Markov Chains: An Introduction/Review

David Sirl
dsirl@maths.uq.edu.au
http://www.maths.uq.edu.au/~dsirl/

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

Modified and shortened by Longin Jan Latecki
latecki@temple.edu

Andrei A. Markov (1856 – 1922)

Random Processes

A random process is a collection of random variables indexed by some set I, taking values in some set S.

- I is the index set, usually time, e.g. $\mathbb{Z}^+$, $\mathbb{R}$, $\mathbb{R}^+$.
- S is the state space, e.g. $\mathbb{Z}^+$, $\mathbb{R}^n$, $\{1, 2, \ldots, n\}$, $\{a, b, c\}$.

We classify random processes according to both the index set (discrete or continuous) and the state space (finite, countable or uncountable/continuous).

Markov Processes

- A random process is called a *Markov Process* if, conditional on the current state of the process, its future is independent of its past.
- More formally, $X(t)$ is Markovian if has the following property:

$$\begin{aligned} &\mathbb{P}(X(t_n) = j_n \,|\, X(t_{n-1}) = j_{n-1} \,, \, \ldots \,, \, X(t_1) = j_1) \\ = \; &\mathbb{P}(X(t_n) = j_n \,|\, X(t_{n-1}) = j_{n-1}) \end{aligned}$$

for all finite sequences of times $t_1 < \ldots < t_n \in I$ and of states $j_1, \ldots, j_n \in S$.

Time Homogeneity

A Markov chain $(X(t))$ is said to be *time-homogeneous* if

$$\mathbb{P}(X(s+t) = j \mid X(s) = i)$$

is independent of s. When this holds, putting $s = 0$ gives

$$\mathbb{P}(X(s+t) = j \mid X(s) = i) = \mathbb{P}(X(t) = j \mid X(0) = i).$$

Time Homogeneity

A Markov chain $(X(t))$ is said to be *time-homogeneous* if

$$\mathbb{P}(X(s+t) = j \,|\, X(s) = i)$$

is independent of s. When this holds, putting $s = 0$ gives

$$\mathbb{P}(X(s+t) = j \,|\, X(s) = i) = \mathbb{P}(X(t) = j \,|\, X(0) = i).$$

Probabilities depend on elapsed time, not absolute time.

Discrete-time Markov chains

- At time epochs $n = 1, 2, 3, \ldots$ the process changes from one state i to another state j with probability p_{ij}.

Discrete-time Markov chains

- At time epochs $n = 1, 2, 3, \ldots$ the process changes from one state i to another state j with probability p_{ij}.
- We write the *one-step transition matrix* $P = (p_{ij},\ i, j \in S)$.

Discrete-time Markov chains

- At time epochs $n = 1, 2, 3, \ldots$ the process changes from one state i to another state j with probability p_{ij}.
- We write the *one-step transition matrix* $P = (p_{ij},\ i, j \in S)$.
- Example: a frog hopping on 3 rocks. Put $S = \{1, 2, 3\}$.

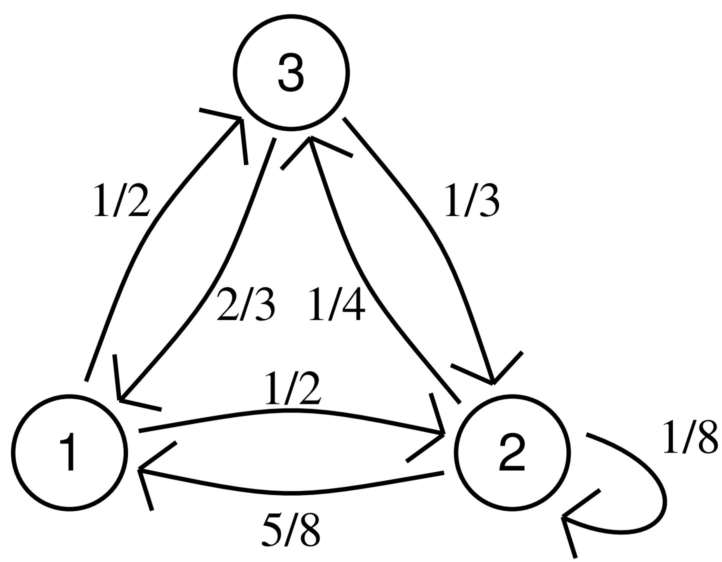
$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

DTMC example

- Example: A frog hopping on 3 rocks. Put $S = \{1, 2, 3\}$.

