STAT3004 – Probability Models & Stochastic Processes

Project 1

Student Name: Louis Yang

Due Date: 29/04/2020

Questions/Tasks:

1. Equation (4.1.4) represents the expected numbers in the Greenwood model.

(a) Derive these equations.

In the Greenwood model, it is assumed that the cause of infection is not related to the number of infectives so that α can be regarded simply as the probability of non-infection. In this case the number of infectives y_t at time t is determined by x_{t-1} and x_t as $y_t = x_{t-1} - x_t$.

As such, for a Greenwood model equation (4.1.2):

$$\mathbb{E}[X_{t+1}|X_t] = \alpha X_t$$

Remembering that for random variables X and Y:

$$\mathbb{E}X = \mathbb{E}\mathbb{E}[X|Y]$$

As a result, by taking expected values on both sides:

$$\mathbb{E}X_{t+1} = \alpha \mathbb{E}X_t$$

Iterating this gives:

$$\mathbb{E}[X_t|X_0 = x_0] = \alpha^t x_0$$

On the other hand:

$$Y_t = X_{t-1} - X_t$$

Accordingly, by taking expected values on both sides:

$$\mathbb{E}Y_{t+1} = \alpha \mathbb{E}(X_{t-1} - X_t)$$

Expanding:

$$\mathbb{E}Y_{t+1} = \alpha \mathbb{E}X_{t-1} - \alpha \mathbb{E}X_t$$

Iterating this gives:

$$\mathbb{E}[Y_t|X_0 = x_0] = \alpha^{t-1} x_0 - \alpha^t x_0$$

Simplifying:

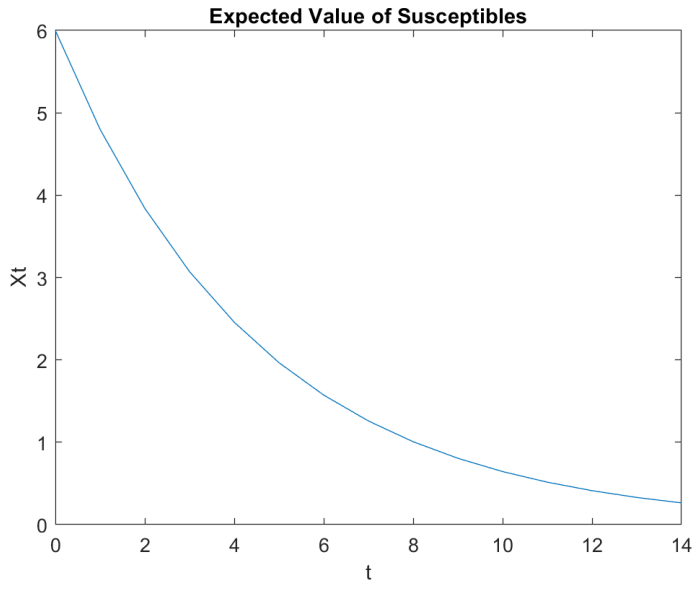
$$\mathbb{E}[Y_t|X_0 = x_0] = \alpha^{t-1}(1-\alpha) x_0$$

(b) Assume that $x_0 = 6$ and $\alpha = 0.8$. Then plot these expected values for some sensible time horizon.

Taking time t to be equal to 14 (a sensible time horizon as it represents the number of days a person is required to be in quarantine), a plot can be produced regarding the two equations (4.1.4). This is shown below:

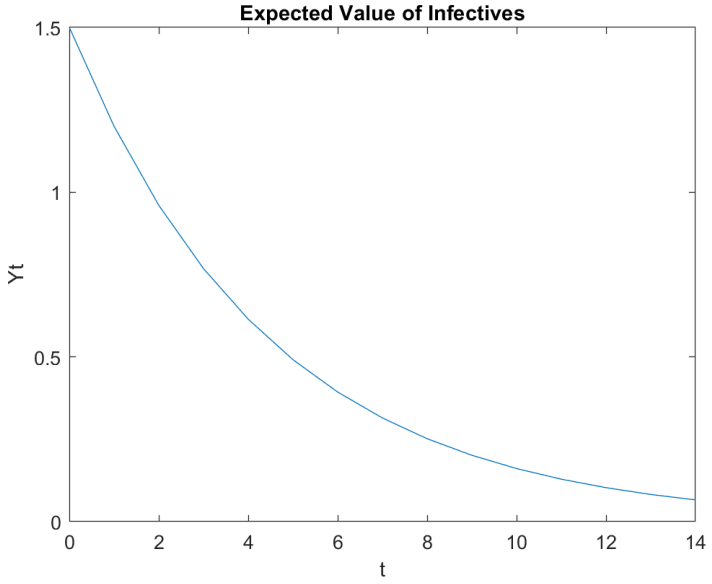
$$\therefore \mathbb{E}[X_t|X_0 = x_0] = \alpha^t x_0$$

$$\therefore \mathbb{E}[X_t|X_0 = x_0] = 0.8^t 6 \text{ for } 0 \leq t \leq 14$$



$$\therefore \mathbb{E}[Y_t|X_0 = x_0] = \alpha^{t-1}(1-\alpha)x_0$$

$$\therefore \mathbb{E}[Y_t|X_0 = x_0] = 0.8^{t-1}(1-0.8)6 \text{ for } 0 \leq t \leq 14$$



---

2. Equation (4.2.1) presents a recursion for the expected number of susceptibles and infected in the Reed-Frost model.

(a) Derive these equations.

