Statistics 222, Spatial Statistics.

<u>Outline for the day:</u>

1. Kernel smoothing.

2. F, G, J, K, and L functions.

3. Projects.

4. Exercises, code and entering data.

5. van Lieshout pp. 23-26.

1. Kernel smoothing.
A simple way to start summarizing a spatial point process is by kernel smoothing.
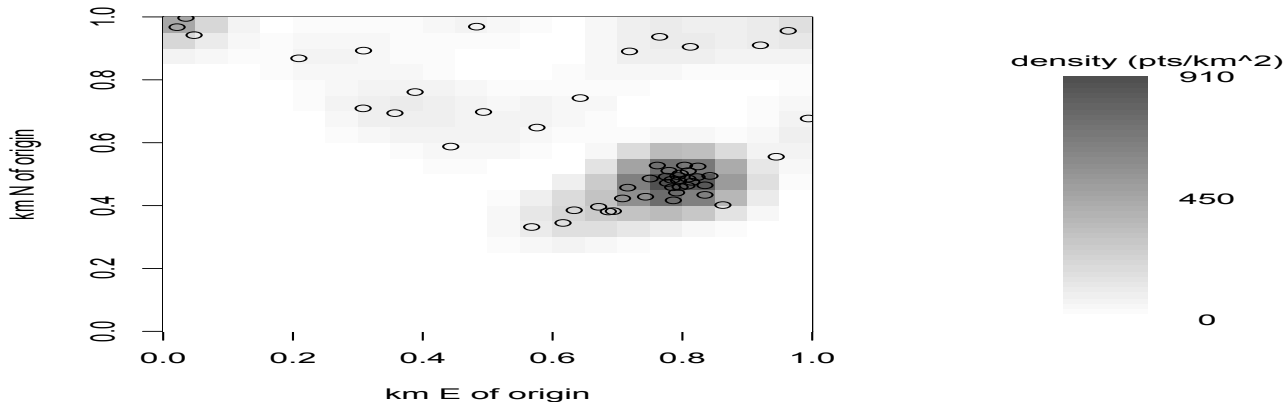
Suppose your observation region is B.
Let k(x,y) be a spatial density function, called a kernel, and construct, for each location (x,y),
$\lambda^{\wedge}(x,y) = \int_B k((x',y') - (x,y)) \, dN(x',y') / \rho(x,y)$,
where $\rho(x,y) = \int_B k((x',y') - (x,y)) \, dx' \, dy'$ is an edge correction term.

The resulting function $\lambda^{\wedge}(x,y)$ is a natural estimator of $\lambda(x,y)$ and, when N is a Poisson process, can be an asymptotically unbiased estimator of $\lambda(x,y)$.

2. F,G,H,J,K, and L functions.

Let F(r) be the probability that the distance from a randomly chosen location to its nearest *point* of the process is $\leq$ r.

Let G(r) be the probability that the distance from a randomly chosen *point* to its nearest neighbor is $\leq$ r.

F is the empty space function and G is the nearest neighbor distribution function.

Matern (1971) showed that for a homogeneous Poisson process, $F(r) = G(r) = 1 - \exp(-\lambda \pi r^2)$.

Marie-Collette van Lieshout

Let $J(r) = (1-G(r)) / (1-F(r))$, for any r such that F(r)<1.

