Discrete Random Processes (EDRP)

Lecture 4

A quick recap

Previously on the lectures:

- long-term behaviour of Markov chains
- a **limiting distribution** – a probability distribution λ such that for all i and j, $\lim_{n\to\infty}(\mathbf{P}^n)_{ij} = \lambda_j$,
  - it does not always exist,
  - if it exists, them for any initial distribution α, $\lim_n \alpha\mathbf{P}^n = \lambda$,
- a **stationary distribution** – a probability distribution π such that $\pi\mathbf{P} = \pi$.
  - there are Markov chains with more than one stationary distribution,
  - there are Markov chains with unique stationary distributions that are not limiting distributions,
- the limiting and stationary distribution may coincide in some situations.

A quick recap - cont'd

Questions about stationary and limiting distributions for stochastic matrices:

- Does every Markov chain have a stationary distribution π?
  - Yes - if the state space is *finite* (if infinite - not always).
- When is the stationary distribution *unique*?
- When can we conclude that

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix},$$

  and hence that for all initial distributions α

$$\lim_{n\to\infty} \alpha \mathbf{P}^n = \pi?$$

A quick recap - cont'd

Definition 1

- A matrix $\mathbf{M}$ is said to be **positive** if all the entries of $\mathbf{M}$ are positive. We write $\mathbf{M} > 0$. Similarly, write $x > 0$ for a vector x with all positive entries.
- A transition matrix $\mathbf{P}$ is said to be **regular** if some power of $\mathbf{P}$ is positive. That is, $\mathbf{P}^n > 0$ for some $n \geq 1$.

Theorem 2

A Markov chain whose transition matrix $\mathbf{P}$ *is regular has a limiting distribution, which is the unique, positive, stationary distribution of the chain.*

Stationary distributions, eigenvalues & eigenvectors

Stationary distributions, eigenvalues & eigenvectors

The stationary distribution of a Markov chain is closely related to the eigenvectors and eigenvalues of the transition matrix.

Recall: if $\mathbf{A}$ is a square $n \times n$ matrix then

- if there is a *column* vector $\mathbf{x} \neq \mathbf{0}$ such that

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

  for $\lambda \in \mathbb{C}$, then λ is called a **eigenvalue of A** with corresponding **right eigenvector x**;

- a **left eigenvector of A** is a *row* vector $\mathbf{y}$, which satisfies

$$\mathbf{y}\mathbf{A} = \mu \mathbf{y}$$

  for some $\mu \in \mathbb{C}$;

- a left eigenvector of $\mathbf{A}$ is a right eigenvector of $\mathbf{A}^T$.

If

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

has an eigenvalue λ, then the corresponding eigenvectors satisfy

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

equivalently,

$$\begin{bmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The last equation can be written compactly as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0},$$

with $\mathbf{I}$ being the identity matrix, and $\mathbf{0}$ - the column vector of all 0s. From Cramer's rule, a linear system of equations has nontrivial solutions iff the determinant vanishes, so the solutions of the above equation are given by

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

This equation is known as the **characteristic equation of A**, and the left-hand side is known as the **characteristic polynomial**.

Example 1

Find eigenvalues and eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}.$$

An eigenvector of a matrix $\mathbf{A}$ is a non-zero vector which does not change direction when multiplied by $\mathbf{A}$. Applying $\mathbf{A}$ to the eigenvector only *scales* the eigenvector by the eigenvalue.

The connection between stationary distributions and eigenvectors: since any stationary distribution π satisfies $\pi\mathbf{P} = \pi$, it follows that

π is a left eigenvector of $\mathbf{P}$ corresponding to eigenvalue $\lambda = 1$.

If a Markov chain has a unique stationary distribution, then the distribution is a right (column) eigenvector of $\mathbf{P}^T$ corresponding to eigenvalue $\lambda = 1$.

Markov chains – relationships between states

Communication

The long-term behavior of a Markov chain is related to how often states are visited. Now we will look more closely at the relationship between states and how reachable, or accessible, groups of states are from each other.

Definition 3

We say that state $j \in S$ is **accessible** from state $i \in S$ if

$$(\mathbf{P}^n)_{ij} > 0 \text{ for some } n \geq 0.$$

We say that states i and j **communicate** if i is accessible from j, and j is accesible from i (notation: $i \leftrightarrow j$).

