I. MARKOV CHAINS

In the present chapter we will introduce the mathematical description of time-dependent random phenomena. We will begin treating the simple case in which the time evolution can be represented as a sequence of steps in discrete time, and the random variables describing the quantities evolving randomly in discrete time take values in a discrete space.

In the following, we will consider a probability space $(\Omega, \mathcal{F}, P)$ and a set E **at most countable**, which we will call **state space**.

All the random variables X which will be dealt with are measurable functions $X : \Omega \to E$ with discrete density:

$$E \ni k \rightsquigarrow v_k \stackrel{def}{=} P(X = k), \quad v_k \geq 0, \quad \sum_{k \in E} v_k = 1 \tag{1}$$

As discussed in the first chapter, a discrete density uniquely defines a law: having in mind (1), for the sake of simplicity, we will call v the law of X with innocuous abuse of notation.

Random processes with discrete time and discrete state space can be interpreted as random walks on the points of E. To the purpose of describing such processes we must know the *transition probability* from a generic point $k \in E$ to another one. Therefore, a central ingredient in out treatment is represented by the following:

Definizione 1 *A* **transition matrix** $\underline{\mathcal{P}}$ *on* E *is a matrix with real-valued coefficients, such that:*

1. $\forall\, i, j \in E \quad 0 \leq \mathcal{P}_{i \to j} \leq 1$
2. $\forall\, i \in E \quad \sum_{j=1}^{N} \mathcal{P}_{i \to j} = 1$

where the symbol $\mathcal{P}_{i \to j}$, $i, j \in E$ *denotes the matrix elements of* $\underline{\mathcal{P}}$.

The above requirements enable us to interpret $\mathcal{P}_{i \to j}$ as the probability of moving from $i \in E$ to $j \in E$ in one time step: the second one, in particular, corresponds to asking that the probability of transitioning from i to any other state is 1. We now give the fundamental:

Definizione 2 *Given a law* v *on* E *and a transition matrix* $\underline{\mathcal{P}}$ *on* E, *we call* **homogeneous Markov chain** *with sample space* E, *with initial law* v *and transition matrix* $\underline{\mathcal{P}}$, *a family:* $\{X_n\}_{n \geq 0}$ *of random variables* $X_n : \Omega \to E$ *such that:*

1. X_0 *has law* v
2. *whenever conditional probabilities make sense:*

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \\ = P(X_{n+1} = j | X_n = i) = \mathcal{P}_{i \to j} \end{aligned} \tag{2}$$

We stress that $P(X_{n+1} = j | X_n = i)$ has been assumed independent of n, whence the adjective **homogeneous** in definition (2).

II. THE RANDOM WALK ON $\mathbb{Z}$

To clarify ideas, we begin with a remarkable example. We let $E = \mathbb{Z}$ be the set of integer numbers, and $\{\xi_m\}_{m\in\mathbb{N}}$ a family of independent and identically distributed random variables taking values in $\{\pm 1\}$ such that:

$$P(\xi_m = \pm 1) = \frac{1}{2} \tag{3}$$

We define:

$$X_0 = 0, \quad X_n = \xi_1 + \xi_2 + \cdots + \xi_n \tag{4}$$

X_n has the natural interpretation of position, at time n, of a walker starting from 0 and moving left or right with probability 1/2 at each time step.

With this position, we have defined an homogeneous Markov chain with initial law v, $\mathbb{Z} \ni k \rightsquigarrow v_k = \delta_{k,0}$ and transition matrix:

$$\mathcal{P}_{i\to j} = \begin{cases} \frac{1}{2}, & |i-j| = 1 \\ 0, & otherwise \end{cases} \tag{5}$$

In fact, choosing by construction $i_0 = 0$ and $i, i_{n-1}, \ldots, i_1$ *nearest neighbours*):

$$\begin{aligned} & P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = 0) = \\ & = P(X_n + \xi_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = 0) = \\ & = \tfrac{P(X_n+\xi_{n+1}=j, X_n=i, X_{n-1}=i_{n-1},\ldots,X_0=0)}{P(X_n=i,X_{n-1}=i_{n-1},\ldots,X_0=0)} = \\ & = \tfrac{P(\xi_{n+1}=j-i, \xi_n=i-i_{n-1}\ldots,X_0=0)}{P(\xi_n=i-i_{n-1}\ldots,X_0=0)} = \\ & = P(\xi_{n+1} = j-i) = \mathcal{P}_{i\to j} \end{aligned} \tag{6}$$

where the fact that ξ_{n+1} is independent on all ξ_m with $m \le n$ has been used.

In the following, we will show that knowledge of the initial law and of the transition matrix makes possible to compute all the properties of the chain.