J > 1 indicates inhibition, and J < 1 indicates clustering.
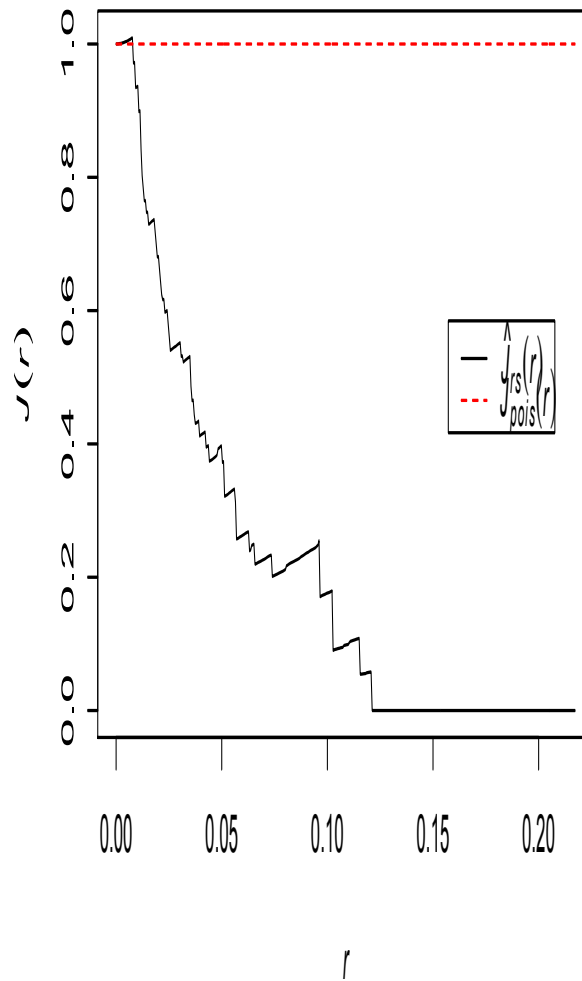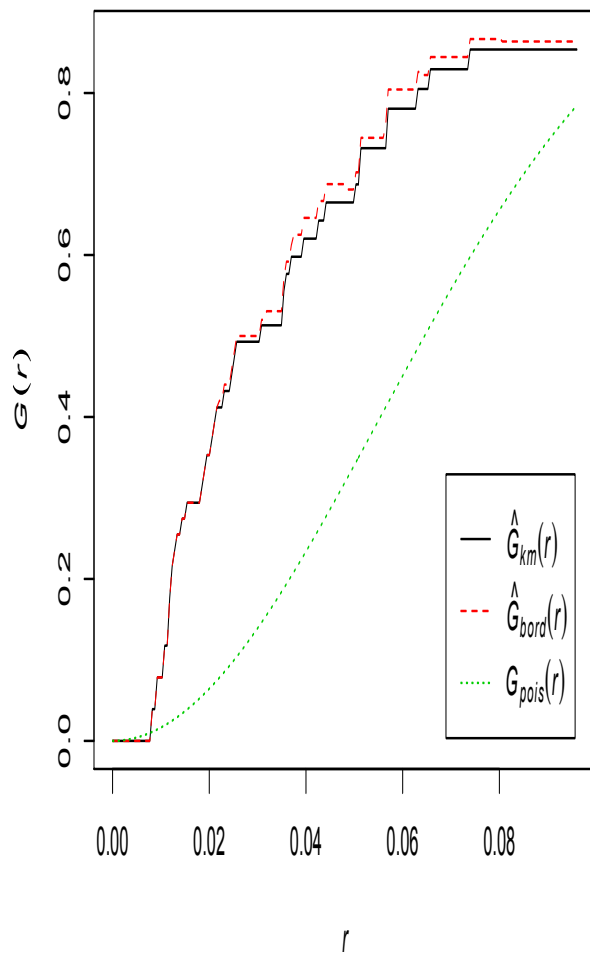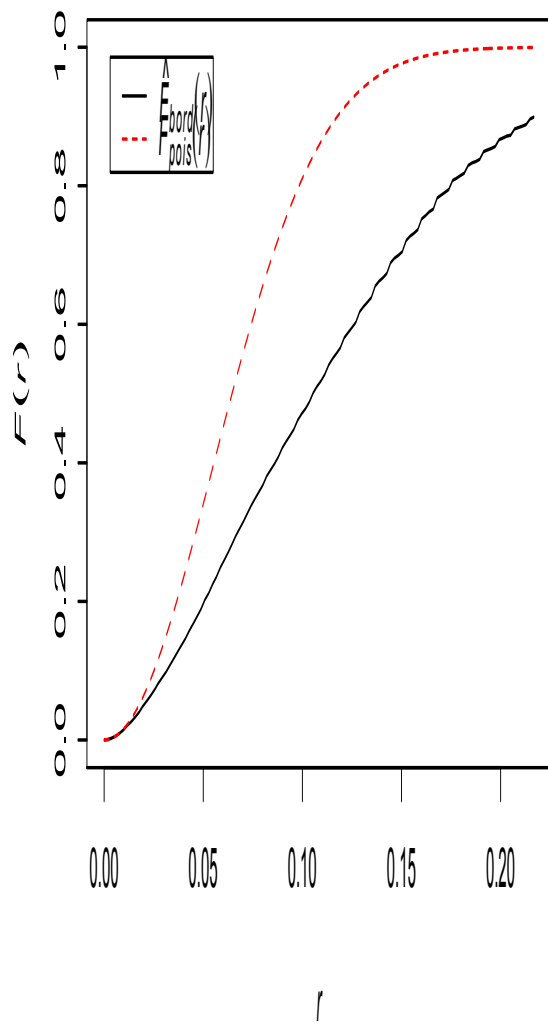
For a stationary Poisson process with rate μ, let

$K(r) = 1/\mu$ E(# of other points within distance *r* of a randomly chosen point).

K is the reduced 2nd moment measure or *Ripley*'s K-function (Ripley, 1976).

van Lieshout, M.C. (2010). A *J*–function for inhomogeneous point processes. *Statistica Neerlandica*, 65(2), 183-201.

and references therein gives extensions to the inhomogeneous Poisson process and to marked point processes.

F, G, J, K, and L functions, continued.

