1 Stochastic processes

In its most general expression, a stochastic process is simply a collection of random variables $\{X_t, t \in I\}$. The index t often represents time, and the set I is the index set of the process. The most common index set are $I = \{0, 1, 2, ...\}$ representing discrete time, and $I = [0, \infty)$ representing the continuous time. Discrete-time stochastic processes are sequences of random variables. Continuous-time stochastic processes are uncountable collections of random variables.

The random variables of a stochastic process take values in a common state space S, either discrete or continuous. A stochastic process is specified by its index and state spaces, and by the dependency relations among its random variables.

Example 1.1. (A discrete-time continuous-state stochastic process). An air-monitoring station in southern California records oxidant concentration levels every hour in order to monitor smog pollution. If it is assumed that hourly concentration levels are governed by some random mechanism, then the station's data can be considered a realization of a stochastic process $X_0, X_1, ...,$, where X_k is the oxidant concentration level at the k-th hour. The time variable is discrete. Since concentration levels take a continuum of values, the state space is continuous.

Example 1.2. (Continuous time, discrete state space) Dany receives text messages at random times day and night. Let X_t be the number of texts he receives up to time t. Then $\{X_t, t \in [0, \infty]\}$ is a continuous-time stochastic process with discrete state space $\{0, 1, 2, ...\}$.

If we observe the prices of a stock over a period of time, we can clearly see that the stock price goes up and down randomly and we cannot predict exactly the stock price at any future time. Thus the stock price S_t at each future time t can be considered as a random variable.

If we collect all these prices over a period of time $[0, T]$, we obtain a stochastic process for the stock price. If we only observe the closed price at the end of each trading day, then we have a discrete collection $(S_t)_{t=0,1,2,...}$, which is called a discrete-time stochastic process. If we observe all intraday prices (all prices during a trading day), we obtain a continuous collection $(S_t)_{0 \le t \le T}$, which is called a continuous-time stochastic process. In short, a stochastic process is a collection of random variables on the same probability space, representing the evolution of randomness over time.

To avoid multiple subscripts, we sometimes use $X(t)$ or X_t to denote the stochastic process $(X_t)_{t \ge 0}$. We are mainly interested in continuous-time stochastic process, with the index set $[0, \infty)$, representing for time-axis.

Definition. A stochastic process on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, denoted by $(X_t)_{t \in I}$, is a function of two variables with domain $I \times \Omega$ and range $\mathcal{S} \subset \mathbb{R}$ and is expressed as:

$$\begin{aligned} X : \quad & I \times \Omega \to \mathcal{S} \\ & (t, \omega) \to X(t, \omega) \end{aligned}$$

where I is the non-empty time index set and the range $\mathcal{S} \subset \mathbb{R}$ is the state space of the process. If I is a continuous set then $(X_t)_{t \in I}$ is called a continuous-time process; otherwise it is called a discrete-time process. Also if $\mathcal{S}$ is a discrete set, then $(X_t)_{t \in I}$ is called a discrete-state process; otherwise $(X_t)_{t \in I}$ is a continuous-state process.

Example 1. $(X_t)_{t \in I}$ is a discrete-time process if X_t represents the closing price of a stock on the t-trading day and $I = \{1, 2, \cdots\}$.

Example 2. $(X_t)_{t\in I}$ is a continuous-time process if X_t represents the intra-day price of a stock at time t and say $I = [9:30am, 3:00pm]$ on July 2017.

Example 3. $(X_t)_{t\in I}$ is a discrete-state process if X_t represents the total number of heads in the first t flips of a coin and $I = \{1, 2, \cdots\}$.

Example 4. $(X_t)_{t\in I}$ is a continuous-state process if X_t represents the stock price at time t, and is normally distributed with mean 0 and variance t^2.

Let $(X_t)_{t\geq 0}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. For each fixed t, we have

$$\begin{aligned} X_t: \qquad & \Omega \to \mathcal{S} \\ & \omega \to X_t(\omega) = X(t,\omega) \end{aligned}$$

Thus X_t is a random variable for each fixed time t.

On the contrary, for each fixed outcome $\omega \in \Omega$, we have

$$\begin{aligned} X(\omega): \qquad & I \to \mathcal{S} \\ & t \to X(t,\omega) \end{aligned}$$

Thus $X(\omega)$ is a deterministic function of time, which is called the sample path (realization, trajectory, sample function). For each fixed out come ω, a continuous sample path of the process $(X_t)_{t\geq 0}$ is defined in the ordinary calculus sense. That is $\lim\limits_{s\to t} X(s,\omega) = X(t,\omega)$ for each $t > 0$.

A stochastic process $(X_t)_{t\geq 0}$ is said to be almost surely continuous (or simply continuous) if almost surely all sample paths are continuous. That is,

$$\begin{aligned} X(\omega): \qquad & I \to \mathcal{S} \\ & t \to X(t,\omega) \end{aligned}$$

is a continuous sample path for almost surely (a.s) all $\omega \in \Omega$, which means that

$$P(\omega \in \Omega | X(\omega) \text{ is a continuous sample path}) = 1.$$

The term “Almost surely” here means the probability of that is 1.

2 Filtration: the evolution of information

Let $(S_t)_{t\geq 0}$ be a stock price process. In normal circumstances, the market price of a stock is the price at which a buyer and a seller agree to trade. Thus, the change of the market price of a stock is caused by the supply and demand, which reflect expectations of the company’s profitability. To figure out such expectations, one has to figure out how any news about the company will be interpreted by traders and investors. Therefore, we need a time-evolving information structure to study random process of a stock price. Another part of the motivation of establishing such a structure comes from the mathematical object itself: if the increment $S_{t+h} - S_t$ is assumed to be independent of t (to make full description of the process X mathematically manageable), then what we really say is that $S_{t+h} - S_t$ is independent information up to time t. This observation suggests we describe the evolution of information as information propagation over time. The mathematical concept that serves this purpose is called filtration.

