MVE550 - Obligatory assignment 1

Thijs Van Meurs - Pierre-Eliot Jourdan

December 29, 2021

Question 1

a)

The number of customers contacting Lisa's business each day is Poisson(λ), λ being our parameter (expected number of daily customers).

Our initial data is $x_1 = 4$ and $x_2 = 1$ (4 customers on Monday, 1 on Tuesday).

The prior probability density for λ is Gamma(6,2). For the computation of the posterior distribution given the data, let's apply the formula for the Poisson Gamma conjugacy (appendix C of the Bayesian inference compendium) :

The posterior $\pi(\lambda|x_1)$ is Gamma(6+x_1, 2+1) = Gamma(10;3).

The posterior $\pi(\lambda|x_1, x_2)$ is Gamma(10+x_2, 3+1) = Gamma(11;4).

The resulting posterior is then a Gamma(11;4).

Let's, now, compute the probabilities of the new values. We use the equation 1.12 of the Bayesian inference compendium (page 14). We have :

$$\pi(y_{new}|y) = \frac{\pi(y_{new}|\lambda)\pi(\lambda|y)}{\pi(\lambda|y_{new}, y)} = \frac{e^{-\lambda}\lambda^{y_{new}}/y_{new}! * 4^{11}\lambda^{10}e^{-4\lambda}/\Gamma(11)}{5^{11+y_{new}}\lambda^{11+y_{new}-1}e^{-5\lambda})/\Gamma(11+y_{new})} = \frac{4^{11}\Gamma(11+y_{new})}{5^{11+y_{new}}\Gamma(11)y_{new}!}$$

Then, we can compute the predicted probabilities for getting 0,1,2...,10 or more customers on Wednesday :

$$\pi(0|y) = 0.0859$$

$$\pi(1|y) = 0.1890$$

$$\pi(2|y) = 0.2268$$

$$\pi(3|y) = 0.1965$$

$$\pi(4|y) = 0.1376$$

$$\pi(5|y) = 0.0825$$

$$\pi(6|y) = 0.0440$$

$$\pi(7|y) = 0.0213$$

$$\pi(8|y) = 0.0096$$

$$\pi(9|y) = 0.0041$$

$$\pi(10|y) = 0.0016$$

$$\pi(>10|y) = 1 - \sum_{i=0}^{10} \pi(i|y) = 0.0011$$

(We used the fact that for all integer $n > 0, \Gamma(n) = (n-1)!$).

b)

In this question, we consider $y_{new} = 3$ (3 customers on Wednesday). We are going to apply a discretization of the λ variable to re-compute the exact probability for getting exactly 3 customers on Wednesday.

We know that : $\pi(y_{new} = 3|\lambda)$ is $Poisson(2;\lambda)$, $\pi(\lambda)$ is $Gamma(6;2)$ and $\pi(\lambda|y_{new} = 3)$ is $Gamma(6 + y_{new}; 2 + 1) = Gamma(9;3)$.

We will apply the discretization method seen in the section 1.5 of the Compendium (see R code). Let's generate $k = 1000$ values for λ in the interval [0;7] (the interval was chosen according to the observation of the Gamma(6,2) distribution).

The prior density density $\pi(\lambda)$ can reasonably be approximated by a discrete distribution on $\lambda_1, ..., \lambda_k$ specified by :

$$a_i = Pr(\lambda = \lambda_i) = \frac{\pi(\lambda_i)}{\sum_{j=1}^{k} \pi(\lambda_j)} = \frac{Gamma(\lambda_i; 6, 2)}{\sum_{j=1}^{k} Gamma(\lambda_j; 6, 2)}$$

Let's note $b_i = \pi(y|\lambda_i) = Poisson(4, \lambda_i) * Poisson(1, \lambda_i)$ for i=1,...,k we can then approximate the posterior with a discrete distribution on $\lambda_1, ..., \lambda_k$ specified by :

$$c_i = Pr(\lambda = \lambda_i|y) = \frac{a_i * b_i}{\sum_{j=1}^{k} a_j * b_j}$$

Finally, for a specific value of y_{new} ($y_{new} = 3$), we may approximate the predictive distribution as

$$\pi(y_{new} = 3|y) = \sum_{i=1}^{k} \pi(y_{new} = 3|\lambda_i) * Pr(\lambda = \lambda_i|y) = \sum_{i=1}^{k} \pi(y_{new} = 3|\lambda_i) * c_i = \sum_{i=1}^{k} Poisson(3, \lambda_i) * c_i = 0.2071.$$

