Introduction to Markov Chains
(Chapter 9)

Michael Tsiang

Stats 102C: Introduction to Monte Carlo Methods

UCLA

Acknowledgements: Qing Zhou

Outline

Introduction

Our main goal in using Monte Carlo methods thus far has been to simulate iid random variables.

We usually are interested in a target distribution $f(x)$.

- To generate samples from $f(x)$:
  - The inverse CDF method
  - Rejection sampling
- To estimate $E_f[h(X)]$:
  - Importance sampling

Introduction

- For complicated or high-dimensional distributions, generating iid samples may be extremely inefficient.

- **Markov Chain Monte Carlo (MCMC)** methods use Markov chains to simulate *correlated* samples that are (approximately) from a target distribution.

- MCMC methods were developed for Bayesian statistics, where posterior densities can have complicated forms and are not easy to sample from directly.

- Quantities of interest, such as the posterior mean or posterior variance, may also be difficult to compute analytically.

Outline

Definitions

relationship of stochastic process.

Definition

A **(discrete-time) Markov chain** $\{X_t : t = 0, 1, 2, \dots\}$ is a stochastic process (i.e., a sequence of random variables) that satisfies the **Markov property**: only depend on last sequence

$$P(X_{n+1} = j | \underbrace{X_0 = i_0, \dots, X_{n-1} = i_{n-1},}_{} X_n = i) = P(X_{n+1} = j | X_n = i)$$

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow \cdots \longrightarrow \boxed{X_n} \longrightarrow X_{n+1}$$

chain

Definition

The **state space** of a Markov chain is the collection of all possible values for $X_0, X_1, \dots$.

Definitions

- The Markov property means that the probability that the chain moves to state j on the next step only depends on the current state i, not on where the chain has been previously.
- We will only consider Markov chains with countable or finite state spaces (i.e., discrete-state discrete-time Markov chains).
- Without loss of generality, we can write the state space as $\{0, 1, 2, \ldots\}$ (if countable) or $\{0, 1, 2, \ldots, N\}$ (if finite).

Definitions

$$P_{ij} = \Big(\overset{\text{future}}{\underset{n+1}{P_j}} \;\Big|\; \overset{\text{current}}{\underset{n}{P_i}}\Big)$$

$i\ j \in S$

(time-homogenous)

Definition

The **(one-step) transition probabilities** for a Markov chain $\{X_t : t = 0, 1, 2, \ldots\}$ are defined as the conditional probabilities

$$P_{ij} := P(X_{n+1} = j | X_n = i),$$

for all n and all states i and j in the state space.

Definitions

It is often convenient to describe the transition probabilities by a **transition matrix**

P_{ij}

$$\begin{array}{cl}
\mathbb{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}, & \begin{array}{l} \rightarrow \text{Sum} = 1 \\ \rightarrow \text{Sum} = 1 \\ \\ \\ \\ \end{array}
\end{array}$$

where properties ① ②

$n \times n$

1. All the entries are non-negative: (≥ 0)

$$P_{ij} \geq 0, \text{ for all } i, j.$$

2. The sum of each row is 1:

$$\sum_{j=0}^{\infty} P_{ij} = \sum_{j=0}^{\infty} P(X_{n+1} = j | X_n = i) = 1, \text{ for all } i.$$

Sum(row) = 1

Definitions

$$\text{states: } \quad 1 \xrightarrow{p(1,2)} 2 \xrightarrow{p(2,3)} 3 \qquad p(1,2) \geq 0$$

Another way to visualize the transition probabilities of a Markov chain is with a **transition state diagram**:

- Each state in the state space is represented by a node/vertex.
- Each nonzero transition probability P_{ij} is represented by an arrow from vertex i to vertex j.

Outline

Example 1: Chance of Rain

- Suppose the weather on any given day in Los Angeles is either sunny or rainy.
- One way to predict tomorrow's weather is to predict that it will be the same as it is today.
- We can model the weather as a two-state Markov chain, with state space $\{0, 1\}$, where $0 =$ "sunny" and $1 =$ "rainy".

