Module 2

Absorbing Markov chain

Topics: Absorbing Markov chain, Gambler's ruin problem, Branching chain

Absorbing Markov chain

Definition: We begin our analysis of the long-run behavior of Markov chains with a special kind of Markov chain known as the absorbing Markov chain. In the population model of previous module, we saw that $p_{0,0} = 1$ signifying that state 0 is unescapable. Such states are called absorbing states. A Markov chain with at least one absorbing state and a plausible way of reaching some absorbing state from every non-absorbing state is called an absorbing Markov chain. Formally, a Markov chain $\{X_n : n = 0,1,2, \dots\}$ is absorbing if

(i) $A \coloneqq \{j \in \Omega : p_{j,j} = 1\} \neq \emptyset$ and

(ii) For all $i \in A^c$, $\exists n < \infty$ s.t. $p_{i,j}^{(n)} > 0$ for some $j \in A$

In the above definition, A denotes the set of absorbing states. Let us examine if the Markov chain in the population model is absorbing or not. Here, $A = \{0\} \neq \emptyset$, and for every $i \in A^c$, $p_{i,0} = q^i > 0$. You may wonder that the probability is zero for $i = \infty$, but $\infty \notin A^c$. For a specific $i \in A^c$, no matter how large, $q^i > 0$. Thus, we have an absorbing Markov chain. Another absorbing Markov chain with finite states is shown to the right. This one is an example of the gambler's ruin problem, which will analyze in the next section. In this case, $A = \{0,4\}$, $p_{1,0} = 0.4$, $p_{2,4}^{(2)} = 0.36$, $p_{3,4} = 0.6$, satisfying the definition.

p_{ij}	0	2	4	6	8	$\cdots$
0	1	0	0	0	0	$\cdots$
2	q^2	$2pq$	p^2	0	0	$\cdots$
4	q^4	$\binom{4}{1}pq^3$	$\binom{4}{2}p^2q^2$	$\binom{4}{3}p^3q$	p^4	$\cdots$
6	q^6	$\binom{6}{1}pq^5$	$\binom{6}{2}p^2q^4$	$\binom{6}{3}p^3q^3$	$\binom{6}{4}p^4q^2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

p_{ij}	0	1	2	3	4
0	1	0	0	0	0
1	0.4	0	0.6	0	0
2	0	0.4	0	0.6	0
3	0	0	0.4	0	0.6
4	0	0	0	0	1

Absorption in the long-run: When we think of the long-run behavior of an absorbing Markov chain, intuition suggests that the chain would be absorbed in one of the absorbing states, i.e., the chain would reach some absorbing state and never escape from it. This intuition is mostly correct, but with some exceptions.

Result: *Absorption is guaranteed in a finite-state absorbing Markov chain.*

In the gambler's ruin problem, consider $X_0 = 1$. The earliest time till absorption from state-1 is 1, and the probability of absorption at $n = 1$ is: 0.4 via path $1 \to 0$. If absorption does not take place at $n = 1$, which has probability of 0.6, then $X_1 = 2$ and the Markov chain restarts from state-2. The earliest time till absorption from state-2 is 2, and the absorption probability at $n = 2$ is: $0.4 \times 0.4 + 0.6 \times 0.6 = 0.52$ via paths $2 \to 1 \to 0$ and $2 \to 3 \to 4$. If absorption does not take place at $n = 2$, which has probability of 0.48, then $X_2 = 2$ and the Markov chain restarts from state-2 once more. This process repeats indefinitely. Then probability that absorption never takes place is: $0.6 \times 0.48 \times 0.48 \times 0.48 \times \cdots = 0$. The same happens even when $X_0 = 2$ or 3. Thus, absorption is guaranteed here.

Let us formalize the above argument. For an arbitrary $i \in A^c$, let $n_i = \min\left\{n: p_{i,j}^{(n)} > 0 \text{ for some } j \in A\right\}$ denote the earliest time till absorption starting from state-i, and $p_i = \sum_{j \in A} p_{i,j}^{(n_i)}$ denote the probability of absorption at the earliest possible time from state-i. Note that $n_i < \infty$ and $p_i > 0$ for all $i \in A^c$. Let $X_0 = i_0$. If $i_0 \in A$, then absorption has already taken place. For $i_0 \in A^c$, at $n = n_{i_0}$, the Markov chain is either absorbed with probability p_{i_0} or not absorbed with probability $q_{i_0} = 1 - p_{i_0} < 1$. In the latter case, $X_{n_{i_0}} \in A^c$. Let $X_{n_{i_0}} = i_1$. Due to the Markov property, the chain 'restarts' from state $i_1 \in A^c$. At $n = n_{i_1}$, the chain is either absorbed with probability p_{i_1} or not absorbed with probability $q_{i_1} = 1 - p_{i_1} < 1$. In the later case, $X_{n_{i_0}} \in A^c$, and the Markov chain restarts from some state $i_2 \in A^c$ once more. The same sequence of events repeats again and again. Then the probability that absorption never takes place is: $q_{i_0} \times q_{i_1} \times q_{i_2} \times \cdots$. Note that $i_0, i_1, i_2, \ldots$ need not be unique.

Let $q = \max\{q_i : i \in A^c\}$. For a finite-state absorbing Markov chain, A^c is finite, and thus, the maximum exists; moreover, $q < 1$ as $q_i < 1\ \forall i \in A^c$. Then the probability of no absorption $q_{i_0} \times q_{i_1} \times q_{i_2} \times \cdots \le q \times q \times q \times \cdots = 0$ as $q < 1$. Hence, absorption is guaranteed. For the infinite-state case, A^c can be infinite and then q may be undefined and the supremum may be 1. Then we won't be able to bound $q_{i_0} \times q_{i_1} \times q_{i_2} \times \cdots$ by 0 from above. An example of this is shown below. It's a variation of the gambler's ruin problem.

p_{ij}	0	1	2	3	4	$\cdots$
0	1	0	0	0	0	$\cdots$
1	α	0	β	0	0	$\cdots$
2	0	α	0	β	0	$\cdots$
3	0	0	α	0	β	$\cdots$
4	0	0	0	α	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Note that $\alpha \in (0,1)$ and $\beta = 1 - \alpha$. Here, $A = \{0\}$ and $A^c = \{1,2,3,\ldots\}$. $n_i = i$ and $p_i = \alpha^i\ \ \forall i \in A^c$. Then $\max\{q_i : i \in A^c\} = \max\left\{1 - \alpha^i : i = 1,2,3,\ldots\right\}$ does not exist and supremum is 1. So, we cannot tell for sure if absorption is guarantee or not. Later, we will show that absorption is guaranteed when $\alpha \ge 0.5$, and it is not when $\alpha < 0.5$.

Absorption probabilities: Probability of absorption into different absorbing states is of main objectives in the analysis of absorbing Markov chains. Let us define $T = \min\{n: X_n \in A\}$, the time till absorption. It is a random variable. For example, in the gambler's ruin problem (the finite-state case) with $X_0 = 2$, T takes values $2,4,6,\ldots$. Using T, we define probability of absorption into different absorbing states as: $u_{ij} = P(X_T = j | X_0 = i)$ for $j \in A$ where i is the starting state. Note that for $i \in A$, $T = 0$ as the chain is already absorbed, and $u_{ii} = 1$. For $i \in A^c$, we can obtain u_{ij} for $j \in A$ by first-step analysis, which is presented below.

$$
\begin{aligned}
u_{ij} = P(X_T = j | X_0 = i) &= \sum_{k \in \Omega} P(X_T = j, X_1 = k | X_0 = i) \\
&= \sum_{k \in \Omega} P(X_T = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) \\
&= \sum_{k \in \Omega} p_{ik} P(X_T = j | X_0 = k) \text{, by Markov property} \\
&= \sum_{k \in \mathrm{A}} p_{ik} u_{kj} + \sum_{k \in \mathrm{A}^c} p_{ik} u_{kj} = p_{ij} + \sum_{k \in \mathrm{A}^c} p_{ik} u_{kj}
\end{aligned}
$$

The above expression tells that the probability of absorption from $i \in A^c$ into $j \in A$ is the probability of going from i to j directly (i.e., in one-step) plus the probability of going from i to j indirectly (via different states in A^c). It shall be noted that the above expression is valid irrespective of absorption is guaranteed or not. If the absorption is guaranteed, then we have another expression (by the 3rd axiom): $\sum_{j \in A} u_{ij} = 1$ for all i.

