Solutions to Homework 4 - Markov Chains

1) Transforming a random process into a Markov chain. Consider the random process $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ with state space $S = \{0, 1, 2\}$. Suppose

$$P\left[X_{n+1} = j \,\middle|\, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right] = \left\{ \begin{array}{ll} P^e_{ij}, & n \text{ even} \\ P^o_{ij}, & n \text{ odd} \end{array} \right. \tag{1}$$

where $\sum_{j=0}^{2} P^e_{ij} = \sum_{j=0}^{2} P^o_{ij} = 1$, $i \in S$. As defined, $X_{\mathbb{N}}$ is not a Markov chain (MC) because the probabilities in (1) depend on the time step n (through whether n is even or odd). This contradicts the basic defining property of a MC, that requires sole dependence on i and j.

The state space of $X_{\mathbb{N}}$ is $S = \{0, 1, 2\}$ and we consider the augmented state space $\overline{S} = \{0, 1, 2, \overline{0}, \overline{1}, \overline{2}\}$ where $0, 1, 2$ are the original states when n is even and $\overline{0}, \overline{1}, \overline{2}$ accounts for the same respective states when n is odd. In this case we obtain a MC as desired, with transition probability matrix

$$\mathbf{P} = \left(\begin{array}{cccccc} 0 & 0 & 0 & P^e_{00} & P^e_{01} & P^e_{02} \\ 0 & 0 & 0 & P^e_{10} & P^e_{11} & P^e_{12} \\ 0 & 0 & 0 & P^e_{20} & P^e_{21} & P^e_{22} \\ P^o_{00} & P^o_{01} & P^o_{02} & 0 & 0 & 0 \\ P^o_{10} & P^o_{11} & P^o_{12} & 0 & 0 & 0 \\ P^o_{20} & P^o_{21} & P^o_{22} & 0 & 0 & 0 \end{array} \right).$$

2) Coin flips. Suppose that coin 1 has probability 0.7 of coming up heads whereas coin 2 has probability 0.6 of coming up heads. If the coin flipped today comes up heads, then we select coin 1 to flip tomorrow, otherwise we select coin 2 to flip tomorrow. The coin initially flipped is equally likely to be coin 1 or coin 2.

Let X_n be the coin flipped at day n. Given the previous definition of the problem, clearly $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ constitutes a MC with state space is $S = \{1, 2\}$. The transition matrix is

$$\mathbf{P} = \left(\begin{array}{cc} 0.7 & 0.3 \\ 0.6 & 0.4 \end{array} \right).$$

The initial distribution is $P[X_0 = 1] = 1/2$ and $P[X_0 = 2] = 1/2$ which we can arrange in a column vector

$$\mathbf{p}(0) = [1/2, 1/2]^T.$$

We are interested in computing the probability that the coin flipped on the third day after the initial flip is coin 1. Given our notation, this can be translated as computing the unconditional probability $P[X_3 = 1]$. To that end, we will need to compute the three-step transition probability matrix

$$\mathbf{P}^{(3)} = \mathbf{P}^3 = \mathbf{P} \times \mathbf{P} \times \mathbf{P} = \left(\begin{array}{cc} 0.667 & 0.333 \\ 0.666 & 0.334 \end{array} \right).$$

Using total probability we have

$$P[X_3 = 1] = P\left[X_3 = 1 \,\middle|\, X_0 = 1\right] P[X_0 = 1] + P\left[X_3 = 1 \,\middle|\, X_0 = 2\right] P[X_0 = 2] = P^3_{11} P[X_0 = 1] + P^3_{21} P[X_0 = 2]$$

which corresponds to the first entry of the vector

$$\mathbf{p}(3) = \left(\mathbf{P}^3\right)^T \mathbf{p}(0) = \left(\begin{array}{cc} 0.667 & 0.666 \\ 0.333 & 0.334 \end{array} \right) \left(\begin{array}{c} 1/2 \\ 1/2 \end{array} \right) = \left(\begin{array}{c} 0.6665 \\ 0.3335 \end{array} \right).$$

Accordingly, the answer is $P[X_3 = 1] = 0.6665$.

Moving on to the next question, suppose that the coin flipped on Monday comes up heads. We want to compute the probability that the coin flipped on Friday of the same week also comes up heads. Let n_0 denote the Monday when the coin was flipped. We know that $X_{n_0+1} = 1$, because the flip that took place at time n_0 came up heads. The Friday of interest corresponds to time $n_0 + 4$ and so we are asked to compute

