STAT3004: Project 1

Isabel Peters

Question 1

Equation (4.1.4) presents the expected numbers in the Greenwood model.

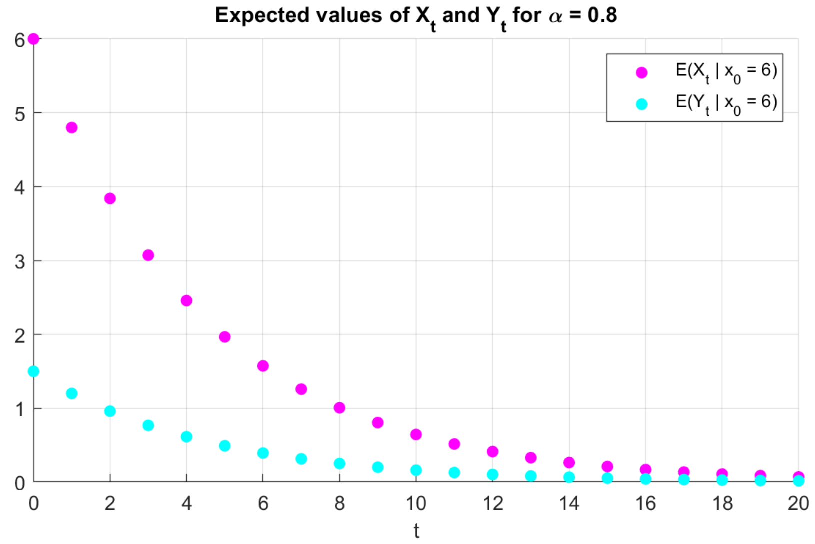
a) Derive these equations. As per equation (4.1.2), we know $X_t \sim Binom(X_{t-1},\ \alpha)$, which has expectation αX_{t-1}. Thus we have

$$\begin{aligned}
\mathbb{E}(X_t \mid X_0 = x_0) &= \mathbb{E}(\mathbb{E}(X_t \mid X_{t-1},\ X_0 = x_0)) \qquad \text{(Law of total expectation)}\\
&= \mathbb{E}(\alpha X_{t-1} \mid X_0 = x_0)\\
&= \alpha\mathbb{E}(X_{t-1} \mid X_0 = x_0)\\
&= \alpha\mathbb{E}(\mathbb{E}(X_{t-1} \mid X_{t-2},\ X_0 = x_0))\\
&= \alpha^2\mathbb{E}(X_{t-2} \mid X_0 = x_0)\\
&\vdots\\
&= \alpha^t\mathbb{E}(X_{t-t} \mid X_0 = x_0)\\
&= \alpha^t\mathbb{E}(X_0 \mid X_0 = x_0)\\
&= \alpha^t\mathbb{E}(x_0)\\
&= \alpha^t x_0, \quad \text{as required.}
\end{aligned}$$

We know $Y_t = X_{t-1} - X_t$. Thus we have

$$\begin{aligned}
\mathbb{E}(Y_t \mid X_0 = x_0) &= \mathbb{E}(X_{t-1} - X_t \mid X_0 = x_0)\\
&= \mathbb{E}(X_{t-1} \mid X_0 = x_0) - \mathbb{E}(X_t \mid X_0 = x_0)\\
&\qquad \text{(by linearity of expectation)}\\
&= \alpha^{t-1}x_0 - \alpha^t x_0\\
&= \alpha^{t-1}(1-\alpha)x_0, \quad \text{as required.}
\end{aligned}$$

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



MatLab code:

```
x_0 = 6;
alpha = 0.8;
t = 0:20;
% Calculate expected values
E_Xt = alpha.^t * x_0;
E_Yt = alpha.^(t-1) * (1-alpha) * x_0;
% Plot
scatter(t, E_Xt, 'filled', 'm');
hold on;
scatter(t, E_Yt, 'filled', 'c');
```

Question 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.

As per equation (4.1.3), we know $X_{t+1} \sim Binom(X_t,\ \alpha^{Y_t})$. If $(X,Y)_t = (x,y)_t$, then X_{t+1} is simply a $Binom(x_t,\ \alpha^{y_t})$ random variable, the expectation of which is $x_t\alpha^{y_t}$. Thus we have

$$\mathbb{E}(X_{t+1} \mid (X,Y)_t = (x,y)_t) = \mathbb{E}\left(\binom{x_t}{x_{t+1}}\alpha^{y_t x_{t+1}}(1-\alpha^{y_t})^{(x_t - x_{t+1})}\right) = x_t\alpha^{y_t}$$

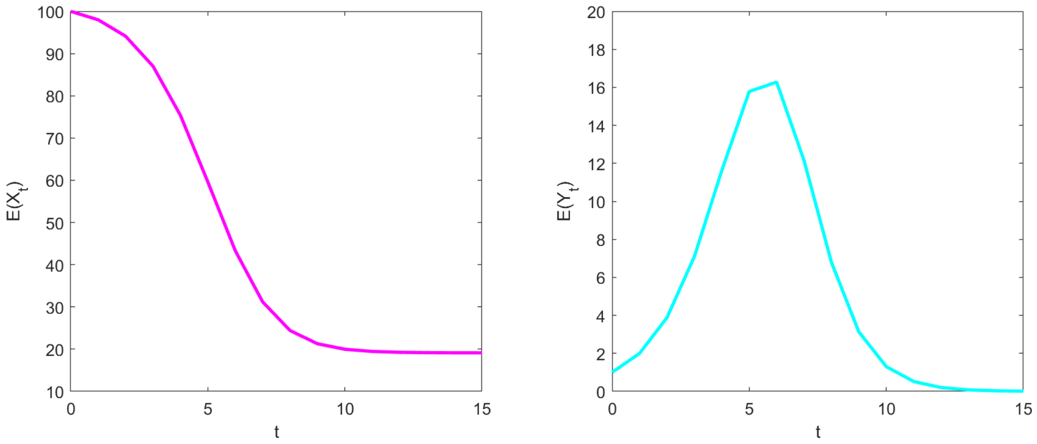
We then note $Y_{t+1} = X_t - X_{t+1}$. So we have

