Markov Chains

Lecture #5

Contents

1 Stochastic Process

Sotchastic Process

- A stochastic process X_t (or $X(t)$) is a family of random variables indexed by a parameter t (usually the time). Formally, a stochastic process is a mapping from the sample space S to functions of t. With each element e of S is associated a function $X_t(e)$ where e is a realization of the sotochastic process (also trajectory or sample path).
  - For a given value of e, $X_t(e)$ is a function of time
  - For a given value of t, $X_t(e)$ is a random variable
  - For a given value of e and t, $X_t(e)$ is a (fixed) number
- For any stochastic process
  - State spasce: the set of possible values of X_t

- Parameter space: the set of values of t
- Stochastic processes can be classified according to whether these spaces are discrete or continuous
- According to the type of the parameter space one speaks about discrete time or continuous time stochastic processes.
- Discrete time stochastic processes are also called random sequences.

- In considering stochastic processes we are often interested in quantities like:
  - Time-dependent distribution: defines the probability that X_t takes a value in a particular subset of S at a given instant t
  - Stationary distribution: defines the probability that X_t takes a value in a particular subset of S as t $\to \infty$ (assuming the limit exists)
  - The relationships between $X_s$and X_t for differnet times s and t (e.g. covariance or correlation of X_s and X_t)
  - Hitting probability: the probability that a given state is S will ever be entered
  - First passage time: the instant at which the stochastic process first time enters a given state or set of states starting from a given initial state

2 Markov Process

- A stochastic process is called a Markov process when it has the Markov property

$$P\{X_{t_n} \leq x_n | X_{t_{n-1}} = x_{n-1}, ..., X_1 = x_1\} = P\{X_{t_n} \leq x_n | X_{t_{n-1}} = x_{n-1}\} \, \forall n, \, \forall t_1 < ... < t_n$$

  that is to say that

  - the future path of a Markov process, given its current state ($X_{t_{n-1}}$) and the past history before t_{n-1} , depends only on the current state (not on how this state has been reached).
  - The current state contains all the information (summary of the past) that is needed to characterize the future (stochastic) behaviour of the process.
  - Given the state of the process at an instant its future and past are independent.

3 Markov Chain

- In this course, we will use Markov Chain to refer to a Markov process that is discrete in both time and state. Such a process will be also denoted as DTMC (Discrete time Markov Chain)
- A Markov chain is thus the process that represents the evolution of the process X_n states in a discrete time index $n = 0, 1, ...$
- We also focus on time homogeneous Markov Processes which feature stationary probablistic evolution, that is to say the transition probability does not depend on n

- A Markov process of this kind is **characterized** by the (one-step) transition probabilities (transition from state i to state j)

$$p_{ij} = P(X_n = j | X_{n-1} = i)$$

- Hence, the **probability of a state path** can be expressed as $P\{X_o = i_o, ..., X_n = i_n\} = P\{X_o = i_o\} p_{i_o i_1} p_{i_1 i_2} ... p_{i_{n-1} i_n}$

$$\begin{aligned} \mathrm{P}\{X_0 = i_0, X_1 = i_1\} &= \underbrace{\mathrm{P}\{X_1 = i_1 | X_0 = i_0\}}_{p_{i_0,i_1}} \mathrm{P}\{X_0 = i_0\} \\ \mathrm{P}\{X_0 = i_0, X_1 = i_1, X_2 = i_2\} &= \underbrace{\mathrm{P}\{X_2 = i_2 | X_1 = i_1, X_0 = i_0\}}_{p_{i_1,i_2}} \underbrace{\mathrm{P}\{X_1 = i_1, X_0 = i_0\}}_{p_{i_0,i_1}\mathrm{P}\{X_0 = i_0\}} \\ &= \mathrm{P}\{X_0 = i_0\}\, p_{i_0,i_1}\, p_{i_1,i_2} \end{aligned}$$

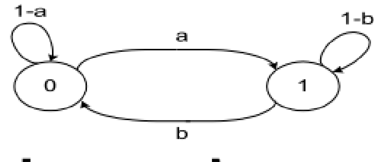
3.1 Single-Step Conditional Transition Probability