Let us calculate $u_{2,j}$ for $j \in \{0,4\}$ for the gambler's ruin problem (the finite-state case).

$$u_{2,0} = p_{2,0} + \sum_{k \in \{1,2,3\}} p_{2,k} u_{k,0} = p_{2,1} u_{1,0} + p_{2,3} u_{3,0} = 0.4 u_{1,0} + 0.6 u_{3,0} \quad \cdots (1)$$

$$u_{1,0} = p_{1,0} + \sum_{k \in \{1,2,3\}} p_{1,k} u_{k,0} = 0.4 + p_{1,2} u_{2,0} = 0.4 + 0.6 u_{2,0} \quad \cdots (2)$$

$$u_{3,0} = p_{3,0} + \sum_{k \in \{1,2,3\}} p_{3,k} u_{k,0} = p_{3,2} u_{2,0} = 0.4 u_{2,0} \quad \cdots (3)$$

Using (2) and (3) in (1), $u_{2,0} = 0.4 u_{1,0} + 0.6 u_{3,0} = 0.4\big(0.4 + 0.6 u_{2,0}\big) + 0.6\big(0.4 u_{2,0}\big) \Rightarrow u_{2,0} = 4/13$. Since absorption is guaranteed in this case, $u_{2,4} = 1 - u_{2,0} = 9/13$, which can also be obtained from the first-step analysis. For the infinite-state cases, we can set up the equations (infinitely many of those), but solution won't be easy.

Expected Time till absorption: Apart from absorption probabilities, time till absorption T is also of interest. In the context of the population model, it is the time till termination of the infection or extinction of the population. We can use the first-step analysis to obtain $v_i = E[T | X_0 = i]$, the expected time till absorption after starting in state-i. Note that $v_i = 0$ for $i \in A$, as $T = 0$ in such cases. For $i \in A^c$, v_i is obtained as follows.

$$
\begin{aligned}
v_i = E[T | X_0 = i] = E_{X_1}\big[E[T | X_1 = k, X_0 = i]\big] &= \sum_{k \in \Omega} E[T | X_1 = k, X_0 = i] P(X_1 = k | X_0 = i) \\
&= \sum_{k \in \Omega} p_{ik} E[1 + T | X_0 = k] \text{, as } T | X_1 = k, X_0 = i \equiv 1 + T | X_0 = k \\
&= \sum_{k \in \Omega} p_{ik} (1 + E[T | X_0 = k]) = \sum_{k \in \Omega} p_{ik} + \sum_{k \in \Omega} p_{ik} v_k \\
&= 1 + \sum_{k \in \mathrm{A}} p_{ik} v_k + \sum_{k \in \mathrm{A}^c} p_{ik} v_k = 1 + \sum_{k \in \mathrm{A}^c} p_{ik} v_k
\end{aligned}
$$

The above expression tells that the expected time till absorption from $i \in A^c$ is 1 unit of time (to leave state-i) plus the expected time till absorption from the next state (which is in A^c). It shall be noted that the calculation of v_i is meaningful only when absorption is guaranteed, i.e., $\sum_{j\in A} u_{ij} = 1$; otherwise, $T|X_0 = i$ can be ∞ with positive probability, i.e., $1 - \sum_{j\in A} u_{ij}$, and then $v_i = \infty$ (or does not exist). For the gambler's ruin problem,

$$v_2 = 1 + \sum_{k\in\{1,2,3\}} p_{2,k}v_k = 1 + p_{2,1}v_1 + p_{2,3}v_3 = 1 + 0.4v_1 + 0.6v_3 \quad \cdots (1)$$

$$v_1 = 1 + \sum_{k\in\{1,2,3\}} p_{1,k}v_k = 1 + p_{1,2}v_2 = 1 + 0.6v_2 \quad \cdots (2)$$

$$v_3 = 1 + \sum_{k\in\{1,2,3\}} p_{3,k}v_k = 1 + p_{3,2}v_2 = 1 + 0.4v_2 \quad \cdots (3)$$

Using (2) and (3) in (1), $v_2 = 1 + 0.4(1 + 0.6v_2) + 0.6(1 + 0.4v_2) \Rightarrow v_2 = 50/13$.

<u>First-step analysis</u>: First-step analysis can be used to determine other quantities of interest. Let us obtain how many times an absorbing Markov chain visits a particular non-absorbing state before absorption. Let $v_{ik} = E[\sum_{n=0}^{T-1} \mathbb{1}(X_n = k|X_0 = i)]$ for $k \in A^c$ denote the expected number of visits to non-absorbing state k after starting in state-i (until absorption happens). For $i \in A$, $v_{ik} = 0$ for all $k \in A^c$. For $i \in A^c$, v_{ik} is obtained as:

$$\begin{aligned}
v_{ik} = E\left[\sum_{n=0}^{T-1} \mathbb{1}(X_n = k|X_0 = i)\right] &= E_{X_1}\left[E\left[\sum_{n=0}^{T-1} \mathbb{1}(X_n = k|X_1 = j, X_0 = i)\right]\right] \\
&= \sum_{j\in\Omega} E\left[\sum_{n=0}^{T-1} \mathbb{1}(X_n = k|X_1 = j, X_0 = i)\right] P(X_1 = j|X_0 = i) \\
&= \sum_{j\in\Omega} p_{ij} E\left[\mathbb{1}(X_0 = k|X_1 = j, X_0 = i) + \sum_{n=1}^{T-1} \mathbb{1}(X_n = k|X_1 = j, X_0 = i)\right] \\
&= \sum_{j\in\Omega} p_{ij} E\left[\mathbb{1}(k = i) + \sum_{n=0}^{T-1} \mathbb{1}(X_n = k|X_0 = j)\right], \text{by Markov property} \\
&= \sum_{j\in\Omega} p_{ij}\delta_{ik} + \sum_{j\in\Omega} p_{ij}v_{jk} = \delta_{ik} + \sum_{j\in A} p_{ij}v_{jk} + \sum_{j\in A^c} p_{ij}v_{jk} \\
&= \delta_{ik} + \sum_{j\in A^c} p_{ij}v_{jk}, \text{where } \delta_{ik} = \mathbb{1}(k = i)
\end{aligned}$$

We included $n = 0$ in the definition of v_{ik} because it ensures equality of total number of visits to all non-absorbing states and the time till absorption. Note that $\sum_{k\in A^c}\sum_{n=0}^{T-1} \mathbb{1}(X_n = k|X_0 = i) = \sum_{n=0}^{T-1}\sum_{k\in A^c} \mathbb{1}(X_n = k|X_0 = i) = \sum_{n=0}^{T-1} 1|X_0 = i = T|X_0 = i$.

Gambler's ruin problem

<u>The setup</u>: Consider a player is playing a game of repeated gambling against an opponent. In each round of the game, the player wins with probability $p \in (0,1)$ and loses with probability $q = 1 - p \in (0,1)$. A win increases wealth of the player by 1 unit and a loss decreases it by 1 unit. The player begins with $a > 0$ units of wealth and the opponent with $A - a > 0$ units. The game can end either with the player winning all or losing all. This kind of setting appears most often in casinos, and the questions of interest here are: (i) what is the probability of win or ruin of the player and (ii) what is the expected duration of the game? We can answer these questions using first-step analysis of absorbing Markov chain.

