Chalmers-University of Gothenburg Department of Mathematical Sciences

Probability, Statistics and Risk, MVE240

Project 7

In the project description we sketch the analysis of the problem we expect you to do. (Obviously you are welcome to do more.) To pass the project a short report should be written and handed in to the project supervisor. In addition the group should present their results in class. The presentation should take about 15 minutes. Please include a short introduction which will facilitate for other students to understand the results of the project. (Do not assume that the audience knows the subject.)

1 Introduction - spatial Poisson processes

You have previously been acquainted with homogenous Poisson point processes (PPP) in time, Section 2.6.1 where PPP was called Poisson stream. Later on, in Section 7.5, homogenous PPP in space were introduced. (Homogenous means that the intensity of points is constant.) In that section the model was illustrated by two data sets; the places were V-bombs hit London during the World War II; and locations of Japanese black pines, see Figure 7.7. In examples one tested a hypothesis whether the locations tended to cluster or could be consider as random, i.e. modelled by PPP with constant intensity.

In order to check the hypothesis one divided a region into small squares, count the number of dots (bombs, trees) in each small region and then test whether the counts are Poisson distributed with the same mean. The χ^2-test was used to test the hypothesis. The average number of dots divided by the area of small region is the intensity of points.

In this project you will study (use) non-homogenous PPP, i.e when the mean of counts in small regions varies significantly. In such a case the test described above would, with high probability, lead to rejection of homogeneity (constant intensity) hypothesis.

2 Non homogenous PPP

Let consider a simple region $W = [0, a] \times [0, b]$ and denote by N_W the number of points in the region. We turn now to the "hierarchical" definition of PPP:

- Step 1: Assume that $N_W \in \mathrm{Po}(m)$ with positive but finite mean m.
- Step 2: Choose a two dimensional pdf $f(x, y) \geq 0$ for $(x, y) \in W$ and zero otherwise. Denote by (X, Y) the random variables having pdf equal to $f(x, y)$.
- Step 3: Given that $N_W = n > 0$ let place independently n points (x_i, y_i) which are obtained by independent repetitions of (X, Y). In other words the coordinates of points are independent random variables having joint pdf $f(x, y)$, $0 \leq x \leq a$ and $0 \leq y \leq b$.

The intensity of points $\lambda(x, y) = m \cdot f(x, y)$. Since $\int f(x, y)\, dx\, dy = 1$ thus $\int_W \lambda(x, y)\, dx\, dy = m$. For the homogenous PPP on W, X is uniformly distributed on $[0, a]$ while Y uniformly distributed on $[0, b]$ and X and Y are independent, $f(x, y) = 1/(a \cdot b)$ while $\lambda(x, y) = m/(a \cdot b)$.

The following script will generate a homogenous PPP. Let $a = 5$ while $b = 10$ and assume that there are in average 50 points in the region. Then

```
>> m = 50;
>> noEvents = poissrnd(m);
>> coords = rand(noEvents,2);
>> figure(1), hold on
>> a=5;b=10;
>> axis([0 a 0 b]);
>> plot(a*coords(:,1),b*coords(:,2),'.')
```

will generate a plot with homogenous PPP.

2.1 PPP with separable intensity

The simplest non-homogenous PPP is obtained by taking X and Y independent but not uniformly distributed. In such a case the intensity $\lambda(x, y) = m \cdot f(x) \cdot f(y)$. Principally any PPP with separable intensity can be obtained from the homogenous PPP by changing the scales. Compute the intensity $\lambda(x, y)$ of the PP that is simulated using the following script:

```
>> m = 50;
>> noEvents = poissrnd(m);
>> coords = rand(noEvents,2);
>> figure(1), hold on
>> a=5;b=10;
>> axis([0 a 0 b]);
>> X=a*sqrt(coords(:,1)); Y=b*coords(:,2).^2;
>> plot(X,Y,'*')
```

- Plot histograms of X, Y and compare them with $f(x)$ and $f(y)$. Conclussions.
- Propose and describe a method to estimate the intensity $\lambda(x, y)$ for separable PPP.

2.2 Analyse real data and estimate risk for failure

Suppose that X is a position of a defect, $X \in [0, 125]$ cm, (for example location of the corrosion damage along the pipeline) while Y, $Y \in [0, +\infty]$, is the severity of the damage measured by a suitable quantity. The following image presents results of laboratory investigation of deterioration of material in 1.25 meter long part od a pipeline.

```
>> i=1;
>> datafile=['rb' num2str(i,'%2d') '.txt'];
>> data=load(datafile);
>> n=length(data);
>> Y=data(:,1); X=data(:,2);
>> figure
>> plot(x,z,'*')
```

If the image can be considered as an outcome of PPP with separable intensity then, in principle, one has n independent observations of (X, Y). In addition X and Y are independent.
For the data

- Use the χ^2-test to test the independence hypothesis, see Lecture 7 for example. Here, for example one could split x in 6 classes and y in 2. You can choose other divisions.

- Propose a suitable model for the intensity $\lambda(x, y)$.
- Suppose that the true componet is 10 meter long and a risk for hazard is high when the damage indicator y exceeds level 50000 units. Use the proposed model to find distribution of the number points in a component where damage exceeded the critical level. This will require extrapolation of the estimated intensity $\lambda(x, y)$ to the region $W = [0,\, 1000] \times [50000,\, +\infty]$. Obviously one assumes that $N_W \in \mathrm{Po}(m)$, hence, what is needed is computation (estimate) of $m = \int_W \lambda(x, y)\, dx\, dy$.