2. Markov Chains

2.1 Stochastic Process

- A *stochastic process* $\{X(t), t \in T\}$ is a collection of random variables. That is, for each $t \in T$,$X(t)$ is a random variable.

- The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the state of the process at time t. For example, $X(t)$ might equal the total number of customers that have entered a supermarket by time t; or the number of customers in the supermarket at time t; or the total amount of sales that have been recorded in the market by time t; etc.

- The set T is called the index set of the process. When T is a countable set the stochastic process is said to be a *discrete-time process*. If T is an interval of the real line, the stochastic process is said to be a *continuous-time process*.

- The state space of a stochastic process is defined as the set of all possible values that the random variables $X(t)$ can assume.

- In applications to physics for example one can mention phase transitions, atomic emission phenomena, etc. In biology the time behaviuor of live beings is often subject to randomness, at least when the observer only handles partial information.

- This point is of importance, as it shows that the notion of randomness is linked to the concept of information: what appears random to an observer may not be random to another observer equipped with more information.

- In finance the importance of modeling time-dependent random phenomena is quite clear as no one can make fully accurate predictions for the future moves of risky assets.

- The concrete outcome of such modeling lies in the computation of *expectations or expected values* , which often turn out to be more useful than the probability values themselves.

- A stochastic process $\{X(t), t \geq 0\}$ is said to be a *stationary process or strong stationary process* if for all $n, s, t_1, \ldots, t_n$, the random vectors $X(t_1), \ldots, X(t_n)$ and $X(t_1 + s), \ldots, X(t_n + s)$ have same joint distribution. $X(t)$ is defined to be a *weakly stationary process* if $\mathrm{E}X(t) = c$ and $\mathrm{Cov}(X(t), X(t+s)) = R(s)$ does not dependent on t, where $R(s)$ is called as the *autocovariance* of the process.

2.2 Markov Chains: Introduction

- Let X_n denote its value in time period n, and suppose we want to make a probability model for the sequence of successive values $X_0, X_1, X_2 \ldots$ The simplest model would probably be to assume that the X_n are independent random variables, but often such an assumption is clearly unjustified.

- For instance, starting at some time suppose that X_n represents the price of one share of some security, such as Google, at the end of n additional trading days. Then it certainly seems unreasonable to suppose that the price at the end of day $n+1$ is independent of the prices on days $n, n-1, n-2$ and so on down to day 0.

- However, it might be reasonable to suppose that the price at the end of trading day $n+1$ depends on the previous end-of-day prices only through the price at the end of day n. That is, it might be reasonable to assume that the conditional distribution of X_{n+1} given all the past end-of-day prices $X_n, X_{n-1}, \ldots, X_0$ depends on these past prices only through the price at the end of day n.

- Let $\{X_n, n = 0, 1, 2, \ldots\}$ be a stochastic process that takes on a finite or countable number of possible values. If $X_n = i$, then the process is said to be in state i at time n. We suppose that whenever the process is in state i, there is a fixed probability P_{ij} that it will next be in state j. That is, we suppose that

$$P(X_{n+1} = j | X_n = i, x_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) = P_{ij}$$

  for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geq 0$. Such a stochastic process is known as a *Markov chain*.

- The value P_{ij} represents the probability that the process will, when in state i, next make a transition into state j, so we have

$$p_{ij} \geq 0, i, j \geq 0, \sum_{j=0}^{\infty} p_{ij} = 1, i = 0, 1, \ldots$$

- Let $\mathbf{P}$ denote the matrix of one-step transition probabilities P_{ij} , so that

$$\mathbf{P} = \left\| \begin{array}{cccc} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{array} \right\|$$

- **Example 1.** (Forecasting the Weather). Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain tomorrow with probability β.

$$\mathbf{P} = \left\| \begin{array}{cc} \alpha & 1-\alpha \\ \beta & 1-\beta \end{array} \right\|$$

- **Example 2**. (Transforming a Process into a Markov Chain). Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2.

$$\mathbf{P} = \left\| \begin{array}{cccc} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{array} \right\|$$

- **Example 3**. (A Gambling Model). Consider a gambler who, at each play of the game, either wins $\$1$ with probability p or loses $\$1$ with probability $1 - p$. If we suppose that our gambler quits playing either when he goes broke or he attains a fortune of $\$$N, then the gambler's fortune is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1}, i = 1, 2, \ldots, N - 1,$$

$$P_{00} = P_{NN} = 1$$

States 0 and N are called *absorbing states* since once entered they are never left. Note that the preceding is a finite state random walk with absorbing barriers (states 0 and N).

