Markov Chains

Jan Bouda

FI MU

May 3, 2012

Objectives

- Introduce Markov Chains
  – powerful tool for special random processes
- Stationary Distribution
- Random Walks

Stochastic Process

Definition (Stochastic Process)

A collection of random variables $X = \{X_t \mid t \in T\}$ is called a stochastic process. The index t often represents time; X_t is called the state of X at time t.

Example

A gambler is playing a fair coin-flip game: wins 1 Kč if head, loses 1 Kč if tail. Let

- X_0 denote a gambler's initial money
- X_t denote a gambler's money after t flips
  $\Rightarrow \{X_t \mid t \in \{0, 1, 2, \dots\}\}$ is a stochastic process

Stochastic Process

Definition

If X_t assumes values from a finite set, then the process is a finite stochastic process.

Definition

If T (where the index t is chosen) is countably infinite, the process is a discrete time process.

Question:

In the previous example about a gambler's money, is the process finite? Is the process discrete time?

Markov Chain

Definition

A discrete time stochastic process $X = \{X_0, X_1, X_2, \dots\}$ is a Markov chain if

$$\begin{aligned}\Pr(X_t = a | X_{t-1} = b, X_{t-2} = a_{t-2}, \dots, X_0 = a_0) \\ = \Pr(X_t = a | X_{t-1} = b) = P_{b,a}\end{aligned}$$

That is, the value of X_t depends on the value of X_{t-1}, but not on the history of how we arrived at X_{t-1} with that value

Question:

In the example about a gambler's money is the process a Markov chain?

Markov Chain

In other words, if X is a Markov chain, then

$$
\begin{aligned}
&\Pr(X_1 = a | X_0 = b) = P_{b,a} \\
&\Pr(X_2 = a | X_1 = b) = P_{b,a} \\
&\dots
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow P_{b,a} \quad &= \Pr(X_1 = a | X_0 = b) \\
&= \Pr(X_2 = a | X_1 = b) \\
&= \Pr(X_3 = a | X_2 = b) = \dots
\end{aligned}
$$

Markov Chain

- Next, we focus our study on Markov chain whose state space (the set of values that X_t can take) is finite
- So, without loss of generality, we label the states in the state space by $0, 1, 2, \ldots, n$
- The probability $P_{i,j} = \Pr(X_t = j \mid X_{t-1} = i)$ is the probability that the process moves from state i to state j in one step

Transition Matrix

- The definition of Markov chain implies that we can define it using a one-step transition matrix P with

$$P_{i,j} = \Pr(X_t = j \mid X_{t-1} = i)$$

Question: For a particular i, what is $\sum_j P_{i,j}$?

Transition Matrix

- The transition matrix representation of a Markov chain is very convenient for computing the distribution of future states of the process
- Let $p_i(t)$ denote the probability that the process is at state i at time t

Question: Can we compute $p_i(t)$ from the transition matrix P assuming we know $p_0(t-1), p_1(t-1), \ldots$?

Transition Matrix

The value of $p_i(t)$ can be expressed as

$$p_i(t) := p_0(t-1)P_{0,i} + p_1(t-1)P_{1,i} + \cdots + p_n(t-1)P_{n,i}$$

In other words, let $\langle p(t) \rangle$ denote the vector

$$\langle p(t) \rangle = (p_0(t), p_1(t), \ldots, p_n(t))$$

Then, we have

$$\langle p(t) \rangle = \langle p(t-1) \rangle P$$

Transition Matrix

- For any m, we define the m-step transition matrix

$$P_{i,j}^{(m)} = \Pr(X_{t+m} = j \mid X_t = i),$$

which is the probability that we move from state i to state j in exactly m steps

- It is easy to check that $P^{(2)} = P^2$, $P^{(3)} = P \cdot P^{(2)} = P^3$, and in general, $P^{(m)} = P^m$

