# Brief notes about discrete time Markov Chains

Based on [N] ; [MP]

MARKOV CHAINS - J.R. NORRIS - Cambridge University Press 1997

A FIRST COURSE IN PROBABILITY AND MARKOV CHAINS
G. MODICA & L. POGGIOLINI - Wiley 2013

## (1) DEFINITIONS

### 1.1 STOCHASTIC VECTORS AND MATRICES

( $x(t) = S_i$ ⟹ the Markov chain visits i at time t )

A Markov chain is defined on a N-dimensional discrete state space (N can be $+\infty$) through the

stochastic matrix $W_{ij}$ : transition probability from j to i ( $\Rightarrow W_{ij} \geq 0 \;\forall i,j$ )

in a stochastic matrix columnes are normalized like probabilities: $\sum_i W_{ij} = 1 \quad \forall j$

stochastic vector $p_i^n$ : probability of being in state i at time n ( $\Rightarrow p_i \geq 0 \;\forall i$ )

stochastic vectors are normalized like probabilities: $\sum_i p_i = 1$

It can be proved that:

- W stochastic matrix, $\vec{p}$ stochastic vector
  $\Rightarrow W\vec{p}$ is a stochastic vector
- W stochastic matrix $\Rightarrow W^n$ is a stochastic matrix

## 1.2 STOCHASTIC UPDATE RULE

$$\vec{p}^{\,n+1} = W \vec{p}^{\,n} \longrightarrow p_i^{n+1} = \Sigma_j W_{ij} p_j^n$$

$$\vec{p}^{\,n} = W^n \vec{p}^{\,0} \longrightarrow p_i^{n} = \Sigma_j (W^n)_{ij} p_j^0$$

## 1.3 ACCESSIBILITY, IRREDUCIBILITY

- State $i$ is <u>accessible</u> from state $j$ if $\exists\, n \,/\, (W^n)_{ij} > 0$
  (it is possible to visit $i$ starting from $j$ in a finite time)

- The Markov chain is <u>irreducible</u> if all states are accessible from any other state
  (this $\not\Rightarrow \exists\, n \,/\, (W_{ij})^n > 0 \ \forall\, i,j$)

- EXAMPLE: the transition matrix $\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$ defines a reducible Markov chain: $(W^n)_{21} = 0 \ \forall n$ and state 2 is not accessible from state 1

## 1.4 PERIOD; APERIODICITY

- <u>period</u> of state $i$: $\tau_i = \mathrm{gcd}\{ n > 0 \,/\, (W^n)_{ii} > 0 \}$
  $\tau_i = 1 \Rightarrow$ state $i$ is <u>aperiodic</u> (gcd = greatest common divisor)

- If a Markov chain is irreducible $\Rightarrow$ all its states share the same period
  $\rightarrow T$ is the period of the Markov chain ($\tau_i = T \ \forall i$)

- A Markov chain is <u>aperiodic</u> if all its states have period 1: $\tau_i = 1 \ \forall i$

## 1.5 RECURRENT, TRANSIENT, POSITIVE RECURRENT

- $T_i$ = first return time to $i$ ; $T_i = \inf\{t \geq 1 \,/\, x(t) = S_i\}$

$T_i$ is a random variable (given that $X(0) = S_i$)

- $q_i^{(n)}$ is the associated probability: $q_i^{(n)}$ = prob. that $T_i = n$

$\sum_{1}^{\infty} {}_n\, q_i^{(n)}$ $= 1 \Rightarrow$ state $i$ is <u>persistent</u> (or recurrent)

↳ (we are SURE to go back to $i$ within a finite time)

$< 1 \Rightarrow$ state $i$ is <u>transient</u>

non zero prob. of not going back to $i$ in a finite time

- <u>mean recurrence time</u>: $\langle T_i \rangle = \sum_{1}^{\infty} {}_n\, n\, q_i^{(n)}$

(can be defined only for persistent states)

