

STATISTICAL DATA ANALYSIS AND MODELING OF SKIN DISEASES

Ezzatollah Mohammadi (MSc)^{a,*}, Dr. H. Şebnem Düzgün^b

^aSuite 9, 3rd floor, Arman Building, Arman Alley, Motahari St., Tehran, Iran, mohammadi15238@itc.nl

Tel: +98 261 340 99 42, +98 21 88 44 38 86, Fax: +98 21 88 41 17 00, Mobile: +98 912 121 22 19

^bMiddle East Technical University (METU), Ankara, Turkey, Geodetic and Geographic Information Technologies (GGIT)

Tel: +90-312-210 2668, Fax: +90-312-210 5822, duzgun@metu.edu.tr

Commission VI, WG VI/1, VI/2

Key words: E-learning, statistical data, analysis and modeling, skin diseases, CSR tests

Abstract:

Skin diseases are categorized in different types and one type which we deal with is a kind of fungus infections on the skin. This spotted pattern makes their home in moist area of the body where the skin is wet, between toes, under the breasts, or on sternum skin, under the neck. In this paper we focus on to the later part of the skin which in men is exposed to the pendant ingredients in the air (polluted air).

In this research we do not go to the medicine treatment aspects and only concentrate on the spatial distribution of spotted pattern on the surface of the skin. We take the photo of infected part of the skin and assign a coordinate system to the spots in order to explore the density of them. The point pattern of the spot's spatial distribution is tested if it is random, regular, or clustered. The exploration and modeling of the point pattern could help developing appropriate treatment and amount of the medicine to be applied, especially making advises through e-media, e.g. e-mail.

As the science in every aspect is developed, the population of our planet is increased. Trading, traveling, and immigrating are the main ways of spreading diseases from city to city, state to state, country to country, and definitely continent to continent. Hence we need the easiest, fastest, and cheapest way of treatment. Certainly e-treatment can serve in this way.

Our data is a photo of skin infected by spotted pattern. We marked the spots with red circle in order to transfer the photo to software with appropriate coordinate system to read the coordinates of spots to be put in database. We analyze the spotted pattern regardless of the type of disease, even if the initiative search leads us to name the type of disease "fungus", which does not goes deep in the skin.

To visualize data, first we use quadrat density method in R. What we gain from quadrat method helps to get a general idea of the distribution and density of point pattern on the skin surface. What we perceive from this process and analysis is the intensity of point pattern on skin which is exposed to the pendant ingredients in the air.

Then we use kernel estimation to obtain a smooth estimate of probability density from the observed area and do nearest neighbor distance tests for Complete Spatial Randomness (CSR). Moreover the CSR test is applied by using K-function and L-function for the red signs to see if the observed pattern is clustered, regular or random. CSR tests indicate clustering in red signs. The most infected areas on the skin (indicating clustering) are exposed to the pendant ingredients and polluted air which normally is not covered by cloths.

Concluding, in our case study, we focus on skin diseases with the aid of E-learning in spatial data analysis and modeling for any decision to be taken by the physicians/doctors in order to reduce the time of treatment and the amount of medicine to be used. The treatment can be done form distance in a reliable way.

1- Introduction:

Skin diseases are categorized in different types and one type which we deal with is a kind of fungus infections on the skin. This point pattern makes their home in moist area of the body where the skin is wet, between toes, under the breasts, or on sternum skin, under the neck. In this paper we focus on to the

later part of the skin in which men is exposed to the pendant ingredients in the air (polluted air).

In this research we do not go to the medicine treatment aspects and only concentrate on the spatial distribution of point pattern

on the surface of the skin. We take the photo of infected part of the skin and then assign the spots to a coordinate system in order to test if the density and distribution of the point pattern is random, regular, or clustered.

Our assumption is that this could provide useful information in medicine for making decisions about the treatment and required amount of the medicine. Hence the main research questions are:

Do the sick have to use the medicine equally in all infected parts of the skin? Or in the clustered areas for instance use more pomade and in randomly scattered area use less?

Therefore, the basic aim is to investigate the spatial pattern of the skin diseases in depth with the aid of spatial data analysis and modeling techniques for any decision to be taken by the physicians/doctors for reducing the time of treatment and the amount of medicine to be used.

