Spatial Point Analysis

Sudhakar Brar

May 17, 2021

Contents

1 Data **i**

2 Intensity **i**

3 Poisson Models **ii**

3.1 Test for complete spatial randomness ii

3.2 Maximum Likelihood for Poisson Process for Inhomogenous Processes ii

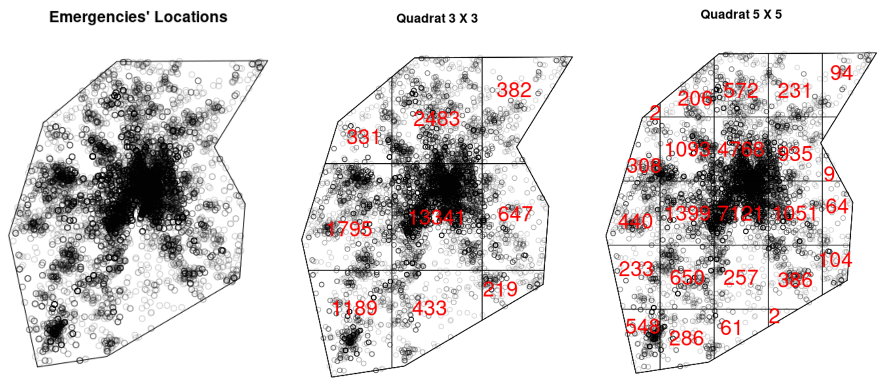
4 Interaction **iii**

5 Conclusion **v**

6 References **vi**

1 Data

The database is of emergencies regarding firemen which is provided by the fire department SDIS 31. It contains the location and characteresitcs (time, duration, number of firemen, etc) of the emergencies during 2004 in the surroundings of Toulouse. It consists of 20820 points and 5433 unique points. The area of the concerned region is 619.79 km^2. The points in the dataset are in meters. They were converted into kilometers. Since, there are many points which have been duplicated, hence, we add a random gaussian noise to the X co-ordinates of the data. This gives us 20820 unique data points. Now, plotting the data (Figure 1 (a)), we can see that there might be some inhomogeneity in our data. Since, a lot of points are concentrated in the middle of the region. We check for this in the next section.



(a) (b) (c)

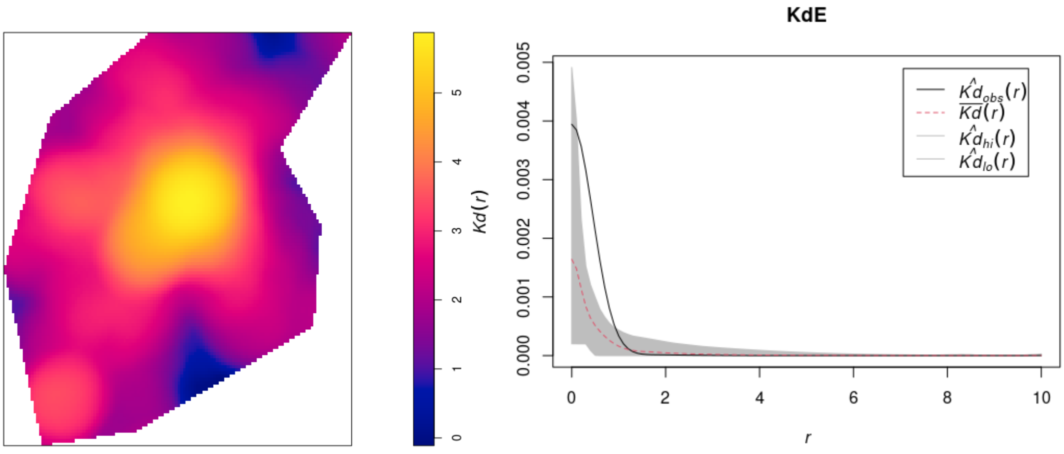
Figure 1: (a) Locations of Emergencies (b) Quadrat count plot (3 X 3)
(c) Quadrat count plot (3 X 3)

2 Intensity

From the above figure (Figure 1 (a)), we can see that many points are concentrated in the middle. This might lead us to suspect that there is inhomogeniety in our data. Hence, we will now plot 2 different quadrats for our data, below. In Figure 1 (b) & (c), we can see that the number of points in a quadrat vary a lot. Hence, we strongly suspect that there might be inhomogeneity in our data. However, since this is not a statistical test, we cannot say definitely if there is inhomogeniety in our data.

To further check for the issue of homogeneity, we will plot *Kernel Smoothed Intensity*, more specifically, the logarithm transformation of Kernel Smoothed Intensity. In the figure 2, we can see that the intensity is highest in the middle and becomes relatively lower near the boundaries. This can also be seen in the figure 1 (a), where there are maximum number of points in the middle quadrat. Hence, again we suspect that there might be homogeneity in our data.

