Week 4: Markov chains
Branching processes and mitochondrial DNA
Solutions

A Is the total number of women a Markov chain? Yes, the process $\{X_n\}_{n\in\mathbb{N}}$ is an MC because it has the Markov property, i.e., it is "memoryless." Indeed, the number of women in a generation depends only on the number of women in the previous generation, since only they give birth to the girls that form the new generation.

The transition probability P_{0j}, P_{1j}, P_{i0}, and P_{ii} are given by

$$P_{0j} = \mathbb{P}\left[X_{n+1} = j \mid X_n = 0\right] = \begin{cases} 0, & j \neq 0 \\ 1, & j = 0 \end{cases}$$ [no mother, no descendent]

$$P_{1j} = \mathbb{P}\left[X_{n+1} = j \mid X_n = 1\right] = \mathbb{P}\left[D_1 = j\right] = p_j, \; \forall j$$ [the number of women in next generation is the number of daughters of the only mother in the current generation]

$$P_{i0} = \mathbb{P}\left[X_{n+1} = 0 \mid X_n = i\right] = \prod_{k=1}^{i} \mathbb{P}\left[D_k = 0\right] = \prod_{k=1}^{i} p_0 = {p_0}^i$$ [none of the mothers should bear daughters and these events are assumed to be independent]

The probability P_{ii} of staying in the same state can be bounded by finding a specific event for which the MC would not change state. For instance, if each of the mothers in a generation has exactly one daughter, then the MC would remain in the same state. This is not the only condition for which this happens, but it allows us to derive a lower bound as in

$$P_{ii} = \mathbb{P}\left[X_{n+1} = i \mid X_n = i\right] > \prod_{k=1}^{i} \mathbb{P}\left[D_k = 1\right] = p_1^i > 0.$$

Finally, note that this MC is not recurrent. The reason is that 0 is an absorbing state, i.e., once there is no mother, there will never again be children. Formally, it holds for all $i > 0$ that $P_{i0} = p_0^i > 0$, i.e., there is a positive probability of jumping to state 0 from every state. However, $P_{0j} = 0$ for any $j \neq 0$. In other words, we can go into state 0 but we can never come out of it. More systematically, the state 0 forms a recurrent class. However, all other states $i > 0$ form a transient class. Hence, the MC cannot be recurrent.

B Is the number of women of a certain DNA type a Markov chain? The obstacle to the process $\{X_{nr}\}_{n\in\mathbb{N}}$ being Markovian is the state zero. The issue is that $X_{nr} = 0$ can mean one of two things: (i) either type r has become extinct or (ii) type r has not occurred yet. Moreover, P_{0j} is affected by this information that is not captured only by the state being 0. Indeed, if type r has become extinct, then $P_{0j} = 0$ for all $j > 0$. On the other hand, if type r has not yet existed, i.e., the number of types is less than r, then P_{0j} may be positive for all j. Therefore, the memoryless property does not hold and $\{X_{nr}\}_{n\in\mathbb{N}}$ is not a MC.

In contrast, the process $\{\hat{X}_{ir}\}_{i\in\mathbb{N}}$ is an MC since it has the memoryless property, now that state zero can only occur if type r has become extinct.

The transition probabilities P_{0j} and P_{1j} are given by

$$P_{0j} = \mathbb{P}\left[\hat{X}_{(n+1)r} = j \mid \hat{X}_{nr} = 0\right] = \begin{cases} 0, & j \neq 0 \\ 1, & j = 0 \end{cases}$$

$$P_{1j} = \mathbb{P}\left[\hat{X}_{(n+1)r} = j \mid \hat{X}_{nr} = 1\right] = \begin{cases} (1-q)p_j, & j \neq 0 \\ p_0 + (1-p_0)q, & j = 0 \end{cases}$$

We next show how to calculate this last probability by finding P_{i0}. The probability that a generation has no women of type r given that the previous generation has i women is the same as the probability of each of the i women either (i) having no daughters or (ii) have daughters of another type (mutation). We can evaluate this using total probability and considering each of these event separately:

$$\begin{aligned} P_{i0} &= \mathbb{P}\left[\hat{X}_{(n+1)r} = 0 \mid \hat{X}_{nr} = i\right] \\ &= \mathbb{P}\left[\hat{X}_{(n+1)r} = 0 \mid \hat{X}_{nr} = i, \text{mutation}\right] \mathbb{P}\left[\text{mutation}\right] \\ &+ \mathbb{P}\left[\hat{X}_{(n+1)r} = 0 \mid \hat{X}_{nr} = i, \text{no mutation}\right] \mathbb{P}\left[\text{no mutation}\right] \\ &= 1 \times q + p_0 \times (1-q) \end{aligned}$$

Finally, we can use the same approach as in part A to obtain that $P_{ii} > p_1^i(1-q) > 0$. Moreover, for the same reason as part A, this MC is not recurrent.

