MVE550 - Obligatory assignment 3

Pierre-Eliot Jourdan - Thijs Van Meurs

January 3, 2022

Question 1

a)

Let's consider a square [0, 1] × [0, 1] where trees are placed according to a spatial Poisson process with parameter $\lambda = 36$. In this question, we want to compute the probability that there are 6 trees or more in the area A = [0.2, 0.6] × [0.2, 0.6] (see figure 1 below).

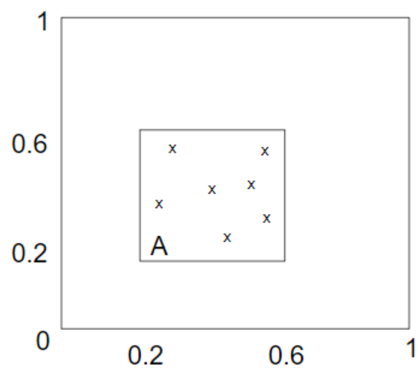

Figure 1: Square at question 1a.

Let's note T_A the random variable corresponding to the number of trees in the area A. According to the rules of Spatial Poisson processes (Lecture 9), T_A has a Poisson distribution with parameter $\lambda|A|$ with $|A| = (0.6 - 0.2)^2 = 0.16$ (the perimeter of the area A). Thus we have the following expression that we can compute using R (see R code) :

$$P(T_A \geq 6) = 1 - P(T_A \leq 5) = 1 - ppois(5; 0.16 * \lambda) = 0.5150434$$

b)

In this question, we want to compute the probability that there are exactly 4 trees in the square A = [0.2, 0.6] × [0.2, 0.6] and at the same time exactly 4 trees in the square B = [0.4, 0.8] × [0.4, 0.8].

On the previous figure, we define 3 new zones Z1, Z2 and Z3 (see figure 2 below ...):

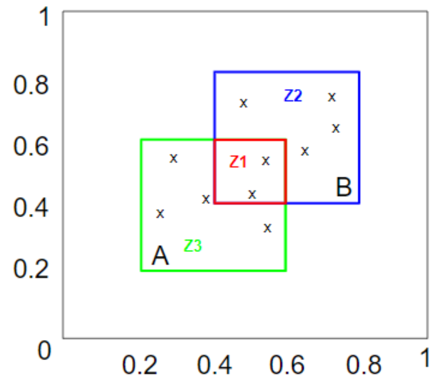

Figure 2: Square at question 1b.

We notice that : $|Z1| = (0.6 - 0.4)^2 = 0.04$ and $|Z2| = |Z3| = 3 * |Z1| = 0.12$.

Let's note T_{Z1}, T_{Z2} and T_{Z3} the random variable corresponding to the number of trees in the 3 aforementioned areas respectively. We can then apply conditional probabilities on how many trees there are in the zone Z1

$$P(T_{Z1}+T_{Z2}=4, T_{Z1}+T_{Z3}=4) = \sum_{k=0}^{4} P(T_{Z1}=k) * P(T_{Z2}=4-k|T_{Z1}=k) * P(T_{Z3}=4-k|T_{Z1}=k)$$

$$= \sum_{k=0}^{4} P(T_{Z1}=k) * (P(T_{Z2}=4-k|T_{Z1}=k))^2$$

as T_{Z2} and T_{Z3} are independent and have the same perimeter so they follow the same Poisson law with parameter $0.12 * \lambda$. Also, T_{Z1} has a Poisson distribution with parameter $0.04 * \lambda$. Finally, we can compute the probability using R:

$$P(T_{Z1}+T_{Z2}=4, T_{Z1}+T_{Z3}=4) = \sum_{k=0}^{4} dpois(k, 0.04*36) * (dpois(4-k, 0.12*36))^2 = 0.02390222.$$

c)

In this question, we wrote a R code to place trees in a $[0, 1] \times [0, 1]$ square according to a spatial Poisson process with parameter $\lambda = 36$. We used a similar procedure as the simulation of a Spatial Poisson process in chapter 6 of Dobrow. Figure 5 shows the scatter plot of $N \sim Poisson(\lambda)$ trees, which are normally distributed over the range $[0, 1]$. λ is 36 and N turned out to be 33.