In the case of fire emergencies, it might be the case that the fire emergencies with higher workloads are located close to each other. Hence, we use the envelope of K_d to determine this. We take the 5% highest emergencies which caused the highest workloads and see if at a certain distance (0 km to 10 km), the probability another serious emergency is greater thant that of finding an emergency regardless of its workload (a similar methodology to Macron (2015)). From the Figure 2 (b) we can see that this probability is quite negligible.



(a) (b)

Figure 2: (a) Kernel Smoothened Intensity (b) Envelope of K_d

3 Poisson Models

3.1 Test for complete spatial randomness

Under the null hypothesis of this test, we assume that our spatial data is in a complete random fashion, i.e., it is a homogenous spatial Poisson process. Since, we did see some evidence regarding inhomogeneity in our data, we will use this test to determine if there is or isn't any of it. Below (Table 1) we tabulate the results of 5 different tests. We use five different tests since the results of the test are highly dependent on the number of quadrats. Hence, we select different kinds of quadrats. We can see that the *P-Value* in all the 5 tests is 0. Hence, we are able to reject the null hypothesis.

Table 1: CSR Test Result

Quadrat	χ^2 Stat	DoF	P-Value
3 X 3	36690.48	8	0
4 X 4	34282.22	15	0
5 X 5	46648.97	23	0
6 X 6	71216.81	32	0
7 X 7	69902.31	45	0

3.2 Maximum Likelihood for Poisson Process for Inhomogenous Processes

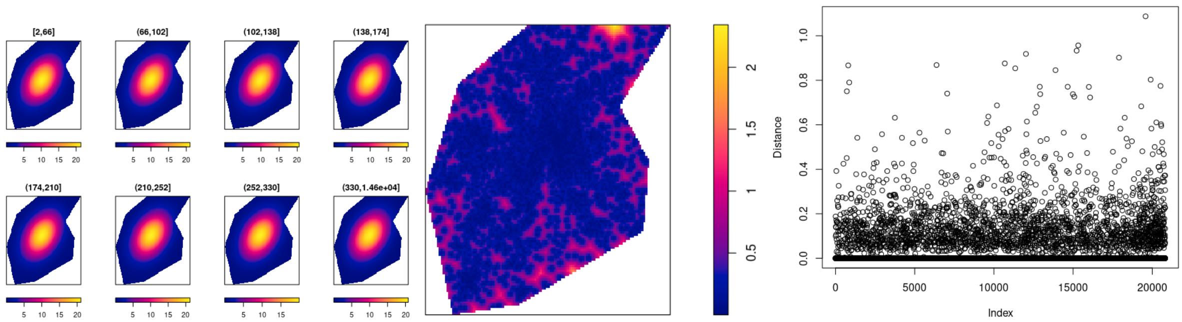
In this section we will different types of poisson models and see which ones tend to perform better than others. Before that we divide our data into eight equal parts. This is to decretize the workload variable which will help in judging if the workload has any impact on the estimated intensity. We use 6 different modes to assess the realtionship on which intensity depends.

We also perform a Likelihood ratio test which follows a χ^2 distribution. In this likelihood ratio test, we test the different models against the assumption that the intensity is homogenous. We present the results of this test in Table 2.

From the results, we can see that when the intensity depends just on the x co-ordinate or on its higher degree polynomial, the relationship in the models (1, 3, 4) seems to be significant. However, when it just depends the y co-ordinate or on its higher degree polynomial, the relationship in models (2, 6) don't seem to be significant. It is only in the model 7 where the degree of y co-ordinate is, where the relationship in the model seems to be significant. However, when we use x and y co-ordinates together, the relationship in the models (5, 8, 9) still seems to be significant. Since, we previously suspected that the intensity might depend also depend on the workload of a fire emergency, hence, we also test for models which have workload as a covariate in them (8, 9, 10). We can see in model 8 that when intensity just depends on the workload, the relationship between intensity and workload is not significant. But when we add x and y co-ordinates (9) or their higher degree polynomial (10), the relationship then seems to be significant.