Example 1: Chance of Rain

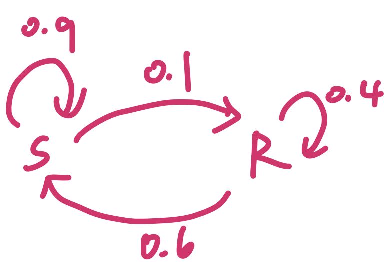
2 states MC

$\pi_0(S, R) =$

	S	R
S	0.9	0.1
R	0.4	0.6

- Suppose if it is sunny today, our prediction that it will be sunny tomorrow is correct 90% of the time.
- Suppose if it is rainy today, our prediction that it will be rainy tomorrow is correct 60% of the time.
- The transition matrix for this Markov chain is then

$$\mathbb{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}.$$

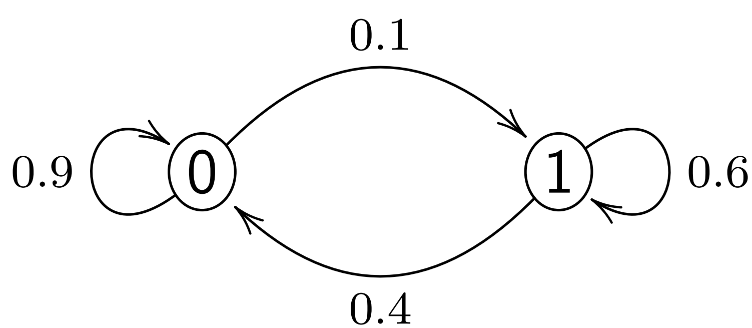


Example 1: Chance of Rain

The transition matrix

$$\mathbb{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$$

can be expressed as a transition state diagram by:



Example 2: The Ehrenfest Urn Model

- The **Ehrenfest urn model** is a classical mathematical model for diffusion of molecules through a membrane.
- Suppose we have two urns labeled A and B that contain a total of N balls (molecules).
- At each step, a ball is randomly chosen from the N balls and moved to the other urn (a molecule diffuses at random through the membrane).

Example 2: The Ehrenfest Urn Model

$\boxed{A}$ $\boxed{B}$

\# ball

- Let X_n denote the number of balls in urn A at step n.
- The possible values of X_n are $\{0, 1, 2, \ldots, N\}$. $0 - N$
- The sequence $\{X_0, X_1, X_2, \ldots, X_n, \ldots\}$ is a Markov chain with state space $\{0, 1, 2, \ldots, N\}$.
- What are the transition probabilities?

	0 — N
0	
$\vdots$	
N	

Example 2: The Ehrenfest Urn Model

Suppose there are i balls in urn A at step n (i.e., $X_n = i$).

拿走 放进

- If we know $X_n = i$, then $X_{n+1} \in \{i - 1, i + 1\}$.

$$X_{n+1} = \begin{cases} X_n - 1 & 50\% \\ X_n - 1 & 50\% \end{cases}$$

- If a ball is chosen from urn A, then $X_{n+1} = i - 1$.
- If a ball is chosen from urn B, then $X_{n+1} = i + 1$.
- Since the ball is chosen uniformly among all N balls, then the probability that a ball is chosen from a particular urn is the proportion of balls that are in that urn.

Example 2: The Ehrenfest Urn Model

The transition probabilities are therefore:

- $P(X_{n+1} = i - 1 | X_n = i) = \frac{i}{N}$ $p(i, i-1) = p(\text{lose}) = p(\text{take from A}) = \frac{i(A)}{N}$
- $P(X_{n+1} = i + 1 | X_n = i) = \frac{N - i}{N} = 1 - \frac{i}{N}$ $p(i, i+1) = p(\text{gain}) = p(\text{get from B}) = 1 - \frac{i}{N}$ $= \frac{N-i\,(B)}{N}$
- $P(X_{n+1} = j | X_n = i) = 0$, for $j \notin \{i - 1, i + 1\}$

$p(i, i+2) = p(i, i-2) = \cdots = 0$

Example 2: The Ehrenfest Urn Model

Suppose $N = 4$. The transition matrix is then

$$\mathbb{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} & P_{04} \\ P_{10} & P_{11} & P_{12} & P_{13} & P_{14} \\ P_{20} & P_{21} & P_{22} & P_{23} & P_{24} \\ P_{30} & P_{31} & P_{32} & P_{33} & P_{34} \\ P_{40} & P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

