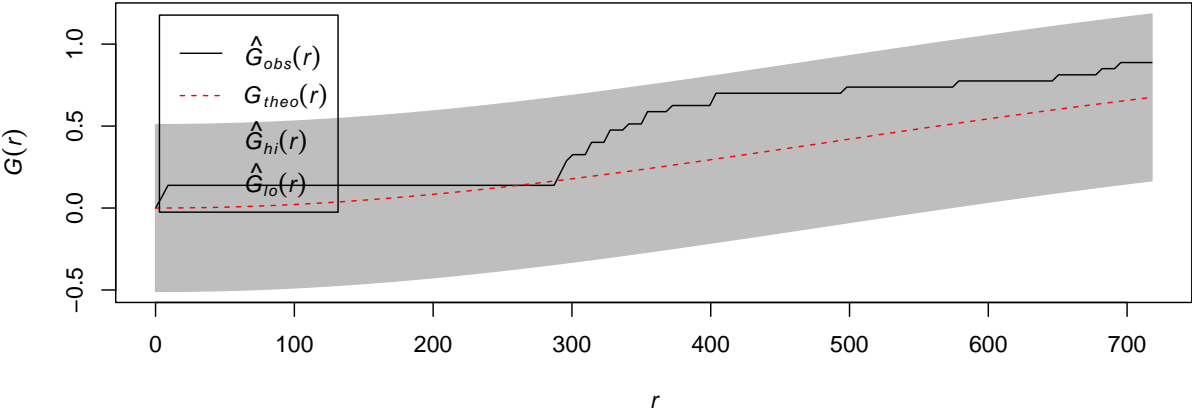
G function Envelope, Unaffected Population
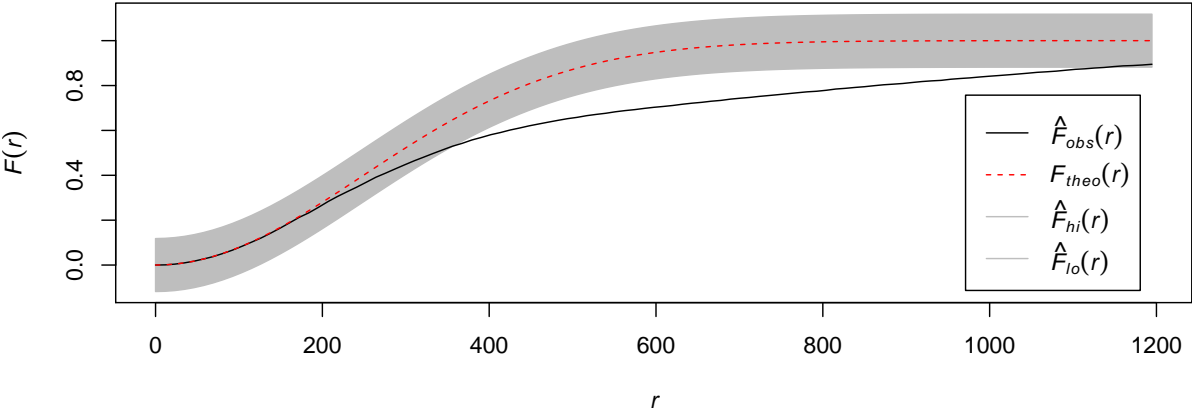
Problem 1c:

Now let's consider a more likely window: that these graves represent a complete excavation of the area in which they appear, and that area is irregularly shaped. Since we don't know what it is, we can choose a rough polygon outline to surround the points. You can create such an outline by plotting the locations and using the locator function; again, see help(owin) for the details of how to specify a polygon boundary. Create two new ppp objects with this new window.

Again, I created these two ppp objects, please find the code in the appendix.

Problem 1d:

Repeat step (b) for the new datasets. What changes? Can you explain the reason for this, based on the form of the test statistics?