C System simulation.

Refer to parts D and E.

D Simulation tests one.

The MATLAB code for the simulation experiment is given below.

```
% Delete all variables and close figures
clear all
close all

X0 = 100;            % Number of individuals in the first generation
max_t = 50;          % Time limit of the simulation
max_types = 1000;    % Maximum number of types (a safely large number,
                     % MATLAB will reallocate the vector if this is not enough)

mu = 1.05;           % Poisson process rate
q = 10^-2;           % Rate of mutation

% Preallocate output vectors
X = zeros(max_types,max_t);                   % Women per type at each instant
number_of_types = zeros(max_t,1);             % Number of types per instant
number_of_extinct_types = zeros(max_t,1);     % Number of extinct types per instant

% Initialization
X(1:X0,1) = 1;                % Start with X0 women, one of each type
number_of_types(1) = X0;

% Simulation
for n = 2:max_t
    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);   % Draw number of daughters
            mutation = binornd(1,q,1,1);    % Draw mutation indicator

            % Check if a mutation occured
            if mutation == 1
                % Daughters are of a new type
                number_of_types(n) = number_of_types(n) + 1;
                X(number_of_types(n),n) = daughters;
            else
                % Daughters are of same type as mother
                X(type,n) = X(type,n) + daughters;
            end
        end

        % Check if type has gone extinct
        if X(type,n) == 0
            number_of_extinct_types(n) = number_of_extinct_types(n) + 1;
        end
    end
end


% Number of women per type
h1 = figure();
stairs(1:max_t, X', 'LineWidth', 2);
xlabel('Generation');
ylabel('Number of women per type');
grid;

% Number existing and extinct types
h2 = figure();
plot(1:max_t, [number_of_types - number_of_extinct_types, number_of_extinct_types], 'LineWidth', 2);
xlabel('Generation');
ylabel('Number of types');
grid;
legend('Number of types in the population', 'Number of extinct types');

% Final number of women per type
h3 = figure();
bar(1:number_of_types(end), X(1:number_of_types(end), max_t));
xlabel('Types');
ylabel('Number of women');