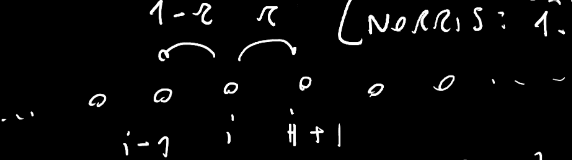
- $\langle T_i \rangle < +\infty \Rightarrow$ state $i$ is <u>positive-recurrent</u>

- A Markov chain is recurrent (or positive-recurrent) if all its states are recurrent (or positive-recurrent)

> N finite: a Markov chain is irreducible
> (M-P: th 5.7)) $\Rightarrow$ the Markov chain is positive-recurrent

$\Rightarrow$ the notions of IRREDUCIBLE, RECURRENT, POSITIVE-RECURRENT are distinct only in the $N = \infty$ case

- Example: unbiased random walk in $d = 1$ on an infinite lattice $\mathbb{Z}$ ($W_{i+1,i} = r$; $W_{i-1,i} = 1-r$) [NORRIS: example 1.6.1]




all states are accessible from any other in a finite time

$\Rightarrow$ irreducible Markov chain (if $0 < r < 1$)

but all states are TRANSIENT in the biased case ($r \neq 1/2$)

and all states are RECURRENT, but NOT POSITIVE-RECURRENT in the unbiased case ($r = 1/2$) [NORRIS, example 1.7.8]

## 2) STATIONARY DISTRIBUTIONS

### 2.1 EXISTENCE (N finite)

$$\exists\, \vec{w}^* / W \vec{w}^* = \vec{w}^* \rightarrow \sum_j W_{ij} w_j^* = w_i^*$$

($\vec{w}$ right eigenvector of $W$ with eigenvalue 1)

because $w^*_{e,i} = 1$ is left eigenvector of $W$ with eigenvalue 1

$$\left( \sum_j w^*_{e,j} \cdot W_{ji} = \sum_j W_{ji} = 1 = w^*_{e,i} \rightarrow \vec{w}_e^* = \vec{w}_e^* W \right)$$

and thus, for finite $N$, a right eigenvector exists with the same eigenvalue $\Rightarrow$ **ANY FINITE STOCHASTIC MATRIX ADMITS AT LEAST ONE STATIONARY DISTRIBUTION**

### 2.2 UNIQUENESS

- A Markov chain irreducible and positive-recurrent admits a unique invariant distribution $w_i^* = 1/\langle T_i \rangle$

(NORRIS, theorem 1.7.7)

- The condition of positive recurrence is crucial to normalize the invariant distribution. If the Markov chain is irreducible and recurrent, there exists a unique invariant "measure", but the measure cannot in general be normalized to obtain a distribution

(NORRIS, theorem 1.7.6).

- For finite $N$, any IRREDUCIBLE stochastic matrix admits a unique invariant distribution $\left[\begin{matrix}\text{irreducible} \\ \Downarrow \\ \text{positive-recurrent}\end{matrix}\right]$

## 2.3 CONVERGENCE

- For a Markov chain irreducible and aperiodic, with a stationary distribution $W_i^*$:

$$p_i^n \xrightarrow[n\to\infty]{} W_i^* \ \forall i \quad \text{and} \ (W^n)_{ij} \xrightarrow[n\to\infty]{} W_i^* \ \forall i,j$$

and for any choice of the initial condition $\vec{p}^{\,0}$:

$$\left( p_i^n = \textstyle\sum_j (W^n)_{ij}\, p_j^0 \xrightarrow[n\to\infty]{} \sum_j W_i^* p_j^0 = W_i^* \sum_j p_j = W_i^* \right)$$

[NORRIS: theorem 1.8.3]

$\left\{ \begin{array}{l} \text{Irreducible + positive-recurrent} \Rightarrow \text{unique } W_i^* \text{ exists} \\ \text{Aperiodic} \Rightarrow \text{convergence to } W_i^* \text{ for any initial condition} \end{array} \right.$