2.3. Chapman-Kolmogorov Equations

We have already defined the one-step transition probabilities P_{ij} . We now define the n-step transition probabilities P_{ij}^n to be the probability that a process in state i will be in state j after n additional transitions. That is,

$$P_{ij}^n = P(X_{n+k} = j | X_k = i), n \geq 0, i, j \geq 0$$

- The Chapman-Kolmogorov equations provide a method for computing these n-step transition probabilities. These equations are

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m, \quad \text{for all } n, m \geq 0 \quad \text{all } i, j.$$

- If we let $\mathbf{P}(n)$ denote the matrix of n-step transition probabilities P_{ij}^n , then the above equation asserts that

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

Example 4. Consider Example 1 in which the weather is considered as a two-state Markov chain. If $\alpha = 0.7$ and $\beta = 0.4$, then calculate the probability that it will rain four days from today given that it is raining today.

$$\mathbf{P}^{(4)} = \left\| \begin{array}{cc} 0.7 & 0.3 \\ 0.4 & 0.6 \end{array} \right\|^4 = \left\| \begin{array}{cc} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{array} \right\|$$

Example 5. Consider Example 2. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

$$\mathbf{P}^{(2)} = \left\| \begin{array}{cccc} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{array} \right\|^2 = \left\| \begin{array}{cccc} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{array} \right\|$$

Since rain on Thursday is equivalent to the process being in either state 0 or state 1 on Thursday, the desired probability is given by $P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61$

Example 6. In a sequence of independent flips of a fair coin, let N denote the number of flips until there is a run of three consecutive heads. Find (a) $P(N \leq 8)$ and (b) $P(N = 8)$. (Solution: (a) $107/256 \approx .4180$. (b) $1/2P_{0,2}^7$.)

Remark. So far, all of the probabilities we have considered are conditional probabilities. For instance, P_{ij}^n is the probability that the state at time n is j given that the initial state at time 0 is i. If the unconditional distribution of the state at time n is desired, it is necessary to specify the probability distribution of the initial state. Let us denote this by

$$\alpha_i = P(X_0 = i), i \geq 0, \sum_{i=0}^{\infty} \alpha_i = 1$$

All unconditional probabilities may be computed by conditioning on the initial state. That is,

$$P(X_n = j) = \sum_{i=0}^{\infty} P(X_n = j | X_0 = i) P(X_0 = i) = \sum_{i=0}^{\infty} P_{ij}^n \alpha_i.$$

2.4 Classification of States

- State j is said to be accessible from state i if $P_{ij}^n > 0$ for some $n \geq 0$. Note that this implies that state j is accessible from state i if and only if, starting in i, it is possible that the process will ever enter state j. This is true since if j is not accessible from i, then

$$\begin{aligned} P(\text{ever be in } j|\text{start in } i) &= P\left\{\bigcup_{n=0}^{\infty}(X_n = j)\Big|X_0 = i\right\} \\ &\leq \sum_{n=0}^{\infty} P(X_n = j|X_0 = i) = \sum_{n=0}^{\infty} P_{ij}^n = 0 \end{aligned}$$

- Two states i and j that are accessible to each other are said to communicate, and we write $i \leftrightarrow j$.

Note that any state communicates with itself since, by definition,

$$P_{ii}^0 = P\{X_0 = i | X_0 = i\} = 1$$

The relation of communication satisfies the following three properties:

1. State i communicates with state i, all $i \geq 0$.
2. If state i communicates with state j, then state j communicates with state i.
3. If state i communicates with state j, and state j communicates with state k, then state i communicates with state k by that if $P_{ij}^n > 0, P_{jk}^m > 0$, then

$$P_{ik}^{n+m} = \sum_{r=0}^{\infty} P_{ir}^n P_{rk}^m \geq P_{ij}^n P_{jk}^m > 0.$$

- Two states that communicate are said to be in the same class. Any two classes of states are either identical or disjoint. The concept of communication divides the state space up into a number of separate classes. The Markov chain is said to be *irreducible* if there is only one class, that is, if all states communicate with each other.

Example 7. Consider a Markov chain consisting of the four states 0, 1, 2, 3 and having transition probability matrix

$$\mathbf{P} = \left\| \begin{matrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{matrix} \right\|$$

The classes of this Markov chain are $\{0, 1\}, \{2\}$, and $\{3\}$. Note that while state 0 (or 1) is accessible from state 2, the reverse is not true. Since state 3 is an absorbing state, that is, $P_{33} = 1$, no other state is accessible from it.

- For any state i we let f_i denote the probability that, starting in state i, the process will ever reenter state i. State i is said to be *recurrent* if $f_i = 1$ and *transient* if $f_i < 1$.
- If state i is recurrent then, starting in state i, the process will reenter state i again and again and again, in fact, infinitely often.
- if state i is transient then, starting in state i, the number of time periods that the process will be in state i has a geometric distribution with finite mean $1/(1 - f_i)$.

Proposition 1. State i is

$$\text{recurrent if } \sum_{n=1}^{\infty} P_{ii}^{n} = \infty$$

$$\text{transient if } \sum_{n=1}^{\infty} P_{ii}^{n} < \infty$$

Corollary 1. If state i is recurrent, and state i communicates with state j , then state j is recurrent.

