Markov Chains (2023/24)

Giambattista Giacomin

5045 Bât. Sophie Germain (Campus PRG)
Email address: giacomin@lpsm.paris
URL: https://www.lpsm.paris/users/giacomin/index

Contents

CHAPTER 1

Markov chains: definitions, basic properties and examples

1. Basic concepts

Like *always* in probability, we work on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that is, at the same time, crucial and useless. At a certain point it will be very useful to make $(\Omega, \mathcal{F}, \mathbb{P})$ explicit, and we will make it explicit, but for most of the time it is just abstract nonsense.

To define a Markov Chain (MC) we need to provide a *state space* E and a probability kernel $\mathtt{p}$:

- the state space E is just a set, but it comes with its own σ-algebra $\mathcal{E}$ that tells us which subsets of E are measurable: so $(E, \mathcal{E})$ is a measurable space;
- $\mathtt{p}$ is an application from $E \times \mathcal{E}$ such that
  1. $\mathtt{p}(x, \cdot)$ is a probability on $(E, \mathcal{E})$ for every $x \in E$;
  2. $\mathtt{p}(\cdot, A) : E \to \mathbb{R}$ is a measurable function for every $A \in \mathcal{E}$ (the measurable subsets of $\mathbb{R}$ are the Borel subsets of $\mathbb{R}$).

A Markov Chain with state space E and probability kernel $\mathtt{p}$ (in short: $(E, \mathtt{p})$-MC, when E is obvious we just write $\mathtt{p}$-MC, sometimes we omit $\mathtt{p}$ too) is a sequence $(X_n)_{n=0,1,\dots}$ of random variables taking values in E with the property that for every n and every $A \in \mathcal{E}$

$$\mathbb{P}\left(X_{n+1} \in A \,\middle|\, \mathcal{F}_n\right) = \mathtt{p}\left(X_n, A\right) , \tag{1.1}$$

where $\mathcal{F}_n = \sigma\left(X_0, X_1, \dots, X_n\right)$ is the σ-algebra generated by $X_0, X_1, \dots, X_n$. The notation $\mathcal{F}_n \prec \mathcal{F}$, where $\prec$ means simply that $\mathcal{F}_n \subset \mathcal{F}$, but it reminds us that both $\mathcal{F}_n$ and $\mathcal{F}$ are σ-algebras.

REMARK 1.1. *Since, by the Tower Property of conditional expectation,* (1.1) *directly yields*

$$\mathtt{p}\left(X_n, A\right) = \mathbb{P}\left(X_{n+1} \in A \,\middle|\, X_n\right) , \tag{1.2}$$

one may be tempted to think that (1.1) *is equivalent to*

$$\mathbb{P}\left(X_{n+1} \in A \,\middle|\, \mathcal{F}_n\right) = \mathbb{P}\left(X_{n+1} \in A \,\middle|\, X_n\right) , \tag{1.3}$$

but there is an hidden information in (1.2)*: it is saying that the right-hand side (of* (1.2)*) depends only on X_n and A. Notably, there is no direct dependence on n itself! We call just MC a process that satisfies* (1.3) *for every n: it is a more general process because it may be time inhomogeneous (see Exercise 2.3 for examples). We will consider such a case only occasionally and the fact that we sometimes write MC for a* $\mathtt{p}$*-MC should not lead to confusion.*

REMARK 1.2. *It may sometimes be practical to consider MC with respect to more general filtrations* $(\mathcal{G}_n)$. *Since we want that* (X_n) *is adapted (i.e.,* X_n *is* $\mathcal{F}_n$*-measurable for every* n*), we require that* $\mathcal{G}_n \succ \mathcal{F}_n$ *for every* n. *Then we say that* (X_n) *is a* $(\mathtt{p}, (\mathcal{G})_n)$*-MC if* (1.1) *holds for every* n *with* $\mathcal{F}_n$ *replaced by* $\mathcal{G}_n$. *It is straightforward to check that a* $(\mathtt{p}, (\mathcal{G})_n)$*-MC is a* $\mathtt{p}$*-MC. Analogous generalization holds for (inhomogeneous) MC.*

We stress also that the equality in (1.1) is meant only *almost surely* because a priori the left-hand side is defined only almost surely.