$$\begin{aligned}\mathbb{E}(Y_{t+1} \mid (X,Y)_t = (x,y)_t) &= \mathbb{E}(X_t - X_{t+1} \mid (X,Y)_t = (x,y)_t)\\ &= \mathbb{E}(X_t \mid (X,Y)_t = (x,y)_t) - \mathbb{E}(X_{t+1} \mid (X,Y)_t = (x,y)_t)\\ &\quad \text{(by linearity of expectation)}\\ &= \mathbb{E}(x_t) - \mathbb{E}(X_{t+1} \mid (X,Y)_t = (x,y)_t)\\ &= x_t - x_t\alpha^{y_t}\\ &= x_t(1-\alpha^{y_t})\end{aligned}$$

Combining these two equations, we have $\mathbb{E}((X,Y)_{t+1} \mid (X,Y)_t = (x,y)_t) = (x_t\alpha^{y_t},\ x_t(1-\alpha^{y_t}))$, as required.

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].

Reproduction of figure 4.2:



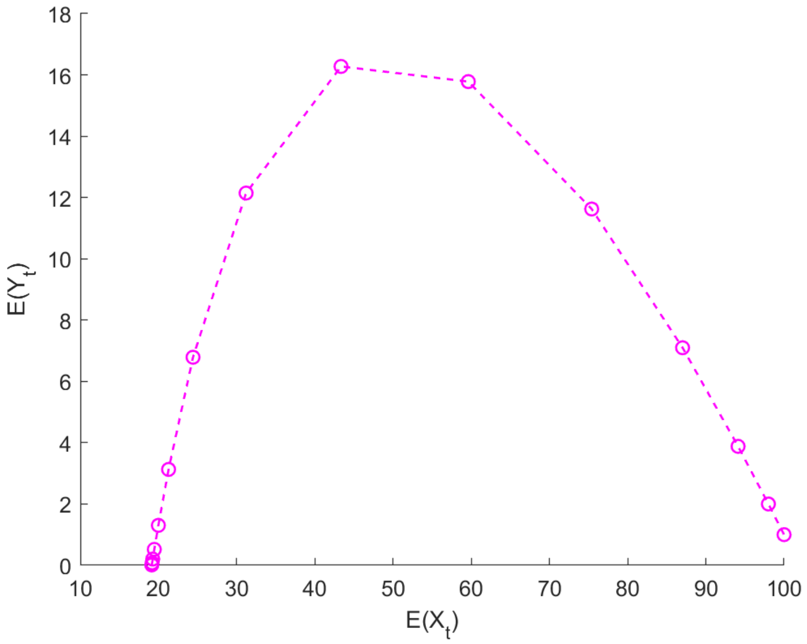
The above figure was produced by the following MatLab code:

```
x_0 = 100;
y_0 = 1;
alpha = 0.98;

E_Xt = zeros(1, 16);
E_Yt = zeros(1, 16);
E_Xt(1) = x_0;
E_Yt(1) = y_0;
for i = 1:15
    E_Xt(i+1) = E_Xt(i) * alpha^E_Yt(i);
    E_Yt(i+1) = E_Xt(i) * (1 - alpha^E_Yt(i));
end

t = 0:15;
subplot(1,2,1);
plot(t, E_Xt, 'm', 'linewidth',2);
subplot(1,2,2);
plot(t, E_Yt, 'c', 'linewidth',2);
```

We then plot these expected values against each other:

```
scatter(E_Xt, E_Yt, 'm', 'linewidth',1)
hold on;
plot(E_Xt, E_Yt, 'm--', 'linewidth',1)
ylabel('E(Y_t)');
xlabel('E(X_t)');
```

This code produces the plot:



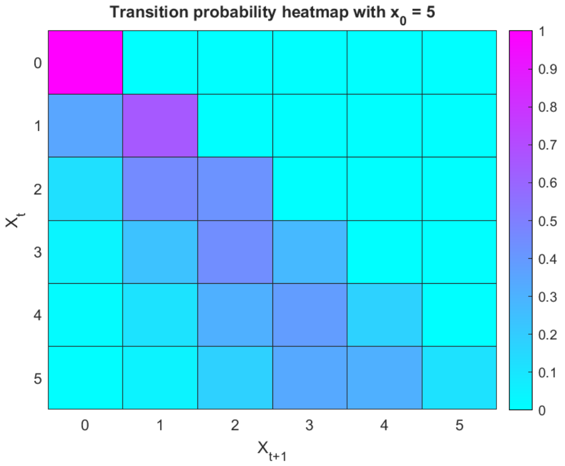
Question 3

Consider the Markov Chain $X_t, t = 0, 1, 2, ...$ defined by $X_0 = x_0$ and the transition probability matrix as in equation (4.1.5) for the Greenwood model. The state space is $0, 1, 2, ..., x0$.