Nevertheless, it must be observed that, for this particular Markov chain, it is possible to compute the law of X_n with the following reasoning: let us introduce, for each time step, two integer-valued non-negative random variables R_n and L_n defined as:

$$R_n + L_n = n, \quad R_n - L_n = X_n \tag{7}$$

The straightforward interpretation of R_n and L_n is the number of right and left steps made by the walker before time n. Therefore:

$$R_n = \frac{1}{2}(X_n + n) \tag{8}$$

and thus:

$$P(X_n = k) = P\left(R_n = \frac{1}{2}(k+n)\right) \tag{9}$$

with the caution that, if $\frac{1}{2}(k+n)$ is not integer, $P(X_n = k) = 0$.

Since R_n is the number of right steps of the n total steps, it follows a binomial law with parameter $1/2$ (number of successes in n trials). In conclusion, if $\frac{1}{2}(k+n)$ is a non-negative integer less than n:

$$P(X_n = k) = \binom{n}{\frac{1}{2}(k+n)} \frac{1}{2^n} \tag{10}$$

We take the opportunity to observe that:

$$\begin{aligned} P(X_{n+1} = j) &= P\left(\bigcup_{h \in \mathbb{Z}} \{X_{n+1} = j, X_n = h\} \right) = \\ &= \sum_{h \in \mathbb{Z}} P(X_{n+1} = j, X_n = h) = \sum_{h \in \mathbb{Z}} P(X_{n+1} = j | X_n = h) P(X_n = h) = \\ &= \frac{1}{2} \left(P(X_n = j-1) + P(X_n = j+1) \right) \end{aligned} \tag{11}$$

Subtracting $P(X_n = j)$ from both members yields:

$$P(X_{n+1} = j) - P(X_n = j) = \frac{1}{2} \left(P(X_n = j-1) + P(X_n = j+1) - P(X_n = j) \right) \tag{12}$$

Equation has the form of a partial differential equation with time and space finite differences instead of derivatives, closely resembling the heat equation. Another analogy with the heat equation is represented by the following equalities, resulting from easy calculations:

$$\begin{aligned} E[X_n] &= \sum_{i=0}^{n} E[\xi_i] = 0 \\ var(X_n) &= \sum_{i=0}^{n} var(\xi_i) = n \end{aligned} \tag{13}$$

The mean distance covered by the walker scales with the square root of the number of steps, a typical property of diffusion processes.

The connection between a stochastic process and a partial differential equation is not a coincidence, but the first appearence of a general relationship between two apparently disconnected fields of Mathematics, which we will explore in detail in the next chapters.

III. RECURSIVE MARKOV CHAIN

The random walk on $\mathbb{Z}$ exemplifies a general procedure for constructing explicitly Markov chains. Let X_0 be a given random variable, and $\{U_m\}_{m \in \mathbb{N}}$ a sequence of independent and identically distributed uniform random variables in $(0,1)$. Let moreover $h : E \times (0,1) \to E$ be a function, until now arbitrary. Define:

$$X_{n+1} = h\left(X_n, U_{n+1}\right) \tag{14}$$

Repeting the calculation of the previous chapter it is easily concluded that $\{X_n\}_n$ is a Markov chain. Its transition matrix is readily obtained:

$$\begin{gathered} P(X_{n+1}=j|X_n=i) = \frac{P(h(X_n,U_{n+1})=j,X_n=i)}{P(X_n=i)} = \\ = \frac{P(h(i,U_{n+1})=j,X_n=i)}{P(X_n=i)} = P(h\,(i,U_{n+1})=j) = \mathcal{P}_{i\to j} \end{gathered} \tag{15}$$

This result is very useful since, having at our disposal a random number generator, the problem of simulating a Markov chain with initial law v and transition matrix $\underline{\mathcal{P}}$ is solved samping the initial state with probability v and iteratively applying $X_{n+1} = h\,(X_n, U_{n+1})$ where $h : E \times (0,1) \to E$ is a function such that $P(h\,(i,U)=j) = \mathcal{P}_{i\to j}$ if U is uniform in $(0,1)$.

Nota 3 From now on we will limit to the case in which the set E is finite. For the sake of simplicity, we will write $E = \{1,\dots,N\}$. Probabilities on the set E, when useful, will be identified with row vectors $\underline{v} = (v_1,\dots,v_N) \in \mathbb{R}^N$, $v_i \geq 0$, $\sum_{i=1}^N v_i = 1$.

