TS: MARKOV CHAINS

Consider a sequence of RVs $X_1, X_2, X_3, \ldots$ When these RVs are considered independent, you can't directly use them to model (i) Evolution of the size of a population
(ii) Stock price
(iii) DNA sequence / Sequence of symbols from an alphabet $\mathcal{A}$.

↳ Indeed, you may want to attribute a manuscript to some author, by (naively) assuming that consecutive letters are independent. Back to the frequency interpretation of probability, you can check the independence assumption by taking a long enough empirical text $\tau = \{a_1, \ldots, a_n\}$, $a_i \in \mathcal{A}$, and put

$$n_{\ell_i} = \#\{j \leq n \mid a_j = \ell_i\}$$
$$n_{\ell_1 \ell_2} = \#\{j \leq n-1 \mid a_j a_{j+1} = \ell_1 \ell_2\}$$

= frequency of some letters $\ell_i \in \mathcal{A}$, $i = 1, 2$.

If τ = random text, we expect that

$$\frac{n_{\ell_i}}{n} \simeq \mathbb{P}(\text{letter} = \ell_i) \quad \& \quad \frac{n_{\ell_1 \ell_2}}{n} \simeq \mathbb{P}(\text{two conc. letters} = \ell_1 \ell_2)$$

If consecutive letters were independent, then we would have that

$$\frac{n_{\ell_1 \ell_2}}{n} \simeq \frac{n_{\ell_1}}{n} \times \frac{n_{\ell_2}}{n}.$$

↳ But this almost never holds. For example, taking $\mathcal{A} = \{a, b, c, \ldots, y, z\}$ and $\ell_1 = \ell_2 = a$, then for a plain English text, we would get $\mathbb{P}(\text{letter} = a) \simeq 0.08$, while the pair 'aa' is extremely rare ($= 0$ unless you have in your hands a book about mammals in southern Africa: aardvarks & aardwolfs)

→ In many situations, the use of independent sequences is not justified.
When we have a sequence of RVs (defined on a common probability space), we call it a <u>STOCHASTIC PROCESS</u> (SP) or a <u>RANDOM PROCESS</u>.

→ The set S where the X_n's take values is called the <u>STATE SPACE</u> of the process. Points $i \in S$ are called the <u>STATES</u> of the process.
In this Chapter, we focus on discrete time processes with finite $S = \{s_1, \ldots, s_m\} = \{1, \ldots, m\}$ & infinite S.

Def: An SP $\{X_n, n \geq 0\}$ is called a <u>MARKOV CHAIN</u> (MC) if, given the present state of the process, its future does not depend on its past:

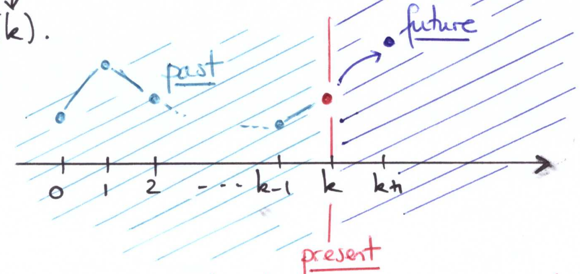
$$\mathbb{P}(\underbrace{X_{k+1} = \ell}_{\text{future}} \mid \underbrace{X_0 = i_0, \ldots, X_{k-1} = i_{k-1}}_{\text{past}}, \underbrace{X_k = j}_{\text{present}})$$
$$= \mathbb{P}(\underbrace{X_{k+1} = \ell}_{\text{future}} \mid \underbrace{X_k = j}_{\text{present}}); \quad i_m \in S, \ m = 0, \ldots, k-1. \quad j, \ell \in S$$

Denote this by $p_{j\ell}$ (k).
time index

If they do not depend on k, the MC is said to be <u>HOMOGENEOUS</u>:

$$p_{j\ell}(k) = p_{j\ell}.$$

called <u>TRANSITION PROBABILITIES</u>.



knowledge of the present state is enough (⚠ The EXACT PRESENT) to predict the future.

Examples. (i) iid RVs :

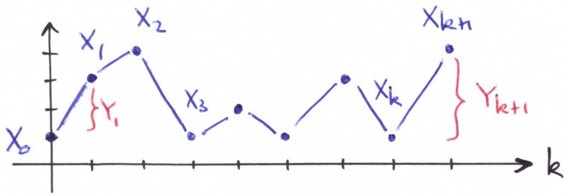
$$\mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_k, \dots, X_0 = i_0)$$
$$= \mathbb{P}(X_{k+1} = i_{k+1})$$
$$= \mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_k)$$

$\Rightarrow$ A sequence of iid RVs is a Markov Chain.

(ii) Random walks : • $X_{k+1} = X_k + Y_{k+1}$, $k \geq 0$

• X_0 = some initial value

• Y_k = iid RVs (integer valued).



Nb= State Space S is infinite.

$$\mathbb{P}(X_{k+1} = i_{k+1} \mid \underline{X_k = i_k}, \dots, X_0 = i_0)$$
$$\| \qquad\qquad \updownarrow$$
$$X_k + Y_{k+1} = i_k + Y_{k+1}$$

$$= \mathbb{P}(Y_{k+1} = i_{k+1} - i_k \mid X_k = i_k, \cancel{\dots, X_0 = i_0})$$

remove, from the independence of the Y's.

$$= \mathbb{P}(Y_{k+1} = i_{k+1} - i_k)$$ Using a similar reasoning

$$= \mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_k).$$

$\Rightarrow$ Random Walks are indeed Markov Chains.

Any MC $\{X_k\}$ is $\equiv X_{k+1} = f(X_k, Y_{k+1})$ for some function f and iid RVs Y_k.

Homogeneous MC.

The transition probabilities are usually put together into a matrix :

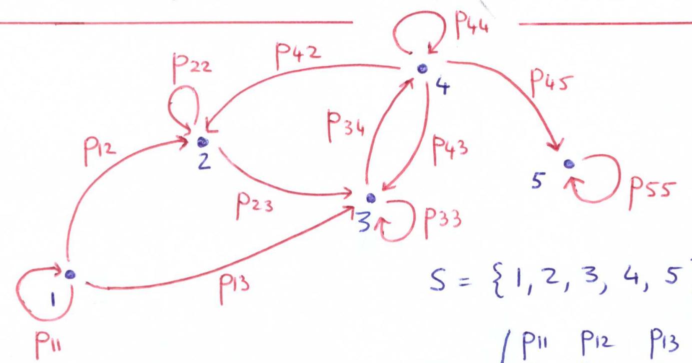
$$P = (p_{j\ell}) = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ p_{21} & \cdots & p_{2m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}$$

called TRANSITION MATRIX.

$(m \times m)$ matrix when the state space S is finite.

When S is (countably) infinite, think of P as a "square infinite" matrix.

Can be represented using a transition diagram :



$S = \{1, 2, 3, 4, 5\}$

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & 0 & 0 \\ 0 & p_{22} & p_{23} & 0 & 0 \\ 0 & 0 & p_{33} & p_{34} & 0 \\ 0 & 0 & p_{43} & p_{44} & p_{45} \\ 0 & 0 & 0 & 0 & p_{55} \end{pmatrix}$$

Remark: • If you leave state 1, you never return to it.