Let Z_n for $n = 1,2,3, \dots$ denote the wealth earned by the player in the n-th round. Clearly, $P(Z_n = 1) = p$ and $P(Z_n = -1) = q$ for all n. It's also reasonable to consider $Z_1, Z_2, Z_3, \dots$ independent. Let X_n for $n = 1,2,3, \dots$ denote the wealth of the player after n-th round, and $X_0 = a$ denote the initial wealth. Then $X_{n+1} = X_n + Z_{n+1}$ for $n \geq 0$. If X_n is known, then randomness of X_{n+1} is due to Z_{n+1}, and due to independence of $Z_1, Z_2, Z_3, \dots$, $X_0, X_1, \dots, X_{n-1}$ cannot influence X_{n+1} even indirectly. Hence, Markov property holds and $\{X_n : n = 0,1,2, \dots\}$ is a Markov chain. Its transition probabilities are shown below.

p_{ij}	0	1	2	3	$\cdots$	$A-1$	A
0	1	0	0	0	$\cdots$	0	0
1	q	0	p	0	$\cdots$	0	0
2	0	q	0	p	$\cdots$	0	0
3	0	0	q	0	$\cdots$	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$A-1$	0	0	0	0	$\cdots$	0	p
A	0	0	0	0	$\cdots$	0	1

At $X_n = 0$ and $X_n = A$, the game ends, but the Markov chain continues with 0 and A being absorbing states. Other states are non-absorbing and $p_{i,0}^{(i)} = q^i > 0$ for $i = 1,2, \dots, A-1$. So, we have an absorbing Markov chain. Absorption is guaranteed due to finite number of states.

We calculate probability of win $u_{a,A} = P(X_T = A | X_0 = a)$, where $T = \min\{n: X_n \in \{0, A\}\}$ denotes the length of the game. Then the probability of ruin $u_{a,0} = 1 - u_{a,A}$. We also calculate expected length of the game $v_a = E[T | X_0 = a]$

Probability of win: Here, we calculate $u_{a,A}$ for all $a \in \Omega$. First, note that $u_{0,A} = 0$ and $u_{A,A} = 1$. For $a \in \{1,2, \dots, A-1\}$, by the first-step analysis,

$$u_{a,A} = p_{a,A} + \sum_{k=1}^{A-1} p_{a,k} u_{k,A} = p u_{a+1,A} + q u_{a-1,A}$$

Verify the above expression, particularly for $a = 1$ and $a = A - 1$.

$$\begin{aligned}
&u_{a,A} = p u_{a+1,A} + q u_{a-1,A} \Rightarrow u_{a+1,A} - u_{a,A} = (q/p)\big(u_{a,A} - u_{a-1,A}\big) \text{ for } a = 1,2, \dots, A-1 \\
&\Rightarrow u_{a+1,A} - u_{a,A} = (q/p)^2\big(u_{a-1,A} - u_{a-2,A}\big), \text{by the above equation for } a-1 \\
&\qquad\qquad\qquad\quad = (q/p)^3\big(u_{a-2,A} - u_{a-3,A}\big), \text{by the above equation for } a-2 \\
&\qquad\qquad\qquad\quad \vdots \\
&\qquad\qquad\qquad\quad = (q/p)^a\big(u_{1,A} - u_{0,A}\big), \text{by the above equation for } 1 \\
&\qquad\qquad\qquad\quad = (q/p)^a u_{1,A}, \text{as } u_{0,A} = 0
\end{aligned}$$

Note that $u_{a+1,A} - u_{a,A} = (q/p)^a u_{1,A}$ is valid for $a = 1,2, \dots, A-1$. Let us add the first $a - 1$ of these equations (corresponding to $a-1, a-2, \dots, 1$). We get

$$\begin{aligned}
&\big(u_{a,A} - u_{a-1,A}\big) + \big(u_{a-1,A} - u_{a-2,A}\big) + \big(u_{a-2,A} - u_{a-3,A}\big) + \cdots + \big(u_{3,A} - u_{2,A}\big) + \big(u_{2,A} - u_{1,A}\big) \\
&\qquad = \{(q/p)^{a-1} + (q/p)^{a-2} + (q/p)^{a-3} + \cdots + (q/p)^2 + (q/p)^1\} u_{1,A} \\
&\Rightarrow u_{a,A} = u_{1,A} \sum_{k=0}^{a-1} (q/p)^k = \begin{cases} u_{1,A} \dfrac{1 - (q/p)^a}{1 - q/p} & \text{if } p \neq q \\ u_{1,A} a & \text{if } p = q = 1/2 \end{cases}
\end{aligned}$$

Because of the way the above expression is obtained, it is valid for $a = 2,3, \dots, A$. Check for the case $a = A$. In addition, it trivially holds for $a = 0,1$. Since $u_{A,A} = 1$,

$$u_{A,A} = \begin{cases} u_{1,A} \dfrac{1 - (q/p)^A}{1 - q/p} & \text{if } p \neq q \\ u_{1,A} A & \text{if } p = q = 1/2 \end{cases} = 1 \Rightarrow u_{1,A} = \begin{cases} \dfrac{1 - q/p}{1 - (q/p)^A} & \text{if } p \neq q \\ 1/A & \text{if } p = q = 1/2 \end{cases}$$

$$\Rightarrow u_{a,A} = \begin{cases} \dfrac{1 - (q/p)^a}{1 - (q/p)^A} & \text{if } p \neq q \\ a/A & \text{if } p = q = 1/2 \end{cases} \text{ for } a = 0,1,2, \dots, A$$

Expected length of the game: Here we calculate v_a for $a = 0,1,2, \dots, A$. First, note that $v_0 = v_A = 0$. For $a = 1,2, \dots, A-1$, by the first-step analysis,

$$v_a = 1 + \sum_{k=1}^{A-1} p_{a,k} v_k = 1 + p v_{a+1} + q v_{a-1}$$

Verify the above expression, particularly for $a = 1$ and $a = A - 1$.

$$v_a = 1 + p v_{a+1} + q v_{a-1} \Rightarrow v_{a+1} - v_a = (q/p)(v_a - v_{a-1}) - 1/p \;\text{ for } a = 1,2, \dots, A-1$$

$$\Rightarrow v_{a+1} - v_a = (q/p)^2 (v_{a-1} - v_{a-2}) - (1/p)(q/p + 1), \text{by the above equation for } a-1$$

$$= (q/p)^3 (v_{a-2} - v_{a-3}) - \frac{1}{p} \sum_{k=0}^{2} (q/p)^k, \text{by the above equation for } a-2$$

$$\vdots$$

$$= (q/p)^a (v_1 - v_0) - \frac{1}{p} \sum_{k=0}^{a-1} (q/p)^k, \text{by the above equation for } 1$$

$$= (q/p)^a v_1 - \frac{1}{p} \sum_{k=0}^{a-1} (q/p)^k, \text{as } v_0 = 0$$

$$\Rightarrow v_{a+1} - v_a = \begin{cases} \left(\frac{q}{p}\right)^a v_1 - \frac{1}{p} \cdot \frac{1 - (q/p)^a}{1 - q/p} = \left(\frac{q}{p}\right)^a v_1 - \frac{1 - (q/p)^a}{p - q} & \text{if } p \neq q \\ v_1 - 2a \;\; \text{if } p = q = 1/2 \end{cases}$$

Note that the above expression is valid for $a = 1,2, \dots, A-1$. Let us add the first $a-1$ of these equations (corresponding to $a-1, a-2, \dots, 1$). We get

$$(v_a - v_{a-1}) + (v_{a-1} - v_{a-2}) + (v_{a-2} - v_{a-3}) + \cdots + (v_3 - v_2) + (v_2 - v_1) = v_a - v_1$$

$$= \begin{cases} v_1 \sum_{k=1}^{a-1} \left(\frac{q}{p}\right)^k - \frac{1}{p-q}\left[(a-1) - \sum_{k=1}^{a-1}\left(\frac{q}{p}\right)^k\right] & \text{if } p \neq q \\ v_1(a-1) - 2\sum_{k=1}^{a-1} k \;\; \text{if } p = q = 1/2 \end{cases}$$

$$\Rightarrow v_a = \begin{cases} v_1 \sum_{k=0}^{a-1} \left(\frac{q}{p}\right)^k - \frac{1}{p-q}\left[a - \sum_{k=0}^{a-1}\left(\frac{q}{p}\right)^k\right] = \left(v_1 + \frac{1}{p-q}\right)\sum_{k=0}^{a-1}\left(\frac{q}{p}\right)^k - \frac{a}{p-q} \\ = \left(v_1 + \frac{1}{p-q}\right)\frac{1 - (q/p)^a}{1 - q/p} - \frac{a}{p-q} \;\; \text{if } p \neq q \\ v_1 a - a(a-1) \;\; \text{if } p = q = 1/2 \end{cases}$$