Now that I have changed the window to polygons, the data is showing strong evidence for complete spatial randomness. The observed line is very close to the mean simulated line, especially in the F function.
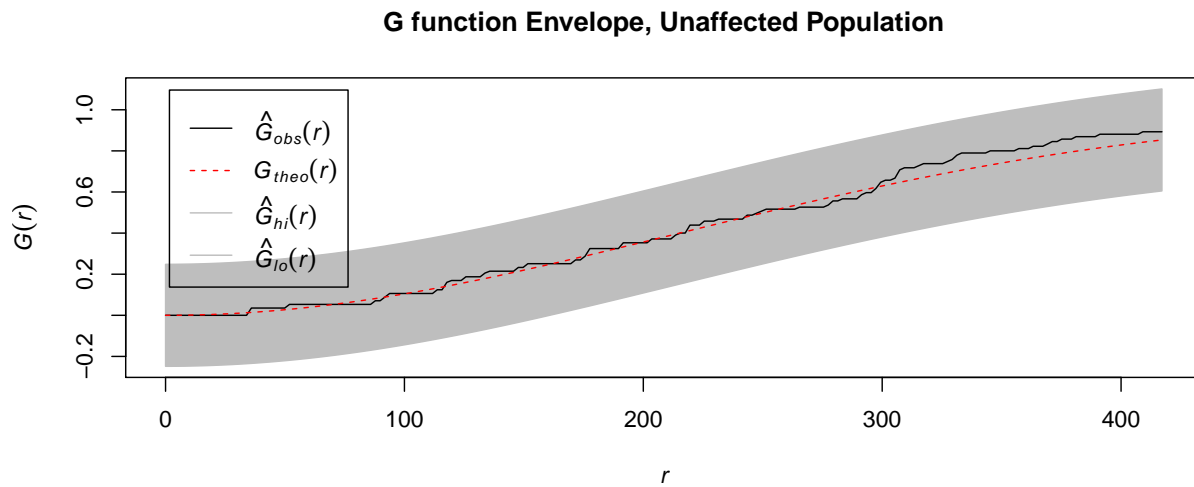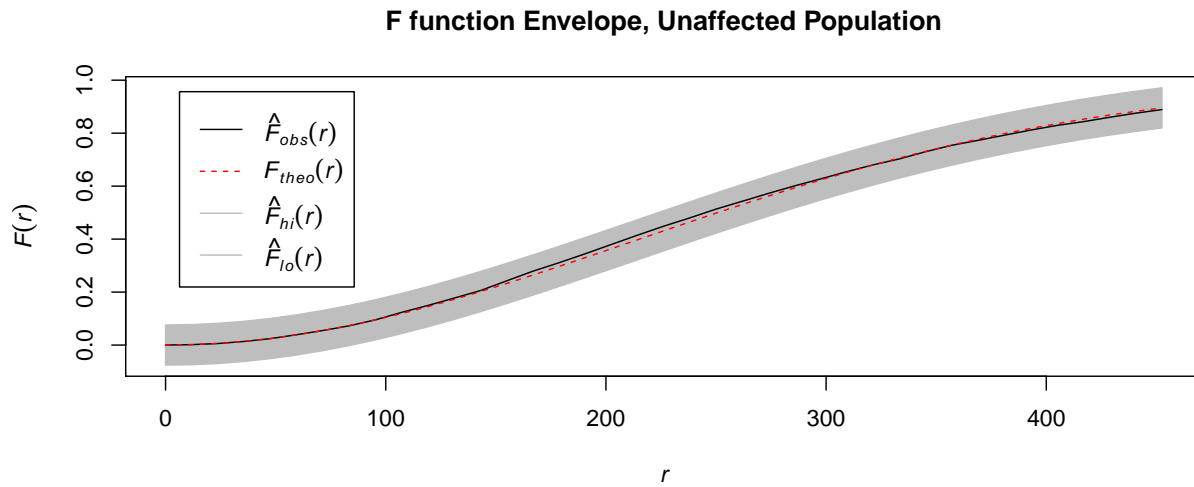
Can you explain the reason for this, based on the form of the test statistics?