• If you get to state 5, you never leave it.

$\Rightarrow$ States 1 and 5 are of a very different nature. More on this soon.

Elements of a transition matrix are such that 5

(i) $p_{j\ell} \geq 0 \quad \forall j, \ell \in S$

(ii) $\sum_{\ell \in S} p_{j\ell} = 1 \quad , \forall j \in S,$ row sums to one.

A matrix satisfying (i) and (ii) is called a STOCHASTIC MATRIX. Note that for any Stochastic Matrix, there always exist a MC whose transition matrix coincides with the given stochastic matrix.

Examples. (i) iid RVs.

Let $\{X_k\}$ be a sequence of i.i.d. RVs s.t.

$\mathbb{P}(X_k = j) = \pi_j \quad , \quad j = 1, \dots, m.$

Then

$$P = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \pi_1 & \pi_2 & \cdots & \pi_m \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{pmatrix}.$$

(ii) Random Walk.

Assume that the Y_k introduced on page 3 satisfy:

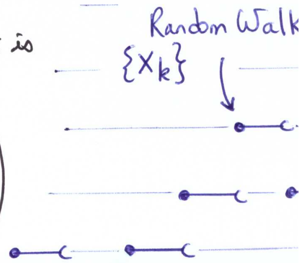
$$Y_k = \begin{cases} +1 & \text{w.p } p \\ -1 & \text{w.p } 1-p =: q \end{cases}.$$

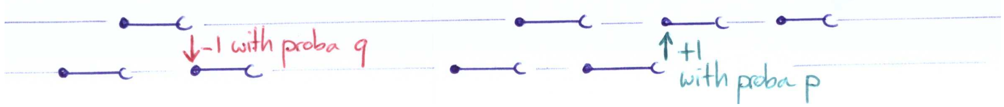
The transition probabilities are:

$$p_{j\ell} = \begin{cases} p & \text{if } \ell = j+1 \\ 1-p & \text{if } \ell = j-1 \\ 0 & \text{otherwise} \end{cases},$$

and the (doubly-infinite) transition matrix is

$$P = \begin{pmatrix} & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & 0 & p & 0 & 0 & 0 & \cdots \\ \cdots & q & 0 & p & 0 & 0 & \cdots \\ \cdots & 0 & q & 0 & p & 0 & \cdots \\ \cdots & 0 & 0 & q & 0 & p & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$





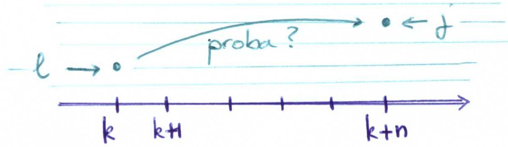
Summary: Ingredients needed to characterise the dynamics of a homogeneous MC are 6

(i) Initial distribution $p = \{p_j\}$; $p_j = \mathbb{P}(X_0 = j)$, $j \in S$

(ii) Transition matrix $\mathbb{P} = (p_{j\ell})$.

↳ What can we say about n-step transition probabilities?

$\mathbb{P}(X_{k+n} = j \mid X_k = \ell)$



$$p^{(n)}_{\ell j} := \mathbb{P}(X_{k+n} = j \mid X_k = \ell)$$ homogeneous MC.

$$= \mathbb{P}(X_n = j \mid X_0 = \ell)$$

$$= \sum_{i \in S} \mathbb{P}(X_n = j, X_m = i \mid X_0 = \ell) \qquad 1 \leq m \leq n-1$$

$$= \sum_{i \in S} \mathbb{P}(X_n = j \mid X_m = i, \cancel{X_0 = \ell}) \, \mathbb{P}(X_m = i \mid X_0 = \ell)$$

(Markov property)

$$= \sum_{i \in S} \mathbb{P}(X_n = j \mid X_m = i) \, \mathbb{P}(X_m = i \mid X_0 = \ell)$$

We arrive at the so-called CHAPMAN-KOLMOGOROV equation.

$$\boxed{\begin{aligned} p^{(n)}_{\ell j} &= \sum_{i \in S} p^{(m)}_{\ell i} \, p^{(n-m)}_{ij} \\ P^{(n)} &= P^{(m)} P^{(n-m)}, \quad \text{where } P^{(n)} := \left(p^{(n)}_{\ell j}\right) \end{aligned}}$$

Consequence is that $P^{(n)} = P^{(n-1)} P^{(1)} = P^{(n-1)} P = P^{(n-2)} P^2 = \dots = P^n$; ie the n-step transition probability is simply the n-th power of the transition matrix: $P^{(n)} = P^n$.

↳ What about the distribution of X_n ?

$$\mathbb{P}(X_n = j) = \sum_{i \in S} \mathbb{P}(X_n = j \mid X_0 = i)\, \mathbb{P}(X_0 = i)$$
$$= \sum_{i \in S} p_i\, p_{ij}^{(n)}$$
$$= \Big(\underset{(1\times m)}{p}\, \underset{(m \times m)}{P^{(n)}} \Big)_j = (p P^n)_j$$

$\Rightarrow \mathbb{P}(X_n = j)$ is the j-th element of the vector $p P^n$.

↳ What about the probability of a trajectory $j_1 \cdots j_n$ given the initial state was $X_0 = j$?

$$\mathbb{P}(X_n = j_n, X_{n-1} = j_{n-1}, \ldots, X_1 = j_1 \mid X_0 = j)$$
$$= \mathbb{P}(X_n = j_n \mid X_{n-1} = j_{n-1}) \times \cdots \times \mathbb{P}(X_1 = j_1 \mid X_0 = j)$$
$$= P_{j j_1} P_{j_1 j_2} \times \cdots \times P_{j_{n-1} j_n}.$$

__Examples__. (i) __iid RVs__.

$$P^{(2)} = P^2 = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{pmatrix} \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{pmatrix} = P$$

(ii) __Random Walk__.

$$P^2 = \begin{pmatrix} & \vdots & \vdots & \vdots & \vdots & \\ \cdots & 2pq & 0 & p^2 & 0 & \cdots \\ \cdots & 0 & 2pq & 0 & p^2 & \cdots \\ \cdots & q^2 & 0 & 2pq & 0 & \cdots \\ \cdots & 0 & q^2 & 0 & 2pq & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

Indeed, jumps have distribution $Y_1 + Y_2 = \begin{cases} -2 & \text{w.p } q^2 \\ 0 & \text{w.p } 2pq \\ +2 & \text{w.p } p^2 \end{cases}$

