Markov Chains

ISE/OR 560 Fall 2022

Hong Wan

September 28, 2022

North Carolina State University

NC STATE UNIVERSITY

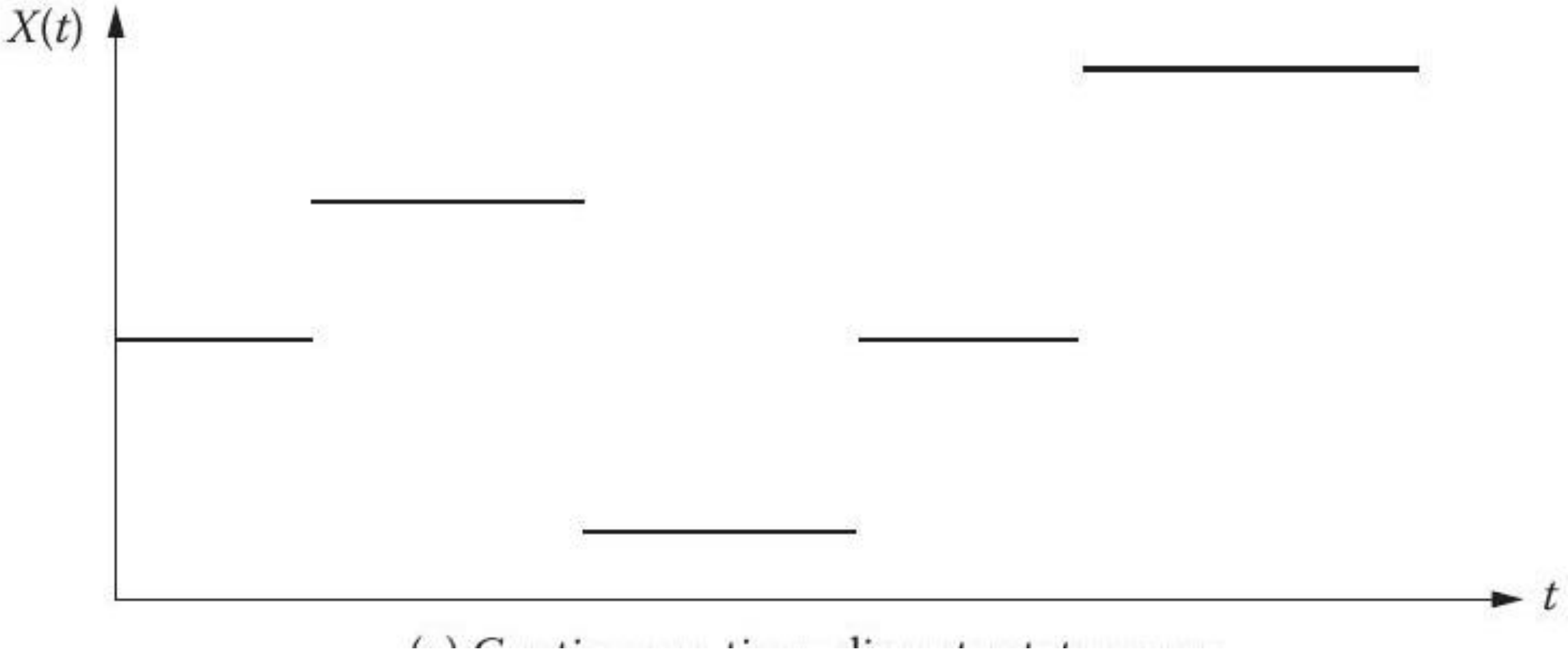
Stochastic System

System that evolves randomly in time..

A stochastic process is a collection of random variables $\{X(\tau), \tau \in T\}$, indexed by the parameter τ taking values in the parameter set T. The random variables take values in the set S, called the **state-space** of the stochastic process.