A filtered probability space is a quadruple $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathcal{P})$ where $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space and $\{\mathcal{F}_t\}$ is a filtration. A filtration $\mathcal{F}_t$ is a non-decreasing collection of σ-algebras $\{\mathcal{F}_t \subset \mathcal{F}\}_{t\geq 0}$ such that $\mathcal{F}_s \subset \mathcal{F}_t, \ \forall 0 \leq s \leq t$.

Intuition. Filtration functions like a filter of information flow to control information propagation. For our purpose, it is sufficient to know the following:

- $\mathcal{F}_t$ represents the information available at time t.
- The condition $\mathcal{F}_s \subset \mathcal{F}_t, \forall 0 \leq s \leq t$ ensures that the amount of information grows as time evolves and that no information is lost with increasing time. It means that whatever information available at time s is still available at time t as long as $t \geq s$.

The σ-algebra $\mathcal{F}_t$ can be called as "information set".

Example. We consider an oversimplified stock price behavior described by a binomial tree with two periods $[t_0, t_1]$ and $[t_1, t_2]$ such that in each period, the stock price either goes up by a factor $u > 1$ with probability p or goes down by a factor $d < 1$ with probability $1 - p$. The corresponding probability space is represented by $(\Omega, \mathcal{F}, \mathcal{P})$, where: the sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, where $\omega_1 = UU, \omega_2 = UD, \omega_3 = DU, \omega_4 = DD$, the σ-algebra $\mathcal{F} = 2^\Omega$ containing all subsets of Ω.
Let $\mathcal{F}_{t_0} = \{\emptyset, \Omega\}, \mathcal{F}_{t_1} = \{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\}, \mathcal{F}_{t_2} = \mathcal{F}$. Then $(\mathcal{F}_t)_{t=0,1,2}$ is a filtration. Note that as t increases, information set $\mathcal{F}_t$ becomes finer and reveals more information about the evolution of stock price in the following sense:

- at $t = t_0$ we have no information available about stock price movements (experiment outcomes)
- at $t = t_1$ we know whether the stock will go up or go down, thus either the event $\{\omega_1, \omega_2\}$ or $\{\omega_3, \omega_4\}$ occur, but we still do not know which exactly outcome occurs.
- at $t = t_2$ we know exactly which outcome ω_i occurs.

The filtration $(\mathcal{F}_t)_{t\geq 0}$ is called a natural filtration of the process $(X_t)_{t\geq 0}$ if $\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t), t \geq 0$, i.e., $\mathcal{F}_t$ is the σ-algebra generated by random variables $X_s, 0 \leq s \leq t$.

3 Adapted processes

A stochastic process $(X_t)_{t\geq 0}$ defined on a filter probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathcal{P})$ is said to be adapted if X_t is $\mathcal{F}_t$-measurable for each $t \geq 0$, i.e., $\sigma(X_t) \in \mathcal{F}_t$ (X_t can be completely determined from the information available in $\mathcal{F}_t$).

For each $t > s$, X_t may not be $\mathcal{F}_s$-measurable, i.e., at time s, X_t is considered unknown (we cannot compute the probability of events described by X_t based on information available at any time earlier than t).

INTUITION. The notion of adaptedness can be interpreted as inability to have knowledge about future events. For this reason, an adapted process is also called non-anticipating because the propagation or progressive relation of information under adaptedness allows no anticipation of future information.

Example. A stochastic process is always adapted to its natural filtration.

INTUITION OF FILTER PROBABILITY SPACE. A filter probability space can be used to represent an economy, where each simple event $\omega \in \Omega$ represents an economic state, and the mathematical concept of adapted process $(X_t)_{t\geq 0}$ to the filtration can be used to represent the evolution of a security price and to ensure that events such as $\{X_t \leq a\} = \{\omega \in \Omega | X_t(\omega) \leq a\}$ are not anticipated at any time $s < t$. Also, note that the concept of filtration suggests that the economic state $\omega \in \Omega$ are not instantaneous states but represent entire possible histories of the economy.

4 Discrete Markov process

Consider a game with a playing board consisting of squares numbered 1-4 arranged in a circle. A player starts at square 1. At each turn, the player flip a coin and moves forward or backward if the coin lands heads (H) or tails (T), respectively. The player keeps moving forward or backward the board according to the result of the coin flip.

Let X_k be the position the player lands on after k moves, with $X_0 = 1$. Assume that the player successively flip the coin as H,H,T. The first four positions are then $(X_0, X_1, X_2, X_3) = (1, 2, 3, 2)$. Given this information, what can be said about the player's next position X_4?

Even though we know the player's full past history of moves, the only information relevant for predicting their future position is the their most recent location X_3. Since $X_3 = 2$, then necessarily $X_4 \in \{1, 3\}$, with equal probability. Formally,

$$P(X_4 = j | X_0 = 1, X_1 = 2, X_2 = 3, X_3 = 2) = P(X_4 = j | X_3 = 2) = \frac{1}{2},$$

for $j = 1, 3$. Given the player's most recent location X_3, their future position X_4 is independent of past history X_0, X_1, X_2.

The sequence of such positions $X_0, X_1, ...$ is a stochastic process called discrete Markov process (Markov chain). The game illustrates the essential property of a Markov chain: the future, given the present, is independent of the past.