(iii) Suppose that the transition matrix of some homogeneous MC is given by

$$\bullet\ P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}.$$

You may check that

$$\bullet\ P^2 = \begin{pmatrix} 0.34 & 0.39 & 0.27 \\ 0.22 & 0.36 & 0.42 \\ 0.17 & 0.34 & 0.49 \end{pmatrix} \begin{array}{l} \leftarrow \text{distribution of } X_2 \text{ given } X_0 = 1 \\ \leftarrow \text{———''——— } X_2 \text{ — } X_0 = 2 \\ \leftarrow \text{———''——— } X_2 \text{ — } X_0 = 3 \end{array}$$

$$\bullet\ P^4 = \begin{pmatrix} 0.2478 & 0.3648 & 0.3879 \\ 0.2254 & 0.3582 & 0.4164 \\ 0.2159 & 0.3553 & 0.4288 \end{pmatrix}$$

$$\bullet\ P^{10} = \begin{pmatrix} 0.226548 & 0.358531 & 0.414922 \\ 0.226409 & 0.358489 & 0.415103 \\ 0.226348 & 0.358470 & 0.415181 \end{pmatrix}$$

$$\bullet\ P^{20} = \begin{pmatrix} 0.226415 & 0.358491 & 0.415094 \\ 0.226415 & 0.358491 & 0.415094 \\ 0.226415 & 0.358491 & 0.415094 \end{pmatrix} \leftarrow \text{rows are identical !}$$

$\Rightarrow$ It appears that the n-step transition matrix $P^{(n)} = P^n$ converges to a matrix of the form

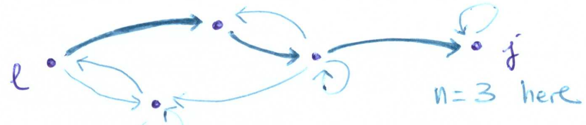
$$P^n \longrightarrow \Pi = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \end{pmatrix}, \quad \text{as } n \to \infty.$$

__Interpretation__: The i-th row of P^n gives the conditional distribution of X_n given $X_0 = i$. These distributions appear to be almost equal to each other for large n.
"The MC forgets" where it started.

↳ We want to derive conditions for the convergence $P^n \to \Pi$. To understand the long-term behaviour of a MC, one first has to take a detour and __CLASSIFY__ the states of a MC (recall example page 5).

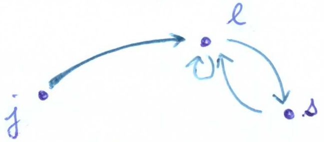
I _ CLASSIFICATION OF STATES

• A State $j \in S$ is said to be ACCESSIBLE from $\ell \in S$ (and we write $\ell \mapsto j$) if, for some $n>0$, one has $p^{(n)}_{\ell j} > 0$



[In other words, there is a path $\ell = i_0, i_1, \dots, i_n = j$ of length n such that $p_{i_0 i_1} p_{i_1 i_2} \times \cdots \times p_{i_{n-1} i_n} > 0$]

• A state $j \in S$ is called NON-ESSENTIAL if there exists a state $\ell \in S$ such that $j \mapsto \ell$, but $\ell \not\mapsto j$.
That is, for some $n>0$, $p^{(n)}_{j\ell} > 0$, but $p^{(m)}_{\ell j} = 0$ $\forall m \geq 1$.



The state j is non-essential: you can go from j to ℓ, but how would you go back to j from ℓ?

[In other words, we can reach ℓ from j, but we cannot reach j from ℓ]

→ P^n has a non-zero (j, ℓ) entry while
→ P^m has a zero (ℓ, j) entry $\forall m \geq$

Collection of all non-essential states is denoted S_{NE}.

• Otherwise, the state $j \in S$ is said to be ESSENTIAL: if $j \mapsto \ell$, then $\ell \mapsto j$. If two states ℓ and j satisfy this, we say that they COMMUNICATE. Denote by S_E the set of all essential states.
↑ write $j \leftrightarrow \ell$.

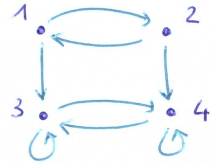
The states ℓ and s in the diagram above communicate.

Take Away Message

Essential & Non Essential states are about accessibility.

Example In the MC with

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix},$$



states 1 and 2 are non-essential, while states 3 and 4 are essential.

• For a non-essential state, the MC never visits j after some random time. In opposite, a state j is called ABSORBING if, having arrived at state j, the MC never leaves it.

• Let $j \in S_E$, and write $S(j)$ = class of all states communicating with j.

Satisfies:

• Reflexivity $j \leftrightarrow j$
↗ j is absorbing and thus $p_{jj} = 1$, $\Rightarrow j \leftrightarrow j$
↘ $\exists k$ s.t. $j \leftrightarrow k$ and $\exists n_1, n_2$ s.t. $p^{(n_1)}_{jk} > 0$, $p^{(n_2)}_{kj} > 0$.
$\Rightarrow p^{(n_1+n_2)}_{jj} \geq p^{(n_1)}_{jk} p^{(n_2)}_{kj} > 0$.

• Symmetry $j \leftrightarrow k \Leftrightarrow k \leftrightarrow j$

• Transitivity: If $j \leftrightarrow k$ and $k \leftrightarrow \ell$, then $j \leftrightarrow \ell$.

We have an EQUIVALENT CLASS

Points in S(j) are said to be equivalent.

↓

Important: the classes S(j) for different $j \in S_E$ are either identical or disjoint.

(11)

Indeed, if for $j, k \in S_E$ $S(j) \cap S(k) \neq \phi$, then $\exists i \in S_E$ such that $j \leftrightarrow i$ and $k \leftrightarrow i$.

Now, if $v \in S(j)$, i.e. $v \leftrightarrow j$ then by transitivity we get that $v \leftrightarrow i$, so that $v \in S(i) \Rightarrow S(j) \subset S(i)$.

Similarly, if $u \in S(i)$, then $u \in S(j) \Rightarrow S(i) \subset S(j)$, and we conclude that $S(j) = S(i)$.

Similarly, we can show that $S(k) = S(i) \Rightarrow S(j) = S(k)$.

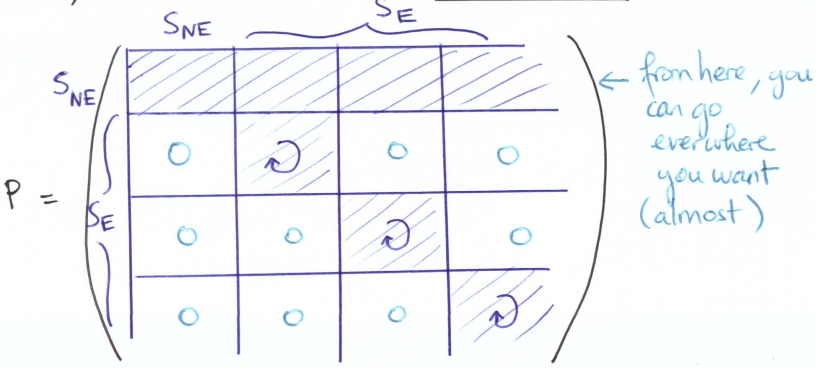
$\Rightarrow$ We can divide the state space into a number of disjoint classes.

- class S_{NE} of non essential states, and
- its complement, the class S_E of essential states, which can be further divided into closed classes $S_1, S_2, \dots$ of communicating states.

If the chain enters it, it will never leave it.

$\Rightarrow$ The transition matrix can be divided into blocks.

RK: If a MC has only one class; i.e. all states communicate with each other, the chain is said to be IRREDUCIBLE.



symbol ↻ ≡ communication within a class

(12)

Example.