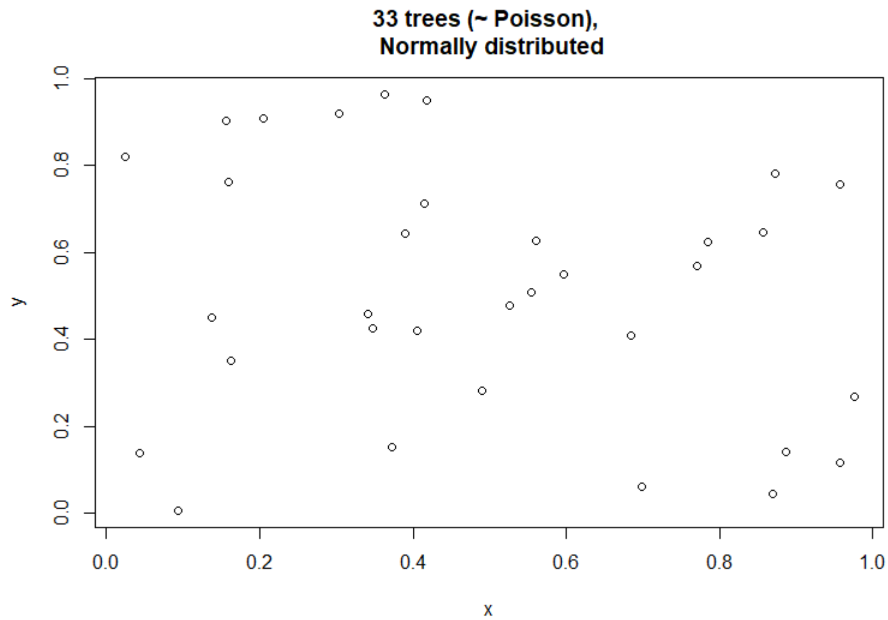

Figure 3: N trees placed in $[0,1] \times [0,1]$ grid.

d)

In this question, we assume that λ has the prior $\pi(\lambda) \propto_\lambda \frac{1}{\lambda}$. Our data are those illustrated in the Figure 1 of the instructions' pdf document. 36 trees are observed in a square of size 1. Let's determine the posterior for λ.

Like in the first question of Assignment 2, we notice that the given prior can be written as a Gamma(0,0) distribution. Then we can apply the Poisson Gamma conjugacy, we find that the posterior for λ is a Gamma(36,1) distribution.

As far as the R code is concerned, we just need to modify the initialization of λ which will now be a random value from the Gamma(36,1) distribution, rather than 36.

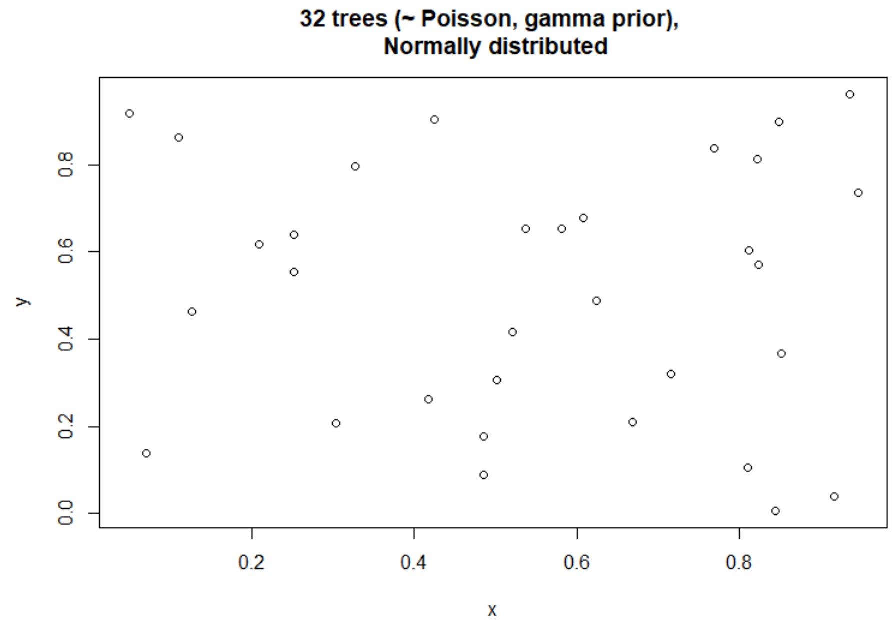

Figure 4: 33 trees placed in $[0,1] \times [0,1]$ grid, in accordance to a $Gamma(36,1)$ posterior.

