MA 424 - SIMULATIONS PROJECT

14-04-2022

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

SUBMITTED BY
CANDIDATE ID - 39719

Table of Contents

Acknowledgement

I hereby declare that the report presented here is an authentic record of my own work. I have though referred to code provided during seminars and in the solutions to the mock project in order to build my solutions. I used code from seminars in Question 2 for Acceptance/Rejection method and MCMC algorithm. Further, I referred to code in mock solutions for arriving at an estimate for the second and third part in Question 1 and it has been acknowledged in the code as well.

Overview

Carsafe is an insurance company dealing in automobile insurance. They are looking to analyze their downside risk by analyzing their customer's arrival, departure and accident patterns. Their analysts have come up with these numbers and have asked to estimate the probability of ending up with a capital amount of less than £30000 at the end of the year. Carsafe also wants an estimate of the capital at the end of 12 months, along with a 95% confidence interval of the estimate.

Question 1 Part (I) – Carsafe in the next 12 months

(I) We start by defining our variables :

- **t_A -** time of next arrival of customer (*value between 0 and 12*)
- **t_D -** time of departure for each customer (*value between their arrival time and 12*)
- **t_X -** time of next accident for each customer (*value between their arrival and departure*)
- **N_A** - number of arrivals
- **N_D** - number of departures
- **N_X** - number of accidents encountered by Carsafe customers during the year

To keep track of above events we store our event lists and state-time pairs as follows :

State-time pairs : The states that define our system our arrival and departure, and we store them as a (nx2) matrix

Event List : (T_A,T_D) , it keeps track of arrival and departure time for each of the customer that ever becomes a system of Carsafe.

The following cases affect the variables above –

Case 1 : Arrival of Customer in Customer

Case 2 : Departure of Customer from Carsafe

Case 3 : Occurrence of an accident for a customer

Arrivals occur with a Poisson process, with rate $\lambda = 3$. The departures follow a non-homogeneous Poisson process, with the departure times depending on the time elapsed. Hence, we use the arrival times to generate the corresponding departure times. If the departure time generated is more than 12, we set it equal to 12 (assuming the customer does not leave in the given time period).

The pseudocode for each event case is as follows –

Pseudocode -

Initialize :

1. Set t = 0
2. Set N_A , N_D & N_X = 0
3. Set n = 0
4. Set claim = 0

Case 1 : Arrival of customer

1. Using t = 0, generate a uniform random variable (say, U) and set t = t - 1/ λ *log(U)
2. Set N_A = N_A + 1
3. Set n = n + 1
4. Record the arrival time in vector T_A
5. Set t = t_A for generation of the next arrival time

Case 2 : Departure of customer

1. Using arrival time of each customer, generate a Poisson random variable (say, t) as above and a uniform random variable (say, U2). Compare U2 with ratio mu_f (t) / mu
2. If U2 <= above ratio, then we accept the value for departure (t_D)
3. If t_D <12
   a. Set N_D = N_D + 1
   b. Set n = n - 1
4. Else set t_D = 12 (assuming the customer does not depart)
5. Record the departure time in vector T_D

Case 3 : Occurrence of accident

1. Using time of arrival of each customer, generate a Poisson random variable as in case 1 to arrive at an accident time t_X
2. Compare t_X with the corresponding departure time t_D, if t_X < t_D then,
   a. N_X = N_X + 1
   b. Record t_X in vector T_X
   c. Continue generating accidents, repeating a) and b) as above till t_X >t_D
3. If t_X > t_D, move to next customer and generate its first accident time

Final Calculations :

1. Calculate time_in_system = (T_D - T_A)
2. Calculate revenue = sum((time_in_system) * 300)
3. We calculate claim as follows :
   a. Generate a uniform random variable(say, P).
   b. If P<0.6, generate a random integer (say, X) between 0 and 10 using floor(runif(1)*11)
   c. Calculate claim_amt = 300*X + 500
   d. If P>0.6, claim_amt = 0
4. We calculate amount at the end of the year = c_0 + revenue – claim, where c_0 is initial capital = 50000
5. Store this value of amount in amount_tuple

We implement the above in R and run for K = 1000 iterations, and in each iteration, we count the number of times the capital at the end of the year goes below 30,000. We count these occurrences and divide by K to get **probability = 0.169**.

Question 1 Part (II) – Carsafe in the next 12 months

If the sample standard deviation of the estimator $\overline{\boldsymbol{amount}}$ is d, then we have: $\boldsymbol{Pr}\{|\ \overline{\boldsymbol{amount}} - \boldsymbol{E}(\boldsymbol{amount})| > \boldsymbol{c} * \boldsymbol{d}\ \} = \mathbf{2}\ (\mathbf{1} - \ \boldsymbol{\Phi}(\mathbf{c}))$ from the central limit theorem.