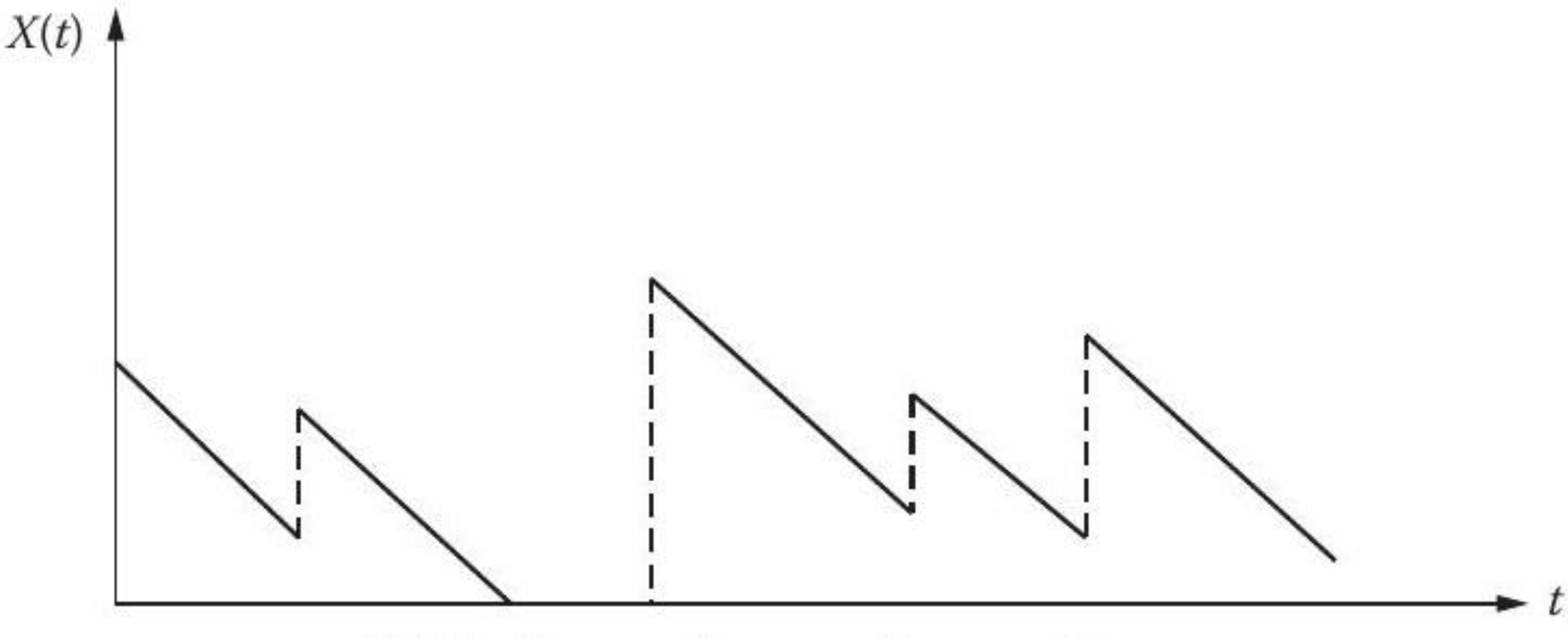
- Suppose we observe this system at discrete time points $n = 0, 1, 2, \cdots$. Let X_n be the state of the system at time n. We say that $\{X_n, n \geq 0\}$ is a *discrete-time stochastic process*
- If the system is observed continuously in time, with $X(t)$ being its state at time t, then $\{X(t), t \geq 0\}$ is described by a continuous time stochastic process.

