Rlab: random spatial index (point processes)

Kristine Dinh

3/16/2020

I. Exploration

1. [2 pts] Create a `SpatialPolygons*` object named `kansas` that represents the boundary of the state of Kansas (hint: we've used the `tigris` package to do things like this in the past). Provide executable code.

```
states <- states(cb = TRUE, resolution = "20m")

kansas <- states[states$NAME == "Kansas", ]
```

2. [2 pts] Make a new column in the `dataframe earthquakes` called `POSIX` that converts Date to a `POSIX*` class object. What is the most recent observation?

The most recent observation is:

```
##     Date     Time Magnitude      POSIX
## 1 5/3/16 20:20:55       2.6 2016-05-03
```

3. [3 pts] How many earthquakes were there in 2002? How many in 2012? 2015?

```
##   year number_of_earthquakes
## 1 2002                     4
## 2 2015                   449
```

- 2012 doesn't have any earthquakes

4. [2 pts] Make a new column in the dataframe wells called `COMPLETION_POSIX` that converts the completion date for each well into a `POSIX*` class object. When was the most recent well completed?

```
wells@data$COMPLETION_POSIX %<>% as.POSIXct()
```

- The most recent well completed was on 2020-02-10

5. [2 pts] Which year had the largest number of wells completed? How many wells were completed that year?

```
##   COMPLETION_year number_of_wells
## 1            2014            2876
```

6. [3 pts] Make new versions of `kansas`, `earthquakes`, and `wells` called `kansas_utm`, and `earthquakes_utm`, `wells_utm` that change the CRS to the Universal Transverse Mercator coordinate system, zone 13. Provide executable code. Compare your first few coordinate values to the ones below to make sure you've transformed correctly.

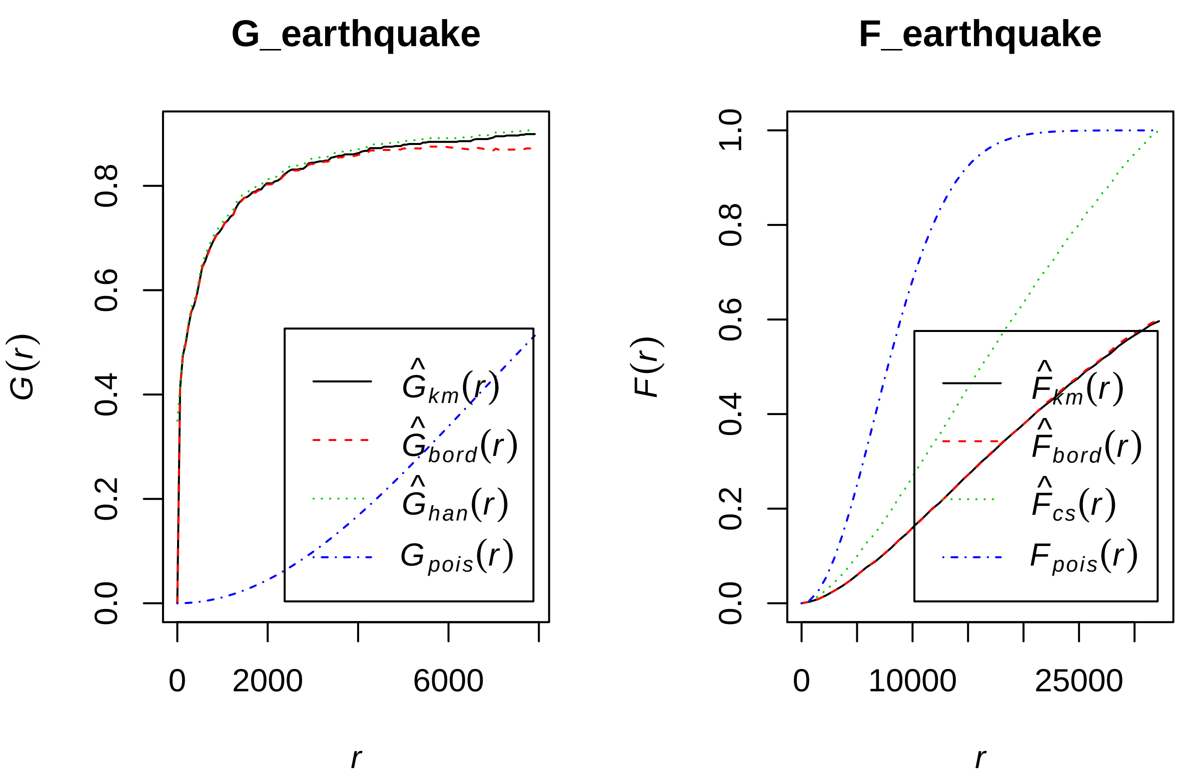
```
CRS_new <- CRS("+proj=utm +zone=13")

kansas_utm <- spTransform(kansas, CRS_new)

earthquakes_utm <- spTransform(earthquakes, CRS_new)

wells_utm <- spTransform(wells, CRS_new)
```