In the Reed-Frost model, an individual susceptible at time t is still susceptible at time $t + 1$ only if contact with all Y_t infectives is avoided (more pointedly, if infections contact is avoided). This event has probability α^{Y_t}, independently for each of the X_t individuals susceptible at t.

This can be summarised such that:

$$X_{t+1} = \text{Bin}(X_t, \alpha^{Y_t})$$

Hence:

$$\mathbb{E}[X_{t+1}|(X,Y)_t = (x,y)_t] = \mathbb{E}X_t\alpha^{Y_t}$$

Iterating this gives:

$$\mathbb{E}[X_{t+1}|(X,Y)_t = (x,y)_t] = x_t\alpha^{y_t}$$

In contrast:

$$Y_t = X_{t-1} - X_t$$

Moreover:

$$Y_{t+1} = \text{Bin}(X_t, (1-\alpha^{Y_t})$$

Thus:

$$\mathbb{E}[Y_{t+1}|(X,Y)_t = (x,y)_t] = \mathbb{E}X_t(1-\alpha^{Y_t})$$

Iterating this gives:

$$\mathbb{E}[Y_{t+1}|(X,Y)_t = (x,y)_t] = x_t(1-\alpha^{y_t})$$

Referring to equation (4.1.3), a deterministic analogue can then be presented:

$$\mathbb{E}[(X,Y)_{t+1}|(X,Y)_t = (x,y)_t] = (\mathbb{E}X_t\alpha^{Y_t}, \mathbb{E}X_t(1-\alpha^{Y_t}))$$

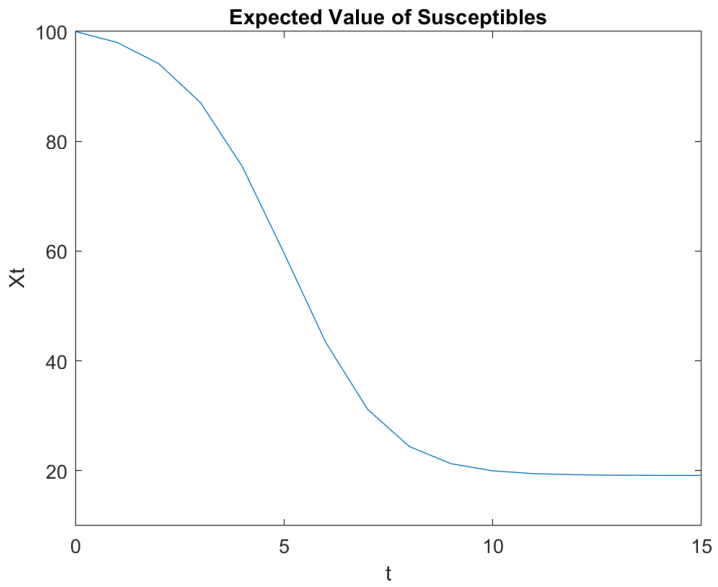
Iterating this gives:

$$\mathbb{E}[(X,Y)_{t+1}|(X,Y)_t = (x,y)_t] = (x_t\alpha^{y_t}, x_t(1-\alpha^{y_t}))$$

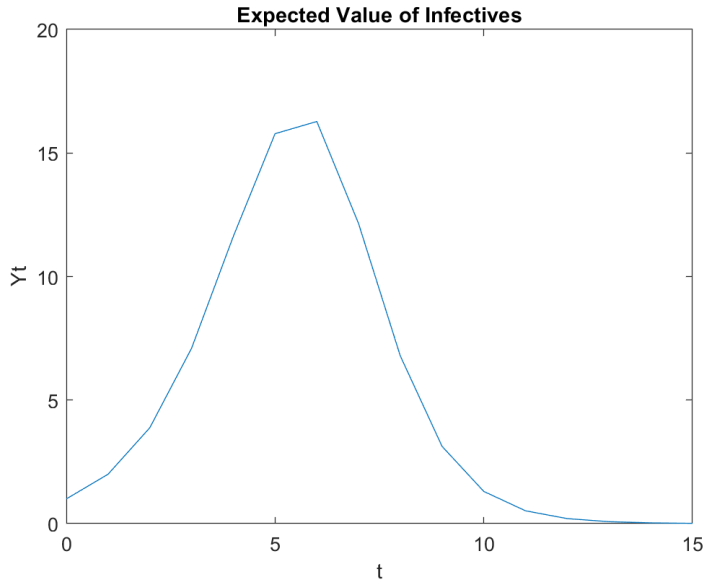
(b) Reproduce Figure 4.2 and also plot the trajectory of expected values, jointly on the (X, Y) plane in a similar manner to Figure 10.1 of [SWJ-10] (there, the plot is for a predator pray model).