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

- We can gain some insight by drawing a picture:



DTMCs: n-step probabilities

- We have P, which tells us what happens over one time step; lets work out what happens over two time steps:

$$
\begin{aligned}
p_{ij}^{(2)} &= \mathbb{P}(X_2 = j \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_1 = k \mid X_0 = i)\, \mathbb{P}(X_2 = j \mid X_1 = k\,,\, X_0 = i) \\
&= \sum_{k \in S} p_{ik} p_{kj}.
\end{aligned}
$$

$$= \sum_k P(X_2 = j, X_1 = k \mid X_0 = i)$$

DTMCs: n-step probabilities

- We have P, which tells us what happens over one time step; lets work out what happens over two time steps:

$$
\begin{aligned}
p_{ij}^{(2)} &= \mathbb{P}(X_2 = j \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_1 = k \mid X_0 = i)\, \mathbb{P}(X_2 = j \mid X_1 = k\,,\, X_0 = i) \\
&= \sum_{k \in S} p_{ik} p_{kj}.
\end{aligned}
$$

- So $P^{(2)} = PP = P^2$.

DTMCs: n-step probabilities

- We have P, which tells us what happens over one time step; lets work out what happens over two time steps:

$$
\begin{aligned}
p_{ij}^{(2)} &= \mathbb{P}(X_2 = j \,|\, X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_1 = k \,|\, X_0 = i)\, \mathbb{P}(X_2 = j \,|\, X_1 = k\,,\, X_0 = i) \\
&= \sum_{k \in S} p_{ik} p_{kj}.
\end{aligned}
$$

- So $P^{(2)} = PP = P^2$.
- Similarly, $P^{(3)} = P^2 P = P^3$ and $P^{(n)} = P^n$.

DTMC: Arbitrary initial distributions

- We may wish to start the chain according to some initial distribution $\pi^{(0)}$.

DTMC: Arbitrary initial distributions

- We may wish to start the chain according to some initial distribution $\pi^{(0)}$.
- We can then calculate the state probabilities $\pi^{(n)} = (\pi_j^{(n)},\ j \in S)$ of being in state j at time n as follows:

$$\pi_j^{(n)} = \sum_{k \in S} \mathbb{P}(X_0 = k)\, \mathbb{P}(X_n = j \,|\, X_0 = k)$$

$$= \sum_{k \in S} \pi_j^{(0)} p_{ij}^{(n)}.$$

DTMC: Arbitrary initial distributions

- We may wish to start the chain according to some initial distribution $\pi^{(0)}$.
- We can then calculate the state probabilities $\pi^{(n)} = (\pi_j^{(n)},\ j \in S)$ of being in state j at time n as follows:

$$\begin{aligned}\pi_j^{(n)} &= \sum_{k\in S} \mathbb{P}(X_0 = k)\, \mathbb{P}(X_n = j \,|\, X_0 = k) \\ &= \sum_{k\in S} \pi_j^{(0)} p_{ij}^{(n)}.\end{aligned}$$

- Or, in matrix notation, $\pi^{(n)} = \pi^{(0)} P^n$; similarly we can show that $\pi^{(n+1)} = \pi^{(n)} P$.

Class structure

- We say that a state i leads to j (written $i \to j$) if it is possible to get from i to j in some finite number of jumps: $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

Class structure

- We say that a state i leads to j (written $i \to j$) if it is possible to get from i to j in some finite number of jumps: $p_{ij}^{(n)} > 0$ for some $n \geq 0$.
- We say that i communicates with j (written $i \leftrightarrow j$) if $i \to j$ and $j \to i$.