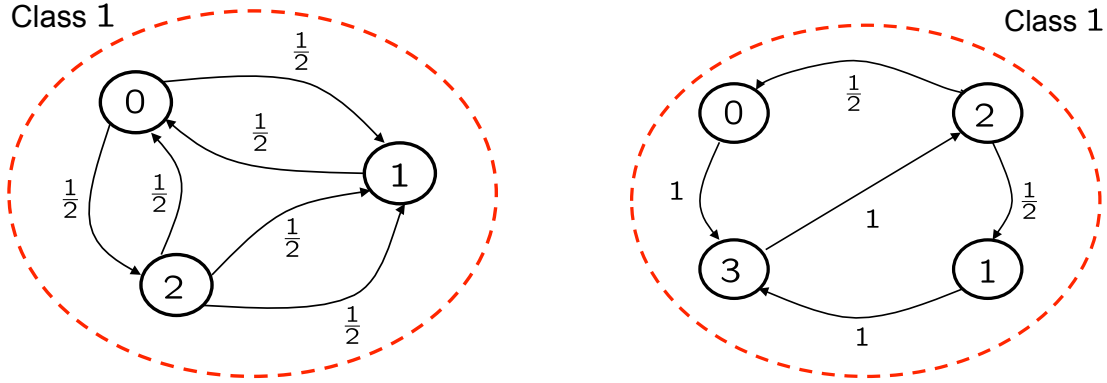$$P\left[X_{n_0+5} = 1 \,\middle|\, X_{n_0+1} = 1\right] = P^4_{11}.$$

Fig. 1. State transition diagrams for the MCs with transition probility matrices (left) $\mathbf{P}_1$ and (right) $\mathbf{P}_2$.

The four-step transition probability matrix is

$$\mathbf{P}^{(4)} = \mathbf{P}^4 = \mathbf{P} \times \mathbf{P}^3 = \begin{pmatrix} 0.6667 & 0.3333 \\ 0.6666 & 0.3334 \end{pmatrix}$$

and therefore the desired result is $\mathrm{P}\left[X_{n_0+5} = 1 \,\middle|\, X_{n_0+1} = 1\right] = 0.6667$.

3) Positive transition matrices remain positive. Let $\mathbf{P}$ be the transition probability matrix of a MC with s states, say. Suppose that for some positive integer r, $\mathbf{P}^r$ has all positive entries. We want to show that so does $\mathbf{P}^n$ for $n \geq r$. By an induction argument, it is clear that it suffices to show that $\mathbf{P}^{r+1} = \mathbf{P} \times \mathbf{P}^r$ has all positive entries.

Let $\mathbf{p}_i^T$ denote the i-th row vector of the transition matrix $\mathbf{P}$ (with entries $P_{ik} \in [0, 1]$ for all $k = 1, \ldots, s$), and let $\mathbf{p}_j^r$ denote the j-th column vector of the r-step probability transition matrix $\mathbf{P}^r$ (with entries $P_{kj}^r \in (0, 1]$ for all $k = 1, \ldots, s$). The i, j-th entry of the product $\mathbf{P}^{r+1} = \mathbf{P} \times \mathbf{P}^r$ is given by the inner product

$$P_{ij}^{r+1} = \mathbf{p}_i^T \mathbf{p}_j^r = \sum_{k=1}^{s} P_{ik} P_{kj}^r = \sum_{k \text{ s.t. } P_{ik} > 0} P_{ik} P_{kj}^r > 0.$$

The probability P_{ij}^{r+1} is positive because the last summation only involves strictly positive summands. (Notice that not all P_{ik}, $k = 1, \ldots s$, can be identically zero because we must have $\sum_{k=1}^{s} P_{ik} = 1$.) It is clear that the previous result holds for all i, j with $i = 1, \ldots, s$ and $j = 1, \ldots, s$, establishing that $\mathbf{P}^{r+1}$ will have all positive entries as desired.

4) Communication classes. A MC with state space $\mathcal{S}_1 = \{0, 1, 2\}$ has the following transition matrix

$$\mathbf{P}_1 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

The information in $\mathbf{P}_1$ is also summarized in the state transition diagram shown in Fig. 1 (left). From the diagram it is clear that all states communicate, and therefore they all belong to a single class $\mathcal{C}_1 = \{0, 1, 2\}$. The MC is irreducible, and all the states in the class have to be recurrent because the state space is finite.

A second MC is considered with state space $\mathcal{S}_2 = \{0, 1, 2, 3\}$. The transition probability matrix is

$$\mathbf{P}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and the corresponding state transition diagram is also shown in Fig. 1 (right). As in the first case we observe that all states communicate, and therefore they all belong to a singles class $\mathcal{C}_1 = \{0, 1, 2, 3\}$. The MC is irreducible, and all the states in the class have to be recurrent because the state space is finite.
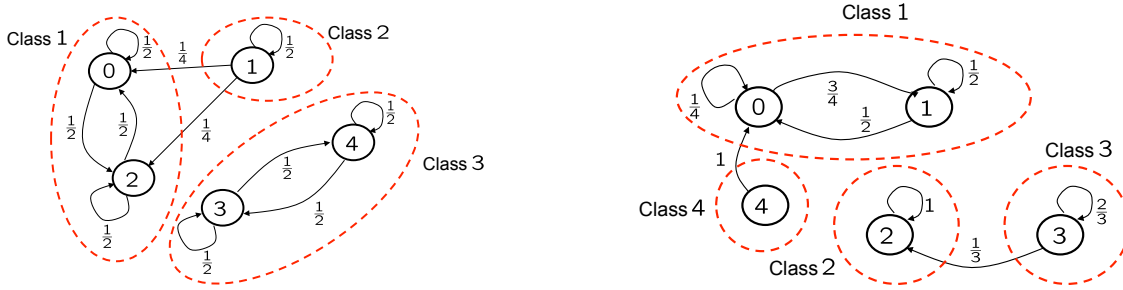
Fig. 2. State transition diagrams for the MCs with transition probility matrices (left) $\mathbf{P}_3$ and (right) $\mathbf{P}_4$.