More importantly, note that if we set $\mu(A) := \mathbb{P}(X_0 \in A)$, then μ is a probability on $(E, \mathcal{E})$ and by the Tower Property of conditional expectation

$$\begin{aligned}\mathbb{P}(X_0 \in A_0, X_1 \in A_1) &= \mathbb{E}\left[\mathbb{P}\left(X_1 \in A_1 \middle| \mathcal{F}_0\right) \mathbf{1}_{\{X_0 \in A_0\}}\right] \\ &= \int_{A_0} \left(\int_{A_1} \mathtt{p}(x_0, \mathrm{d}x_1) \right) \mu(\mathrm{d}x_0) = \int_{A_0} \int_{A_1} \mu(\mathrm{d}x_0) \mathtt{p}(x_0, \mathrm{d}x_1),\end{aligned} \tag{1.4}$$

for every A_0 and $A_1 \in \mathcal{E}$. Of course this generalizes to

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n) = \int_{A_0} \int_{A_1} \cdots \int_{A_n} \mu(\mathrm{d}x_0) \mathtt{p}(x_0, \mathrm{d}x_1) \mathtt{p}(x_1, \mathrm{d}x_2) \ldots \mathtt{p}(x_{n-1}, \mathrm{d}x_n), \tag{1.5}$$

for every $A_0, \ldots, A_n \in \mathcal{E}$.

PROPOSITION 1.3. X *is a* $\mathtt{p}$*-MC with* $X_0 \sim \mu$ *(i.e., the law of* X_0 *is* μ*) if and only if* (1.5) *holds for every* $n = 0, 1, \ldots$ *and for every* $A_0, \ldots, A_n \in \mathcal{E}$.

PROOF. Useful exercise. □

Before giving examples of MC's let us give the following result, which is central for us.

PROPOSITION 1.4. $(\xi_j)_{j=1,2,\ldots}$ *is an IID sequence of random variables that take values in a measurable space* $(E', \mathcal{E}')$ *and* $h : E \times E' \to E$ *is measurable. If* X_0 *is independent of* (ξ_j) *and if we set recursively* $X_{n+1} = h(X_n, \xi_{n+1})$, $n = 0, 1, \ldots$, *we have that* (X_n) *is a* $(E, \mathtt{p})$*-MC with*

$$\mathtt{p}(x, A) := \mathbb{P}\left(h(x, \xi_1) \in A\right), \tag{1.6}$$

for every $x \in E$ *and every* $A \in \mathcal{E}$.

We take this occasion to point out that when dealing with product spaces we use the product σ-algebra which is the σ-algebra that contains the product topology: in particular, the σ-algebra that equips $E \times E'$ is the smallest σ-algebra that contains the sets $A \times A'$ with $A \in \mathcal{E}$ and $A' \in \mathcal{E}'$.

Essentially without loss of generality we can choose $E' = \mathbb{R}$ or $E' = (0, 1)$, but sometimes if is practical to deal with more general spaces (and we will see it with the first examples). Moreover if we introduce the notation $h_\xi(x) = h(x, \xi)$ we have the convenient notation

$$X_n = h_{\xi_n} \circ h_{\xi_{n-1}} \circ \ldots \circ h_{\xi_1}(X_0) , \tag{1.7}$$

so that X_n is just the result of applying n random functions to the initial condition X_0. And (1.7) is one of the most efficient ways to simulate a Markov chain.