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

We have $S_{NE} = \{4\}$ and $S_E = \{1, 2, 3, 5\}$, with

$S_1 = S(1) = S(2) = \{1, 2\}$ and

$S_2 = S(3) = S(5) = \{3, 5\}$.

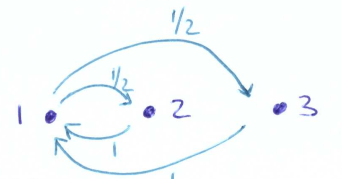
- In S_1, states 1 and 2 are visited with proba 1/2
- In S_2, states 3 and 5 are systematically visited one after another $\Rightarrow$ no randomness.

$\Rightarrow$ Classes S_1 and S_2 are of a very different nature.
We need further classification.

- A state j is said to be PERIODIC with PERIOD $d > 1$ if $p_{jj}^{(n)} = 0$ for $n \neq kd$, $k = 1, 2, \dots$, and d is the largest integer with this property.
  d = Greatest Common Divisor (GCD) of all n s.t. $p_{jj}^{(n)} > 0$.
  If the GCD = 1, the state is called APERIODIC.

Examples (i) Chain with states 1, 2, 3. All states (communicating) have period $d = 2$.

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$



Indeed, $p_{11}^{(2)} > 0$, $p_{22}^{(2)} > 0$, $p_{33}^{(2)} > 0$

$d = 2$ is the GCD of all n s.t. $p_{jj}^{(n)} > 0$

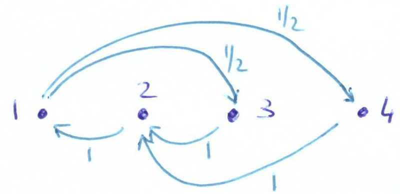
Also, $p_{11}^{(4)} > 0$, $p_{22}^{(4)} > 0$, $p_{33}^{(4)} > 0$

[In fact $p_{11}^{(2)} = 1$, $p_{22}^{(2)} = p_{33}^{(2)} = 1/2$

And $p_{11}^{(3)} = p_{22}^{(3)} = p_{33}^{(3)} = 0$

(ii) Here, all (communicating) states have $d=3$:

$$P = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$



Rk: • It is a general fact: all communicating states must have the same period (→ cf. Solidarity theorem p. 18)

• If you slightly change P and put $p_{21} = 1-\varepsilon$, $p_{22} = \varepsilon$, for $\varepsilon > 0$, then the MC becomes aperiodic.

⇒ In the presence of a "loop" (ie. a strictly positive term on the diagonal), the associated state is aperiodic, always. All states communicating with it are also aperiodic.

• We now turn our attention to MC with a countably infinite state space: more care is needed. Indeed:

Example. Random walk introduced page 3, with transition matrix given on page 5.

➘ Let $j<k$. Then $p_{jk}^{(k-j)} = p^{k-j} > 0$ ← going up $(k-j)$ times

& $p_{kj}^{(k-j)} = q^{k-j} > 0$ ← going down $(k-j)$ times, so that $j \leftrightarrow k$

⇒ All states are essential and the chain is irreducible.

➘ Walk can return to the starting point in an even # of steps only: $p_{jj}^{(2n+1)} = 0 \quad \forall j \in S,\ n \geq 0$

⇒ The MC is periodic with $d=2$.

➘ Let $p>q$. Then $\dfrac{X_n}{n} = \dfrac{\sum Y_i}{n} \xrightarrow[\text{SLLN}]{a.s} EY_1 = p-q > 0$

↑ assuming $X_0 \equiv 0$.

⇒ $X_n \approx n(p-q)$ and the walk escapes to $+\infty$!

⇒ For any $j \in S$, the walk will visit it only finitely many times ≡ as if j was non-essential.

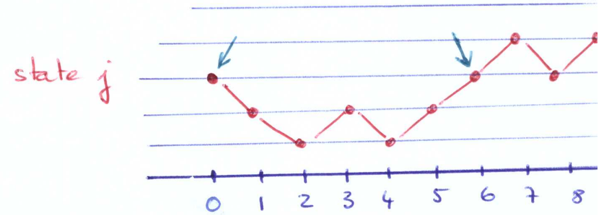
But all states are essential & the chain is irreducible

We need further classification for MCs with countably infinite state space.

• Let $f_j = \mathbb{P}(X_n = j \text{ for some } n > 0 \mid X_0 = j)$

= probability of ever returning to the initial state, given it was j.

A state j is said to be RECURRENT if $f_j = 1$, and TRANSIENT if $f_j < 1$.



For a recurrent state j, you will return to j eventually.

For a transient state, there is a (stly) positive probability that upon leaving j, you never return to it.

We have a simple criterion for recurrence

↓

Theorem 1. Put $q_j = \sum_{n=1}^{\infty} p_{jj}^{(n)} \leq \infty$. Then

$$f_j = \frac{q_j}{1+q_j}\ ; \quad \text{with } f_j = 1 \text{ when } q_j = \infty.$$

Thus, state j is recurrent iff. $q_j = \infty$.

proof: Put

$$f_j^{(n)} = \mathbb{P}(X_n = j,\ X_{n-1} \neq j, \dots, X_1 \neq j \mid X_0 = j)$$

= proba of returning to j at time n for the first time.

We have

$$p_{jj}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = j)$$

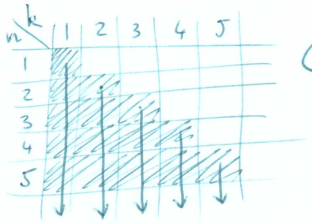
$$= \sum_{k\geq 1}^{n} \mathbb{P}(X_n = j,\ \text{first return time to } j \text{ is } k \mid X_0 = j)$$

$$= \sum_{k=1}^{n} f_j^{(k)} p_{jj}^{(n-k)}$$

$\times z^n$ and take the sum $\sum_{n\geq 1}^{\infty}$

$$\tilde{q}_j(z) := \sum_{n\geq 1}^{\infty} p_{jj}^{(n)} z^n = \sum_{n=1}^{\infty}\sum_{k=1}^{n} z^n f_j^{(k)} p_{jj}^{(n-k)}$$

$$= \sum_{n\geq 1}^{\infty}\sum_{k\geq 1}^{n} z^k f_j^{(k)} z^{n-k} p_{jj}^{(n-k)}$$



$$= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} z^k f_j^{(k)} z^{n-k} p_{jj}^{(n-k)}$$

$$= \sum_{k\geq 1}^{\infty} z^k f_j^{(k)} \sum_{n-k=0}^{\infty} z^{n-k} p_{jj}^{(n-k)}$$

$$= \tilde{f}_j(z)\left(1 + \tilde{q}_j(z)\right)$$

$$\Rightarrow \tilde{f}_j(z) = \frac{\tilde{q}_j(z)}{1+\tilde{q}_j(z)} \xrightarrow[z\uparrow 1]{} \frac{q_j}{1+q_j},$$

since for a non-neg sequence $\{a_n\}$, $\lim_{z\uparrow 1} \tilde{a}(z) := \lim_{z\uparrow 1} \sum_{n\geq 1}^{\infty} a_n z^n = \sum_{n\geq 1}^{\infty} a_n$.