grid;


%%% Export figure %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
set(h1,'Color','w');
export_fig(h1, '-q101', '-pdf', 'HW4_D1.pdf');

set(h2,'Color','w');
export_fig(h2, '-q101', '-pdf', 'HW4_D2.pdf');

set(h3,'Color','w');
export_fig(h3, '-q101', '-pdf', 'HW4_D3.pdf');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The results are show in Figures **??**–**??**.

E Simulation tests two. Using the same code as above, modifying only the parameters q and X_0, we obtain Fig. **??**–**??**.

F Expected value of the number of direct line female descendants. The exercise states that

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

Using (1), we obtain that the expected number of individuals in generation $n+1$ is

$$\begin{aligned}
\mathbb{E}\left[X_{n+1}\right] = \mathbb{E}\left[\sum_{i=1}^{X_n} D_i\right] &= \mathbb{E}_{X_n}\left[\mathbb{E}\left[\sum_{i=1}^{X_n} D_i \mid X_n\right]\right] \\
&= \mathbb{E}_{X_n}\left[\sum_{i=1}^{X_n} \mathbb{E}\left[D_i \mid X_n\right]\right] \\
&= \mathbb{E}_{X_n}\left[\sum_{i=1}^{X_n} \mathbb{E}\left[D_i\right]\right] \\
&= \mathbb{E}_{X_n}\left[\sum_{i=1}^{X_n} \nu\right] \\
&= \nu\, \mathbb{E}_{X_n}\left[\sum_{i=1}^{X_n} 1\right] \\
&= \nu\, \mathbb{E}\left[X_n\right]
\end{aligned}$$

Thus, we have the recursion $\mathbb{E}\left[X_{n+1}\right] = \nu\, \mathbb{E}\left[X_n\right]$ with initial state $\mathbb{E}\left[X_0\right] = X_0$, give that X_0 is a deterministic quantity. The desired result follows readily. Note that we could have used Wald's equation from the beginning to get the recursion.

For parts D and E, we have that $\nu = 1.05$, so that $\mu_n = X_0 \times 1.05^n$. In other words, the expected size of the population should grow exponentially with n. This is consistent with Figures **??** and **??**: although most of the types go extinct, the numbers for those who do not increase *exponentially* fast.

Similar to previous part or directly from Wald's equation, we obtain

$$\mathbb{E}\left[X_{nr}\right] = \mathbb{E}\left[\sum_{i=1}^{X_{nr-1}} D_{ir}\right] = \mathbb{E}\left[X_{nr-1}\right]\mathbb{E}\left[D_{ir}\right],$$

which gives us

$$\mathbb{E}\left[X_{nr}\right] = \nu_r^n\, \mathbb{E}\, X_{0r} = (1-q)^n \nu^n \times 1. \tag{2}$$