Definition 4.1. Let S be a discrete set. A Markov chain is a sequence of random variables $X_0, X_1, ...$ taking values in S with the property that

$$P(X_{n+1} = j | X_0 = x_0, ..., X_{n-1} = x_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i),$$

for all $x_0, x_1, ..., x_{n-1}, i, j \in S$ and $n \geq 0$. The set S is the state space of the Markov chain.

We often use descriptive language to describe the evolution of a Markov chain. For instance, if $X_n = i$, we say that the chain visits state i, or hits i, at time n.

In this course, we usually consider time-homogeneous Markov chain, unless stated otherwise. A Markov chain is time-homogeneous if $P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) = P_{ij}$, for all $n \geq 0$. In other words, the one-step transition probabilities from a state i to a state j depend only on i and j, but not depend on when the transition occurs (i.e., not depend on n). The matrix P, with elements P_{ij}, is called transition matrix, or Markov matrix (a matrix with entries are the one-step transition probabilities from one stage to another one). If the state S has k elements, then the transition matrix is a square kxk matrix. If the state is countably infinite, the transition matrix is infinite.

Example 4.1. For simple board game Markov chain, the sample space is $S = \{1, 2, 3, 4, \}$, with the transition matrix $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{Vmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{Vmatrix} \end{array}$.

.

The entries of every Markov transition matrix P are nonnegative, and each row sums to 1. Such a matrix is called stochastic matrix. This is because from one state, the process will move to another accessible states.

Example 4.2. (Stochastic matrix.) A stochastic matrix is a square matrix P, which satisfies: all of its entries are nonnegative and each row sums to 1.

Example 4.3. (Chained to the weather.) Some winter days in Minnesota it seems like the snow will never stop. A Minnesotan's view of winter might be described by the following transition matrix for a weather Markov chain $\mathbf{P} = \begin{array}{c} \\ r \\ s \\ c \end{array}\begin{array}{c} \begin{array}{ccc} r & s & c \end{array} \\ \begin{Vmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{Vmatrix} \end{array}$, where r, s and c denote rain, snow and clear, respectively. For this model, no matter what the weather today, there is always at least 60% chance that it will snow tomorrow. In particular, if it snows today, then there is 80% it will continue snow tomorrow.

Example 4.4. (I.i.d sequence). An independent and identically distributed sequence of random variables is trivially a Markov chain. Assume $X_0, X_1, ...$ is an i.i.d sequence that takes values in $\{1, 2, ..., k\}$ with $P(X_n = j) = p_j$, for $j = 1, 2, ..., k$ and $n \geq 0$, where $p_1 + p_2 + ... + p_n = 1$. By independence, $P(X_1 = j|X_0 = i) = P(X_1 = j) = P(X_n = j) = p_j$. The transition matrix is $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & ... & k \end{array} \\ \begin{Vmatrix} p_1 & p_2 & ... & p_k \\ p_1 & p_2 & ... & p_k \\ p_1 & p_2 & ... & p_k \\ p_1 & p_2 & ... & p_k \end{Vmatrix} \end{array}$.

Example 4.5. (Gambler's ruin). In each round of a gambling game a player either win \$1, with probability p, or loses \$1, with probability $1 - p$. The gambler starts with k. The game stops when the player either loses all their money, or gains a total of n dollars $(n > k)$.

The gambler's successive fortunes form a Markov chain on $\{0, 1, ..., n\}$ with $X_0 = k$ and transition matrix given by:

$$P_{ij} = \begin{cases} p \quad \text{if} \quad j = i + 1, 0 < i < n, \\ 1 - p \quad \text{if} \quad j = i - 1, 0 < i < n, \\ 1 \quad \text{if} \quad i = j = 0, \text{or } i = j = n, \\ 0, \quad \text{otherwise}. \end{cases}$$

4.1 Basic computations

A powerful feature of Markov chains is the ability to use matrix algebra for computing probabilities. To use matrix methods, we consider probability distributions as vectors.

A probability vector is a row vector of non-negative numbers that sum to 1. Bold Greek letters α, γ, π are use to denote such vectors.

Assume that X is a discrete random variable with $P(X = j) = \alpha_j$, for $j = 1, 2, ...$ Then $\alpha = (\alpha_1, ..., \alpha_n)$ is a probability vector. We say that the distribution of X is α.

For a Markov chain $X_0, X_1, ...$, the distribution of X_0 is called the initial distribution of the chain. In particular, if $\alpha = (\alpha_1, ..., \alpha_j, ...)$ is the initial distribution then $P(X_0 = j) = \alpha_j$.

4.1.1 n-steps transition probabilities

For states i and j, $n \geq 1$, $P(X_n = j|X_0 = i)$ is the probability that the chain started at stage i and hits state j in n steps. The n-step transition matrix of the Markov chain includes ij-th entries $P(X_n = j|X_0 = i)$. In particular, if $n = 1$ then this is the usual (one-step) transition matrix P.

For $n \geq 1$, one of the central computational results of Markov chains is that n-step transition matrix is precisely P^n, the nth matrix power of the one-step transition matrix.

Theorem 1. Let $X_0, X_1, ...$ be a Markov chain with transition matrix P. The matrix P^n is the n-step transition matrix of the chain. For $n \geq 0$, $P^n_{ij} = P(X_n = j|X_0 = i), \quad \forall\, i, j$.

Important note 4.1. Do not confuse P^n_{ij}, the ijth entry of the matrix P^n, with $(P_{ij})^n$, the number P_{ij} raised to nth power. In particular P^0 is the identity matrix.

Example 4.6. For gambler's ruin, assume that the gambler's initial state is \$3 and the gambler plays until either gaining \$8 or going bust. At each play the gambler wins \$1, with probability 0.6, or loses \$1, with probability 0.4. Find the gambler's expected fortune after four plays.