Because of the way the above expression is obtained, it's valid for $a = 2,3, \dots, A$. Check for the case $a = A$. In addition, it trivially holds for $a = 0,1$. Since $v_A = 0$,

$$v_A = \begin{cases} \left(v_1 + \frac{1}{p-q}\right)\frac{1 - (q/p)^A}{1 - q/p} - \frac{A}{p-q} = 0 \Rightarrow v_1 = \frac{A}{p-q}\frac{1 - q/p}{1 - (q/p)^A} - \frac{1}{p-q} \;\; \text{if } p \neq q \\ v_1 A - A(A-1) = 0 \Rightarrow v_1 = A - 1 \;\text{if } p = q = 1/2 \end{cases}$$

$$\Rightarrow v_a = \begin{cases} \left(\frac{A}{p-q}\frac{1-q/p}{1-(q/p)^A}\right)\frac{1-(q/p)^a}{1-q/p} - \frac{a}{p-q} = \frac{1}{p-q}\left[A\left(\frac{1-(q/p)^a}{1-(q/p)^A}\right) - a\right] \text{ if } p \neq q \\ a(A-a) \text{ if } p = q = 1/2 \end{cases} \quad \text{for } a = 0,1,2,\dots,A$$

<u>Game against infinitely wealthy adversary</u>: Let us consider a limiting case of the gambler's ruin problem. Let the opponent be infinitely wealthy, i.e., $A - a \to \infty \equiv A \to \infty$. Actually, the opponent need not have infinite wealth; it should be able to arrange additional wealth whenever needed. Now, win is no more a possibility for the player and the game either ends with the player's ruin or it continues forever. The Markov chain now have infinite states and 0 is the only absorbing state. Absorption is no more guaranteed, as there are infinite states. Now, $u_{a,A}$ represents probability of unending game.

$$\lim_{A\to\infty} u_{a,A} = \begin{cases} \lim_{A\to\infty} \frac{1-(q/p)^a}{1-(q/p)^A} = 0 \text{ if } p < q \equiv p < 1/2 \\ \lim_{A\to\infty} a/A = 0 \text{ if } p = q \equiv p = 1/2 \\ \lim_{A\to\infty} \frac{1-(q/p)^a}{1-(q/p)^A} = 1 - \left(\frac{q}{p}\right)^a \text{ if } p > q \equiv p > 1/2 \end{cases} \quad \text{for finite } a$$

$$\lim_{A\to\infty} v_a = \begin{cases} \lim_{A\to\infty} \frac{1}{p-q}\left[A\left(\frac{1-(q/p)^a}{1-(q/p)^A}\right) - a\right] = \frac{a}{q-p} \text{ if } p < q \equiv p < 1/2 \\ \lim_{A\to\infty} a(A-a) = \infty \text{ if } p = q \equiv p = 1/2 \\ \lim_{A\to\infty} \frac{1}{p-q}\left[A\left(\frac{1-(q/p)^a}{1-(q/p)^A}\right) - a\right] = \infty \text{ if } p > q \equiv p > 1/2 \end{cases} \quad \text{for finite } a$$

If $p \leq 1/2$, then $u_{a,A} = 0$, i.e., ruin is guaranteed, and if $p > 1/2$, then the game may never end, which has a probability of $1 - (q/p)^a$. For $p > 1/2$, since the game may not end, v_a is naturally undefined, i.e., $v_a = \infty$. Among the cases when the game surely ends (with player's ruin), $v_a = a/(q-p)$ for $p < 1/2$, but for $p = 1/2$, even though the game is guaranteed to terminate, the expectation does not exist, i.e., $v_a = \infty$.

Branching chain

<u>Modelling as Markov chain</u>: Consider a population growth model, where each member lives for a fixed duration of time and produces random number of offspring at the end of life. Let X_n denote population size in the n-th generation for $n = 0,1,2,\dots$. Note that X_0 is the size of the initial population. Let $Z_k^{(n)}$ denote the offspring produced by the k-th members of n-th population for $k = 1,2,\dots,X_n$ and $n \geq 0$. These are *iid* random variables for all k, n, i.e., all members across populations behave (in terms of offspring production) independently and identically. Furthermore, these are independent of $X_0, X_1, X_2, \dots$, i.e., population size does not have any influence on offspring production. These assumptions may appear rather restrictive, but the insights offered by this simplified model are very useful.

Given the above description, $X_{n+1} = Z_1^{(n)} + Z_2^{(n)} + \cdots + Z_{X_n}^{(n)}$ for all n. If X_n is known, then randomness of X_{n+1} is due to $Z_1^{(n)}, Z_2^{(n)}, Z_3^{(n)}, \ldots$. Since these are independent of other Z's and $X_0, X_1, X_2, \ldots$, then $X_0, X_1, \ldots, X_{n-1}$ cannot influence X_{n+1}. So, Markov property holds and $\{X_n : n = 0,1,2,\ldots\}$ is a Markov chain, also known as the branching chain. Let us consider a generic mass function for $Z_k^{(n)}$ for all k, n: $P\left(Z_k^{(n)} = l\right) = a_l$ for $l = 0,1,2,\ldots$ where $a_0 > 0$, $a_l \geq 0\ \forall l \geq 1$, and $\sum_{l=0}^{\infty} a_l = 1$. Note that $a_0 > 0$ is a necessary condition for the possibility of eventual extinction of the population. If $a_0 = 0$, then the population only grows and eventually explodes. Let us try to construct the transition probability matrix.

p_{ij}	0	1	2	3	4	$\cdots$
0	1		0	0	0	$\cdots$
1	a_0	a_1	a_2	a_3	a_4	$\cdots$
2	a_0^2	$2a_0a_1$	$2a_0a_2 + a_1^2$	$2a_0a_3 + 2a_1a_2$	$2a_0a_4 + 2a_1a_3 + a_2^2$	$\cdots$
3	a_0^3	$3a_0^2a_1$	$3a_0^2a_2 + 3a_0a_1^2$	$3a_0^2a_3 + 6a_0a_1a_2 + a_1^3$	$3a_0^2a_4 + 6a_0a_1a_3 + 3a_0a_2^2 + 3a_1^2a_2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

It's evident that the transition probabilities contain more and more terms as i, j increases and there is no simple patter. If we consider specific mass function for $Z_k^{(n)}$, only in a few simple cases the pattern can be identified. One such case is: $a_0, a_2 > 0$ and rest are zero, i.e., either no offspring is produced, or two offspring are produced. We considered this example in the previous module. However, if we just make $a_1 > 0$, the pattern in the transition probability matrix vanishes. Verify this yourself.

Despite the difficulty in the calculation of transition probabilities, we can observe that the branching chain is an absorbing Markov chain. 0 is the only absorbing state and $p_{i,0} = a_0^i > 0$ for all $i \geq 1$. However, we are not sure about absorption due to infinite states. The main question here is the probability of absorption or extinction $u_{i,0}$ for $i \geq 1$, and the expected time till extinction v_i whenever $u_{i,0} = 1$. Unlike the gambler's ruin problem, we cannot use the first-step analysis as the transition probabilities are not easy to calculate. So, we take a different approach and answer the questions.