2- Problem definition:

As the science in every aspect is developed, the population of our planet is increased and trading, traveling, and immigrating are the main factors/elements of spreading disease from city to city, state to state, country to country, and definitely continent to continent.

We focus on point pattern on skin as it suits to spatial data analyses and visually it is exposed to the others and sick people are socio-evasive.

3- Data:

The first data is a photo of skin infected by fungus (figure 1). We marked the spots with red circle in order to transfer the photo to software with appropriate coordinate system to read the coordinates of spots to be put in the database.

We analyze the point pattern regardless of the type of disease, even if the initiative search leads us to the name and the type of fungus, which does not go deep in the skin.

4- Analyzing point pattern:

Visualization is the first step in any point pattern analyses, which is basically a dot map of the spatial phenomena. It provides developing appropriate hypothesis for the related point pattern. As can be seen from figure 2 the northern part of the study region, which is the body part close to neck indicates some clustered pattern.



Figure 1, skin diseases, spotted pattern enhanced with red circles

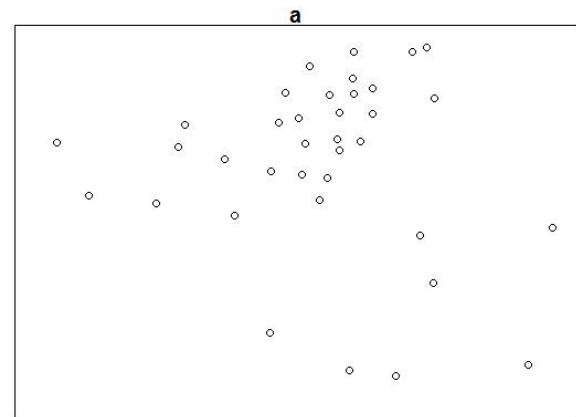


Figure 2, visualization of spotted pattern in R environment

5- Exploring the point pattern by density analyses:

The next step in spatial point pattern analyses is exploring the pattern's density. The simplest method of finding the density is quadrat analyses, where the study region is partitioned into quadrats and the number of points in each quadrat is mapped. As we see in figure 3 the densest quadrat has value 9 (9 spots) and the next dense quadrats are 6, 4, and 3. In figure 4 in down left, density is 0 (dark blue color) and as goes further to up, density increases which is enhanced with light yellow. Distribution of density in figure 4 is created based on the quadrat values in figure 3. What we perceive from this processing and analyzing is that intensity of spotted pattern on skin is in a place which is exposed to the pendant ingredients in the air, (see figure 1).

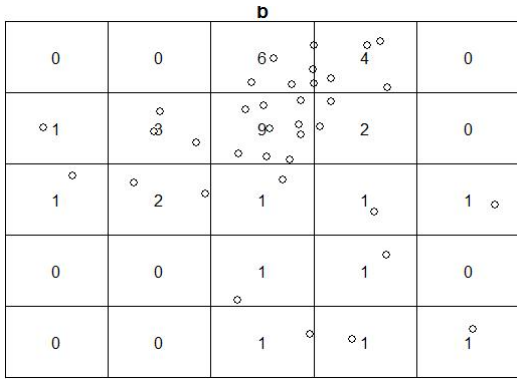


Figure 3, quadrat method, 5X5, in R

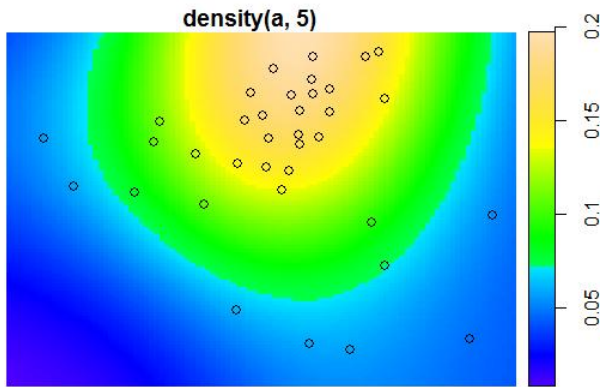


Figure 4, density of fungi (plural of fungus) on the infected skin

In figures 5 to 8 we see the described procedure in above to see the density in quadrats 3X3 and 2X2.