F, G, J, K, and L functions, continued.

K(r) = 1/μ E(# of other points within distance *r* of a randomly chosen point).
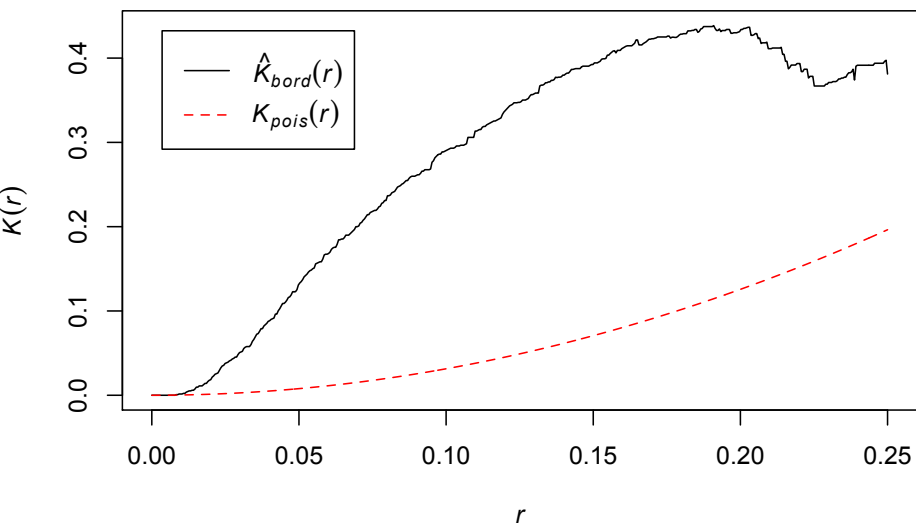
K is estimated in the obvious way using data, but various edge correction ideas are available.

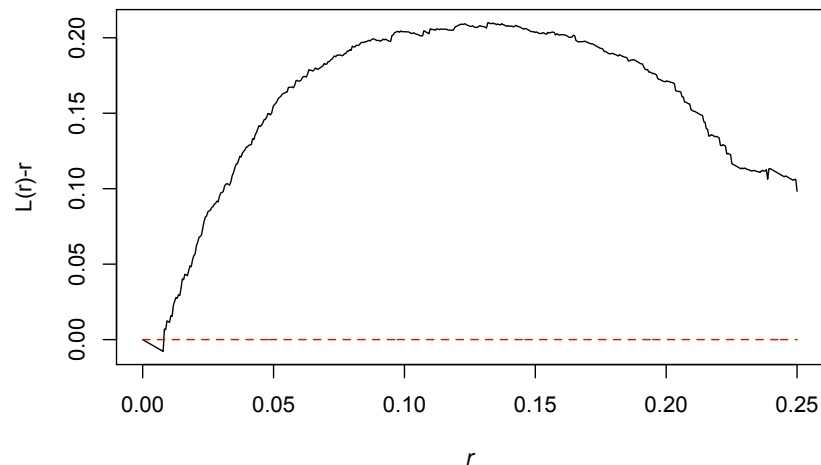For a stationary Poisson process in R^2, $K(r) = \pi r^2$, so one may consider
$L(r) = \sqrt{(K(r)/\pi)}$.
For a stationary Poisson process in R^2, $L(r) - r = 0$ and $L^{\wedge}(r)-r$ should be approx. 0.

3. Project notes.

Oral Presentation (10%) last 2 weeks of class.

Written project (45%) due Sat, Jun8, 8pm, by email to me at frederic@stat.ucla.edu .

Find a spatial-temporal point process dataset and analyze it using some of the relevant methods described in class. Your report should contain about 4-6 pages of text, followed by as many figures as appropriate. You may include as many figures as you like, but the text itself should not exceed 6 pages, double-spaced. In selecting your dataset, choose something that interests you, and try to choose a dataset where the quantification of clustering or inhibition of the locations of the points would be of interest. Begin your paper with an introduction, a description of your data and how they were obtained, and a summary of the main questions to be addressed in your paper. Then summarize your analyses, paying special attention to the plausibility of assumptions you are making. Conclude by assessing how effective the methods you used were in helping to answer your main questions. Oral presentations of project results will take place on the last 2 weeks of class. These will involve simply presenting a clear and concise 10 minute summary of your dataset including a couple of your main results. Do not feel you have to show all the results from your paper in your oral presentation! You should only show around 8-10 slides in your presentation. Email me a pdf of your slides the night before your presentation, by 8pm, so I can have them ready before class starts. You will use my computer in class.

More notes about the projects.
The oral presentations will be 10 minutes each in total. You will receive one overall grade based on the oral report and written report, combined.