- The single step transition probability matrix, denoted as P, represents a ***conditional probability*** matrix that describes the transition probability form state i (identifed by the row id) and state j (identifed by column id)

$$P = \begin{pmatrix} p_{oo} & p_{o1} & p_{02} & .... & ..... \\ p_{12} & p_{22} & p_{23} & .... & .... \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix}$$

  - Since the system always goes to some state, the sum of the row probabilities is 1. A matrix with non-negative elements such that the sum of each row equals 1 is called a ***stochastic matrix.***
  - One can easily show that the product of two stochastic matrices is a stochastic matrix.

- It is very common to represent Matrix P with a graph called state diagram in which
  - the vertix represents the state and
  - teh edge representing the single step transition probability between the connected verticies.



$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \quad 0 \leq a, b \leq 1$$

Figure 1: 2-state DTMC for a cascade of binary comm. channels. Signal values: '0' or '1' form the state values.

3.2 n-Step Conditional Transition Probabilities Matrix

- Typically, the probability that the system, initially in state i, will in state j after two steps is $\sum_k p_{ik} P_{kj}$ to take into account all paths via an intermediate state k.
  - Clearly this the element {i, j} of the matrix P^2
  - Similarly, one finds that the n-step transition probability matrix P^n
- Denote its elements by $p_{ij}^{(n)}$ (the subscript refers to the number of steps). Since it holds that $P^n = P^m \cdot P^{n-m}$ ($0 \leq$ m $\leq$ n), we can write in component form

$$\underbrace{p_{ij}^{(n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n-m)}}_{Chapman-Kolmogorov\, equation}$$

- Chapman-Kolmogorov equation simply expresses the law of total probability, where the transition in n steps from state i to state j is conditioned on the system being in state k after m steps.

3.3 Unconditional State Probabilities $\pi^{(n)}$ Vector

- The unconditional state probability $\pi_i^{(n)}$ representes the probability that the process is in state i at time n; i.e. $P\{X_n = i\}$
- By arranging these probabilities in a vector, we obtain Unconditional State Probabilities $\pi^{(n)}$ Vector $[\pi_1^{(n)}\, \pi_2^{(n)} ....]$
- Note that $\pi^{(n)}$ represents the probabilistic transient behavior of the DTMC
- By the law of total probability we have

$$P\{X_n = i\} = \sum_k P\{X_n = i | X_{n-1} = k\} P\{X_{n-1} = k\}$$

$$\pi_i^{(n)} = \sum_k \pi_k^{(n-1)} p_{ki}$$

$$\overline{\pi}_i^{(n)} = \overline{\pi}_i^{(n-1)} P$$

- and Recursively, we can have

$$\overline{\pi}_i^{(n)} = \overline{\pi}_i^{(0)} P^n$$

> The Beauty of Markov Chains
> Given the initial probabilities and by the repeated use of one-step transition probabilities (n-step transition probability matrix) We can determine the n^{th} order pmf for all n

Example

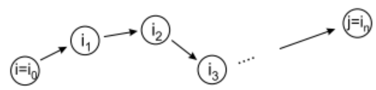
- Derive an expression of n-step transition probability matrix for the binary cascaded channel

$$P^n = \begin{pmatrix} \frac{2}{3} + \frac{1}{3}(\frac{1}{4})^n & \frac{1}{3} - \frac{1}{3}(\frac{1}{4})^n \\ \frac{2}{3} - \frac{2}{3}(\frac{1}{4})^n & \frac{1}{3} + \frac{2}{3}(\frac{1}{4})^n \end{pmatrix}$$

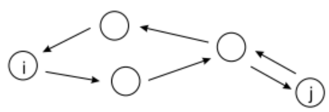
  - what will happen to P^n as $n \to \infty$?
  - Assuming an inital dtate distribution of [p q], what would be the state distribution as $n \to \infty$?