We want $\mathbf{2}(\mathbf{1} - \boldsymbol{\Phi}(\mathbf{c})) = \mathbf{0.1}$ and $\mathbf{c} \times \mathbf{d} = \mathbf{500}$. From the first equality we get $\mathbf{c} = \mathbf{1.65}$ and from the second we get $\mathbf{d} = \frac{\mathbf{500}}{\mathbf{1.65}}$. Thus, we want a sample standard deviation of at most $\mathbf{d} = \mathbf{303.03}$

To do this we run the simulation for at least 100 iterations (for the central limit theorem to work) and after that until we get $\frac{\boldsymbol{S_k}}{\sqrt{\boldsymbol{k}}} < \boldsymbol{d}$ where $\frac{\boldsymbol{S_k}}{\sqrt{\boldsymbol{k}}}$ is the sample s.d. of the estimator after k iterations.

The estimator of $\mathbf{E}(\mathbf{amount})$ here is $\overline{\boldsymbol{amount}} = \mathbf{41515.2739}$. The sample standard deviation of the estimator is $\frac{\boldsymbol{S}}{\sqrt{\boldsymbol{k}}} = \mathbf{302.8948}$ where S is the sample standard deviation and k is the number of iterations. The number of iterations that I run for this was $\boldsymbol{k} = \mathbf{1572}$.

Question 1 Part (III) – Carsafe in the next 12 months

We know from lecture that with probability $\mathbf{1} - \boldsymbol{\alpha}$ the population mean $\overline{\boldsymbol{amount}}$ lies within the interval

$$\overline{\boldsymbol{amount}} \pm \boldsymbol{Z}_{\alpha/2} \frac{\boldsymbol{S_k}}{\sqrt{\boldsymbol{k}}}$$

where is $\frac{\boldsymbol{S_k}}{\sqrt{\boldsymbol{k}}}$ the sample s.d. of the estimator after k iterations. Here, $\boldsymbol{\alpha} = \mathbf{0.05}$ and $\boldsymbol{Z}_{\mathbf{0.025}} = \mathbf{1.96}$ and thus the interval is $\mathbf{41515.2739} \pm \mathbf{1.96} * \mathbf{302.8948}$

Thus, we are **95%** confident that **E(amount)** is in the interval $[\mathbf{40921.6001}, \mathbf{42108.9476}]$.

Question 2 – Two other models to simulate claims

(a) Team 1 –

(I) Proposal distribution used – Uniform random distribution over the set {0,1,2,...,10}. This is appropriate because this distribution shares the same domain as the one proposed by the first team, which is {0,1,2...10}

(II) For a random variable $\boldsymbol{i}$ in {0,1,2,...10}, let $\mathbf{Q}(\mathbf{X} = \mathbf{i}) = \frac{\mathbf{1}}{\mathbf{11}}$ and $\mathbf{P}(\mathbf{X} = \mathbf{i}) = \mathbf{Se^{-i}}$. We have an efficient method for simulating a random variable from distribution $\boldsymbol{Q}$. Let $\boldsymbol{c}$ be a constant such that -

$$\frac{\boldsymbol{p_i}}{\boldsymbol{q_i}} \leq \boldsymbol{c} \;\; \textit{for all } \boldsymbol{i} \textit{ such that } \boldsymbol{p_i} > \mathbf{0}$$

As each iteration independently results in an accepted value with probability $\frac{\mathbf{1}}{\boldsymbol{c}}$, we see that the number of iterations needed is geometric with mean $\mathbf{c}$. We try to maximize this value of $\mathbf{c}$, in order to accept more values, hence the number of iterations needed to run before accepting a value. Hence, $\mathbf{c} = \mathbf{max}\frac{\boldsymbol{p_i}}{\boldsymbol{q_i}}$

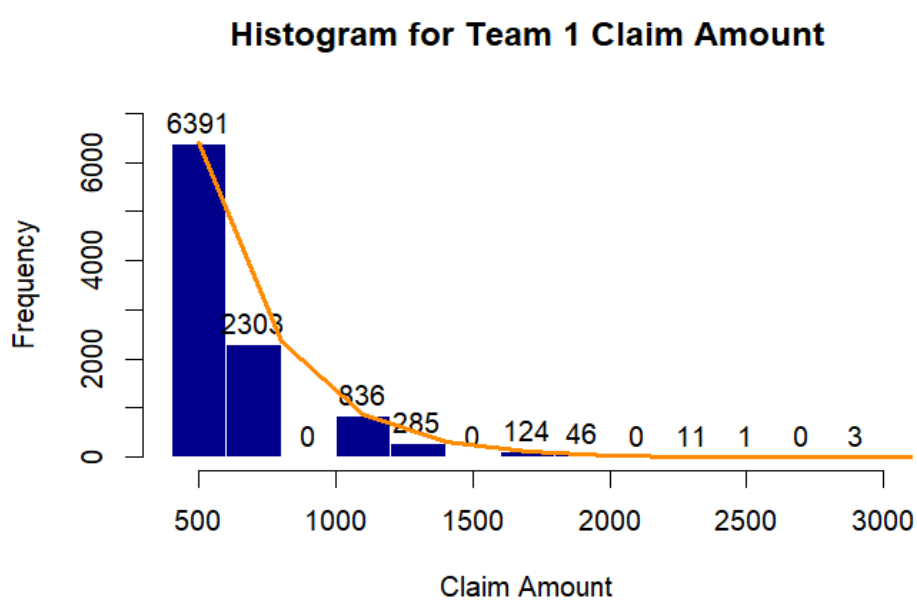
Also, we know that $\sum \boldsymbol{Se^{-i}} = \mathbf{1}, \;\; \textit{for all } i \in \{0,1,2,\dots 10\}$

