Chapter 3

Markov Chain

3.1 Introduction

In the previous chapter independently and identically distributed (iid) random sequences generated from discrete value stochastic process had been discussed. However, this chapter considers discrete-value random sequence $\{X_n|n = 0, 1, 2...\}$ that is not iid. The consideration is on a system in which, X_{n+1} depends on X_n but not on earlier values of the random sequence, $X_0, ..., X_{n-1}$. Keep in mind that each X_n is a discrete random variable with range $S_X = \{0, 1, 2...\}$.

The system that behaves as above can also be explained in a probabilistic sense in a way that, the present state of the system contains all the relevant information needed to predict the future. This kind of behaviour is formulated mathematically by Andrey Markov, that is why this kind of property is named Markovian property.

In the following section, the definition of discrete-time Markov Chain and some examples will be given.

3.2 Discrete-time Markov Chain

Definition (Definition 12.1 Yates & Goodman 2005)
A *discrete time Markov chain* $X_n|n = 0, 1, 2, ...$ is a discrete-time discrete value random sequence such that given $X_0, X_1, ..., X_n$ the next random variable X_{n+1} depends only on X_n which can be described by the following transition probability

$$P[X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0] = P[X_{n+1} = j|X_n = i] = P_{ij}$$

The value of X_n contains all the relevant information that is it summarises all of past history of the system needed to predict the variable X_{n+1} in the random sequence. X_n is called the current or present state of the system at time n, and its sample space is called the *set of states* or *state space*. In general, there is a fixed *transition probability* P_{ij} that the next state will be j given the current state is i. The following theorem state this result.

Theorem 1 (Theorem 12.1 Yates & Goodman 2005)
The transition probabilities P_{ij} of a Markov Chain satisfy

$$P_{ij} \geq 0, \quad \sum_{j=0}^{\infty} p_{ij} = 1.$$

A Markov chain can be represented by a graph with nodes representing the sample space of X_n and directed arcs (i, j) for all pairs of states (i, j) such that $P_{ij} > 0$.

Example 1 (Example 12.1 Yates & Goodman 2005)
The two state Markov chain can be used to model a wide variety of systems that alternate between ON and OFF states. After each unit of time in the OFF state, the system turns on with probability p. After each unit of time in the ON state, the system turns OFF with probability q. Using 0 and 1 as states, what is the Markov chain for the system?

Solution
The MC for the above system

Thus the transition probabilities are $P_{00} =$ $P_{01} =$ $P_{10} =$ $P_{11} =$

Example 2(Example 12.2 Yates & Goodman 2005)
A packet voice communications system transmits digitised speech only during "talk-spurts" when the speaker is talking. In every 10ms interval (referred as time slot) the system decides whether the speaker is talking or silent. When the speaker is talking, a speech packet is generated; otherwise no packet is generated. If the speaker is silent in a slot, then the speaker is talking in the next slot with probability $p = 1/140$. If the speaker is talking in a slot, the speaker is silent in the next slot with probability $q = 1/100$. If states 0 and 1 represent silent and talking, sketch the Markov chain for this packet voice system.

Solution

Example 3(Example 12.4 Yates & Goodman 2005)
In a discrete random walk, a person's position is marked by an integer on the real line. Each unit of time, the person randomly moves one step, either to the right(with probability p) or to the left. Sketch the Markov chain.

Solution

Exercise

Consider a simple weather forecasting model in which we classify the day's weather as either sunny or rainy. On the basis of previous data we have determined that if it is sunny today, there is an 80% chance that it will be sunny tomorrow regardless of the past weather, whereas if it is rainy today there is a 30% chance that it will be rainy tomorrow, regardless of the past. Let X_n be the weather on day n. By labelling sunny as state 1 and rainy as state 2, draw the Markov chain and write down its transition probabilities.

Solution

The graphical representation of Markov chain encourages the use of special terminology like *branch probability, hop and path.*

- Transition probability is also called a *branch probability* as it equals the probability of following the branch from state i to j. The sum of branch probabilities leaving any state i to j must equal to 1.
- A state transition is also called a *hop* as a transition from i to j can be viewed as hopping form i to j on a Markov chain.
- A State sequence results from a sequence of hops in the Markov chainis called a *path.* e.g. A state sequence of i, j, k corresponding to a sequence of states $X_n = i, X_{n+1} = j$, and $X_{n+2} = k$ is a two-hop from i to k. We consider the state sequence i, j, k as a path in the Markov chain if each corresponding state transition has non zero probability.

