Markov String: X_1^n **:** $p\left(x_1^n\right)=p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_k|x_{k-1}\right)p\left(x_{k+1}|x_k\right)\cdots p\left(x_n|x_{n-1}\right)$

- $\equiv$ For all k, $1<k\le n$, $p\left(x_k\middle|x_1^{k-1}\right)=p\left(x_k|x_{k-1}\right)$.
- Ordered substring of Markov string is Markov:
  For $0\le n_1<n_2<\cdots<n_k\le n$, $\left(X_{n_1},X_{n_2},\ldots,X_{n_k}\right)$ is also a Markov string.
- Given X_k (the present), we have X_1^{k-1} (the past) and X_{k+1}^n (the future) are independent:
  $p\left(x_1^{k-1},x_{k+1}^n|x_k\right)=p\left(x_{k+1}^n|x_k\right)p\left(x_1^{k-1}|x_k\right)$.
  - Given X_k, we have X_{k-1} and X_{k+1} are independent: $p\left(x_{k-1},x_{k+1}|x_k\right)=p\left(x_{k-1}|x_k\right)p\left(x_{k+1}|x_k\right)$.
  - For $0\le m_1<m_2<\cdots<m_r<k<m_{r+1}<\cdots m_{r+j}\le n$,
    $p\left(x_{m_1},\ldots,x_{m_r},x_{m_{r+1}},\ldots,x_{m_{r+j}}|x_k\right)=p\left(x_{m_1},\ldots,x_{m_r}|x_k\right)p\left(x_{m_{r+1}},\ldots,x_{m_{r+j}}|x_k\right)$.
- For $1\le k<n$, and $k+1<m\le n$, $p\left(x_{k+1}^m|x_k\right)=p\left(x_{k+1}|x_k\right)\cdots p\left(x_m|x_{m-1}\right)$.
- $p\left(x_{k+1}^m|x_j^k\right)=p\left(x_{k+1}^m|x_k\right)$
- Reverse Markov string is also a Markov string
  $p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_{k+1}|x_k\right)\cdots p\left(x_n|x_{n-1}\right)=p\left(x_n\right)p\left(x_{n-1}|x_n\right)\cdots p\left(x_k|x_{k+1}\right)\cdots p\left(x_1|x_2\right)$
- Ordered grouped Markov string is still a Markov string: Let $k_0=0$, $1\le k_1<k_2<\cdots<k_m=n$. For $1\le\ell\le m$, define $\bar{y}_\ell=\left(x_{k_{\ell-1}+1},x_{k_{\ell-1}+1},\ldots,x_{k_\ell}\right)$. Then, $\bar{y}_1^m$ is a Markov string.

Proof of equivalent statement:

"⇒"

$$p\left(x_k\middle|x_1^{k-1}\right)=\frac{p\left(x_1^k\right)}{p\left(x_1^{k-1}\right)}=\frac{\cancel{p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_{k-2}|x_{k-1}\right)}\,p\left(x_k|x_{k-1}\right)}{\cancel{p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_{k-2}|x_{k-1}\right)}}=p\left(x_k|x_{k-1}\right).$$

"⇒"

$$p\left(x_1^n\right)=p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_n\middle|x_1^{n-1}\right)=p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_n|x_{n-1}\right).$$

Proof of properties

$p\left(x_1^n\right)=p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_k|x_{k-1}\right)p\left(x_{k+1}|x_k\right)\cdots p\left(x_n|x_{n-1}\right)$, then,

a) $p\left(x_1^k\right)=p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_k|x_{k-1}\right)$ for all $0<k\le n$. (X_1^k is also a Markov string)

Proof. $$p\left(x_1^{n-1}\right)=\sum_{x_n}p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_{n-1}|x_{n-2}\right)p\left(x_n|x_{n-1}\right)$$

$$=p\left(x_1\right)p\left(x_2|x_1\right)\cdots p\left(x_{n-1}|x_{n-2}\right)\overbrace{\cancel{\sum_{x_n}p\left(x_n|x_{n-1}\right)}}^{1}$$

By induction, we have $p\left(x_1^k\right)=p\left(x_1^{n-(n-k)}\right)=p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_k|x_{k-1}\right)$.

b) $p\left(x_k^n\right)=p(x_k)\,p\left(x_{k+1}|x_k\right)\cdots p\left(x_n|x_{n-1}\right)$ for all $1\le k<n$. (X_k^n is also a Markov string)

Proof. $p\left(x_2^n\right)=\sum_{x_1} p(x_1)\,p\left(x_2|x_1\right)p\left(x_3|x_2\right)\cdots p\left(x_n|x_{n-1}\right)$

$$=\left(\sum_{x_1} p(x_1)\,p\left(x_2|x_1\right)\right)p\left(x_3|x_2\right)\cdots p\left(x_n|x_{n-1}\right)$$

$$=\left(\sum_{x_1} p(x_1,x_2)\right)p\left(x_3|x_2\right)\cdots p\left(x_n|x_{n-1}\right)=p(x_2)\,p\left(x_3|x_2\right)\cdots p\left(x_n|x_{n-1}\right)$$

By induction, we have $p\left(x_k^n\right)=p(x_k)\,p\left(x_{k+1}|x_k\right)\cdots p\left(x_n|x_{n-1}\right)$.

c) $p\left(x_k^j\right)=p(x_k)\,p\left(x_{k+1}|x_k\right)\cdots p\left(x_j|x_{j-1}\right)$ $1\le k<j\le n$ (X_k^j is also a Markov string)

Proof. *a*) and *b*).

d) $p\left(x_1^{k-1},x_{k+1}^n\right)=p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_{k-1}|x_{k-2}\right)p\left(x_{k+1}|x_{k-1}\right)p\left(x_{k+2}|x_{k+1}\right)\cdots p\left(x_n|x_{n-1}\right)$.

Proof.

$$p\left(x_1^{k-1},x_{k+1}^n\right)$$

$$=\sum_{x_k} p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_k|x_{k-1}\right)p\left(x_{k+1}|x_k\right)\cdots p\left(x_n|x_{n-1}\right)$$

$$=p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_{k-1}|x_{k-2}\right)\left(\sum_{x_k} p\left(x_k|x_{k-1}\right)p\left(x_{k+1}|x_k\right)\right)p\left(x_{k+2}|x_{k+1}\right)\cdots p\left(x_n|x_{n-1}\right)$$

$$=p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_{k-1}|x_{k-2}\right)\left(\sum_{x_k} p\left(x_k|x_{k-1}\right)p\left(x_{k+1}|x_k,x_{k-1}\right)\right)p\left(x_{k+2}|x_{k+1}\right)\cdots p\left(x_n|x_{n-1}\right)$$

$$=p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_{k-1}|x_{k-2}\right)\left(\sum_{x_k} p\left(x_k,x_{k+1}|x_{k-1}\right)\right)p\left(x_{k+2}|x_{k+1}\right)\cdots p\left(x_n|x_{n-1}\right)$$

$$=p(x_1)\,p\left(x_2|x_1\right)\cdots p\left(x_{k-1}|x_{k-2}\right)p\left(x_{k+1}|x_{k-1}\right)p\left(x_{k+2}|x_{k+1}\right)\cdots p\left(x_n|x_{n-1}\right)$$

e) Ordered substring of Markov string is Markov.