- Markov chains irreducible and periodic (period $d$):
the state space can be partitioned in $d$ subsets $C_s$, $s=0,1,\dots,d-1$, such that, if the initial condition $\vec{p}^{\,0}$ is "localized" in $C_0$,

$$\vec{p}_j^{\,nd+s} \xrightarrow[n\to+\infty]{} d/\langle T_j\rangle \ \text{and} \ (W^{nd+s})_{ji} \xrightarrow[n\to\infty]{} d/\langle T_j\rangle \ \forall s, \forall j \in C_s$$

[NORRIS: theorem 1.8.5]

The theorem holds for the case of irreducible non positive-recurrent Markov chains as well, when $\langle T_j \rangle = +\infty$.

The theorem holds for the case of irreducible, aperiodic ($d=1$), non positive-recurrent Markov chains, so that

$$p_j^n \xrightarrow[n\to\infty]{} 1/\langle T_j\rangle \ ;$$

but if $\langle T_j\rangle = +\infty$ $p_j^n$ converges in general to an invariant "measure" that cannot be normalized to yield a distribution:

in the above example (1.5 : random walk on $\mathbb{Z}$)

$\langle T_j \rangle = +\infty \ \forall j$ and $p_j^n \xrightarrow[n\to\infty]{} 0$ : the limit is <u>NOT</u> a distribution

- Example of periodic Markov chains:

random walk on a ring (Livi & Politi, 1.5.2):

- for N even: $d = 2$ ; $C_0$ = "even sites" ; $C_1$ = "odd sites"

starting from even sites: $p_j^{2n} \xrightarrow[n\to\infty]{} 2/N$ for even $j$

$\left( \begin{array}{l} w_j^* = 1/\langle T_j \rangle = 1/N \text{ is the} \\ \text{unique invariant distribution} \end{array} \right)$ $p_j^{2n} \xrightarrow[n\to\infty]{} 2/N$ for odd $j$

for a irreducible, finite (and thus positive-recurrent) Markov chain)

- for $r = 0$ or $r = 1$ (the random walk moves <u>always</u> in the same direction)

$d = N$ and

each site $j$ is a subset $C_j$ in the partition

$p_j^{Nn+j} \xrightarrow[n\to+\infty]{} N/N = 1$ for $p_i^0 = \delta_{i0}$ $\left[ \begin{array}{l} w_j^* = 1/N \text{ is} \\ \text{always the} \\ \text{unique invariant} \\ \text{distribution} \end{array} \right]$

## $\boxed{2.4}$ ERGODICITY

$\phi_j(n)$ = fraction of time spent in state $j$ after $n$ steps

- For irreducible Markov chains:

$\phi_j(n) \xrightarrow[n\to\infty]{} 1/\langle T_j \rangle$ (this again holds in general for $\langle T_j \rangle = +\infty$ with no positive-recurrence)

this is probably why some authors define irreducible Markov chains as "ergodic"

[NORRIS: Theorem 1.10.2]

- For irreducible and positive-recurrent Markov chains for any bounded function $f(X)$ defined on the state space: $\frac{1}{n}\sum_{k=0}^{n-1} f(X(n)) \xrightarrow[n\to+\infty]{} \sum_i f(x_i) W_i^*$

where $W_i^* = 1/\langle T_i \rangle$ is the unique invariant distribution

[NORRIS: theorem 1.10.2]

In practice: time average = ensemble average (at stationarity)

Aperiodicity (and thus convergence of $p_i^n$ to $W_i^*$) is NOT needed for the last theorem to hold.