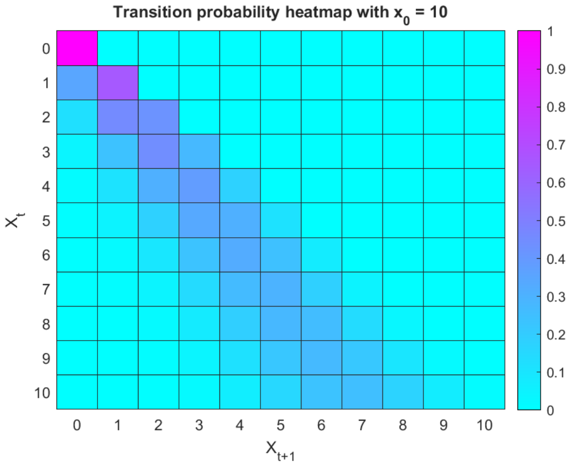
a) Plot a heat-map of this transition probability matrix for $x_0 = 5$, $x_0 = 10$, and $x_0 = 20$ and some $\alpha \in (0.5, 0.9)$ of your choice.

Let $\alpha = 0.65 \in (0.5, 0.9)$. We execute the following code with $x_0 = 5$, 10 and 20, to produce a heatmap for each.

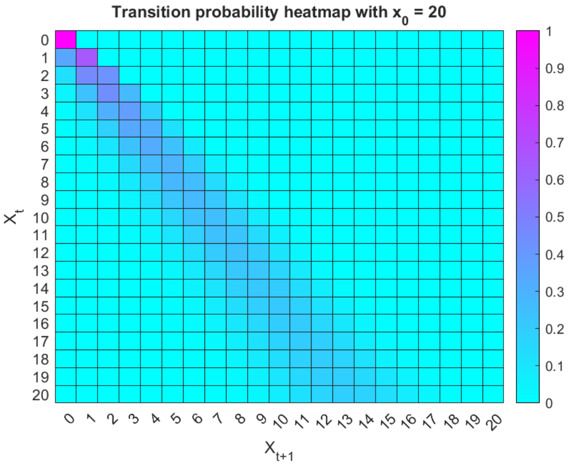
```
x_0 = 5;
alpha = 0.65;

% Calculate the transition probability matrix
P = zeros(x_0+1, x_0+1);
for i = 0:x_0
    for j = 0:i
        P(i+1, j+1) = binopdf(j,i,alpha);
    end
end

% Plot a heatmap of the matrix
T = append('Transition probability heatmap with x_0 = ',num2str(x_0));
heatmap(0:x_0, 0:x_0, P, 'Title',T, 'XLabel','X_{t+1}', 'YLabel','X_t', ...
    'Colormap',cool, 'CellLabelFormat','%.20f');
```



Transition probability heatmap with x_0 = 5



Transition probability heatmap with x_0 = 10



Transition probability heatmap with x_0 = 20

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

Suppose any two states $0 \leq i, j \leq x_0$. We have $p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i) = \binom{i}{j}(1-\alpha)^{i-j}\alpha^j$.

For $0 < \alpha < 1$, we have $\alpha^k > 0 \; \forall k \in \mathbb{Z}$ and $(1-\alpha)^k > 0 \; \forall k \in \mathbb{Z}$. Thus $p_{ij} > 0 \iff \binom{i}{j} > 0$.

If $i > j$, $\binom{i}{j} > 0 \implies p_{ij} > 0$. So state i leads to all states $j < i$.

However, if $i < j$, $\binom{i}{j} = 0 \implies p_{ij} = 0$. So state i cannot transition to any state $j > i$. As this holds for all $0 \leq i \leq x_0$, we have i cannot lead to any state $j > i$.

Thus if $i \neq j$, states i and j do *not* communicate.

Any state communicates with itself by definition, as a state leads to itself in 0 steps. Hence we have $i \leftrightarrow j \iff i = j$, so each state forms its own communicating class. Thus there are $x_0 + 1$ communicating classes.

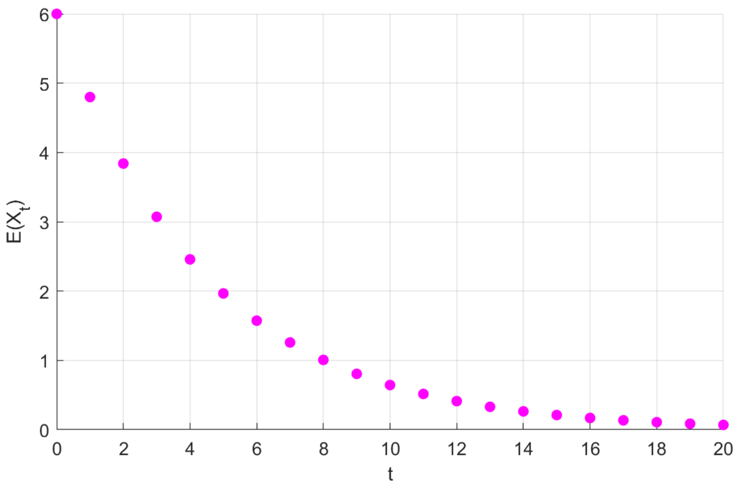
State 0 is a recurrent communicating class as, if we reach this state, we are guaranteed to return to it (in fact, if we reach state 0, we never leave).