Solving for $\mathbf{S}$ we get, $\boldsymbol{S} = \mathbf{0.6321311}$
Finally, solving for $\mathbf{c}$, we get $\mathbf{c} = \mathbf{6.95344}$

(III) *See file Q2_a_iii*

1. We generate a random variable (say, Y) in {0,1,2...10} using runif() function
2. Our acceptance ratio in this case is exp(-x)
3. We generate another uniform random variable (say, U2) and compare this with the acceptance ratio.
4. If U2 <= acceptance ratio, we accept the value else, we go back to step 1

(IV) *See file Q2_a_iv*



(b) Team 2 –

(I) MCMC Algorithm used – We use the **Metropolis Hastings** algorithm to simulate the claim amount. This is because the MH chain proposes where to move next using the initial transition probabilities, then accepts the recommendation with probability $\frac{s_j}{s_i}$, staying in its present state if it is rejected. Here, s_j *and* s_i stationary distributions.

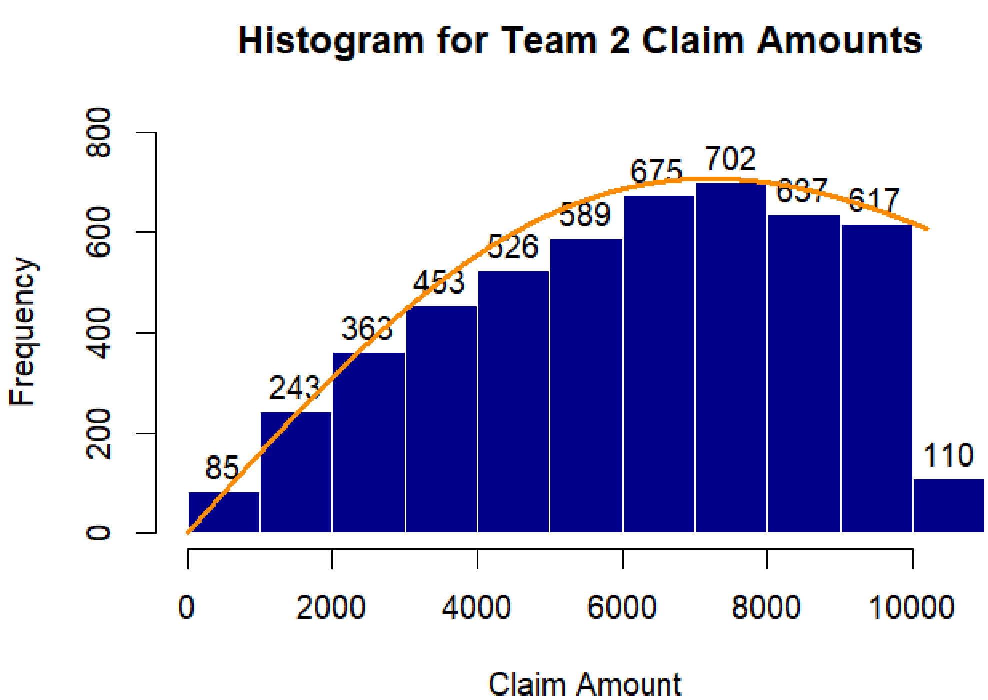
(II) Proposal chain used – Uniform random distribution over the domain $(0,1)$. This is appropriate because this distribution shares the same domain as the one proposed by the second team. That is, the proposed state on the interval $(0, 1)$ is always a fresh Uniform random variable in $(0, 1)$, independent of the current state.

(III) *See file Q2_b_iii*

1. We run a Markov Chain for 10000 states, and start with n = 0 as the first stage
2. We generate a current state of the Markov chain by generating a uniform random variable
3. For n<10000, we propose a new state X_{prop} and we calculate the acceptance probability of going into this state and comparing it with another uniform random variable (say, U).
4. The probability of acceptance is defined as $prob_{accept} = \min(acceptance\ ratio, 1)$ where,

$$acceptance\ ratio = \frac{X_{prop} * exp\left(-(X_{prop}^2)\right)}{X_{current} * exp\left(-(X_{current}^2)\right)}$$

(IV) See file Q2_b_iv



Question 3 Reducing variance of the estimator

Our estimate of the amount at the end of the year $\boldsymbol{E}(\boldsymbol{C}_{12})$ depends on three variables mainly, the time of arrival, the time of departure and the accident time. We can use control variates to reduce the estimate of our amount. Since departures are directly dependent on each arrival time, we can generate them such that they are negatively correlated to the arrival time and thus reduce the variance of our estimate. Further, the number of times a customer undergoes an accident is also directly impacted by the duration that they stay in Carsafe, hence, a similar approach can also be adopted with accident times to reduce the variance of $\boldsymbol{E}(\boldsymbol{C}_{12})$.