Sample Path



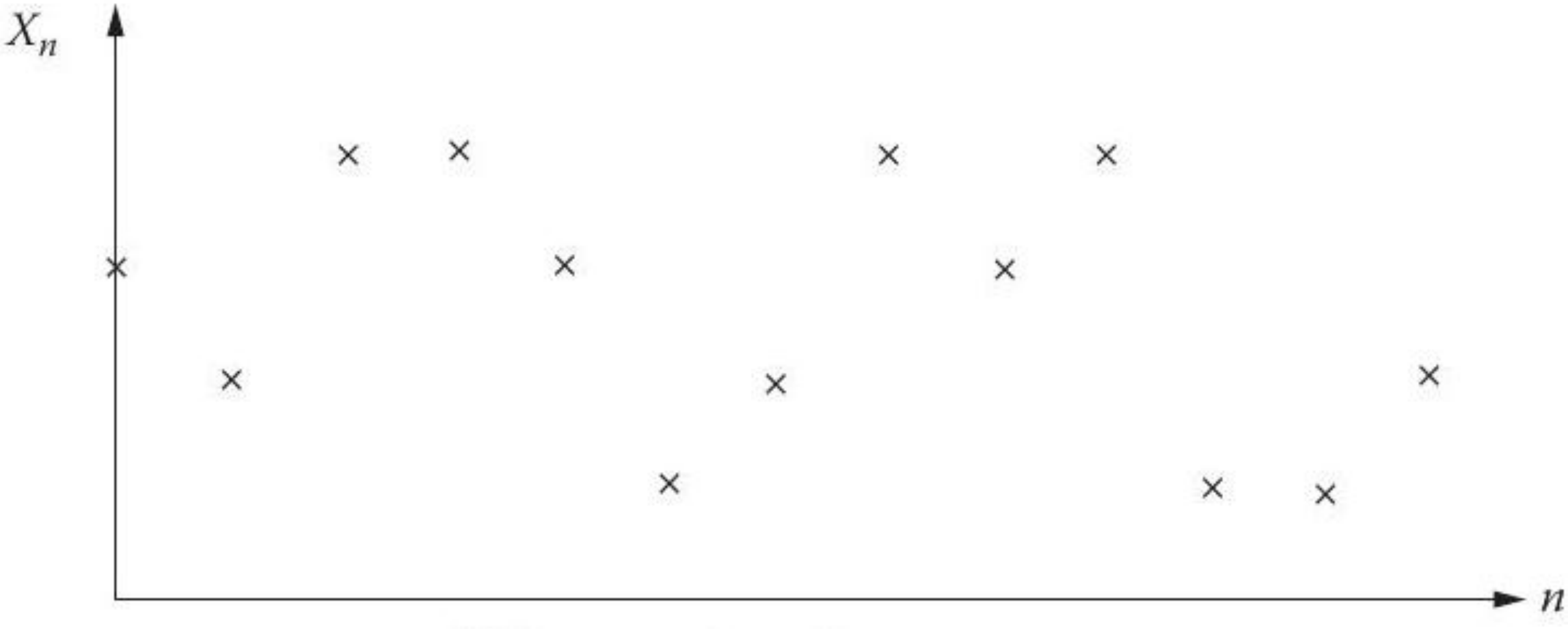
(a) Continuous-time, discrete state-space

Sample Path



(b) Continuous-time, continuous state-space

Sample Path



(c) Discrete-time, discrete state-space

Examples

- Queues: Let $X(t)$ be the number of customers waiting for service in a service facility such as an outpatient clinic.
- Supply Chains. Consider a supply chain of computer printers with three levels: the manufacturer (level 1), the regional warehouse (level 2), and the retail store (level 3). The printers are stored at all three levels. Let $X_i(t) =$ the number of printers at level i, $(1 \leq i \leq 3)$.
- A random walk is an example of a stochastic process in discrete time, where a particle starts at the origin at time 0 and moves one distance left with probability p or one distance unit right with probability $1p$ at each time unit.

CDF Characterization of Stochastic Process

- A single random variable X is completely described by its cumulative distribution function (cdf)

$$F(x) = \mathrm{P}(X \leq x), \quad -\infty < x < \infty.$$

- A multivariate random variable $(X_1, X_2, \cdots, X_n)$ is completely described by its joint cdf

$$F(x_1, x_2, \cdots, x_n) = \mathrm{P}(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n),$$

  for all $-\infty < x_i < \infty$ and $i = 1, 2, \cdots, n$. If the parameter set T is finite, the stochastic process $\{X(\tau), \tau \in T\}$ is a multivariate random variable, and hence is completely described by the joint cdf.

- what about the case when T is not finite?

Suppose the sample paths of $\{X(t), t \geq 0\}$ are, with probability 1 , right continuous with left limits, i.e.,

$$\lim_{s \downarrow t} X(s) = X(t),$$

and $\lim_{s \uparrow t} X(s)$ exists for each t. Furthermore, suppose the sample paths have a finite number of discontinuities in a finite interval of time with probability one. Then $\{X(t), t \geq 0\}$ is completely described by a consistent family of finite dimensional joint cdfs

$$F_{t_1, t_2, \cdots, t_n}(x_1, x_2, \cdots, x_n) = P\left(X\left(t_1\right) \leq x_1, X\left(t_2\right) \leq x_2, \cdots, X\left(t_n\right) \leq x_n\right),$$

for all $-\infty < x_i < \infty, i = 1, \cdots, n, n \geq 1$ and all $0 \leq t_1 < t_2 < \cdots < t_n$.