States $i = 1,\ 2,\ \ldots,\ x_0$ are transient communicating classes, as if we transition to some state $j < i$, we cannot transition back to state i (because state j cannot lead to state $i > j$).

c) The part of equation 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 the power P^t and compute $e_{x_0+1}^\top P^t v$, where e_{x_0+1} is the $x_0 + 1$ long unit vector $[0\ 0\ \ldots\ 1]^\top$ and v is the vector $[0\ 1\ 2\ \ldots\ x_0]^\top$. Compute this numerically and see that the results numerically agree with those in the plot of 1b. Explain why this holds.

We execute the following MatLab code to produce a plot of $\mathbb{E}(X_t)$ via this method.

```
x_0 = 6;
alpha = 0.8;

% Calculate the transition probability matrix
P = zeros(x_0+1, x_0+1);
for i = 0:x_0
    for j = 0:i
        P(i+1, j+1) = binopdf(j,i,alpha);
    end
end

% Calculate E(X_t)
e = zeros(x_0+1, 1);
e(x_0+1) = 1;
v = (0:x_0)';
E_Xt = zeros(21, 1);
for t = 0:20
    E_Xt(t+1) = e' * P^t * v;
end

% Plot
scatter(0:20, E_Xt, 'filled', 'm');
```



This plot is clearly identical to that produced in Q1b.
This method works because we calculate $\mathbb{E}(X_t \mid X_0 = x_0) = \sum_x \mathbb{P}(X_t = x \mid X_0 = x_0) \cdot x$.
$\mathbb{P}(X_t = x \mid X_0 = x_0)$ is given by $e_{x_0+1}P^t$. This is because $P^t = (p_{ij}^t)$ where p_{ij}^t is the probability of getting from state i to state j in t steps. We want the probability we got from state x_0 to any state x in t steps, so we pre-multiply by e_{x_0+1} which represents $\mathbb{P}(X_0 = x_0) = 1$. This gives a row vector of the probabilities required.
We then post-multiply by the column vector v, which contains the 'value' of each of the states. This executes both the multiplying by x and the summing over all values of x, by taking advantage of matrix multiplication. Thus the final answer is in fact $\mathbb{E}(X_t \mid X_0 = x_0)$.

d) Omitted due to time constraints.

Question 4

Consider now the joint distribution of (W, 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 variables T and W describe the "end of the infection". The main aim is to know the probabilities,

$$\Gamma(k, n \mid x_0) = \mathbb{P}\{(W, T) = (k, n) \mid X_0 = x_0,\ Y_0 > 0\},$$

for $k = 0, 1, ..., x_0$ and $n = 1, 2, ...$ These assume that at onset x_0 family members are sick and there is an infection in the household. For all the numerical computations in this question, use $x_0 = 6$ and some fixed $\alpha \in (0.7, 0.9)$ of your choice.

a) Explain equation (4.1.6).

Equation (4.1.6) says

$$p_j^t = \mathbb{P}(X_t = j, Y_t > 0) = \sum_{i=j+1}^{x_0-(t-1)} p_i^{t-1} p_{ij},$$

where $p_0^i = \mathbb{P}(X_0 = i) = \delta_{x_0 i}$

This equation gives the probability that we are in state j at time t, *and* the infection is still occurring ($Y_t > 0$). Clearly, we can decompose this into the probability that we were in some state other state i in the previous time period $t - 1$ (and the infection was still occurring), and then transitioned to state j (and the infection was still occurring):

$$\begin{aligned}\mathbb{P}(X_t = j, Y_t > 0) &= \sum_{i \in S_{t-1}} \mathbb{P}(X_t = j, X_{t-1} = i, Y_t > 0)\\ &= \sum_{i \in S_{t-1}} (\mathbb{P}(X_t = j, Y_t > 0 \mid X_{t-1} = i)\ \mathbb{P}(X_{t-1} = i, Y_t > 0))\\ &= \sum_{i \in S_{t-1}} p_i^{t-1} p_{ij}\end{aligned}$$

where S_{t-1} is the set of all possible states X_{t-1} can be.

We note we must have at least one infection each time period $1, ..., t$ if $Y_t > 0$. So, for non-zero probabilities, $i = x_0 - t - 1$ at most. This also means $j \leq i - 1$, so, for non-zero probabilities, $i = j + 1$ at least. This gives the summation seen in equation (4.1.6):

$$\mathbb{P}(X_t = j, Y_t > 0) = \sum_{i \in S_{t-1}} p_i^{t-1} p_{ij} = \sum_{i=j+1}^{x_0-(t-1)} p_i^{t-1} p_{ij}$$

Finally, we consider $p_0^i = \mathbb{P}(X_0 = i, Y_0 > 0)$. We always take $Y_0 > 0$ (otherwise there is is infection to model), so $p_0^i = \mathbb{P}(X_0 = i)$.

We have $\mathbb{P}(X_0 = i) = 1 \iff i = x_0$ and $\mathbb{P}(X_0 = i) = 1 \iff i \neq x_0$, which is represented by the Kronecker delta function $\delta_{x_0 i}$. Thus $p_0^i = \delta_{x_0 i}$.

b) Use the recursive relationship $\Gamma(k, n \mid x_0) = p^{n-1}_{x_0-k}\alpha^{x_0-k}$ to (numerically) compute $\mathbb{P}(W > 4)$.