Using the equations developed from the Reed-Frost model (4.2.1) as well as the information given that $(N, I) = (100,1)$ and $\alpha = 0.98$, Figure 4.2 can be reproduced. This is shown below:

$$\therefore \mathbb{E}[X_{t+1}|(X, Y)_t = (x, y)_t] = x_t \alpha^{y_t} \text{ for } 0 \leq t \leq 15$$

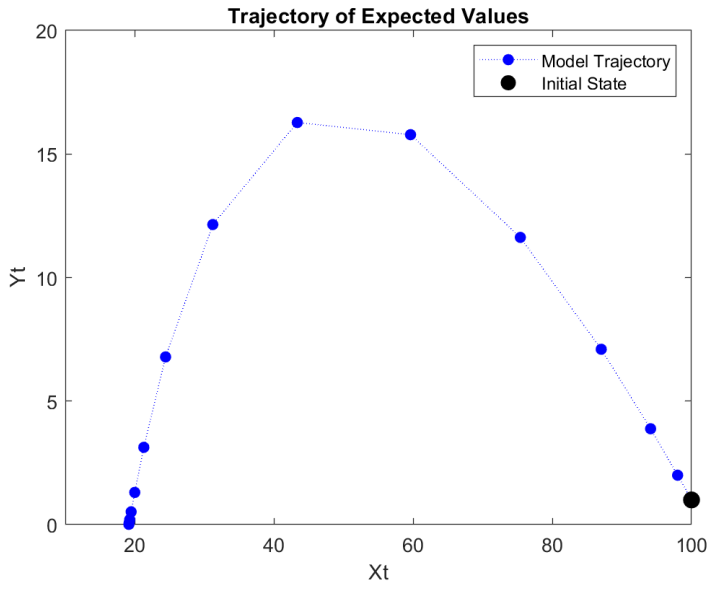


$$\therefore \mathbb{E}[Y_{t+1}|(X, Y)_t = (x, y)_t] = x_t(1 - \alpha^{y_t}) \text{ for } 0 \leq t \leq 15$$



The trajectory of the expected values in a similar manner to Figure 10.1 can also be plotted. This is shown below:

$$\therefore \mathbb{E}[(X,Y)_{t+1}|(X,Y)_t = (x,y)_t] = (x_t\alpha^{y_t}, x_t(1-\alpha^{y_t}))$$

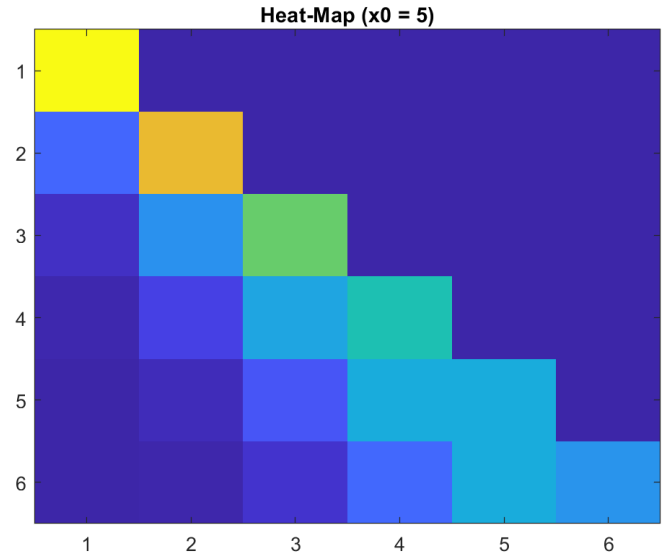


3. Consider the Markov Chain $\{X_t, t = 0, 1, 2, \ldots\}$ defined by $X_0 = x_0$ and the transcript probability matrix as in equation (4.1.5) for the Greenwood model. The state space is $\{0, 1, 2, \ldots x_0\}$.

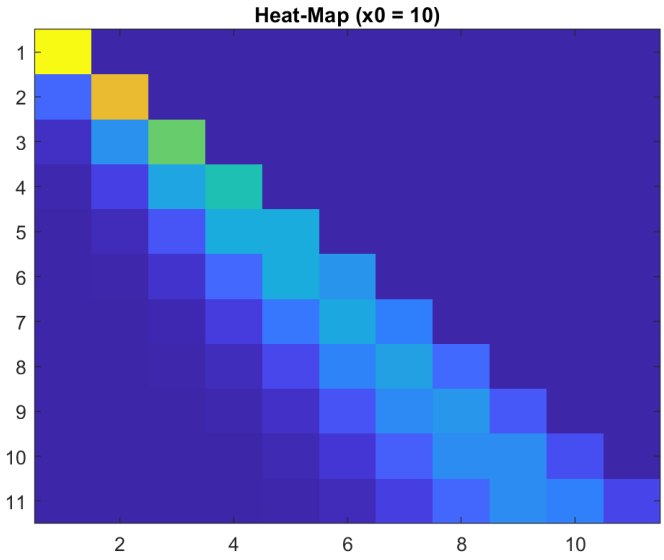
(a) Plot a heat-map of this transition probability matrix for $x_0 = 5, x_0 = 10, \text{and } x_0 = 20$ and some $\alpha \in (0.5, 0.9)$ of your choice.