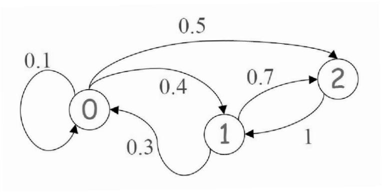
Thus, for any $t \geq 0$ and $m \geq 1$ we have,

$$\langle p(t+m) \rangle = \langle p(t) \rangle P^m$$

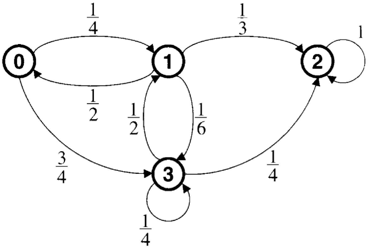
Directed Graph Representation

Markov chain can also be expressed by a directed weighted graph (V, E) such that

- V denotes the state space
- E denotes transition between states with weight of edge (i, j) equal to $P_{i,j}$



Example: Markov Chain & Graph Representation



$$P = \begin{bmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{bmatrix}$$

Consider the probability of going from state 0 to state 3 in exactly 3 steps. From the graph, all possible paths are

$$0-1-0-3, 0-1-3-3, 0-3-1-3, \text{ and } 0-3-3-3$$

Probability of success for each path is: $3/32$, $1/96$, $1/16$ and $3/64$ respectively. Summing up the probabilities we find the total probability is $41/192$.

Example: Markov Chain & Graph Representation

Alternatively, we can compute

$$P^3 = \begin{bmatrix} 3/16 & 7/48 & 29/64 & 41/192 \\ 5/48 & 5/24 & 79/144 & 5/36 \\ 0 & 0 & 1 & 0 \\ 1/16 & 13/96 & 107/192 & 47/192 \end{bmatrix}$$

The entry $P^3_{0,3} = 41/192$ gives the correct answer.

Gambler's Ruin

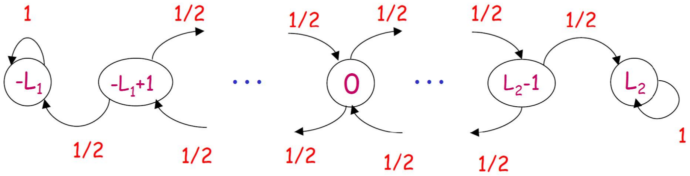
- Discuss Gambler's ruin
  – A study of the game between two gamblers until one is ruined (no money left)
- Introduce stationary distribution
  – and a sufficient condition when a Markov chain has a stationary distribution
- Analyze random walks on a graph

The Game

- Consider two players, one has L_1 Kč and the other has L_2 Kč. Player 1 will continue to throw a fair coin, such that
  – if head appears, he wins 1 Kč
  – if tails appears, he loses 1 Kč
- Suppose the game is played until one player goes bankrupt. What is the probability that Player 1 survives?

The Markov Chain Model

The previous game can be modelled by the following Markov chain:



The Markov Chain Model

- Initially, the chain is at state 0.
- Let $P_j^{(t)}$ denote the probability that after t steps, the chain is at state j
- Also, let q be the probability that the game ends with Player 1 winning L_2 Kč
- We can see that
  (i) $\lim_{t\to\infty} P_j^{(t)} = 0$ for $j \neq -L_1, L_2$
  (ii) $\lim_{t\to\infty} P_j^{(t)} = 1 - q$ for $j = -L_1$
  (iii) $\lim_{t\to\infty} P_j^{(t)} = q$ for $j = L_2$

The Analysis

- Now, let W_t denote the money Player 1 has won after t steps
- By linearity of expectation,

$$E[W_t] = 0$$

- On the other hand,

$$E[W_t] = \sum_j jP_j^{(t)} = 0$$

The Analysis

- By taking limits, we have

$$
\begin{aligned}
0 &= \lim_{t\to\infty} E[W_t] \\
&= \lim_{t\to\infty} \sum_j jP_j^{(t)} \\
&= (-L_1)(1-q) + 0 + 0 + \cdots + 0 + (L_2)q
\end{aligned}
$$

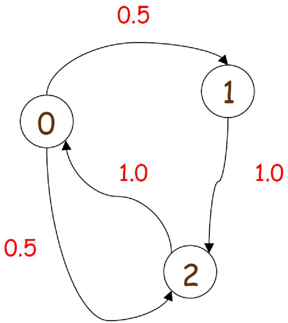
- Re-arranging terms, we obtain

$$q = L_1/(L_1 + L_2)$$

– That is, the probability of winning (or losing) is proportional to the amount of money a player is willing to lose (or win)

Stationary Distribution

Consider the following Markov chain:



- Let $p_j(t)$ denote the probability that the chain is at state j at time t, and let $\langle p(t) \rangle = (p_0(t), p_1(t), p_2(t))$
- Suppose that $\langle p(t) \rangle = (0.4, 0.2, 0.4)$

Question: In this case, what will $\langle p(t+1) \rangle$ be?