F function Envelope, Affected Population

## G function Envelope, Affected Population

F function Envelope, Unaffected Population
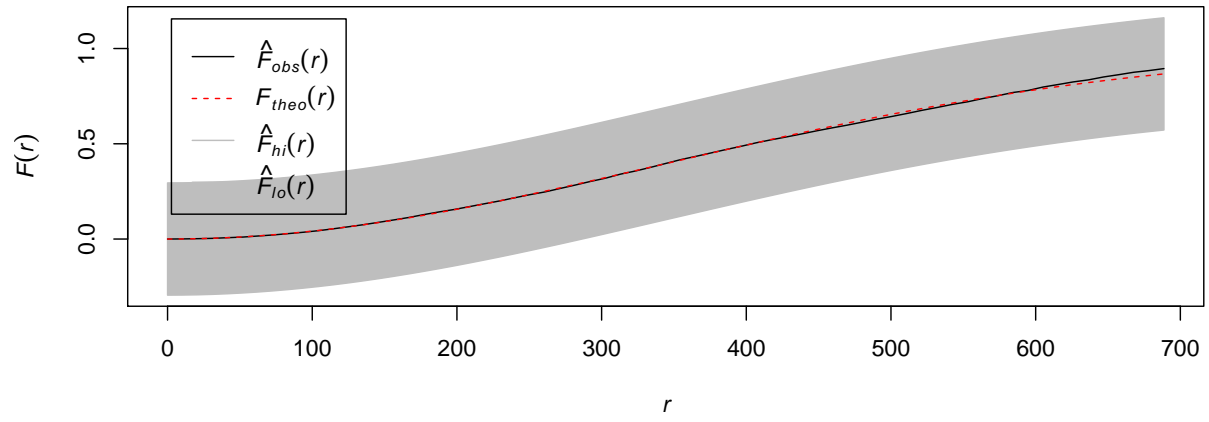
G function Envelope, Unaffected Population
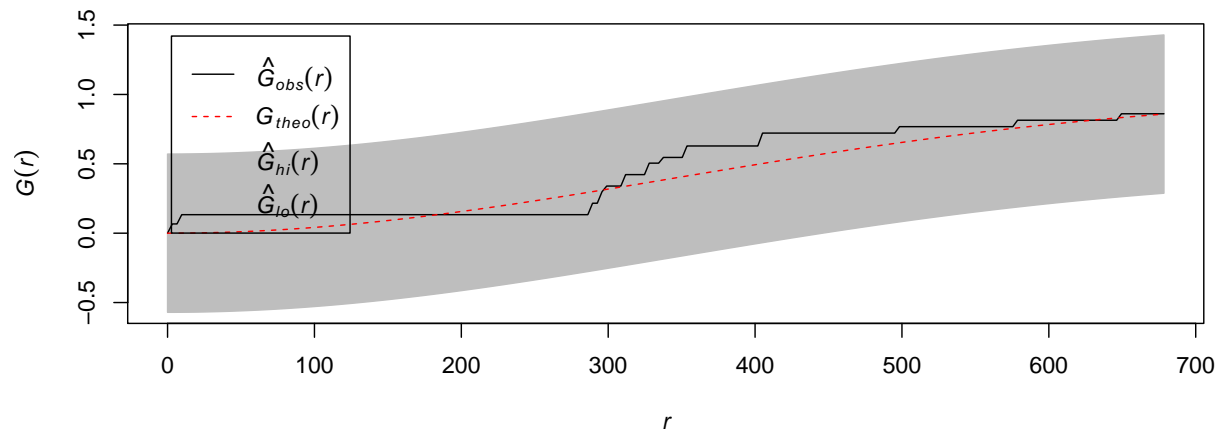
Problem 2:

Simulate four datasets on the unit square, from a homogeneous Poisson process with a rate λ of your choosing. For each one, fit a kernel estimate of the intensity function and plot it with the points overlaid. What do you conclude about interpreting the results of this estimator? In other words, what can go wrong if we fit a spatial point process model without checking for CSR first?

A homogenous Poisson Process is an example of a CSR point process model. If we do not assess for CSR first, we may conclude that there is not complete spatial randomness when we fit models. Even though they are estimators of the same intensity, they give vastly different results in terms of the estimator.
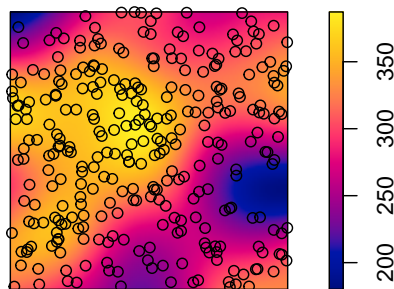
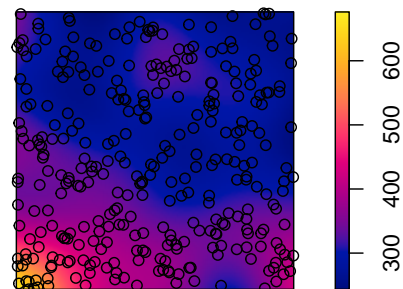**Kernel estimate with Uniform Method on (0,1)x(0,1) when intensity = λ = 300**
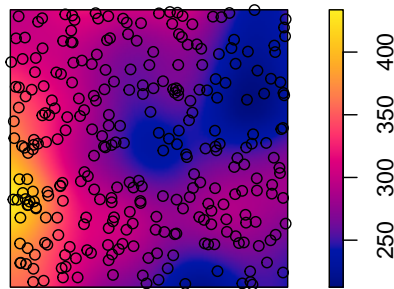
Kernel estimate with Uniform Method on (0,1)x(0,1) when intensity = λ = 300
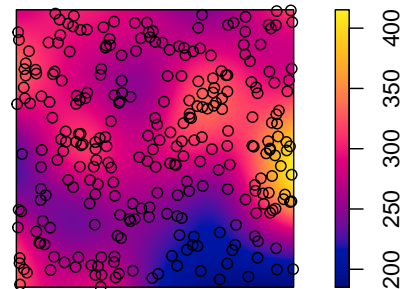
Kernel estimate with Uniform Method on (0,1)x(0,1) when intensity = λ = 300