We will focus on a finite state Markov chain first. Its one-step transition probabilities can be represented by the matrix

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & ... & P_{0K} \\ P_{10} & P_{11} & & \vdots \\ \vdots & & \ddots & \\ P_{K0} & \dots & & P_{KK} \end{bmatrix}.$$

This nonnegative matrix $\mathbf{P}$ with rows sum to 1 is called a *state transition matrix* or a *stochastic matrix*.

Exercises

1. Quiz 12.1 Yates & Goodman 2005
2. Problem 12.3.1
3. Problem 12.3.5

3.3 Discrete Time Markov Chain Dynamics

The dynamics in Markov chain describes the variation of the state over a short time interval starting from a given initial state. They are random processes and we cannot say exactly which sequence of states will follow the initial state. However, there are many applications that need the information of future state given the current state X_m. This prediction of future state X_{n+m} conditioned on present state X_m requires knowledge of conditional probability function of X_{n+m} given X_m. The $n-$step transition probabilities that we will define next contains this information.

Definition (Definition 12.2 Yates & Goodman 2005 pg 448)
For a finite Markov chain, the n-step transition probabilities are given by the matrix $\mathbf{P}(n)$ which has i, jth element

$$P_{ij}(n) = P[X_{n+m} = j | X_m = i].$$

The following result known as Chapman-Kolmogorov equations give a recursive procedure for calculating the n-step transition probabilities. It is based on the observation that going from i to j in $n + m$ steps passes through some state k after n-steps. The other result is equivalent in matrix or vector form.

Theorem 2(Theorem 12.2 Yates & Goodman 2005)

For a finite Markov chain, the n-step transition probabilities satistfy

$$P_{ij}(n+m) = \sum_{k=0}^{K} P_{ik}(n) P_{kj}(m), \qquad \mathbf{P}(n+m) = \mathbf{P}(n)\mathbf{P}(m)$$

By the definition of n- step transition probability,

$$\begin{aligned} P_{ij}(n+m) &= \sum_{k=0}^{K} P[X_{n+m} = j, X_n = k | X_0 = i] \quad (3.1) \\ &= \sum_{k=0}^{K} P[X_n = k | X_0 = i] P[X_{n+m} = j | X_n = k, X_0 = i] \quad (3.2) \\ &= \sum_{k=0}^{K} P_{ik} P[X_{n+m} = j | X_n = k, X_0 = i] \quad (3.3) \end{aligned}$$

By Markov property, $P[X_{n+m} = j | X_n = k, X_0 = i] = P[X_{n+m} = j | X_n = k] = P_{kj}(m)$. This completes the proof.

Theorem 3(Theorem 12.3 Yates & Goodman 2005)
For a finite Markov chain with transition matrix $\mathbf{P}$, the n-th step transition matrix is

$$\mathbf{P}(n) = \mathbf{P}^n.$$

Example 4
Based on Example 2, calculate $\mathbf{P}^2$.

Solution

The following example shows how to calculate the power of matrix using eigenvalues. Given a square matrix $m \times m$,P, there exists an eigenvalue called λ if there is a nonzero vector x such that

$$xP = \lambda x$$

and

$$\det(\lambda I - P) = 0,$$

where x is called left eigenvector of P corresponding to λ. Since $\det(\lambda I - P)$ is a polynomial of order m in λ, there are m eigenvalues of P. P is then can be written as $P = X^{-1}DX$, where $D = \text{diag}[\lambda_1 ... \lambda_m]$. By the property of the matrix,

$$P^n = X^{-1}D^nX$$

.

Example 5 (Example 12.6 Yates & Goodman 2005)
Based on Example 1, find the n-step transition matrix $\mathbf{P}^n$. Given the system is OFF at time 0 what is the probability that the system is OFF at time $n = 33$. Comment on your result.