7. [2 pts] From the explorations so far, do you anticipate any connection between the locations/numbers of new oil & gas wells and the location/numbers of earthquakes? Why or why not?

- I anticipate that there is weak negative connection between numbers of new oil & gas wells and number of earthquakes. Because at the location where there is wells, earthquake tends to happened less.

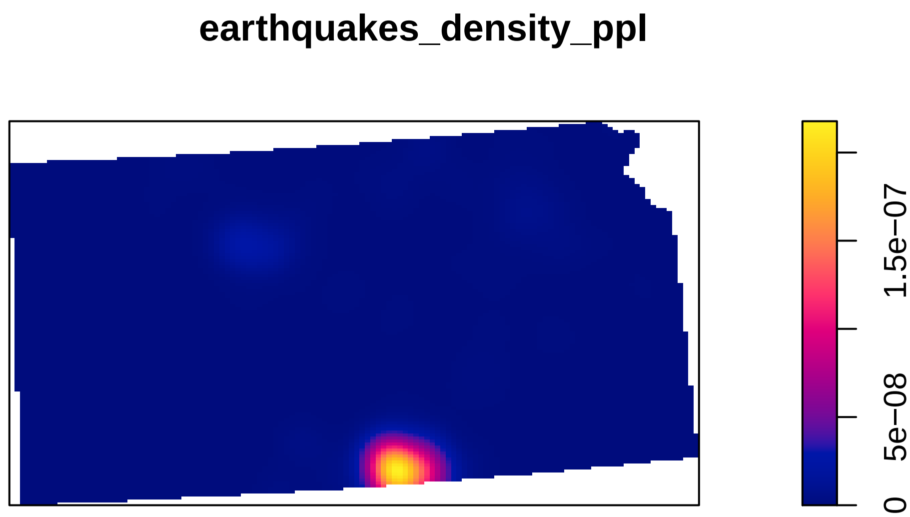
II. First-order properties of earthquakes

8. [4 pts] Compute the empirical G(r) and F(r) functions for the collection of all earthquake locations. Is it reasonable to conclude that the earthquakes arise according to a homogeneous Poisson process? Why or why not?



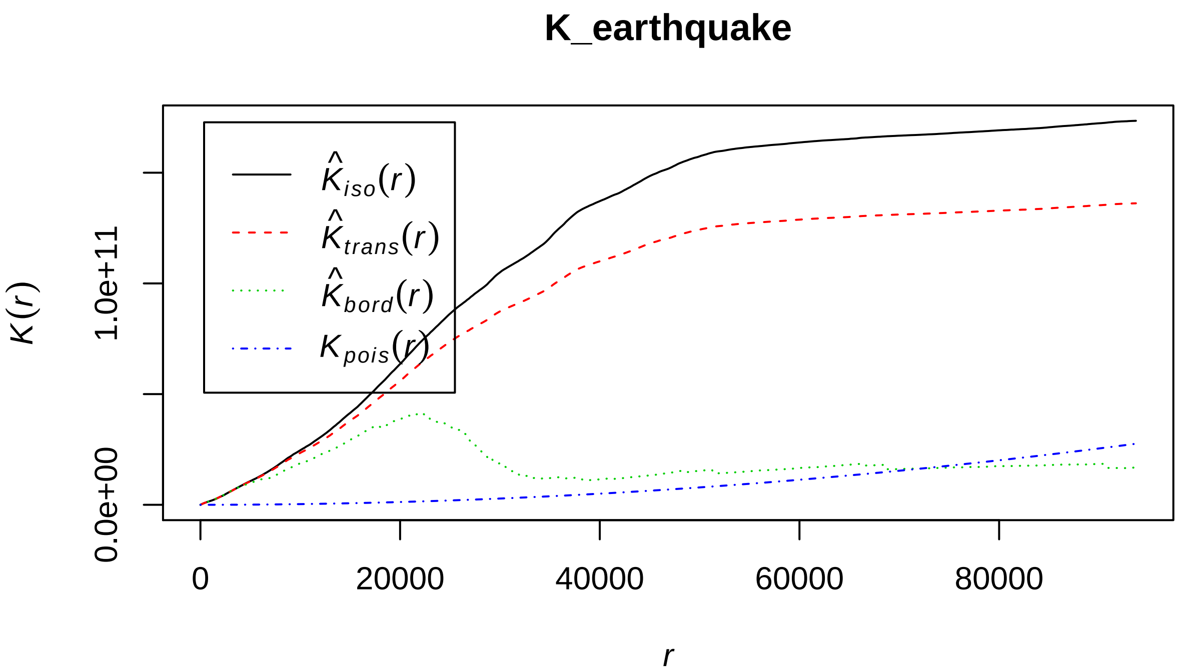
- It's reasonable to conclude that the earthquakes arise according to a homogeneous Poisson process. Because the distant of the lines are too far apart. As seen in plot of G(r) and F(r), we can see that they shows that they lines are homogeneous Poisson.