IV. TRANSITION MATRIX AND INITIAL LAW

We will now show how initial law and transition matrix give an exhaustive knowledge of the corresponding homogeneous Markov chain. We begin computing the law $v^{(1)}$ of X_1:

$$P(X_1=k) = \sum_{h=1}^N P(X_0=h)P(X_1=k|X_0=h) = \sum_{h=1}^N v_h\,\mathcal{P}_{h\to k} \tag{16}$$

which can be written in matrix form recalling that laws can be represented through row vectors in $\mathbb{R}^N$:

$$\underline{v}^{(1)} = \underline{v}\,\underline{\mathcal{P}} \tag{17}$$

At the subsequent instant:

$$\begin{gathered} P(X_2=k) = \sum_{l=1}^N P(X_1=l)P(X_2=k|X_1=l) = \\ = \sum_{l=1}^N P(X_1=l)\,\mathcal{P}_{l\to k} = \sum_{l=1}^N \sum_{h=1}^N v_h\,\mathcal{P}_{h\to l}\,\mathcal{P}_{l\to k} = \\ = \sum_{h=1}^N v_h \sum_{l=1}^N \mathcal{P}_{h\to l}\,\mathcal{P}_{l\to k} \end{gathered} \tag{18}$$

that is:

$$\underline{v}^{(2)} = \underline{v}\,\underline{\mathcal{P}}^2 \tag{19}$$

Iterating this reasoning we easily conclude that the law at instant n is obtained applying to the row vector representing the initial law the n-th power of the transition matrix:

$$\underline{v}^{(n)} = \underline{v}\,\underline{\mathcal{P}}^n \tag{20}$$

It is interesting to observe that, denoting with $\mathcal{P}^{(m)}_{i\to j}$ the matrix elements of the m-th power $\underline{\mathcal{P}}^m$ of the transition matrix, one obtains the *m-step transition probabilities.*

$$\mathcal{P}^{(m)}_{i\to j} = P(X_{n+m}=j|X_n=i) \tag{21}$$

This can be shown iterating the following calculation:

$$\begin{gathered} P(X_{n+m}=j|X_n=i)=\tfrac{P(X_{n+m}=j,X_n=i)}{P(X_n=i)}= \\ \sum_h \tfrac{P(X_{n+m}=j,X_{n+m-1}=h,X_n=i)}{P(X_n=i)}= \\ =\sum_h \tfrac{P(X_{n+m}=j,X_{n+m-1}=h,X_n=i)}{P(X_{n+m-1}=h,X_n=i)}\tfrac{P(X_{n+m-1}=h,X_n=i)}{P(X_n=i)}= \\ =\sum_h P(X_{n+m}=j|X_{n+m-1}=h,X_n=i)P(X_{n+m-1}=h|X_n=i)= \\ =\sum_h \mathcal{P}_{h\to j}P(X_{n+m-1}=h|X_n=i) \end{gathered} \tag{22}$$

We eventually compute the joint laws ot the process in terms of the initial law v and of the transition matrix $\underline{\mathcal{P}}$, that is $P\left(X_{n_1}=i_1,\dots,X_{n_k}=i_k\right)$, $0\le n_1<\dots<n_k$:

$$\begin{gathered} P\left(X_{n_1}=i_1,\dots,X_{n_k}=i_k\right)= \\ =P\left(X_{n_1}=i_1,\dots,X_{n_{k-1}}=i_{k-1}\right)P\left(X_{n_k}=i_k|X_{n_1}=i_1,\dots,X_{n_{k-1}}=i_{k-1}\right)= \\ =P\left(X_{n_1}=i_1,\dots,X_{n_{k-1}}=i_{k-1}\right)\mathcal{P}^{(n_k-n_{k-1})}_{i_{k-1}\to i_k}=\dots= \\ =P\left(X_{n_1}=i_1\right)\mathcal{P}^{(n_2-n_1)}_{i_1\to i_2}\dots\mathcal{P}^{(n_k-n_{k-1})}_{i_{k-1}\to i_k}= \\ =\sum_j v_j\mathcal{P}^{(n_1)}_{j\to i_i}\mathcal{P}^{(n_2-n_1)}_{i_1\to i_2}\dots\mathcal{P}^{(n_k-n_{k-1})}_{i_{k-1}\to i_k} \end{gathered} \tag{23}$$

Nota 4 *Rewriting the* 2- *and* 3- *times joint laws in the form:*

$$P\left(X_{n_1}=i_1,X_{n_3}=i_3\right)=P\left(X_{n_1}=i_1\right)\mathcal{P}^{(n_3-n_1)}_{i_1\to i_3} \tag{24}$$

$$P\left(X_{n_1}=i_1,X_{n_2}=i_2,X_{n_3}=i_3\right)=P\left(X_{n_1}=i_1\right)\mathcal{P}^{(n_2-n_1)}_{i_1\to i_2}\mathcal{P}^{(n_3-n_2)}_{i_2\to 3} \tag{25}$$

the relation:

$$P\left(X_{n_1}=i_1,X_{n_3}=i_3\right)=\sum_{i_2=1}^{N}P\left(X_{n_1}=i_1,X_{n_2}=i_2,X_{n_3}=i_3\right) \tag{26}$$

is equivalent to the followig **Chapman-Kolmogorov equation** *for the m-step transition probability:*

$$\mathcal{P}^{(n_3-n_1)}_{i_1\to i_3}=\sum_{i_2=1}^{N}\mathcal{P}^{(n_2-n_1)}_{i_1\to i_2}\mathcal{P}^{(n_3-n_2)}_{i_2\to 3},\quad 0\le n_1<n_2<n_3 \tag{27}$$