The third considered MC has state space $\mathcal{S}_3 = \{0, 1, 2, 3, 4\}$ and transition probability matrix

$$\mathbf{P}_3 = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

The state transition diagram is shown in Fig. 2 (left). Notice that states 0 and 2 communicate, whereas state 1 is not accessible from any other state except from itself. Finally states 3 and 4 also communicate but are not accessible from any other state. Accordingly, in this case we encounter three communication classes, $\mathcal{C}_1 = \{0, 2\}$, $\mathcal{C}_2 = \{1\}$ and $\mathcal{C}_3 = \{3, 4\}$. The states in the classes $\mathcal{C}_1$ and $\mathcal{C}_3$ are all recurrent because if the MC is initialized in either class, it will stay there forever. On the other hand, state 1 (and hence class $\mathcal{C}_2$) is transient, because starting from state 1 the MC will almost surely leave the state and never return.

For the last MC the state space is $\mathcal{S}_4 = \{0, 1, 2, 3, 4\}$. The transition probability matrix is given by

$$\mathbf{P}_4 = \begin{pmatrix} 1/4 & 3/4 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

while the corresponding state transition diagram is shown in Fig. 2 (right). From the state diagram we observe that states 0 and 1 communicate and form the first class $\mathcal{C}_1 = \{0, 1\}$, whose states are recurrent. State 2 is an absorbing state, therefore it is recurrent and it forms a second class $\mathcal{C}_2 = \{2\}$. Finally states 3 and 4 are not accesible from any other state, and they separately form two more classes $\mathcal{C}_3 = \{3\}$ and $\mathcal{C}_4 = \{4\}$. These states are clearly transient.

5) Multi-step, unconditional, and joint MC probabilities. Suppose that $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ is a MC with state space $S = \{1, 2\}$, transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 1/5 & 4/5 \\ 2/5 & 3/5 \end{pmatrix}$$

and initial distribution $\mathrm{P}\left[X_0 = 1\right] = 3/4$ and $\mathrm{P}\left[X_0 = 2\right] = 1/4$.

We first want to compute the conditional probability $\mathrm{P}\left[X_3 = 1 \,\middle|\, X_1 = 2\right] = P_{21}^2$. To that end, the two-step transition probability matrix $\mathbf{P}^{(2)}$ is

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \mathbf{P} \times \mathbf{P} = \begin{pmatrix} 0.36 & 0.64 \\ 0.32 & 0.68 \end{pmatrix}.$$

Hence, the desired result is $\mathrm{P}\left[X_3 = 1 \,\middle|\, X_1 = 2\right] = P_{21}^2 = 0.32$.

Second, from the Markov property we have

$$\mathrm{P}\left[X_3 = 1 \,\middle|\, X_2 = 1, X_1 = 1, X_0 = 2\right] = \mathrm{P}\left[X_3 = 1 \,\middle|\, X_2 = 1\right] = P_{11} = \frac{1}{5}.$$

Next, the unconditional probability $P[X_2 = 2]$ can be calculated from the initial distribution $\mathbf{p}_0 = [3/4\ 1/4]^T$ and the second row of the transposed two-step transition matrix $(\mathbf{P}^2)^T$. In fact, since

$$P[X_2 = 2] = P[X_2 = 2 \mid X_0 = 1] P[X_0 = 1] + P[X_2 = 2 \mid X_0 = 2] P[X_0 = 2] = P_{12}^2 \times \frac{3}{4} + P_{22}^2 \times \frac{1}{4}$$

we finally obtain

$$P[X_2 = 2] = (0.64\ 0.68) \begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix} = 0.65.$$

To evaluate the required joint probability, use the definition of conditional probability, i.e.,

$$P[X_0 = 1, X_2 = 1] = P[X_2 = 1 \mid X_0 = 1] P[X_0 = 1] = P_{11}^2 \times P[X_0 = 1] = 0.36 \times 0.75 = 0.27.$$

Finally, to find the unconditional expectation $\mathbb{E}[X_2]$ recall we already know $P[X_2 = 2] = 0.65$, which implies $P[X_2 = 1] = 0.35$. Accordingly, the desired expectation is

$$\mathbb{E}[X_2] = 1 \times 0.35 + 2 \times 0.65 = 1.65.$$

6) Something the Markov property does not imply. Suppose that $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ is a MC, $j, i_{n-1}, \ldots, i_1, i_0$ are elements of the state space S, and A is a subset of the state space. In this problem we will find an example for which

$$P[X_{n+1} = j \mid X_n \in A, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0] \neq P[X_{n+1} = j \mid X_n \in A].$$