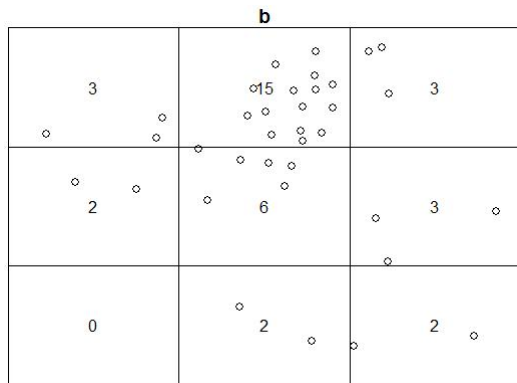


Figure 5, quadrat method, 3X3

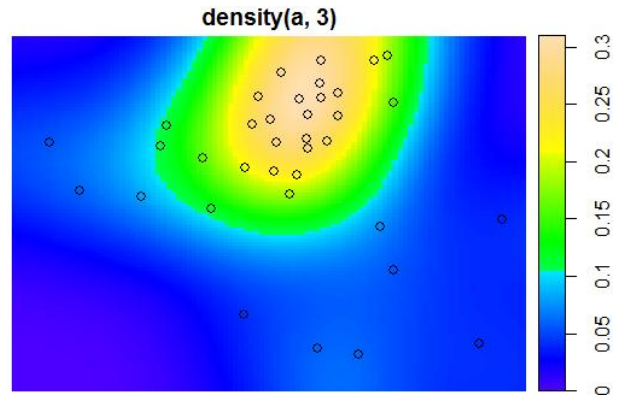


Figure 6, density of spotted pattern based on quadrat 3X3

As it is perceived from figures 3 to 8; the density of fungi are under the neck of the ill person and the more effected area is where is not covered with the cloths and is exposed to the pendant ingredients in the air and also enough light, moist, and fresh air.

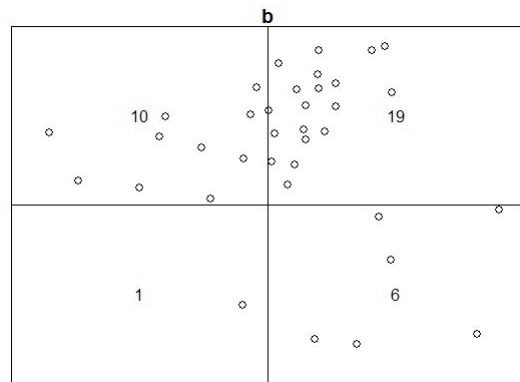


Figure 7, quadrat method, 2X2

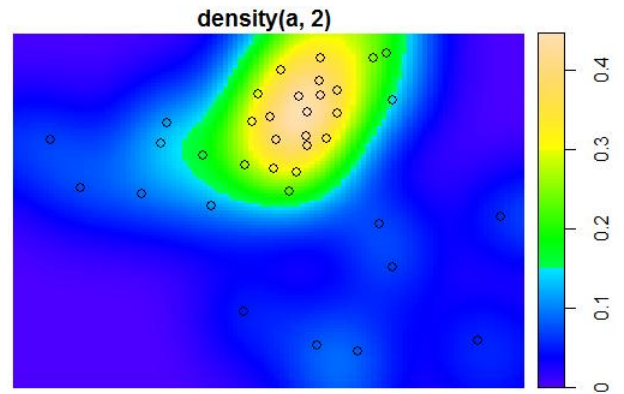


Figure 8, density of potted pattern based on quadrat 2X2

Another way of exploring the density of a point pattern is using Kernel estimation which gives a smoother estimate of density from the observed area (skin).