Class structure

- We say that a state i leads to j (written $i \to j$) if it is possible to get from i to j in some finite number of jumps: $p_{ij}^{(n)} > 0$ for some $n \geq 0$.
- We say that i communicates with j (written $i \leftrightarrow j$) if $i \to j$ and $j \to i$.
- The relation $\leftrightarrow$ partitions the state space into *communicating classes*.

Class structure

- We say that a state i leads to j (written $i \to j$) if it is possible to get from i to j in some finite number of jumps: $p_{ij}^{(n)} > 0$ for some $n \geq 0$.
- We say that i communicates with j (written $i \leftrightarrow j$) if $i \to j$ and $j \to i$.
- The relation $\leftrightarrow$ partitions the state space into *communicating classes*.
- We call the state space *irreducible* if it consists of a single communicating class.

Class structure

- We say that a state i leads to j (written $i \to j$) if it is possible to get from i to j in some finite number of jumps: $p_{ij}^{(n)} > 0$ for some $n \geq 0$.
- We say that i communicates with j (written $i \leftrightarrow j$) if $i \to j$ and $j \to i$.
- The relation $\leftrightarrow$ partitions the state space into *communicating classes*.
- We call the state space *irreducible* if it consists of a single communicating class.
- These properties are easy to determine from a transition probability graph.

Classification of states

- We call a state i *recurrent* or *transient* according as $\mathbb{P}(X_n = i$ for infinitely many $n)$ is equal to one or zero.

Classification of states

- We call a state i *recurrent* or *transient* according as $\mathbb{P}(X_n = i$ for infinitely many $n)$ is equal to one or zero.
  - A recurrent state is a state to which the process always returns.
  - A transient state is a state which the process eventually leaves for ever.

Classification of states

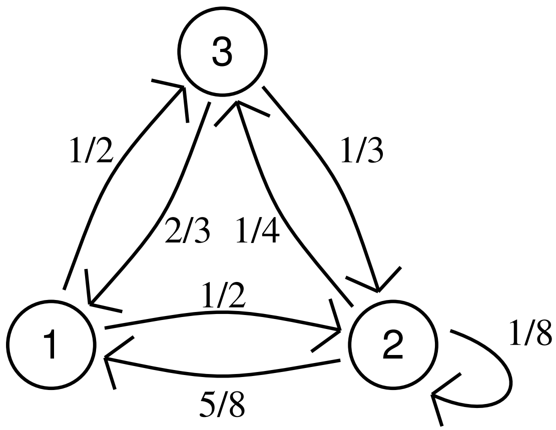
- We call a state i *recurrent* or *transient* according as $\mathbb{P}(X_n = i$ for infinitely many $n)$ is equal to one or zero.
  - A recurrent state is a state to which the process always returns.
  - A transient state is a state which the process eventually leaves for ever.
- Recurrence and transience are class properties; i.e. if two states are in the same communicating class then they are recurrent/transient together.
- We therefore speak of recurrent or transient classes

Classification of states

- We call a state i *recurrent* or *transient* according as $\mathbb{P}(X_n = i$ for infinitely many $n)$ is equal to one or zero.
  - A recurrent state is a state to which the process always returns.
  - A transient state is a state which the process eventually leaves for ever.
- Recurrence and transience are class properties; i.e. if two states are in the same communicating class then they are recurrent/transient together.
- We therefore speak of recurrent or transient classes
- We also assume throughout that no states are *periodic*.