Answer 4.1. Let X_k be the expected fortune after k plays. After four plays, the fortune of the player could be $X_4 = \{0, 1, 2, 3, 4, 5, 6, 7\}$. Thus the gambler's expected fortune after four play is

$$E(X_4|X_0 = 3) = \sum_{i=0}^{7} jP(X_4 = j|X_0 = 3) = \sum_{i=0}^{7} jP^4_{3j},$$

where P is the one-step transition matrix of the Markov chain player fortune. We need to determine the third row of P^4 in order to compute $E(X_4|X_0 = 3)$. The transition matrix P is

$$\mathbf{P} = \begin{array}{c} \\ 0\\1\\2\\3\\4\\5\\6\\7\\8 \end{array}\begin{array}{c} \begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \left\| \begin{array}{ccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0.4 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0.4 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0.4 & 0 & 0.6 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0.4 & 0 & 0.6 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0.4 & 0 & 0.6 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0.6 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0.6\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array} \right\| \end{array},$$

with $\mathbf{P}^4$ is

$$\mathbf{P^4} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{r} \begin{array}{rrrrrrrrr} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \left\| \begin{array}{rrrrrrrrr} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.496 & 0.115 & 0 & 0.259 & 0 & 0.130 & 0 & 0 & 0 \\ 0.237 & 0 & 0.288 & 0 & 0.346 & 0 & 0.130 & 0 & 0 \\ 0.064 & 0.115 & 0 & 0.346 & 0 & 0.346 & 0 & 0.130 & 0 \\ 0.026 & 0 & 0.154 & 0 & 0.346 & 0 & 0.346 & 0 & 0.130 \\ 0 & 0.026 & 0 & 0.154 & 0 & 0.346 & 0 & 0.259 & 0.216 \\ 0 & 0 & 0.026 & 0 & 0.154 & 0 & 0.288 & 0 & 0.533 \\ 0 & 0 & 0 & 0.026 & 0 & 0.115 & 0 & 0.115 & 0.744 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right\| \end{array}.$$

The gambler's expected fortune after four play is

$$E(X_4|X_0 = 3) = 0 * 0.064 + 1 * 0.115 + 2 * 0 + 3 * 0.346 + 4 * 0 + 5 * 0.346 + 6 * 0 + 7 * 0.13 = \$3.79.$$

4.1.2 Distribution of X_n

In general, a Markov chain $X_0, X_1, ...$ is not a sequence of identically distributed random variables. For $n \geq 1$, the marginal distribution of X_n depends on the n-step transition matrix P^n, as well as the initial distribution α. More precisely,

$$P(X_n = j) = \sum_i P(X_n = j|X_0 = i)P(X_0 = i) = \sum_i P^n_{ij}\alpha_i.$$

In other words, αP^n is the distribution of X_n and $P(X_n = j) = (\alpha P^n)_j$, the j-th element of αP^n, $\forall j$.

Example 4.7. Consider the weather chain Markov chain with transition matrix $\mathbf{P} = \begin{array}{c} \\ r \\ s \\ c \end{array} \begin{array}{c} \begin{array}{ccc} r & s & c \end{array} \\ \left\| \begin{array}{ccc} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{array} \right\| \end{array}$

where r, s and c denote rain, snow and clear, respectively. For tomorrow, the meteorologist predicts a 50% chance of snow and a 50% chance of rain. Find the probability that it will snow 2 days after tomorrow.

Answer 4.2. Denote X_k be the weather in k days later. As the ordered states of the chain are rain, snow and clear, the initial distribution is $\alpha = (0.5, 0.5, 0)$, i.e., $P(X_0 = r) = 0.5, P(X_0 = s) = 0.5, P(X_0 = c) = 0$. The desired distribution is of X_2, which is $\alpha\mathbf{P}^2 = (0.5, 0.5, 0) \begin{bmatrix} 0.12 & 0.72 & 0.16 \\ 0.11 & 0.76 & 0.13 \\ 0.11 & 0.72 & 0.17 \end{bmatrix} = (0.115, 0.74, 0.145)$. Thus, the probability that it will snow 2 days after tomorrow is 74%.

4.1.3 Joint distribution

The marginal distribution of X_n, of the Markov chain $(X_n)_{n\geq 1}$, is determined by the initial distribution α and the transition matrix P, as presented in the above section.

In addition α, P determine all the joint distributions of a Markov chain, that is, the joint distribution of any finite subset of $X_0, X_1, ...$ In that sense, the initial distribution and transition matrix give a complete probabilistic description of a Markov chain.

To illustrate, consider an arbitrary joint probability, such as $P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l)$, for some states i, j, k, l. For the underlying event, the chain moves to i in five steps, then to j in one step, then to k in three steps, and then to l in eight steps. Using the equality $P(A \cap B) = P(A|B)P(B)$, the time-homogeneity, the Markov property, we have:

$$\begin{aligned}
& P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l) = P(X_{17} = l | X_5 = i, X_6 = j, X_9 = k) P(X_5 = i, X_6 = j, X_9 = k) \\
= \;& P(X_{17} = l | X_9 = k) P(X_9 = k | X_5 = i, X_6 = j) P(X_5 = i, X_6 = j) = P^8_{kl} P(X_9 = k | X_6 = j) P(X_5 = i, X_6 = j) \\
= \;& P^8_{kl} P^3_{jk} P(X_6 = j | X_5 = i) P(X_5 = i) = P^8_{kl} P^3_{jk} P_{ij} (\alpha P^5)_i
\end{aligned}$$

Example 4.8. Danny's daily lunch choices are modeled by a Markov chain with trainsition matrix $\mathbf{P} = \begin{array}{c|cccc|} & \text{Burrito} & \text{Falafel} & \text{Pizza} & \text{Sushi} \\ \text{Burrito} & 0 & 0.5 & 0.5 & 0 \\ \text{Falafel} & 0.5 & 0 & 0.5 & 0 \\ \text{Pizza} & 0.4 & 0 & 0 & 0.6 \\ \text{Sushi} & 0 & 0.2 & 0.6 & 0.2 \end{array}$. On Sunday, Danny chooses lunch uniformly at random. Find the probability that he chooses sushi on the following Wednesday and Friday, and pizza on Saturday.