Remarks:

(i) Corollary 1 also implies that transience is a class property. For if state i is transient and communicates with state j, then state j must also be transient.

(ii) Corollary 1 along with the result that not all states in a finite Markov chain can be transient leads to the conclusion that all states of a finite irreducible Markov chain are recurrent.

Example 8. Let the Markov chain consisting of the states 0, 1, 2, 3 have the transition probability matrix

$$\mathbf{P} = \left\| \begin{array}{cccc} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right\|$$

Determine which states are transient and which are recurrent. (Solution: It is a simple matter to check that all states communicate and, hence, since this is a finite chain, all states must be recurrent.)

Example 9. Consider the Markov chain having states 0, 1, 2, 3, 4 and

$$\mathbf{P} = \left\| \begin{array}{ccccc} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{array} \right\|$$

(Solution: This chain consists of the three classes $\{0, 1\}, \{2, 3\}$, and $\{4\}$. The first two classes are recurrent and the third transient.)

Example 10. (A Random Walk). Consider a Markov chain whose state space consists of the integers $i = 0, \pm 1, \pm 2, \ldots$, and has transition probabilities given by

$$P_{i,i+1} = p = 1 - P_{i,i-1}, i = 0, \pm 1, \pm 2, \ldots$$

where $0 \leq p \leq 1$. In other words, on each transition the process either moves one step to the right (with probability p) or one step to the left (with probability $1 - p$). One colorful interpretation of this process is that it represents the wanderings of a drunken man as he walks along a straight line. Another is that it represents the winnings of a gambler who on each play of the game either wins or loses one dollar.

Since all states clearly communicate, it follows from Corollary 1 that they are either all transient or all recurrent. So let us consider state 0 and attempt to determine if $\sum_{n=1}^{\infty} P_{00}^{n}$ is finite or infinite.

2.5 Long-Run Proportions and Limiting Probabilities

For pairs of states $i \neq j$, let $f_{i,j}$ denote the probability that the Markov chain, starting in state i, will ever make a transition into state j . That is,

$$f_{i,j} = P(X_n = j \text{ for some } n > 0 | X_0 = i)$$

We then have the following result.

Propoistion 2. If i is recurrent and i communicates with j, then $f_{i,j} = 1$.

If state j is recurrent, let m_j denote the expected number of transitions that it takes the Markov chain when starting in state j to return to that state. That is, with $N_j = \min\{n > 0 : X_n = j\}$ equal to the number of transitions until the Markov chain makes a transition into state j, $m_j = \mathrm{E}[N_j | X_0 = j]$

Definiton. Say that the recurrent state j is positive recurrent if $m_j < \infty$ and say that it is null recurrent if $m_j = \infty$.

Now suppose that the Markov chain is irreducible and recurrent. In this case we now show that the long-run proportion of time that the chain spends in state j is equal to $1/m_j$. That is, letting π_j denote the long-run proportion of time that the Markov chain is in state j , we have the following proposition.

Proposition 3. If the Markov chain is irreducible and recurrent, then for any initial state

$$\pi_j = 1/m_j$$

Because $m_j < \infty$ is equivalent to $1/m_j > 0$, it follows from the preceding that state j is positive recurrent if and only if $\pi_j > 0$. We now exploit this to show that positive recurrence is a class property.

Proposition 4. If i is positive recurrent and $i \leftrightarrow j$ then j is positive recurrent.

- It follows from the preceding result that null recurrence is also a class property.
- An irreducible finite state Markov chain must be positive recurrent.
- The classical example of a null recurrent Markov chain is the one dimensional symmetric random walk.

To determine the long-run proportions $\{\pi_j, j \geq 1\}$, note, because π_i is the long-run proportion of transitions that come from state i, that

$$\pi_i P_{i,j} = \text{long-run proportion of transitions that go from state } i \text{ to state } j$$

Summing the preceding over all i now yields that

$$\pi_j = \sum_i \pi_i P_{i,j}$$

Indeed, the following important theorem can be proven.

Theorem 1. Consider an irreducible Markov chain. If the chain is positive recurrent then the long-run proportions are the unique solution of the equations

$$\pi_j = \sum_i \pi_i P_{i,j}, j \geq 1, \text{and} \quad \sum_j \pi_j = 1$$

Moreover, if there is no solution of the preceding linear equations, then the Markov chain is either transient or null recurrent and all $\pi_j = 0$.