e)

In this question, we implemented the random variable X as described and generated 10000 different values of λ (as done in 1d). We obtain the following histogram for this random sample (of size 10000) from the distribution of X :

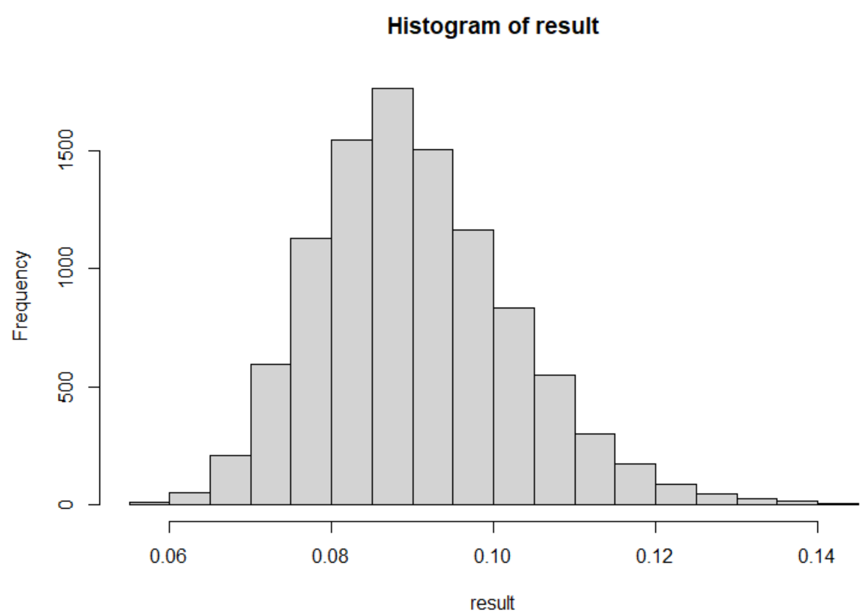

Figure 5: The histogram of 10000 simulations of random variable X

f)

The mean of our distribution from 1e is 0.09035267, with a standard deviation of 0.01214155. This entails the data from Figure 1 in the instructions is 3.74 standard deviations away from our model. This is very unlikely, and it is therefore not probable that the Poisson model has been used to simulate the data. Since the value of X is computed using the minimum distance between samples, high deviations between different sets may occur if the data points are placed at random positions. If the data points are more evenly spread out, i.e. with more consistent distances between the points, the mean value of X should be higher. This should approximate the distribution in Figure 1 of the instructions better.

Question 2

State	1	3	2	3	2	1	2	3	1	2
Duration	1.32	6.21	1.63	2.44	5.11	0.29	2.87	4.3	2.76	1.92

Table 1: The data used in Question 2 (10 consecutive states of the chain with their associated stay duration).

a)

In this question, we study a continuous-time Markov chain with the three possible states 1, 2, and 3. We will consider the following vector of parameters for the Exponential distribution : $q = (q_1, q_2, q_3) = (\frac{1}{5}, 1, \frac{1}{2})$ as well as the following transition matrix :

$$P = \begin{bmatrix} 0 & p_{12} & p_{13} \\ p_{21} & 0 & p_{23} \\ p_{31} & p_{32} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

Let's compute the infinitesimal generator matrix Q. According to Lecture 10 and the use of transition rates, we can compute the coefficients q_{ij} of the matrix Q with the following formula (p_{ij} being the coefficients of the transition matrix P) :

$$q_{ij} = p_{ij} * q_i$$

We know the matrix Q has the following form :

$$Q = \begin{bmatrix} -q_1 & q_{12} & q_{13} \\ q_{21} & -q_2 & q_{23} \\ q_{31} & q_{32} & -q_3 \end{bmatrix}$$

The previous formula gives us the following result : $(q_{12}, q_{13}) = (\frac{1}{10}, \frac{1}{10})$, $(q_{21}, q_{31}) = (\frac{1}{2}, \frac{1}{2})$ and $(q_{31}, q_{32}) = (\frac{1}{4}, \frac{1}{4})$.