G Extinction in probability and almost sure extinction. Similar to (2), we have that if there are X_{0r} women of type r in the zeroth generation, then the average number of type r descendants at generation n is

$$\mu_{nr} = \mathbb{E}\left[X_{nr}\right] = X_{0r}\nu_r^n.$$

Hence, if $\nu_r < 1$ we have that $\mathbb{E}\left[X_{nr}\right] \to 0$ regardless of X_{0r} (which is assumed to be finite). Since X_{nr} is a non-negative random variable, i.e., $X_{nr} \geq 0$, this implies that $\lim_{n\to\infty} \mathbb{P}\left[X_{nr} = 0\right] =$

1. Indeed, by definition

$$\lim_{n\to\infty} \mathbb{E}\left[X_{nr}\right] = \lim_{n\to\infty} \sum_{k=0}^{\infty} k \times \mathbb{P}\left[X_{nr} = k\right] = 0 \Rightarrow \mathbb{P}\left[X_{nr} = k\right] = 0, \text{ for } k \neq 0 \Rightarrow \mathbb{P}\left[X_{nr} = 0\right] = 1.$$

Almost sure convergence is a little more complicated. But in the case of this MC, we can use the fact that the only recurrent state is zero to deduce this part. Indeed, every state except zero is transient and since the MC converges to zero in expectation, it cannot be transient to infinity. This implies that it converges almost surely to a recurrent class, which in this case is the state zero.

H Probability of extinction in m generations. As explained in the exercise, $P_{e1}(1) = p_0$. To obtain a recursive expression for $P_{em}(1)$, we condition on the number of daughters in the first generation, i.e., X_{1r}:

$$P_{em}(1) = \sum_{j=1}^{\infty} \mathbb{P}\left[\text{extinction in } m \text{ generations} \mid X_{1r} = j\right] \mathbb{P}\left[X_{1r} = j\right]$$

The second probability is simply p_j, the probability of the mother having j daughters. The first probability can be written recursively as the probability that each daughter goes extinct in $m-1$ generations. Indeed, this would lead to their mother's type going extinct in m generations. Since each daughter is independent of the others, we can write $\mathbb{P}\left[\text{extinction in } m \text{ generations} \mid X_{1r} = j\right] = \left[P_{e(m-1)}(1)\right]^j$ to obtain

$$P_{em}(1) = \sum_{j=1}^{\infty} \left[P_{e(m-1)}(1)\right]^j p_j.$$