Taking α to be equal to 0.8, a heat-map can be plotted relating to the three x_0 values. This is shown below:

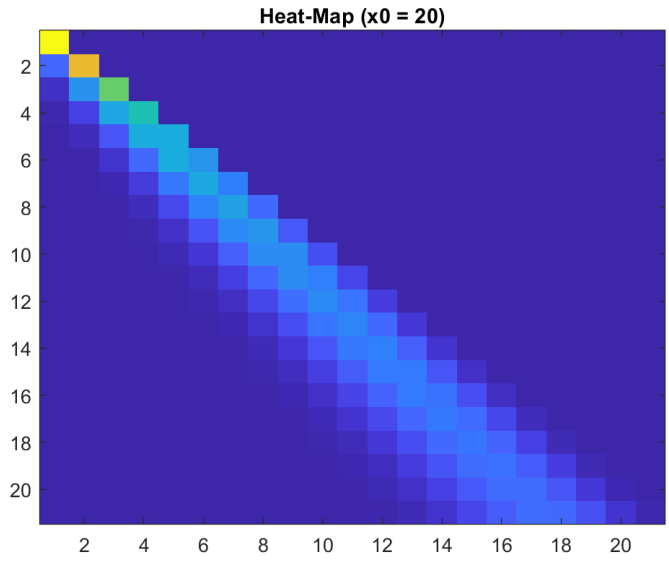
$$\therefore x_0 = 5$$



$$\therefore x_0 = 10$$



$$\therefore x_0 = 20$$



(b) Determine the communicating classes of this Markov chain. How many are there? Which are recurrent? Which are transient?

Two states are said to be communicating, written as $i \leftrightarrow j$, if they are accessible from each other. In order words, $i \leftrightarrow j$ means $i \rightarrow j$ and $j \rightarrow i$. Communication is an equivalence relation. Accordingly, every state communicates with itself, $i \leftrightarrow i$, if $i \leftrightarrow j$, then $j \leftrightarrow i$, if $i \leftrightarrow j$, and $j \leftrightarrow k$, then $i \leftrightarrow k$. And so, based on this definition, the states of a Markov chain can be partitioned into so called communicating classes if members of the same class communicate with each other.

The communicating classes for this Markov chain are simply their own individual state. As such, let communicating classes be denoted by C. The communicating classes of this Markov chain are $C_1 = \{0\}, C_2 = \{1\}, C_3 = \{2\}, C_4 = \{3\}, C_5 = \{4\}, C_6 = \{5\}, C_7 = \{6\}$. In total, there are seven communicating classes.

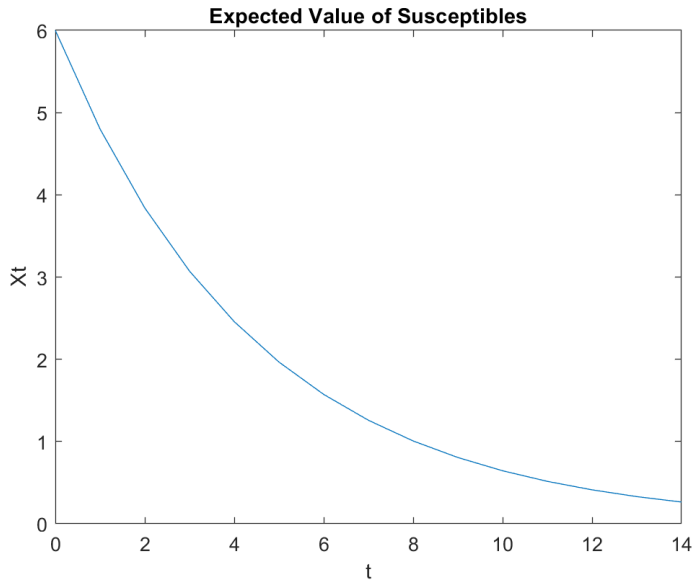
If a state is recurrent, then the chain will always return to that state if at any time it leaves. The only state which is recurrent for this Markov chain is state 0 since it has a probability of 1. Therefore, the chain will visit this state an infinite number of times.

On the other hand, if a state is transient, the chain will not always return to the state if at any time it leaves. Every other state for this Markov chain in that of 1, 2, 3, 4, 5, and 6 are transient. This is because each of these states has a probability of less than 1 for returning.

(c) The part of equation (4.1.4) for X_t, presenting the expected value, can be obtained in a much more cumbersome way to what you did in 1a above. For this, take the power P^t and compute $e_{x_0+1}^T P^t v$, where e_{x_0+1} is the $x_0 + 1$ long unit vector $[0\ 0 \ldots 1]^T$ and v is the vector $[0\ 1\ 2 \ldots x_0]^T$. Compute this numerically and see the results numerically agree with those in the plot 1b. Explain why this holds.