Stationary Distribution

- After some calculations, we get

$$\langle p(t+1)\rangle = (0.4, 0.2, 0.4)$$

  which is the same as $\langle p(t)\rangle$!

- We can see that in the previous example, the Markov chain has entered an 'equilibrium' condition at time t, where

$$\langle p(n)\rangle \text{ remains } (0.4, 0.2, 0.4) \text{ for all } n \geq t$$

  $\rightarrow$ this probability distribution is called a Stationary Distribution

Stationary Distribution

Precisely, let P be the transition matrix of a Markov chain. Then,

Definition

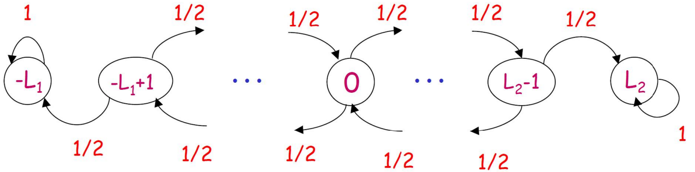
If $\langle p(t+1)\rangle = \langle p(t)\rangle P = \langle p(t)\rangle$, then $\langle p(t)\rangle$ is a stationary distribution of the Markov chain?

Question:

How many stationary distributions can a Markov chain have? Can it be more than one? Can it be none?

Stationary Distribution

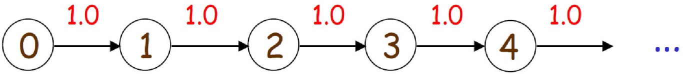
Ans. It can be more that one. For example,



In this case both $(1, 0, 0, \dots, 0)$ and $(0, 0, \dots, 0, 1)$ are stationary distributions

Stationary Distribution

Ans. It can also be none. For example,



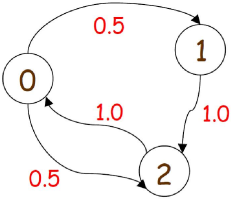
Here, no stationary distributions exists

Question:
Are there some conditions that can be used to tell whether a Markov chain has a unique stationary distribution?

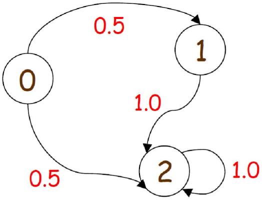
Definition

A Markov chain is irreducible if its directed representation is a strongly connected component. That is, every state j can reach any state k

For example:



irreducible



not irreducible

Special Markov Chains

Definition

A Markov chain is periodic if there exists some state j and some integer $d > 1$ such that

$$\Pr(X_{t+s} = j \mid X_t = j) = 0$$

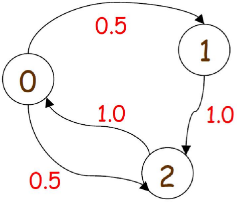
unless s is divisible by d

In other words, once we start at state j, we can only return to j after a multiple of d steps

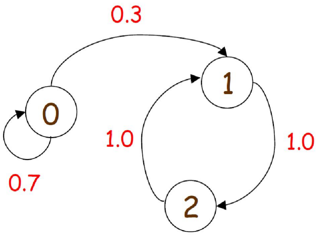
If a Markov chain is not periodic, then it is called aperiodic

Special Markov Chains

For example:



aperiodic



periodic

Sufficient Conditions

Theorem

Suppose a Markov chain is finite with states $0, 1, \ldots, n$. If it is irreducible and aperiodic, then

- *The chain has a unique stationary distribution* $\langle \pi \rangle = (\pi_0, \pi_1, \ldots, \pi_n)$;
- $\pi_k = 1/h_{k,k}$ *where* $h_{k,k}$ *is the expected number of steps to return to state* k, *when starting at state* k