Random sum of random variables: In the review of probability, we obtained: $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$ and $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j)$. If $X_1, X_2, \ldots, X_n$ are independent, then $Cov(X_i, X_j) = 0\ \forall i \neq j$ and $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$. In addition to independence, if $X_1, X_2, \ldots, X_n$ are identically distributed with mean μ and variance σ^2, then $E[\sum_{i=1}^{n} X_i] = n\mu$ and $Var(\sum_{i=1}^{n} X_i) = n\sigma^2$. Here, n is a positive integer.

Consider a positive integer-valued random variable N. Let $X_1, X_2, X_3, \ldots$ denote a sequence of *iid* random variables with mean μ and variance σ^2. Moreover, $X_1, X_2, X_3, \ldots$ are independent of N. We are interested in $\sum_{i=1}^{N} X_i$, a random sum of random variables. We encounter this quantity in the branching chain. Let us obtain mean and variance of $\sum_{i=1}^{N} X_i$.

$$E\left[\sum_{i=1}^{N} X_i\right] = E_N\left[E\left[\sum_{i=1}^{N} X_i \,|N = n\right]\right] = \sum_{n=1}^{\infty} E\left[\sum_{i=1}^{N} X_i \,|N = n\right] p_N(n)$$

$$= \sum_{n=1}^{\infty} E\left[\sum_{i=1}^{n} X_i\right] p_N(n)\text{, due to independence of } X_1, X_2, X_3, \ldots \text{ with } N$$

$$= \sum_{n=1}^{\infty} (n\mu) p_N(n) = \mu \sum_{n=1}^{\infty} n p_N(n) = \mu E[N]$$

$$Var\left(\sum_{i=1}^{N} X_i\right) = E_N\left[Var\left(\sum_{i=1}^{N} X_i \,|N = n\right)\right] + Var_N\left(E\left[\sum_{i=1}^{N} X_i \,|N = n\right]\right)$$

First, $E_N\left[Var\left(\sum_{i=1}^{N} X_i \,|N = n\right)\right] = \sum_{n=1}^{\infty} Var\left(\sum_{i=1}^{N} X_i \,|N = n\right) p_N(n)$

$$= \sum_{n=1}^{\infty} Var\left(\sum_{i=1}^{n} X_i\right) p_N(n)\text{, due to independence of } X_1, X_2, \ldots \text{ with } N$$

$$= \sum_{n=1}^{\infty} (n\sigma^2) p_N(n) = \sigma^2 E[N]$$

Next, $Var_N\left(E\left[\sum_{i=1}^{N} X_i \,|N = n\right]\right) = E_N\left[\left(E\left[\sum_{i=1}^{N} X_i \,|N = n\right]\right)^2\right] - \left(E_N\left[E\left[\sum_{i=1}^{N} X_i \,|N = n\right]\right]\right)^2$

$$= \sum_{n=1}^{\infty} \left(E\left[\sum_{i=1}^{N} X_i \,|N = n\right]\right)^2 p_N(n) - \left(\sum_{n=1}^{\infty} E\left[\sum_{i=1}^{N} X_i \,|N = n\right] p_N(n)\right)^2$$

$$= \sum_{n=1}^{\infty} (n\mu)^2 p_N(n) - \left(\sum_{n=1}^{\infty} (nu) p_N(n)\right)^2\text{, due to independence of } X_1, X_2, \ldots \text{ with } N$$

$$= \mu^2 \sum_{n=1}^{\infty} n^2 p_N(n) - \mu^2 \left(\sum_{n=1}^{\infty} n p_N(n)\right)^2 = \mu^2 (E[N^2] - E^2[N]) = \mu^2 Var(N)$$

Therefore, $Var\left(\sum_{i=1}^{N} X_i\right) = \sigma^2 E[N] + \mu^2 Var(N)$

Mean and variance of population size: In the branching chain, $X_n = Z_1^{(n-1)} + Z_2^{(n-1)} + \cdots + Z_{X_{n-1}}^{(n-1)}$ for $n \geq 1$, where $Z_1^{(n-1)}, Z_2^{(n-1)}, Z_3^{(n-1)}, \ldots$ are *iid* random variables denoting the offspring produced by the 1st, 2nd, 3rd, … members of the $n-1$st generation. These are independent of X_{n-1}, size of the $n-1$st generation, as well. Let μ and σ^2 denote mean and variance of offspring produced by any member of the population in any generation. Then by the formulas for the random sum of random variables,

$$E[X_n] = E\left[\sum_{i=1}^{X_{n-1}} Z_i^{(n-1)}\right] = \mu E[X_{n-1}] \text{ for } n \geq 1 \text{ (}X_0\text{ is the initial population size)}$$
$$= \mu^2 E[X_{n-2}] = \mu^3 E[X_{n-3}] = \cdots = \mu^n E[X_0] = X_0 \mu^n$$

$$\begin{aligned} Var(X_n) = Var\left(\sum_{i=1}^{X_{n-1}} Z_i^{(n-1)}\right) &= \sigma^2 E[X_{n-1}] + \mu^2 Var(X_{n-1}) \text{ for } n \geq 1 \\ &= \sigma^2 E[X_{n-1}] + \mu^2\big(\sigma^2 E[X_{n-2}] + \mu^2 Var(X_{n-2})\big) \\ &= \sigma^2 E[X_{n-1}] + \sigma^2\mu^2 E[X_{n-2}] + \mu^4\big(\sigma^2 E[X_{n-3}] + \mu^2 Var(X_{n-3})\big) \\ &= \sigma^2 E[X_{n-1}] + \sigma^2\mu^2 E[X_{n-2}] + \sigma^2\mu^4 E[X_{n-3}] + \mu^6\big(\sigma^2 E[X_{n-4}] + \mu^2 Var(X_{n-4})\big) \\ &\vdots \\ &= \sigma^2(E[X_{n-1}] + \mu^2 E[X_{n-2}] + \mu^4 E[X_{n-3}] + \cdots + \mu^{2n-2} E[X_0]) + \mu^{2n} Var(X_0) \\ &= \sigma^2 X_0(\mu^{n-1} + \mu^n + \mu^{n+1} + \cdots + \mu^{2n-2}) = \begin{cases} X_0\sigma^2\mu^{n-1}\dfrac{1-\mu^n}{1-\mu} \text{ if } \mu \neq 1 \\ X_0\sigma^2 n \text{ if } \mu = 1 \end{cases} \end{aligned}$$

It appears that the mean number of offspring produced by any member μ determines whether $E[X_n]$ and $Var(X_n)$ increases or decreases with n. In particular,

	$\mu < 1$	$\mu = 1$	$\mu > 1$
$\lim_{n\to\infty} E[X_n]$	0	X_0	∞
$\lim_{n\to\infty} Var(X_n)$	0	∞	∞
Implication for $X_\infty = \lim_{n\to\infty} X_n$ and extinction	X_∞ converges to 0, implying guaranteed extinction.	X_∞ is random with finite mean and infinite variance. Extinction may still be guaranteed, but it would take 'long time'.	X_∞ is random with infinite mean and variance. Extinction is no more guaranteed.

A recurrence relation for extinction probability: Now, we make our first attempt to calculate extinction probability. The intuition that we obtained from the above table seems to suggest that X_0 does not play significant role in the long-run behavior of the branching chain. *So, we take $X_0 = 1$ in the remainder of the analysis*. Later, we will examine the effect of $X_0 > 1$. Let $T = \min\{n: X_n = 0 | X_0 = 1\}$ denote the time till extinction. Let $u_n = P(T \leq n | X_0 = 1) = P(X_n = 0 | X_0 = 1)$ denote the probability of extinction by n-th generation for $n \geq 0$. Then the probability of (eventual) extinction, $u = \lim_{n\to\infty} P(X_n = 0 | X_0 = 1) = \lim_{n\to\infty} u_n$. If we can obtain the values of $u_0, u_1, u_2, \ldots$, then we have u.