Considering the values such that that $x_0 = 6$ and $\alpha = 0.8$, the new equation can be compared to the part of equation (4.1.4) for X_t. By computing this numerically, the expected value can be achieved along the same time horizon with t equalling 14. This is shown below:

$$\therefore e_{x_0+1}^T P^t v \text{ for } 0 \leq t \leq 14$$

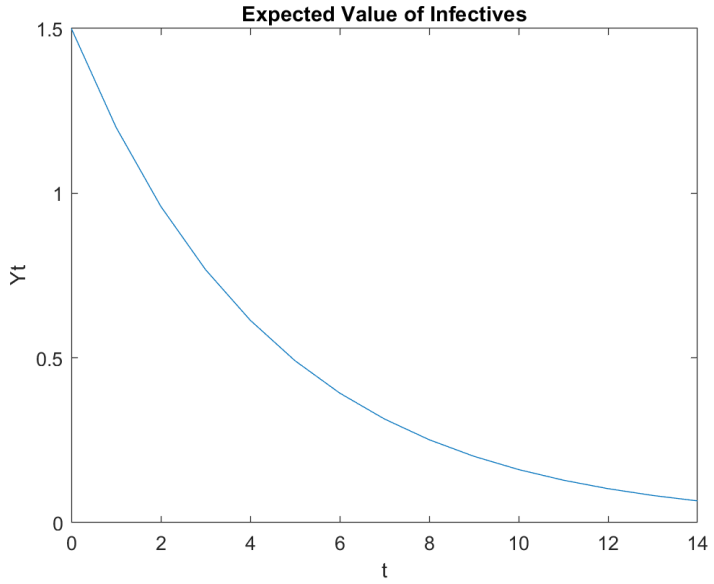


Analysing the plot, the results can be seen to numerically agree with the original equation (4.1.4). This new equation holds because it follows in a similar manner to the part of equation (4.1.4) for X_t in that of $\mathbb{E}[X_t|X_0 = x_0] = \alpha^t x_0$. The matrix P, taking into consideration the α value of 0.8 and time horizon, when multiplied with the vector v, produces the expected values. The vector e then allows for the initial distribution being the initial value $x_0 = 6$ to be considered as well. Accordingly, with all variables accounted for, the plot matches the one produced previously.

(d) Attempt to carry out a similar numerical computation for the expectation of Y_t in equation (4.1.4) and explain your method.

In the same way, by considering the values such that that $x_0 = 6$ and $\alpha = 0.8$, a new equation can be developed and therefore compared to the part of equation (4.1.4) for Y_t. By computing this numerically, the expected value can be achieved along the same time horizon with t equalling 14. This is shown below:

$$\therefore e_{x_0+1}^{T} P^{t-1}(1-a)v \text{ for } 0 \leq t \leq 14$$



Examining the plot, the results can also be seen to numerally agree with the original equation (4.1.4). The explanation for this method is due to this equation for the expectation of Y_t being developed similarly to the one in equation (4.1.4). For reference, the part of equation (4.1.4) for Y_t is $\mathbb{E}[Y_t|X_0 = x_0] = \alpha^{t-1}(1-\alpha)x_0$. The addition of the $(1-a)$ in this equation allows for the infectives to be determined rather than the susceptibles. Along with the matrix P, taking into consideration the α value of 0.8 and time horizon, being multiplied with the vector v, produces the expected values. The vector e then allows for the initial distribution being the initial value $x_0 = 6$ to be considered as well. However, since the time is now (t – 1) instead of just t, this value is shifted. Consequently, with all variables accounted for, the plot matches the one produced previously.

4. Consider now the joint distribution $(\mathbf{W}, \mathbf{T})$ as described in subsection 4.1.1 (dealing with the Greenwood model). Here T is the first time in which there are no infectives and W is the number of susceptibles that have been infected by that time. That is the random variable T and W describe the "end of the infection". The main aim is to know the probabilities,

$$\boldsymbol{\tau(k, n|x_0) = \mathbb{P}((W, T) = (k, n)|X_0 = x_0, Y_0 > 0)},$$

for $\mathbf{k = 0, 1, \ldots, x_0}$ and $\mathbf{n = 1, 2, \ldots}$. These assume that at onset $\mathbf{x_0}$ family members are sick and there is an infection in the household.

For all numerical computations in this question, use $\mathbf{x_0 = 6}$ and some fixed $\boldsymbol{\alpha \in (0.7, 0.9)}$ of your choice.

(a) Explain equation (4.1.6).

For equation (4.1.6):

$$p_j^t \equiv \Pr\{X_t = j, Y_t > 0\} = \sum_{i=j+1}^{x_0-(t-1)} p_i^{t-1}\, p_{ij} \text{ where for any integers } t \geq 1 \text{ and } j = 0, \ldots, i$$

Using the law of total expectation:

$$p_j^t = \sum_{i=j+1}^{x_0-(t-1)} P(X_{t-1} = i, Y_{t-1} > 0)\, P(X_t = j|X_{t-1} = i, Y_{t-1} > 0)$$

This equation (4.1.6) describes the position at state j given a time t. Having a closer investigation, the use of the summation allows for the consideration of all states. It is vital that the probability of different states is added because this studies all possibilities. The lower bound of $i = j + 1$ suggests that there will always be a change from the previous state to the next. This is the case as susceptibles can become infectives but infectives cannot become susceptibles. The upper bound of $x_0 - (t - 1)$ indicates that the number of susceptibles changes with respect to time. Given that $t \geq 1$, this makes sense as the number of susceptibles is at its initial value when $t = 1$. These bounds ultimately indicate that infectives will increase for susceptibles that are decreasing until there are no more to infect.