Computing the Stationary Distribution

One way to compute the stationary distribution of a finite Markov chain is to solve the system of linear equations

$$\langle \pi \rangle P = \langle \pi \rangle$$

For example, given the transition matrix

$$P = \begin{bmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{bmatrix}$$

we have five equations for the four unknowns π_0, π_1, π_2 and π_3 given by $\langle \pi \rangle P = \langle \pi \rangle$ and $\sum \pi_i = 1$

Another technique is to study the cut-sets of a Markov chain

Stationary Distribution & Cut-sets

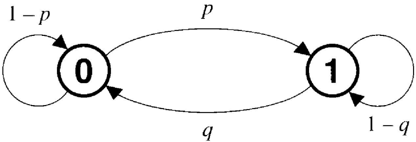
For any state i of the Markov chain, we have

$$\sum_{j \neq i}^{n} \pi_j P_{j,i} = \sum_{j \neq i}^{n} \pi_i P_{i,j}$$

That is, in the stationary distribution the probability that a chain leaves a state equals the probability that it enters a state

Stationary Distribution & Cut-sets

Example:



This Markov chain is used to represent bust errors in communication transmission. The corresponding transition matrix is

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

Solving $\langle \pi \rangle P = \langle \pi \rangle$ yields to system

$$\begin{aligned} \pi_0(1-p) + \pi_1 q &= \pi_0 \\ \pi_0 p + \pi_1(1-q) &= \pi_1 \\ \pi_0 + \pi_1 &= 1 \end{aligned}$$

Stationary Distribution & Cut-sets

Example cont'd:

For these equations, we find the second redundant. The solution is

$$\pi_0 = q/(p+q) \text{ and } \pi_1 = p/(p+q)$$

When $p = .005$ and $q = .1$ in the stationary distribution more that 95% of the bits are received uncorrupted

Using the cut-set formula, we have in the stationary distribution the probability of leaving state 0 must equal the probability of entering 0. Hence

$$\pi_0 p = \pi_1 q$$

Using $\pi_0 + \pi_1 = 1$ yields

$$\pi_0 = q/(p+q) \text{ and } \pi_1 = p/(p+q)$$

Stationary Distribution & Cut-sets

We can summarize this result in the following:

Theorem (10)

Consider a finite, irreducible Markov chain with transition matrix P. If there are nonnegative numbers $\langle\pi\rangle = (\pi_0, \pi_1, \ldots, \pi_n)$ such that $\sum_{j=0}^{n} \pi_i = 1$ and if for any pair of states i, j

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

then $\langle\pi\rangle$ is the stationary distribution corresponding to P

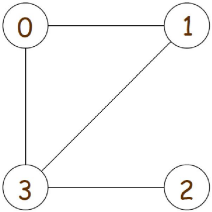
Random Walk

- Let G be a finite, undirected and connected graph
- Let $D(G)$ be a directed graph formed by replacing each undirected edge $\{u, v\}$ of G by two directed edges (u, v) and (v, u)

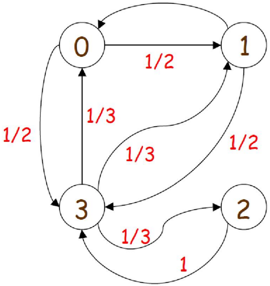
Definition

A random walk on G is a Markov chain whose directed representation is $D(G)$, and for each edge (u, v), the transition probability is $1/\deg(u)$

For example:



G



Representation random walk on G

Random Walk

- Since G is connected, it is easy to check that $D(G)$ is strongly connected
- The lemma below gives a simple criterion for a random walk on G to be aperiodic

Lemma

A random walk on G is aperiodic if and only if G is not bipartite

Random Walk

Consider a random walk on a finite, undirected, connected and non-bipartite graph G. Then G satisfies the conditions of Theorem (10) – and leads to a stationary distribution

The following result shows that this distribution depends only on the degree sequence of the graph!

Theorem

If $G = (V, E)$ is not bipartite, the random walk on G has a unique stationary distribution $\langle\pi\rangle$. Moreover, for the vertex v, the corresponding probability in $\langle\pi\rangle$ is:

$$\pi_v = \deg(v)/(2|E|)$$

Material covered:

- Markov chains
  – Definitions, Gambler's ruin, Graph representation
- Stationary distributions
  – computing the distribution, cut-set technique
- Random Walks
  – Graph representation, definition as Markov chain, implications for the stationary distribution