Markov Chains and Random Walks

RA-MIRI

QT Curs 2020-2021

Stochastic Process

- A stochastic process is a sequence of random variables $\{X_t\}_{t=0}^n$.
- Usually the subindex t refers to time steps and if $t \in \mathbb{N}$, the stochastic process is said to be discrete.
- The random variable X_t is called the state at time t.
- If $n < \infty$ the process is said to be finite, otherwise it is said infinite.
- A stochastic process is used as a model to study the probability of events associated to a random phenomena.

An example: Gambler's Ruin

Model used to evaluate insurance risks.

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.

An example: Gambler's Ruin

Model used to evaluate insurance risks.

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.
- One goal is finding the probability of winning i.e. getting the 1000€.

An example: Gambler's Ruin

Model used to evaluate insurance risks.

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.
- One goal is finding the probability of winning i.e. getting the 1000€.

Notice in this process, once we get 0€ or 1000€, the process stops.

Markov Chain

One simple model of stochastic process is the Markov Chain:

- Markov Chains are defined on a finite set of states (S), where at time t, X_t could be any state in S, together with by the matrix of transition probability for going from each state in S to any other state in S, including the case that the state X_t remains the same at $t+1$.

- In a Markov Chain, at any given time t, the state X_t is determined only by X_{t-1}.
  memoryless: does not remember the history of past events,

Other memoryless stochastic processes are said to be Markovian.

An example: Gambler's Ruin

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.

An example: Gambler's Ruin

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, 1000\}$.

An example: Gambler's Ruin

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, 1000\}$.
- The probability of losing/winning is independent on the state and the time, so this process is a Markov chain.

An example: Gambler's Ruin

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of 100€.
- You keep playing until you loose all your money or you arrive to have 1000€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, 1000\}$.
- The probability of losing/winning is independent on the state and the time, so this process is a Markov chain.
- Observe that the number of states is finite.

Markov-Chains: An important tool for CS

- One of the simplest forms of stochastic dynamics.
- Allows to model stochastic temporal dependencies
- Applications in many areas
  - Surfing the web
  - Design of randomizes algorithms
  - Random walks
  - Machine Learning (Markov Decision Processes)
  - Computer Vision (Markov Random Fields)
  - etc. etc.

Formal definition of Markov Chains

A finite, time-discrete Markov Chain, with finite state $S = \{1, 2, \dots, k\}$ is a stochastic process $\{X_t\}$ s.t. for all $i, j \in S$, and for all $t \geq 0$,

$$\mathbf{Pr}\left[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_t = i\right] = \mathbf{Pr}\left[X_{t+1} = j \mid X_t = i\right].$$

Formal definition of Markov Chains

A finite, time-discrete Markov Chain, with finite state $S = \{1, 2, \ldots, k\}$ is a stochastic process $\{X_t\}$ s.t. for all $i, j \in S$, and for all $t \geq 0$,

$$\mathbf{Pr}\left[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_t = i\right] = \mathbf{Pr}\left[X_{t+1} = j \mid X_t = i\right].$$

We can abstract the time and consider only the probability of moving from state i to state j, as $\mathbf{Pr}\left[X_{t+1} = j \mid X_t = i\right]$

MC: Transition probability matrix

For $v, u \in S$, let $p_{u,v}$ be the probability of going from $u \rightsquigarrow v$ in q steps i.e. $p_{u,v} = \mathbf{Pr}\left[X_{s+1} = v \mid X_s = u\right]$.

MC: Transition probability matrix

For $v, u \in S$, let $p_{u,v}$ be the probability of going from $u \rightsquigarrow v$ in q steps i.e. $p_{u,v} = \mathbf{Pr}\left[X_{s+1} = v \mid X_s = u\right]$.

$P = (p_{u,v})_{u,v \in S}$ is a matrix describing the transition probabilities of the MC

P is called the transition matrix

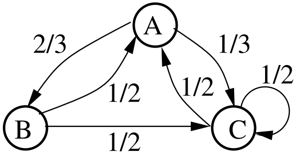
MC: Transition probability matrix

For $v, u \in S$, let $p_{u,v}$ be the probability of going from $u \rightsquigarrow v$ in q steps i.e. $p_{u,v} = \mathbf{Pr}\left[X_{s+1} = v \,|\, X_s = u\right]$.

$P = (p_{u,v})_{u,v \in S}$ is a matrix describing the transition probabilities of the MC

P is called the transition matrix P also defines digraph, possibly

with loops.

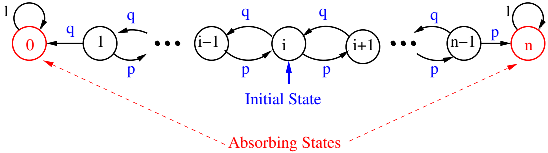


Gambler's Ruin: MC digraph

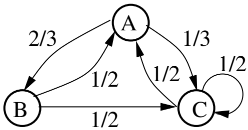
- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of i €and keep playing until you loose all your money or you arrive to have n€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, n\}$.