Kernel estimate with Uniform Method on (0,1)x(0,1) when intensity = λ = 300

Problem 3:

Read through sections 15.2 and 15.3 of the notes by Adrian Baddeley about fitting point process models in R with the spatstat package (you can find it on Canvas). In 15.3, follow along with the R code example for the bei dataset. In particular, find the MLEs for inhomogeneous models with intensity functions:

$$\text{Model 1: } = exp\{\beta_0 + \beta_1 Z(x)\}$$

$$\text{Model 2: } = \beta Z(x)$$

Plot a kernel density estimate of $\lambda(x)$, ignoring covariates, as well as the fitted intensities under the two models. Put them all on the same color scale, and include a sentence or two comparing them. Note: the data for this problem are available from the spatstat package.

I notice that the kernel density estimate of $\lambda(x)$ is fairly different than the fitted intensities under the two models. The two models are relatively similar, just slightly different ranges. However, they include peaks at

6

pretty much the same set of locations.

Kernel Density Estimate　　　　　　　**Model 1**　　　　　　　**Model 2**
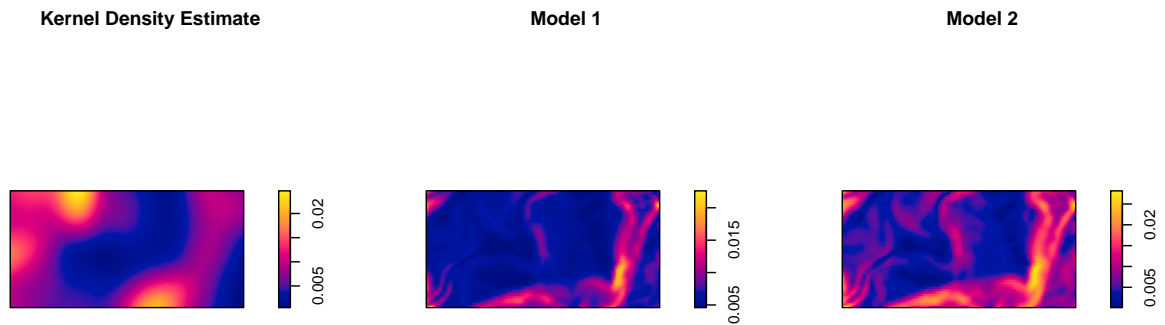
Appendix: R Code

```
###
### Problem 1a
###

#load the data
grave_dat <- read.table("C:/Users/ckell/OneDrive/Penn State/2017-2018/597/spatial_statistics_597/Homewo
#ignore the first column
grave_dat <- grave_dat[,-1]
#label the variables
colnames(grave_dat) <- c("indicator", "x", "y")

#affected dataset
aff_dat <- grave_dat[which(grave_dat$indicator == 1),]
#unaffected dataset
unaff_dat <- grave_dat[which(grave_dat$indicator == 0),]

#what window to use? for now, take the window to be the same for each ppp object
#     the window will be a rectangular region based on the range of x and y for both datasets
window <- owin(xrange = range(grave_dat$x), yrange = range(grave_dat$y))

#create ppp object- affected individuals
aff_ppp <- as.ppp(aff_dat[,c(2:3)], W = window)

#create ppp object- unaffected individuals
unaff_ppp <- as.ppp(unaff_dat[,c(2:3)], W = window)

###
### Problem 1b
###
```

```r
#creating MC sim envelope for affected dataset
#    F function
f_aff <- envelope(aff_ppp, fun = Fest, global = TRUE)

#    G function
g_aff <- envelope(aff_ppp, fun = Gest, global = TRUE)



#creating MC sim envelope for unaffected dataset
#    F function
f_unaff <- envelope(unaff_ppp, fun = Fest, global = TRUE)

#    G function
g_unaff <- envelope(unaff_ppp, fun = Gest, global = TRUE)
```

```r
#creating MC sim plots for affected dataset
#    F function
plot(f_aff, main = "F function Envelope, Affected Population")
#    G function
plot(g_aff, main = "G function Envelope, Affected Population")

#creating MC sim plots for unaffected dataset
#    F function
plot(f_unaff, main = "F function Envelope, Unaffected Population")
#    G function
plot(g_unaff, main = "G function Envelope, Unaffected Population")