Answer 4.3. Consider Danny's lunch choices is a Markov chain $X_0, X_1, ...$ started on Sunday. The lunch choice of Danny on Wednesday, Friday, Saturday would be X_3, X_5, X_6 As on Sunday, Danny chooses lunch uniformly at random, the initial distribution is $\alpha = (1/4, 1/4, 1/4, 1/4)$. The required probability is $P(X_3 = s, X_5 = s, X_6 = p) = (\alpha P^3)_s P^2_{ss} P_{sp}$. We have $\mathbf{P}^2 = \begin{array}{c|cccc|} & \text{Burrito} & \text{Falafel} & \text{Pizza} & \text{Sushi} \\ \text{Burrito} & 0.45 & 0 & 0.25 & 0.3 \\ \text{Falafel} & 0.2 & 0.25 & 0.25 & 0.3 \\ \text{Pizza} & 0 & 0.32 & 0.56 & 0.12 \\ \text{Sushi} & 0.34 & 0.04 & 0.22 & 0.4 \end{array}$ and

$$\mathbf{P}^3 = \begin{array}{c|cccc|} & \text{Burrito} & \text{Falafel} & \text{Pizza} & \text{Sushi} \\ \text{Burrito} & 0.1 & 0.285 & 0.405 & 0.21 \\ \text{Falafel} & 0.225 & 0.16 & 0.405 & 0.21 \\ \text{Pizza} & 0.384 & 0.024 & 0.232 & 0.36 \\ \text{Sushi} & 0.108 & 0.25 & 0.43 & 0.212 \end{array}.$$

Thus the desired probability could be easily computed as 0.05952.

4.1.4 Exercise

1. A Markov chain has transition matrix $\mathbf{P} = \begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & 0.1 & 0.3 & 0.6 \\ 2 & 0 & 0.4 & 0.6 \\ 3 & 0.3 & 0.2 & 0.5 \end{array}$ with initial distribution $\alpha = (0.2, 0.3, 0.5)$.

   Note that $\mathbf{P}_{ij}$ is the probability the chain moves from initial position $X_0 = i$ to $X_1 = j$, i.e., $\mathbf{P}_{ij} = P(X_1 = j | X_0 = i)$. Find the following:
   a) $P(X_7 = 3 | X_6 = 2)$
   b) $P(X_9 = 2 | X_1 = 2, X_5 = 1, X_7 = 3)$
   c) $P(X_0 = 3 | X_1 = 1)$
   d) $E(X_2)$

2. Let $X_0, X_1, ...$ be a Markov chain with transition matrix $\mathbf{P} = \begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & 0 & 1/2 & 1/2 \\ 2 & 1 & 0 & 0 \\ 3 & 1/3 & 1/3 & 1/3 \end{array}$ and initial distribution

   $\alpha = (1/2, 0, 1/2)$. Find the following:

   a) $P(X_2 = 1|X_1 = 3)$

   b) $P(X_1 = 3, X_2 = 1)$

   c) $P(X_1 = 3|X_2 = 1)$

   d) $P(X_9 = 1|X_1 = 3, X_4 = 1, X_7 = 2)$.

3. For the general two-state chain with transition matrix $\mathbf{P} = \begin{array}{c|cc|} & a & b \\ a & 1-p & p \\ b & q & 1-q \end{array}$ and initial distribution

   $\alpha = (\alpha_1, \alpha_2)$, find the following:

   a) the two-step transition matrix

   b) the distribution of X_1.

4. Consider a random walk on $\{0, \dots, k\}$, which moves one-step left or right with respective probabilities p and $1-p$ with $p > 0$. If the walk is at 0 it transitions to 1 on the next step. If the walk is at k it transitions to $k-1$ on the next step. This is called random walk with reflecting boundaries. Assume that $k = 3, q = 1/4, p = 3/4$, and the initial distribution is uniform. For the following, use technology if needed.

   (a) Exhibit the transition matrix.
   (b) Find $P(X_7 = 1|X_0 = 3, X_2 = 2, X_4 = 2)$.
   (c) Find $P(X_3 = 1, X_5 = 3)$.

5. A fair tetrahedron die has four faces labeled $1, 2, 3$, and 4. In repeated independent rolls of the die $R_0, R_1, ...$, (R_i can only receive values in $1, 2, 3, 4$). Let $X_n = \max\{R_0, ..., R_n\}$ be the maximum value after $n+1$ rolls, for $n \geq 0$. For each roll, the chance of appearing each face is equally likely.

   (a) Give an intuitive argument for why $X_0, X_1, ...$, is a Markov chain, and exhibit the transition matrix.
   (b) Find $P(X_3 \geq 3)$.

6. Let $X_0, X_1, \dots$ be a Markov chain with transition matrix $\mathbf{P}$. Let $Y_n = X_{3n}$, for $n = 0, 1, 2, \dots$. Show that $Y_0, Y_1, ...$ is a Markov chain and exhibit its transition matrix.

7. * You start with five dice. Roll all the dice and put aside those dice that come up 6. Then, roll the remaining dice, putting aside those dice that come up 6. And so on. Let X_n be the number of dice that are sixes after n rolls.
   (a) Describe the transition matrix P for this Markov chain.