Example 11. Consider Example 3 in which the mood of an individual is considered as a three-state Markov chain having a transition probability matrix

$$\mathbf{P} = \left\| \begin{array}{ccc} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{array} \right\|$$

In the long run, what proportion of time is the process in each of the three states? (Solution: $\pi_0 = 21/62, \pi_1 = 23/62, \pi_2 = 18/62$)

Example 12. (A Model of Class Mobility). A problem of interest to sociologists is to determine the proportion of society that has an upper- or lower-class occupation. One possible mathematical model would be to assume that transitions between social classes of the successive generations in a family can be regarded as transitions of a Markov chain. That is, we assume that the occupation of a child depends only on his or her parent's occupation. Let us suppose that such a model is appropriate and that the transition probability matrix is given by

$$\mathbf{P} = \left\| \begin{array}{ccc} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.70 & 0.25 \\ 0.01 & 0.50 & 0.49 \end{array} \right\|$$

(Solution: $\pi_0 = 0.07, \pi_1 = 0.62, \pi_2 = 0.31$)

Example 13. (The Hardy-Weinberg Law and a Markov Chain in Genetics). Consider a large population of individuals, each of whom possesses a particular pair of genes, of which each individual gene is classified as being of type A or type a. Assume that the proportions of individuals whose gene pairs are AA,aa, or Aa are, respectively, p_0, q_0, and $r_0(p_0 + q_0 + r_0 = 1)$. When two individuals mate, each contributes one of his or her genes, chosen at random, to the resultant offspring. Assuming that the mating occurs at random, in that each individual is equally likely to mate with any other individual, we are interested in determining the proportions of individuals in the next generation whose genes are AA, aa, or Aa. Calling these proportions p, q, and r, they are easily obtained by focusing attention on an individual of the next generation and then determining the probabilities for the gene pair of that individual.

The long run proportions $\pi_j, j \geq 0$, are often called *stationary probabilities*. The reason being that if the initial state is chosen according to the probabilities $\pi_j, j \geq 0$, then the probability of being in state j at any time n is also equal to π_j . That is, if

$$P\{X_0 = j\} = \pi_j, j \geq 0$$

then

$$P\{X_n = j\} = \pi_j \text{ for all } n, j \geq 0$$

The preceding is easily proven by induction, for it is true when $n = 0$ and if we suppose it true for $n - 1$, then writing

$$P\{X_n = j\} = \sum_i P\{X_n = j | X_{n-1} = i\} P\{X_{n-1} = i\} = \sum_i P_{ij} \pi_i = \pi_j$$

by the induction hypothesis by Theorem 1

Example 14. Suppose the numbers of families that check into a hotel on successive days are independent Poisson random variables with mean λ. Also suppose that the number of days that a family stays in the hotel is a geometric random variable with parameter $p, 0 < p < 1$. (Thus, a family who spent the previous night in the hotel will, independently of how long they have already spent in the hotel, check out the next day with probability p.) Also suppose that all families act independently of each other. Under these conditions it is easy to see that if X_n denotes the number of families that are checked in the hotel at the beginning of day n then $\{X_n, n \geq 0\}$ is a Markov chain. Find

(a) the transition probabilities of this Markov chain;

(b) $\mathrm{E}[X_n|X_0 = i]$;

(c) the stationary probabilities of this Markov chain.

Proposition 3. Let $\{X_n, n \geq 1\}$ be an irreducible Markov chain with stationary probabilities $\pi_j, j \geq 0$, and let r be a bounded function on the state space. Then, with probability 1,

$$\lim_{N \to \infty} \frac{\sum_{n=1}^{N} r(X_n)}{N} = \sum_{j=0}^{\infty} r(j)\pi_j$$

Limiting Probabilities

we considered a two-state Markov chain with transition probability matrix

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

Then

$$\mathbf{P}^{(4)} = \left\| \begin{array}{cc} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{array} \right\| \approx \mathbf{P}^{(8)} = \left\| \begin{array}{cc} 0.571 & 0.429 \\ 0.571 & 0.429 \end{array} \right\|$$

- Indeed, it seems that P_{ij}^n is converging to some value as $n \to \infty$, with this value not depending on i, thus making it appear that these long-run proportions may also be limiting probabilities.

- Although this is indeed the case for the preceding chain, it is not always true that the long-run proportions are also limiting probabilities. To see why not, consider a two-state Markov chain having

$$P_{0,1} = P_{1,0} = 1$$

- In general, a chain that can only return to a state in a multiple of $d > 1$ steps (where d = 2 in the preceding example) is said to be *periodic* and does not have limiting probabilities.

- However, for an irreducible chain that is not periodic, and such chains are called *aperiodic*, the limiting probabilities will always exist and will not depend on the initial state.

- Moreover, the limiting probability that the chain will be in state j will equal π_j , the long-run proportion of time the chain is in state j.

- An irreducible, positive recurrent, aperiodic Markov chain is said to be *ergodic*.

2.6 Some Applications

The Gambler's Ruin Problem

Consider a gambler who at each play of the game has probability p of winning one unit and probability $q = 1 - p$ of losing one unit. Assuming that successive plays of the game are independent, what is the probability that, starting with i units, the gambler's fortune will reach N before reaching 0?

Example 15. Suppose Max and Patty decide to flip pennies; the one coming closest to the wall wins. Patty, being the better player, has a probability 0.6 of winning on each flip. (a) If Patty starts with five pennies and Max with ten, what is the probability that Patty will wipe Max out? (b) What if Patty starts with 10 and Max with 20?