For the next part, recall that each of the X_{0r} descendant lines are independent since they go down different branching trees. Hence, their extinctions are also independent events. Since type r going extinct in m steps is the intersection of each of the event that X_{0r} lineage goes extinct in m steps, we can write $P_{em}(x)$ as a product of the $P_{em}(1)$, namely $P_{em}(x) = \left[P_{em}(1)\right]^x$.

I Probability of eventual extinction. Again, we will use the law of total probability to find a recursion by conditioning on the number of daughters in the first generation. Explicitly,

$$P_e(1) = \sum_{j=1}^{\infty} \mathbb{P}\left[\text{extinction} \mid X_{1r} = j\right] \mathbb{P}\left[X_{1r} = j\right].$$

The thing to notice here is that the probability of the type of each direct descendant ever going extinct is the same as the probability of their mother's type ever going extinct, which gives us the recursion:

$$P_e(1) = \sum_{j=1}^{\infty} \left[P_e(1)\right]^j p_j.$$

Once again, we can use the fact that the extinction of each of the X_{0r} women are independent event, so that the probability of their intersection is the product of their individual probabilities. Therefore, it holds that $P_e(x) = \left[P_e(1)\right]^x$.

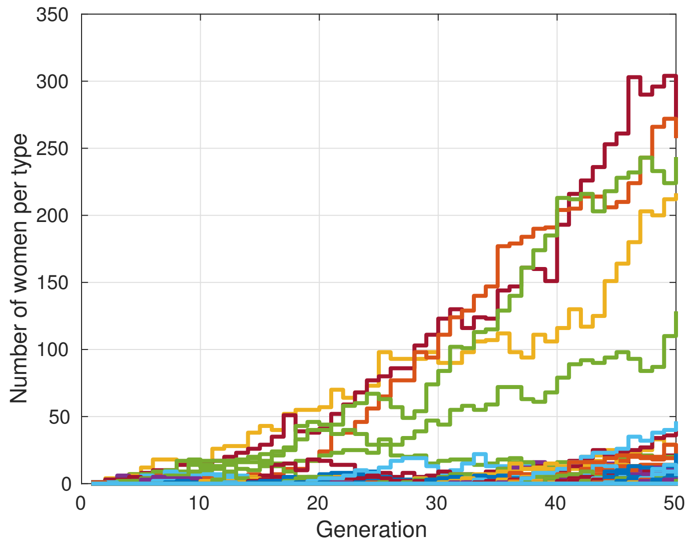

Figure 1: Number of women per types over 50 generations for $X_0 = 100$ women of different types and mutation rate $q = 10^{-2}$ (part D).

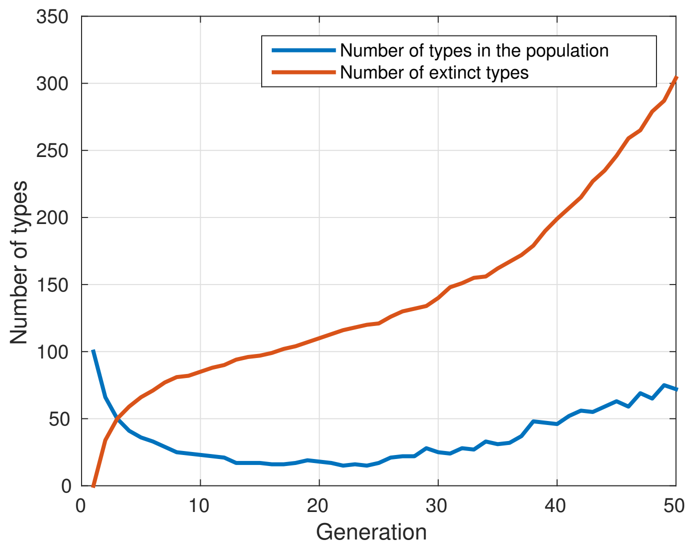

Figure 2: Number of existing types and extinct types over 50 generations for $X_0 = 100$ women of different types and mutation rate $q = 10^{-2}$ (part D).

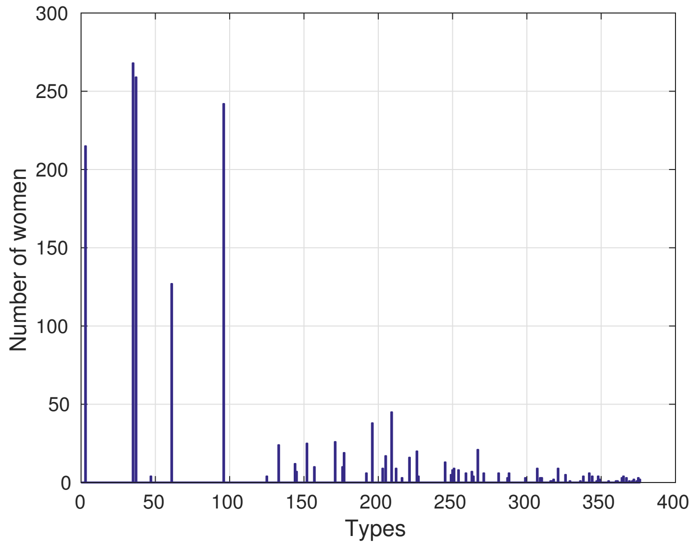

Figure 3: Final number of women per type after 50 generations for $X_0 = 100$ women of different types and mutation rate $q = 10^{-2}$ (part D).

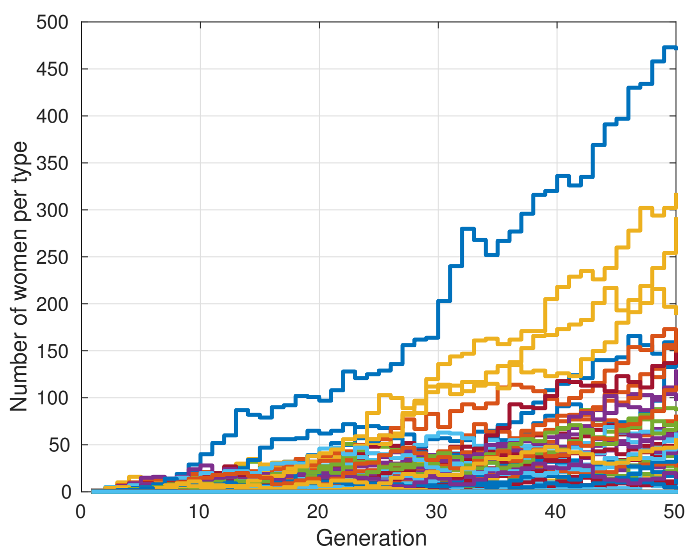

Figure 4: Number of women per types over 50 generations for $X_0 = 400$ women of different types and mutation rate $q = 0$ (part E).

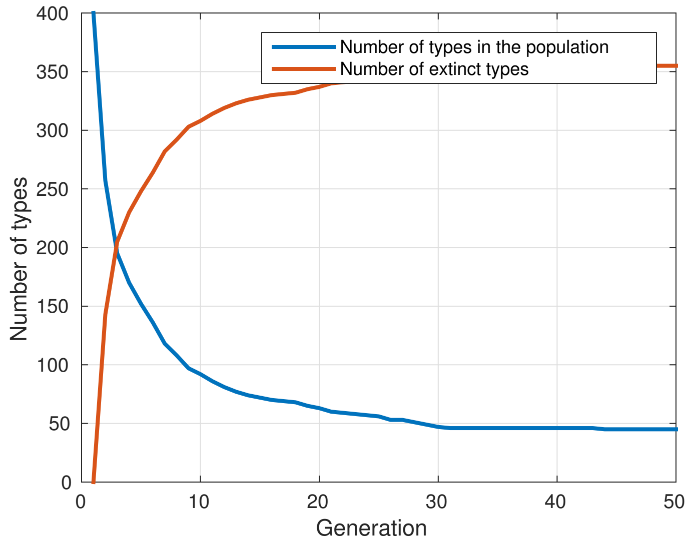

Figure 5: Number of existing types and extinct types over 50 generations for $X_0 = 400$ women of different types and mutation rate $q = 0$ (part E).

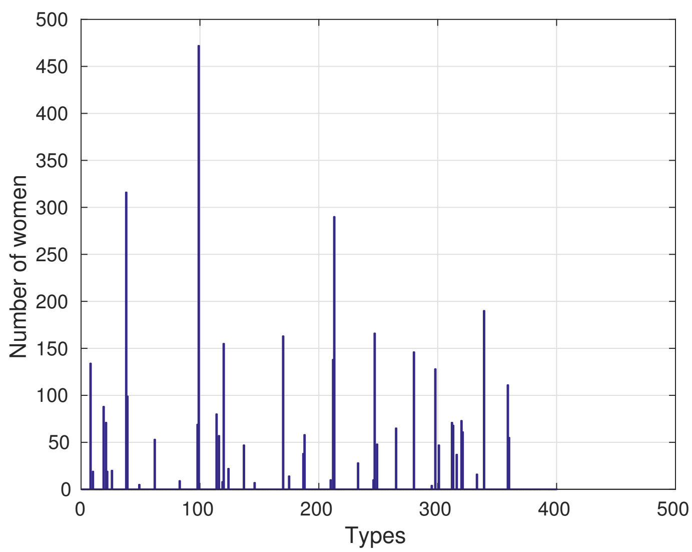

Figure 6: Final number of women per type after 50 generations for $X_0 = 400$ women of different types and mutation rate $q = 0$ (part E).