Proof. *c*) and *d*). Use *c*) to cut the beginning and the end part of the string. Use *d*) to take any amount from the middle.

f) Given x_k, we have x_{k-1} and x_{k+1} are independent: $p\left(x_{k-1},x_{k+1}|x_k\right)=p\left(x_{k-1}|x_k\right)p\left(x_{k+1}|x_k\right)$

Proof. $p\left(x_{k-1},x_{k+1}|x_k\right)=\dfrac{p\left(x_{k-1},x_k,x_{k+1}\right)}{p(x_k)}=\dfrac{p(x_{k-1})\,p\left(x_k|x_{k-1}\right)p\left(x_{k+1}|x_k\right)}{p(x_k)}$

$$=\frac{p\left(x_{k-1},x_k\right)}{p(x_k)}p\left(x_{k+1}|x_k\right)=p\left(x_{k-1}|x_k\right)p\left(x_{k+1}|x_k\right)$$

g) For $j<k<m$, $p\left(x_j, x_m \middle| x_k\right) = p\left(x_j \middle| x_k\right) p\left(x_m \middle| x_k\right)$.

Proof. From *e*), $\left(Y_1, Y_2, Y_3\right) = \left(X_j, X_k, X_m\right)$ is a Markov string. Apply *f*) to complete the proof.

h) For $1 \le k < n$, and $k+1 < m \le n$, $p\left(x_{k+1}^m \middle| x_k\right) = p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_m \middle| x_{m-1}\right)$.

Proof. From *e*), X_k^m is a Markov string. So, $p\left(x_k^m\right) = p\left(x_k\right) p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_m \middle| x_{m-1}\right)$. And therefore,

$$p\left(x_{k+1}^m \middle| x_k\right) = \frac{p\left(x_k^m\right)}{p\left(x_k\right)} = p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_m \middle| x_{m-1}\right).$$

i) $p\left(x_{k+1}^m \middle| x_j^k\right) = p\left(x_{k+1}^m \middle| x_k\right)$.

Proof. $$p\left(x_{k+1}^m \middle| x_j^k\right) = \frac{p\left(x_j^m\right)}{p\left(x_j^k\right)} = \frac{\cancel{p\left(x_j\right)}\, \cancel{p\left(x_{j+1} \middle| x_j\right)} \cdots \cancel{p\left(x_k \middle| x_{k-1}\right)}\, p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_m \middle| x_{m-1}\right)}{\cancel{p\left(x_j\right)}\, \cancel{p\left(x_{j+1} \middle| x_j\right)} \cdots \cancel{p\left(x_k \middle| x_{k-1}\right)}}$$

$$= p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_m \middle| x_{m-1}\right) = p\left(x_{k+1}^m \middle| x_k\right)$$

, where the last equality applies *h*) directly.

j) For $j+1<k<m-1$, $p\left(x_j^{k-1}, x_{k+1}^m \middle| x_k\right) = p\left(x_j^{k-1} \middle| x_k\right) p\left(x_{k+1}^m \middle| x_k\right)$.

Proof. $$p\left(x_j^{k-1}, x_{k+1}^m \middle| x_k\right) = \frac{p\left(x_j^{k-1}, x_k, x_{k+1}^m\right)}{p\left(x_k\right)} = \frac{p\left(x_j^{k-1}, x_k\right)}{p\left(x_k\right)} p\left(x_{k+1}^m \middle| x_j^{k-1}, x_k\right)$$

$$= p\left(x_j^{k-1} \middle| x_k\right) p\left(x_{k+1}^m \middle| x_k\right)$$

, where the last equality uses *i*).

ℓ) For $0 \le m_1 < m_2 < \cdots < m_r < k < m_{r+1} < \cdots m_{r+j} \le n$,

$$p\left(x_{m_1}, \ldots, x_{m_r}, x_{m_{r+1}}, \ldots, x_{m_{r+j}} \middle| x_k\right) = p\left(x_{m_1}, \ldots, x_{m_r} \middle| x_k\right) p\left(x_{m_{r+1}}, \ldots, x_{m_{r+j}} \middle| x_k\right).$$

Proof. Because $\left(x_{m_1}, \ldots, x_{m_r}, x_k, x_{m_{r+1}}, \ldots, x_{m_{r+j}}\right)$ is an ordered substring of a Markov string, it is therefore a Markov string. The property follows from *j*).

m) Reverse of a stationary Markov string is also a Markov string

$$p\left(x_1\right) p\left(x_2 \middle| x_1\right) \cdots p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_n \middle| x_{n-1}\right) = p\left(x_n\right) p\left(x_{n-1} \middle| x_n\right) \cdots p\left(x_k \middle| x_{k+1}\right) \cdots p\left(x_1 \middle| x_2\right)$$

Proof.

$$
\begin{aligned}
p\left(x_1^n\right) &= p\left(x_1\right) p\left(x_2 \middle| x_1\right) \cdots p\left(x_{k+1} \middle| x_k\right) \cdots p\left(x_n \middle| x_{n-1}\right) \\
&= \cancel{p\left(x_1\right)} \frac{p\left(x_1, x_2\right)}{\cancel{p\left(x_1\right)}} \cdots \frac{p\left(x_k, x_{k+1}\right)}{p\left(x_k\right)} \cdots \frac{p\left(x_{n-1}, x_n\right)}{p\left(x_{n-1}\right)} \\
&= \frac{p\left(x_1, x_2\right) \cdots p\left(x_k, x_{k+1}\right) \cdots p\left(x_{n-1}, x_n\right)}{p\left(x_2\right) \cdots p\left(x_{k+1}\right) \cdots p\left(x_{n-1}\right)} \frac{p\left(x_n\right)}{p\left(x_n\right)} \\
&= \frac{p\left(x_1, x_2\right)}{p\left(x_2\right)} \cdots \frac{p\left(x_k, x_{k+1}\right)}{p\left(x_{k+1}\right)} \cdots \frac{p\left(x_{n-1}, x_n\right)}{p\left(x_n\right)} p\left(x_n\right) \\
&= p\left(x_1 \middle| x_2\right) \cdots p\left(x_k \middle| x_{k+1}\right) \cdots p\left(x_{n-1} \middle| x_n\right) p\left(x_n\right)
\end{aligned}
$$

o) Ordered grouped Markov string is still a Markov string.