=1

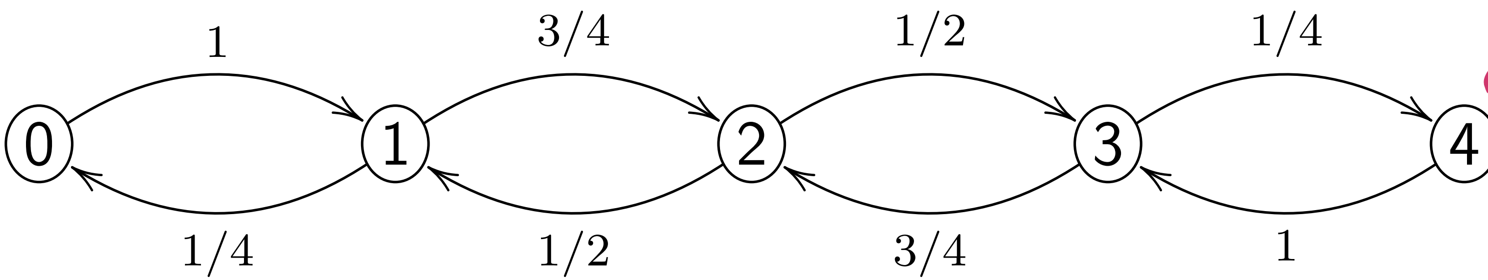
sum=1

where

$$P_{ij} = P(X_{n+1} = j | X_n = i) = \begin{cases} \dfrac{i}{N} & \text{if } j = i-1, \quad \text{lose} \\ \dfrac{N-i}{N} & \text{if } j = i+1, \quad \text{gain} \\ 0 & \text{otherwise.} \quad \text{other} \end{cases}$$

Example 2: The Ehrenfest Urn Model

A transition state diagram for the Ehrenfest urn model, for $N = 4$:



periodic

stationary

Example 3: Random Walk

- A **random walk** is a discrete-time stochastic process that is widely used to model the path an object or particle takes as it moves through space.
- Some applications:
  - The path a particle takes as it moves through a liquid or gas (this is a continuous-time process called **Brownian motion**)
  - The path an animal takes as it searches for food
  - Polymer configurations (self-avoiding walks)
  - A gambler's winnings/losings
  - Stock prices

Example 3: Random Walk

We will focus on one-dimensional random walks.

Definition

A **random walk** is a stochastic process

$$\{X_0, X_1, X_2, \ldots, X_n, \ldots\},$$

defined on the integers $\mathbb{Z}$, such that:

1. The walk starts at 0: $X_0 = 0$. initial money (center)
2. At each step, the random walk moves to the right 1 unit with probability p and moves to the left 1 unit with probability q (so $p + q = 1$).

p

q

$$\cdots\ -4 \quad -3 \xleftarrow{q} -2 \xleftarrow{q} -1 \xleftarrow{q} 0 \xrightarrow{p} 1 \xrightarrow{p} 2 \xrightarrow{p} 3 \quad 4 \ \cdots$$

- Since $X_0 = 0$, then

$$X_1 = \begin{cases} 1 & \text{with probability } p, \\ -1 & \text{with probability } q. \end{cases}$$

- Suppose the current state is $X_n = i$, for $i \in \mathbb{Z}$. Then

$$X_{n+1} = \begin{cases} i+1 & \text{with probability } p, \\ i-1 & \text{with probability } q. \end{cases}$$

- The random walk is a Markov chain, with transition probabilities

$$P(X_{n+1} = j | X_n = i) = \begin{cases} p & \text{if } j = i+1, \\ q & \text{if } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

Example 3: Random Walk

- The random walk can also be expressed as a sum of iid random variables.

Sum of random varible

- Let $Y_1, Y_2, \ldots, Y_n, \ldots$ be iid random variables such that

$$\boxed{P(Y_i = 1) = p} \quad \text{and} \quad \boxed{P(Y_i = -1) = q,}$$

where $p + q = 1$.

- Define the Markov chain $\{X_0, X_1, X_2, \ldots, X_n, \ldots\}$ by:

$$X_0 = 0$$

$$X_n = X_{n-1} + Y_n, \quad \text{for } n = 1, 2, \ldots$$

$Y_n = \begin{cases} +1 \\ -1 \end{cases} \sim$ Bernoulli

$X_{n+1} = X_n + Y_{n+1}$ (depended on current and Y_i)

Example 3: Random Walk

$\therefore X_n = \underbrace{\underbrace{\underbrace{X_0 + Y_1}_{X_1} + Y_2}_{X_2} + Y_3}_{X_3} + \cdots + Y_n = X_0 + \sum_{1}^{n} Y_i$

- So

$$X_1 = X_0 + Y_1 = \begin{cases} X_0 + 1 & \text{if } Y_1 = 1\text{, with pr. } p, \\ X_0 - 1 & \text{if } Y_1 = -1\text{, with pr. } q. \end{cases}$$

- In general,

$$X_n = X_{n-1} + Y_n = \begin{cases} X_{n-1} + 1 & \text{if } Y_n = 1\text{, with pr. } p, \\ X_{n-1} - 1 & \text{if } Y_n = -1\text{, with pr. } q. \end{cases}$$

- The transition probabilities are the same, so this defines the same Markov chain as the random walk.