The property (27)*, quite natural in the present context since it is known that the m-step transition probability is obtained computing the m-th power of the transition matrix, will turn out of great importance when dealing with continous-time Markov processes taking values in* $\mathbb{R}^d$*.*

V. INVARIANT LAWS

Given a probability distribution π on E, which as seen before can be represented with a row vector $\underline{\pi}=(\pi_1,\dots,\pi_N)\in\mathbb{R}^N$, and a homogeneous Markov chain with transition matrix $\underline{\mathcal{P}}=\left\{\mathcal{P}_{i\to j}\right\}_{i,j}$ and initial law v, we will say that π is **invariant** provided that:

$$\underline{\pi}=\underline{\pi}\,\underline{\mathcal{P}},\quad i.e\quad \pi_k=\sum_{h\in E}\pi_h\,\mathcal{P}_{h\to k} \tag{28}$$

We stress that if the initial law v is invariant, X_n has law v for all n: all the X_n have the same law, giving rise to a **stationary Markov process**.

We now prove an important result:

Teorema 5 (Markov-Kakutani) *Any transition matrix $\underline{\mathcal{P}}$ admits has at least an invariant law.*

Dimostrazione 1 *We first observe that there is a one-to-one correspondence between probabilities on E ad points in the set: set:*

$$S = \left\{ \underline{x} \in \mathbb{R}^N : 0 \leq x_i \leq 1, \sum_{i=1}^{N} x_i = 1 \right\} \tag{29}$$

S is a closed and limited set in $\mathbb{R}^N$, and therefore compact: by virtue of Bolzano-Weierstrass theorem, any sequence in S has a convergent subsequence. Given a generic point $\underline{x} \in S$, consider the sequence:

$$\underline{x}_n = \frac{1}{n} \sum_{k=0}^{n-1} \underline{x}\, \underline{\mathcal{P}}^k \tag{30}$$

Obviously $\underline{x}_n$ has non-negative components. Moreover, $\underline{x}_n \in S$ as the following simple calculation shows:

$$\sum_i x_{n,i} = \frac{1}{n} \sum_{k=0}^{n-1} \sum_h \sum_i x_h \mathcal{P}^{(k)}_{h\to i} = \frac{1}{n} \sum_{k=0}^{n-1} \sum_h x_h = 1 \tag{31}$$

where we have taken into account the fact that $\mathcal{P}^{(k)}_{h\to i}$ is the probability of moving from h to i in k steps, and therefore $\sum_i \mathcal{P}^{(k)}_{h\to i} = 1$.

Since $\{\underline{x}_n\}_n \subset S$ it has a subsequence: $\left\{\underline{x}_{n_k}\right\}_{n_k}$ converging to a point $\underline{\pi} \in S$. We observe that:

$$\underline{x}_{n_k} - \underline{x}_{n_k} \underline{\mathcal{P}} = \frac{1}{n_k} \left(\sum_{h=0}^{n_k-1} \underline{x}\, \underline{\mathcal{P}}^h - \sum_{h=0}^{n_k-1} \underline{x}\, \underline{\mathcal{P}}^{h+1} \right) = \frac{1}{n_k} \left(\underline{x} - \underline{x}\, \underline{\mathcal{P}}^{n_k} \right) \tag{32}$$

and since the quantity $\underline{x} - \underline{x}\, \underline{\mathcal{P}}^{n_k}$ is limited by construction:

$$\underline{\pi} - \underline{\pi}\, \underline{\mathcal{P}} = \lim_{k\to+\infty} \left(\underline{x}_{n_k} - \underline{x}_{n_k} \underline{\mathcal{P}} \right) = \lim_{k\to+\infty} \frac{1}{n_k} \left(\underline{x} - \underline{x}\, \underline{\mathcal{P}}^{n_k} \right) = 0 \tag{33}$$

which completes the proof.

We observe that the proof of Markov-Kakutani's theorem is constructive: any sequence $\underline{x}_n$, once thinned out, converges to an invariant law, which may be not unique. Since the point $\underline{x} = \underline{x}_0$ is completely arbitrary, it can be chosen $x_h = \delta_{h,i}$, corresponding to a probability distribution concentrated at the point i. Were that the case:

$$x_{n,j} = \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{P}^{(k)}_{i\to j} \tag{34}$$

If $\{X_n\}$ is the Markov chain with initial law $\underline{x}$, concentrated at the point i with transition matrix $\underline{\mathcal{P}}$, we know that:

$$\mathcal{P}^{(k)}_{i\to j} = P(X_k = j) \tag{35}$$

Therefore:

$$x_{n,j} = \frac{1}{n}\sum_{k=0}^{n-1}\mathcal{P}^{(k)}_{i\to j} = \frac{1}{n}\sum_{k=0}^{n-1}P(X_k = j) = E\left[\frac{1}{n}\sum_{k=0}^{n-1}1_{\{X_k=j\}}\right] \tag{36}$$

and $x_{n,j}$ coincides with the expectation of the random variable:

$$\frac{1}{n}\sum_{k=0}^{n-1}1_{\{X_k=j\}} \tag{37}$$

representing the fraction of time the process has spent in the state j before the n-th time step. Remarkably, for large n the expectation of such random variable approximates the j-th component of one invariant law.