Solution

The state transition matrix is

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

The eigenvalues of $\mathbf{P}$ are $\lambda_1 = 1$ and $\lambda_2 = 1 - (p+q)$. Since p and q are probabilities, $|\lambda_2| \leq 1$. $\mathbf{P}$ can be expressed in the diagonalised form

$$\mathbf{P} = \mathbf{X^{-1}DX} = \begin{bmatrix} 1 & \frac{-p}{p+q} \\ 1 & \frac{q}{p+q} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ -1 & 1 \end{bmatrix}.$$

The n-th step transition matrix is given by

$$\mathbf{P^n} = \begin{bmatrix} P_{00}(n) & P_{01}(n) \\ P_{10}(n) & P_{11}(n) \end{bmatrix} = \mathbf{X^{-1}D^nX} = \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{\lambda_2^n}{p+q} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}$$

Given the system is OFF at time 0, the probability that the system will go OFF at $n = 33$ is as below

$$P_{00}(33) = \frac{q}{p+q} + \frac{\lambda_2^{33} p}{p+q} = \frac{q + (1-(p+q))^{33} p}{p+q}$$

The nth step transition matrix describe the evolution of probabilities in a Markov chain. The calculation of n step transition probability is nontrivial.

Sometimes, we do not need the complete set of n step transition probability but only a few element in the n step transition matrix, for example we only need to know $P[X_n = i]$. Let use the following notation to represent the state probabilities at time n.

$$\{p_j(n) | j = 0, 1, 2, ...K\} \text{ where } p_j(n) = P[X_n = j]$$

equivalently,

$$\mathbf{p}(n) = [p_0(n) \dots p_K(n)]^T.$$

Definition (Defintion 12.3 Yates & Goodman, 2005)
A vector $\mathbf{p}$= $[p_0(n) \dots p_K(n)]^T$ is a state probability vector if $\sum_{j=0}^{K} p_j = 1$, and each element p_j is nonnegative.

The following theorem shows how to calculate the state probability vector $\mathbf{p}(n)$ at any time n with initial state probability vector $\mathbf{p}(0)$ of a Markov chain.

Theorem 4(Theorem 12.4 Yates & Goodman, 2005)
The state probabilities $p_j(n)$ at time n can be found by either one iteration with n-step transition probabilities:

$$p_j(n) = \sum_{i=0}^{K} p_i(0)P_{ij}(n), \qquad \mathbf{p}^T(n) = \mathbf{p}^T(0)\mathbf{P}^n$$

OR n iterations with one-step transition probabilities:

$$p_j(n) = \sum_{i=0}^{K} p_i(n-1)P_{ij}, \qquad \mathbf{p}^T(n) = \mathbf{p}^T(n-1)\mathbf{P}$$

Proof
From the definition of n-step transition probabilities,

$$p_j(n) = P[X_n = j] = \sum_{i=0}^{K} P[X_n = j|X_0 = i]P[X_0 = i] = \sum_{i=0}^{K} P_{ij}(n)p_i(0)$$

From the defintion of transition probability,

$$p_j(n) = P[X_n = j] = \sum_{i=0}^{K} P[X_n = j|X_{n-1} = i]P[X_{n-1} = i] = \sum_{i=0}^{K} P_{ij}p_i(n-1)$$

Example 6(Example 12.7 Yates & Goodman 2005)
For the two state Markov chain in Example 1 with initial state probabilities $\mathbf{p}(0) = [p_0 p_1]$, find the state probability vector $\mathbf{p}(n)$.

Solution

$\mathbf{p}^T(n) = \mathbf{p}^T(0)\mathbf{P}^n$ and by using $\mathbf{P}^n$ found in Example 5, the state probabilities at time n is

$$\mathbf{p}^T(n) = [p_0(n)p_1(n)] = [\frac{q}{p+q} \quad \frac{p}{p+q}] + \lambda_2^n[\frac{p_0p - p_1q}{p+q} \quad \frac{-p_0p - p_1q}{p+q}]$$

Exercises Quiz 12.2, Problem 12.2.1, Problem 12.2.2

3.3.1 Limiting State Probabilities for a Finite Markov Chain

In this subsection we will analyse the Markov chain by studying the behaviour of state probability vector $\mathbf{p}(n)$ in the long run that is as $n \to \infty$.

