MVE550 2021 Lecture 3
Dobrow Chapter 2 and start of Chapter 3

Petter Mostad

Chalmers University

November 7, 2022

Some words you need to learn about Markov chains

MARKOV CHAIN, STATE SPACE, TIME-HOMOGENEOUS, TRANSITION MATRIX, STOCHASTIC MATRIX, LIMITING DISTRIBUTION, STATIONARY DISTRIBUTION, POSITIVE MATRIX, REGULAR TRANSITION MATRIX, RANDOM WALK, TRANSITION GRAPH, WEIGHTED GRAPH, ACCESSIBLE STATES, COMMUNICATING STATES, EQUIVALENCE RELATION, COMMUNICATION CLASSES, IRREDUCIBILITY, RECURRENT STATES, TRANSIENT STATES, CLOSED COMMUNICATION CLASSES, CANONICAL DECOMPOSITION, IRREDUCIBLE MARKOV CHAINS, POSITIVE RECURRENT STATES, NULL RECURRENT STATES, PERIODICITY, APERIODIC, ERGODIC MARKOV CHAINS, TIME REVERSIBILITY, DETAILED BALANCE CONDITION, ABSORBING STATES, ABSORBING MARKOV CHAINS, FUNDAMENTAL MATRIX, ...

Overview

- Definition and examples of (discrete time, discrete state-space) Markov chains.
- Basic computations
- Investigating long term evolution using powers of matrices or simulation
- Induction
- Limiting and stationary distributions

Example

- Consider a game: At each time step i you are at positions 1, 2, or 3.
- We write $X_i = 1$, $X_i = 2$, or $X_i = 3$ for $i = 0, 1, 2, \ldots$.
- At each time step, you move to a higher number (or from 3 to 1) with probability p, or stay put with probability $1 - p$.
- The transitions can be specified with

$$
\begin{array}{ccc}
\Pr(X_{i+1} = 1 \mid X_i = 1) = 1 - p & \Pr(X_{i+1} = 2 \mid X_i = 1) = p & \Pr(X_{i+1} = 3 \mid X_i = 1) = 0 \\
\Pr(X_{i+1} = 1 \mid X_i = 2) = 0 & \Pr(X_{i+1} = 2 \mid X_i = 2) = 1 - p & \Pr(X_{i+1} = 3 \mid X_i = 2) = p \\
\Pr(X_{i+1} = 1 \mid X_i = 3) = p & \Pr(X_{i+1} = 2 \mid X_i = 3) = 0 & \Pr(X_{i+1} = 3 \mid X_i = 3) = 1 - p
\end{array}
$$

- A more succinct specification is with the *transition matrix*:

$$
P = \begin{bmatrix} 1 - p & p & 0 \\ 0 & 1 - p & p \\ p & 0 & 1 - p \end{bmatrix}.
$$

- The sequence $X_0, X_1, X_2, \ldots,$ is an example of a *Markov chain* (see definition on next overhead).

Definition of a Markov chain

Let S be a discrete set (not necessarily finite), called the *state space*. A *Markov chain* is a sequence of random variables $X_0, X_1, \dots$ taking values in S, with the property

$$\pi(X_{n+1} \mid X_0, X_1, \dots, X_n) = \pi(X_{n+1} \mid X_n)$$

for all $n \geq 1$.

- The chain is *time-homogeneous* if, for all $n > 0$,

  $$\pi(X_{n+1} \mid X_n) = \pi(X_1 \mid X_0)$$

  (We will generally assume this).
- The *transition matrix* is defined with

  $$P_{ij} = \pi(X_1 = j \mid X_0 = i)$$

- A *stochastic matrix* is a real matrix P with non-negative entries, satisfying $P\mathbf{1}^t = \mathbf{1}^t$, where $\mathbf{1}$ is a row vector consisting only of 1's.
- All transition matrices are stochastic matrices, and all stochastic matrices can be used as transition matrices.

Basic computations

- If v is a vector describing the probability distribution of states at stage k, then vP is the vector describing the probability distribution of states at stage $k+1$.
- If v is a vector describing the distribution of states at stage k, then vP^n is the vector describing the distribution of states at stage $k+n$.
- Thus the probability to go from state i to state j in n steps is given by $(P^n)_{ij}$. (We write P^n_{ij})
- Chapman-Kolmogorov relationship $P^{n+m}_{ij} = \sum_k P^m_{ik} P^n_{kj}$ is derived from $P^{m+n} = P^m P^n$.
- The probability of being at i_1 at stage n_1, and then at i_2 in stage n_2, and so on up to i_k at stage n_k, with $n_1 < n_2 < \cdots < n_k$, is given by the product of corresponding entries of powers of the transition matrix:

$$\left(p_0 P^{n_1}\right)_{i_1} \left(P^{n_2 - n_1}\right)_{i_1 i_2} \left(P^{n_3 - n_2}\right)_{i_2 i_3} \cdots \left(P^{n_k - n_{k-1}}\right)_{i_{k-1} i_k}$$

where p_0 is the distribution of states for X_0.