Random Walk

Let $\{X_n, n \geq 1\}$ be be a sequence of independent and identically distributed (iid) random variables with common distribution $F(\cdot)$. Define

$$S_0 = 0, S_n = X_1 + \cdots + X_n, \quad n \geq 1.$$

The stochastic process $\{S_n, n \geq 0\}$ is called a random walk. It is also completely characterized by $F(\cdot)$, since the joint distribution of $(S_0, S_1, \cdots, S_n)$ is completely determined by that of $(X_1, X_2, \cdots, X_n)$, which, is determined by $F(\cdot)$ that is defined as

$$F_n(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i),$$

for all $-\infty < x_i < \infty, i = 1, 2, \cdots, n$, and $n \geq 1$.

Consider a system that is modeled by a discrete-time stochastic process $\{X_n, n \geq 0\}$ with a countable state-space S, say $\{0, 1, 2, \cdots\}$. Consider a fixed value of n that we shall call **"the present time"** or just the "present." Then X_n is called **the present (state) of the system**, $\{X_0, X_1, \cdots, X_{n-1}\}$ is called **the past of the system**, and $\{X_{n+1}, X_{n+2}, \cdots\}$ is called **the future of the system**. If $X_n = i$ and $X_{n+1} = j$, we say that the system has **jumped (or made a transition)** from state i to state j from time n to $n + 1$.

Markov Properties

- If the present state of the system is known, the future of the system is independent of its past.
- Intuition: the past affects the future only through the present.

Definition of Discrete-Time Markov Chain: A stochastic process $\{X_n, n \geq 0\}$ with countable state-space S is called a DTMC if

1. for all $n \geq 0, X_n \in S$,
2. for all $n \geq 0$, and $i, j \in S$

Time-Homogeneous DTMC

Time-Homogeneous DTMC. A DTMC $\{X_n, n \geq 0\}$ with countable state-space S is said to be time-homogeneous if

Transition Probability

Let

$$P = [p_{i,j}]$$

denote the matrix of the conditional probabilities $p_{i,j}$. We call $p_{i,j}$ the transition probability from state i to state j. The matrix P is called the one-step transition probability matrix or just the transition probability matrix. When S is finite, say $S = \{1, 2, \cdots, m\}$, one can display P as a matrix as follows:

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,m-1} & p_{1,m} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,m-1} & p_{2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{m-1,1} & p_{m-1,2} & \cdots & p_{m-1,m-1} & p_{m-1,m} \\ p_{m,1} & p_{m,2} & \cdots & p_{m,m-1} & p_{m,m} \end{bmatrix}$$

Frame Title

Stochastic Matrix. A square matrix $P = \left[p_{i,j}\right]$ is called stochastic if

$$\text{(i).} \quad p_{i,j} \geq 0 \quad \text{for all } i, j \in S,$$

$$\text{(ii).} \quad \sum_{j \in S} p_{i,j} = 1 \quad \text{for all } i \in S.$$

Stochastic Matrix

Stochastic Matrix. A square matrix $P = \left[p_{i,j}\right]$ is called stochastic if

$$
\begin{aligned}
&\text{(i).} \quad && p_{i,j} \geq 0 \quad \text{for all } i, j \in S, \\
&\text{(ii).} \quad && \sum_{j \in S} p_{i,j} = 1 \quad \text{for all } i \in S.
\end{aligned}
$$

We can prove that the one-step transition probability matrix of a DTMC is stochastic.

Initial Distribution of the DTMC

Suppose we specify the distribution of X_0 externally. Let

$$a_i = \mathrm{P}\left(X_0 = i\right), \quad i \in S,$$

and

$$a = [a_i]_{i \in S}$$

be a row vector representing the probability mass function (pmf) of X_0. We say that a is the initial distribution of the DTMC.

Theorem: A DTMC $\{X_n, n \geq 0\}$ is completely described by its initial distribution a and the transition probability matrix P.