Let $k_0 = 0$, $1 \le k_1 < k_2 < \cdots < k_m = n$. For $1 \le \ell \le m$, define $\vec{y}_\ell = \left(x_{k_{\ell-1}+1}, x_{k_{\ell-1}+1}, \ldots, x_{k_\ell}\right)$. Then, $\vec{y}_1^m$ is a Markov string.

Proof. From *i*), we have 1) $P\left(\vec{y}_\ell \middle| \vec{y}_1^{\ell-1}\right) = P\left(x_{k_{\ell-1}+1}^{k_\ell} \middle| x_1^{k_{\ell-1}}\right) = P\left(x_{k_{\ell-1}+1}^{k_\ell} \middle| x_{k_{\ell-1}}\right)$ and 2) $P\left(\vec{y}_\ell \middle| \vec{y}_{\ell-1}\right) = P\left(x_{k_{\ell-1}+1}^{k_\ell} \middle| x_{k_{\ell-2}+1}^{k_{\ell-1}}\right) = P\left(x_{k_{\ell-1}+1}^{k_\ell} \middle| x_{k_{\ell-1}}\right)$. Hence, $P\left(\vec{y}_\ell \middle| \vec{y}_1^{\ell-1}\right) = P\left(\vec{y}_\ell \middle| \vec{y}_{\ell-1}\right)$.

MARKOV CHAINS (MC)

- $\forall i > 1 \;\; P\left(A_i \middle| \bigcap_{j<i} A_j\right) = P\left(A_i \middle| A_{i-1}\right)$
- $P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \prod_{k=2}^{n} P\left(A_k \middle| A_{k-1}\right)$
- **one-step transition probabilities** $\boxed{P_{i,j} = \Pr\left[\text{Next state is } j \middle| \text{Current state is } i\right] = p_{i \to j} = p_{i \to j} = p\left(j \middle| i\right)}$
  - generally, depend on the time (*n*) at which the step is taking place
- **transition probability matrix** $P = \left[P_{i,j}\right]$
  - To write *P*, start from the first row (currently in first sate), fill in each column of this first row with probability that it will go to other state. (So, use out arrow).
  - $\boxed{\text{Row sums} = 1}$. Column sums don't.
- X_k or $X(k)$
  - Discrete random variable
  - Denote the state of the chain at time k.
- $P_{i,j}(m,n) = \Pr\left[\text{MC will be in state j at time n} \middle| \text{MC is in state i at time m}\right]$

  $= \Pr\left[X(n) = j \middle| X(m) = \text{i}\right]$
- $P(m,n) = \left[P_{i,j}(m,n)\right]$

- $\{X(k)\}$ is a **Markov chain** if $\forall r \in \mathbb{N}/\{1\}$, for any r times (moments, instants) $k_1 < k_2 < \ldots < k_r$ and $\forall$ sequence $j_1, \ldots, j_r$ of states,

$$P\left[X(k_r) = j_r \middle| X(k_{r-1}) = j_{r-1}, X(k_{r-2}) = j_{r-2}, \ldots, X(k_1) = j_1\right] = P\left[X(k_r) = j_r \middle| X(k_{r-1}) = j_{r-1}\right]$$

  - "Given the present, the future becomes independent of the past."
- **Chapman-Kolmogorov Equations**
  - For time indices $m < u < n$, $\boxed{P(m,n) = P(m,u)P(u,n)}$

    Proof.

$$P_{i,j}(m,n) = P\left[X(n) = j \middle| X(m) = i\right] = \sum_k P\left(X(n) = j, X(u) = k \middle| X(m) = i\right)$$

k^{th} term in the sum is the probability of all paths from i at time m to j at time n that pass through k at time u

Use $P(A \cap B | C) = P(B|C)P(A|B \cap C)$.

$$P_{i,j}(m,n) = \sum_k P\left(X(n) = j \middle| X(u) = k, X(m) = i\right) P\left(X(u) = k \middle| X(m) = i\right)$$

$$= \sum_k P\left(X(n) = j \middle| X(u) = k\right) P\left(X(u) = k \middle| X(m) = i\right)$$

; from Markov property.

$$= \sum_k P_{k,j}(u,n) P_{i,k}(m,u) = \sum_k P_{i,k}(m,u) P_{k,j}(u,n)$$

- All transition matrices $P(m,n)$ governing $(n-m)$-step transition probabilities can be expressed in terms of 1-step transition matrices $\{P(k,k+1)\}$ in the case of MC.

$$P(m,n) = \prod_{k=m}^{n-1} P(k,k+1)$$

Proof $P(m,n) = P(m,m+1)P(m+1,n)$

$$= P(m,m+1)P(m+1,m+2)P(m+2,n)$$

$$= P(m,m+1)P(m+1,m+2)\cdots P(n-1,n)$$

- Let $\underline{p}(m) = (p_0(m), p_1(m), \ldots)$ be the row vector of state probabilities at time m. Then, $\underline{p}(n) = \underline{p}(m)P(m,n)$ [1]

Homogeneous MC with 1-step transition matrix *P*

- MC is **homogeneous** if all one-step transition matrices are all identical $\equiv \exists P \ \forall k \ P(k, k+1) = P$.

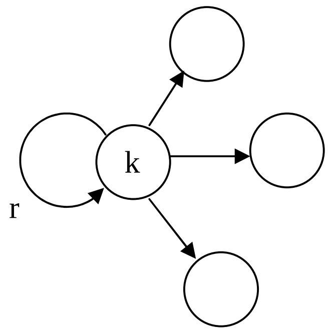
For homogeneous MC,

- $P(m,n) = P^{n-m}$

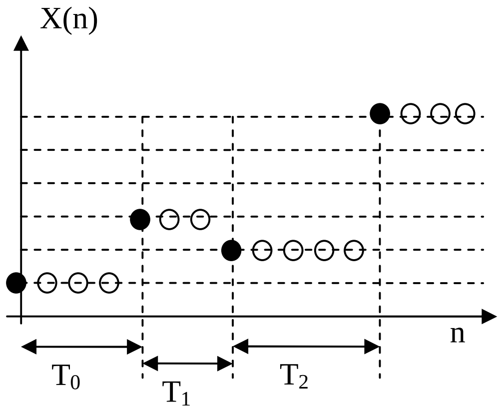
[1] Caution: $\underline{x}$ here means row vector

- The (i,j) entry in P^{n-m} is positive if and only if there is a path of length n-m from state i to state j.