Table 2: Likelihood Ratio Test

	Model	No. of Coefficient	DoF	Deviance	P-value
1	~x	2	1	18.4	1.829e-05***
2	~y	2	0	345.2	
3	~polynom(x, 2)	3	1	10456.0	<2.2e-16***
4	~polynom(x, 3)	4	1	1577.2	<2.2e-16***
5	~x + y	3	-1	-12030.2	<2.2e-16***
6	~polynom(y, 2)	3	0	12214.5	
7	~polynom(y, 5)	6	3	6055.9	<2.2e-16***
8	~Workload	10	2	-18260.0	
9	~Workload + x + y	10	4	366.6	<2.2e-16***
10	~Workload + polynom(x, y, 2)	13	3	26669.4	<2.2e-16***



(a) (b) (c)

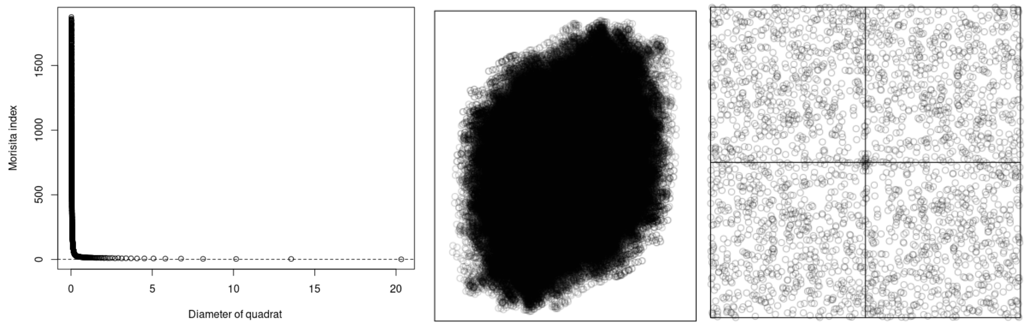
Figure 3: (a) SE of Model 10 (b) Empty Distances (c) Nearest Neighbour Distances

The figure on the left (Figure 3 (a)) shows the different plots for standard error of different mark classes. We can see that near the boundaries the Standard Errors tend to be small in every class. However, near the centre they tend to increase.

In Figure 3 (b) and (c) we plot the empty distances and the nearest neighbour distances. But these distances plots do not potray a lot. Hence, we will now use the Ripley K function to assess graphically the deviations from spatial homogeniety. Next (in Figure 4 (a)) we take a look at Morishita Index Plot which does seems to suggest clustering. It is difficult to make sense of Fry Plots (Figure 4 (b) & (c)) because the number of points are so many and zoomed plot doesn't really helps.

4 Interaction

Since the amount of points is so large, we have reduced to the emergencies which were between January and June. Assuming that the assumption of *CSR* holds true, we can make plots for Ripley's K function and L plot which can help us in determining if the point process is clustered, independent or random. From both the plots (Figure 5 (a) & (b)) we can see that the $\hat{K}_{obs}$ & $\hat{L}_{obs}$ (black) are both above the $\hat{K}_{pois}$ & $\hat{L}_{pois}$ (blue), respectively, hence, the points do seem to be clustering at certain distances.



(a) (b) (c)

Figure 4: (a) Morishita Index Plot (b) Fry Plot (c) Zoomed Fry Plot

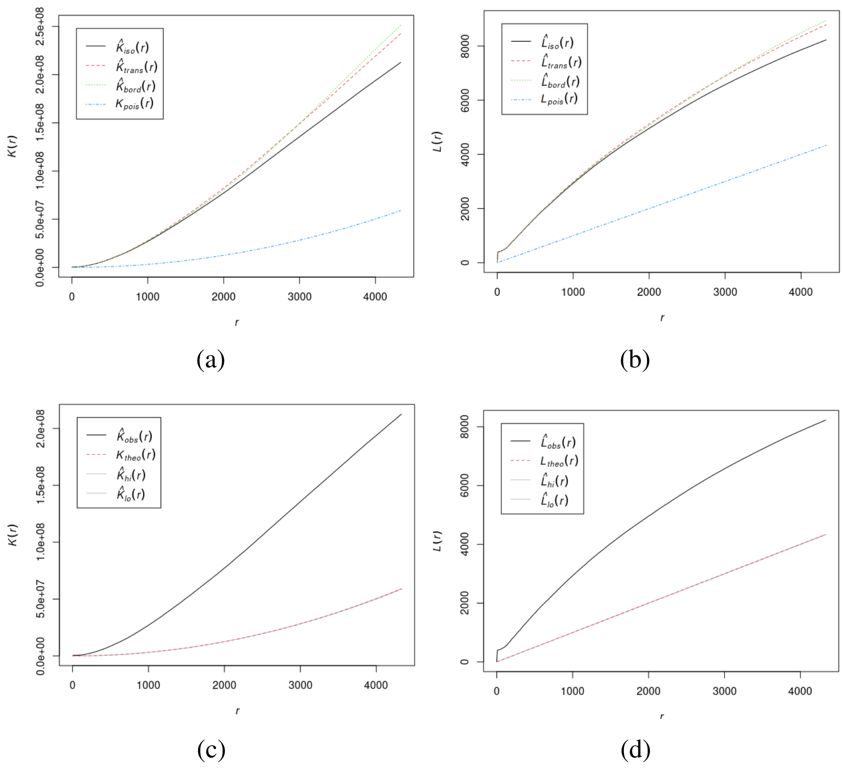
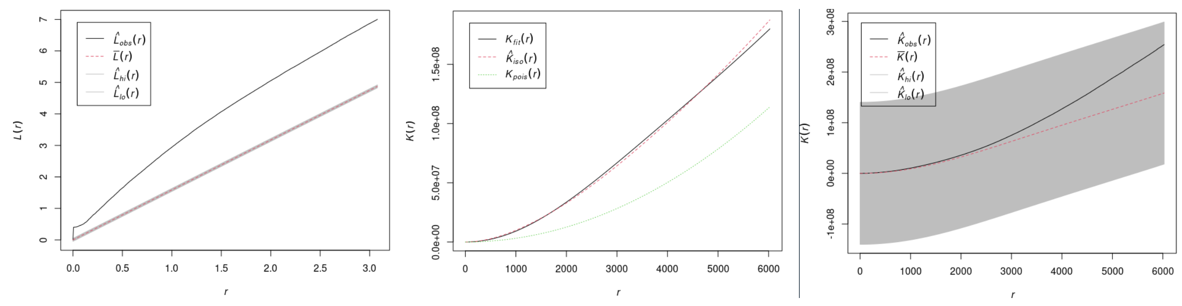