$u_n = P(T \leq n | X_0 = 1)$ is the distribution function value of T at n. We know that distribution function value of any random variable is 1 as we approach ∞. Then extinction probability $u = \lim_{n\to\infty} u_n = \lim_{n\to\infty} P(T \leq n | X_0 = 1)$ must be 1. However, this argument is flawed as T

may not be a 'proper' random variable. We defined random variable as a real-valued function on the sample space of a random experiment. The definition ensures that distribution function value of a random variable is 1 as we approach ∞. Consider a random experiment where we toss a fair coin, and if head shows, we stop, else we continue tossing indefinitely. The number of tosses is 1 with probability $1/2$ and ∞ with probability $1/2$. Here the distribution function value is $1/2$ as we approach ∞. This is an example of 'improper' random variable. We have seen this in the gambler's ruin problem against infinitely wealthy adversary with $p > 1/2$. A random variable X is 'proper' if $\lim_{x\to\infty} F(x) = 1$, else it's 'improper'. Expected value of an 'improper' random variable is undefined or ∞. Expected value of a 'proper' random variable can be ∞ too, as seen in the gambler's ruin problem against infinitely wealthy adversary with $p = 1/2$. In the branching chain, we don't know if extinction is guaranteed or not, i.e., if T is a 'proper' random variable or not. So, we can't say $\lim_{n\to\infty} u_n = 1$.

Let us apply the first-step analysis to $u_n = P(X_n = 0|X_0 = 1)$. Earlier, we used it to obtain absorption probability through the transition probabilities, which are not available to us for the branching chain. Here, we take an indirect route of obtaining extinction probability u via $u_0, u_1, u_2, \ldots$. First, note that $u_0 = P(X_0 = 0|X_0 = 1) = 0$. For $n \geq 1$,

$$u_n = P(X_n = 0|X_0 = 1) = \sum_{k=0}^{\infty} P(X_n = 0, X_1 = k|X_0 = 1)$$

$$= \sum_{k=0}^{\infty} P(X_n = 0|X_1 = k, X_0 = 1) \cdot P(X_1 = k|X_0 = 1)$$

$$= \sum_{k=0}^{\infty} P\left(Z_1^{(0)} = k\right) \cdot P(X_{n-1} = 0|X_0 = k) = \sum_{k=0}^{\infty} a_k u_{n-1}^k \text{, by Markov property}$$

In the above derivation, we wrote $P(X_{n-1} = 0|X_0 = k) = u_{n-1}^k$ for $k \geq 0$. It is correct for $k = 0$ and 1, as $P(X_{n-1} = 0|X_0 = 0) = 1 = u_{n-1}^0$ and $P(X_{n-1} = 0|X_0 = 1) = u_{n-1}$ by the definition of u_n. For $k \geq 2$, we can consider the branching chain with $X_0 = k$ as the sum of k number of independent and identical branching chains with $X_0 = 1$, each emerging from a member of the 0th generation in the original chain. Let $\left\{X_n^{(i)}: n = 0,1,2, \ldots\right\}$ denote i-th such chain for $i = 1,2, \ldots, k$. Then

$$P(X_{n-1} = 0|X_0 = k) = P\left(X_{n-1}^{(1)} + X_{n-1}^{(2)} + \cdots + X_{n-1}^{(k)} = 0|X_0^{(i)} = 1 \text{ for } i = 1,2, \ldots, k\right)$$

$$= P\left(X_{n-1}^{(1)} = 0|X_0^{(1)} = 1\right) \cdot P\left(X_{n-1}^{(2)} = 0|X_0^{(2)} = 1\right) \cdots P\left(X_{n-1}^{(k)} = 0|X_0^{(k)} = 1\right)$$

$$= P(X_{n-1} = 0|X_0 = k) \cdot P(X_{n-1} = 0|X_0 = k) \cdots P(X_{n-1} = 0|X_0 = k) = u_{n-1}^k$$

due to independence and due to identicalness.

With $u_0 = 0$ and $u_n = \sum_{k=0}^{\infty} a_k u_{n-1}^k$ for $n \geq 1$, we can iteratively calculate $u_0, u_1, u_2, \ldots$ for any branching chain and obtain $u = \lim_{n\to\infty} u_n$. Let us consider a simple example. Let $a_0 = p \in (0,1)$, $a_1 = q = 1 - p$, and $a_2 = a_3 = \cdots = 0$ in a branching chain, i.e., each member produces at most 1 offspring. Then $u_n = \sum_{k=0}^{1} a_k u_{n-1}^k = p + q u_{n-1}$ for $n \geq 1$.

$u_0 = 0$
$u_1 = p + qu_0 = p$
$u_2 = p + qu_1 = p + qp$
$u_3 = p + qu_2 = p + qp + q^2p$
$\vdots$

$$u_n = p + qu_{n-1} = p + qp + q^2p + \cdots + q^{n-1}p = p\frac{1-q^n}{1-q} = 1 - q^n \text{ for } n \geq 1$$

Then $u = \lim_{n\to\infty} u_n = 1 - \lim_{n\to\infty} q^n = 1$. i.e., extinction is guaranteed.

Expected time till extinction: We can obtain mas function of T from $u_0, u_1, u_2, \ldots$ as:

$P(T = n|X_0 = 1) = P(T \leq n|X_0 = 1) - P(T \leq n - 1|X_0 = 1) = u_n - u_{n-1}$ for $n \geq 1$

Then $v = E[T|X_0 = 1] = \sum_{n=1}^{\infty} n(u_n - u_{n-1})$. Note that this calculation makes sense only when $u = 1$. For the above example $u_n = 1 - q^n$ for $n \geq 0$. Then

$$v = \sum_{n=1}^{\infty} n(u_n - u_{n-1}) = \sum_{n=1}^{\infty} n(q^{n-1} - q^n) = p\sum_{n=1}^{\infty} nq^{n-1} = p\sum_{n=1}^{\infty}\frac{dq^n}{dq} = p\frac{d}{dq}\left(\sum_{n=1}^{\infty} q^n\right)$$
$$= p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\frac{1}{(1-q)^2} = 1/p$$

There is an alternate way of obtaining expected time till extinction: $v = \sum_{n=0}^{\infty}(1 - u_n)$. With this formula for the example problem, $v = \sum_{n=0}^{\infty} q^n = 1/p$. This formula is based on the general formula: $E[N] = \sum_{n=0}^{\infty} P(N > n)$ for non-negative integer valued random variable N. It has a continuous analogue: $E[X] = \int_0^{\infty} P(X > x)dx = \int_0^{\infty}\big(1 - F(x)\big)dx$ for non-negative continuous random variable X. Verify these yourself.

Effect of $X_0 > 1$: We have discussed about a way to calculate extinction probability u and the expected time till extinction v when $X_0 = 1$. If $X_0 > 1$, say $X_0 = k \geq 2$, then we can consider the branching chain as the sum of k number of independent and identical branching chains with $X_0 = 1$, each emerging from a member of the 0^{th} generation in the original chain. Let $\left\{X_n^{(i)} : n = 0,1,2,\ldots\right\}$ denote i-th such chain for $i = 1,2,\ldots,X_0$. Then

$$\begin{aligned}
(u|X_0 = k) &= \lim_{n\to\infty} P(X_n = 0|X_0 = k) \\
&= \lim_{n\to\infty} P\left(X_n^{(1)} + X_n^{(2)} + \cdots + X_n^{(k)} = 0|X_0^{(i)} = 1 \text{ for } i = 1,2,\ldots,k\right) \\
&= \lim_{n\to\infty} \prod_{i=1}^{k} P\left(X_n^{(i)} = 0|X_0^{(i)} = 1\right), \text{due to independence} \\
&= \prod_{i=1}^{k} \lim_{n\to\infty} P\left(X_n^{(i)} = 0|X_0^{(i)} = 1\right), \text{exchange is possible as } k \text{ is finite} \\
&= \prod_{i=1}^{k} u = u^k, \text{due to identicalness}
\end{aligned}$$

where u denotes the extinction probability for $X_0 = 1$. If $u = 1$, then $(u|X_0 = k) = 1$ for all $k \geq 2$. Thus, greater initial population cannot alter eventual extinction if that is the case with $X_0 = 1$. If $u < 1$, then $(u|X_0 = k)$ decreases with k, but it is positive for all $k \geq 2$, i.e., the possibility of extinction cannot be eliminated by increasing initial population size.