Theorem 4

Communication is an equivalence relation, which means it satisfies the following three properties:

- (**reflexive**) *every state communicates with itself,*
- (**symmetric**) *if i communicates with j, then j communicates with i,*
- (**transitive**) *if i communicates with j, and j communicates with k, then i communicates with k.*

Since communication is an equivalence relation, the state space S can be partitioned into equivalence classes, called **communication classes**. That is, the state space can be divided into disjoint subsets, each of whose states communicate with each other but do not communicate with any states outside their class.

Communication - cont'd

Example 2

Find the communication classes for the Markov chain with the state space $S = \{a, b, c, d, e\}$ *and transition matrix*

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

Irreducibility

Definition 5

A Markov chain is called **irreducible** if it has exactly one communication class. That is, all states communicate with each other.

Example 3

Are the Markov chains with the following transition matrices irreducible?

$$(a) \quad \mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \end{bmatrix} \qquad (b) \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 2/3 \\ 0 & 2/3 & 1/3 \end{bmatrix}$$

Recurrence & transience

Recurrence and transience

Example 4

Consider a transition matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

of a Markov chain on $S = \{1, 2, 3\}$.

1. *Find the communication classes.*
2. *What is the probability for the chain started in i to revisit i, where $i = 1, 2, 3$?*

Recurrence and transience - cont'd

It turns out that the states of a Markov chain, as this example illustrates, exhibit one of two contrasting behaviors. For the chain started in a given state, the chain either

- revisits that state, with probability 1, or
- there is a positive probability that the chain will never revisit that state.

Recurrence and transience - cont'd

Given a Markov chain $(X_n)_{n \in \mathbb{N}_0}$, let

$$T_j = \min\{n > 0 : X_n = j\}$$

be the **first passage time to state** j. Let

$$f_j = \mathbb{P}(T_j < \infty | X_0 = j)$$

be the probability that the chain started in j eventually returns to j.

Definition 6

State j is said to be **recurrent** if $f_j = 1$. It is said to be **transient** if $f_j < 1$.

So if j is recurrent, then the Markov chain started in j eventually revisits j. If j is transient, then there is a positive probability that the chain started in j never returns to j.

Recurrence and transience - cont'd

Whether or not a state is eventually revisited is strongly related to how often that state is visited.
For the chain started in i, let

$$I_n = \begin{cases} 1, & \text{if } X_n = j \\ 0, & \text{otherwise.} \end{cases}$$

Then $\sum_{n=0}^{\infty} I_n$ is the number of visits to j (of the chain started in i). One can show that the **expected number of visits to j** (for the chain started in i) is

$$\mathbb{E}\left(\sum_{n=0}^{\infty} I_n\right) = \sum_{n=0}^{\infty} (\mathbf{P}^n)_{ij},$$

where the infinite sum may possibly diverge to $+\infty$.

Recurrence and transience - cont'd

One can prove that

- state j is recurrent if and only if

$$\sum_{n=0}^{\infty} (\mathbf{P}^n)_{jj} = \infty$$

  (that is, if the expected number of visits to j of the chain started in j, is infinite);

- state j is transient if and only if

$$\sum_{n=0}^{\infty} (\mathbf{P}^n)_{jj} < \infty$$

  (that is, if the expected number of visits to j of the chain started in j, is finite).

Recurrence and transience - cont'd

One can also prove that

- the states of a communication class are either all recurrent or all transient,
- for a finite irreducible Markov chain, all states are recurrent (this is not true for infinite chains).

Limit theorem for finite irreducible Markov chains

Limit theorem for finite irreducible Markov chains

Recall that $T_j = \min\{n > 0 : X_n = j\}$ is the first passage time to state j. For each state j let

$$\mu_j = \mathbb{E}\left[\, T_j \,|\, X_0 = j\right]$$

be **the expected return time to** j.

Theorem 7

Assume $(X_n)_{n\in\mathbb{N}_0}$ is a finite irreducible Markov chain.
Then

- *there exists a unique positive stationary distribution π,*
- *the expected return time to j is finite for every $j \in S$,*
- *for every $j \in S$*

$$(\pi)_j = \frac{1}{\mu_j}.$$

The theorem does not assert that there exists a limiting distribution. Therefore, under the assumptions of the theorem *we cannot* conclude that

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix},$$

and hence we can't say that

$$\lim_{n\to\infty} \alpha \mathbf{P}^n = \pi$$

for all initial distributions α.

Nevertheless, the theorem is useful as it allows to find the expected return times to states (by finding the stationary distribution).

Finding the expected return time

Example 5

Consider a Markov chain with state space $S = \{a, b, c\}$ *and transition matrix*

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

Compute the expected return times $\mathbb{E}\left[\, T_x \middle| X_0 = x\right]$ *for* $x = a, b, c$.