Such an example can be constructed as follows; suppose that a single state i that belongs to the subset A is accessible from the state i_{n-1}. So for this particular MC, the event of interest satisfies

$$\{X_{n+1} = j \mid X_n \in A, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$$

because we know that given $X_{n-1} = i_{n-1}$, the only accessible state from A at time instant n (and generally any other time instant) is i. So from the Markov property we have

$$P[X_{n+1} = j | X_n \in A, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = P[X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = P_{ij}$$

Suppose that $A = S$, then because $\{X_n \in S\}$ is the sure event we have

$$P[X_{n+1} = j \mid X_n \in A] = P[X_{n+1} = j \mid X_n \in S] = P[X_{n+1} = j] = p_j(n+1)$$

which is an uncoditional probability that depends on the initial distribution $\mathbf{p}(0)$, and the time instant $n + 1$. It should be clear now that the quantities P_{ij} and $p_j(n+1)$ will be different in general, and thus making the required point.

For a concrete example, consider a MC with state space $S = \{1, 2\}$, transition matrix

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}$$

and initial distribution $P[X_0 = 1] = 1/4$, $P[X_0 = 2] = 3/4$. Pick $A = \{1, 2\}$ and note that only state $1 \in A$ is accessible from state 2. Therefore we can compute, for $n = 1$ say

$$P[X_2 = 1 \mid X_1 \in A, X_0 = 2] = P[X_2 = 1 \mid X_1 = 1, X_0 = 2] = P_{11} = \frac{1}{2}.$$

On the other hand,

$$P[X_2 = 1 \mid X_1 \in A] = P[X_2 = 1] = P_{11}^2 P[X_0 = 1] + P_{21}^2 P[X_0 = 2] = \frac{3}{4} \times \frac{1}{4} + \frac{1}{2} \times \frac{3}{4} = \frac{9}{16}.$$

7) A branching process.

A) Is the count of women a Markov chain? The process $X_{\mathbb{N}}$ which indicates the total amount of women across generations is clearly a MC with state space $S = \{0, 1, 2, \ldots\}$, since it satisfies the Markov property. To see this, note that considering a particular generation, the number of women in the next one depends exclusively on the

number of women in the present generation and is independent of the past history of women population. Formally, we can write

$$X_{n+1} = \sum_{i=1}^{X_n} D_i \quad \text{if } X_n > 0$$

and $X_{n+1} = 0$ if $X_n = 0$.

Exploring some of the transition probabilities of this MC it is immediate to realize that

$$P_{0j} = \mathrm{P}\left[X_{n+1} = j | X_n = 0\right] = \begin{cases} 0 & \text{if } j \neq 0 \\ 1 & \text{if } j = 0 \end{cases} \tag{2}$$

since once the total population of women vanishes it cannot increase in the future. Similarly, we can assert that

$$P_{1j} = \mathrm{P}\left[X_{n+1} = j | X_n = 1\right] = \mathrm{P}\left[\sum_{i=1}^{X_n} D_i = j | X_n = 1\right] = \mathrm{P}\left[D_1 = j\right] = p_j, \ j \geq 0$$

The above expression has a simple interpretation, the amount of women in the next generation depends only on the number of daughters the only woman in the present generation has. Likewise, for $i = 2$ we can write

$$P_{2j} = \mathrm{P}\left[\sum_{i=1}^{X_n} D_i = j | X_n = 2\right] = \mathrm{P}\left[D_1 + D_2 = j\right] = \sum_{k=0}^{j} p_k p_{j-k}$$

which corresponds to the probability of having j women in the next generation if there are two in the present one. This can be obtained as the sum of the probabilities that the first woman has no daughters and the second one has j, plus the first woman having one daughter and the second one having $j-1$, and so on. The assumed independence between D_1 and D_2 allows us to write this as the product of both probabilities. Hence, we see that calculating P_{2j} requires solving a discrete convolution between the pmf of D_1 and the pmf of D_2 (which are the same function since the D_i are i.i.d.). In general, calculating P_{ij} requires solving an i-term convolution whose analytical difficulty depends on the distribution of D_i. (See also exercise 7 in the practice midterm for the general expression.)

Nevertheless, we can still characterize even more transition probabilities. For example,

$$P_{i0} := \mathrm{P}\left[X_{n+1} = 0 | X_n = i\right] = \prod_{k=1}^{i} \mathrm{P}\left[D_k = 0\right] = \prod_{k=1}^{i} p_0 = p_0^i$$

is the probability of the population becoming extinct in the next generation given that there are i individuals in the current one. For this to happen, every woman must bear no daughters and, by independence, the above expression follows.

We can also deduce that the probability of a state transitioning into itself is positive for all states i. We already know this is true for $i = 0$ and, for $i > 0$ note that one possible way of maintaining the same amount of women through a generation is that every woman has exactly one daughter. This motivates the following inequality

$$P_{ii} = \mathrm{P}\left[X_{n+1} = i \,\middle|\, X_n = i\right] \geq p_1^i > 0, \ i \geq 0.$$

Here we are assuming that the probability of bearing exactly one daughter is strictly positive.