(b) Find the probability of getting all sixes by the third play.

(c) What do you expect P^{100} to look like? Use R to confirm your answer.

8. Given a Markov chain with the following transition matrix $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{Vmatrix} 0.5 & 0.25 & 0.25 \\ 1/3 & 0 & 2/3 \\ 0.5 & 0.5 & 0 \end{Vmatrix} \end{array}$, and $P(X_1 = 1) = P(X_1 = 2) = 0.25$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

4.2 Long-term behavior of a discrete Markov chain

In any stochastic or deterministic process, the long-term behavior of the system is often of interest. In particular, we are often interested in knowing the probability that the chain will hit a state j after a long run.

Definition 4.2. (Limiting distribution.) Let $X_0, X_1, ...$ be a Markov chain with transition matrix P. A limiting distribution for the Markov chain is a probability distribution λ with the property that for all i and j, $\lim\limits_{n\to\infty} P_{ij}^n = \lambda_j$ (for n large enough, the n-step transition probabilities from state i to state j do not depend on the starting position i). We interpret λ_j as the long-term probability that the chain hits state j.

Important note 4.2. The definition of limiting distribution is equivalent to each of the following:

- For any initial distribution, and for all j, $\lim\limits_{n\to\infty} P(X_n = j) = \lambda_j$. In words, vector $\lambda = (\lambda_j)$ is the distribution of X_n.
- For any initial distribution α, $\lim\limits_{n\to\infty} \alpha P^n = \lambda$.
- $\lim\limits_{n\to\infty} P^n = A$, where A is a stochastic matrix all of whose rows are equal to λ.

If a limiting distribution exists, a quick and dirty numerical method to find it is to take high matrix powers of the transition matrix until one obtains an obvious limiting matrix with equal rows. The common row is the limiting distribution.

Example 4.9. (Two-state Markov chain) The transition matrix for a general two-state chain is $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{Vmatrix} 1-p & p \\ q & 1-q \end{Vmatrix} \end{array}$, for $0 \le p, q \le 1$. Find the limiting distribution of the chain.

Answer 4.4. There are two cases needed to consider: $p + q = 1$ or $p + q \neq 1$.

1. Consider the case $p + q = 1$. The matrix P is then $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{Vmatrix} 1-p & p \\ 1-p & p \end{Vmatrix} \end{array}$. It is also easy to prove that $P^n = P$, for all $n \in \mathbb{N}$. Thus $\lim\limits_{n\to\infty} P^n = P$ and $\lambda = (1-p, p)$ is the limiting distribution for the Markov chain.

2. Consider the case $p + q \neq 1$. To compute $\mathbf{P}^n$, we first diagonalize the matrix.

   The characteristic polynomial of $\mathbf{P}$ is

$$|P - \lambda I| = \begin{vmatrix} 1-p-\lambda & p \\ q & 1-q-\lambda \end{vmatrix} = \lambda^2 - (2-p-q)\lambda - p - q + 1.$$

The polynomial has two roots: $\lambda_1 = 1, \lambda_2 = 1-p-q$. The two corresponding eigen vectors are $u_1 = (1,1)$ and $u_2 = (p,-q)$. Let $T = \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix}$, then the inverse matrix of T is $T^{-1} = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{1}{p+q} & \frac{-1}{p+q} \end{bmatrix}$. We now can express $\mathbf{P}$ as $\mathbf{P} = TDT^{-1}$, with $D = \begin{bmatrix} 1 & 0 \\ 0 & 1-p-q \end{bmatrix}$. As a result, we can easily compute $\mathbf{P}^n$ as follows:

$$\mathbf{P}^n = TD^nT^{-1} = \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{bmatrix} \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{1}{p+q} & \frac{-1}{p+q} \end{bmatrix} = \begin{bmatrix} \frac{q+p(1-p-q)^n}{p+q} & \frac{p(1-(1-p-q)^n)}{p+q} \\ \frac{q(1-(1-p-q)^n)}{p+q} & \frac{p+q(1-p-q)^n}{p+q} \end{bmatrix}.$$

As $0 \leq p, q \leq 1$ and $p+q \neq 1$, we deduce that $|p+q-1| < 1$. Thus $\lim\limits_{n\to\infty}(1-(1-p-q)^n) = 1$ and $\mathbf{P}^n = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix}$. The limiting distribution of the chain is $\alpha = (\frac{q}{p+q}, \frac{p}{p+q})$.

Example 4.10. Management of the New Fangled Soft-drink Company believes that the probability of a customer purchasing Red Pop or the company's major competition, Super Cola, is based on the customer's most recent purchase. Suppose that the following transition probabilities are appropriate: $\mathbf{P} = \begin{array}{c|cc|} & \text{Red Pop} & \text{Super Cola} \\ \text{Red Pop} & 0.9 & 0.1 \\ \text{Super Cola} & 0.1 & 0.9 \end{array}$.

a. Consider a customer who last purchased Red Pop. What is the probability that this customer purchases Red Pop on the second purchase?
b. Consider a customer who will buy Red Pop or Super Cola with probability 0.6 or 0.4, respectively. What is the probability of purchase Red Pop after 20th purchases?
c. What is the long-run market share for each of these two products?
d. A Red Pop advertising campaign is being planned to increase the probability of attracting Super Cola customers. Management believes that the new campaign will increase to 0.15 the probability of a customer switching from Super Cola to Red Pop. What is the projected effect of the advertising campaign on the market shares?