With the above characterization, time till extinction in the original chain is maximum of the times till extinction for the branching chains emerging from members of the 0th generation in the original chain, denoted by $T^{(1)}, T^{(2)}, \dots, T^{(k)}$. Then

$$(T|X_0 = k) = \left(\max\left\{T^{(1)}, T^{(2)}, \dots, T^{(k)}\right\} \middle| X_0^{(i)} = 1 \text{ for } i = 1,2,\dots,k\right)$$

$T^{(1)}, T^{(2)}, \dots, T^{(k)}$ are *iid* random variables with distribution function values $u_0, u_1, u_2, \dots$ at $0,1,2,\dots$, which we obtained recursively for $X_0 = 1$. Then for $n \geq 0$,

$$\begin{aligned} P(T \leq n|X_0 = k) &= P\left(\max\left\{T^{(1)}, T^{(2)}, \dots, T^{(k)}\right\} \leq n|X_0^{(i)} = 1 \text{ for } i = 1,2,\dots,k\right) \\ &= P\left(T^{(i)} \leq n|X_0^{(i)} = 1 \text{ for } i = 1,2,\dots,k\right), \text{by equivalence of events} \\ &= \prod_{i=1}^{k} P\left(T^{(i)} \leq n|X_0^{(i)} = 1\right), \text{due to independence} \\ &= \prod_{i=1}^{k} u_n = u_n^k, \text{due to identicalness} \end{aligned}$$

Hence, $E[T|X_0 = k] = \sum_{n=1}^{\infty} n\left(u_n^k - u_{n-1}^k\right) = \sum_{n=0}^{\infty}(1 - u_n^k)$. Since $u_n < 1$ for all n (*Why?*), then u_n^k decreases with k, and thus, $E[T|X_0 = k] = \sum_{n=0}^{\infty}(1 - u_n^k)$ increases with k. Hence, larger initial population can delay extinction, even though it cannot stop it.

<u>Going beyond the recurrence relation</u>: $u_n = \sum_{k=0}^{\infty} a_k u_{n-1}^k$ for $n \geq 1$ with $u_0 = 0$ and $u = \lim_{n\to\infty} u_n$ is sufficient to calculate extinction probability for any given branching chain. However, it is not sufficient to extract insights. We need to go beyond the recurrence relation. Taking limits of both sides of $u_n = \sum_{k=0}^{\infty} a_k u_{n-1}^k$ as $n \to \infty$, we get

$$u = \lim_{n\to\infty} u_n = \lim_{n\to\infty} \sum_{k=0}^{\infty} a_k u_{n-1}^k = \sum_{k=0}^{\infty} a_k \lim_{n\to\infty} u_{n-1}^k = \sum_{k=0}^{\infty} a_k \left(\lim_{n\to\infty} u_{n-1}\right)^k = \sum_{k=0}^{\infty} a_k u^k$$

$u = \sum_{k=0}^{\infty} a_k u^k$ is computationally much simpler than the original recursive relation. In its derivation, we changed the order of limit and sum, which is always permitted for finite sum. For infinite sum, which is the case here, the change of order may not be permitted. Consider two cases: $\lim_{n\to\infty} \sum_{k=0}^{\infty} 1/n^k$ and $\lim_{n\to\infty} \sum_{k=0}^{\infty} 1/n$. In the first case,

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{1}{n^k} = \lim_{n\to\infty} \frac{1}{1 - 1/n} = 1, \text{which is same as}$$

$$\sum_{k=0}^{\infty} \lim_{n\to\infty} \frac{1}{n^k} = \lim_{n\to\infty} \frac{1}{n^0} + \sum_{k=1}^{\infty} \lim_{n\to\infty} \frac{1}{n^k} = 1 + \sum_{k=1}^{\infty} 0 = 1$$

So, change of order is permitted in the first case. In the second case,

$$\lim_{n\to\infty}\sum_{k=0}^{\infty}\frac{1}{n}=\lim_{n\to\infty}\lim_{m\to\infty}\sum_{k=0}^{m}\frac{1}{n}=\lim_{n,m\to\infty}\frac{m+1}{n}=1, \text{which is not same as}$$

$$\sum_{k=0}^{\infty}\lim_{n\to\infty}\frac{1}{n}=\sum_{k=0}^{\infty}0=0; \text{thus, chage of order is not permitted here.}$$

There are some results, known as the convergence theorems, that lays out conditions for the change of order of limit and sum to be possible. One such theorem is *bounded convergence theorem*, which will be invoked number of times in this course.

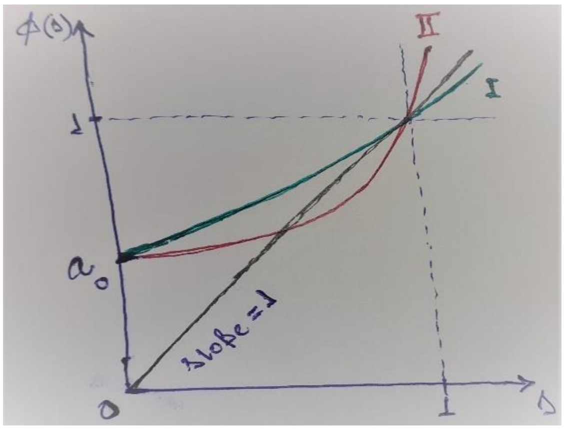
Let $a(\cdot)$ be a real valued function defined on unbounded Ω such that $\sum_{x\in\Omega}|a(x)|$ exists. Let $b_n(\cdot)$ for $n=0,1,2,\dots$ be a sequence of real valued function on Ω such that $|b_n(x)|\le 1$ for all $x\in\Omega$ and $n=0,1,2,\dots$ and $\lim_{n\to\infty}b_n(x)$ exist for all $x\in\Omega$. Let $\lim_{n\to\infty}b_n(x)=b(x)$ for some real valued function $b(\cdot)$ on Ω. In this set-up, $\lim_{n\to\infty}\sum_{x\in\Omega}a(x)b_n(x)$ is same as $\sum_{x\in\Omega}a(x)\lim_{n\to\infty}b_n(x)=\sum_{x\in\Omega}a(x)b(x)$.

In $\lim_{n\to\infty}\sum_{k=0}^{\infty}a_k u_{n-1}^k$, $\Omega=\{0,1,2,\dots\}$, $a(k)=a_k$ and $b_n(k)=u_{n-1}^k$ for $k\in\Omega$. Here, Ω is unbounded, $\sum_{k\in\Omega}|a(k)|=\sum_{k=0}^{\infty}a_k=1$, i.e., it exists, $|b_n(x)|=u_{n-1}^k\le 1\ \forall k,n$ (as u_{n-1} is probability) and $\lim_{n\to\infty}b_n(x)=\lim_{n\to\infty}u_{n-1}^k=u^k$ exists for all k. Bounded convergence theorem applies here, and thus, the order of limit and sum can be changed. It shall be noted that the theorem does not say anything when the underlying conditions are not met.