First we create a function that recursively calculates p^t_j as per equation (4.1.6), for some x_0 and α.

```
function p = Eqn_4_1_6(j, t, x_0, alpha)
    if t == 0
        % Probabiltiy of x_t = j when t = 0
        if j == x_0
            p = 1;
        else
            p = 0;
        end
    else
        % Probability of all possible paths that could result in x_t = j,
        p = 0;
        for i = (j+1):(x_0 - (t-1))
            p = p + (Eqn_4_1_6(i, t-1, x_0, alpha) * binopdf(j,i,alpha));
        end
    end
end
```

We then note that $\mathbb{P}(W > 4) = \mathbb{P}(W = 5) + \mathbb{P}(W = 6)$, as W is at most equal to $x_0 = 6$ (all susceptibles become infected).
We also note that the slowest an infection can progress is when one susceptible becomes infected each time period until time x_0, when $X_{x_0} = 0$. There are then no susceptibles left to infect, so $X_{x_0+1} = 0 = X_{x_0}$, so $x_0 + 1 = T$. Thus the longest the infection can last / the maximum value for T is 7 time periods. The minimum value for T is 1 (where no infections occur so $X_1 = x_0 = X_0$).
Thus we have

$$\begin{aligned}\mathbb{P}(W > 4) &= \sum_{t=1}^{7}\Big(\mathbb{P}\{(W, T) = (5, t) \mid X_0 = 6, Y_0 > 0\} + \mathbb{P}\{(W, T) = (5, t) \mid X_0 = 6, Y_0 > 0\}\Big)\\ &= \sum_{t=1}^{7}\big(\Gamma(5, t \mid 6) + \Gamma(6, t \mid 6)\big)\end{aligned}$$

We can then use the equation (4.1.6) function to calculate $\mathbb{P}(W > 4)$ this way:

```
x_0 = 6;
alpha = 0.85;

Gamma = @(k, n) Eqn_4_1_6(x_0-k, n-1, x_0, alpha) * alpha^(x_0 - k);

prob = 0;
for t = 1:7
    for w = 5:6
        prob = prob + Gamma(w,t);
    end
end
```

This script returns

```
prob =
   0.051583700268074
```

Thus numerically we have $\underline{\underline{\mathbb{P}(W > 4) = 0.0515837}}$.

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

To do a Monte-Carlo simulation, we create a function that generates one possible outcome of the Greenwood model, for some x_0 and α, and returns the values of W and T.

```
function [w, t] = Greenwood_sim(x_0, alpha)
    t = 0;
    x_t = x_0;
    x_tminus1 = -1;
    % x_0 > 0 so we get at least one step
    while x_t ~= x_tminus1
        t = t+1;
        x_tminus1 = x_t;
        x_t = binornd(x_tminus1, alpha);
    end
    % x_t = x_tminus1, so infection has ended at time t
    w = x_0 - x_t;
end
```

We then execute 10^6 of these simulations, and estimate $\mathbb{P}(W > 4)$ by the proportion of simulations which have $W > 4$, as so:

```
x_0 = 6;
alpha = 0.85;
N = 10^6; % Total number of sim.s to run

gt4 = 0; % To record the number of sim.s with W > 4
for n = 1:N
    [W, T] = Greenwood_sim(x_0, alpha);
    % Add a count for all that have W > 4
    if W > 4
        gt4 = gt4 + 1;
    end
end

% P(W > 4) is approx. by the proportion of the N total sim.s that had W > 4
prob = gt4/N;
```

A run of this script returned

```
prob =
   0.051664
```

Thus we have an estimate of $\underline{\underline{\mathbb{P}(W > 4) \approx 0.051664}}$.

Note that this agrees with the numerical answer from Q4b ($\mathbb{P}(W > 4) = 0.0515837$) until the third decimal place, and is only 1 point away at the fourth decimal place.

d) Omitted due to time constraints.

Question 5

Consider the Markov chain for the Reed-Frost model with transition probability matrix as in (4.2.2). For all the numerical computations in this question, use x0 = 6, some y0 of your choice, and some fixed $\alpha \in (0.7, 0.9)$ of your choice.

a) What is the state space?

We have $\alpha = 0.85$, as in Q4, and let $y_0 = 2$.
The Reed-Frost model says we should have the state space $\{(x, y) \mid 0 \leq x, y \leq x_0\}$. However, we have $X_0 = x_0 = 6$ and $Y_0 = y_0 = 2$, and we know $X_t = X_{t+1} + Y_{t+1}$, for $t \geq 0$. Hence, once unreachable states are removed, we have the state space:

$$S = \{(x, y) \mid 0 \leq x, y \leq x_0,\ 0 \leq x + y \leq x_0\} \cup \{(6, 2)\}$$

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

A state of the form $\{(x, 0)\} \in S$ leads only to itself. This is because there are no infectives to spread the infection, so the number of susceptibles will not change. Thus these states each form their own communicating class. These communicating classes are recurrent as once we enter them we always return (in fact, we never leave).

A state of the form $\{(x, y) \mid y \neq 0\} \in S$ leads to all states $\{(x', y') \mid 0 \leq x' + y' \leq x\}$. However, none of these states will lead back to (x, y) (since we would require $x' = x + y \implies y = 0$). Any state communicates with itself by definition, as a state leads to itself in 0 steps. Thus these states also each form there own communicating class. These communicating classes are transient as once we leave them we may never return (in fact, we cannot return).