Answer 4.5. Let X_k denote the k-th purchase of the customer.
a. The required probability is $P(X_2 = \text{Red Pop}|X_0 = \text{Red Pop}) = \mathbf{P}^2_{11} = 0.82$
b. Given $\alpha = (0.6, 0.4)$ be the distribution of X_0. The required probability is $P(X_{20} = \text{Red Pop}) = (\alpha\mathbf{P}^{20})_1 = 0.501$
c. The long-run market share for each of these two products is the limiting distribution of the Markov chain, which is $\lambda = (\frac{0.1}{0.1+0.1}, \frac{0.1}{0.1+0.1}) = (0.5, 0.5)$.
d. The new probability transition matrix is $\mathbf{P} = \begin{array}{c|cc|} & \text{Red Pop} & \text{Super Cola} \\ \text{Red Pop} & 0.9 & 0.1 \\ \text{Super Cola} & 0.25 & 0.75 \end{array}$. The long-run market share is then $\lambda = (\frac{0.25}{0.1+0.25}, \frac{0.1}{0.1+0.25}) = (0.714, 0.286)$

Example 4.11. Consider a Markov chain with transition matrix $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Find the limiting distribution of the Markov chain.

Answer 4.6. We have $\mathbf{P}^n = \mathbf{P}$ and the rows of $\mathbf{P}$ are different. Thus, there is no limiting distribution of the chain.

Finding limiting distribution of a Markov chain become more difficult when the transition matrix has a high order (greater than 2). We can find it using the stationary distribution.

4.2.1 Stationary distribution

Definition 4.3. Let $X_0, X_1, ...$ be a Markov chain with transition matrix $\mathbf{P}$. A stationary distribution is a probability distribution π, which satisfies $\pi = \pi\mathbf{P}$.

Important note 4.3. If a stationary distribution π is the initial distribution, then all of the X_n have the same distribution π. In other words, $X_0, X_1, ...$ is a sequence of identically distributed random variables. The random variables in the sequence may not be independent and the dependency structure between successive random variables in a Markov chain is governed by the transition matrix, regardless of the initial distribution.

The name stationary comes from the fact that if the chain starts in its stationary distribution, then it stays in that distribution. We refer to the stationary Markov chain or the Markov chain in its stationarity for the chain started in its stationary distribution.

If $X_0, X_1, ...$ is a stationary Markov chain, then for any $n > 0$, the sequence $X_n, X_{n+1}, ...$ is also a stationary Markov chain with the same transition matrix and stationary distribution as the original chain.

Theorem 2. Assume π be the limiting distribution of a Markov chain with transition matrix $\mathbf{P}$. Then π is a stationary distribution.

Proof. Assume that π be the limiting distribution. For any initial distribution α, we have

$$\pi = \lim_{n\to\infty} \alpha\mathbf{P}^n = \lim_{n\to\infty} \alpha(\mathbf{P}^{n-1}\mathbf{P}) = \lim_{n\to\infty} (\alpha\mathbf{P}^{n-1})\mathbf{P} = \pi\mathbf{P},$$

which uses the fact that $\lim\limits_{n\to\infty} x_n = x$ then $\lim\limits_{n\to\infty} x_{n-1} = x$. □

Important note 4.4. Unfortunately, the converse of the Theorem is not true, i.e., stationary distributions are not necessary limiting distributions.

Example 4.12. Consider a Markov chain with transition matrix $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. To find the stationary distribution of the chain we solve the equation $\pi\mathbf{P} = \pi$ and obtain the unique solution $\pi = (0.5, 0.5)$. The stationary distribution is uniform on each state. However, the chain has no limiting distribution. The process evolves by flip-flopping back and forth between states.

Important note 4.5. There are Markov chains with more than one stationary distribution; there are Markov chains with unique stationary distribution that are not limiting distributions; and there are even Markov chains that do not have stationary distributions. However, a large and important class of Markov chains has unique stationary distributions that are limiting distribution of the chain. We now characterize such chains.

Definition 4.4. (Positive matrices.) A matrix M is said to be positive if all entries of M are positive. We write $M > 0$. Similarly, we write $x > 0$ for a vector x with all positive entries.

Definition 4.5. (Regular matrices.) A transition matrix $\mathbf{P}$ is said to be regular if some power of $\mathbf{P}$ is positive. That is $\mathbf{P}^n > 0$, for some $n \geq 1$.

Example 4.13. Matrix $\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ is regular since $\mathbf{P}^4 = \begin{bmatrix} 9/16 & 5/16 & 1/8 \\ 1/4 & 3/8 & 3/8 \\ 1/2 & 5/16 & 3/16 \end{bmatrix}$ is positive.

Theorem 3. (Limit theorem for regular Markov chains.) A Markov chain whose transition matrix **P** is regular has a limiting distribution, which is unique, positive, stationary distribution of the chain. That is, there exists a unique probability vector $\pi > 0$ such that $\lim_{n\to\infty} P_{ij}^n = \pi_j$ and $\pi\mathbf{P} = \pi$.

From the above theorem, we now have a procedure to compute the limiting distribution of a Markov chain.

1. Check whether **P** is a regular matrix.
2. Solve the linear equation system $\pi\mathbf{P} = \pi$.
3. Conclude the solution π is the limiting distribution.

Example 4.14. Consider the general two-state chain, with the transition probability matrix $\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$, $0 < p, q < 1$. Find the limiting distribution of the chain.

Answer 4.7. As $0 < p, q < 1$, **P** is a positive matrix, thus **P** is a regular matrix. There exists one unique limiting distribution π, which is the solution of the following equation $\pi\mathbf{P} = \pi$ or the equivalent system $\begin{cases} (1-p)\pi_1 + q\pi_2 = \pi_1 \\ p\pi_1 + (1-q)\pi_2 = \pi_2 \\ \pi_1 + \pi_2 = 1 \end{cases}$. It could be easily solved that $\pi = (\frac{q}{p+q}, \frac{p}{p+q})$.