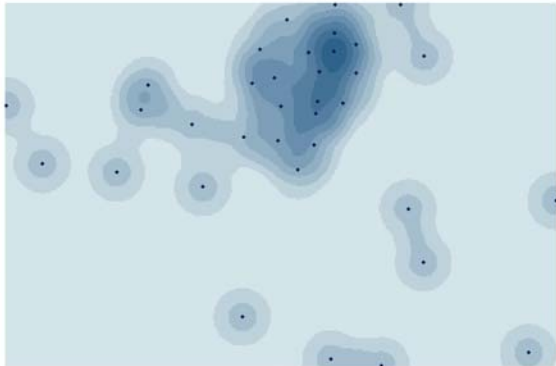


Figure 9, Kernel estimation, bandwidth 2 units and cell size 0.05 units

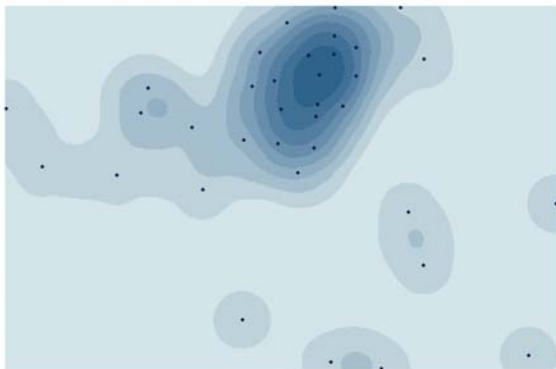


Figure 10, Kernel estimation, bandwidth 3 units and cell size 0.05 units

Figures 9, 10, and 11 are Kernel estimation for point pattern (fungus infections on the skin) with various bandwidths. Figures 9-11 suggest a cluster of spots near the neck area.

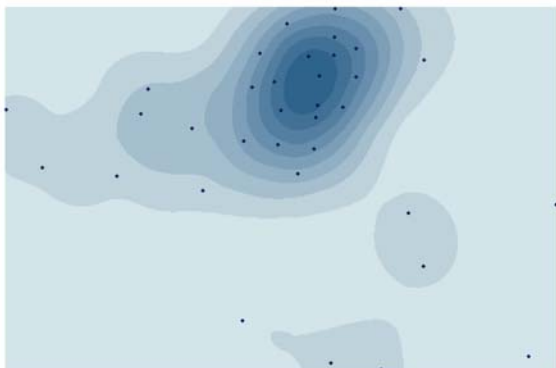


Figure 11, Kernel estimation, bandwidth 3 units and cell size 0.05 units

6- Testing the significance of the point pattern:

After exploring the point pattern, the next step is to statistically test the significance of the observed pattern.

The complete spatial randomness (CSR) forms the basis of a reference hypothesis, in which the observed pattern is tested against.

Hence CSR tests are implemented to see whether these red signs in figure 1 are clustered, regular or random. The K function and the transformed version of it called L function, is one of the tools for testing against CSR. The peaks of L function indicate clustering at the corresponding distances. The observed pattern is tested by comparing it with a random pattern created by simulations. The K and L functions for the observed pattern and the random realizations which are performed by simulations are illustrated in figures 12 and 13, respectively. In these figures, red and black curves are maximum and minimum simulated envelopes, respectively. The green curve is the estimated one. As the estimated envelope is most of the time above the simulated envelopes, especially for shorter distances, it can be said that fungi spots have a clustered pattern.