Proof of Proposition 1.4. From (1.7) we see that X_n is measurable with respect to $\sigma(X_0, \xi_1, \ldots, \xi_n)$. This implies both that $\mathcal{F}_n \prec \sigma(X_0, \xi_1, \ldots, \xi_n)$ and that X_n and ξ_{n+1} are independent. And of course ξ_{n+1} is independent of $\sigma(X_0, \xi_1, \ldots, \xi_n)$. Therefore

$$\mathbb{E}\left[\mathbf{1}_{\{h(X_n,\xi_{n+1})\in A\}} \,\middle|\, X_0, \xi_1, \ldots, \xi_n\right] = \mathtt{p}(X_n, A) , \tag{1.8}$$

where $\mathtt{p}(x, A)$ is defined in (1.6). By the Tower Property we conclude the proof:

$$\begin{aligned} \mathbb{P}\left(X_{n+1} \in A \,\middle|\, \mathcal{F}_n\right) &= \mathbb{E}\left[\mathbf{1}_{\{h(X_n,\xi_{n+1})\in A\}} \,\middle|\, \mathcal{F}_n\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{h(X_n,\xi_{n+1})\in A\}} \,\middle|\, X_0, \xi_1, \ldots, \xi_n\right] \,\middle|\, \mathcal{F}_n\right] \\ &= \mathtt{p}(X_n, A) . \end{aligned} \tag{1.9}$$

$\square$

Exercise 1.5. *The process that we have defined iteratively by means of a function h and the IID sequence (ξ_n) is called* Random Dynamical System*. So, Proposition 1.4 tells us that a Random Dynamical System is a $\mathtt{p}$-MC. The converse is true in great generality. Namely: under very mild assumptions (on the state space E), given a $\mathtt{p}$-MC (X_n) there exists a Random Dynamical System that coincides in law with (X_n). The exercise consists in proving such a statement for $E = \mathbb{R}$.*

In order to give the first examples of Markov chains let us consider for the moment just the class of random walks: this is a very limited context, but it contains already a lot of examples.

(1) If $E = E' = \mathbb{R}^d$ and $h(x, y) = x + y$ then the arising MC is just a random walk on $\mathbb{R}^d$: for $n = 1, 2, \ldots$

$$X_n = X_0 + \sum_{j=1}^{n} \xi_j , \tag{1.10}$$

and $\mathtt{p}(x, \cdot)$ coincides with the law of $x + \xi_1$.

(2) If $E = E' = \mathbb{Z}^d$ and $h(x, y) = x + y$ then the arising MC is just a random walk on $\mathbb{Z}^d$: note that (1.10) still holds

(3) If $E = E' = \mathbb{Z}$, $d = 1$ and $\mathbb{P}(\xi_1 = +1) = 1 - \mathbb{P}(\xi_1 = -1) = p$ then X_n is just a one dimensional simple random walk (simple refers that it jumps just to nearest neighbors). If $p = 1/2$ we speak of *simple symmetric random walk*.

(4) Random walks are naturally defined for example on a graph (N, L) where N is a (finite or countably infinite) set and L is a subset of N^2. N is the set of *nodes* (or *sites*) and L is the sent of *links*. We say that the graph is symmetric if $(x, y) \in \mathrm{L}$

implies $(y,x) \in \mathtt{L}$. If $n_x := |\{y : (x,y) \in \mathtt{L}\}| < \infty$ for every $x \in \mathtt{N}$, so we can write the set $\{y : (x,y) \in \mathtt{L}\}$ as $\{y_{x,1}, y_{x,2}, \ldots, y_{x,n_x}\}$, we define for $u \in (0,1)$

$$h(x,u) = \sum_{j=1}^{n_x} y_{x,j} \mathbf{1}_{((j-1)/n_x, j/n_x]}(u)\,. \tag{1.11}$$

This way if (ξ_j) is an IID sequence of variables that are uniformly distributed over $(0,1)$ (notation: $\mathcal{U}(0,1)$), then, given $X_0 \in \mathtt{N}$, $X_{n+1} = h(X_n, U_{j+1})$ defines the *simple random walk* on the graph $(\mathtt{N}, \mathtt{L})$, which is a MC with state space $E = \mathtt{N}$. In particular, if $\mathtt{N} = \mathbb{Z}$ and $\mathtt{L} = \{(x,y) \in \mathbb{Z}^2 : |x-y| = 1\}$, then $(\mathtt{N}, \mathtt{L})$ is a symmetric graph and the MC we have just defined is the simple symmetric random on $\mathbb{Z}$.