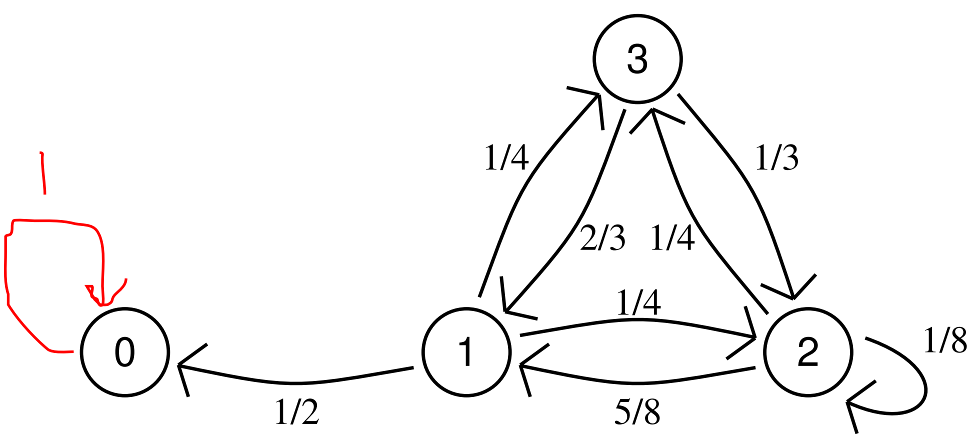
DTMCs: Two examples

- S irreducible:



$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

- $S = \{0\} \cup C$, where C is a transient class:



$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

DTMCs: Quantities of interest

Quantities of interest include:

- Hitting probabilities.
- Expected hitting times.
- Limiting (stationary) distributions.
- Limiting conditional (quasistationary) distributions.

DTMCs: Hitting probabilities

Let α_i be the probability of hitting state 1 starting in state i.

- Clearly $\alpha_1 = 1$; and for $i \neq 1$,

$$\begin{aligned}\alpha_i &= \mathbb{P}(\text{hit } 1 \,|\, \text{start in } i) \\ &= \sum_{k \in S} \mathbb{P}(X_1 = k \,|\, X_0 = i)\ \mathbb{P}(\text{hit } 1 \,|\, \text{start in } k) \\ &= \sum_{k \in S} p_{ik}\alpha_k\end{aligned}$$

DTMCs: Hitting probabilities

Let α_i be the probability of hitting state 1 starting in state i.

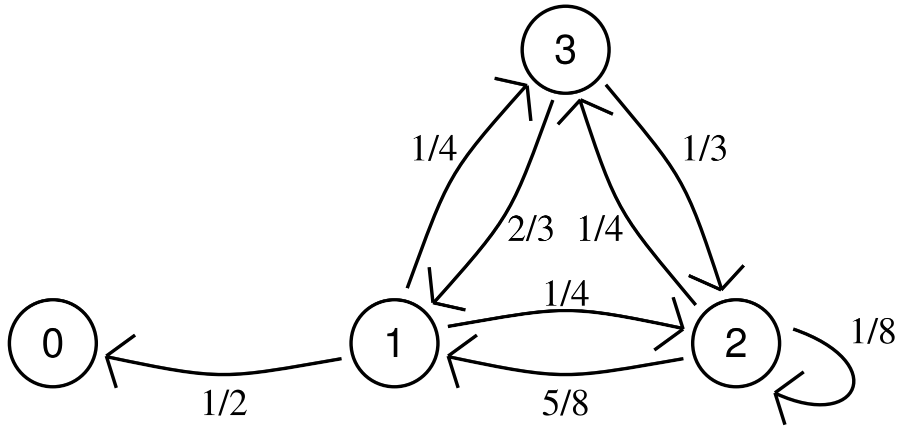
- Clearly $\alpha_1 = 1$; and for $i \neq 1$,

$$\begin{aligned}\alpha_i &= \mathbb{P}(\text{hit } 1 \,|\, \text{start in } i) \\ &= \sum_{k \in S} \mathbb{P}(X_1 = k \,|\, X_0 = i)\ \mathbb{P}(\text{hit } 1 \,|\, \text{start in } k) \\ &= \sum_{k \in S} p_{ik}\alpha_k\end{aligned}$$

- Sometimes there may be more than one solution $\alpha = (\alpha_i,\ i \in S)$ to this system of equations.

  If this is the case, then the hitting probabilites are given by the *minimal* such solution.