Transition Diagram

Transition Diagram. Consider a DTMC $\{X_n, n \geq 0\}$ on state-space $\{1, 2, 3\}$ with the following transition probability matrix:

$$P = \begin{bmatrix} .1 & .2 & .7 \\ .6 & 0 & .4 \\ .4 & 0 & .6 \end{bmatrix}$$

Transition Diagram

Probability Calculation

Joint Distributions. Let $\{X_n, n \geq 0\}$ be a DTMC on state-space $\{1, 2, 3, 4\}$ with the transition probability matrix given below:

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.6 & 0.2 & 0.1 & 0.1 \end{bmatrix}$$

The initial distribution is

$$a = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}.$$

1. Compute $\mathrm{P}(X_3 = 4, X_2 = 1, X_1 = 3, X_0 = 1)$.
2. Compute $\mathrm{P}(X_3 = 4, X_2 = 1, X_1 = 3)$.

Probability Calculation

Probability Calculation

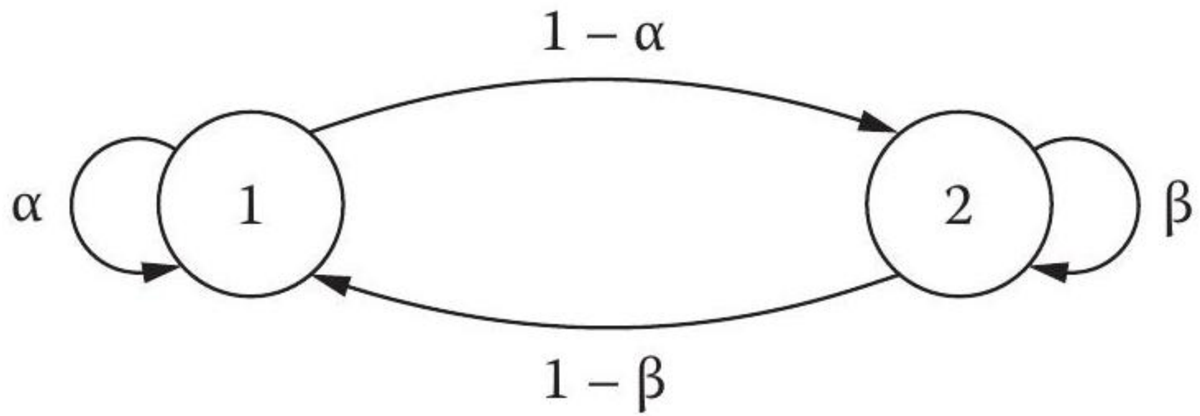
Two-State DTMC

Two-State DTMC. One of the simplest DTMCs is one with two states, labeled 1 and 2. Thus $S = \{1, 2\}$. Such a DTMC has a transition matrix as follows:

$$\begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix},$$

where $0 \le \alpha, \beta \le 1$.

The transition diagram:



Weather Model

Consider a simple weather 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.

Weather Model

Clinical Trials:

Suppose two drugs are available to treat a particular disease, and we need to determine which of the two drugs is more effective. This is generally accomplished by conducting clinical trials of the two drugs on actual patients. Here we describe a clinical trial setup that is useful if the response of a patient to the administered drug is sufficiently quick, and can be classified as "effective" or "ineffective." Suppose drug i is effective with probability $p_i, i = 1, 2$. In practice the values of p_1 and p_2 are unknown, and the aim is to determine if $p_1 \geq p_2$ or $p_2 \geq p_1$. Ethical reasons compel us to use the better drug on more patients. This is achieved by using the play the winner rule as follows.

The initial patient (indexed as patient 0) is given either drug 1 or 2 at random. If the nth patient is given drug $i (i = 1, 2)$ and it is observed to be effective for that patient, then the same drug is given to the $(n + 1)$-st patient; if it is observed to be ineffective then the $(n + 1)$-st patient is given the other drug. Thus we stick with a drug as long as its results are good; when we get a bad result, we switch to the other drug - hence the name ”play the winner.”

Clinical Trials

Gambler's Ruin: Consider two gamblers, A and B, who have a combined fortune of N dollars. They bet one dollar each on the toss of a coin. If the coin turns up heads, A wins a dollar from B, and if the coin turns up tails, B wins a dollar from A. Suppose the successive coin tosses are independent, and the coin turns up heads with probability p and tails with probability $q = 1 - p$. The game ends when either A or B is broke (or ruined).