When E is finite or countably infinite the σ-algebra that we use is simply the set of all subsets of E. In this case we set

$$Q(x,y) := \mathtt{p}(x, \{y\})\,, \tag{1.12}$$

and Q is a *stochastic matrix* in the sense that

(1) $Q(x,y) \geq 0$ for every $x, y \in E$;
(2) $\sum_{y \in E} Q(x,y) = 1$ for every $x \in E$.

Of course, knowing Q is equivalent to knowing $\mathtt{p}$.

As an example, the stochastic matrix associated to the random walk on a graph is $Q(x,y) = 1/n_x$ for every y such that $(x,y) \in \mathtt{L}$, and $Q(x,y) = 0$ otherwise.

2. The Markov property

It is practical (for certain proofs) to introduce a *canonical space* in which to represent a p-MC with state space E. The canonical space is simply $\Omega := E^{\{0,1,\ldots\}}$, equipped with the product topology and the corresponding σ-algebra, that we denote (as usual) by $\mathcal{F}$. $\mathcal{F}$ can be characterized as the smallest σ-algebra that contains the *cylindric events*, that is the events of the form $A_1 \times A_2 \times \ldots$ with $A_j \in \mathcal{E}$ for every j and for which there exists j_0 such that $A_j = E$ for every $j \geq j_0$. Note that the class of cylinder events is stable under intersection. So the class of cylinder events forms what in measure theory is a π-system and if two probabilities coincide on a π-system, then they coincide on the whole σ-algebra (i.e., they are the same probability). A concise and approachable treatment of these issue and, in general, to the measure theory we need may be found in [**8**, Ch. 1].

If μ is a probability on $(E, \mathcal{E})$ and if $\mathtt{p}$ is a probability kernel, then we can define a probability on the canonical space by stipulating that

$$\mathbb{P}_\mu\left(A_0 \times A_1 \times \ldots \times A_k \times E \times E \times \ldots\right) = \int_{A_0 \times A_1 \times \ldots \times A_{k-1}} \mu(\,\mathrm{d}x_0)\mathtt{p}(x_0, \,\mathrm{d}x_1) \ldots \mathtt{p}(x_{k-1}, A_k)\,. \tag{1.13}$$

Standard (non trivial, but intuitive) extension theorems from measure theory guarantee that there exists a $\mathbb{P}_\mu$ (on the whole product σ-algebra for which (1.13) holds for every cylinder set. And such a probability is unique because the cylinder sets form a π-system. We call this probability the *canonical probability* for the MC with initial condition μ and transition kernel $\mathtt{p}$.

When $\mu = \delta_x$ we write $\mathbb{P}_x$ for $\mathbb{P}_{\delta_x}$. Moreover, when the initial datum is evident or not important to be underlined, we simply write $\mathbb{P}$ for $\mathbb{P}_\mu$. Of course $\mathbb{E}_\mu$ and $\mathbb{E}_x$ are the corresponding expectations.

Consider now the *canonical projections* $X_j(\omega) := \omega_j$, $j = 0, 1, \ldots$. Note that X_j is a random variable on the canonical probability space $(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ and that it follows directly from the definitions that $(X_j)_{j=0,1,\ldots}$ is a p-MC with $X_0 \sim \mu$.

We point out however that we use $\mathbb{P}_\mu$, $\mathbb{P}_\mu$, ext… also when the probability space is not canonical and $\mathbb{P}_\mu$ is the law of the $\mathtt{p}$-MC (X_n) with $X_0 \sim \mu$.

Let us also introduce the translation operator on the canonical space: for $n = 0, 1, \ldots$

$$(\theta_n \omega)_j \,=\, \omega_{n+j}\,, \tag{1.14}$$

for $j = 0, 1, \ldots$. For A an event in the canonical space we have, with standard notation for the pre-image of a set under the action of a function, $\theta_k^{-1} A = \{\omega : \theta_k \omega \in A\}$ (and one directly checks that θ_k is measurable, i.e. $\theta_k^{-1} A \in \mathcal{F}$: in fact, $\theta_k^{-1} A$ is cylindrical if A is).