3.4 Classification of states of a Markov chain

- State i leads to state j (written i → j), if there is a path $i_0 = i, i_1, ..., i_n = j$ such that all the transition probabilities are positive,$p_{i_k,i_{k+1}} > 0$, $k = 0, ..., n-1$. Then $(P^n)_{ij} > 0$.



- States i and j communicate (written i ↔ j), if i → j and j → i.



- Communication is an equivalence relation: the states can be grouped into equivalent classes so that
  - within each class all the states communicate with each other
  - two states from two different classes never communicate which each other

> If all the states of a Markov chain belong to the same communicating class, the Markov chain is said to be ***irreducible.***

- A set of states is ***closed***, if none of its states leads to any of the states outside the set.

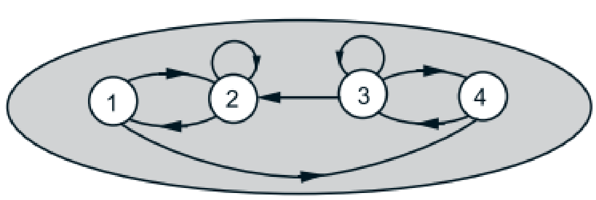


Figure 2: Single class of recurrent states

- A single state which alone forms a closed set is called an absorbing state
  - for an absorbing state we have $p_{i,i} = 1$

- one may reach an absorbing state from other states, but one cannot get out of it

- Each state is either **transient or recurrent**.
  - A state i is transient if there is a nonzero probability that the Markov chain will not return to that state again.
  - A state i is recurrent if the probability of returning to the state is = 1.
    i.e. with certainty, the system sometimes returns to the state.
    * Recurrent states are further classified according to the expectation of the time T_{ii} it takes to return to the state.
      - positive recurrent expectation of first return time $< \infty$
      If the first return time of state i can only be a multiple of an integer d $>$ 1 the state i is called ***periodic***. Otherwise the state is ***aperiodic***.
      - null recurrent expectation of first return time $= \infty$

Type	# visits	E[T_{ii}]
Transient	$< \infty$	∞
Null Recurrent	∞	∞
Positive Recurrent	∞	$< \infty$

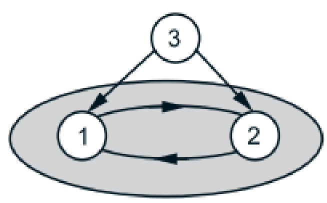


Figure 3: Single class of recurrent states (1 and 2) and one transient state (3)

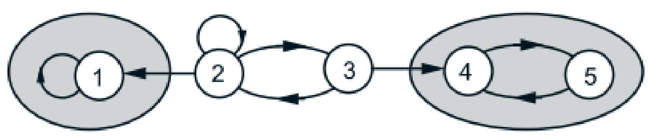


Figure 4: Two classes of recurrent states (class of state1 and class of states 4 and 5) and two transient states (2 and 3)

3.5 State Sojurn Time

- The number of steps the system consecutively stays in state i is geometrically distributed
- The distribution of state Sojurn time is ~Geo(1-p_{ii}) because the exit from the state occurs with the probability (1-p_{ii})

Theorem: Steady-state probability distributions in a DTMC
In an irreducible and aperiodic DTMC with positive recurrent (finite) states:

- the limiting distribution $\overline{\pi} = lim_{n\to\infty}\overline{\pi}(n) = lim_{n\to\infty}p_{i,j}(n)$ does exist
- $\overline{\pi}$is independent of the initial probability distribution π_0;
- $\overline{\pi}$ is the unique stationary probability distribution (AKA, the steady-state probability vector, equilibrium distribution).
- $\overline{\pi}$ can be estimated by solving

$$\overline{\pi}P = \overline{\pi} \;\; and \;\; \sum_i \pi_i = 1$$

- Note that the $\overline{\pi}$ can be estimated as the eigenvector of the matrix P belonging to the eigenvalue 1.
- Note that π_j defines the proportion of time (steps) the system stays in state j.