Urn Model

Consider two urns labeled A and B, containing a total of N white balls and N red balls among them. An experiment consists of picking one ball at random from each urn and interchanging them. This experiment is repeated in an independent fashion. Let X_n be the number of white balls in urn A after n repetitions of the experiment. Assume that initially urn A contains all the white balls, and urn B contains all the red balls. Thus $X_0 = N$. Note that X_n tells us precisely the contents of the two urns after n experiments: if $X_n = i$, urn A contains i white balls and $N - i$ red balls; and urn B contains $N - i$ white balls and i red balls.

Random Walk Revisited: State-Dependency

Consider a particle that moves randomly on a doubly infinite one-dimensional lattice where the lattice points are labeled $\cdots, -2, -1, 0, 1, 2, \cdots$. The particle moves by taking steps of size 0 or 1 or -1 as follows: if it is on site i at time n, then at time $n+1$ it moves to site $i+1$ with probability p_i, or to site $i-1$ with probability q_i, or stays at site i with probability $r_i = 1 - p_i - q_i$, independent of its motion up to time n. Let X_n be the position of the particle (the label of the site occupied by the particle) at time n. Thus $\{X_n, n \geq 0\}$ is a DTMC on state-space $S = \{0, \pm 1, \pm 2, \cdots\}$ with transition probabilities given by

$$p_{i,i+1} = p_i, \quad p_{i,i-1} = q_i, \quad p_{i,i} = r_i, \quad i \in S.$$

Random Walk Revisited

This random walk is NOT space-homogeneous, since the step size distribution depends upon where the particle is.

Special Case 1: When $r_i = 0, \quad p_i = p, \quad q_i = q$ for all $i \in S$, the random walk is space-homogeneous, and we call it a **simple random walk**.

Special Case 2: For a simple random walk, if $p = q = 1/2$, we call it a **simple symmetric random walk**.

Success Runs

Consider a game where a coin is tossed repeatedly in an independent fashion. Whenever the coin turns up heads, which happens with probability p, the player wins a dollar. Whenever the coin turns up tails, which happens with probability $q = 1 - p$, the player loses all his winnings so far. Let X_n denote the player's fortune after the n-th toss. We have

$$\left(X_{n+1} = \begin{cases} 0 & \text{with probability } q \\ X_n + 1 & \text{with probability } p. \end{cases} \right.$$

How do I show $P(X_{n+1} \mid X_n) = P(X_{n+1} \mid X_n, X_{n-1}, \ldots, X_0)$

Finance: Stock Fluctuation

Let X_n be the value of a stock at time n (this could be a day, or a minute). We assume that $\{X_n, n \geq 0\}$ is a stochastic process with state-space $(0, \infty)$, not necessarily discrete, although stock values are reported in integer cents. Define the return in period n as

$$R_n = \frac{X_n - X_{n-1}}{X_{n-1}}, \quad n \geq 1.$$

Thus $R_3 = .1$ implies that the stock value increased by 10% from period two to three, $R_2 = -.05$ is equivalent to saying that the stock value decreased by 5% from period one to two. From this definition it follows that

$$X_n = X_0 \prod_{i=1}^{n} (1 + R_i), \quad n \geq 1$$

Marginal Distribution

Let $\{X_n, n \geq 0\}$ be a DTMC on state-space $S = \{0, 1, 2, \cdots\}$ with transition probability matrix P and initial distribution a. In this section we shall study the distribution of X_n. Let the pmf of X_n be denoted by

$$a_j^{(n)} = \mathrm{P}\left(X_n = j\right), \quad j \in S, n \geq 0.$$

Clearly $a_j^{(0)} = a_j$ is the initial distribution. By using the law of total probability we get

why X_0?

$$\begin{aligned} \mathrm{P}\left(X_n = j\right) &= \sum_{i \in S} \mathrm{P}\left(X_n = j \mid X_0 = i\right) \mathrm{P}\left(X_0 = i\right) \\ &= \sum_{i \in S} \mathrm{P}\left(X_n = j \mid X_0 = i\right) a_i \\ &= \sum_{i \in S} a_i p_{i,j}^{(n)} \end{aligned}$$

where

$$p_{i,j}^{(n)} = \mathrm{P}\left(X_n = j \mid X_0 = i\right), \quad i, j \in S, n \geq 0.$$