Another interpretation of the expected return time

Assume that the chain is irreducible so every state is recurrent. The chain started at j will eventually revisit j with probability 1. Once it hits j, the chain begins anew and behaves as if a new version of the chain started at j. We say the Markov chain *regenerates itself*. (This intuitive behavior is known as **the strong Markov property**.)

This implies that the expected return time to a given state can also be interpreted more generally as **the expected time between visits to this state**.

Periodicity

Periodicity

Finite irreducible Markov chains have unique, positive stationary distributions. However, they may not have limiting distributions.

Example 6

If

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

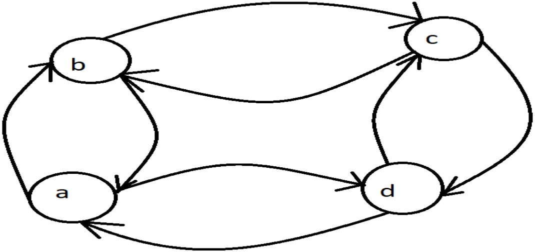
then Markov chain $(X_n)_n$ *on* $S = \{1, 2\}$

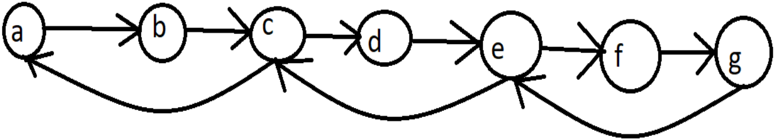
- *is irreducible,*
- *has unique stationary distribution* $\pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} > 0$*,*
- *has no limiting distribution, since the chain flip-flops back and forth between* 1 *and* 2*.*

It is precisely the finite irreducible Markov chains that do not exhibit this type of periodic behavior, which have limiting distributions.

Periodicity - cont'd

For a Markov chain started in state i, let us consider the set of times when the chain can return to i. See the following graphs:





Periodicity - cont'd

For the chain described by the first graph:

- the set of possible return times from any state is $\{2, 4, 6, 8, \ldots\}$;
- the chain started from any state returns to that state in *multiples of two* steps.

For the chain described by the second graph:

- the set of possible return times from any state is $\{3, 6, 9, 12, \ldots\}$;
- the chain started from any state returns to that state in *multiples of three* steps.

Periodicity - cont'd

Definition 8

Period of state $i \in S$, denoted $d(i)$, is the greatest common divisor of the set of possible return times to i:

$$d(i) = \gcd\{n > 0 : (\mathbf{P}^n)_{ii} > 0\}.$$

If $d(i) = 1$ then state i is said to be **aperiodic**. If the set above is empty, we define $d(i) = +\infty$.

Recall that the greatest common divisor (gcd) of a set of positive integers is the largest integer that divides all the numbers of the set without a remainder.

Periodicity - cont'd

- Thus returns from a state $i \in S$ to i can only occur in *multiples of* $d(i)$ steps.
- One can prove that periodicity, similar to recurrence and transience, is a class property, which means that the states of a communication class all have the same period.
- This implies that all states in an irreducible Markov chain have the same period.

Definition 9

A Markov chain is **periodic** if it is irreducible and all states have period greater than 1. A Markov chain is **aperiodic** if it is irreducible and all states have period equal to 1.

Periodicity - a sufficient condition

- Note that any state i with the property that $(\mathbf{P})_{ii} > 0$ is necessarily aperiodic.
- Thus, a sufficient condition for an *irreducible* Markov chain to be aperiodic is that $(\mathbf{P})_{ii} > 0$ for some $i \in S$.
- Clearly, this is not a necessary condition.

Limit theorem for finite irreducible and aperiodic Markov chains

Limit theorem for finite irreducible and aperiodic Markov chains

Theorem 10

Assume $(X_n)_n$ is a finite, irreducible and aperiodic Markov chain. Then

- *there exists a unique positive stationary distribution π,*
- *the expected return time to j is finite for every $j \in S$,*
- *for every $j \in S$, $(\pi)_j = 1/\mu_j$,*
- *π is the limiting distribution of the chain, that is*

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix}, \ \textit{and} \ \lim_{n\to\infty} \alpha\mathbf{P}^n = \pi$$

for any initial distribution α.

This means that it is precisely the class of finite, irreducible and aperiodic Markov chains that have positive limiting distributions.

Recall that we have stated the same assertion for the class of Markov chains with *regular* transition matrices. It can be proved that a finite Markov chain is irreducible and aperiodic if and only if its transition matrix is regular.