Gambler's Ruin: MC digraph

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of i € and keep playing until you loose all your money or you arrive to have n€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, n\}$.



Transition matrix: Example



$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \end{array} = P$$

Notice the entry (u, v) in P denotes the probability of going from $u \to v$ in one step.

Notice, in a MC the transition matrix is stochastic, so sum of transitions out of any state must be 1.

Longer transition probabilities

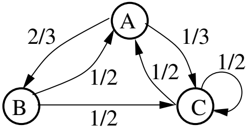
For $v, u \in S$, let $p_{u,v}^t$ be the probability of going from $u \rightsquigarrow v$ in exactly t steps i.e. $p_{u,v}^t = \mathbf{Pr}\left[X_{s+t} = v \,|\, X_s = u\right]$.

Formally for $s \geq 0$ and $t > 1$, $p_{u,v}^t = \mathbf{Pr}\left[X_{s+t} = v \,|\, X_s = u\right]$.

A times, we may use i $P_{u,v}^t$ to indicate entry (u, v)in the matrix P, i.e $p_{u,v}^t = P_{u,v}^t = \mathbf{Pr}\left[X_{s+t} = v \,|\, X_s = u\right]$.

How can we relate P^t with P?

The powers of the transition matrix



$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \end{array} = P$$

In ex. $\mathbf{Pr}[X_1 = C | X_0 = A] = P^1_{A,C} = 1/3$.
$\mathbf{Pr}[X_2 = C | X_0 = A] = P^1_{AB} P^1_{BC} + P^1_{AC} P^1_{CC} = 1/3 + 1/6 = P^2_{A,C}$

In general, assume a MC with k states and transition matrix P, let $u, v \in S$:

- What is the $\mathbf{Pr}[X_1 = u | X_0 = v]$, i.e. $= P_{v,u}$?
- What is the $\mathbf{Pr}[X_2 = u | X_0 = v] = P^2_{v,u}$?

The powers of the transition matrix

Use Law Total Probability+ Markov property:

$$\begin{aligned}\mathbf{Pr}\left[X_2 = u | X_0 = v\right] &= \sum_{w=1}^{m} \mathbf{Pr}\left[X_1 = w | X_0 = v\right] \mathbf{Pr}\left[X_2 = u | X_1 = w\right] \\ &= \sum_{w=1}^{m} P_{v,w} P_{w,u} = P^2_{v,u}.\end{aligned}$$

In general $\mathbf{Pr}\left[X_t = w | X_0 = v\right] = P^t_{v,u}$ and $\mathbf{Pr}\left[X_{k+t} = w | X_k = v\right] = P^t_{v,u}$.

The argument can be generalized to

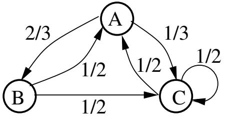
Given the transition matrix P of a MC, then for any $t > 1$,

$$P^t = P \cdot P^{t-1}.$$

Notice the entry (u, v) in P^t denotes the probability of going from $u \to v$ in t steps.

Distributions at time t

To fix the initial state, we consider a random variable X_0, assigning to S an initial distribution π_0, which is a row vector indicating at $t = 0$ the probability of being in the corresponding state.
For example, in the MC:



we may consider,

$$\begin{array}{ccc} A & B & C \\ (0 & 0.3 & 0.6) \end{array} = \pi_0$$

Distributions at time t

Starting with an initial distribution π_0, we can compute the state distribution π_t (on S) at time t,

For a state v,

$$\begin{aligned}\pi_t[v] =& \mathbf{Pr}\left[X_t = v\right] \\ =& \sum_{u \in S} \mathbf{Pr}\left[X_0 = u\right] \mathbf{Pr}\left[X_t = v | X_0 = u\right] \\ =& \sum_{u \in S} \pi_0[u] P^t_{v,u}.\end{aligned}$$

i.e. $\pi_t[y]$ is the probability at step t the system is in state y.

Therefore, $\pi_t = \pi_0 P^t$ and $\pi_{s+t} = \pi_s P^t$.

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of i € and keep playing until you loose all your money or you arrive to have n€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, n\}$.

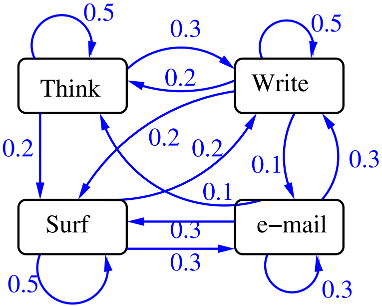
Gambler's Ruin: Exercise

- You place bets of 1€. With probability p, you gain 1€, and with probability $q = 1 - p$ you loose your 1€ bet.
- You start with an initial amount of i €and keep playing until you loose all your money or you arrive to have n€.
- We have a state for each possible amount of money you can accumulate $S = \{0, 1, \ldots, n\}$.
- Which is the initial distribution π_0?
- And, the state distribution at time $t = 3$?