A Model for Algorithm Efficiency

The following optimization problem is called a linear program:

$$\text{minimize } \mathbf{cx}, \text{subject to } \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq 0$$

where $\mathbf{A}$ is an $m \times n$ matrix of fixed constants; $\mathbf{c} = (c_1, \ldots, c_n)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ are vectors of fixed constants; and $\mathbf{x} = (x_1, \ldots, x_n)$ is the n-vector of nonnegative values that is to be chosen to minimize $\mathbf{cx} = \sum_{i=1}^{n} c_i x_i$. Supposing that $n > m$, it can be shown that the optimal $\mathbf{x}$ can always be chosen to have at least $n - m$ components equal to 0-that is, it can always be taken to be one of the so-called extreme points of the feasibility region.

The simplex algorithm solves this linear program by moving from an extreme point of the feasibility region to a better (in terms of the objective function $\mathbf{cx}$) extreme point (via the pivot operation) until the optimal is reached. Because there can be as many as $N = C_n^m$ such extreme points, it would seem that this method might take many m iterations, but, surprisingly to some, this does not appear to be the case in practice.

4.7 Mean Time Spend in Transient States

Consider now a finite state Markov chain and suppose that the states are numbered so that $T = \{1, 2, \ldots, t\}$ denotes the set of transient states. Let

$$\mathbf{P}_T = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1t} \\ \vdots & \vdots & \vdots & \vdots \\ P_{t1} & P_{t2} & \cdots & P_{tt} \end{bmatrix}$$

For transient states i and j , let s_{ij} denote the expected number of time periods that the Markov chain is in state j, given that it starts in state i. Let $\delta_{i,j} = 1$ when $i = j$ and let it be 0 otherwise. Condition on the initial transition to obtain

$$s_{ij} = \delta_{i,j} + \sum_k P_{ik} s_{kj} = \delta_{i,j} + \sum_{k=1}^{t} P_{ik} s_{kj}$$

where the final equality follows since it is impossible to go from a recurrent to a transient state, implying that $s_{kj} = 0$ when k is a recurrent state.

Let S denote the matrix of values $s_{ij}, i, j = 1, \ldots, t$. That is,

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1t} \\ \vdots & \vdots & \vdots & \vdots \\ s_{t1} & s_{t2} & \cdots & s_{tt} \end{bmatrix}$$

So the above equation can be written as

$$\mathbf{S} = \mathbf{I} + \mathbf{P}_T \mathbf{S} \quad \text{and} \quad (\mathbf{I} - \mathbf{P}_T)\mathbf{S} = \mathbf{I}$$

Hence

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1}$$

That is, the quantities $s_{ij}, i \in T, j \in T$, can be obtained by inverting the matrix $\mathbf{I} - \mathbf{P}_T$.

- For $i \in T, j \in T$, the quantity f_{ij} , equal to the probability that the Markov chain ever makes a transition into state j given that it starts in state i, is easily determined from $\mathbf{P}_T$. To determine the relationship, let us start by deriving an expression for s_{ij} by conditioning on whether state j is ever entered. This yields

$$\begin{aligned} s_{ij} &= \mathrm{E}[\text{time in } j|\text{start in } i, \text{ever transit to } j] f_{ij} \\ &\quad + \mathrm{E}[\text{time in } j|\text{start in } i, \text{never transit to } j](1 - fij) \\ &= (\delta_{i,j} + s_{jj}) f_{ij} + \delta_{i,j}(1 - f_{ij}) = \delta_{i,j} + f_{ij} s_{jj} \end{aligned}$$

So

$$f_{ij} = \frac{s_{ij} - \delta_{i,j}}{s_{jj}}.$$

Example 16. Consider the gambler's ruin problem with $p = 0.4$ and $N = 7$. Starting with 3 units, determine

(a) the expected amount of time the gambler has 5 units. (Solution: 0.9228)

(b) the expected amount of time the gambler has 2 units. (Solution: 2.3677)

(c) what is the probability that the gambler ever has a fortune of 1? (Solution: 0.8797)

Suppose we are interested in the expected time until the Markov chain enters some sets of states A, which need not be the set of recurrent states. We can reduce this back to the previous situation by making all states in A absorbing states. That is, reset the transition probabilities of states in A to satisfy

$$P_{i,i} = 1, i \in A$$

This transforms the states of A into recurrent states, and transforms any state outside of A from which an eventual transition into A is possible into a transient state. Thus, our previous approach can be used.

4.8 Branching Process

In this section we consider a class of Markov chains, known as *branching processes*, which have a wide variety of applications in the biological, sociological, and engineering sciences.

- Consider a population consisting of individuals able to produce offspring of the same kind. Suppose that each individual will, by the end of its lifetime, have produced j new offspring with probability $P_j, j \geq 0$, independently of the numbers produced by other individuals.