In summary:
$\{(x, 0)\} \in S$ each form their own recurrent communicating class.
$\{(x, y) \mid y \neq 0\} \in S$ each form their own transient communicating class.

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

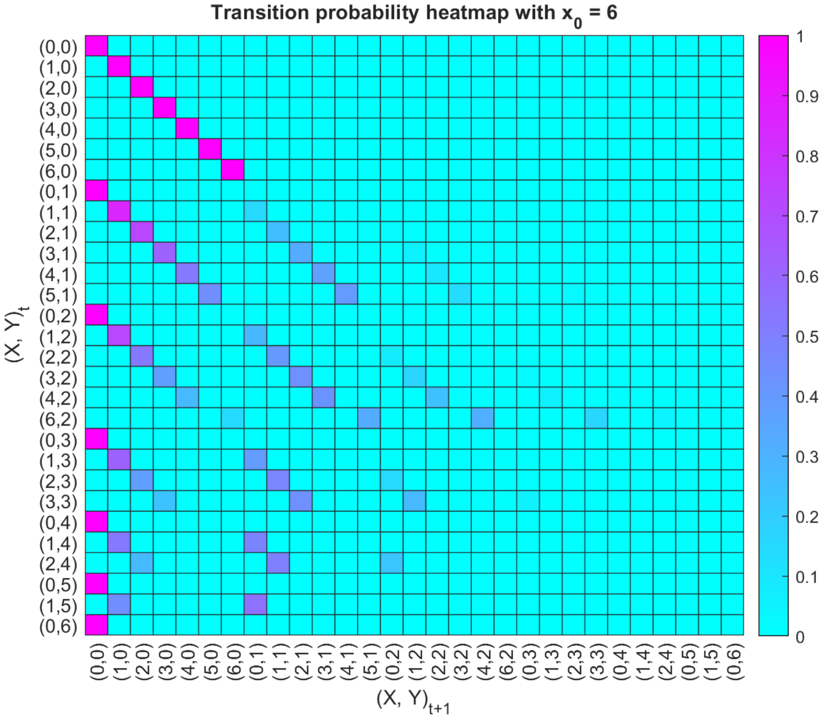
First we construct a Reed-Frost transition probability matrix, as per equation (4.2.2):

```
function P = R_F_trans_mat(x_0, alpha)
    P = zeros((x_0 + 1)^2, (x_0 + 1)^2);
    row = 0;
    for i = 0:x_0
        for k = 0:x_0
            row = row+1;
            col = 0;
            for j = 0:x_0
                for l = 0:x_0
                    col = col+1;
                    if l+j == k
                        P(row, col) = binopdf(l,k,alpha^i);
                    end
                end
            end
        end
    end
end
```

We then remove all columns and rows that correspond to states not in our state space. That is those with $x + y \geq 6$ (but we keep our starting state $(6, 2)$.

```
function P = R_F_trans_mat_y0(x_0, y_0, alpha)
    RF_rows_cols = R_F_trans_mat(x_0, alpha);
    vals = strings(0);

    RF_cols = []; % A new matrix with all the unwanted rows removed
    row = 0;
    for i = 0:x_0
        for k = 0:x_0
            row = row+1;
            if k + i <= x_0 || (k == x_0 && i == y_0)
                RF_cols = [RF_cols; RF_rows_cols(row, :)];
                vals = [vals; append('(',num2str(k),',',num2str(i),')')];
            end
        end
    end
    P = []; % The final matrix with all the unwanted columns removed too
    col = 0;
    for j = 0:x_0
        for l = 0:x_0
            col = col+1;
            if l + j <= x_0 || (l == x_0 && j == y_0)
                P = [P, RF_cols(:, col)];
            end
        end
    end

    heatmap(vals, vals, P, 'XLabel','X_{t+1}', 'YLabel','X_t', ...
        'Colormap',cool, 'CellLabelFormat','%.3g');
end
```

Calling `R_F_trans_mat_y0(6, 2, 0.85)` returns the heatmap:



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

As in Q4c, we first create a function that generates one possible outcome of the Reed-Frost model, for some x_0, y_0 and α, and returns the values of W and T.

```
function [w, t] = RF_sim_y0(x_0, y_0, alpha)
    t = 0;
    x_t = x_0;
    y_t = y_0;
    x_tminus1 = -1;
    % x_0 > 0 so we get at least one step
    y_tminus1 = 1; % just for setup
    while x_t ~= x_tminus1
        t = t+1;
        x_tminus1 = x_t;
        y_tminus1 = y_t;
        x_t = binornd(x_tminus1, alpha^y_tminus1);
        y_t = x_tminus1 - x_t;
    end
    % x_t = x_tminus1, so infection has ended at time t
    w = x_0 - x_t;
end
```

We then execute 10^6 of these simulations, and estimate $\mathbb{P}(W > 4)$ by the proportion of simulations which have $W > 4$, as so:

```
x_0 = 6;
y_0 = 2;
alpha = 0.85;
N = 10^6; % Total number of sim.s to run

gt4 = 0; % To record the number of sim.s with W > 4
for n = 1:N
    [W, T] = RF_sim_y0(x_0, y_0, alpha);
    % Add a count for all that have W > 4
    if W > 4
        gt4 = gt4 + 1;
    end
end

% P(W > 4) is approx. by the proportion of the N total sim.s that had W > 4
prob = gt4/N
```