Definition (Definition 12.4)

Given a finite Markov chain with initial state probability vector $\mathbf{p}(0)$, the limiting state probailities, when they exist, are defined as the vector, $\pi = \lim_{n\to\infty} \mathbf{p}(\mathrm{n})$.

Note The jth element, π_j of **pi** is the probability the system will be in state j in the distant future.

Example 7(Example 12.8 Yates & Goodman 2005)
Based on Example 2, what is the limiting state probability vector $[\pi_0\pi_1]^T = \lim_{n\to\infty} \mathbf{p}(n)$.

Solution

Given $\mathbf{p}^T(0) = [p_0 \ \ p_1]^T$,

$$\mathbf{p}^T(n) = [p_0(n)p_1(n)] = [\frac{q}{p+q} \quad \frac{p}{p+q}] + \lambda_2^n[\frac{p_0p - p_1q}{p+q} \quad \frac{-p_0p \ - p_1q}{p+q}]$$

substitute $p = \frac{1}{140}$ and $q = \frac{1}{100}$, this gives the value of $\lambda_2 = 1 - (p+q) = \frac{344}{350}$.

$$\mathbf{p}^T(n) = \left[\frac{7}{12}\frac{5}{12}\right] + \lambda_2^n\left[\frac{5}{12}p_0 - \frac{7}{12}p_1 \ \frac{-5}{12}p_0 + \frac{7}{12}p_1\right].$$

Since $|\lambda_2| < 1$, the limiting state probabilities are

$$\lim_{n\to\infty} \mathbf{p}^T(n) = \left[\frac{7}{12} \ \frac{5}{12}\right]$$

The limiting state probabilities are the same regardless of its initial state probabilities.

Theorem 5 (Theorem 12.5 Yates & Goodman 2005)
If a finite Markov chain with transition matrix $\mathbf{P}$ and initial state probability $\mathbf{p}(0)$ has limiting state probability vector $\pi = \lim_{n\to\infty} \mathbf{p}(n)$ then

$$\pi^{\mathbf{T}} = \pi^{\mathbf{T}}\mathbf{P}.$$

The limiting state probability vector need to satisfy certain constraints. Its existence depends on the existence of the limit as well as whether or not depend on the initial state probability vector. That is there are three possibilities of a finite Markov chain

1. $\lim_{n\to\infty} \mathbf{p}(n)$ exists, and independent of initial state probability vector $\mathbf{p}(0)$
2. $\lim_{n\to\infty} \mathbf{p}(n)$ exists, but depends on $\mathbf{p}(0)$
3. $\lim_{n\to\infty} \mathbf{p}(n)$ does not exist.

All these depend on the stationary probability vector that we will see its definition below.

Definition(Definition Yates & Goodman 2005)

For a finite Markov chain with transition probability matrix **P**, a state probability vector π is stationary if $\pi^T = \pi^T P$

Theorem 6 (Yates & Goodman 2005) shows that a finite Markov chain X_n is a stationary stochastic process. ie at every time t, the same random variable can be observed.

For case (1) - Markov chain **P** has unique stationary probability vector
For case (2) - Markov chain **P** has multiple stationary probability vector
For case (3) - Markov chain **P** does not have stationary probability vector

Example 12.10 Yates & Goodman (2005) describes these three possibilities.

Exercises Quiz 12.3, Problem 12.3.1 [hint: $|\pi_j - p_j| \leq 0.01\pi_j$; ans:267].

3.3.2 State Classification

It is important for us to see the classification of the state in Markov chain as it helps us to determine the class of its limiting state probability vector. That is we can see whether the state probability vector converges to a unique stationary probability and independent of the initial state probability vector $\mathbf{p}(0)$.

Definition (Defintion 12.6 Yates & Goodman 2005)
State j is **accessible** from state i, written as $i \to j$, if $Pij(n) > 0$ for some $n > 0$.

Definition (Defintion 12.7 Yates & Goodman 2005)
States i and j **communicate**, written as $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

Note State i always communicate with itself.