In fact: $\frac{1}{n}\sum_{k=0}^{n-1} f(X(n)) \xrightarrow[n\to\infty]{} \frac{1}{N}\sum_i f(x_i) \quad \left(W_i^* = \frac{1}{N}\right)$

for a random walk on a ring, also in the periodic cases

# 3 DETAILED BALANCE

## 3.1 Detailed balance implies stationarity

- A Markov chain satisfying the detailed balance condition: $\underbrace{q_j W_{ij} = q_i W_{ji} \ \forall i,j}$ is said to be REVERSIBLE

for a distribution $q_i$

- $q_i$ satisfies DB for $W \Rightarrow q_i$ is stationary for $W$

$$\sum_i q_j W_{ij} = \sum_i q_i W_{ji} \Rightarrow q_j = \sum_i W_{ji} q_i$$

$$\left(\sum_i W_{ij} = 1\right) \qquad \underbrace{W\vec{q} = \vec{q}}$$

- Reversibility ⟶ invariance under time reversal
$\Updownarrow$
equilibrium : STRONGER condition than stationarity

- ERGODIC Markov chain : $\left\{ \begin{array}{l} \text{irreducible} \\ \text{aperiodic} \\ \text{positive-recurrent} \end{array} \right.$

- D.B $\not\Leftrightarrow$ ERGODIC Markov chains
  - Random walk on a ring with $N$ odd and bias $(r \neq 1/2)$ is an example of ergodic MC where D.B. does not hold
  - D.B. may hold for reducible M.C :

$$\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix} \; ; \; W_{21} = 0 \quad \text{D.B:} \; p_2^* W_{12} = p_1^* W_{21} = 0$$
$$\Rightarrow p_2^* = 0 \; ; \; p_1^* = 1 \longrightarrow W \vec{p}^* = \vec{p}^*$$

and convergence still holds: $\vec{p}^n \xrightarrow[n \to \infty]{} \vec{p}^*$ for any initial condition

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \; ; \; W_{21} = W_{12} = 0 \; ; \; \text{D.B.:} \; p_2^* W_{12} = p_1^* W_{21}$$

holds for any choice of $p_1^*, p_2^* \,/\, p_1^* + p_2^* = 1$

- D.B. + irreducibility $\Rightarrow$ unique equilibrium distribution
[Appendix C ; Livi & Politi] (aperiodicity then ensures convergence)
  - But careful: unbiased RW on $\mathbb{Z}$ $(W_{i+1,i} = W_{i-1,i})$

D.B. would imply $p_i \cancel{W_{i+1,i}} = p_{i+1} \cancel{W_{i,i+1}} \Rightarrow p_i = \text{const}$

but $p_i = \text{const}$ CANNOT be normalized, so

D.B cannot hold for any distribution

- N finite : if D. B. holds $\Rightarrow$ W has N real eigenvalues

$$p_i^* W_{ji} = p_j^* W_{ij} \Rightarrow A_{ji} = A_{ij}, \text{ with } A_{ij} = \sqrt{\frac{p_j^*}{p_i^*}}\, W_{ij}$$

The symmetric matrix A is a similarity transform of W

$\Downarrow$ N real eigenvalues $\quad$ $\Downarrow$ A, W have the same eigenvalues

$q_i$ eigenvector of A : $\sum_j A_{ij}\, q_j = \lambda\, q_i$ (eigenvalue $\lambda$)

$$\Rightarrow \sum_j \sqrt{p_j^*/p_i^*}\; W_{ij}\, q_j = \lambda\, q_i \rightarrow \sum_j W_{ij}\, q_j' = \lambda\, q_i'$$

with $q_i' = \sqrt{p_i^*}\, q_i$ eigenvalue of W (eigenvalue $\lambda$)

$q_i, q_i'$ are real-component eigenvector

- In general, a finite $(N \times N)$ stochastic matrix W does not have N eigenvalues.