To compute the invariant law(s), the following problem must be solved:

$$\pi_j = \sum_{i=1}^{n}\pi_i\,\mathcal{P}_{i\to j} \tag{38}$$

To this purpose, the following interesting result holds, which represent a sufficient condition for a law π to be invariant:

Teorema 6 *If a law π satisfies the* **detailed balance equation***:*

$$\pi_i\,\mathcal{P}_{i\to j} = \pi_j\,\mathcal{P}_{j\to i}, \quad \forall i,j\in E \tag{39}$$

then it is invariant.

Dimostrazione 2 *The proof is extremely simple:*

$$\sum_{i=1}^{n}\pi_i\,\mathcal{P}_{i\to j} = \sum_{i=1}^{n}\pi_j\,\mathcal{P}_{j\to i} = \pi_j \tag{40}$$

A transition matrix may have, in general, infinite stationary laws: in fact, as a simple calculation shows, if π and π' are distinct stationary laws for $\underline{\mathcal{P}}$, any convex linear combination of π and π' is still a stationary law for $\underline{\mathcal{P}}$.

It is therefore very interesting to investigate the uniqueness of the invariant law. To this purpose, we introduce the following:

Definizione 7 *Let $\underline{\mathcal{P}} = \{\mathcal{P}_{i\to j}\}_{i,j}$ be the transition matrix of a homogeneous Markov chain.*

1. *$\underline{\mathcal{P}} = \{\mathcal{P}_{i\to j}\}$ is* **irreducibile** *if for all $i,j\in E$ there exists an integer $m = m(i,j) > 0$ such that $\mathcal{P}^{(m)}_{i\to j} > 0$.*

2. $\underline{\mathcal{P}} = \{\mathcal{P}_{i\to j}\}$ *is* **regular** *if there exists a number* $m > 0$ *such that* $\mathcal{P}^{(m)}_{i\to j} > 0$ *for all* $i, j \in E$.

A regular trasition matrix is always irreducible, but the converse is not true in general. Nevertheless, the following result holds:

Lemma 8 *If a transition matrix is irreducible, and there exists* $h \in E$ *such that* $\mathcal{P}_{h\to h} > 0$*, it is regular.*

Dimostrazione 3 *If for all* $i, j \in E$ *there exists* $m = m(i, j) > 0$ *such that* $\mathcal{P}^{(m)}_{i\to j} > 0$*, chosen* $s = \max_{i,j\in E} m(i, j)$ *we have* $\mathcal{P}^{(2s)}_{l\to k} > 0$ *for all* $l, k \in E$*, as the following inequality makes clear:*

$$\mathcal{P}^{(2s)}_{l\to k} \geq \mathcal{P}^{(n(l,h))}_{l\to h}\mathcal{P}_{h\to h}\dots\mathcal{P}_{h\to h}\mathcal{P}^{(n(h,k))}_{h\to k} > 0 \tag{41}$$

in which the term $\mathcal{P}_{h\to h}$ *appears* $2s - n(l, h) - n(h, k)$ *times.*

Nota 9 *At a first glance, it might seem very difficult to verify whether a chain is irreducible or not, but there exist observations that can considerably simplify the calculations involved: chosen two states* i, j*, if the chain is irreducible there exists* $m > 0$*, in general depending on the couple* (i, j) *of interest, such that* $\mathcal{P}^{(m)}_{i\to j} > 0$*; since the transition matrix has non-negative elements, this corresponds to the existence of at least one* $(m-1)$*-tuple of states* $k_1, \dots, k_{m-1}$ *such that:*

$$0 < \mathcal{P}_{i\to k_1}\mathcal{P}_{k_1\to k_2}\dots\mathcal{P}_{k_{m-1}\to j} \leq \mathcal{P}^{(m)}_{i\to j} \tag{42}$$

Intuitively, it is necessary to move from i *to* j *passing through points in* E *making steps with non-zero transition probability. A crucial role is played by elements of the transition matrix which are often denoted with stars in standard Markov chains textbooks, for instance:*

$$\underline{\mathcal{P}} = \begin{pmatrix} 0 & \star & 0 & 0 & 0 & 0 \\ 0 & \star & 0 & 0 & \star & 0 \\ 0 & 0 & 0 & 0 & \star & 0 \\ 0 & \star & 0 & 0 & \star & 0 \\ \star & \star & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \star \end{pmatrix} \tag{43}$$

We leave to the reader the task of convincing that the above matrix is not irreducible.