A run of this script returned

```
prob =
   0.225828
```

Thus we have an estimate of $\mathbb{P}(W > 4) \approx 0.225828$ for the Reed-Frost model with $x_0 = 6,\ y_0 = 2$ and $\alpha = 0.85$.
Note that the numerical answer from Q4b for the Greenwood model was $\mathbb{P}(W > 4) \approx 0.0515837$. The Reed-Frost answer is about 4 times this. This is to be expected as the Reed-Frost model takes into account the number of infectives. As we start with $y_0 = 2$, within the first step there is a higher probability of infection than in the Greenwood model. This is true for any state in Reed-Frost with $Y_t > 1$, which results in a higher probability that we get a large infection ($W > 4$).

Question 6: Modelling Task

To avoid confusion between x_0 and X_0, I will be referring to the carrying capacity of the motel as c.

A day at the COVID-19 Motel:

- Test results arrive.
  - Guests who test positive - and were thus infectious the previous night - are removed to a health facility.
  - If all guests test negative, they are released into the town.
  - If the motel is now empty - either via all guests testing positive and being removed, or via all guests testing negative and being released - a new set of c guests arrive.
- Guests are tested.
- Potential infection occurs.
  - Note that no infection occurs between sets of guests, only within one particular set.

We take the variables X_t and Y_t to be the number of susceptibles and infectives, respectively, in the evening of day t. We create a Markov chain for this situation by adapting the Reed-Frost model with $x_0 = c$. We have $p = 0.1$ and $\beta = 0.05$, so $\alpha = 1 - p\beta = 0.995$.

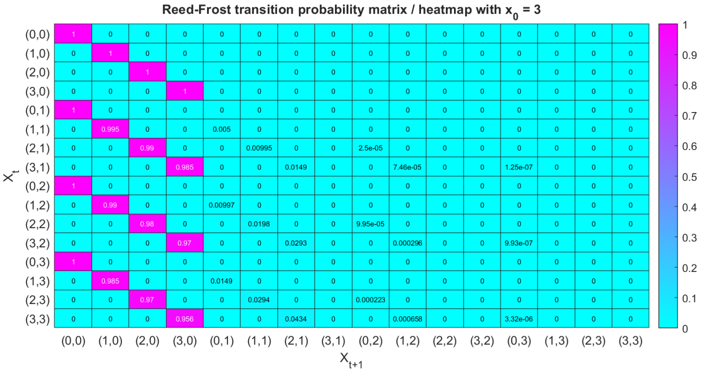
The Reed-Frost model has a state space of $\{(x, y) \mid 0 \leq x, y \leq x_0\}$. We restrict this state space to only those tuples with $1 \leq x + y \leq c$, as we will always have between 1 and c guests *total* in the hotel. Let us call this modified state space $S = \{(x, y) \mid 0 \leq x, y \leq c,\ 1 \leq x + y \leq c\}$.

We then modify the transition probabilities from those states with $x = 0$ or $y = 0$ (noting $(0, 0)$ has been excluded from the state space). These are the states which lead to a new set of guests, which requires $X_{t+1} + Y_{t+1}$ to equal c. We know "the probability of new arrivals being infected by COVID-19, is $\eta = 0.05$". So if $X_t = 0$ or $Y_t = 0$, Y_{t+1} has binomial distribution with $n = c$ and $p = \eta = 0.05$, and X_t, is then given by $c - Y_t$. This means we have the transition probability for $(x, y), (x', y') \in S$, when $x = 0$ or $y = 0$, of

$$p_{(x,y),(x',y')} = \begin{cases} \binom{c}{y'}(1-\eta)^{x'}\eta^{y'}, & \text{if } x' + y' = c \\ 0, & \text{otherwise} \end{cases}$$

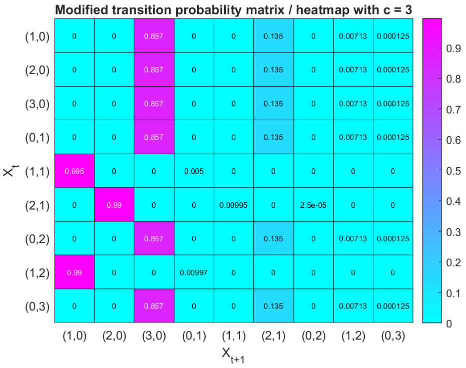
All other transition probabilities remain the same as in the Reed-Frost model.

For example, here is a heatmap of the Reed-Frost transition probability matrix for $x_0 = 3$ with $\alpha = 0.995$, and the modified version of this matrix ($c = 3,\ \alpha = 0.995,\ \eta = 0.05$):