Definition (Definition 12.8 Yates & Goodman 2005)
A communicating class is a nonempty subset of states C such that, if $i \in C$, then $j \in C$ if and only if $i \leftrightarrow j$.

Example 8 (Example 12.11 Yates & Goodman 2005)
In the following Markov Chain, the branches corresponding to transition probabilities $P_{ij} > 0$. For this chain, identify the communicating classes.

3 communicating classes: $C_1 = \{0, 1, 2\}$, $C_2 = \{4, 5, 6\}$ $C_3 = \{3\}$

Definition (Definition 12.9 Yates & Goodman 2005) State i has period d if d is the largest integer such that $P_{ii}(n) = 0$ whenever n is not divisible by d. If $d = 1$, then state i is called aperiodic.

Alternative definition: State i has period d if d is the largest integer such that $P_{ii}(n) > 0$ whenever n is divisible by d. If $d = 1$, then state i is called aperiodic.

Theorem 8 Theorem 12.7 (Yates & Goodman 2005) says that all states in a communicating class have the same period.

Example
Give an example to describe this theorem.

Definition (Definition 12.10 Yates & Goodman 2005)
In a finite Markov chain, a state i is **transient** if there exists a state j such that $i \rightarrow j$ but $j \rightarrow i$(not accessible from j); otherwise, if no such state j exists, then state i is **recurrent**.

Definition (Definition 12.11 Yates & Goodman 2005)
A Markov chain is irreducible if there is only one communicating class.

The following theorems show some results on state classifications. Firstly, if the two states communicate, then both must either be recurrent or transient.

Theorem 9(Theorem 12.8 Yates & Goodman, 2005)
If i is recurrent and $i \leftrightarrow j$, then j is recurrent.

Proof
Consider a state, k, need to show that j and k communicate $[j \rightarrow k \Rightarrow k \rightarrow j]$.
Since $i \leftrightarrow j$ and $j \rightarrow k$,there is a path from i to k via j. Thus $i \rightarrow k$. Since i is recurrent, i must be accesible from k, $(k \rightarrow i)$.
Since $i \leftrightarrow j$, there is a path from j to i and then to k.

Think If i is transient and $i \leftrightarrow j$, then j is transient.

Theorem 10(Theorem 12.9 Yates & Goodman, 2005)
If state i is transient, then N_i, the number of visits to state i over all time, has expected value $E[N_i] \leq \infty$.

Example 8(Example 12.14 Yates & Goodman, 2005)
For the following MC, identify each communicating class and indicate whether it is transient or recurrent.

Theorem 11(Theorem 12.10 Yates & Goodman, 2005)
A finite state MC always has a recurrent communicating class.

Exercises Quiz 12.4

In the following MC, all transitions with nonzero probability are shown.

1. What are the communicating classes?
2. For each communicating class, identify whether the states are periodic or aperiodic.
3. For each communicating class, identify whether the states are transient or recurrent.

3.3.3 Limit Theorems for Irreducible Finite MCs

Theorem 12 (Theorem 12.11 Yates & Goodman 2005)
For an irreducible, aperiodic, finite MC with states $\{0, 1, ...\}$, the limiting n-th step transition matrix

$$\mathbf{lim_{n\to}P}^n = \mathbf{1}\pi^T \begin{bmatrix} \pi_0 & \pi_1 & ... & \pi_K \\ \pi_0 & \pi_1 & ... & \pi_K \\ \vdots & & \ddots & \\ \pi_0 & \pi_1 & ... & \pi_K \end{bmatrix}.$$

where ($\mathbf{1}$) is the column vector $[1 \; ... \; 1]^T$ and $\pi = [\pi_0 \; ... \; \pi_K]^T$ is the unique vector satisfying $\pi^{\mathbf{T}} = \pi^{\mathbf{T}}\mathbf{P}, \qquad \pi^{\mathbf{T}}\mathbf{1} = \mathbf{1}$.

Theorem 13(Theorem 12.12 Yates & Goodman 2005)
For an irreducible, aperiodic, finite Markov chain with transition matrix $\mathbf{P}$ and initial state probability vector $\mathbf{p}(0)$, $\lim_{n\to\infty} \mathbf{p}(n) = \pi$

Example 9(Example 12.16 Yates & Goodman 2005)
Based on MC for Example 6, use the above theorem to calculate the stationary probabilities $[\pi_0 \;\; \pi_1]$.