On the other hand, it is easy to realize that, given an irreducible transition matrix, every other transition matrix having at least the same configuration of stars is irreducible.

We are now ready to show the following fundamental result:

Teorema 10 (Markov) *If a transition matrix is regular, it admits a unique invariant law* π *and, for all initial laws* v*:*

$$\pi_j = \lim_{n\to+\infty} \left(\underline{v}\,\underline{\mathcal{P}}^n\right)_j \tag{44}$$

Dimostrazione 4 *By virtue of Markov-Kakutani's theorem we already know that an invariant law* π *exists. By definition:*

$$\underline{\pi} = \underline{\pi}\,\underline{\mathcal{P}}, \quad \sum_k \pi_k = 1, \quad 0 \leq \pi_k \leq 1 \tag{45}$$

Consider the 1*-dimensional vector subspace of* $\mathbb{C}^N$ *spanned by* π*:*

$$\mathcal{V}_\pi = \left\{ \underline{u} \in \mathbb{C}^N | \underline{u} = t\underline{\pi},\, t \in \mathbb{C} \right\} \tag{46}$$

As already discussed, $\underline{\pi} \in \mathcal{V}_\pi$ *is an eigenvector of* $\underline{\mathcal{P}}$*. We now define another subspace:*

$$\mathcal{V}_0 = \left\{ \underline{y} \in \mathbb{C}^N | \sum_k y_k = 0 \right\} \tag{47}$$

having dimension $M-1$*, and of course:*

$$\mathcal{V}_0 \cap \mathcal{V}_\pi = \{\underline{0}\} \tag{48}$$

since elements in $\mathcal{V}_\pi$ *are such that the sum of their components equals* t*.* $\mathbb{C}^N$ *can be expressed as direct sum of* $\mathcal{V}_0$ *and* $\mathcal{V}_\pi$*: every element* $\underline{v} \in \mathbb{C}^N$ *can be uniquely written as:*

$$\underline{v} = t\underline{\pi} + \underline{y}, \quad \underline{y} \in \mathcal{V}_0 \tag{49}$$

The eigenvalue equation for $\underline{\mathcal{P}}$ *in* $\mathbb{C}^N$ *has the form:*

$$\underline{v}\,\underline{\mathcal{P}} = \lambda\,\underline{v}, \quad \lambda \in \mathbb{C} \tag{50}$$

We observe that the map:

$$\mathcal{V}_0 \ni \underline{y} \rightsquigarrow \underline{y}\,\underline{\mathcal{P}} \tag{51}$$

leaves invariant $\mathcal{V}_0$*, since:*

$$\sum_k (\underline{y}\,\underline{\mathcal{P}})_k = \sum_k \sum_i y_i \mathcal{P}_{i\to k} = \sum_i y_i \sum_k \mathcal{P}_{i\to k} = \sum_i y_i = 0 \tag{52}$$

Therefore, if an element $\underline{v}$ *satisfies* (50)*:*

$$\underline{v}\,\underline{\mathcal{P}} = t\underline{\pi}\,\underline{\mathcal{P}} + \underline{y}\,\underline{\mathcal{P}} = \lambda t\underline{\pi} + \lambda\underline{y} \tag{53}$$

Since $\underline{y}\,\underline{\mathcal{P}} \in \mathcal{V}_0$*, and the representation of* $\underline{v}\,\underline{\mathcal{P}}$ *as an element of the direct sum of* $\mathcal{V}_0$ *and* $\mathcal{V}_\pi$ *is unique:*

$$\begin{cases} t\underline{\pi}\,\underline{\mathcal{P}} = \lambda t\underline{\pi} \\ \underline{y}\,\underline{\mathcal{P}} = \lambda\underline{y} \end{cases} \tag{54}$$

If $\underline{y} = 0$ *we find the already known eigenvalue* π *relative to the eigevalue* 1*. We now discuss the eigenvalue equation:*

$$\underline{y}\,\underline{\mathcal{P}} = \lambda\underline{y}, \quad i.e. \quad \lambda y_i = \sum_k y_k\,\mathcal{P}_{k\to i} \tag{55}$$

with $\mathcal{V}_0 \ni \underline{y} \neq 0$*. Computing the moduli and summing over* i *yields:*

$$|\lambda| \sum_i |y_i| = \sum_i |\sum_k y_k \, \mathcal{P}_{k\to i}| \leq \sum_i \sum_k |y_k| \mathcal{P}_{k\to i} = \sum_k |y_k| \tag{56}$$

that is:

$$|\lambda| \leq 1 \tag{57}$$

The arguments presented so far hold for a generic trasition matrix, not necessarily regular.