Reed-Frost transition probability matrix / heatmap with x_0 = 3

X_t \ X_{t+1}	(0,0)	(1,0)	(2,0)	(3,0)	(0,1)	(1,1)	(2,1)	(3,1)	(0,2)	(1,2)	(2,2)	(3,2)	(0,3)	(1,3)	(2,3)	(3,3)
(0,0)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,0)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(2,0)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
(3,0)	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
(0,1)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,1)	0	0.995	0	0	0.005	0	0	0	0	0	0	0	0	0	0	0
(2,1)	0	0	0.99	0	0	0.00995	0	0	2.5e-05	0	0	0	0	0	0	0
(3,1)	0	0	0	0.985	0	0	0.0149	0	0	7.46e-05	0	0	1.25e-07	0	0	0
(0,2)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,2)	0	0.99	0	0	0.00997	0	0	0	0	0	0	0	0	0	0	0
(2,2)	0	0	0.98	0	0	0.0198	0	0	9.95e-05	0	0	0	0	0	0	0
(3,2)	0	0	0	0.97	0	0	0.0293	0	0	0.000296	0	0	9.93e-07	0	0	0
(0,3)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,3)	0	0.985	0	0	0.0149	0	0	0	0	0	0	0	0	0	0	0
(2,3)	0	0	0.97	0	0	0.0294	0	0	0.000223	0	0	0	0	0	0	0
(3,3)	0	0	0	0.956	0	0	0.0434	0	0	0.000658	0	0	3.32e-06	0	0	0



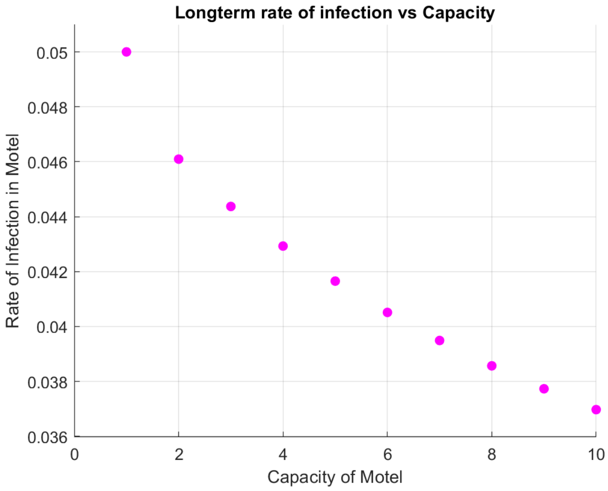
Modified transition probability matrix / heatmap with c = 3

X_t \ X_{t+1}	(1,0)	(2,0)	(3,0)	(0,1)	(1,1)	(2,1)	(0,2)	(1,2)	(0,3)
(1,0)	0	0	0.857	0	0	0.135	0	0.00713	0.000125
(2,0)	0	0	0.857	0	0	0.135	0	0.00713	0.000125
(3,0)	0	0	0.857	0	0	0.135	0	0.00713	0.000125
(0,1)	0	0	0.857	0	0	0.135	0	0.00713	0.000125
(1,1)	0.995	0	0	0.005	0	0	0	0	0
(2,1)	0	0.99	0	0	0.00995	0	2.5e-05	0	0
(0,2)	0	0	0.857	0	0	0.135	0	0.00713	0.000125
(1,2)	0.99	0	0	0.00997	0	0	0	0	0
(0,3)	0	0	0.857	0	0	0.135	0	0.00713	0.000125

Note the modified transition probability matrix is irreducible as every state can lead to all other states (one communicating class). Thus we can calculate the stationary distribution of this matrix. This is easily computed in MatLab using eigenvalues, for $\alpha = 0.995, \eta = 0.05$, and $c = 1, ..., 10$ (as required):

```
[P, S] = Q6_trans_mat(c, alpha, eta, 0);
% P is transition probability matrix, S is state space
% Want eigenvector, V, for eigenvalue = 1, s.t. P'v = v
[v, d] = eigs(P', 1, 1);
% Normalise the eigenvector to get stationary distribution, xP = x
x = transpose(abs(v/norm(v,1))); % row vector
infect_prop = S(:,2)./sum(S,2); % y/(X + y) for each state  (col. vec.)
avg_infect_prop = x*infect_prop; % E(y/(x + y) | c)
```

Tabulating and graphing the output of this code for $c = 1, ..., 10$ we get:



Capacity (c)	**Rate of infection**
1	0.05
2	0.046096
3	0.044372
4	0.042929
5	0.041652
6	0.040512
7	0.03949
8	0.038567
9	0.037732
10	0.036973

As expected, for $c > 1$, the average rate of infection for motel guests is the same or lower than $\eta = 0.05$ (the rate of infection in new arrivals). As c increases, the rate of infection decreases non-linearly.

A Monte-Carlo simulation was also run (until $t = 10^5$) and produced these similar results:

Capacity (c)	**Rate of infection**
1	0.0474
2	0.0461
3	0.0428
4	0.04362
5	0.04077
6	0.04002
7	0.03988
8	0.03882
9	0.03701
10	0.03742

Many parts of this model are unrealistic. Firstly, it was assumed all new arrivals have independent probability η of being infective. In reality, people live and travel in family groups, so these new arrivals will likely not have independent probabilities of having COVID-19.

Similarly, the Reed-Frost model assumes that, whilst in the motel, the probability of contact is $p = 0.1$. This would likely be increased between guests who know each other, even if they stay in separate rooms. It also assumes that there is no immunity to COVID-19, i.e. each susceptible is equally likely ($\beta = 0.05$) to contract the disease upon contact with an infective, which is probably not true. However, as an average, these assumptions are quite reasonable.

Finally, the main assumption of this modelling scenario is that the we are in an extended time period where the probability of having COVID-19 in a new arrival remains constant. If we consider that guests are only released into the town when they are known to *not* be infectious, it is clear that the infection rate in the town would be decreasing. If multiple towns employ the same strategy, then the overall rate of infection in the general population should decrease.