Corollary 1: If a state j is recurrent, and $j \leftrightarrow k$, then k is also recurrent.

proof: By definition, $j \leftrightarrow k$ means that $\exists s,t \geq 1$ s.t $p_{jk}^{(s)} > 0$ and $p_{kj}^{(t)} > 0$; so that $\alpha = p_{jk}^{(s)} p_{kj}^{(t)} > 0$.

Also,

$$p_{jj}^{(s+n+t)} = \mathbb{P}(X_{s+n+t} = j \mid X_0 = j)$$

$$\geq \mathbb{P}(X_{s+n+t} = j,\ X_{s+n} = k,\ X_s = k \mid X_0 = j)$$

$$= p_{jk}^{(s)} p_{kk}^{(n)} p_{kj}^{(t)}$$

$$= \alpha\, p_{kk}^{(n)}.$$

Similarly,

$$p_{kk}^{(s+n+t)} = \mathbb{P}(X_{s+n+t} = k \mid X_0 = k)$$

$$\geq \mathbb{P}(X_{s+n+t} = k,\ X_{t+n} = j,\ X_t = j \mid X_0 = k)$$

$$= p_{kj}^{(t)} p_{jj}^{(n)} p_{jk}^{(s)}$$

$$= \alpha\, p_{jj}^{(n)}.$$

$\Rightarrow$ For $n > u := s+t$, we have

$$\alpha\, p_{jj}^{(n-u)} \leq p_{kk}^{(n)} \leq \frac{1}{\alpha} p_{jj}^{(n+u)}$$

$\Rightarrow$ The series $\sum_n p_{jj}^{(n)}$ and $\sum_n p_{kk}^{(n)}$ converge / diverge simultaneously.

If j is recurrent, the theorem on page 14 implies that $\sum_n p_{jj}^{(n)}$ diverges; so that the series $\sum_n p_{kk}^{(n)}$ diverges, and applying the theorem again, we conclude that k is recurrent.

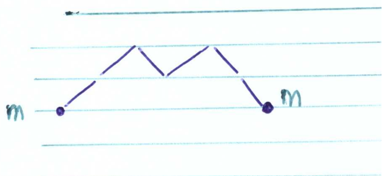
Remarks (i) In our random walk, we already saw that all states are transient.

(ii) This is not possible with a finite state MC: once a state is essential, it is recurrent. Indeed, in a finite MC, not all states can be transient; hence, at least one is recurrent; and all states communicating with it are also recurrent from the previous corollary; thus there can be no transient essential states in a finite MC.

Example. Random Walk.

- We apply Corollary 1 to the RW example.
- First, let's compute $p_{jj}^{(m)}$.

m odd, $p_{jj}^{(2n+1)} = 0$. look →



m even, $p_{jj}^{(2n)}$ = binomial probability of having exactly n successes in $2n$ trials.

$$= p_{00}^{(2n)} = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!\,n!}(pq)^n.$$

Use STIRLING FORMULA for large n:

$$n! \simeq \sqrt{2\pi n}\; n^n e^{-n}$$

- For asymmetric walks; $p \neq q$, we have

$$p_{jj}^{(2n)} \simeq \frac{\sqrt{2\pi 2n}\,(2n)^{2n} e^{-2n}}{(\sqrt{2\pi n}\, n^n e^{-n})^2}(pq)^n = \frac{1}{\sqrt{\pi n}}(4pq)^n.$$

Since $4pq<1$ for $p\neq q$, $p_{jj}^{(n)}$ vanishes geometrically fast and $\sum_n p_{jj}^{(n)}$ converges; so that the RW is transient (compare with p.13/14)

- For $p=q=\frac{1}{2}$, we have $p_{jj}^{(2n)} \simeq \frac{1}{\sqrt{\pi n}}$; the series diverges and the RW is recurrent.

Theorem 2: SOLIDARITY THEOREM

In any closed class $S_r \subset S$ of a MC $\{X_n\}$ with state space S, all the states $j \in S_r$ are

(i) Either recurrent or transient, and

(ii) Either periodic with a common period $d>1$, or aperiodic.

If the states from a class S_r are periodic with a period $d>1$, then one has the partition

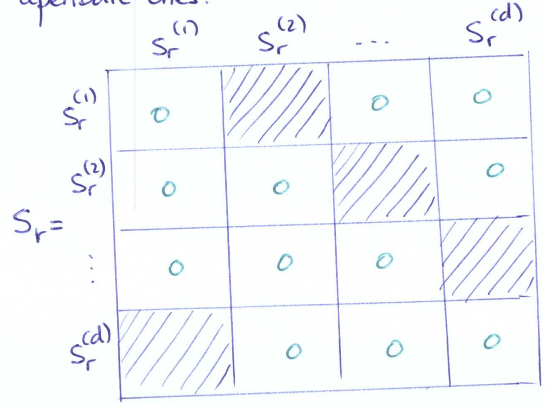
$$S_r = S_r^{(1)} + S_r^{(2)} + \cdots + S_r^{(d)}$$

From the subclass $S_r^{(i)}$, the MC goes w.p.1 to $S_r^{(i+1)}$, and from $S_r^{(d)}$ proceeds to $S_r^{(1)}$.

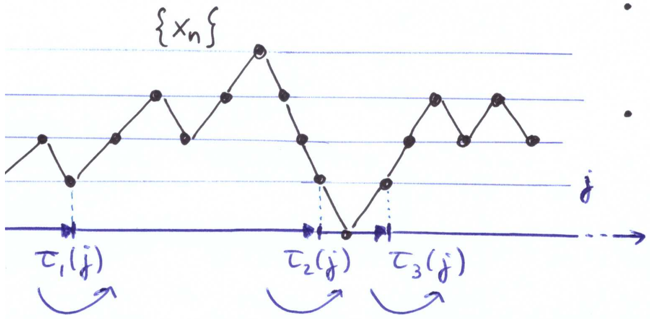
Rk: In our example on the bottom of page 12, the MC is periodic with $d=2$, and
$S_r^{(1)} = \{1\}$
$S_r^{(2)} = \{2, 3\}$

For a chain with transition matrix P^d (≡ original chain sampled at times $n = m + kd$, $k = 1, 2, 3, \ldots$), each subclass $S_r^{(i)}$ becomes a closed aperiodic class.

⇒ We can reduce the study of periodic MC to that of aperiodic ones.



II – LIMITING BEHAVIOUR OF MC.



- Let $\{X_n\} = $ MC starting at some $j \in S$.
- Put $\tau_i(j)$, $i = 1, 2, \ldots$ $=$ times between successive visits to the state j.

These RVs break the MC into independent cycles

- j transient $\Rightarrow$ $\tau_i = \infty$ w. positive proba.
- j recurrent $\Rightarrow$ $\tau_i < \infty$ w.p. 1.

↳ Expect some 'statistical regularity' i.e. the relative frequency of visits to j will tend to a certain number.

[It turns out, as we shall see, that we require that the mean time between successive visits to j is finite, $\mu_j := E\{\tau_i(j)\} < \infty$, to have statistical regularity].

Example: Consider a finite + irreducible MC.