While state 0 (extinction) is an absorbing state and hence clearly recurrent, for all other states $i \neq 0$ there is a positive probability p_0^i of going from i to 0 which is an absorbing state. Accordingly, every state different from 0 is a transient state and the MC is neither irreducible nor recurrent.

B) Is the count of women classified by mitochondrial DNA type a Markov chain? Consider now the amount of women X_{rn} with mitochondrial DNA of type r, the process $X_{r\mathbb{N}}$ is *not* a MC. To see this, focus on the state 0, that is, the state that indicates that there are no women of type r in the current generation. This may happen as a consequence of two very different situations. It might be the case that the type r existed in the past and is now extinct, in which case, $\mathrm{P}\left[X_{rm} > 0 \,\middle|\, X_{rn} = 0\right] = 0$, for all $m > n$. However, it might be the case that the mutation type r has not been created yet. In this case, there is positive probability of having $X_{rm} > 0$ for some future generation. Consequently, the transition probabilities out of the state 0 depend on the past. Since the process does not satisfy the Markov property, it is not a MC. Note that this is the only state that presents this inconvenience.

Suppose now that we are given the information that at some time n, $X_{rn} > 0$. Define the random process $X_{r,n:\infty} := X_{rn}, X_{r,n+1}, \ldots$ that starts at time n with the information that $X_{rn} > 0$. This process is clearly a MC

since the state 0 has now only one interpretation, i.e. an absorbing state as the one described for the process $X_{\mathbb{N}}$ with $P_{0j} = 0$ for any $j \neq 0$ and $P_{00} = 1$. Slightly abusing notation, here P_{ij} denotes the transition probability from state i to state j in the process $X_{r,n:\infty}$. The transition probabilities out of the state 1 are different from the ones in $X_{\mathbb{N}}$. In particular,

$$P_{1j} = (1 - q)p_j, \ j > 0$$

meaning, the probability of having exactly j daughters without a mutation in their mitochondrial DNA. For transitioning from 1 to 0 we have

$$P_{10} = p_0 + (1 - p_0)q = q + (1 - q)p_0$$

where both expressions are trivially equal but they arise from different arguments. The first one expresses that the population of women of type r becomes extinct if one of two mutually exclusive events occur: the woman has no daughters or she has daughters and they suffer a DNA mutation. The second expression indicates the probability of the union of the events that the daughters present the mutation or that they do not present the mutation but the woman bears no daughter. Note that $P_{10} \neq p_0 + q$ since the events of having no daughters and presenting the mutation in the DNA are not disjoint.

Appealing to independence, we readily obtain

$$P_{i0} = P_{10}^i = \left(p_0 + (1 - p_0)q\right)^i, \ i \geq 0$$

since for the population of type r to become extinct every woman must have either no daughters or some daughters and present the mutation.

By following a reasoning analogous to the one performed for $X_{\mathbb{N}}$ in Part A, we can see that there is a strictly positive probability of self transitioning for all states i, that is,

$$P_{ii} \geq ((1 - q)p_1)^i > 0, \ i \geq 0.$$

Similarly, recurrence can be analyzed to find that 0 is the only recurrent state, just as in the previous process. Again, the MC is neither irreducible nor recurrent.

C) System simulation. Next, you can find a Matlab script to perform the system simulation, and generate the plots requested for Parts *D* and *E*.

```
close all; clear;clc;
X_o = 100;
max_t = 50;
max_types = 1000; % a safely large number (matlab reallocates size if this
                  % is not enough)
mu = 1.05;
q = 10^-2;

X=zeros(max_types, max_t); % preallocate vector for population size
number_of_types=zeros(1, max_t); % preallocate vector for number of types
X(1:X_o,1) = 1;
number_of_types(1)=X_o; %initialize first generation
number_of_extinct_types=zeros(1,max_t);


for n=2:max_t
    disp(['n=',num2str(n)])
    number_of_types(n)=number_of_types(n-1);
    for type = 1:number_of_types(n-1);
        for i = 1:X(type,n-1)
            daughters = poissrnd(mu,1,1); % daw number of daughters
            mutation = binornd(1,q,1,1); % draw mutation indicator
            if mutation
```

```
                    number_of_types(n) = number_of_types(n)+1;
                    X(number_of_types(n),n) = daughters;
                else
                    X(type,n) = X(type,n) + daughters;
                end % if-else
            end % i
            if X(type,n)== 0
                number_of_extinct_types(n)=number_of_extinct_types(n)+1;
            end % if
        end % type
end % n

figure
plot(1:max_t, X)
xlabel('generation','FontSize',14)
ylabel('number of women of each type','FontSize',14)
title('$q=10^{-2},$ $X_{0}=100,$ $n=50$','FontSize',14,'Interpreter','latex')

figure
stairs(1:max_t, X')
xlabel('generation','FontSize',14)
ylabel('number of women of each type','FontSize',14)
title('$q=10^{-2},$ $X_{0}=100,$ $n=50$','FontSize',14,'Interpreter','latex')

figure
stairs(1:max_t, [number_of_types;number_of_extinct_types]','LineWidth',2)
xlabel('generation','FontSize',14)
ylabel('number of women of each type','FontSize',14)
title('$q=10^{-2},$ $X_{0}=100,$ $n=50$','FontSize',14,'Interpreter','latex')
axis([0 50 0 number_of_types(end)])
legend('number of types','number of extinct types','Location','Best')

figure
bar(1:number_of_types(end), X(1:number_of_types(end),max_t),'r')
xlabel('types','FontSize',14)
ylabel('number of women of each type','FontSize',14)
title('Histogram of the final number of women of each type','FontSize',14)
```