Example: Hitting Probabilities



$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

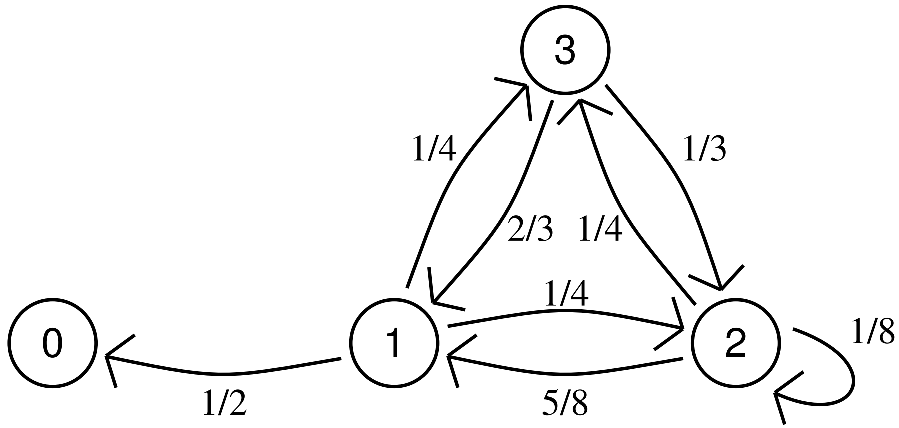
Let α_i be the probability of hitting state 3 starting in state i.

So $\alpha_3 = 1$ and $\alpha_i = \sum_k p_{ik}\alpha_k$:

$$\alpha_0 = \alpha_0$$
$$\alpha_1 = \tfrac{1}{2}\alpha_0 + \tfrac{1}{4}\alpha_2 + \tfrac{1}{4}\alpha_3$$
$$\alpha_2 = \tfrac{5}{8}\alpha_1 + \tfrac{1}{8}\alpha_2 + \tfrac{1}{4}\alpha_3$$

relation to Label Propagation

Example: Hitting Probabilities



$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Let α_i be the probability of hitting state 3 starting in state i.

$$\alpha = \begin{pmatrix} 0 \\ \frac{9}{23} \\ \frac{13}{23} \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0.39 \\ 0.57 \\ 1 \end{pmatrix}.$$

DTMCs: The Limiting Distribution

Assume that the state space is irreducible, aperiodic and recurrent.

- What happens to the state probabilities $\pi_j^{(n)}$ as $n \to \infty$?

DTMCs: The Limiting Distribution

Assume that the state space is irreducible, aperiodic and recurrent.

- What happens to the state probabilities $\pi_j^{(n)}$ as $n \to \infty$?
- We know that $\pi^{(n+1)} = \pi^{(n)} P$.

DTMCs: The Limiting Distribution

Assume that the state space is irreducible, aperiodic and recurrent.

- What happens to the state probabilities $\pi_j^{(n)}$ as $n \to \infty$?
- We know that $\pi^{(n+1)} = \pi^{(n)} P$.
- So if there is a limiting distribution π, it must satisfy

$$\pi = \pi P \qquad \text{(and } \textstyle\sum_i \pi_i = 1\text{)}.$$

(Such a distribution is called *stationary*.)

DTMCs: The Limiting Distribution

Assume that the state space is irreducible, aperiodic and recurrent.

- What happens to the state probabilities $\pi_j^{(n)}$ as $n \to \infty$?
- We know that $\pi^{(n+1)} = \pi^{(n)} P$.
- So if there is a limiting distribution π, it must satisfy

$$\pi = \pi P \qquad (\text{and } \textstyle\sum_i \pi_i = 1).$$

  (Such a distribution is called *stationary*.)
- This limiting distribution does not depend on the initial distribution.