- Claim: the RV $\tau = \tau_i(j)$ not only has a finite mean for any state $j \in S = \{1, \ldots, m\}$, but also has an exponentially fast vanishing tail;
  $P(\tau > n) < C a^n$ for some $a < 1$, $C < \infty$.

- Indeed, set $q_k(n) = P(X_n \neq j, \ldots, X_1 \neq j \mid X_0 = k)$

Note that

$$P(\tau > n) = P(X_n \neq j, \ldots, X_1 \neq j \mid X_0 = j) \leq q(n) := \max_k q_k(n)$$

$\Rightarrow$

$$q_k(n+t) \overset{\text{Law Total Probability}}{=} \sum_{r \neq j} P(X_{n+t} \neq j, \ldots, X_{n+1} \neq j, X_n = r, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = k)$$

$$\overset{\text{Markov Property}}{=} \sum_{r \neq j} P(X_{n+t} \neq j, \ldots, X_{n+1} \neq j \mid X_n = r) \times P(X_n = r, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = k)$$

$$= \sum_{r \neq j} q_r(t)\, P(X_n = r, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = k)$$

$$\leq q(t) \sum_{r \neq j} P(X_n = r, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = k)$$

$$= q(t)\, q_k(n)$$

$q_k(n+t) \leq q(t)\, q(n)$, $\forall k \Rightarrow q(n+t) \leq q(t)\, q(n)$.

$\Rightarrow$ For $n = rm$, $q(n) \leq [q(m)]^r = \left[q(m)^{1/m}\right]^n$.

- It remains to show that $q(m) < 1$.
  (then one can put $a = q(m)^{1/m}$ and $P(\tau > n) \leq q(n) \leq a^n$,
  $n = rm$

  if n is not a multiple of m, then need to consider an extra constant $C < \infty$ on the RHS.

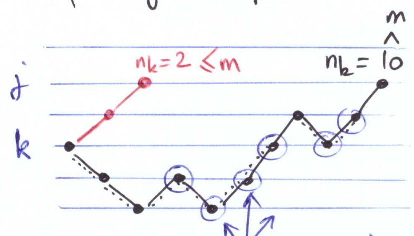
- Since $q(m) = \max_k q_k(m)$ by definition, consider an arbitrary state $k \in S$. We want to show that $\forall k \in S$, $q_k(m) < 1$.

Since the MC is irreducible, $\forall k \in S$, $\exists n_k < \infty$ s.t. you go from state k to state j in n_k steps with strictly positive probability, which is equivalent to

$$q_k(n_k) = P(X_{n_k} \neq j, \dots, X_1 \neq j \mid X_0 = k) < 1.$$

If $n_k > m$, among the n_k states forming this path, there must be repetitions as we have only m states.

$\Rightarrow$ $\exists$ cycles inside this path, and removing them will only increase the probability of the shorter path still leading from k to j.



$\Rightarrow$ The minimum n_k such that

$$P_{k i_1} P_{i_1 i_2} \cdots P_{i_{n_k - 1} j} > 0$$ is less than m, and thus

$\forall k \in S$, $q_k(m) < 1$. ■

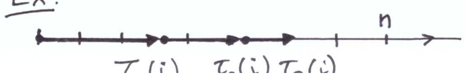
When the mean return time is finite, $\mu_j = E\{\tau_i(j)\} < \infty$, the SLLN ensures that

$$\frac{\tau_1 + \dots + \tau_v}{v} \simeq \mu_j = E\{\tau_i(j)\} \quad \text{iid } \forall i$$

so that $\underbrace{\tau_1 + \dots + \tau_v}_{} \simeq \mu_j v$.

$\Rightarrow$ After n ~~times~~ steps, the number v_n of visits to state j satisfies $n \simeq \mu_j v_n$, so that the relative frequency of visits to state j is $\frac{v_n}{n} \simeq \mu_j^{-1}$

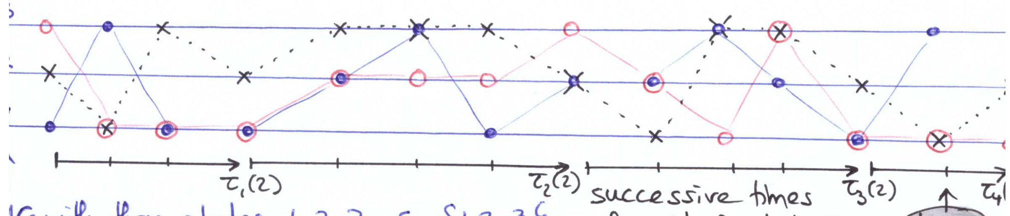
Ex:



$n \simeq \tau_1(j) + \tau_2(j) + \tau_3(j) \simeq 3\mu_j$

Terminology.

Recurrent states j with $\mu_j < \infty$ are called POSITIVE and those with $\mu_j = \infty$ ~~are~~ are called NULL since the limiting frequencies of visiting them are positive and null, respectively. □



MCs with three states 1, 2, 3, $S = \{1, 2, 3\}$, starting from 1, 2, and 3, respectively ($\bullet$, x, o):

—•— MC starting from state '1'

-x- MC '2'

-o- MC '3'

successive times of visit of state 2, for the MC starting at 2 (black crosses x)

↓

It turns out that by time n,

$$\frac{v_n^{(2)}}{n} \simeq \frac{1}{\mu_2}, \quad \bullet\ \mu_2 = E\,\tau(2),$$

$\bullet$ $v_n^{(2)}$ = # visit of state 2 by time n.

Similarly,

$$\frac{v_n^{(1)}}{n} \simeq \frac{1}{\mu_1}; \quad \frac{v_n^{(3)}}{n} \simeq \frac{1}{\mu_3}.$$

But then, what would you get if ~~the second~~ each process, ~~say~~, were actually iid observations?

↳ $\mathbb{P}$(obtaining a '1') $\overset{\text{SLLN}}{\simeq} \frac{\#\text{obs} = 1}{n} = \frac{v_n^{(1)}}{n}$ ← for process $\bullet$

Similarly, expect

$\mathbb{P}$(obtaining a 'j') $\simeq \frac{\#\text{obs} = j}{n} = \frac{v_n^{(j)}}{n}$ $(j = 1, 2, 3)$

$\Rightarrow$ You would hope that something similar holds for MC as well; that is the quantities $1/\mu_j$ correspond (for large n) to some probability of being in state j. And indeed, ...

Theorem. If a state j is aperiodic and $\mu_j = E\,\tau_1(j) < \infty$ (23)
then, for any $i \leftrightarrow j$,

$$p_{ij}^{(n)} \longrightarrow \frac{1}{\mu_j} \quad \text{as } n \to \infty.$$

If $\mu_j = \infty$, then $p_{ij}^{(n)} \to 0$.

Particularly interesting if the limit is the same, starting from any $i \in S$, which leads to the next definition:

i.e. when all states $i \in S$ communicate with each other; ie when there is only one class of essential states.

Def: If, in a MC $\{X_n\}$, the limits
- $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$
- $\sum_j \pi_j = 1$

exist and do not depend on i, then the MC is said to be <u>ERGODIC</u>.