- $\underline{p}(n) = \underline{p}(m) P^{n-m}$
- $\underline{p}(n) = \underline{p}(0) P^{n}$
- T = **holding time** in state k = the random duration of a stay in state k. $T \in \{1, 2, 3 \ldots\}$.

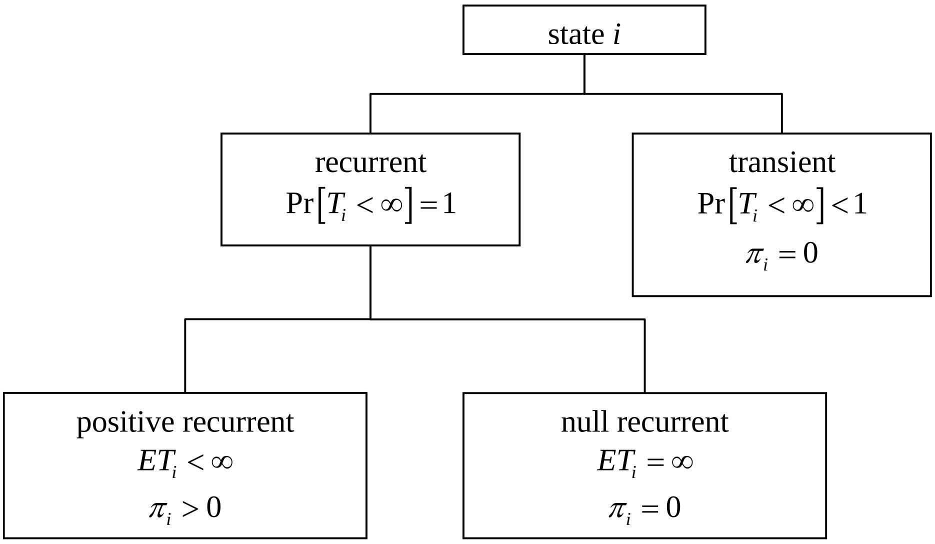


- Let r = probability, if any, on the state's self loop. T is geometrically distributed with parameter r.
- $\Pr[T = t]$
  = Pr[exactly t-1 traversals of self loop before leaving]
  = $(1-r) r^{t-1}$
- Holding time $\{T_k\}$ of a discrete-time homogeneous Markov chain.
  $T_0 = \min\{\ell \geq 1 : X(\ell) \neq X(0)\}$
  $T_k = \min\{\ell \geq 1 : X(T_0 + \ldots + T_{k-1} + \ell) \neq X(T_0 + \ldots + T_{k-1})\}$ for $k \geq 1$
  - $\{T_k\}$ are geometric random variable that are independent if one is given the sequence of states.



  - $\{T_k\}$ are not identically distributed because, in general, the parameter r of the geometric distribution varies with the state.

Classes of states



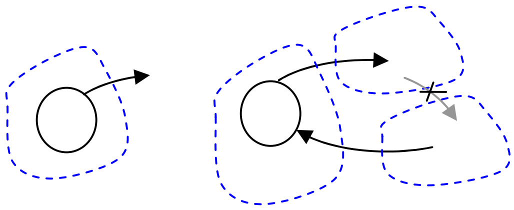
- **Def**:
  - If $p_{ij}(n) > 0$ for some integer $n \geq 1$, state j is **accessible** from state i, and we write $i \to j$.

    (There is a sequence of transitions from i to j that has nonzero probabilitiy.)
  - If $i \to j$ and $j \to i$, then states i and j **communicate**, and we write $i \leftrightarrow j$.
  - Two states belong to the same (communicating) **class** if they communicate with each other.
- Properties
  - If $i \to j$ and $j \to k$, then $i \to k$.

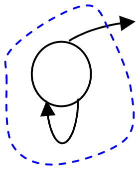
    Proof. If $i \to j$ then there is a nonzero-probability path from i to j. Similarly, $j \to k$ implies that there is a nonzero-probability path from j to k. The combined paths form a nonzero-probability path from i to k.
  - $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$
  - $\{i \leftrightarrow j$ and $j \leftrightarrow k\}$ implies $\{i \leftrightarrow k\}$
  - Two different classes of states must be disjoint.

    Proof. Having a state in common would imply that the states from both classes communicate with each other.
- The **communicating class of state *i*** is $C(i) = \{j : i \leftrightarrow j\}$
  - If $C(i) = \varnothing$, i is a **non-return state**

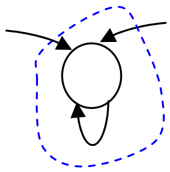
    (once out to another state, never return here.)
    - Ex.



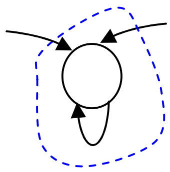
  - *Caution*: if $P_{i,i} > 0$, then state i is <u>not</u> a non-return state since $i \in C(i)$
- A non-empty class C of states is a **communicating class** if $C = C(i)$ for some i.
  - If C_1 and C_2 are communicating classes,
    Then either $C_1 = C_2$ or $C_1 \cap C_2 = \varnothing$.
  - All states in a communicating class have the same period, also called the **class's period**.
  - Ex. 1-state, transient communicating class:



  - Ex. 1-state, absorbing communicating class:



- The states of a Markov chain consist of one or more disjoint communication classes.
- The state space of a MC can be **decompose**d into a union of the form $C_1 \cup C_2 \cup C_3 \cup \ldots$
  Where the C_i are disjoint and each is either a communicating class or contains exactly one non-return state.
- MC chain is **irreducible** or **indecomposable** if
  - ≡ all pairs of states communicates
  - ≡ entire state space consists of one communicating class
  - Ex. reducible (decomposable) if has 1-state, absorbing communicating class



- Either of the following is a sufficient condition of an irreducible MC to be aperiodic
  - $P_{ii} > 0$ for some i (self-loop)
  - $P^n > 0$ for some n (common path length for any state pair)
- Define an indicator function: $I_i(X) = \begin{cases} 1, & X = i \\ 0, & \text{otherwise} \end{cases}$
- Def: Suppose we start a Markov chain in a recurrent state i at time $n = 0$.

- Let $T_i(1), T_i(1)+T_i(2), \ldots$ be the times when the process returns to state *i*, where $T_i(k)$ is the time that elapses between the (k-1)$^{\text{th}}$ and k^{th} returns.
- π_i = **the long-term proportion of time spent in state *i***.