DTMCs: The Limiting Distribution

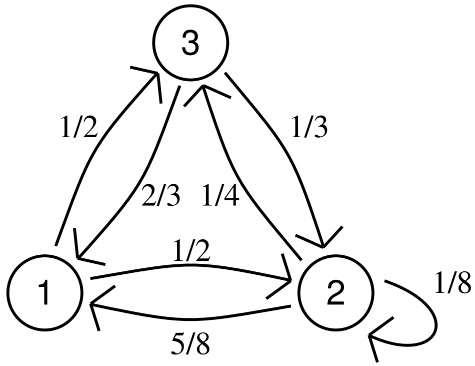
Assume that the state space is irreducible, aperiodic and recurrent.

- What happens to the state probabilities $\pi_j^{(n)}$ as $n \to \infty$?
- We know that $\pi^{(n+1)} = \pi^{(n)} P$.
- So if there is a limiting distribution π, it must satisfy

$$\pi = \pi P \qquad \text{(and } \sum_i \pi_i = 1\text{)}.$$

  (Such a distribution is called *stationary*.)
- This limiting distribution does not depend on the initial distribution.
- When the state space is infinite, it may happen that $\pi_j^{(n)} \to 0$ for all j.

Example: The Limiting Distribution



$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Substituting P into $\pi = \pi P$ gives

relation to PageRank

$$\pi_1 = \tfrac{5}{8}\pi_2 + \tfrac{2}{3}\pi_3,$$
$$\pi_2 = \tfrac{1}{2}\pi_1 + \tfrac{1}{8}\pi_2 + \tfrac{1}{3}\pi_3,$$
$$\pi_3 = \tfrac{1}{2}\pi_1 + \tfrac{1}{4}\pi_2,$$

which together with $\sum_i \pi_i = 1$ yields

$$\pi = \begin{pmatrix} \frac{38}{97} & \frac{32}{97} & \frac{27}{97} \end{pmatrix} \approx \begin{pmatrix} 0.39 & 0.33 & 0.28 \end{pmatrix}.$$

DTMCs: The Limiting Conditional Dist'n

Assume that the state space is consists of an absorbing state and a transient class ($S = \{0\} \cup C$).

- The limiting distribution is $(1, 0, 0, \ldots)$.

DTMCs: The Limiting Conditional Dist'n

Assume that the state space is consists of an absorbing state and a transient class ($S = \{0\} \cup C$).

- The limiting distribution is $(1, 0, 0, \ldots)$.
- Instead of looking at the limiting behaviour of

$$\mathbb{P}(X_n = j \,|\, X_0 = i) = p_{ij}^{(n)},$$

we need to look at

$$\mathbb{P}(X_n = j \,|\, X_n \neq 0\,,\ X_0 = i) = \frac{p_{ij}^{(n)}}{1 - p_{i0}^{(n)}}$$

for $i, j \in C$.

DTMCs: The Limiting Conditional Dist'n

- It turns out we need a solution $m = (m_i,\ i \in C)$ of

$$mP_C = rm,$$

for some $r \in (0, 1)$.

DTMCs: The Limiting Conditional Dist'n

- It turns out we need a solution $m = (m_i,\ i \in C)$ of

$$mP_C = rm,$$

  for some $r \in (0, 1)$.
- If C is a finite set, there is a unique such r.

DTMCs: The Limiting Conditional Dist'n

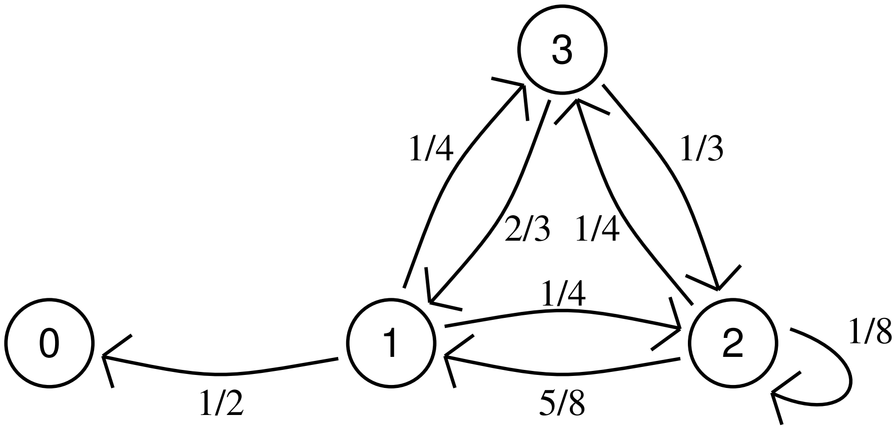
- It turns out we need a solution $m = (m_i,\ i \in C)$ of

$$mP_C = rm,$$

  for some $r \in (0, 1)$.