The probability p_i^{t-1}, expanding to be $P(X_{t-1} = i, Y_{t-1} > 0)$, presents the number of susceptibles for the previous state i at time $t - 1$. In a similar fashion, the probability p_{ij}, expanding to be $P(X_t = j, |X_{t-1} = i, Y_{t-1} > 0)$, presents the number of susceptibles for the current state j given the previous state i. Moreover, the multiplication of these two probabilities presents the number of susceptibles for the next state j at time $t + 1$.

(b) Use the recursive relationship $\boldsymbol{\tau(k, n|x_0) = p_{x_0-k}^{n-1} \alpha^{x_0-k}}$ to (numerically) compute $\boldsymbol{\mathbb{P}(W > 4)}$.

By using the recursive relationship, $\mathbb{P}(W > 4)$ comes to a value of 0.13029835750312502. When rounded to two decimal places, this becomes 0.13 (13%). Based off the initial condition of x_0 being 6 and α equalling to 0.8, this value is reasonable since α represents the probability that there is no infection due to any single infective. Given that α is quite high, the probability for the number of susceptibles to get infected and be greater than 4 would not result in a high value. In this case, the probability is 0.13 or simply 13%.

(c) Compare your numerical result to an estimate obtained by a Monte-Carlo simulation creating 10^6 repeated trajectories and using those to estimate $\mathbb{P}(W > 4)$.

By using a Monte-Carlo simulation, $\mathbb{P}(W > 4)$ can vary in different values. Although slightly different results are produced each simulation, the difference between these values are very minute. Accordingly, one realisation suggested that $\mathbb{P}(W > 4)$ comes to a value of 0.130324. When rounded to two decimal places, this becomes 0.13 (13%). Comparing to the recursive relationship, the value rounded to two decimal places is the same. When comparing the exact values, there is only a difference of less than 0.00003. The reason why the Monte-Carlo simulation is somewhat different to the recursive relationship is because it depends on the variance of the estimator being used for each run. Despite this, the numerical result of the recursive relationship and the Monte-Carlo simulation can be considered to verify each other.

(d) Attempt to reproduce the PGF computations in subsection 4.1.1 to then obtain the same numerical result (this item is longer and slightly more challenging).

By using PGF computations, $\mathbb{P}(W > 4)$ comes to a value of 0.130298357503125. When rounded to two decimal places, this becomes 0.13 (13%). Comparing to the recursive relationship and Monte-Carlo simulation, the value rounded to two decimal places is the same. When comparing the exact values with the recursive relationship, the only difference is the number of decimal places. On the other hand, when comparing the exact values with the Monte-Carlo simulation, there is only a difference of less than 0.00003. As such, the numerical result of the recursive relationship, Monte-Carlo simulation, and PGF computations can be considered to verify one another.

---

5. Consider the Markov chain for the Reed-Frost model with transition probability matrix as in (4.2.2).

For all numerical computations in this question, use $x_0 = 6$ and some fixed $\alpha \in (0.7, 0.9)$ of your choice (use the same α which you used for the previous question).

(a) What is the state space?

By assuming that x_0 is the total size of the population, the state space will be limited. This is because the when summing the number of susceptibles and infectives in any given state, the value must be less than or equal to x_0. To elaborate, taking $x_0 = 6$, (6,6) and (5,5) etc. will not be possible states. On the other hand, states in that of (6,0), (5,0) and (5,1) etc. are possible.

Therefore, the state space is:

$$\{(x, y) | 0 \leq (x, y) \leq x_0\} \forall (x + y) \leq x_0$$

(b) Try to describe the communicating classes in a compact manner? If not possible, constrain to a small fixed x_0.

The recurrent communicating classes are:

$$\{(x, 0) | 0 \leq x \leq x_0\}$$

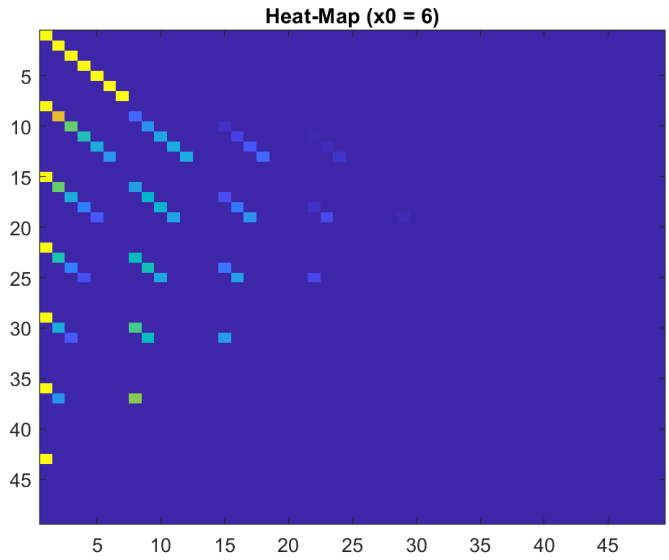
The transient communicating classes are:

$$\{(x, y) | 0 \leq x \leq x_0, 1 \leq y \leq x_0\} \forall (x + y) \leq x_0$$

(c) Plot a heat-map similarly to 3a (you may want to use block-matrices in your software).