And so, in view of the Theorem above, the only thing that can go wrong is that the limiting probabilities do not sum to 1 (in the case of a single essential state); which happens when the mean return time is ∞ (they sum to 0).
$\Rightarrow$ A criterion for ~~ergodicity~~ ergodicity; ie $\exists$ of a proba distrib $(\pi_1, \pi_2, \dots) =: \pi$ should exclude such cases.
[BTW, how would you find π?]

Well, when ergodicity takes place, the n-step transition matrices P^n converge (as $n\to\infty$), to a matrix of the form

$$\Pi = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \cdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

$\equiv$

independent trials: transition to the next step occurs with a proba which is indpt of ~~whole~~ current state.
[X_1 and X_{1+n} are $\approx$ independent for large n]

The distribution $\pi = (\pi_1, \pi_2, \dots)$ is called the <u>STATIONARY DISTRIBUTION</u> of $\{X_n\}$ (24)

a.k.a. the <u>Steady-State</u> or <u>Equilibrium</u> distribution.

<u>Important Remark</u>: For any initial distribution p $(X_0 \sim p)$, we have that

$$P(X_n = j) = (p P^n)_j = \sum_i p_i p_{ij}^{(n)} \longrightarrow \sum_i p_i \pi_j = \pi_j$$

page 7

i.e. $\boxed{p P^n \to \pi}$ (*)

An ergodic MC 'forgets' its original state.

$\Rightarrow$ When the distribution of X_0 coincides with π, the MC is said to be <u>STATIONARY</u> (or, in equilibrium).

In fact, the <u>process</u> X_t $(t \geq 0)$ is stricly stationary; its fdd are invariant by translation.

Also, we get from (*) that

$$\pi = \lim_{n\to\infty} p P^n = \lim_{n\to\infty} p P^{n-1} P = \left(\lim_{n\to\infty} p P^{n-1}\right) P = \pi P;$$

$\boxed{\pi = \pi P}$

Any ~~dis~~ satisfying this equation is called <u>INVARIANT</u> for the MC $\{X_n\}$.

<u>Theorem</u>. A Markov Chain is ergodic <u>iff</u>

(i) there is a single non-empty closed class of essential states, and the class is aperiodic.

(ii) $\exists$ a state j_0 s.t. for the recurrence time $\tau_1(j_0)$ to j_0 one has $E\{\tau_1(j_0) \mid X_0 = j_0\} < \infty$

There is more

For an ergodic MC, the limits $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$ form a unique solution to the system of linear equations

$$(S) = \begin{cases} \pi = \pi P \\ \sum_j \pi_j = 1 \end{cases}$$

Moreover, $\pi_j = 1/\mu_j$, where $\mu_j = E\{\tau_1(j) \mid X_0 = j\}$.

There is more

On the other hand, if a MC satisfies (i) and (S) has a unique solution $\pi_j \geq 0$, then the MC is ergodic and the solution is the chain's stationary distribution.

As a corollary, we see ergodicity for a finite state MC is equivalent to condition (i), since, as stated on pages 19/20/21, the return times have an exponentially fast vanishing tail.

↳ Cv towards π is also exp. fast : $|p_{ij}^{(n)} - \pi_j| < Ce^{n}$.

Remarks: (i) for a MC with finite state space $S = \{1, .., m\}$, there are $(m+1)$ equations in the system (S). However, the rank of the system $\pi = \pi P$ is $(m-1)$ as P is a stochastic matrix : all row sums are equal to 1.

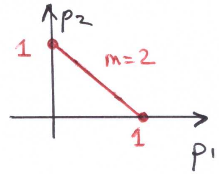
$\Rightarrow$ $(P - I)$ has rows summing to 0.

$\Rightarrow$ Practically, this means that we can always discard one of the equations from $\pi = \pi P$.

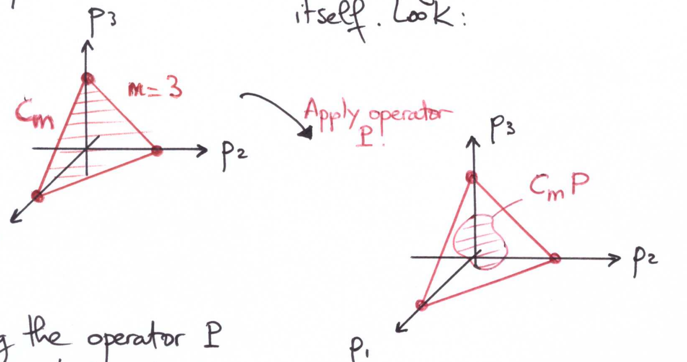
(ii) Geometrical interpretation.

Set of all proba. distributions on $\{1, .., m\}$ is an $(m-1)$ dimensional SIMPLEX C_m ; its vertices being the degenerate distributions (one $p_i = 1$, and the rest $= 0$)

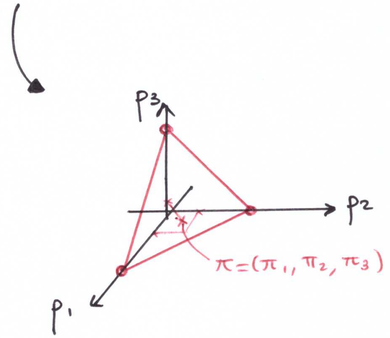
Simplex C_m:



It turns out that P is a CONTRACTION on C_m: the image $C_m P = \{pP \mid p \in C_m\}$ is strictly 'smaller' than C_m itself. Look:

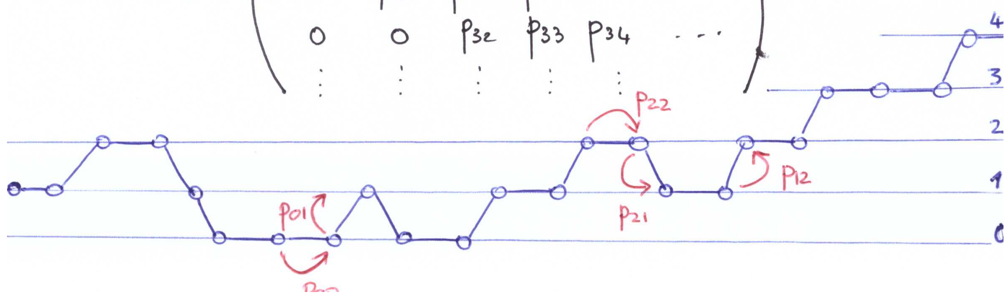


And so, applying the operator P recursively, the initial set C_m keeps contracting, until it vanishes to a single point, corresponding to the stationary distribution.



Example: positive random walk with jumps 0, ±1 ; with transition matrix

$$P \begin{pmatrix} p_{00} & p_{01} & 0 & 0 & 0 & \cdots \\ p_{10} & p_{11} & p_{12} & 0 & 0 & \cdots \\ 0 & p_{21} & p_{22} & p_{23} & 0 & \cdots \\ 0 & 0 & p_{32} & p_{33} & p_{34} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$



Assume
- $p_{j,j+1} > 0 \quad \forall j > 0$
- $p_{0,1} > 0$
- at least one of the diagonal elements $p_{jj} \geq 0$.

Then the MC is irreducible and aperiodic.

