Markov Chains

Marcus Gallagher

ENGG7302
Advanced Computational Techniques in Engineering

Outline

What is a Markov Chain?

- A Markov chain is a discrete-time stochastic process, i.e. a sequence of random variables $X_1, X_2, X_3, \ldots$ with the Markov property:
  - The next state depends only on the present state and not how we arrived in this state

$$P(X_{n+1} = x | X_n = x_n, \ldots, X_1 = x_1) = P(X_{n+1} = x | X_n = x_n)$$

- The **state space** of the chain is the set S of possible values X_i.
- A Markov chain can be represented as a directed graph, with states X_i as nodes and edges denoting transition probabilities between states.

Examples

- Example: simple weather model (G&S Ex.11.1)

- Example: 1-D random walk waiting lines (queues).

- Example: branching process
- Each member of a population has a probability of creating 0,1,2,... members which will make up the next generation.
- The size of the n^{th} generation is a Markov chain.
- Examples include nuclear chain reactions, survival of family surnames, gene mutations.

- Example: waiting lines (queues).
- Customers or jobs arrive randomly and wait for service.
- If the server is free the jobs gets immediate service, otherwise they join a queue (e.g FIFO).
- Busy period: total duration of constant service starting at t=0.
- We might be interested in:
  - The total number of jobs in the busy period
  - Duration of the busy period
  - Probability of busy period ending

- It turns out that queues can be universally characterised depending on the type of interarrival distribution, service time distribution and the number of servers active.
- Let x_n be the number of jobs waiting in line at time instant $t = n$ when the n^{th} customer departs after completing service, then x_n forms a Markov chain.
- In fact, this is an embedded Markov chain because it involves observing some underlying stochastic process $x(t)$ at certain time instances.
- x(t) is not necessarily markovian.

Time-homogeneous Markov chains

- A finite, time-homogeneous Markov chain is specified by an initial distribution $P(X_0 = x)$ and a transition probability matrix, **P**:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} \\ p_{21} & \ddots & & \\ \vdots & & \ddots & \\ p_{i1} & & & p_{ij} \end{bmatrix}$$

- Each row of **P** must sum to 1.

- Most of the examples in this lecture are in fact THMCs.
- More examples: binary communication channel, random walks, queues.

Markov Chain Properties

- The probability of moving from state i to state j in n time steps:
$$p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$
- A single transition is then:
$$p_{ij} = P(X_1 = j | X_0 = i)$$

- To start the chain we must specify an **initial distribution** $P(X_0 = x)$.
- The evolution of the chain through one time step is given by:

$$P(X_n = j) = \sum_{r \in S} p_{rj} P(X_{n-1} = r) = \sum_{r \in S} p_{rj}^{(n)} P(X_0 = r)$$

- The marginal distribution $P(X_n = x)$ is the distribution over states at some time n.

- A state j is said to be **accessible** from state i if, given we are in state i, there is a non-zero probability that at some time in the future, we will be in state j:

$$\exists n : P(X_n = j | X_0 = i) > 0$$

- A state i is said to **communicate** with state j if it is true that both i is accessible from j and that j is accessible from i.
  - A set of states C is said to be a **communicating class** if every pair of states in C communicates with each other.
  - A communicating class is **closed** if the probability of leaving the class is zero.
- A Markov chain is said to be **irreducible** if its state space is a communicating class (i.e. it is possible to get to any state from any state).

- A state i has **period** k if any return to state i must occur in some multiple of k time steps and k is the largest number with this property:

$$k = \mathrm{gcd}\{n : P(X_n = i | X_0 = i) > 0\}$$

- If $k = 1$, then the state is said to be **aperiodic**.
- It can be shown that every state in a communicating class must have the same period.

- A state i is said to be **transient** if, given that we start in state i, there is a non-zero probability that we will never return back to i. Formally, let T_i be the random variable denoting the next return time to state i (the **hitting time**):

$$T_i = \min\{n : X_n = i | X_0 = i\}$$

- then state i is transient if T_i is finite with some probability: $P(T_i < \infty) < 1$.

- If a state i is not transient (i.e. it has finite hitting time with probability 1), then it is said to be **recurrent** or **persistent**.
- While the hitting time is finite in this case, it might not have a finite average. Let M_i be the expected (average) return time, $M_i = E[T_i]$.
  - State i is **positive recurrent** if M_i is finite. Otherwise, state i is **null recurrent**.
- It can be shown that a state is recurrent if and only if $\sum_{n=0}^{\infty} p_{ij}^{(n)} = \infty$.

- A state i is called **absorbing** if it is impossible to leave this state. Therefore, the state i is absorbing if and only if $p_{ii} = 1$ and p_{ij} for $i \neq j$.
- Examples: drunkard's walk, reflecting barriers, cyclic random walks

Ergodicity

- The general idea of ergodicity is important and easy to understand. Unfortunately it is a little complicated to characterise in terms of other Markov chain properties.
- A Markov chain is called **ergodic** if it is possible to go from every state to every other state (not nescessarily in one move).
- Now...
  - A set of states is ergodic if it is closed but no proper subset is closed.
  - A state i will then be **ergodic** if it is aperiodic and positive recurrent.
  - If all states in a Markov chain are ergodic, then the chain is said to be ergodic.
- So, if a Markov chain is irreducible, it must be positive recurrent. Then if it is aperiodic, it **must** be ergodic. Alternatively, if it is periodic, it may or may not be ergodic.

- If there is a number k such that every state can reach every other state in exactly k steps, then the Markov chain is called **regular**.
- Not every ergodic Markov chain is regular (but many are!).
- If a chain is aperiodic, it is also regular.

- For a THMC, a vector π is a **stationary or equilibrium distribution** if its entries sum to 1 and satisfy

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}$$

- An irreducible chain has a stationary distribution iff all of its states are positive-recurrent. In that case, π is unique and is related to the expected return time:

$$\pi_j = \frac{1}{M_j}$$

- Further, if the chain is both irreducible and aperiodic, then for any i and j

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{M_j}$$

- There is no assumption on the starting distribution; the chain converges to the stationary distribution regardless of where it began.

- We can also say that the stationary distribution is a (row) vector which satisfies the equation

$$\pi = \pi \mathbf{P}$$

  i.e. the stationary distribution π is a normalized left eigenvector of the transition matrix associated with the eigenvalue 1.
- An ergodic Markov chain can have only one equilibrium distribution.
- It can be shown that a homogeneous Markov chain will be ergodic, subject only to weak restrictions on the invariant distribution and the transition probabilities.

- π can be viewed as a fixed point of the linear (hence continuous) transformation on the unit simplex associated to the matrix $\mathbf{P}$. As any continuous transformation in the unit simplex has a fixed point, a stationary distribution always exists, but is not guaranteed to be unique, in general. However, if the Markov chain is irreducible and aperiodic, then there is a unique stationary distribution π.
- In addition, $\mathbf{P}^k$ converges to a rank-one matrix in which each row is the stationary distribution π

$$\lim_{k\to\infty} \mathbf{P}^k = \mathbf{1}\pi$$

So the Markov chain forgets where it began (initial distribution) over time and converges towards its stationary distribution.

- A sufficient (but not necessary) condition for ensuring that the required distribution π is invariant is to choose the transition probabilities to satisfy the property of **detailed balance**:

$$p(x_n)p_{nm} = p(x_m)p_{mn}$$

- A Markov chain that respects detailed balance is said to be **reversible**.