$\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$ is in Jordan form : only 1 eigenvalue $(=1)$

eigenvector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

In general, the eigenvalues does not need to be real, except for $\lambda = 1$; but since $W_{ij} \in \mathbb{R}$, complex eigenvalues are present in complex conjugate pairs $(\lambda, \lambda^*)$. For example : RW on a ring

eigenvalues $\lambda_j = \cos\left(\frac{2\pi j}{N}\right) + i\,(1-2r)\sin\left(\frac{2\pi j}{N}\right)$ for $j = 0, 1, \dots, N-1$

$\lambda_j$ and $\lambda_{N-j}$ $(j > 0)$ are complex conjugate pairs ; $\left(\begin{array}{l}\text{if } N \text{ even ;} \\ \lambda_{N/2} = -1 \in \mathbb{R}\end{array}\right)$

# 4 PERRON-FROBENIUS THEOREM (finite matrices)

## 4.1 Regular matrices

- A REGULAR stochastic matrix : $\exists n / (W^n)_{ij} > 0 \ \forall i,j$
  (or PRIMITIVE)

Obviously REGULAR $\Rightarrow$ IRREDUCIBLE

- Perron-Frobenius theorem (for regular stochastic matrices)
  - the eigenvalue with maximum modulus is $\underbrace{\lambda^{(1)} = 1}$
  - $\lambda^{(1)}$ is non degenerate
  - The right eigenvector has all positive components: $W_i^{(1)} > 0$
  - The left eigenvector is $W_{\ell,i}^{(1)} = 1 \ \forall i$ $\left( \sum_i W_{\ell,i}^{(1)} W_i^{(1)} = 1 \right)$
  - for any other eigenvalue $\lambda$ (in general complex):
    $|\lambda| < 1$
  - No other eigenvector exists with all positive components

- W REGULAR $\Leftrightarrow$ W IRREDUCIBLE and APERIODIC

(M.P. : exercise 5.84)

infact under these conditions uniqueness of and convergence to the stationary distribution $W_i^{(1)}$ are granted

## 4.2 EXTENSION TO PERIODIC MATRICES

- Generalization of Perron-Frobenius theorem to stochastic *irreducible* matrices:

- $\lambda^{(1)} = 1$ is a non degenerate eigenvalue
- the right eigenvector has all positive components: $w_i^{(1)} > 0$
- the left eigenvector is $w_{e,i}^{(1)} = 1 \ \forall i$ $\left( \sum_i w_{e,i}^{(1)} \cdot w_i^{(1)} = 1 \right)$
- no other eigenvector exist with positive components
- $d$ is the period of $W$: $W$ has exactly $d$ eigenvalues with modulus 1: $\lambda^{(j)} = \exp\left( \frac{i(j-1)2\pi}{d} \right)$ (the $d$-th roots of 1: $\left[\lambda^{(j)}\right]^d = 1$, $j = 1, \dots, d$

- Example with the RW on a ring:
  - $N$ = even $\to d = 2$

$$j = 1 \to \lambda_1 = 1$$
$$j = N/2 \to \lambda_{N/2} = -1$$

the solutions of $\lambda^2 = 1$

eigenvalues

$$\lambda_j = \cos\left(\frac{2\pi j}{N}\right) + i(1-2r)\sin\left(\frac{2\pi j}{N}\right)$$
$$j = 0, 1, \dots, N-1$$

  - $r = 0$ or $r = 1 \to d = N$

$$\lambda_j = \cos\left(\frac{2\pi j}{N}\right) \pm i \sin\left(\frac{2\pi j}{N}\right) = \underbrace{\exp\left(\pm i \frac{2\pi j}{N}\right)}$$

the $N$ solutions of $\lambda^N = 1$

- $W$ stochastic, irreducible matrix:

$W$ with period $d$ ($d \geq 2$) $\Rightarrow W^d$ is reducible

[MP: exercise 5.85]

Example: RW on a ring with $r = 0$

$$W = \begin{pmatrix} 0 & 1 & 0 & \cdots & & & 0 \\ 0 & 0 & 1 & 0 & \cdots & & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & & & & & \ddots & 1 \\ 1 & 0 & \cdots & & & \cdots & 0 \end{pmatrix} ; \quad W^N = \mathbb{1}$$