Figure 5: (a) K Plot (b) L Plot (c) K Envelope (d) L Envelope

From both the envelope plots (Figure 5 (c) & (d)) we can see that the $\hat{K}_{obs}$ & $\hat{L}_{obs}$ (black) are both above the $\hat{K}_{theo}$ & $\hat{L}_{theo}$, respectively. The *hi* and *lo* interval in the plots are very near the *theo* line. Hence, the estimated K function lies above the typical K function of a completely random pattern.

Below (Figure 6 (a)), now we plot the L plot which is adjusted for Inhomogeniety. We can still see that the observed values are still above the simulated values and the simulated interval. This might be due to clustering of points.

In Figure 6 (b), we can see that when we a non-Poisson model such as *Thomas* process, our fitted values ($\hat{K}_{fit}$) are above the poisson values ($\hat{K}_{pois}$). Hence, our point pattern is clearly different. In Figure 6 (c), we

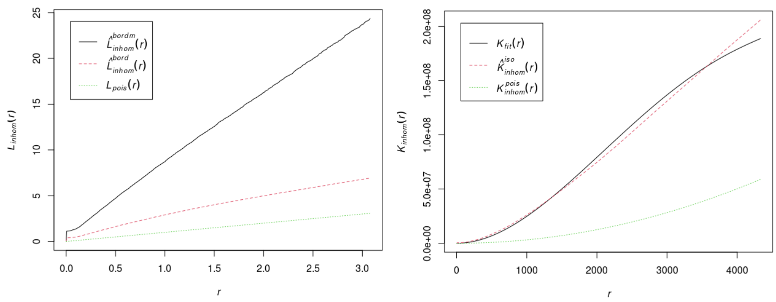


(a) (b) (c)

Figure 6: (a) L Plot adjusted for Inhomogeneity (b) K plot for a Thomas process
(c) K Envelope for a Thomas Process

can see that our fitted values lie within the confidence interval. This fits our assumption that our data process is non-Poisson.

In Figure 7 (a), we can see that even after adjusting for inhomogeneity, our fitted values are still above the poisson values. Comparing it to a non-Poisson (*Matern Cluster*) process which allows for dependence and inhomogeneity, we can see that in Figure 7 (b) our fitted values ($\hat{K}_{fit}$) are above the poisson values($\hat{K}^{pois}_{inhom}$). But they are close to Thomas process' values ($\hat{K}^{iso}_{inhom}$).



(a) (b)

Figure 7: (a) L Plot adjusted for Inhomogeneity (b) K plot for a Thomas process

5 Conclusion

Understanding Fire Emergencies by the way of spatial point analysis was quite informative. We saw that there was inhomogeneity in the data, i.e. their intensity is more in some regions than others. We also saw that the intensity of the fire emergencies also depends on the amount of workload. Also, there was definite clustering in the data. The data also follows an imhomogenous process, one of which was explored.

6 References

Bonneu F (2007). "Exploring and Modeling Fire Department Emergencies with a SpatioTemporal Marked Point Process." Case Studies in Business, Industry and Government Statistics, 1(2), 139–152

Marcon, Eric & Traissac, Stéphane & Puech, Florence & Gabriel, Lang. (2015). Tools to Characterize Point Patterns: dbmss for R. Journal of statistical software. 67. 1-15. 10.18637/jss.v067.c03.

Matt Dowle and Arun Srinivasan (2020). data.table: Extension of ‘data.frame‘. R package version 1.13.6. https://CRAN.R-project.org/package=data.table

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.