Your dataset, which you will find yourselves, on the web,
can be anything you choose, but it should be:
a) spatial or spatial-temporal point process data.
You should have between n=50 and 5000 observations.
b) something of genuine interest to you, and where you have more knowledge than an average person.
It is fine with me if you enter your data yourself using your keypad. In this case show the actual data and your approximation.

You can have covariate information or not, but the points should be of interest on their own. Analyze the data using the methods we have talked about in class, including doing kernel smoothing, estimating standard functions like the J function and L function, and fit models by maximum likelihood. Comment on your fitted model.

Your final project should be submitted to me in pdf by email to frederic@stat.ucla.edu. Do not send them via ccle. They are all due the same date, regardless when your oral presentation is.

More notes about the projects.
For the oral projects,

Rule 1: 8-10 minutes per oral report, plus questions at the end. I will cut you off if you go over 10min. Have someone in the audience help you with the time if you want. I will tell you when there are 2 min left, and then I will tell you to stop.

Rule 2: Everyone must be respectful and quiet during other people's talks. You can ask clarifying questions but keep deep questions until the end.

Rule 3: Speak very slowly in the beginning. Give us a sense of your data. Assume that the listener knows everything about point processes, but knows nothing about the subject matter. Tell us what the methods say about your data. Emphasize the results more than the methods. Make sure to go slowly and clearly in the start so that the listener really understands what your data are.

More notes about the projects.

Rule 4: Speculate and generalize but use careful language. Say "It seems" or "appears" rather than "is" when it comes to speculative statements or models. For example, you might say "The points appear roughly homogeneous" or "a Hawkes model seems to fit well" but not "The data are uniformly distributed" or "The data come from a homogeneous Poisson model".

Rule 5: Start with an introduction explaining what your data are, how you got them, and why they are interesting (roughly 2-3 minutes), then show your results as clearly as possible, with figures preferred (roughly 5 minutes), and then conclude (roughly 2-3 minutes). In your conclusion, mention the limitations of your analysis and speculate about what might make a future analysis better, if you had infinite time. This might include collecting more data, or getting data on more variables, as well as more sophisticated statistical methods.

For your written reports, rules 4-5 apply again. The entire written project should be 4-6 pages, plus figures. Have just the text in the beginning, and then the figures at the end. Please do not spend any time trying to embed the figures into the text. Email your pdf document to me, at frederic@stat.ucla.edu .

4. Exercises.

Suppose N is a spatial point process with clustering for distances $\leq d$.
Let F(r) be the empty space function and let G(r) be the nearest neighbor distribution function.
Which of the following is true.

a. F(d) = G(d).
b. F(d) < G(d).
c. F(d) > G(d)

Entering data and kernel smoothing example.

First, input 54 points using the mouse.
n = 54
plot(c(0,1),c(0,1),type="n",xlab="longitude",ylab="latitude",
                    main="locations")
  x1 = rep(0,n)
  y1 = rep(0,n)
  for(i in 1:n){
  z1 = locator(1)
  x1[i] = z1$x
  y1[i] = z1$y
  points(x1[i],y1[i])
  }

PLOT THE POINTS WITH A 2D KERNEL SMOOTHING IN GREYSCALE PLUS A LEGEND

```
library(splancs)
bdw = sqrt(bw.nrd0(x1)^2+bw.nrd0(y1)^2)  ## possible default bandwidth
b1 = as.points(x1,y1)
bdry = matrix(c(0,0,1,0,1,1,0,1,0,0),ncol=2,byrow=T)
z = kernel2d(b1,bdry,bdw)
attributes(z)
par(mfrow=c(1,2))
image(z,col=gray((64:20)/64),xlab="km E of origin",ylab="km N of
    origin")
points(b1)
x4 = seq(min(z$z),max(z$z),length=100)
```

```r
plot(c(0,10),c(.8*min(x4),1.2*max(x4)),type="n",
        axes=F,xlab="",ylab="")
image(c(-1:1),x4,matrix(rep(x4,2),ncol=100,byrow=T),
add=T,col=gray((64:20)/64))
text(2,min(x4),as.character(signif(min(x4),2)),cex=1)
text(2,(max(x4)+min(x4))/2,
        as.character(signif((max(x4)+min(x4))/2,2)),cex=1)
text(2,max(x4),as.character(signif(max(x4),2)),cex=1)
mtext(s=3,l=-3,at=1,"density (pts/km^2)")

library(spatstat)
b2 = as.ppp(b1,c(0,1,0,1))
k = Kest(b2,correction="border")

plot(k, main="K function")

plot(k, sqrt(./pi)-r ~ r, ylab="L(r)-r", main="L function",legend=F)
```

```
par(mfrow=c(1,3))
plot(Fest(b2,correction="border"),main="")
plot(Gest(b2,correction="border"),main="")
plot(Jest(b2,correction="border"),main="")
```