PROPOSITION 2.1 (Markov Property). *We work on a generic probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (not necessarily the canonical one). If (X_n) is a Markov chain and A is an event in the canonical space, then for every $k = 0, 1, \ldots$ we have that $\mathbb{P}(\,\mathrm{d}\omega)$-a.s.*

$$\mathbb{P}\left((X_n)_{n=0,1,\ldots} \in \theta_k^{-1} A \,\middle|\, \mathcal{F}_k\right)(\omega) \,=\, \mathbb{P}_{X_k(\omega)}(A)\,, \tag{1.15}$$

and we recall that $\mathcal{F}_k = \sigma(X_0, X_1, \ldots X_k)$.

Possibly more intuitively, (1.15) can be rewritten as

$$\mathbb{P}\left((X_{n+k})_{n=0,1,\ldots} \in A \,\middle|\, \mathcal{F}_k\right)(\omega) \,=\, \mathbb{P}_{X_k(\omega)}(A)\,. \tag{1.16}$$

PROOF. By definition of conditional expectation and by standard results of measure theory it suffices to check that

$$\begin{aligned}\mathbb{P}\left(X_j \in B_j \text{ for } j = 0, \ldots, k \text{ and } X_j \in A_j \text{ for } j = k, \ldots, j+m\right) \,=&\\ \mathbb{E}\left[\mathbf{1}_{\{X_j \in B_j \text{ for } j=0,\ldots,k\}} \mathbb{P}_{X_k(\omega)}\left(X_j \in A_j \text{ for } j = 0, \ldots, m\right)\right]&\,,\end{aligned} \tag{1.17}$$

for every $m = 0, 1, \ldots$ and for every choice of events $B_0, \ldots, B_k$ and $A_0, \ldots, A_m$ in $\mathcal{E}$. But, by (1.5), both the left- and right-hand side of (1.17) can be written explicitly as

$$\int_{B_0} \cdots \int_{B_{k-1}} \int_{B_k \cap A_0} \int_{A_1} \cdots \int_{A_m} \mu(\,\mathrm{d}x_0)\mathtt{p}(x_0,\,\mathrm{d}x_1) \ldots \mathtt{p}(x_{k+m-1},\,\mathrm{d}x_{k+m})\,, \tag{1.18}$$

with μ the law of X_0. This is straightforward for the left-hand side. For the right-hand side one should just remark that (1.5) implies via a standard approximation procedure that

$$\mathbb{E}\left[\mathbf{1}_{\{X_0\in A_0, X_1\in A_1,\ldots,X_{k-1}\in A_{k-1}\}} f(X_k)\right] =$$
$$\int_{A_0}\int_{A_1}\cdots\int_{A_{k-1}}\int_E \mu(\,\mathrm{d}x_0)\mathtt{p}(x_0,\,\mathrm{d}x_1)\mathtt{p}(x_1,\,\mathrm{d}x_2)\ldots\mathtt{p}(x_{k-1},\,\mathrm{d}x_k)f(x_k)\,, \quad (1.19)$$

for every bounded measurable function $f : E \to \mathbb{R}$. □

Remark 2.2. *An equivalent way to formulate the Markov property, i.e. Proposition 2.1, is for example to say that for every bounded bounded measurable function* $h : E^{\{0,1,\ldots\}} \to \mathbb{R}$

$$\mathbb{E}\left[h(X_k, X_{k+1},\ldots)\,\middle|\,\mathcal{F}_k\right](\omega) \;=\; \mathbb{E}_{X_k(\omega)}\left[h(X_0, X_1,\ldots)\right]. \quad (1.20)$$

Moreover, if we work on the canonical space and the variables X_j *are the canonical projections, we can state the Markov property by saying that for every positive (or bounded) random variable* Y *which is measurable with respect to* $\sigma(X_0, X_1,\ldots)$

$$\mathbb{E}\left[Y\circ\theta_k\,\middle|\,\mathcal{F}_k\right](\omega) \;=\; \mathbb{E}_{X_k(\omega)}\left[Y\right], \quad (1.21)$$