For a branching chain with $a_2=1-a_0$ and rest zero, $u=\sum_{k=0}^{\infty}a_k u^k\Rightarrow u=a_0+a_2u^2\Rightarrow u=\left(1\pm\sqrt{1-4a_0a_2}\right)/2a_2$. For $a_0=0.6$, $u=\left(1\pm\sqrt{1-0.96}\right)/0.8=1.5,1$; since $u\le 1$, $u=1$ is the solution, i.e., extinction is guaranteed. For $a_0=0.5$, $u=\left(1\pm\sqrt{1-1}\right)/1=1$, i.e., extinction is guaranteed. For $a_0=0.4$, $u=\left(1\pm\sqrt{1-0.96}\right)/1.2=1,2/3$; both the solutions are valid probabilities values. It's not immediately clear which one is extinction probability. Our intuition from the analysis of mean and variance suggests that $u=2/3$ is the extinction probability. It can be confirmed from the recurrence relations. Observe that 1 is a candidate solution of $u=\sum_{k=0}^{\infty}a_k u^k$ in all three cases. This observation is true in general, as: $\sum_{k=0}^{\infty}a_k u^k|_{u=1}=\sum_{k=0}^{\infty}a_k=1$. So, when we solve $u=\sum_{k=0}^{\infty}a_k u^k$ for a branching chain, we always get 1 as a solution. If there is no other solution in $[0,1)$, then $u=1$, i.e., extinction is guaranteed. If we get a solution in $[0,1)$, we are not sure about u. In the following sections, we show that the smallest solution in $[0,1]$ is the extinction probability.

Guaranteed extinction if $\mu\le 1$: Let $\phi(s)=\sum_{k=0}^{\infty}a_k s^k$, where s is a real valued parameter. Then extinction probability u is a solution of $s=\phi(s)$. For a non-negative integer valued random variable N, $E[s^N]=\sum_{n=0}^{\infty}s^n p_N(n)$ for suitable $s\in\mathbb{R}$ is known as the probability generating function of N. Like the moment generating function 'generates' moments, the probability generating function generates probability masses. $p_N(n) = (1/n!)\left.\frac{d^n\phi_N(s)}{ds^n}\right|_{s=0}$ for all $n \geq 1$. $\phi(s)$ is the probability generating function of $Z_k^{(n)}$ for all k, n.

Let us examine $\phi(s) = \sum_{k=0}^{\infty} a_k s^k$. First, note that $\phi(0) = a_0 > 0$ and $\phi(1) = 1$. $\frac{d\phi(s)}{ds} = \sum_{k=1}^{\infty} a_k k s^{k-1} > 0$ for $s \in (0, \infty)$, as $a_0 < 1$ (making $a_k > 0$ for at least one $k \geq 1$). So, $\phi(s)$ is a strictly increasing function in $(0, \infty)$. $\frac{d^2\phi(s)}{ds^2} = \sum_{k=2}^{\infty} a_k k(k-1)s^{k-2} \geq 0$ for $s \in (0, \infty)$. So, $\phi(s)$ is a convex function in $(0, \infty)$. Combining these three observations, we see that $\phi(s)$ increases from a_0 to 1 as s goes from 0 to 1, and the nature of this increase is convex, as shown in the diagram below.



All convex-increasing functions from $(0, a_0)$ to $(1,1)$ can be classified into two groups. The first group (shown in green) consists of those functions that stay above the unit slope line passing through the origin (shown in black) for all $s \in [0,1)$. The second group (shown in red) consists of the rest, i.e., those functions that touch or cross the unit slope line through the origin for at least one $s \in [0,1)$.

Let us find a way to identify the type of a given $\phi(s)$. $\left.\frac{d\phi(s)}{ds}\right|_{s=1} = \sum_{k=1}^{\infty} a_k k \cdot 1^k = \mu$, i.e., slope of $\phi(s)$ at $s = 1$ is same as the mean number of offspring produced by a member. The first group stays above the unit slope line for all $s \in [0,1)$ and touches or crosses the line at $s = 1$, as every $\phi(s)$ passes through $(1,1)$. Clearly, $\mu \leq 1$ for the first group. Then second group, touches or crosses the unit slope line for at least one $s \in [0,1)$ and then again at $s = 1$. Since $\phi(s)$ is convex, $\phi(s)$ touches or crosses the unit slope line exactly once in $[0,1)$. Since it touches/crosses the unit slope line again at $s = 1$ and it's a smooth function, it must cross the unit slope line at both the places. At $s = 1$, where $\phi(s)$ is crossing the unit slope line for the second time, it is coming from below the line and goes above the line after crossing it. Therefore, slope of $\phi(s)$ at $s = 1$ must be greater than unity, i.e., $\mu > 1$.

Solutions of $s = \phi(s)$ are essentially the points where $\phi(s)$ touches or crosses the unit slope line passing through the origin. We are interested in the solutions that are in $[0,1]$. For $\mu \leq 1$, $\phi(s)$ touches/crosses the unit slope line only once in $[0,1]$, at $s = 1$. Since u is a solution of $\phi(s)$, $u = 1$. So, extinction is guaranteed when $\mu \leq 1$.

Extinction probability when $\mu > 1$: If $\mu > 1$, $\phi(s)$ crosses the unit slope line twice in $[0,1]$, first at some point in $[0,1)$ and then at $s = 1$. Here, we have two candidates for u. Let w represents an arbitrary candidate. So, $w = \phi(w)$. Note that $w > 0$, as $\phi(0) = a_0 > 0$. Since

$\phi(s)$ is increasing in $[0,\infty)$, $w = \phi(w) \geq \phi(0)$. From the recurrence relations, $u_0 = 0$ and $u_n = \sum_{k=0}^{\infty} a_k u_{n-1}^k = \phi(u_{n-1})$ for all $n \geq 1$. Thus, $w \geq \phi(0) = \phi(u_0) = u_1$. Now using the same logic, $w = \phi(w) \geq \phi(u_1) = u_2$, and then $w = \phi(w) \geq \phi(u_2) = u_3$, and this process continues. Then $w \geq u_n$ for all $n \geq 0$. Therefore, $w \geq \lim_{n\to\infty} u_n \Rightarrow w \geq u$. So, an arbitrary solution of $s = \phi(s)$ in $[0,1]$ is at least as large as the extinction probability, which too is a solution of $s = \phi(s)$. These two conditions are satisfied only when the extinction probability is the smallest of the candidate solutions. Clearly, $u < 1$ when $\mu > 1$.

If we are interested in the extinction probability of a branching chain, then we need to find μ first. If $\mu \leq 1$, then extinction is guaranteed. If $\mu > 1$, then extinction probability is the only solution of $s = \phi(s)$ in $(0,1)$. If there is any difficulty in solving $s = \phi(s)$ analytically, we can simply employ a binary search as $\phi(s)$ is strictly increasing in $(0,1)$.

Finiteness of time till extinction for $\mu < 1$: When extinction is guaranteed, i.e., when $\mu \leq 1$, we are also interested in the expected time till extinction. Earlier, we calculated it using: $v = \sum_{n=1}^{\infty} n(u_n - u_{n-1}) = \sum_{n=0}^{\infty}(1 - u_n)$. We don't have a simplified formula here, but we can show that $v < \infty$ for $\mu < 1$. With $X_0 = 1$,

$$E[X_n] = \mu^n = \sum_{k=1}^{\infty} kP(X_n = k) \geq \sum_{k=1}^{\infty} P(X_n = k) = 1 - P(X_n = 0) = 1 - u_n \ \forall n$$

$$\Rightarrow v = \sum_{n=0}^{\infty}(1 - u_n) \leq \sum_{n=0}^{\infty} \mu^n = \begin{cases} 1/(1-\mu) < \infty & \text{if } \mu < 1 \\ \infty & \text{if } \mu = 1 \end{cases}$$

For $\mu = 1$, extinction is guaranteed, but time till extinction does not have expected value, i.e., $v = \infty$. However, a formal proof is unknown to me.

Practice problems

Book-1: Introduction to Probability Models by Sheldon Ross [10th edition]
Book-2: An Introduction to Stochastic Modeling by Taylor and Karlin [3rd edition]

Absorbing Markov Chains
Book-1, Chapter-4, Exercise No. 57, 61
Book-2, Chapter-III, Exercise No. 4.4, 4.8 and Problem No. 4.4, 4.12

Branching Chain
Book-1, Chapter-4, Exercise No. 64
Book-2, Chapter-III, Exercise No. 8.3 and Problem No. 9.1, 9.6