- We suppose that $P_j < 1$ for all $j \geq 0$. The number of individuals initially present, denoted by X_0, is called the size of the zeroth generation. All offspring of the zeroth generation constitute the first generation and their number is denoted by X_1. In general, let X_n denote the size of the nth generation.

- It follows that $\{X_n, n = 0, 1, \ldots\}$ is a Markov chain having as its state space the set of nonnegative integers.

- Note that state 0 is a recurrent state, since clearly $P_{00} = 1$. Also, if $P_0 > 0$, all other states are transient. This follows since $P_{i0} = P_0^i$, which implies that starting with i individuals there is a positive probability of at least P_0^i that no later generation will ever consist of i individuals.

- Moreover, since any finite set of transient states $\{1, 2, \ldots, n\}$ will be visited only finitely often, this leads to the important conclusion that, if $P_0 > 0$, then the population will either die out or its size will converge to infinity.

- The mean number of offspring of a single individual is denoted by

$$\mu = \sum_{j=0}^{\infty} j P_j$$

and hence the variance of the number of offspring produced by a single individual should be

$$\sigma^2 = \sum_{j=0}^{\infty} (j - \mu)^2 P_j$$

- Let us suppose that $X_0 = 1$, that is, initially there is a single individual present. We calculate $\mathrm{E}[X_n]$ and Var(X_n) by first noting that we may write

$$X_n = \sum_{i=1}^{X_{n-1}} Z_i$$

where Z_i represents the number of offspring of the ith individual of the $(n-1)$st generation. By conditioning on X_{n-1}, we obtain

$$\begin{aligned}
\mathrm{E}[X_n] &= \mathrm{E}[\mathrm{E}[X_n|X_{n-1}]] = \mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^{X_{n-1}} Z_i|X_{n-1}\right]\right] \\
&= \mathrm{E}[X_{n-1}\mu] = \mu\mathrm{E}[X_{n-1}] = \mu^2\mathrm{E}[X_{n-2}] = \cdots = \mu^{n-1}\mathrm{E}[X_1] = \mu^n
\end{aligned}$$

- Similarly, $\mathsf{Var}(X_n)$ may be obtained by using the conditional variance formula

$$\mathsf{Var}(X_n) = \mathrm{E}(\mathsf{Var}(X_n|X_{n-1})) + \mathsf{Var}(\mathrm{E}(X_n|X_{n-1}))$$

- Now, given X_{n-1}, X_n is just the sum of X_{n-1} independent random variables each having the distribution $\{P_j, j \geq 0\}$. Hence,

$$\mathrm{E}(X_n|X_{n-1}) = X_{n-1}\mu, \quad \mathsf{Var}(X_n|X_{n-1}) = X_{n-1}\sigma^2$$

- The conditional variance formula now yields

$$\begin{aligned}
\mathsf{Var}(X_n) &= \mathrm{E}[X_{n-1}\sigma^2] + \mathsf{Var}(X_{n-1}\mu) = \sigma^2\mu^{n-1} + \mu^2\mathsf{Var}(X_{n-1}) \\
&= \sigma^2\mu^{n-1} + \mu^2(\sigma^2\mu^{n-2} + \mu^2\mathsf{Var}(X_{n-2})) \\
&= \sigma^2(\mu^{n-1} + \mu^n) + \sigma^4\mathsf{Var}(X_{n-2}) \\
&= \cdots\cdots \\
&= \sigma^2(\mu^{n-1} + \mu^n + \cdots + \mu^{2n-2}) + \mu^{2n}\mathsf{Var}(X_0) \\
&= \sigma^2(\mu^{n-1} + \mu^n + \cdots + \mu^{2n-2})
\end{aligned}$$

- Therefore,

$$\text{Var}(X_n) = \begin{cases} \sigma^2 \mu^{n-1} \left(\frac{1-\mu^n}{1-\mu} \right), & \text{if } \mu \neq 1 \\ n\sigma^2, & \text{if } \mu = 1 \end{cases}$$

Let π_0 denote the probability that the population will eventually die out (under the assumption that $X_0 = 1$). More formally,

$$\pi_0 = \lim_{n \to \infty} P(X_n = 0 | X_0 = 1)$$

- We first note that $\pi_0 = 1$ if $\mu < 1$ since

$$\mu^n = \mathrm{E}[X_n] = \sum_{j=1}^{\infty} jP(X_n = j) \geq \sum_{j=1}^{\infty} P(X_n = j) = P(X_n \geq 1)$$

Since $\mu_n \to 0$ when $\mu < 1$, it follows that $P\{X_n \geq 1\} \to 0$,and hence $P\{X_n = 0\} \to 1$.

- In fact, it can be shown that $\pi_0 = 1$ even when $\mu = 1$.