It is called the n-step transition probability, since it is the probability

$$a_j^{(n)} = \mathrm{P}\left(X_n = j\right), \quad j \in S, n \geq 0.$$

is called the n-step transition probability, since it is the probability of going from state i to state j in n transitions. We have

$$p_{i,j}^{(0)} = \mathrm{P}\left(X_0 = j \mid X_0 = i\right) = \delta_{i,j}, \quad i, j \in S,$$

where $\delta_{i,j}$ is one if $i = j$ and zero otherwise, and

$$p_{i,j}^{(1)} = \mathrm{P}\left(X_1 = j \mid X_0 = i\right) = p_{i,j}, \quad i, j \in S.$$

Chapman-Kolmogorov Equation

Theorem: the n-step transition probabilities satisfy the following equations:

$$p_{i,j}^{(n)} = \sum_{r \in S} p_{i,r}^{(k)} p_{r,j}^{(n-k)}, \quad i, j \in S,$$

where k is a fixed integer such that $0 \leq k \leq n$.

Matrix Format

Define the n-step transition probability matrix as

$$P^{(n)} = \left[p_{i,j}^{(n)}\right].$$

The Chapman-Kolmogorov equations can be written in matrix form as

$$P^{(n)} = P^{(k)} P^{(n-k)}, \quad 0 \leq k \leq n.$$

Implication

The n-Step Transition Probability Matrix.

$$P^{(n)} = P^n,$$

where P^n is the n-th power of P.

Probability Mass Function of X_n.

$$a^{(n)} = aP^n, \quad n \geq 0$$

Let $\{X_n, n \geq 0\}$ be the DTMC with the following initial distribution and transition matrix. Compute the pmf of X_4.

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.6 & 0.2 & 0.1 & 0.1 \end{bmatrix}$$

The initial distribution is

$$a = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}.$$

You can use the following website for matrix calculation.
https://matrix.reshish.com/powCalculation.php

Occupancy Times

Let $\{X_n, n \geq 0\}$ be a DTMC on state-space S with transition probability matrix P. In this section we compute the expected time spent by the DTMC in various states. Let $V_j^{(n)}$ be the number of visits to state j by the DTMC over $\{0, 1, 2, \cdots, n\}$. Note that we count the visit at time 0 , that is, $V_j^{(0)} = 1$ if $X_0 = j$, and zero otherwise. Define

$$M_{i,j}^{(n)} = \mathrm{E}\left(V_j^{(n)} \mid X_0 = i\right), \quad i, j \in S, n \geq 0.$$

$M_{i,j}^{(n)}$ is called the occupancy time of state j up to time n starting from state i. Define the occupancy times matrix as

$$M^{(n)} = \left[M_{i,j}^{(n)}\right].$$

Occupancy Times

$$M^{(n)} = \sum_{r=0}^{n} P^r, \quad n \geq 0,$$

where $P^0 = I$, the identity matrix .

For the previous example, can we calculate $M^{(4)}$?

Two-State DTMC

Consider the two-state DTMC we discussed before. Assume $\alpha + \beta < 2$. The n-step transition probability matrix of the DTMC was given as

$$\begin{bmatrix} 1-\beta & 1-\alpha \\ 1-\beta & 1-\alpha \end{bmatrix} + \frac{1-(\alpha+\beta-1)^{(n+1)}}{(2-\alpha-\beta)^2} \begin{bmatrix} 1-\alpha & \alpha-1 \\ \beta-1 & 1-\beta \end{bmatrix}$$

Using that, and a bit of algebra, we see that the occupancy matrix for the two-state DTMC is given by

→ same dimension as prob. trans matrix.

$$M^{(n)} = \frac{n+1}{2-\alpha-\beta} \begin{bmatrix} 1-\beta & 1-\alpha \\ 1-\beta & 1-\alpha \end{bmatrix} + \frac{1-(\alpha+\beta-1)^{(n+1)}}{(2-\alpha-\beta)^2} \begin{bmatrix} 1-\alpha & \alpha-1 \\ \beta-1 & 1-\beta \end{bmatrix}$$

Thus, if the DTMC starts in state 1 , the expected number of times it visits state 2 up to time n is given by $M^{(n)}_{1,2}$. → Expected # of times visited state 2

In practice, it is rarely possible to compute the occupancy matrix analytically.