- $\left(T_i(k)\right)_{k=1}^{\infty}$ form an iid sequence since each return time is independent of previous return times. Let $T_i(k) \sim T_i$ and the **mean recurrence time** $ET_i = E\left[T_i(k)\right]$.

- **Def**: State *i* is said to be
  - **recurrent** if suppose we start a Markov chain in state *i*
    - ≡ (1) $\Pr\left[\text{ever returning to state } i\right] = 1$. ≡ $\Pr\left[T_i < \infty\right] = 1$.
    - ≡ (2) the state reoccurs an infinite number of times. $\left(E\left[\sum_{n=1}^{\infty} I_i(X_n) \middle| X_0 = i\right] = \infty\right)$.
    - ≡ (3) $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$
  - **transient** if suppose we start a Markov chain in state *i*
    - ≡ (1) $\Pr\left[\text{ever returning to state } i\right] < 1$.

      There exists some probability of going out of state *i* and never come back.

      ≡ $\Pr\left[T_i < \infty\right] < 1$ ≡ $\Pr\left[T_i = \infty\right] > 0$.
    - ≡ (2) the state does not reoccur after some finite number of returns. $\left(E\left[\sum_{n=1}^{\infty} I_i(X_n) \middle| X_0 = i\right] < \infty\right)$.
    - ≡ (3) $\sum_{n=1}^{\infty} p_{ii}(n) < \infty$.

Proof. "(1) ≡ (2)" Each reoccurrence of the state can be viewed as Bernoulli. Let γ = $\Pr\left[\text{ever returning to state } i\right] < 1$. The probability of having at least *n* reoccurrences is γ^n. When $\gamma < 1$, $\lim_{n\to\infty} \gamma^n = 0$. Therefore, a transient state reoccurs only a finite number of times. When $\gamma = 1$, $\lim_{n\to\infty} \gamma^n = 1$.

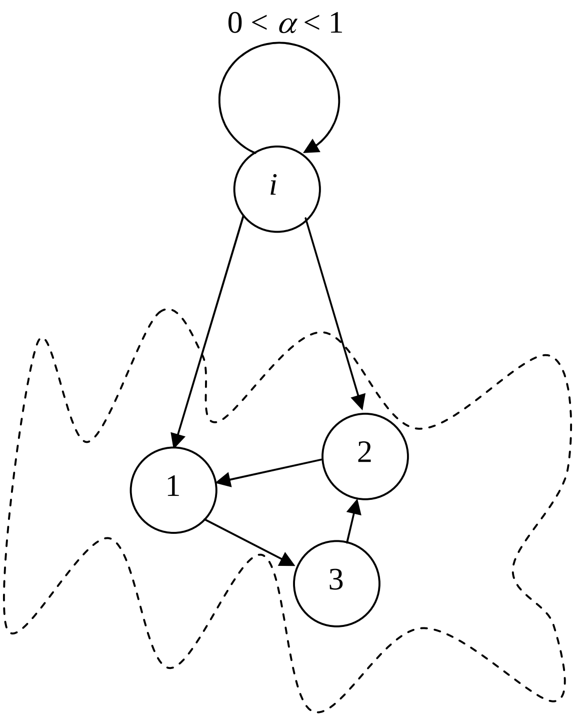
"(2) ≡ (3)" We'll show that

Expected number of returns to state *i* = $E\left[\sum_{n=1}^{\infty} I_i(X_n) \middle| X_0 = i\right] = \sum_{n=1}^{\infty} p_{ii}(n)$.

To see this, note that $E\left[I_i(X_n) \middle| X_0 = i\right] = \Pr\left[X_n = i \middle| X_0 = i\right] = p_{ii}(n)$. Hence

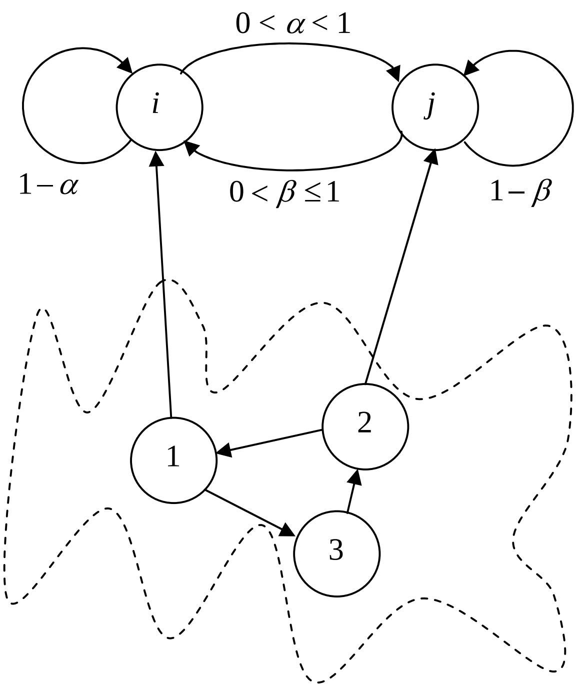
$$E\left[\sum_{n=1}^{\infty} I_i(X_n) \middle| X_0 = i\right] = \sum_{n=1}^{\infty} E\left[I_i(X_n) \middle| X_0 = i\right] = \sum_{n=1}^{\infty} p_{ii}(n).$$

- Ex.



State i is transient.

$$\sum_{n=1}^{\infty} p_{ii}(n) = \sum_{n=1}^{\infty} \alpha^n = \frac{\alpha}{1-\alpha} < \infty .$$



State i is recurrent.

$$\sum_{n=1}^{\infty} p_{ii}(n) = \sum_{n=1}^{\infty} \frac{1}{\alpha+\beta}\left(\beta + \alpha(1-\alpha-\beta)^n\right)$$
$$\geq \sum_{n=1}^{\infty} \frac{\beta}{\alpha+\beta} = \infty$$

- Recurrence and transience is a class property.
  If state i is recurrent and $i \leftrightarrow j$, the state j is also recurrent.

Proof. If state i is recurrent, then all states in its class will be visited eventually as the process returns to i over and over again.

From $i \leftrightarrow j$, we know that $\exists m,n \;\; p_{ji}(m), p_{ij}(n) > 0$. Note that $p_{jj}(m+n+k) \geq p_{ji}(m)\, p_{ii}(k)\, p_{ij}(n)$ because the right side of the inequality is just one of the possible paths to get from j back to j. Hence, we have

$$\sum_{k=1}^{\infty} p_{jj}(k) \geq \sum_{k=1}^{\infty} p_{jj}(m+n+k) \geq \sum_{k=1}^{\infty} p_{ji}(m)\, p_{ii}(k)\, p_{ij}(n)$$
$$= p_{ji}(m)\, p_{ij}(n) \left(\sum_{k=1}^{\infty} p_{ii}(k) \right) = \infty$$

Since all states in a class communicate, if one of the states is recurrent, all of them are recurrent.