Taking α to be equal to 0.8, a heat-map can be plotted relating to $x_0 = 6$. This is shown below:

$$\therefore x_0 = 6$$



(d) Run a Monte-Carlo simulation to obtain an estimate for $\mathbb{P}(\mathbf{W} > \mathbf{4})$ similarly to 4c. How does the result compare to 4c? Explain why.

By using a Monte-Carlo simulation, $\mathbb{P}(W > 4)$ comes to a value of 0.225978. When rounded to two decimal places, this becomes 0.23 (23%). Assuming $y_0 = 1$, this results in the initial condition of x_0 being 5. With α equalling to 0.8, this value is reasonable since α represents the probability that there is no infection due to any single infective. Given that α is quite high, the probability for the number of susceptibles to get infected and be greater than 4 would not result in a particularly high value. In this case, the probability is 0.23 or simply 23%.

In saying that, comparing to the result rounded to two decimal places in that of 0.13 (13%) when x_0 equals 6, the value differs by 0.10 (10%). The reason of this difference is due to the necessity to assume the value of y_0. In doing so, this increases the probability as there is now a lower number of susceptibles and a higher number of infectives to start off with. The increase in probability is almost double the original value.

Scenario

Coronavirus, or otherwise known as COVID-19, is an infectious disease that has heavily impacted millions around the world. With no recent event in history having such a profound and pervasive effect, the outbreak of this virus has forced the whole world to be in isolation. Classified as a pandemic, it is crucial that a way to slow and prevent the spread of this disease is found. Given that a vaccine will take some time to become available, the importance and effectiveness of quarantining must be investigated. Considering a small motel in a regional town with capacity for up to x_0 people, the aim of this report is to determine the long term expected rate of infections per day coming out of this motel. By using the Reed-Frost model, an analysis can be carried out to critically measure the overall impact of this unprecedented circumstance.

Model

The probability of new arrivals being infected by COVID-19 is said to be $\eta = 0.05$. Additionally, the infections inside the motel is said to follow the Reed-Frost model with $p = 0.1$ and $\beta = 0.05$. For this model, p is the probability of contact between an infective and a susceptible, whereas β represents the probability of the contact being an infection. Using these two values, the probability that there is no infection due to any single infective can be determined. This is shown below:

$$\therefore 1 - p + p(1 - \beta) = 1 - p\beta = a$$

$$\therefore 1 - 0.1 + 0.1(1 - 0.05) = 1 - 0.1 \cdot 0.05 = 0.995$$

Taking $\eta = 0.05$ and $\alpha = 0.995$, a binomial distribution can be used to calculate the long term expected rate of infections per day. Considering the initial conditions, the successive states can be modelled by implementing a sequence of binomial random variables. With enough simulations, a steady state can be reached. Through a Monte-Carlo simulation, where it terminates when the number of infectives equals 0, this can be achieved.

By using the Reed-Frost model, it is important to know that $X_{t+1} = \text{Bin}(X_t, \alpha^{Y_t})$, where $\text{Bin}(n, \pi)$ denotes a random variable with the binomial distribution $\left\{\binom{n}{j}\pi^j(1-\pi)^{n-j}, j = 0, \dots, n\right\}$. Considering that $y_{t+1} = x_t - x_{t+1}$, the matrix of one-step transition probabilities is also given by $p(x, y)_t, (x, y)_{t+1} = \binom{x_t}{x_{t+1}} \alpha^{y_t x_{t+1}}(1 - \alpha^{y_t})^{y_{t+1}}$. Looking at the transition probability matrix, the connection between each state can be determined. As such, the stationary distribution can be developed based off each transition probability matrix that is used for the model.

For the expected time between replacements, this is more vital when looking at discrete time. However, over a long period, the rate of infections stabilises and is consistent within this model.

Assumptions

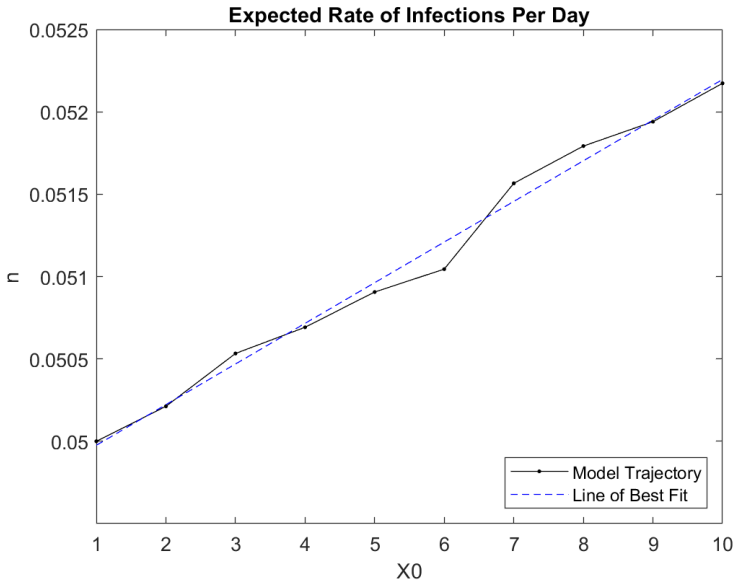
The model being developed has many assumptions that needs to be addressed. The first assumption, as stated before, is that the rate of infections per person into the regional town is at $\eta = 0.05$. As such, it is also required to assume that there is not any immunity to COVID-19. This prevents the consideration of individuals that are unaffected by coronavirus. The infections inside the motel, as previously mentioned, are also assumed to follow the Reed-Frost model with $p = 0.1$ and $\beta = 0.05$. From this, an even distribution between each susceptible is assumed for calculations. It is also noted that when describing the spread of disease whose infectious period is relatively short in comparison with the latent period, it is convenient to use a discrete-time model with the latent period as a unit of time. This is especially the case for smaller populations.