- When $\mu > 1$, it turns out that $\pi_0 < 1$, and an equation determining π_0 may be derived by conditioning on the number of offspring of the initial individual, as follows:

$$\pi_0 = P(\text{population dies out}) = \sum_{j=0}^{\infty} P(\text{population dies out}|X_1 = j)P_j = \pi_0^j P_j$$

  In fact when $\mu > 1$, it can be shown that π_0 is the smallest positive number satisfying the above equation.

Example 17. (a) If $P_0 = \frac{1}{2}, P_1 = \frac{1}{4}, P_2 = \frac{1}{4}$, then determine π_0. (b) If $P_0 = \frac{1}{4}, P_1 = \frac{1}{4}, P_2 = \frac{1}{2}$, then determine π_0. (c) For (a) and (b), what is the probability that the population will die out if it initially consists of n individuals? (Solution: (a) 1, (b) $\frac{1}{2}$, (c) for (a) 1, for (b) $\frac{1}{2^n}$)

9. Continuous-Time Markov Chains

Suppose we have a continuous-time stochastic process $\{X(t), t \geq 0\}$ taking on values in the set of nonnegative integers. we say that the process $\{X(t), t \geq 0\}$ is a *continuous-time Markov chain* if for all $s, t \geq 0$ and nonnegative integers $i, j, x(u), 0 \leq u \leq s$

$$P\{X(t+s) = j | X(s) = i, X(u) = x(u), 0 \leq u < s\} = P\{X(t+s) = j | X(s) = i\}$$

In addition,

$$P\{X(t+s) = j | X(s) = i\}$$

is independent of s, then the continuous-time Markov chain is said to have *stationary or homogeneous transition probabilities*.

Suppose that a continuous-time Markov chain enters state i at some time, say, time 0, and suppose that the process does not leave state i (that is, a transition does not occur) during the next ten minutes. What is the probability that the process will not leave state i during the following five minutes?

Since the process is in state i at time 10 it follows, by the Markovian property, that the probability that it remains in that state during the interval $[10, 15]$ is just the (unconditional) probability that it stays in state i for at least five minutes. That is, if we let T_i denote the amount of time that the process stays in state i before making a transition into a different state, then

$$P\{T_i > 15 | T_i > 10\} = P\{T_i > 5\} \text{ or } P\{T_i > s + t | T_i > s\} = P\{T_i > t\}$$

for all $s, t \geq 0$. Hence, the random variable T_i is *memoryless* and must thus be exponentially distributed $f(T_i) = \lambda e^{-\lambda t}, t \geq 0$ and else $f(T_i) = 0$.

10. The exponential distribution and the Poisson Process

A continuous random variable X is said to have an exponential distribution with parameter λ, $\lambda > 0$, if its probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

or, equivalently, if its cdf is given by

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- The mean of the exponential distribution, $\mathrm{E}[X]$, is given by

$$\mathrm{E}[x] = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

- The moment generating function $\phi(t)$ of the exponential distribution is given by

$$\phi(t) = \mathrm{E}[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t}, \text{for } t < \lambda$$

- Then

$$\mathrm{E}[X^2] = \frac{d^2}{dt^2}\phi(t)|_{t=0} = \frac{2\lambda}{(\lambda - t)^3}|_{t=0} = \frac{2}{\lambda^2}$$

and then

$$\mathsf{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

A random variable X is said to be without memory, or *memoryless* , if

$$P\{X > s + t | X > t\} = P\{X > s\}, \quad \text{for all} \quad s, t \geq 0$$

It follows that exponentially distributed random variables are memoryless with

$$P\{X > s + t\} = P(X > s)P(X > t)$$

Example 18. Suppose that the amount of time one spends in a bank is exponentially 1 distributed with mean ten minutes, that is, $\lambda = 10$. What is the probability that a customer will spend more than fifteen minutes in the bank? What is the probability that a customer will spend more than fifteen minutes in the bank given that she is still in the bank after ten minutes? (Solution: (1) 0.223 (2) 0.607)

A stochastic process $\{N(t), t \geq 0\}$ is said to be a *counting process* if $N(t)$ represents the total number of "events" that occur by time t. From its definition we see that for a counting process $N(t)$ must satisfy:

(i) $N(t) \geq 0$

(ii) $N(t)$ is integer valued.

(iii) If $s < t$ then $N(s) \geq N(t)$.

(iv) For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$

- A counting process is said to possess *independent increments* if the numbers of events that occur in disjoint time intervals are independent.

- A counting process is said to possess *stationary increments* if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval.

Definition 1. The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* having rate $\lambda > 0$ if

(i) $N(0) = 0$

(ii) The process has independent increments

(iii) The number of events in any interval of length t is Poisson distributed with mean λt. That is for all $s, t \geq 0$

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!},\ n = 0, 1, \ldots$$

It means that a Poisson Process has stationary increments and also that

$$\mathrm{E}N(t) = \lambda t$$

which explains why λ is called the rate of the process.