D) Simulation tests one. Here we use the script in Part *C* to run a simulation with $q = 10^{-2}$ as the rate of mutation, generation 0 initialized with $X_0 = 100$ individuals all having different mitochondrial DNA types. The branching process evolves for a horizon of $n_{\max} = 50$ generations. The results are depicted in in Fig. 3.

E) Simulation tests two. We repeat Part *D* with rate of mutation $q = 0$ and $X_0 = 400$ individuals of different types. This tells you about the chances of your direct female line surviving into the next 10 centuries. The results are depicted in in Fig. 4. Notice that most of the types go extinct, a few have a moderate number of individuals and 1 or 2 have a large number of individuals. This means that far into the future, most of your direct female lines will be extinct, but one of you will have a very large number of survivors. But who among you is the one surviving into the future is determined by chance.

F) Expected value of the number of direct line female descendants. As mentioned in Part *A*, the number of individuals in the $n + 1$-st generation can be written as

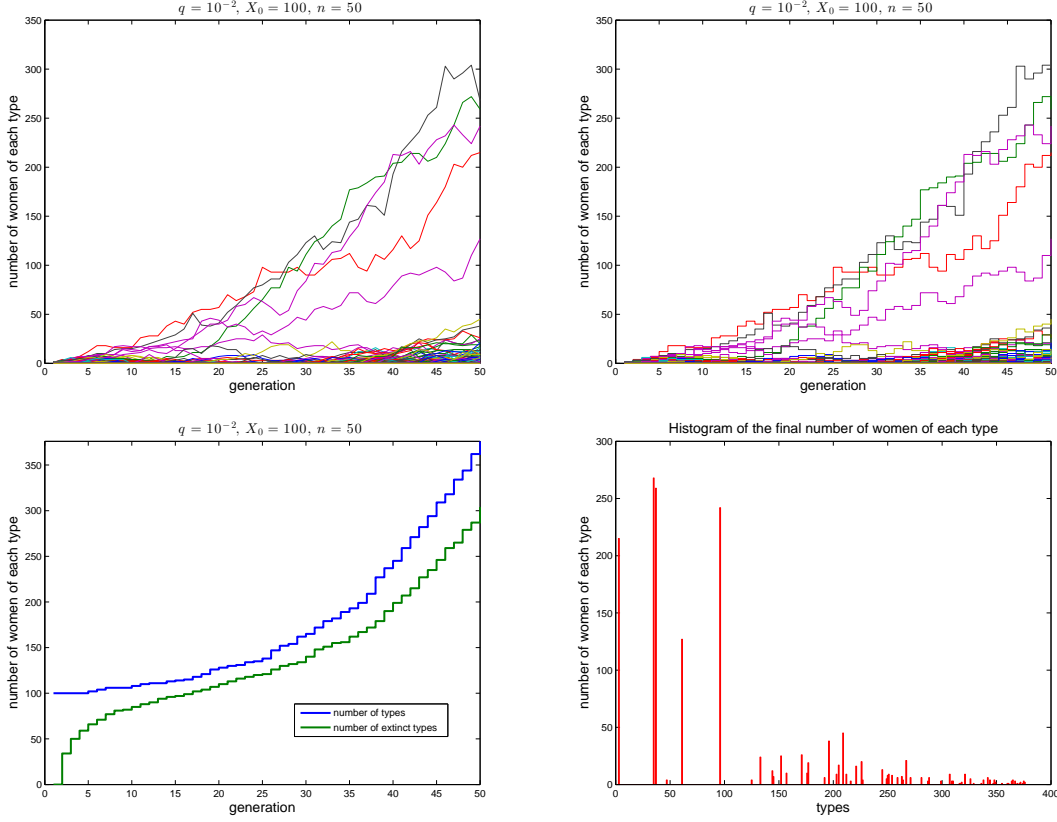$$X_{n+1} = \sum_{i=1}^{X_n} D_i. \tag{3}$$

Fig. 3. Simulation results starting from 100 types and after 50 generations, where $q = 0.01$. (Part D). (top-left) Evolution of types (using plot); (top-right) Evolution of types (using stairs); (bottom-left) Number of overall types and extinct types over generations; and (bottom-right) Histogram of the final number of each type.