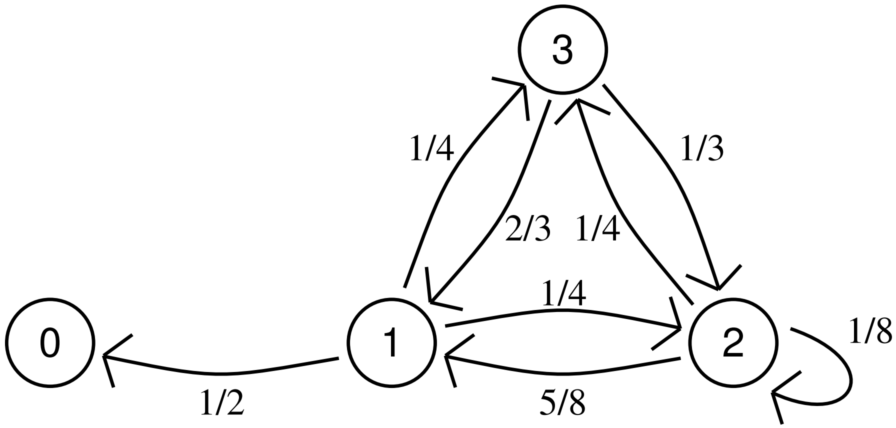
- If C is a finite set, there is a unique such r.
- If C is infinite, there is $r^* \in (0, 1)$ such that all r in the interval $[r^*, 1)$ are admissible; and the solution corresponding to $r = r^*$ is the LCD.

Example: Limiting Conditional Dist'n



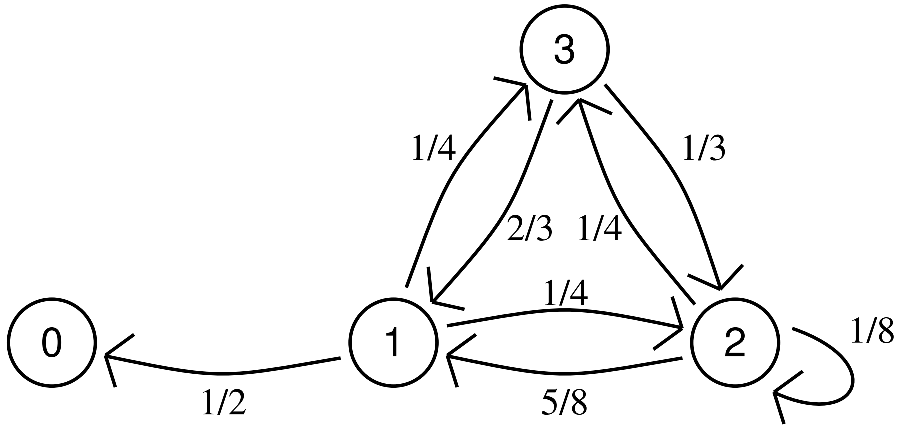
$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Example: Limiting Conditional Dist'n



$$P_C = \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Example: Limiting Conditional Dist'n



$$P_C = \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Solving $mP_C = rm$, we get

$r_1 \approx 0.773$ and $m \approx (0.45,\ 0.30,\ 0.24)$

DTMCs: Summary

From the one-step transition probabilities we can calculate:

- n-step transition probabilities,
- hitting probabilities,
- expected hitting times,
- limiting distributions, and
- limiting conditional distributions.