Now, suppose one of the states in a class is transient. This class cannot have any recurrent state; otherwise, the rest of the states have to be recurrent, including the one assumed to be transient. Therefore, transience is also a class property.

- The states of a finite-state, irreducible Markov chain are all recurrent.

Proof. Irreducible Markov chain consists of a single communication class. Therefore, either all its states are transient or all its states are recurrent. Because the numbers in the cahin is finite, it is impossible for all of the states to be transient.

Priodicity

- **Def**:
  - A state S_j of a Markov chain is **periodic** if $\exists \ell \in \mathbb{N}/\{1\}$, such that $\forall k \in \mathbb{N}$ $p_{jj}(k)=0$ whenever k is not an integer multiple of ℓ.
    (Note, however, that we are not asserting that $p_{jj}(k) \neq 0$ when k is an integer multiple of ℓ.)
  - The **period** of a periodic state S_j is the smallest integer $\ell > 1$ such that $p_{jj}(k)=0$ whenever k is not an integer multiple of ℓ.
  - A state S_j of a Markov chain is an **aperiodic** state if it is not periodic. A state S_j is not periodic if $\forall \ell \in \mathbb{N}/\{1\}$, $\exists k \in \mathbb{N}$ such that k is not an integer multiple of ℓ and $p_{jj}(k) \neq 0$.
- Properties
  - A periodic state can have one and only one period.
  - If any single state in a communicating class is periodic and has period d, then all states in that class are periodic and have period d.
  - If one state in a communicating class is aperiodic, then all the states in that class must be aperiodic.

Proof. If a state S_i in the class is periodic of period d, then all the states in the class would have to be periodic with period d. Since S_j belongs to the class, S_j would have to be periodic. Since S_j is known to be aperiodic, we have a contradiction caused by assuming that some state in the class is periodic.

- Test for aperiodicity

  (a) $p_{jj}(1) \neq 0 \Rightarrow$ state S_j is aperiodic.

  (b) If $\exists k_1, k_2 \in \mathbb{N}$ such that $p_{jj}(k_1), p_{jj}(k_2) > 0$ and such that k_1 and k_2 have no common divisors other than 1, the state S_j is aperiodic.

Proof. (a) If $p_{jj}(1) > 0$, then $\forall \ell \in \mathbb{N}/\{1\}$ $\exists k \in \mathbb{N}$, namely $k = 1$, which is not an integer multiple of ℓ and for which $p_{jj}(1) \neq 0$.

Proof. (b) $\forall \ell \in \mathbb{N}/\{1\}$, one or both of k_1 and k_2 is not an integer multiple of ℓ. Otherwise, $\ell \neq 1$ is a common divisor of k_1 and k_2 which contradicts the assumption that that k_1 and k_2 have no common divisors other than 1.

- Def:
  - A communicating class is periodic of period d if any one state in the class is periodic and has period d.

- If any state of an irreducible Markov chain is periodic with period *d*, then we say that the Markov chain itself is periodic with period *d*.
- An irreducible Markov chain that has one periodic state is called an aperiodic, irreducible Markov chain.

  All the states of an aperiodic irreducible Markov chain are aperiodic states.

- If a Markov chain is irreducible, and if one state is periodic with period *d*, then all states of that irreducible Markov chain is periodic with period *d*.
- Periodic irreducible Markov chains do not achieve probabilistic equilibrium

Limiting Probabilities

- If all the states in a Markov chain are transient, then all the state probabilities approach zero as $n \to \infty$.
- If a Markov chain has some transient classes and some recurrent classes, then eventually the process enters and remains thereafter in one of the recurrent classes.

  Therefore, we can concentrate on individual recurrent classes whenstudying the limiting probabilities of a chain.

- MC chain possesses a **limiting distribution**, $\underline{\pi}$, if

  for all (independent of) initial conditions $\underline{p}(0)$, $\lim_{n\to\infty} \underline{p}(n) = \lim_{n\to\infty} \underline{p}(0) P^n = \underline{\pi}$

  - forget its initial conditions in the sense that $\underline{p}(n) \to \underline{\pi}$ as $n \to \infty$ regardless of $\underline{p}(0)$

- $\underline{\pi}$ is a limiting distribution if and only if $\lim_{n\to\infty} \left(P^n\right)_{ij} = \pi_j$ for all states *i* and *j*

- If $\underline{p} = \underline{p}P$, we say $\underline{p}$ is an **equilibrium distribution**.
  - Every finite-state MC has at least one equilibrium distribution.
  - If a limiting distribution exists, it is the unique equilibrium distribution
  - Every finite state irreducible homogeneous MC has a unique equilibrium distribution.
  - MC that has 2 or more absorbing state has infinite # of equilibrium distribution. (including ones that are distributed among the absorbing states in any way.)

Limiting Probabilities of irreducible Markov chain

- The proportion of time spent in state *i* after *k* returns to *i* is $\dfrac{k}{T_i(1) + T_i(2) + \cdots + T_i(k)}$.

  Since the state is recurrent, the process returns to state *i* an infinite number of times. By the law of large number, with probability one, $\lim_{k\to\infty} \dfrac{T_i(1) + T_i(2) + \cdots + T_i(k)}{k} = ET_i$.

  So, the long-term proportion of time spent in state *i* approaches $\boxed{\dfrac{1}{ET_i} := \pi_i}$.

- Def:

- State i is **positive recurrent** if $ET_i < \infty$.

  $\Rightarrow \pi_i > 0$.
- State i is **null recurrent** if $ET_i = \infty$.

  $\Rightarrow \pi_i = 0$.
- Ergodic states is a positive recurrent, aperiodic state.
- An **ergodic Markov** chain is an irreducible, aperiodic, positive recurrent Markov chain.

- Positive and null recurrence are class properties.
- $\left\{\pi_i = \frac{1}{ET_i} 0\right\}$ satisfies the equations that define the stationary state pmf: $\forall j \ \pi_j = \sum_i \pi_i P_{ij}$, and $1 = \sum_i \pi_i$.

  Proof. Since π_i is the proportion of time spent in state i, then $\pi_i P_{ij}$ is the proportion of time in which state j follows i. If we sum over all i, we then obtain the long-term proportion of time in state j, π_j .
- For Markov chains that exhibit stationary behavior, the n-step transition matrix approaches a fixed matrix of equal rows as $n \to \infty$. $\left(\forall i \ \forall j \ p_{ij}(n) \to \pi_j\right)$. The rows of this limiting matrix consisted of a pmf that satisfies $\forall j \ \pi_j = \sum_i \pi_i P_{ij}$, and $1 = \sum_i \pi_i$.