Hence the expression of the infinitesimal generator matrix Q :

$$Q = \begin{bmatrix} -\frac{1}{5} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \end{bmatrix}$$

Now, let's compute the limiting distribution $\pi = (\pi_1, \pi_2, \pi_3)$ for this chain. According to Lecture 10, the limiting distribution of such a chain can be found solving the equation : $\pi * Q = 0$. And we have another condition on π : $\pi_1 + \pi_2 + \pi_3 = 1$.

We obtain the following system :

$$\frac{-1}{5}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{4}\pi_3 = 0$$
$$\frac{1}{10}\pi_1 - \pi_2 + \frac{1}{4}\pi_3 = 0$$

$$\frac{1}{10}\pi_1 + \frac{1}{2}\pi_2 - \frac{1}{2}\pi_3 = 0$$
$$\pi_1 + \pi_2 + \pi_3 = 1$$

The solution of the previous system is the limiting distribution and we have :

$$\pi = (\frac{5}{8}, \frac{1}{8}, \frac{1}{4})$$

b)

In this question, we want to use the data shown in Table 1 to learn about the parameters of the chain. For that, we will compute the joint posterior distribution for all the parameters, $q_1, q_2, q_3, p_{12}, p_{21}, p_{31}$.

As far as the priors are concerned, we use the following ones : $\pi(q_1) \propto_{q_1} \frac{1}{q_1}$, $\pi(q_2) \propto_{q_2} \frac{1}{q_2}$ and $\pi(q_3) \propto_{q_3} \frac{1}{q_3}$ so q_1,q_2 and q_3 are Gamma(0,0) distributed. We can apply the Exponential-Gamma conjugacy and obtain the following posteriors : q_1 is $Gamma(3, 1.32+0.29+2.76) = Gamma(3, 4.37)$, q_2 is $Gamma(4, 1.63 + 5.11 + 2.87 + 1.92) = Gamma(4, 11.53)$, q_3 is $Gamma(3, 6.21 + 2.44 + 4.3) = Gamma(3, 12.95)$.

We also know that p_{12},p_{21} and p_{13} are each $Beta(\frac{1}{2}, \frac{1}{2})$ distributed. Then, we can apply the Beta Binomial conjugacy and obtain the following posteriors : p_{12} is $Beta(\frac{1}{2}+2, \frac{1}{2}+3-2) = Beta(2.5, 1.5)$, p_{21} is $Beta(\frac{1}{2}+1, \frac{1}{2}+3-1) = Beta(1.5, 2.5)$, p_{31} is $Beta(\frac{1}{2}+1, \frac{1}{2}+3-1) = Beta(1.5, 2.5)$.

Hence the joint posterior distribution for all the parameters (obtained by multiplying the posteriors previously computed) :

$$\pi(q_1, q_2, q_3, p_{12}, p_{21}, p_{31}) = Gamma(q_1; 3, 4.37) * Gamma(q_2; 4, 11.53) * Gamma(q_3, 3, 12.95)$$
$$*Beta(p_{12}; 2.5, 1.5) * Beta(p_{21}; 1.5, 2.5) * Beta(p_{31}; 1.5, 2.5)$$

c)

When using the posteriors for all the values in the P matrix, we evidently miss one posterior value per row. To obtain the lacking values, one needs to simply subtract the value from 1. E.g. $p_{13} = 1 - p_{12}$. This ensures the row summing to 1 (or in the Q matrix: to 0). An example of the simulation can be found here:

$$Q_{sim} = \begin{bmatrix} -0.66035442 & 0.3860066 & 0.2743478 \\ 0.03621913 & -0.2522633 & 0.2160442 \\ 0.04171213 & 0.1269886 & -0.1687007 \end{bmatrix}$$

References