9. [4 pts] Estimate the intensity function for the collection of all earthquake locations. Make a plot of the intensity surface.



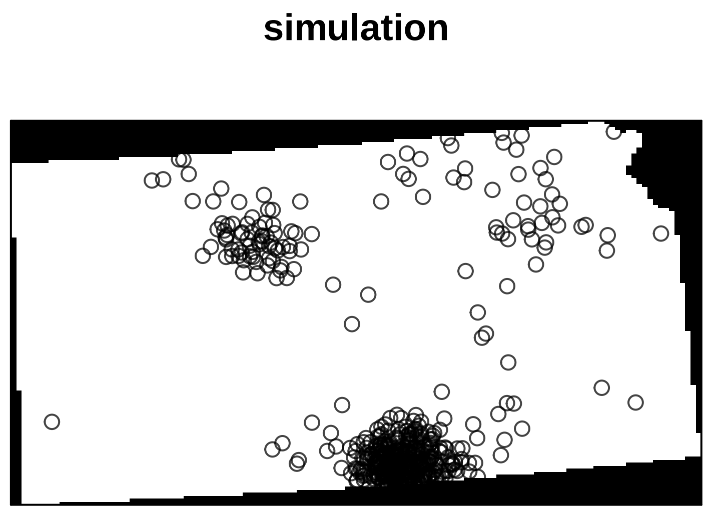
III. Second-order properties of earthquakes

10. [2 pts] Compute the empirical K(r) function for the collection of all earthquake locations. How does it compare to the theoretical function for an independent homogeneous Poisson process?