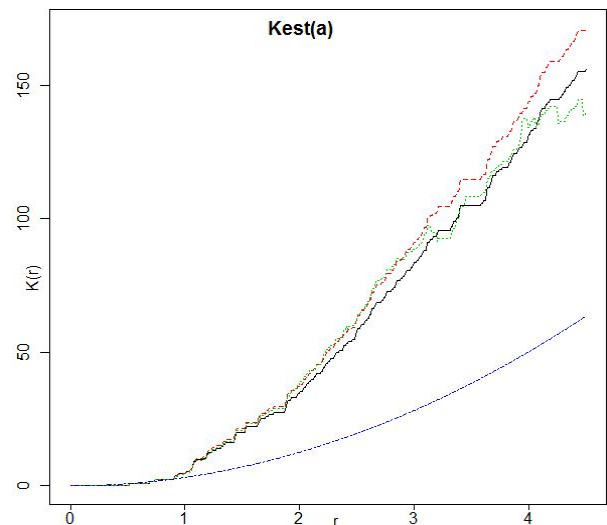


Figure 12, CSR test of K function shows clustering

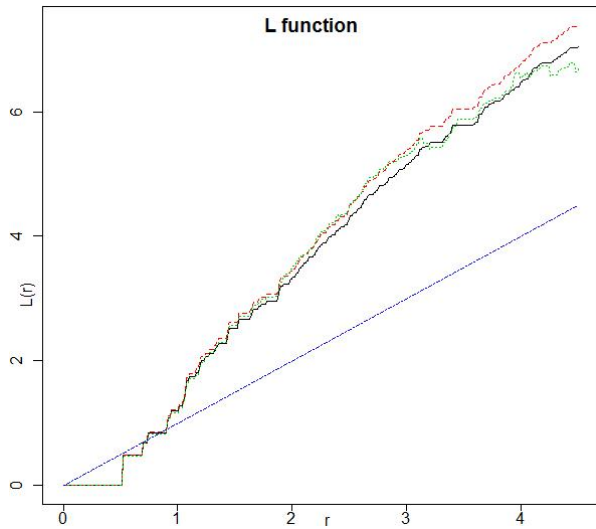


Figure 13, CSR test of L function indicates clustering

Now we go further to do more tests of point pattern distribution to get better idea of distribution of events.

According to the Figure 14, till distance 8, spotted pattern is clustered and figure 15 shows clustering to distance 7.

Here distances are discussed considering the photo scale in Figure 1.

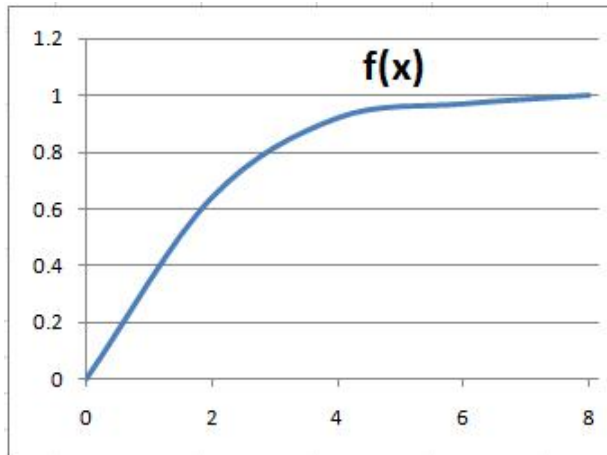


Figure 14, CSR test shows clustering

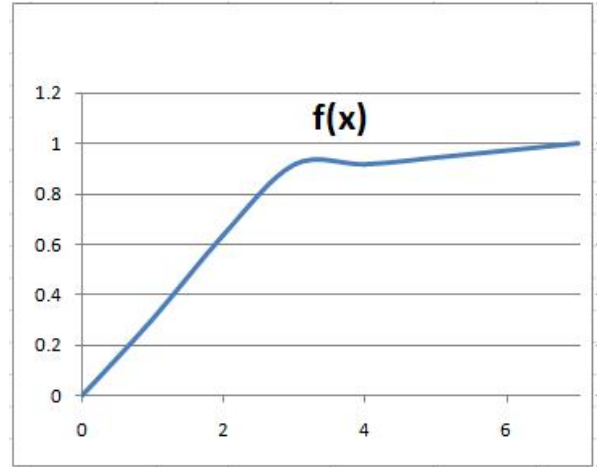


Figure 15, CSR test shows clustering

7- Conclusion:

Density analysis and CSR tests indicates clustering in the pattern of skin diseases. The most infected areas on the skin are exposed to the pendant ingredients and polluted air which normally is not covered by cloths.

Via E-Media it is possible to reduce the time of treatment and the amount of medicine to be used. The treatment can be done form distance in a reliable way (E-treatment).

8- References:

- 1- Baddeley, A., Turner, R., Spatial Point Pattern analysis, model-fitting, simulation, tests, the spatstat package, Version 1.14-9 December 17, 2008
- 2- Bailey, T. C., Gatrell, A. C., (1995), interactive spatial data analysis, Pearson Education Limited, Edinburgh Gate, Harlow, Eddex CM20 2JE, England
- 3- Düzgün, H.S.B., (2008) class's lecture notes, METU, Ankara, Turkey
- 4- Ribeiro Jr., P.J. & Diggle, P.J. (2001) geoR: A, package for geostatistical analysis. R-NEWS, Vol 1, No 2, 15-18. ISSN 1609-3631
- 4- Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., (2008) Statistical Analysis and Modeling of Spatial Point Patterns, Jhan Wiley &Son, Ltd