Example MC: Writing a research paper

Recall that Markov Chains are given either by a weighted digraph, where the edge weights are the transition probabilities, or by the $|S| \times |S|$ transition probability matrix P,

Example: Writing a paper $S = \{r, w, e, s\}$



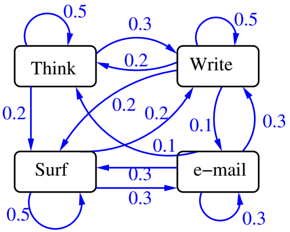
$$
\begin{array}{c c}
 & \begin{array}{cccc} r & w & e & s \end{array} \\
\begin{array}{c} r \\ w \\ e \\ s \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}
\end{array}
$$

More on the Markovian property

Notice the memoryless property does not mean that X_{t+1} is independent from $X_0, X_1, \ldots, X_{t-1}$.

(For instance notice that intuitively we have:
$\mathbf{Pr}\left[\text{Thinking at } t+1\right] < \mathbf{Pr}\left[\text{Thinking at } t \mid \text{Thinking at } t-1\right]$).

But, the dependencies of X_t on $X_0, \ldots, X_{t-1}$, are all captured by X_{t-1}.



Example of writing a paper

$\mathbf{Pr}\left[X_2 = s | X_0 = r\right]$ is the probability that, at $t = 2$, we are in state s, starting in state r.

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} 0.31 & 0.34 & 0.09 & 0.26 \\ 0.21 & 0.38 & 0.14 & 0.27 \\ 0.14 & 0.33 & 0.21 & 0.32 \\ 0.07 & 0.29 & 0.26 & 0.38 \end{pmatrix} & \begin{matrix} r \\ w \\ e \\ s \end{matrix} \end{array}$$

$\mathbf{Pr}\left[X_1 = s | X_0 = r\right] = 0.07.$

Distribution on states

Recall π_t is the prob. distribution at time t over S.

For our example of writing a paper, if $t = 0$ (after waking up):

$$\begin{array}{cccccc} & & r & w & e & s \\ \pi_0 = & (& 0.2 & 0 & 0.3 & 0.5) \end{array}$$

$$\begin{pmatrix} 0.2 & 0 & 0.3 & 0.5 \end{pmatrix} \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.13 & 0.25 & 0.24 & 0.38 \end{pmatrix} = \pi_1$$

Therefore, we have $\pi_t = \pi_0 \times P^t$ and $\pi_{k+t} = \pi_k \times P^t$

Notice $\pi_t = (\pi_t[r], \pi_t[w], \pi_t[e], \pi_t[s])$

Stationary distributions: Writing a paper

- Suppose in the writing a paper example, the t is measured in minutes.
- To see how the Markov chain will evolve after 20 minutes i.e. $\mathbf{Pr}\left[X_{20} = s | X_0 = r\right]$ we must compute P^{20}, and to see if 5' later $\mathbf{Pr}\left[X_{25} = s | X_{20} = s\right]$.
- Matrices P^{20} and P^{25} may be almost identical.
- This indicates that in the long run, the starting state doesn't really matter,
- which implies that after a sufficiently long t: $\pi_t = \pi_{t+k}$, doesn't change when you do further steps.
- That is, for sufficient large t, the vector distribution converges to a π, $\pi_{t+1} = \pi_t P$, $\Rightarrow \pi = \pi P$. The stationary distribution

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}^{20} = \begin{pmatrix} 0.1707318219 & 0.1707318219 & 0.18157173275 & 0.3116530652 \\ 0.1707317681 & 0.3360433708 & 0.18157177167 & 0.3116530893 \\ 0.1707316811 & 0.3360433559 & 0.18157183465 & 0.3116531282 \\ 0.1707315941 & 0.3360433410 & 0.18157189762 & 0.3116531671 \end{pmatrix}$$

$$\begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}^{25} = \begin{pmatrix} 0.170732 & 0.336043 & 0.181572 & 0.311653 \\ 0.170732 & 0.336043 & 0.181572 & 0.311653 \\ 0.170732 & 0.336043 & 0.181572 & 0.311653 \\ 0.170732 & 0.336043 & 0.181572 & 0.311653 \end{pmatrix}$$

Therefore $\pi = (0.170732, 0.336043, 0.181572, 0.311653)$ is the stationary distribution of the writing a paper's problem.

Stationary distributions

A probability vector π is called a stationary distribution over S for P if it satisfies the stationary equations

$$\pi = \pi P.$$

Notice that if a MC has an stationary distribution π that means than after running a certain time the MC, then π the PMF for every r.v. X_i, $0 \leq i \leq n$.