From (3) we can compute the expected value of the total population in the n-th generation, i.e. $\mathbb{E}[X_n] := \mu_n$. Since we are dealing with a summation, the linearity of the expected value operator will be a useful tool. However, it is important to note that X_{n-1} as well as D_i are random variables. Hence, we are dealing with a summation of a random number of RVs, also known as compound RV. The good news is that X_{n-1} and D_i are independent. This is a direct consequence of the problem definition, since the child bearing probability of a given woman does not depend on the total population of her generation. To determine μ_n, begin by conditioning on the value of X_{n-1} to arrive at

$$\mathbb{E}\left[\sum_{i=1}^{X_{n-1}} D_i \,\Big|\, X_{n-1} = k\right] = \mathbb{E}\left[\sum_{i=1}^{k} D_i \,\Big|\, X_{n-1} = k\right] = \mathbb{E}\left[\sum_{i=1}^{k} D_i\right] = \sum_{i=1}^{k} \mathbb{E}[D_i] = \sum_{i=1}^{k} \nu = k\nu.$$

Now using iterated expectations we can determine $\mathbb{E}[X_n]$ through the following recursive relationship

$$\mathbb{E}[X_n] = \sum_{k=0}^{\infty} \mathbb{E}[X_n | X_{n-1} = k] \mathrm{P}[X_{n-1} = k] = \sum_{k=0}^{\infty} k\nu \mathrm{P}[X_{n-1} = k] = \nu \mathbb{E}[X_{n-1}].$$

We can apply this result repeatedly, since its true for all $n > 0$,

$$\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}]\nu = \mathbb{E}[X_{n-2}]\nu^2 = \cdots = \mathbb{E}[X_0]\nu^n = X_0\nu^n,$$

where $\mathbb{E}[X_0] = X_0$. Note that in this way we have written the expected value of the total population in any future generation as a function of the generation, the initial population and the average fertility of every woman. In particular, if the number of daughters of every woman is modeled as Poisson with parameter λ as we have done so far, we have
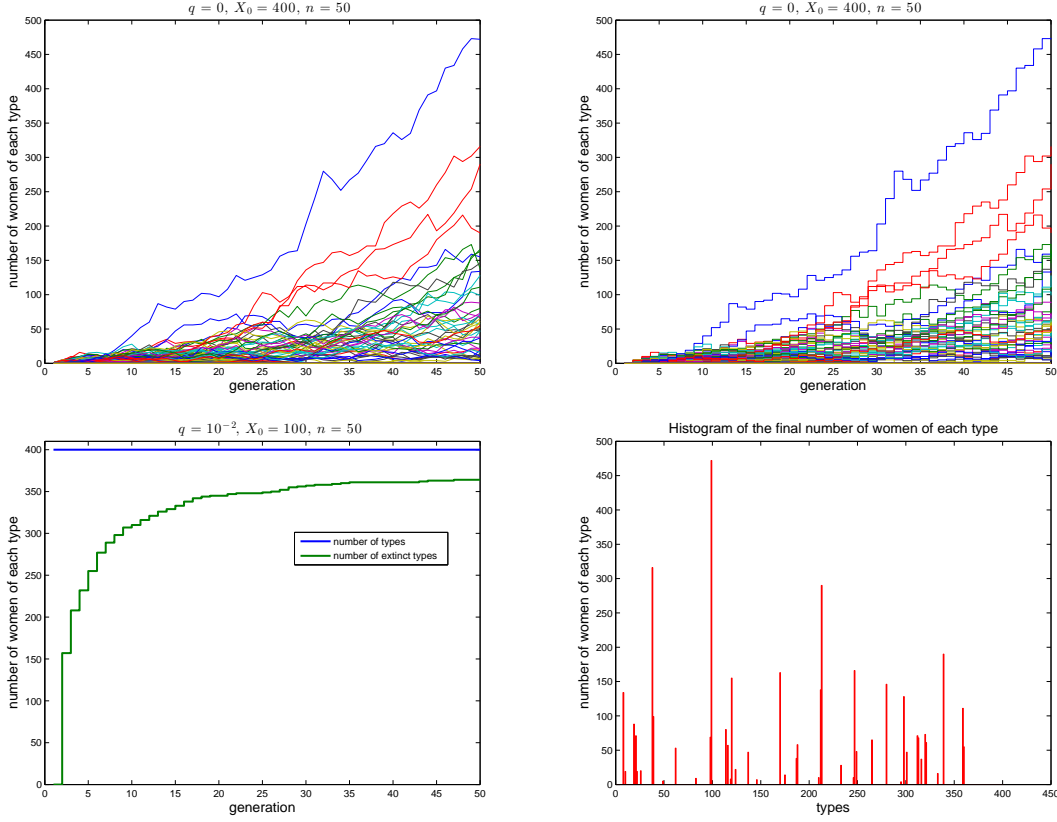
$$\mathbb{E}[X_n] = X_0\lambda^n. \tag{4}$$

Fig. 4. Simulation results starting from 400 types and after 50 generations, where $q = 0$. (Part *E*). (top-left) Evolution of types (using plot); (top-right) Evolution of types (using stairs); (bottom-left) Number of overall types and extinct types over generations; and (bottom-right) Histogram of the final number of each type.