Stationary distribution $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ is solution to the system

$$(S) = \begin{cases} \pi_0 = p_{00}\,\pi_0 + p_{10}\,\pi_1 \\ \pi_1 = p_{01}\,\pi_0 + p_{11}\,\pi_1 + p_{21}\,\pi_2 \\ \pi_2 = p_{12}\,\pi_1 + p_{22}\,\pi_2 + p_{32}\,\pi_3 \\ \dots / \dots \end{cases}$$

The first equation yields $\pi_1 = \dfrac{(1-p_{00})}{p_{10}}\pi_0 = \dfrac{p_{01}}{p_{10}}\pi_0$.

Substituting this expression into the second equation yields

$$\pi_2 = \frac{1}{p_{21}}\left\{ \underbrace{(1-p_{11})}_{p_{10}+p_{12}}\pi_1 - p_{01}\pi_0 \right\} = p_{01}\left[\{p_{10}+p_{12}\} - p_{10}\right]\frac{\pi_0}{p_{21}p_{10}}$$

$$\Rightarrow \pi_2 = \frac{p_{01}p_{12}}{p_{21}p_{10}}\pi_0$$

The general formula is

$$\begin{cases} \pi_j = K_j\,\pi_0 \\ K_j = \dfrac{p_{01}\,p_{12}\cdots p_{j-1,j}}{p_{j,j-1}\cdots p_{21}\,p_{10}} \end{cases}, \quad K_0 = 1$$

Moreover, $\sum_{j\geq 0} \pi_j = 1 \Rightarrow 1 = \pi_0 \sum_{j \geq 0} K_j$, so that

$$\boxed{\pi_j = \frac{K_j}{\sum_{j\geq 0} K_j} \quad \text{if} \quad \sum_{j\geq 0} K_j < \infty}$$

If $\sum K_j = \infty$, then
- ↗ MC is transient (e.g. $p_{jj} = 0$, $p_{j,j+1} = p > 1/2$, $j \geq 1$) i.e. positive drift.
- ↘ MC is null-recurrent ($p = 1/2$; zero drift) ■

→ Calculation of mean recurrence times is usually hard to get directly ⇒ deriving ergodicity from Theorem page 24/25 is usually difficult.

→ Next, we present a criterion to check for ergodicity.

– **Theorem** [FOSTER CRITERION]

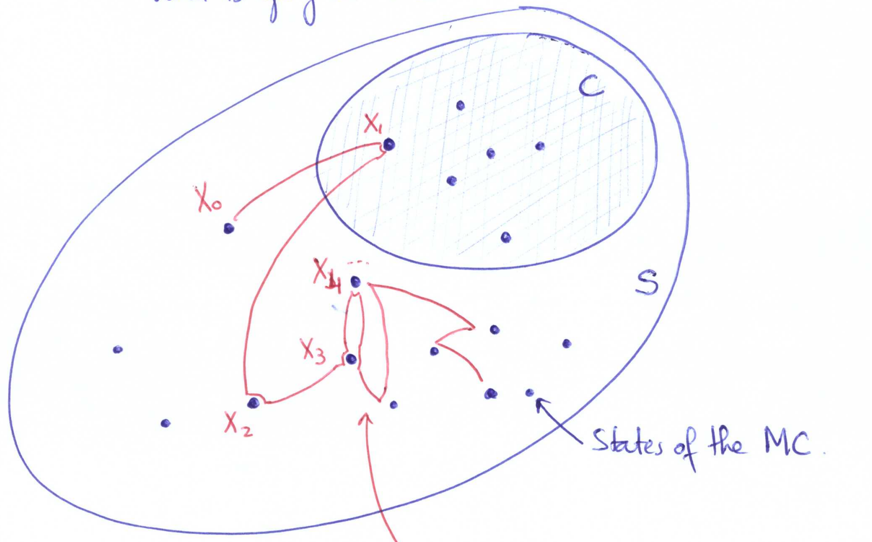
Let $\{X_n\}$ be an irreducible MC.

Suppose there exists a non-negative function V on S, and a finite subset $C \subset S$ such that, for some $\varepsilon > 0$,

$$E\{V(X_1) - V(j) \mid X_0 = j\} < \begin{cases} \infty & \text{for } j \in C \\ -\varepsilon & \text{for } j \notin C. \end{cases}$$

Then $\{X_n\}$ is ergodic.

What is going on here?



If the MC stays here, then the function V is decreased (on average) by some amount ε each step; which would eventually make V negative, which is impossible by assumption ⇒ the chain must regularly visit C, making it ergodic.

Criterion for ergodicity presented on pages 24 and 28 are for irreducible chains.

↳ What can we say about <u>reducible</u> chains ?

* Once a MC enters S_E, it stays there forever. Moreover, if there is more than one closed class of essential states, the MC remains in the first closed class S_r it enters. Denote by P_r the corresponding submatrix of P.

Class S_r = aperiodic + positive recurrent

The 'subchain' is ergodic and

$P(X_n = j \mid X_0 \in S_r) \to \pi_j^{(r)}, \quad j \in S_r$

for some $\pi_j^{(r)} \geq 0$; $\sum_{j \in S_r} \pi_j^{(r)} = 1$.

The limiting distribution of the MC is a mixture of the distrib $\pi^{(r)}$ on the disjoint sets S_r, with weights

$a_r = P(X_\tau \in S_r)$;

$\tau := \min\{n \geq 0 \mid X_n \in S_E\}$

Class S_r = aperiodic + null recurrent

ie there is no $j \in S_r$ with finite mean recurrence time.

There is no limiting proba distribution for $\{X_n\}$;

$P_{ij}^{(n)} \to 0 \quad \forall i, j \in S_r$

S_r = periodic (period $d > 1$)

Then one needs to consider the initial chain, sampled every d units of time, and cyclic subclasses $S_r^{(\ell)}$, $\ell = 1, \ldots, d$.

Then

$P(X_{nd+k} = j \mid X_0 \in S_r^{(\ell)}) \to \tilde{\pi}_j^{(r)}$

for $j \in S_r^{(\ell + k \ (\mathrm{mod}\ d))}$

[For non-neg $\tilde{\pi}_j^{(r)}$ which sum to one.]

Remarks on page 22 suggest that, for n large, the proportion of time the MC stays in a particular regime is given by π_j. We state this more formally in the next theorem. ~~We also~~ In addition, we state a CLT type result for MC:

<u>Theorem</u>: If a MC $\{X_n\}$ is ergodic with stationary distribution π then, for the number of visits $V_j(n) = \#\{t \leq n : X_t = j\}$ to any fixed state j of the MC,

$$\frac{V_j(n)}{n} \to \pi_j \quad \text{as } n \to \infty.$$

Moreover, if $\pi_j > 0$ and $\sigma^2 = \mathrm{Var}(\tau_1(j) \mid X_0 = j) < \infty$, then $\forall k \in S$, uniformly in $x \in \mathbb{R}$,

$$\mathbb{P}\left\{ \frac{V_j(n) - n\pi_j}{\sigma\sqrt{n\pi_j^3}} \leq x \;\middle|\; X_0 = k \right\} \to \Phi(x),$$

as $n \to \infty$.