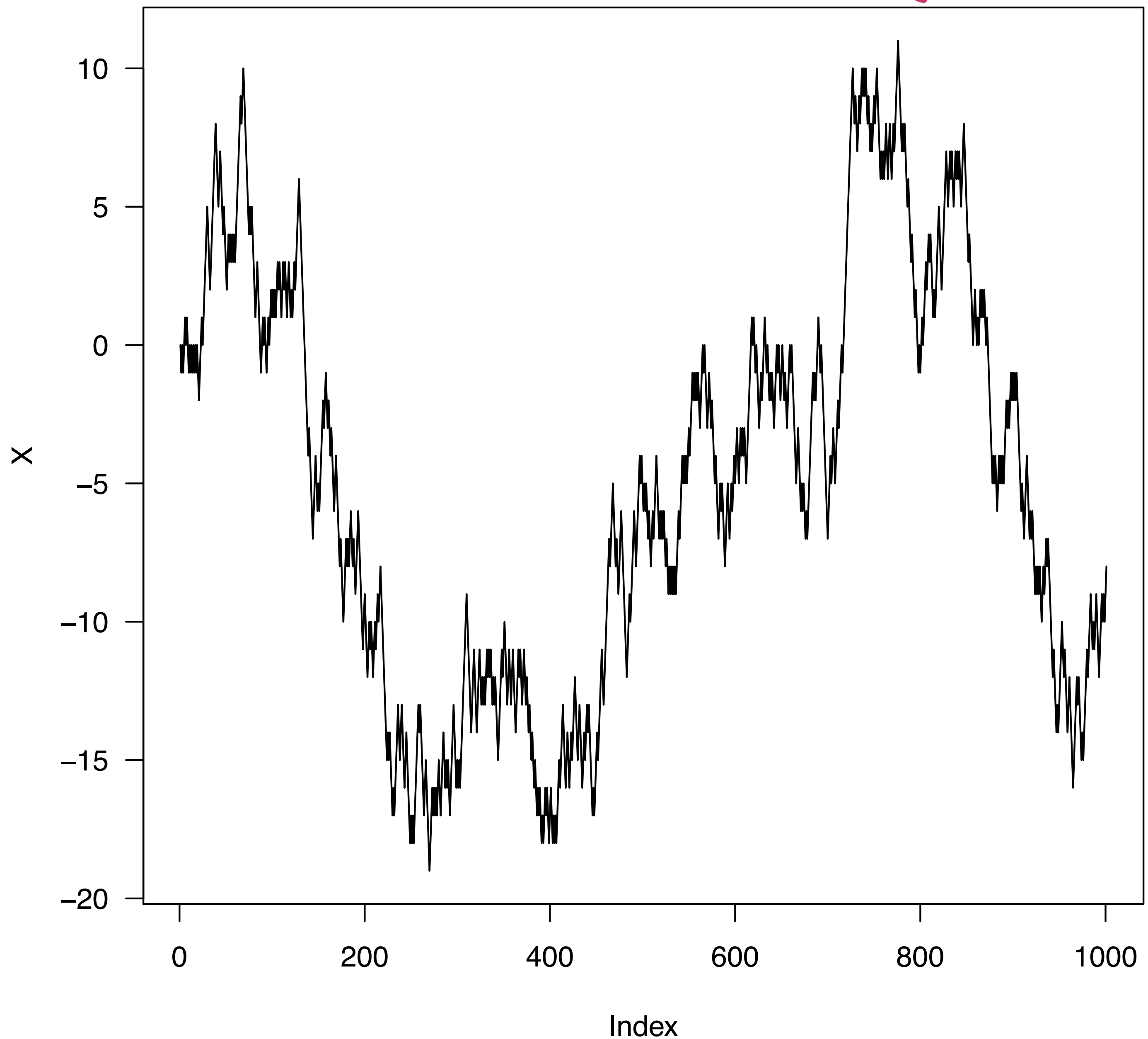
Example 3: Random Walk

R Code to generate a random walk:

```
> set.seed(999) # for reproduceability

> n <- 1000 # specify length of random walk

> p <- 0.5 # specify P(Y = 1)

> # Generate n iid samples from Y
> Y <- sample(c(1, -1),
+             size = n, replace = TRUE, prob = c(p, 1 - p)
+             )

> # Compute the random walk X
> X <- c(0, cumsum(Y))

> # Plot the random walk over time
> plot(X, type = "l", las = 1)
```

Example 3: Random Walk

don't converge



(times)