For an arbitrary regular transition matrix $\underline{\mathcal{P}}$ *there exists a number* $m > 0$ *such that* $\underline{\mathcal{P}}^m$*, which is itself a transition matrix, has only positive elements. When applied to* $\underline{\mathcal{P}}^m$*,* (56) *holds as a strict inequality.*

In fact, the modulus of the sum of several complex numbers equals the sum of their modules if and only if all the addends share the same argument. By virtue of the hypothesis $\sum_k y_k = 0$*, the components* y_k *cannot have the same argument; on the other hand, the arguments of* $y_k \, \mathcal{P}^m_{k\to i}$ *and* y_k *being equal:*

$$|\lambda| < 1 \tag{58}$$

For an arbitrary regular transition matrix $\underline{\mathcal{P}}$ *the linear operator:*

$$\mathcal{V}_0 \ni \underline{y} \rightsquigarrow \underline{y} \, \underline{\mathcal{P}} \in \mathcal{V}_0 \tag{59}$$

has eigenvalues with modulus strictly smaller than 1*.*

For an arbitrary initial law v*, we can write:*

$$\underline{v} \, \underline{\mathcal{P}}^n = (\underline{\pi} + \underline{v} - \underline{\pi}) \, \underline{\mathcal{P}}^n = \underline{\pi} + (\underline{v} - \underline{\pi}) \, \underline{\mathcal{P}}^n \tag{60}$$

And since $\underline{v} - \underline{\pi} \in \mathcal{V}_0$ *the previously discussed properties of the eigenvalues of the linear operator* (59) *make possible to conclude that:*

$$\lim_{n\to+\infty} \underline{v} \, \underline{\mathcal{P}}^n = \underline{\pi} + \lim_{n\to+\infty} \left\{ (\underline{v} - \underline{\pi}) \, \underline{\mathcal{P}}^n \right\} = \underline{\pi} \tag{61}$$

which completes the proof.

We have recalled the fact that if an $m \times m$ *complex-valued matrix* $\underline{A}$ *has eigenvalues with modulus strictly smaller than* 1*:*

$$\lim_{n\to+\infty} \underline{A}^n \, \underline{x} = 0, \quad \forall \underline{x} \in \mathbb{C}^m \tag{62}$$

as a consequence of Gel'fand's formula:

$$\max\left\{|\lambda| \, , \lambda \ \textit{autovalore di} \ \underline{A}\right\} = \lim_{n\to+\infty} ||\underline{A}^n||^{\frac{1}{n}} \tag{63}$$

for which we remind the reader to Functional Analysis textbooks.

VI. METROPOLIS ALGORITHM

Consider a given probability distribution π on E: we can ask ourselves whether there exists a transition matrix $\underline{\mathcal{P}}$ such that, for all initial laws v:

$$\pi_j = \lim_{n\to+\infty} \left(\underline{v} \, \underline{\mathcal{P}}^n\right)_j \tag{64}$$

Were that the case, we could construct a Markov chain $\{X_n\}_n$ with law converging, as n tends to infinity, to π in the sense precised by (64).

As the reader might have guessed, this possibility has deep implications in the field of simulations.

To this purpose, it turns out to be necessary to assume that $\pi_j > 0$ for all the states $j \in E$ and that π is not the uniform distribution.

Let now $\underline{\mathcal{Q}} = \{\mathcal{Q}_{i\to j}\}$ be a symmetric and irreducible transition matrix, $\mathcal{Q}_{i\to j} = \mathcal{Q}_{j\to i}$, subject to no other restrictions, and define:

$$\mathcal{P}_{i\to j} = \begin{cases} \mathcal{Q}_{i\to j}, & i \neq j,\ \pi_j \geq \pi_i \\ \mathcal{Q}_{i\to j}\frac{\pi_j}{\pi_i}, & i \neq j,\ \pi_j < \pi_i \\ 1 - \sum_{j\neq i} \mathcal{P}_{i\to j}, & i = j \end{cases} \tag{65}$$

Teorema 11 (N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, 1 *If $\pi_j > 0$ for all the states $j \in E$ and π is not uniform, for all the initial laws v the Markov chain $\{X_n\}_n$ with initial law v and transition matrix $\underline{\mathcal{P}}$ is regular, and has π as unique invariant distribution. Consequently:*

$$\pi_j = \lim_{n\to+\infty} \left(\underline{v}\,\underline{\mathcal{P}}^n\right)_j = \lim_{n\to+\infty} P(X_n = j) \tag{66}$$

Dimostrazione 5 *It is sufficient to show that π satisfies the detailed balance equation with $\underline{P}$ as transition matrix. Chosen a couple of states (i,j) such that, without loss of generality, $\pi_j \leq \pi_i$, $\mathcal{P}_{i\to j} = \mathcal{Q}_{i\to j}\frac{\pi_j}{\pi_i}$ whereas $\mathcal{P}_{j\to i} = \mathcal{Q}_{j\to i}$ and thus:*

$$\pi_i \mathcal{P}_{i\to j} = \pi_i \mathcal{Q}_{i\to j}\frac{\pi_j}{\pi_i} = \mathcal{Q}_{i\to j}\pi_j = \pi_j \mathcal{P}_{j\to i} \tag{67}$$

where the hypothesis that $\underline{\mathcal{Q}}$ is symmetric has been used. As a consequence of (67) π is invariant.