Results

The results found from this model can be tabulated and then plotted. This is shown below:

x_0	η
1	0.05
2	0.050212
3	0.050533
4	0.050692
5	0.050906
6	0.051045
7	0.051566
8	0.051792
9	0.051940
10	0.052173

*Note that the values for η are rounded to six decimal places.



Analysing the values generated, there seems to be a trend that as the value of x_0 increases, the value of η also increases accordingly. Using a line of best fit, the trajectory of the model can be compared. It is evident from this that there seems to be a linear increase within the x_0 bounds. Although the local behaviour at each point may vary, it is very minute to be able to tell. In saying that, the long-term outlook for this model will develop into a logarithmic function. This is because the expected rate of infections per day for larger x_0 values will eventually be increasing at a decreasing rate. As such, when approaching a rate of 1, the function will plateau. That is, after the period of increase in η values, the state will remain constant over a period of x_0 values.

To further examine the data, although the spread is not uniformly distributed across all parts, it can be interpreted that as more susceptibles are quarantined together, the likelihood of them getting infected increases. This is an interesting discovery as it suggests the idea of quarantining with many others will not be an ideal situation. Despite not being a particularly substantial increase compared to the rate of infections per person into the regional town at $\eta = 0.05$, if the x_0 values were to increase, this may become more concerning in the long run. Given that motels generally hold more than 10 people, this is very likely to be the case. Moreover, it is fundamental to further analyse the model in detail to provide a clearer insight into the possible unrealistic circumstances presented.

Analysis

With the assumptions and results studied, an in-depth analysis shows that many parts of the model can be improved given some sections are unrealistic. The values used within this model can also be revised. By looking at a broader range of x_0, η, p, and β values, a more general idea of how the rate of infection changes can be calculated.

With the rate of infections increasing as more people quarantine together, the implementation of social distancing must also come into play for the prevention of COVID-19. To elaborate, it is essential the population quarantines in smaller groups to make this process effective. With motels generally providing apartments for different groups, it is unlikely that many groups will interact with one another if restrictions are also set. Therefore, the infection rate should decrease as a result. However, just like in this model, if groups interact with each other, the infection rate increases. An example of a real-life situation is the Ruby Princess.

Passengers onboard the Ruby Princess, similarly in this small motel, interact with each other daily. With no guidelines set in place, the transmission of COVID-19, given that there is at least one infective, is easy to pass on from one individual to another. With Ruby Princess being responsible for hundreds of COVID-19 cases and many deaths, the spread is shown to increase at an exponential rate as more people interact (Carruthers & Wootoon, 2020).

Looking at this model, the need to assume that there is not any immunity to COVID-19 is unrealistic. In doing so, prior health conditions of the susceptibles are disregarded. Therefore, this reduces the information in the chance of getting COVID-19 per person. Additionally, it is also unreasonable to not consider personal hygiene in the use of hand sanitisers and face masks. As a result, the rate of infections cannot be contained. Despite not being directly related to the infection rate, the model also does not study the type of individuals in whether they are children or the elderly. As the mortality rate is higher among older people, a more thorough perception can be provided if the likelihood of dying when infected is taken into context (Whiting, 2020).

For future investigations, a further analysis into the testing method of COVID-19 can be explored. In doing so, a confusion matrix, allowing visualisation of the performance of an algorithm can be studied (Narkhede, 2018). This takes into the true positive, true negative, false positive, and false negative values. From this, tests of sensitivity and specificity can be applied based off real-life scenarios (Trevethan, 2017). Although this is beyond the scope of this report, it might potentially give a better insight into the exact values for the long term expected rate of infections per day coming out of this motel.

References

Carruthers, F., & Wootton, H. Financial Review. (2020). Ruby Princess is the most deadly virus ship. Retrieved 29 April 2020, from:

https://www.afr.com/companies/tourism/ruby-princess-is-the-most-deadly-cruise-ship-20200403-p54gwa

Whiting, K. World Economic Forum. (2020). An expert explains: how to help older people through COVID-19 pandemic. Retrieved 29 April 2020, from:

https://www.weforum.org/agenda/2020/03/coronavirus-covid-19-elderly-older-people-health-risk/

Narkhede, S. Medium. (2018). Understanding Confusion Matrix. Retrieved 29 April 2020, from:

https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Trevethan, R. Frontiers. (2017). Sensitivity, Specificity, and Predicted Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Retrieved 29 April 2020, from:

https://www.frontiersin.org/articles/10.3389/fpubh.2017.00307/full

Appendix

All code related documents are in the zip file. The title Project_11b refers to Project 1 1. (b). Otherwise, all plots and numerical computations from the code are presented in this document.