- For an irreducible, aperiodic, and positive recurrent Markov chain, $\forall j \ \lim_{n\to\infty} p_{ij}(n) = \pi_j$ where π_j is the unique nonnegative solution of a pmf that satisfies 1) $\forall j \ \pi_j = \sum_i \pi_i P_{ij}$, and 2) $1 = \sum_i \pi_i$.

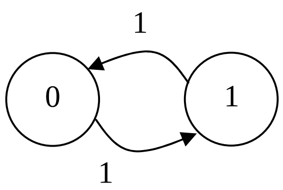
  Hence, for irreducible, aperiodic, and positive recurrent Markov chains, the state probabilities approach steady state values that are independent of the initial condition. These steady state probabilities correspond to the long-term proportion of time spent in the given state.
- For an irreducible, periodic, and positive recurrent Markov chain with period d, $\forall j \ \lim_{n\to\infty} p_{jj}(nd) = d\pi_j$, where π_j is the unique nonnegative solution of a pmf that satisfies 1) $\forall j \ \pi_j = \sum_i \pi_i P_{ij}$, and 2) $1 = \sum_i \pi_i$.

  As before, π_j represents the proportion of time spent in state j. However, the fact that state j is constrained to occur at multiples of d steps implies that the probability of occurrence of the state j is d times greater at the allowable times and zero elsewhere.

- **Markov Theorem**:
  Every finite-state, irreducible, aperiodic MC possesses a limiting distribution.

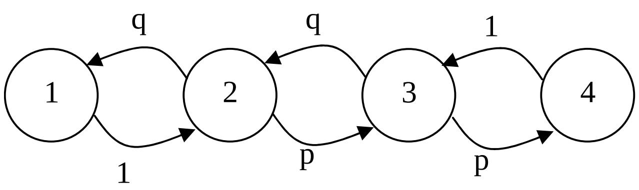
  - If MC is periodic, it doesn't have a limiting distribution.

    Ex. Figure below has a unique equilibrium distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$. But no limiting distribution.



Equilibrium probability

- $\underline{p}$ = the steady-state or equilibrium probability vector
  - $\underline{p} = \underline{p}\, P$ (use "in" arrow)
  - $\underline{p} = (p_1, p_2, \ldots, p_n)$, $\sum_{i=1}^{n} p_i = 1$

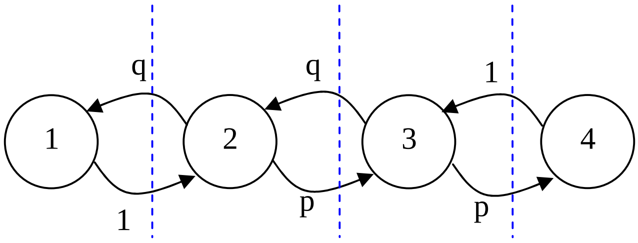


- Out of state: For any state, sum of probabilities on all arrows <u>out</u> of it equals 1. (Not so for arrows <u>in</u>.)
  - $p+q = 1$
- In the long run,
  - p_i = fraction of state transition in the long run that enter state i
  - Consider the arrow <u>**into**</u> each state: $p_j = \sum_i P_{i,j}\, p_i$
    - into state:

      $p_1 = qp_2$
      $p_2 = 1p_1 + qp_3$
      $p_3 = pp_2 + 1p_4$
      $p_4 = pp_3$

      $$\left((p_1, p_2, p_3, p_4) = (p_1, p_2, p_3, p_4)\begin{pmatrix} 0 & 1 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 1 & 0 \end{pmatrix} \right)$$

    - homogeneous.
    - To get solution, add $\sum_{i=1}^{n} p_i = 1$
- Ex. Solving technique for small system



- Use the blue boundary:

$$qp_2 = 1p_1$$
$$qp_3 = pp_2$$
$$1p_4 = pp_3$$

- Set $p_1 = 1$, then $p_2 = \frac{1}{q} p_1 = \frac{1}{q}$. $p_3 = \frac{p}{q} p_2 = \frac{p}{q^2}$. $p_4 = pp_3 = \frac{p^2}{q^2}$.
- $1 + \frac{1}{q} + \frac{p}{q^2} + \frac{p^2}{q^2} = \frac{q^2 + q + p + p^2}{q^2}$. Scale all p_i by one over this.
- Thus

$$p_1 = \frac{q^2}{q^2 + q + p + p^2}, \; p_2 = \frac{q}{q^2 + q + p + p^2}$$
$$p_3 = \frac{p}{q^2 + q + p + p^2}, \; p_4 = \frac{p^2}{q^2 + q + p + p^2}$$

Probability of absorption

- Let *a* be the state(s) for which we seek the probability that absorption eventually will take place there

$$f_i \triangleq \Pr\left[\text{Absoption will (eventually) occur in state } a \middle| \text{Current state is i}\right]$$

  Note that

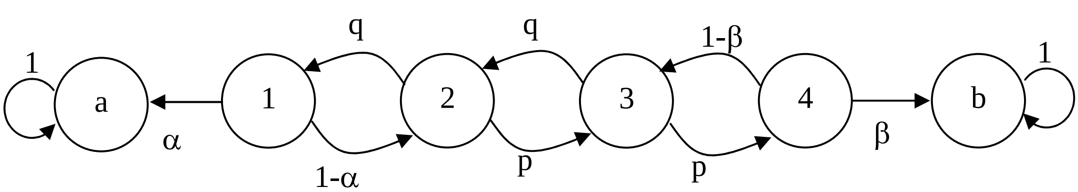
  $f_a = 1$

  $f_i = 0$ if *i* is any absorbing state other than a.

- $\underline{f} = \underline{f} P^T$
- Equation for f_i involves probabilities on arrows going **out** of state *i*.

$$f_i = \sum_j f_j P_{i,j}$$

- Ex



$f_a = 1$
$f_b = 0$
$f_1 = \alpha f_a + (1-\alpha) f_2$
$f_2 = qf_1 + pf_3$
$f_3 = qf_2 + pf_4$
$f_4 = (1-\beta) f_3 + \beta f_b$

- $\sum_i f_i$ do <u>not</u> sum to 1.
- Pr[eventually absorption occurs at a] = $\sum_i p_i f_i$, where p_i's are calculated without the absorbing state(s).