(We computed this result using R).

c)

In this question, instead of discretizing the parameter λ one may apply numerical integration (formula used in section 1.6 of the compedium) :

$$\pi(y_{new} = 3|x_1 = 4, x_2 = 1) = \frac{\int_0^7 \pi(\lambda|y_{new} = 3) * \pi(\lambda|x_1 = 4, x_2 = 1) * \pi(\lambda), d\lambda}{\int_0^7 \pi(\lambda|x_1 = 4, x_2 = 1) * \pi(\lambda), d\lambda}$$

$$= \frac{\int_0^7 Poisson(3;\lambda) * Poisson(4;\lambda) * Poisson(1;\lambda) * Gamma(6;2;\lambda), d\lambda}{\int_0^7 Poisson(4;\lambda) * Poisson(1;\lambda) * Gamma(6;2;\lambda), d\lambda} = 0.1966.$$

The bounds of the integral should be from 0 to infinity. Yet, when we have a look at the density of Gamma(6;2), we notice that almost every generated λs take their value in [0;7]. Hence the choice we made for computing this probability.

Thus, we notice that the result found is closer to the first value found in 1)a) using numerical integration than using discretization of the parameter λ.

d)

In this question, we now consider that the prior density used is a normal density with expectation 3 and standard deviation 2 cut off ($\mu = 3$ and $\sigma = 2$). The prior density (that equals 0 for negative values) is thus :

$$f(x) = \frac{1}{2\sqrt{2\pi}} * e^{-\frac{(x-3)^2}{8}}, 0 < x < +\infty.$$

Let's re-compute the predicted probability of getting exactly 3 customers on Wednesday when using this prior. Like in the previous question, we can apply numerical integration to compute this probability (more accurate) :

$$\pi(y_{new}|y) = \frac{\int_0^6 \pi(\lambda|y_{new}=3) * \pi(\lambda|x_1=4, x_2=1) * f(\lambda), d\lambda}{\int_0^6 \pi(\lambda|x_1=4, x_2=1) * f(\lambda), d\lambda}$$

$$= \frac{\int_0^6 Poisson(3;\lambda) * Poisson(4;\lambda) * Poisson(1;\lambda) * f(\lambda), d\lambda}{\int_0^6 Poisson(4;\lambda) * Poisson(1;\lambda) * f(\lambda), d\lambda} = 0.1908.$$

The prior density equals 0 for negative values and is re-scaled so that the integral below the curve is 1. Again, we notice observing the normal density f that almost every generated λs take their value in [0;6]. Hence our previous computation.

e)

See R code (file 1.e).

We find that the proportion of λs for which the corresponding simulated values for Monday and Tuesday are equal to 4 and 1 is 0.0192.

We find that the proportion of λs for which the corresponding simulated values for Monday, Tuesday and Wednesday are equal to 4, 1 and 3 is 0.003814.

Question 2

In this exercise (question 3.52 in the course literature [Dob16]), we play the game "Snakes and Ladders" on a 3x3 grid with a four-faced die. The lay-out of the playing field is illustrated in figure 1. The rules to this game are relatively simple. If you land on a square which includes the bottom of the ladder, you transition to the square with the top of that ladder. If a square has the top of a 'snake', than in instantly transition to the square with the bottom of the snake. Square 9 is the final square; if you land on this square, you win. A player can only win if he or she rolls the exact number needed to transition to square 9.

a)

The first question regarding the game, is to determine the expected length of the game. In order to do so, we are to compute the transition matrix P before continuing. The matrix can be found here:

$$P = \begin{bmatrix}
0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\
0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

The matrix P includes absorbing state 9, which will not be needed for our calculation. Therefore we will make matrix Q, which excludes state 9.