If we fix $\lambda > 1$ as in our previous simulations, we have an exponential increase in the expected value of the population with time. To illustrate the validity of this result, Figure 5 depicts the total population and the expected value as a function of time for the first of the simulations already analyzed. We see that the theoretical result is validated by this particular realization of the process.

Following an analogous procedure, we arrive to the conclusion that

$$\mu_{rn} = \mathbb{E}\left[X_{rn}\right] = \mathbb{E}\left[X_{r0}\right]\nu_r^n = \nu_r^n = (1-q)^n\nu^n, \tag{5}$$

where assumed that every DNA type is associated with an original population of one woman (this gives the expected number of descendants of any given generation 0 individual).

G) Extinction in probability and almost sure extinction. From Markov's inequality, since the RVs X_{rn} are nonnegative we have that for arbitrary $\epsilon > 0$, and all $n \geq 0$

$$\mathrm{P}\left[X_{rn} \geq \epsilon\right] \leq \frac{\mu_{rn}}{\epsilon} = \frac{\nu_r^n}{\epsilon}.$$

If $\nu_r < 1$, taking limits as $n \to \infty$ yields

$$\lim_{n\to\infty}\mathrm{P}\left[X_{rn} \geq \epsilon\right] \leq \lim_{n\to\infty}\frac{\nu_r^n}{\epsilon} = 0 \;\Rightarrow\; \lim_{n\to\infty}\mathrm{P}\left[X_{rn} < \epsilon\right] = 1.$$

This establishes that for $\nu_r < 1$, type r goes extinct in probability.

H) Probability of eventual extinction $(q = 0)$. Fix $q = 0$. Denote as $P_e(j)$ the probability of eventual extinction of mitochondrial type r, when the number of generation 0 type r individuals is $X_{r0} = j$, i.e.,

$$P_e(j) := \lim_{n\to\infty}\mathrm{P}\left[X_{rn} = 0 \,\middle|\, X_{r0} = j\right] := \mathrm{P}\left[X_{r\infty} = 0 \,\middle|\, X_{r0} = j\right].$$
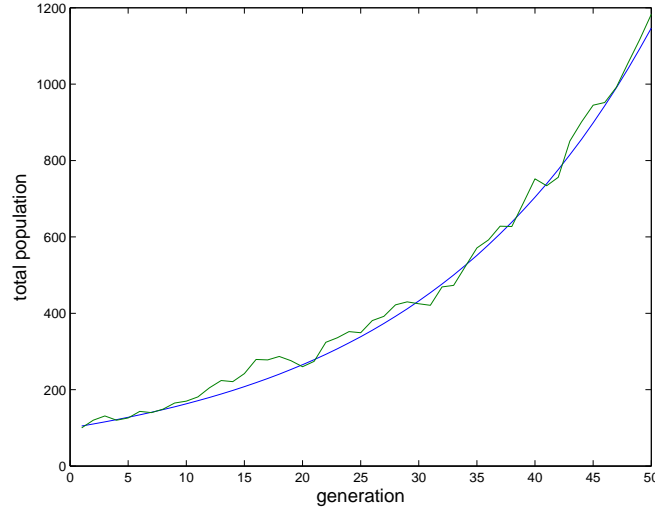
Fig. 5. The expected total population is plotted in blue as described by (4). In green, the total population from the first simulation ($q = 0.01$, $\lambda = 1.05$, $X_0 = 100$) is plotted. The numerical results validate the theoretical analysis (Part *F*).

We already showed that for $\nu < 1$, then $P_e(j) = 1$ independent of j. When $\nu > 1$ is turns out that $P_e(j) < 1$, and the equation determining $P_e(1)$ can be obtained by conditioning on the number of daughters in the first generation, and using the law of total probability. Upon conditioning on $X_{r1} = j$, note that

$$\mathrm{P}\left[X_{r\infty} = 0 \,\middle|\, X_{r1} = j, X_{r0} = 1\right] = \mathrm{P}\left[X_{r\infty} = 0 \,\middle|\, X_{r1} = j\right] = \mathrm{P}\left[X_{r\infty} = 0 \,\middle|\, X_{r0} = 1\right]^j = [P_e(1)]^j.$$

Given that at generation 1 there are j individuals, the key observation here is that from the Markov property the probability of eventual extinction is the probability that each of these j individual lines goes extinct. Because these events are independent, we have that $\mathrm{P}\left[X_{r\infty} = 0 \,\middle|\, X_{r1} = j\right] = [P_e(1)]^j$. Now we apply the law of total probability to obtain

$$P_e(1) = \sum_{j=0}^{\infty} \mathrm{P}\left[X_{r\infty} = 0 \,\middle|\, X_{r1} = j, X_{r0} = 1\right] \mathrm{P}\left[X_{r1} = j \,\middle|\, X_{r0} = 1\right] = \sum_{j=0}^{\infty} [P_e(1)]^j \, p_j$$

as desired. If in general one has $X_{r0} = j$ individuals in generation 0, by independence it follows that

$$P_e(j) = [P_e(1)]^j$$

since eventual extinction of the whole population necessitates the extinction of all j individual lines.