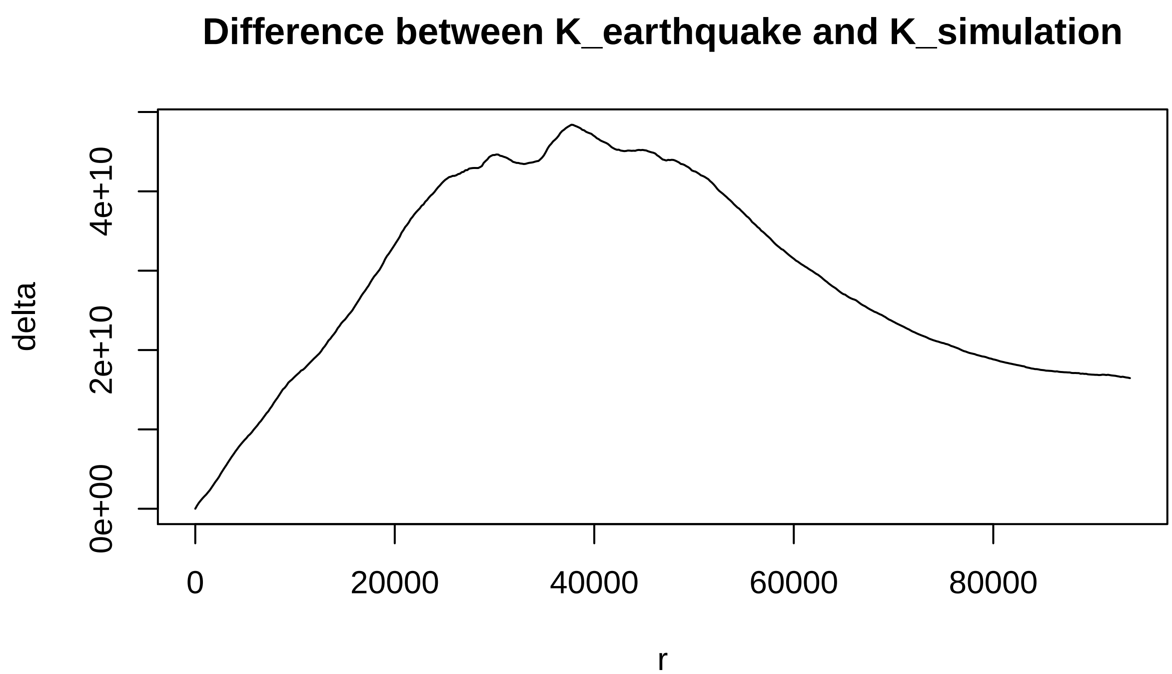


- The lines seems to be similar at about $r < 1000$, but after that, the different lines starting to be wider apart. Comparing to a homogeneous Poisson process, this figure seems to be very different.

11. [4 pts] Use the function `rpoint()` from the `spatstat` package to simulate 500 points from an independent inhomogeneous Poisson process with an intensity function that matches your estimate in question 9. Make a plot of the points with the border of Kansas for reference.



12. [2 pts] Compute the empirical K(r) function for the independent points you just simulated. Plot the difference between the two functions (hint: you will need to set the r argument in Kest() so that the two estimated functions match up).



13. [2 pts] What does the plot suggest about whether or not the earthquake data are independent?

- The plot suggest that earthquake data are dependent.

IV. Cases and controls

14. [2 pts] Use the function `rdist()` in the `fields` package to compute a matrix containing the distances between all locations in `earthquakes_df` (both cases and controls), and `wells`. Name the matrix `d2wells` and provide your executable code.

```
d2wells <- rdist(earthquakes_df@coords, wells_pre2014@coords)
```

15. [2 pts] Fit a generalized linear model with a logit link function that includes a linear effect for distance to the nearest well. Give a 95% CI for the regression coefficient for distance to nearest well.

```
earthquakes_glm <- glm(earthquakes_df$z ~ d2nearest_well, family = binomial(link = "logit"))
```

```
##                            term      estimate    std.error statistic      p.value
## (Intercept)       (Intercept)  0.7784728042 8.967551e-02  8.680997 3.923160e-18
## d2nearest_well d2nearest_well -0.0001183484 1.491111e-05 -7.936929 2.072493e-15
##                        2.5 %        97.5 %
## (Intercept)     0.6053086759  9.570242e-01
## d2nearest_well -0.0001491407 -9.071495e-05
```

16. [2 pts] What does your analysis suggest about a relationship between the locations of oil & gas wells and earthquakes?

- My analysis suggest that there is a negative correlation between the locations of oil & gas wells and earthquakes. In other words, as the distance of locations of oil & gas wells increases by 1-unit, earthquakes will appear on average is $-1.1834841 \times 10^{-4}$

17. [4 pts] Now fit a generalized additive model that accounts for small scale spatial dependence in the residuals like the one we looked at in class for `asthma`. How does the effect of the distance to the nearest well change (if at all)?

```
earthquakes_gam <- gam(earthquakes_df$z ~ d2nearest_well + s(x, y),
                       data = earthquakes_df, family = binomial(link = "logit"))
```

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## earthquakes_df$z ~ d2nearest_well + s(x, y)
##
## Parametric coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.525e+01  1.321e+01  -1.912   0.0559 .
## d2nearest_well  7.503e-05  1.108e-04   0.677   0.4983
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq  p-value
## s(x,y) 26.16   27.2  108.5 1.35e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.93   Deviance explained = 90.1%
## UBRE = -0.81227  Scale est. = 1         n = 1093
```

- The coeficient of `d2nearest_well` changed from negative to postive. Now, the relationship between the distance to the nearest well and earthquakes is not statistically significant anymore.