$$Q = \begin{bmatrix}
0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 \\
0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}$$

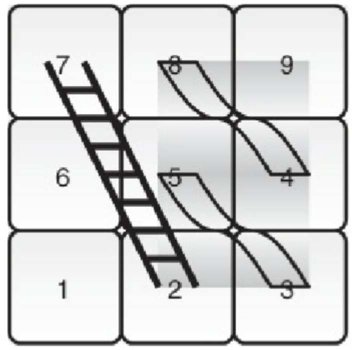

Figure 1: The game 'Snakes and ladders' as is used in question 2. This is a snippet from the literature [Dob16].

An important rule to take into consideration in this game is that the finishing square can only be reached by an exact roll of die otherwise the player stay put on the same square.

According to the shape of our matrix, we can consider the state 9 as the absorbing state in order to extract the matrix Q corresponding to the transient states :

After that, let's compute the fundamental matrix F using RStudio (see code). We know that $F = (I - Q)^{-1} = lim_{n->+\infty} I + Q + ... + Q^n$. Hence :

$$F = \begin{bmatrix} 1 & 0.25 & 0 & 1.21875 & 2.15626 & 0 & 1.125 & 2.875 & 0 \\ 0 & 1.00 & 0 & 1.37500 & 2.12500 & 0 & 1.1666667 & 2.833333 & 0 \\ 0 & 0 & 1 & 0.5 & 1.5 & 0 & 0.666667 & 3.33333 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 & 1.33333 & 2.666667 & 0 \\ 0 & 0.00 & 0 & 1.00000 & 3.00000 & 0 & 1.3333333 & 2.666667 & 0 \\ 0 & 0.00 & 0 & 2.00000 & 2.00000 & 1 & 1.3333333 & 2.666667 & 0 \\ 0 & 0.00 & 0 & 0.50000 & 1.50000 & 0 & 2.0000000 & 2.000000 & 0 \\ 0 & 0.00 & 0 & 0.50000 & 1.50000 & 0 & 0.6666667 & 3.333333 & 0 \\ 0 & 0.00 & 0 & 1.00000 & 3.00000 & 0 & 1.3333333 & 2.666667 & 1 \end{bmatrix}$$

Finally, according to the lecture on absorbing chains (Lecture 4), we know the expected number of steps until absorption is given by the vector $F1^t$. The game starts in state 1, we conclude that the expected length of the game corresponds to the first coefficient of the vector $F1^t$, that is to say :

$$\sum_{j=1}^{k} F_{tj} = 8.625$$

(with initial state $t = 1$ and k number of states). The length from start to finish (state 1 to state 9) is expected to take 8.625 steps.

b)

In this second part of the exercise, we assume that the player is on square 6. Let's find the probability that they will find themselves on square 3 before finishing the game (and being absorbed by state 9). Let's also consider state 3 as an absorbing state. The new matrix Q of the transient states will then look like :

$$Q = \begin{bmatrix} 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and the matrix R will be :

$$R = \begin{bmatrix} \frac{1}{4} & 0 \\ \frac{1}{2} & 0 \\ 0 & 0 \\ \frac{1}{4} & 0 \\ 1 & 0 \\ 0 & \frac{1}{4} \\ 0 & \frac{1}{4} \\ 0 & 0 \end{bmatrix}$$

Let's compute the new fundamental matrix obtained (with RStudio):

$$F = (I - Q)^{-1}$$

The previous computations lead us to compute the absorption probabilities matrix :

$$FR = \begin{bmatrix} 0.609375 & 0.390625 \\ 0.6875 & 0.3125 \\ 0.25 & 0.75 \\ 0.50 & 0.50 \\ 1.00 & 0.00 \\ 0.25 & 0.75 \\ 0.25 & 0.75 \\ 0.50 & 0.50 \end{bmatrix}$$

Thus the probability of players finding themselves in state 3 before finishing the game if they are at state 6 is 0.25.

(It corresponds in the matrix to the (6;1) coefficient).

References

[Dob16] Robert P. Dobrow. *Introduction to Stochastic Processes With R*. John Wiley Sons, Ltd, 2016.