Definition 2. The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with rate $\lambda > 0$ if the following axioms hold:

(i) N(0)=0

(ii) $\{N(t), t \geq 0\}$ has independent increments

(iii) $P(N(t+h) - N(t) = 1) = \lambda h + o(h)$

(iv) $P(N(t+h) - N(t) \geq 2) = o(h)$

Definition 1 and Definition 2 are equivalent.

Example 19. Suppose that people immigrate into a territory according to a Poisson process with rate $\lambda = 2$ per day. (a) Find the probability there are 10 arrivals in the following week (of 7 days). (b) Find the expected number of days until there have been 20 arrivals. (Solution: (a) $e^{-14}(14)^{10}/10!$, (b) 10)

11. Birth and Death Processes

Consider a system whose state at any time is represented by the number of people in the system at that time. Suppose that whenever there are n people in the system, then

(i) new arrivals enter the system at an exponential rate λ_n, and

(ii) people leave the system at an exponential rate μ_n.

That is, whenever there are n persons in the system, then the time until the next arrival is exponentially distributed with mean $1/\lambda_n$ and is independent of the time until the next departure, which is itself exponentially distributed with mean $1/\mu_n$.

Such a system is called a *birth and death process* . The parameters $\{\lambda_n\}_{n=0}^{\infty}$ and $\{\mu_n\}_{n=0}^{\infty}$ are called, respectively, the arrival (or birth) and departure (or death) rates.

Thus, a birth and death process is a continuous-time Markov chain with states $\{0, 1, \ldots\}$ for which transitions from state n may go only to either state $n-1$ or state $n+1$. The relationships between the birth and death rates and the state transition rates and probabilities are

$$v_0 = \lambda_0, v_i = \lambda_i + \mu_i, P_{01} = 1, P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}, P_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}, i > 0$$

Example 20. (A Birth Process with Linear Birth Rate). Consider a population whose members can give birth to new members but cannot die. If each member acts independently of the others and takes an exponentially distributed amount of time, with mean $1/\lambda$, to give birth, then if $X(t)$ is the population size at time t, then $\{X(t), t \geq 0\}$ is a pure birth process with $\lambda_n = n\lambda, n \geq 0$. This follows since if the population consists of n persons and each gives birth at an exponential rate λ, then the total rate at which births occur is $n\lambda$. This pure birth process is known as a Yule process after G. Yule, who used it in his mathematical theory of evolution.

Example 21. (A Linear Growth Model with Immigration). A model in which

$$\begin{aligned} \mu &= n\mu, n \geq 1 \\ \lambda_n &= n\lambda + \theta, n \geq 0 \end{aligned}$$

is called a linear growth process with immigration. Such processes occur naturally in the study of biological reproduction and population growth. Each individual in the population is assumed to give birth at an exponential rate λ; in addition, there is an exponential rate of increase θ of the population due to an external source such as immigration. Hence, the total birth rate where there are n persons in the system is $n\lambda + \theta$. Deaths are assumed to occur at an exponential rate μ for each member of the population, so $\mu_n = n\mu$.

Let $X(t)$ denote the population size at time t. Suppose that $X(0) = i$ and let

$$M(t) = \mathrm{E}[X(t)]$$

We will determine $M(t)$ by deriving and then solving a differential equation that it satisfies.

Example 22. (The Queueing System M/M/1). Suppose that customers arrive at a single-server service station in accordance with a Poisson process having rate λ. That is, the times between successive arrivals are independent exponential random variables having mean $1/\lambda$. Upon arrival, each customer goes directly into service if the server is free; if not, then the customer joins the queue (that is, he waits in line). When the server finishes serving a customer, the customer leaves the system and the next customer in line, if there are any waiting, enters the service. The successive service times are assumed to be independent exponential random variables having mean $1/\mu$.

The preceding is known as the *M/M/1 queueing system*. The first M refers to the fact that the interarrival process is Markovian (since it is a Poisson process) and the second to the fact that the service distribution is exponential (and, hence, Markovian). The 1 refers to the fact that there is a single server.

If we let $X(t)$ denote the number in the system at time t then $\{X(t), t \geq 0\}$ is a birth and death process with $\mu_n = \mu, n \geq 1$ and $\lambda_n = \lambda, n \geq 0$.

Example 23. (A Multiserver Exponential Queueing System). Consider an exponential queueing system in which there are s servers available, each serving at rate μ. An entering customer first waits in line and then goes to the first free server. Assuming arrivals are according to a Poisson process having rate λ, this is a birth and death process with parameters

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq s \\ s\mu, & n > s \end{cases}$$

$$\lambda_n = \lambda, n \geq 0$$

To see why this is true, reason as follows: If there are n customers in the system, where $n \leq s$, then n servers will be busy. Since each of these servers works at rate μ, the total departure rate will be $n\mu$. On the other hand, if there are n customers in the system, where $n > s$, then all s of the servers will be busy, and thus the total departure rate will be $s\mu$. This is known as an *$M/M/s$ queueing model*.