It remains to show that the Markov chain (65) is regular; first, we show that it is irreducible. In fact, if $i \neq j$ and $\mathcal{Q}_{i\to j} > 0$, then, by construction $\mathcal{P}_{i\to j} > 0$; this means that $\underline{\mathcal{P}}$ shares the star structure of $\underline{\mathcal{Q}}$, which is irreducible by definition.

To prove that (65) is regular, by virtue of lemma (8) it is sufficient to show that there exists $i_0 \in E$ such that $\mathcal{P}_{i_0\to i_0} > 0$. Since π is not uniform, there exists a proper subset $M \subset E, M \neq E$ of E on which π takes maximum value; due to the irreducibility of $\underline{\mathcal{Q}}$ the chain can move outside M, and therefore there exist $i_0 \in M$ and $j_0 \in M^c$ such that $\mathcal{Q}_{i_0\to j_0} > 0$ and, by construction, $\pi_{i_0} > \pi_{j_0}$. Moreover, $\mathcal{P}_{i\to j} \leq \mathcal{Q}_{i\to j}$ if $i \neq j$. These intermediate results imply:

$$\begin{gathered} \mathcal{P}_{i_0\to i_0} = 1 - \textstyle\sum_{j\neq i_0} \mathcal{P}_{i_0\to j} = 1 - \sum_{j\neq i_0,j_0} \mathcal{P}_{i_0\to j} - \mathcal{P}_{i_0\to j_0} \geq \\ \geq 1 - \textstyle\sum_{j\neq i_0,j_0} \mathcal{Q}_{i_0\to j} - \mathcal{Q}_{i_0\to j_0}\frac{\pi_{j_0}}{\pi_{i_0}} = \\ = 1 - \textstyle\sum_{j\neq i_0} \mathcal{Q}_{i_0\to j} + \mathcal{Q}_{i_0\to j_0}\left(1 - \frac{\pi_{j_0}}{\pi_{i_0}}\right) = \mathcal{Q}_{i_0\to i_0} + \mathcal{Q}_{i_0\to j_0}\left(1 - \frac{\pi_{j_0}}{\pi_{i_0}}\right) \geq \\ \geq \mathcal{Q}_{i_0\to j_0}\left(1 - \frac{\pi_{j_0}}{\pi_{i_0}}\right) > 0 \end{gathered} \tag{68}$$

that is, the chain is regular by virtue of lemma (8) and admits a unique stationary law by virtue of Markov's theorem.

Metropolis' Theorem is widely used in Physics, where it provides a technique for simulating random variables with law given by:

$$\pi_i = \frac{e^{-\beta \mathcal{H}(i)}}{Z(\beta)}, \quad \mathcal{H} : E \to \mathbb{R}, \quad Z(\beta) = \sum_{i \in E} e^{-\beta \mathcal{H}(i)} \tag{69}$$

E being the configuration space of the classical system under study and $\mathcal{H}$ its Hamiltonian. Notice that the knowledge of $Z(\beta)$ (resulting from an integration procedure which, for large systems of interacting particles, cannot be preformed neither analytically nor numerically) is not necessary for applying (65).

Usually (65) is written, for $i \neq j$, in the form:

$$\mathcal{P}_{i \to j} = \mathcal{Q}_{i \to j} \min\left(1, \frac{\pi_j}{\pi_i}\right) \tag{70}$$

$\mathcal{Q}_{i \to j}$ is a *trial move* that is *accepted or refused* depending on the outcome of a *Metropolis test* controlled by the term $\min\left(1, \frac{\pi_j}{\pi_i}\right)$.

We remind the reader that the hypothesis $\mathcal{Q}_{i \to j} = \mathcal{Q}_{j \to i}$ was framed in the proof of Metropolis' theorem. This hypothesis can be removed, provided that $\mathcal{Q}_{j \to i} > 0$ whenever $\mathcal{Q}_{i \to j} > 0$; in such situation, Metropolis' theorem still holds for the Markov chain:

$$\mathcal{P}_{i \to j} = \mathcal{Q}_{i \to j} \min\left(1, \frac{\pi_j \, \mathcal{Q}_{j \to i}}{\pi_i \, \mathcal{Q}_{i \to j}}\right) \quad i \neq j \tag{71}$$

where it is meant that $\mathcal{P}_{i \to j} = 0$ if $\mathcal{Q}_{i \to j} = 0$, whereas $\mathcal{P}_{i \to i}$ is defined as in (65).