Statistics 222, Spatial Statistics.

Outline for the day:

1. Project notes.
2. Marked G and J functions.
3. Weighted K function.
4. Project order.
5. Kernel smoothing, summary functions, model fitting, and weighted K function for spatial point processes, unmarked and marked, in R.

1. Project notes.
Oral Presentation (10%) -- last 2 weeks of class.
Written project (45%) -- due Sat, Jun 9, 8pm, by email to me.

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 8-10 minute summary of your dataset including a couple of your main results. Do not try to show all the results from your paper in your oral presentation! You should only show around 8-12 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.

1. More notes about the projects.
The oral presentations will be 8-10 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.

You can have covariate information also, 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.

1. More notes about the projects.
For the oral projects,

Rule 1: Do not look at me when you are talking.

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

Rule 3: Everyone must be respectful and quiet during other people's talks. You can ask clarifying questions but keep deep questions until the end.

Rule 4: Send me a pdf version of your slides by 8pm the night before your talk, to frederic@stat.ucla.edu. That way, I can set up the talks in order ahead of time and we won't have to waste time in class waiting for each person to connect their laptop to the projector. About 8-12 slides seems right, though it's fine with me if you want fewer or more. You will use my laptop.

1. More notes about the projects.
For the oral projects,
Rule 5: 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.

Rule 6: 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 7: 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 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.

1. More notes about the projects.

For your written reports, rules 5-7 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. Email your pdf document to me, at frederic@stat.ucla.edu .

2. Marked G and J functions.

$G(r) = P_0$(point within r), where P_0 means given a pt. at 0.
It is estimated with $\hat{G}(r) = 1/n \sum_i 1(\text{there is } j: |\tau_i - \tau_j| \le r)$.
$= 1/n \sum_i 1(\min_{i \ne j} |\tau_i - \tau_j| \le r)$

One could alternatively compute a *marked* G-function
$1/n_1 \sum_i 1(\min_j |\tau_i - \tau_j| \le r)$
where the sum is over the n_1 points τ_i with mark in some range M_1, and the minimum is over the points τ_j with mark in some range M_2.
This is the *marked* or *cross* G-function.
One can similarly define a marked or cross J-function
as $J(r) = (1-G(r)) / (1-F(r))$ accordingly, plugging in the corresponding G function.

Marie-Collette van Lieshout

van Lieshout, M.N.M. (2006). A J-function for marked point patterns. *AISM* 58, 235-259.

3. Weighted K function.

For a stationary Poisson process with rate μ,
$K(r) = 1/\mu\ E$(# of other points within distance *r* of a randomly chosen point).
Estimated via $K4(r) = 1/(\hat{\lambda}\, n) \sum_{i \neq j} (|\tau_i - \tau_j| \le r)\, w(\tau_i, \tau_j)$,
where $\hat{\lambda} = n/|S|$, and $w(\tau_i, \tau_j)$ = 1/proportion of circle centered at i going through j that is in S = border correction term.
If N is inhomogeneous, can instead weight each point by $1/\lambda$,
obtaining $K_w(r) = 1/n \sum_{i \neq j} (|\tau_i - \tau_j| \le r)\, w(\tau_i, \tau_j) / \lambda(\tau_i) / \lambda(\tau_j)$.

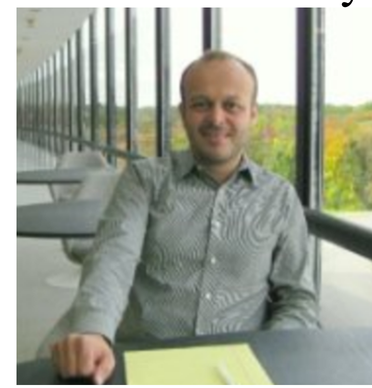
$K_w(r) \sim N(\pi r^2, 2\pi r^2 |S| / E(n)^2)$, if $\inf \lambda = 1$.

Baddeley, A., Møller, J., Waagepetersen, R. (2000). Non and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3), 329-350.

Veen, A. and Schoenberg, F.P. (2006). Assessing spatial point process models for California earthquakes using weighted K-functions: analysis of California earthquakes, in *Case Studies in Spatial Point Process Models*, Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (eds.), Springer, NY, pp. 293-306.

Adelfio, G. and Schoenberg, F.P. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties. *Annals of the Institute of Statistical Mathematics*, 61(4), 929-948.

Adran Baddeley



Alejandro Veen

Giada Adelfio

4. Presentation times.

I will now randomly assign people to presentation times. If you want to change oral presentation dates and times with another person, feel free but let me know.