Expected time to absorption

- $e_i \triangleq E\left(\begin{matrix}\text{Remaining \#transition} \\ \text{until absorption}\end{matrix}\middle|\begin{matrix}\text{Just entered} \\ \text{state i}\end{matrix}\right)$
- e_i = 0 if i is an absorbing state.
- equations: use the **out** arrow
  $e_i = \sum_j P_{i\,to\,j}\left(1+e_j\right) = \sum_j P_{i\,to\,j} + \sum_j P_{i\,to\,j} e_j = 1 + \sum_j P_{i\,to\,j} e_j$
- $\underline{e} = (\underline{1}+\underline{e})P^T$
- note: put 0 (not 1) in P for $P_{i,i}$ where i is an absorbing state. (This means row of 0's for each absorbing state.)
- $\underline{e} = \underline{1}_{1\times n} P^T \left(I - P^T\right)^{-1}$

Time-Reversed Markov Chains

- Let X_n be a stationary ergodic Markov chain (an irreducible, aperiodic, stationary Markov chain) with one-step transition probability matrix $P = \left[p_{ij}\right]$ and stationary state pmf $\left\{\pi_j\right\}$.

  The time-reversed process is also a Markov chain with one-step transition probabilities $\hat{p}_{ij} = \frac{\pi_j}{\pi_i} p_{ji}$

  Proof. $\Pr\left[X_{n-1} = x_{n-1} \middle| X_n^{n+k} = x_n^{n+k}\right] = \frac{\Pr\left[X_{n-1}^{n+k} = x_{n-1}^{n+k}\right]}{\Pr\left[X_n^{n+k} = x_n^{n+k}\right]} = \frac{\pi_{x_{n-1}} p_{x_{n-1},x_n} p_{x_n,x_{n+1}} \cdots p_{x_{n+k-1},x_{n+k}}}{\pi_{x_n} p_{x_n,x_{n+1}} \cdots p_{x_{n+k-1},x_{n+k}}}$

  $$= \frac{\pi_{x_{n-1}} p_{x_{n-1},x_n}}{\pi_{x_n}}$$

  Note also that $\Pr\left[X_{n-1} = x_{n-1} \middle| X_n = x_n\right] = \frac{\pi_{x_{n-1}} p_{x_{n-1},x_n}}{\pi_{x_n}}$. Hence,

  $$\Pr\left[X_{n-1} = x_{n-1} \middle| X_n^{n+k} = x_n^{n+k}\right] = \Pr\left[X_{n-1} = x_{n-1} \middle| X_n = x_n\right].$$

- The forward and reverse process must have the same stationary pmf

  Proof. Since X_n is irreducible and aperiodic, its stationary state probabilities $\{\pi_j\}$ represent the proportion of time that the state is in state *j*. This proportion of time does not depend on whether one goes forward or backward in time, so $\{\pi_j\}$ must also be the stationary pmf for the reverse process.

- Another method for finding the stationary pmf of a discrete-time Markov chain:
  Given $\{p_{ij}\}$, if we can guess a set of transition probabilities $\{q_{ij}\}$ for the reverse process and a pmf $\{\pi_j\}$ so that $\forall i,j \;\; \pi_i q_{ij} = \pi_j p_{ji}$, then the $\{\pi_j\}$ is the stationary pmf for the Markov chain and the $\{q_{ij}\}$ are the transition probabilities for the reverse process.

  Proof. $\forall i,j \;\; \pi_i q_{ij} = \pi_j p_{ji} \Rightarrow \forall i \;\; \sum_j \pi_j p_{ji} = \pi_i \sum_j q_{ij} = \pi_i$. Hence, $\{\pi_j\}$ is the stationary pmf. Because $\{\pi_j\}$ is the stationary pmf, $q_{ij} = \dfrac{\pi_j p_{ji}}{\pi_i}$ is the transition probability for the reverse process.

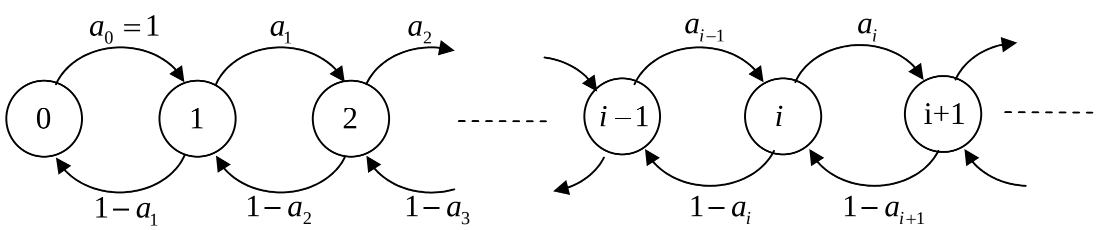
Time-Reversible Markov Chains

- A stationary ergodic Markov chain is said to be **reversible** if
  - ≡ the one-step transition probability matrix of the forward and reverse processes are the same, that is, if $\forall i,j \;\; \hat{p}_{ij} = p_{ij}$.
  - ≡ $\boxed{\forall i,j \;\; \pi_i p_{ij} = \pi_j p_{ji}}$

    (the proportion of transitions from *i* to *j* is equal to the proportion of transitions from *j* to *i*)

  Proof. Use $\hat{p}_{ij} = \dfrac{\pi_j}{\pi_i} p_{ji}$.

Discrete-time birth-and-death process



- Discrete-time birth-and-death processes are reversible

  Proof. For any sample path, the number of transitions from *i* to *i*+1 can differ by at most 1 from the number of transitions from *i*+1 to *i* since the only way to return to *i* is through *i*+1. Thus, the long term proportion of transitions from *i* to *i*+1 is equal to that from *i*+1 to *i*. $\left(\pi_i p_{i,i+1} = \pi_{i+1} p_{i+1,i}\right)$ Since these are the only possible transitions, it follows that birth-and-death processes are reversible because it satisfies $\forall i,j \;\; \pi_i p_{ij} = \pi_j p_{ji}$.

- From $\pi_i p_{i,i+1} = \pi_{i+1} p_{i+1,i}$, we have $\pi_{i+1} = \pi_i \dfrac{p_{i,i+1}}{p_{i+1,i}} = \pi_i \dfrac{a_i}{1-a_{i+1}}$. Hence

$$\pi_j = \frac{a_{j-1}\cdots a_0}{\left(1-a_j\right)\cdots\left(1-a_1\right)}\pi_0 = R_j \pi_0 \text{ where } R_j = \frac{a_{j-1}\cdots a_0}{\left(1-a_j\right)\cdots\left(1-a_1\right)}.$$

If $\sum_{j=0}^{\infty} R_j$ converges, $\pi_0 = \dfrac{1}{\sum_{j=0}^{\infty} R_j}$.