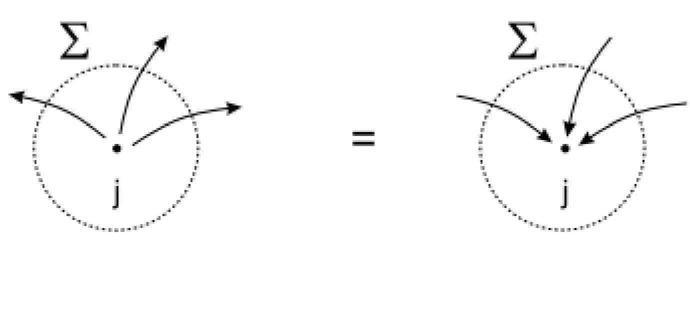
Note. An equilibrium does not mean that nothing happens in the system, but merely that the information on the initial state of the system has been "forgot" or "washed out" because of the stochastic development.

3.6 Gobal Balance Equations

- The equation $\overline{\pi}P = \overline{\pi}$ or $\pi_j = \sum_i \pi_i p_{ij}$ is called (global) balance condition. This equation can be rewritted as

$$\pi_j \sum_k p_{jk} = \sum_i \pi_i p_{ij}$$
$$\sum_k \pi_j p_{jk} = \sum_i \pi_i p_{ij}$$

  - LHS indicates the prob. that the system is in state j and makes a transition to another state
  - RHS indicates the prob. that the system is in another state and makes a transition to state j
  - Hence, at steady state, there are as many exits form state j as there are entries to it (Balance of probability flows).



3.6.1 Solving Balance Equations

- In general, the equilibrium equations are solved as follows (assume a finite state space with n states):
  - Write the balance condition for all but one of the states (n − 1 equations)
    - the equations fix the relative values of the equilibrium probabilities
    - the solution is determined up to a constant factor
  - The last balance equation (automatically satisfied) is replaced by the normalization condition $\sum_i \pi_i = 1$.
- On using the computer, on can solve the system of equations by solving

$$\overline{\pi}(P + E - I) = \mathbf{e}$$

Where E and e are all ones matrix and vector respectively. Note that $\overline{\pi}P = \overline{\pi}$ and $\overline{\pi}E = \mathbf{e}$.

3.6.2 Example: Cascaded Binary Channel

Calculate the stationary distribution for cascaded binary channel

Solution

$$P = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}$$

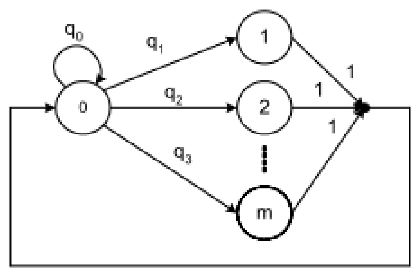
$\overline{\pi}P = \overline{\pi} \rightarrow \pi_o = a\pi_0 + (1-b)\pi_1$

$\pi_1 + \pi_0 = 1$

Solve the two equation together to get π_0, π_1

3.6.3 Example: Computer Program Analysis

A typical computer program: continuous cycle of compute & I/O



The resulting DTMC is irreducible with period =1.

$$P = \begin{bmatrix} q_0 & q_1 & . & . & q_m \\ 1 & 0 & . & . & 0 \\ 1 & 0 & . & . & 0 \\ 1 & 0 & . & . & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\pi_0 = \pi_0 q_0 + \sum_{i=1}^{m} \pi_i$

$\pi_j = \pi_0 q_j$ for all $j = 1, ..., m$

$\sum_i \pi_i = 1$

4 Homework

- A machine can be either working or broken down on a given day. If it is working, it will break down in the next day with probability b, and will continue working with probability $1-b$. If it breaks down on a given day, it will be repaired and be working in the next day with probability r, and will continue to be broken down with probability $1-r$. What is the steady-state probability that the machine is working on a given day?
- If the machine remains broken for a given number of days, despite the repair efforts, it is replaced by a new working machine. What is the steady-state probability that the machine is working on a given day when $l = 3$?