```

```r
###
### Problem 1c
###

#create polygons - affected individuals
plot(aff_dat[,c(2,3)])
aff_poly <- locator()

#create polygons - affected individuals
plot(unaff_dat[,c(2,3)])
unaff_poly <- locator()

save(aff_poly, unaff_poly, file = "C:/Users/ckell/OneDrive/Penn State/2017-2018/597/spatial_statistics_5
```

```r
###
### Problem 1d
###

#creating MC sim envelope for affected dataset
#    F function
f_aff_2 <- envelope(aff_ppp_poly, fun = Fest, global = TRUE)
#    G function
g_aff_2 <- envelope(aff_ppp_poly, fun = Gest, global = TRUE)



#creating MC sim envelope for unaffected dataset
#    F function
```

```r
f_unaff_2 <- envelope(unaff_ppp_poly, nrank = 10,fun = Fest, global = TRUE)
#    G function
g_unaff_2 <- envelope(unaff_ppp_poly, fun = Gest, global = TRUE)

#creating MC sim plots for affected dataset
#    F function
plot(f_aff_2, main = "F function Envelope, Affected Population")
#    G function
plot(g_aff_2, main = "G function Envelope, Affected Population")



#creating MC sim plots for unaffected dataset
#    F function
plot(f_unaff_2, main = "F function Envelope, Unaffected Population")
#    G function
plot(g_unaff_2, main = "G function Envelope, Unaffected Population")

###
### Problem 2
###

#homogenous Poisson Process space and rate
x0=0;x1=1;y0=0;y1=1
lambda0<- 300
par(mfrow= c(1,2))

for(i in 1:4){
  #Simulating for number of events in the rectangle
  n.event=rpois(1,lambda0)

  #calculate x and y
  x.i=runif(n.event,min=x0,max=x1)
  y.i=runif(n.event,min=y0,max=y1)

  dat <- cbind(x.i, y.i)
  dat <- as.data.frame(dat)

  #Plot Homogenous Poisson Process
  #plot(dat,xlim=c(0,1), main=expression(paste("Uniform Method on D when intensity = ", lambda[0]," = 3

  #what window to use? for now, take the window to be the same for each ppp object
  #    the window will be a rectangular region of the unit square (0,1)x(0,1)
  window <- owin(xrange = c(0,1), yrange = c(0,1))


  #create ppp object- affected individuals
  dat_ppp <- as.ppp(dat, W = window)

  plot(density.ppp(dat_ppp), main = expression(paste("Kernel estimate with Uniform \nMethod on (0,1)x(0
  points(dat,xlim=c(0,1))
}
```

```
###
### Problem 3
###

#load the data from the spatstat package
data(bei)
grad <- bei.extra$grad

#plot the data
#plot(grad)

#fit model 1
mod_1 <- ppm(bei, ~slope, covariates = list(slope = grad))
#mod_1

#fit model 2
mod_2 <- ppm(bei, ~offset(log(slope)), covariates = list(slope=grad))
#mod_2

par(mfrow=c(1,3))
# plot a kernel density estimate of lambda(x)
plot(density.ppp(bei), main = "Kernel Density Estimate")

# plot fitted intensity for model 1
plot(mod_1, se = FALSE,ngrid= c(100,100), superimpose = FALSE, main = "Model 1")

# plot fitted intensity for model 2
plot(mod_2, se = FALSE,ngrid= c(100,100), superimpose = FALSE, main = "Model 2")

#NEED TO GET THESE ON THE SAME SCALE
```