Long term evolution: Computing powers of P

When the number of states in S is finite and not too big, we can investigate long term behaviour by computing P^n for large n.

- In some cases, the powers stabilize into a matrix where all rows are identical.
- It may also stabilize without identical rows: Try out $P = I$, the identity matrix!
- Sometimes it does *not* stabilize: Try out, for example

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

- Note that if P is block-diagonal, it may combine several behaviours: If $P = \begin{bmatrix} P_1 & 0 & \dots & 0 \\ 0 & P_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & P_k \end{bmatrix}$ then $P^n = \begin{bmatrix} P_1^n & 0 & \dots & 0 \\ 0 & P_2^n & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & P_k^n \end{bmatrix}$.

Long term evolution: Using simulation

If S is large or infinite, we may instead investigate long term behaviour using *simulation*:

Repeat many times:

- Draw x_0 according to $\pi(x_0)$.
- For i in 1 through n:
  - Draw x_i according to $\pi(x_i \mid x_{i-1})$.

Use the distribution of the x_n to approximate the distribution of X_n.

Proving stuff using induction

1. Formulate a statement $S(n)$ depending on a non-negative integer n.
2. Prove $S(0)$.
3. Prove that if $S(n)$ is true, then $S(n+1)$ is also true.

With this, one may conclude that $S(n)$ is true for all non-negative n.

Example of induction: 2-state Markov chain

- Any 2-state Markov chain has transition matrix $\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ for some $0 < p < 1$, $0 < q < 1$.
- We can prove by induction that, for any $n \geq 0$,

$$\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}^n = \frac{1}{p+q}\left[\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \begin{pmatrix} p & -p \\ -q & q \end{pmatrix}(1-p-q)^n\right].$$

- We can use this to study what happens with the Markov chain when n grows.

Limiting distribution

- A *limiting distribution* for a Markov chain with transition matrix P is a probability vector v such that

$$\lim_{n\to\infty}(P^n)_{ij} = v_j$$

  for all i and j.
- Equivalent formulation: The limit $\lim_{n\to\infty}(P^n)_{ij}$ exists and does not depend on i.
- Equivalent formulation: $\lim_{n\to\infty} P^n$ is a stochastic matrix with all rows identical.
- A Markov chain has either no or one unique limiting distribution. We have seen examples of both cases in examples.
- If a limiting distribution exists, its probabilities correspond to the proportion of time steps the chain spends at each state.

Stationary distribution

- A *stationary distribution* for a Markov chain is a distribution that is unchanged when applying one step of the Markov chain.
- If P is the transition matrix, then a probability vector v represents a stationary distribution if and only if

$$vP = v$$

- A Markov chain can have zero, one, or many stationary distributions.
- Limiting distributions are stationary distributions (but not necessarily vice versa).

Regular transition matrices

- A stochastic matrix P is *positive* if all entries are positive. A stochastic matrix P is *regular* if P^n is positive for some $n > 0$.
- **Limit Theorem for Regular Markov Chains**: If the transition matrix P is regular, the limiting distribution exists. There are no other stationary distributions. The limiting distribution is positive, i.e., all its probabilities are positive.
- Proof in section 3:10 (not part of exam material): One first proves that regular Markov chains are *ergodic* (i.e., irreducible, aperiodic, and all states have finite return times) and then that ergodic Markov chains have a limiting distribution. Two proofs are given:
  - A proof using *coupling*
  - A proof using linear algebra

Finding a stationary distribution

- Find the v satisfying $vP = v$ by
  - solving the linear system $vP = v$.
  - guessing at a v, and showing that $vP = v$.
  - computing an eigenvector for the transpose P^t belonging to the eigenvalue 1.
- Having found a v satisfying $vP = v$; if the transition matrix P is regular, we know v represents the unique limiting distribution and the unique stationary distribution.

Example: Random walks on undirected graphs

- An *undirected graph* consists of *nodes* and *undirected edges* connecting them. (An edge may connect a node with itself).
- An undirected graph defines a *random walk Markov chain* by, at every time step, following one of the edges out of a node, with equal probability. (You also need a starting distribution).
- When the graph is finite, show that the vector u is a stationary distribution, where $u_i = deg(i)/S$, where $deg(i)$ is the number of edges going into edge i and S is the sum of all weights, counting weights on edges between different nodes twice.
- Generalization: A *weighted undirected graph* is a graph with a positive weight at any edge between i and j for all i and j.
- Define the Markov chain by choosing the next node with probabilities according to the weights.
- Show that when the graph is finite, the vector u is a stationary distribution, where $u_i = w(i)/S$, where $w(i)$ is the sum of the weights of the edges going into i, and e is the total sum of all weights, counted as above.