Exercises

1. Problem 12.5.1
   Based on the following state transition matrix, find the stationary distribution π,

$$\mathbf{P} = \begin{bmatrix} 0.998 & 0.001 & 0.001 \\ 0.004 & 0.98 & 0.016 \\ 0.008 & 0.012 & 0.98 \end{bmatrix}.$$

   [ans: [0.7536 0.1159 0.1304]

2. Problem 12.5.2
   Based on the following stochastic matrix, find the stationary probability vector,

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

   [ans: [0.5 0.5 0]]

3. Consider a customer who buys soft-drink at a regular intervals, say once a week, and every time he has to decide among three brands: Fizi, Kola, and 8up. From his buying records so far, it has been determined that his brand of choice in week $n+1$ depends only on his last week's choice (week $n-$th) regardless of previous purchases. Let X_n be the brand he purchases in week n. Then it is obvious that X_n for $n \geq 0$ is a discrete time Markov chain. A transition probability matrix is as follows;

$$\mathbf{P} = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}.$$

   (a) Determine whether this Markov chain is irreducible and aperiodic.
   (b) Find the limiting distribution of brands switching.
   [ans: [0.132 0.319 0.549]]

3.3.4 Branching Process

Stochastic processes have been used in sociology to study the issues of social mobility (how the economic status of the n-th generation affects that of $n+1$-st generation), effect of social traditions and many others. Here we are going to study about the family names propagation. The tool to model this phenomenen is discrete time Markov chain. It is called branching process.
Consider the following family tree.

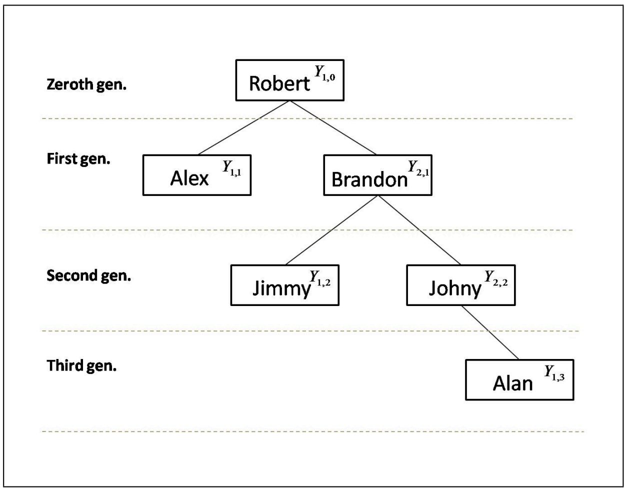

Figure 3.1: Family Tree

X_n- the number of individuals in the n-th generation starting with $X_0 = 1$
$Y_{i,n}$ - the number of male offspring of the i-th person in the n-th generation

$$X_{n+1} = Y_{1,n} + Y_{2,n} + ... + Y_{X_n,n} = \sum_{i=1}^{X_n} Y_{i,n}$$

Suppose $Y_{i,n}$ are iid random variables, then X_n for $n \geq 0$ is a discrete time Markov chain with state space $\{0, 1, 2, ...\}$ and transition probabilities

$$\begin{aligned} P_{ij} &= P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = 1] \\ &= P[\sum_{r=1}^{X_n} Y_{r,n} = j | X_n = i, ..., X_0 = 1] \\ &= P[\sum_{r=1}^{i} Y_{r,n} = j] \end{aligned}$$

This process X_n is called a branching process.
Among questions of interest are What is the probability that eventually the family names become extinct?, How long does it take for the family name to become extinct? How many total males are produced in this family? What is the distribution of the size of the n-th generation?

Some results concerning the expected family size at $n+1$st generation and the probability of family name dies out.

$$E[X_{n+1}] = \mu^{n+1}$$

where μ is the mean number of offspring per individual.

$$\pi_0 = \sum_{j=0}^{\infty} \pi_0^j p_j$$

where p_j is the probability by the end of an individual lifetime of producing j offsprings.