Example 4.15. Find the stationary distribution of the weather Markov chain, with transition matrix $\mathbf{P} = \begin{array}{c|ccc} & r & s & c \\ \hline r & 0.2 & 0.6 & 0.2 \\ s & 0.1 & 0.8 & 0.1 \\ c & 0.1 & 0.6 & 0.3 \end{array}$.

Answer 4.8. As **P** is a positive matrix, thus it is a regular matrix. There exists one unique limiting distribution $\pi = (\pi_1, \pi_2, \pi_3)$, which is the solution of the following equation $\pi\mathbf{P} = \pi$ or the equivalent system $\begin{cases} 0.2\pi_1 + 0.1\pi_2 + 0.1\pi_3 = \pi_1 \\ 0.6\pi_1 + 0.8\pi_2 + 0.6\pi_3 = \pi_2 \\ 0.2\pi_1 + 0.1\pi_2 + 0.3\pi_3 = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$. It could be easily solved that $\pi = (\frac{1}{9}, \frac{3}{4}, \frac{5}{36})$.

R code for solving the linear equations:

```
##create matrix A
A <- matrix(c(-0.8,0.1,0.1,0.6,-0.2,0.6,0.2,0.1,-0.7,1,1,1),ncol=3,byrow=TRUE)
##create matrix b
b <- c(0,0,0,1)
##solve Ax=b
C <- qr.solve(A,b)
```

Example 4.16. After work, Angel goes to the gym and either does aerobics, weights, yoga, or gets a massage. Each day, Angel decides her workout routine based on what she did previous day according to the Markov transition matrix:$\mathbf{P} = \begin{array}{r|rrrr} & \text{Aerobics} & \text{Massage} & \text{Weights} & \text{Yoga} \\ \hline \text{Aerobics} & 0.1 & 0.2 & 0.4 & 0.3 \\ \text{Massage} & 0.4 & 0 & 0.4 & 0.2 \\ \text{Weights} & 0.3 & 0.3 & 0 & 0.4 \\ \text{Yoga} & 0.2 & 0.1 & 0.4 & 0.3 \end{array}$. Find the limiting distribution of the chain.

Answer 4.9. As $\mathbf{P}^2 = \begin{bmatrix} 0.27 & 0.17 & 0.24 & 0.32 \\ 0.2 & 0.22 & 0.24 & 0.34 \\ 0.23 & 0.1 & 0.4 & 0.27 \\ 0.24 & 0.19 & 0.24 & 0.33 \end{bmatrix}$ is a positive matrix, $\mathbf{P}$ is a regular matrix. Thus, the limiting distribution of the Markov chain is also the stationary distribution π of the chain.

The stationary distribution $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ of the chain is a solution of the equation $\pi\mathbf{P} = \pi$ or

$$\begin{cases} 0.1\pi_1 + 0.4\pi_2 + 0.3\pi_3 + 0.2\pi_4 = \pi_1 \\ 0.2\pi_1 + 0\pi_2 + 0.3\pi_3 + 0.1\pi_4 = \pi_2 \\ 0.4\pi_1 + 0.4\pi_2 + 0\pi_3 + 0.4\pi_4 = \pi_3 \\ 0.3\pi_1 + 0.2\pi_2 + 0.4\pi_3 + 0.3\pi_4 = \pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases}$$

R code for solving the system

```
A <- matrix(c(-0.9,0.2,0.4,0.3,1,0.4,-1,0.4,0.2,1,0.3,0.3,-1,0.4,1,0.2,0.1,0.4,-0.7,1),nrow=5)
 b <- c(0,0,0,0,1)
 C <- qr.solve(A,b)
```

The result is

$\pi =$ $\begin{array}{cccc} \text{Aerobics} & \text{Massage} & \text{Weights} & \text{Yoga} \\ 0.2377 & 0.164 & 0.2857 & 0.312 \end{array}$. After a long run, about 31.2% of time, she does Yoga.

4.3 Exercise

1. Consider a Markov chain with transition matrix $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Find the limiting distribution of the Markov chain.

2. Prove that the limiting distribution of the Markov chain having a transition probability matrix $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{ccc} 1/3 & 1/2 & 1/6 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \end{array} \right\| \end{array}$ is $\pi = (9/22, 11/22, 2/22)$.

3. Consider a Markov chain with transition matrix $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \left\| \begin{array}{cccc} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \end{array} \right\| \end{array}$. Find the stationary distribution of the chain.

4. * Find the stationary distribution of a Markov chain having transition matrix $\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left\| \begin{array}{ccc} 1-a & a & 0 \\ 0 & 1-b & b \\ c & 0 & 1-c \end{array} \right\| \end{array}$, with $0 < a, b, c < 1$.

5. * Assume that a Markov chain has transition matrix $\mathbf{P} = \begin{array}{c|ccc|} & 1 & 2 & 3 \\ 1 & 0 & 1-p & p \\ 2 & p & 0 & 1-p \\ 3 & 1-p & p & 0 \end{array}$. Find the limiting distribution.

6. * Prove that the limiting distribution of the Markov chain having a transition probability matrix $\mathbf{P} = \begin{array}{c|cccc|} & 1 & 2 & 3 & 4 \\ 1 & p & 1-p & 0 & 0 \\ 2 & (1-p)/2 & p & (1-p)/2 & 0 \\ 3 & 0 & (1-p)/2 & p & (1-p)/2 \\ 4 & 0 & 0 & 1-p & p \end{array}$ is $\pi = (1/6, 2/6, 2/6, 1/6)$.