$\mathbb{P}(\,\mathrm{d}\omega)$*-a.s..*

We exploit the notations that we have in our hands to remark that $\mathtt{p}(x,B) = \mathbb{P}_x(X_1\in B)$ and also that if we introduce the important notation

$$\mathtt{p}_k\,(x,B) \;:=\; \mathbb{P}_x\,(X_k\in B)\;. \quad (1.22)$$

Note that $\mathtt{p}_k$ is a probability kernel. Moreover

$$\mathtt{p}_\star \;:=\; \sum_{k=1}^\infty 2^{-k}\mathtt{p}_k\,, \quad (1.23)$$

is also a probability kernel (that will come very handy later on).

Exercise 2.3. *Show that if* (X_n) *is a* $\mathtt{p}$*-MC, then for every* $k\in\{1,2,\ldots\}$ *the process* $(X_{nk})_{n=0,1,\ldots}$ *is a* $\mathtt{p}_k$*-MC. Show moreover that given any sequence of integer numbers* $n_0, n_1,\ldots$*, with* $n_0\geq 0$ *and* $n_{j+1}\geq n_j$ *for every* j*, then* (X_{n_j}) *is a MC: in general, this MC is inhomegenousneous. This way we have constructed a large family of inhomogeneous Markov chains.*

3. The Strong Markov property

The notion of stopping time is linked to the filtration that we have chosen for our space: in our case $(\mathcal{F}_n)$ is the natural filtration of the MC we are considering, but it could be a larger filtration. A stopping time τ (with respect to the filtration $(\mathcal{F}_n)$) is a random variable that takes values in $\{0,1,\ldots,\infty\}$ with the property that

$$\{\tau\leq k\}\in\mathcal{F}_k \;\text{ for every } k=0,1,\ldots\,. \quad (1.24)$$

Note that, with $\mathcal{F}_\infty$ the smallest σ-algebra that contains $\cup_n\mathcal{F}_n$, we have that $\{\tau = \infty\}\in\mathcal{F}_\infty$. Moreover one directly checks that (1.24) is equivalent to asking that

$\{\tau = k\} \in \mathcal{F}_k$ for every $k = 0, 1, \ldots$ For τ a stopping time, we introduce also the σ-algebra $\mathcal{F}_\tau$ that *contains what happened up to time* τ. The (mathematical) definition is:

$$\mathcal{F}_\tau \;:=\; \{A \in \mathcal{F} :\; A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for } n = 0, 1, 2, \ldots\} \;. \tag{1.25}$$

The easy exercise of verifying hat $\mathcal{F}_\tau$ is a σ-algebra is highly advised. In the direction of justifying the informal interpretation we have given of $\mathcal{F}_\tau$ we remark that if $\tau = n$ (that is, if τ is a constant, hence trivially a stopping time), then $\mathcal{F}_\tau = \mathcal{F}_n$. Moreover (one more useful exercise!) if τ' is another stopping time and if $\tau' \geq \tau$, then $\mathcal{F}_\tau \prec \mathcal{F}_{\tau'}$. Finally, if (Y_n) is adapted to $(\mathcal{F}_n)$ and if $\tau < \infty$ a.s. we can introduce $Y_\tau(\omega) := Y_{\tau(\omega)}(\omega)$ for every $\omega \in \{\tau < \infty\}$. If $\tau(\omega) = \infty$ and the Y random variables are real we can (for example) set $Y_\tau(\omega) := 0$. With this choice we have $Y_\tau = \sum_{n=0,1,\ldots} Y_n \mathbf{1}_{\{n\}}(\tau)$, where the same over an empty set gives 0. Since for every Borel set B we have $\{Y_\tau \in B\} \cap \{\tau = n\} = \{Y_n \in B\} \cap \{\tau = n\} \in \mathcal{F}_n$ we see that $\{Y_\tau \in B\} \in \mathcal{F}_\tau$. Therefore Y_τ is $\mathcal{F}_\tau$ measurable and this is possibly the strongest argument to say that $\mathcal{F}_\tau$ *contains the what happened up to time* τ.

Important examples of stopping times include the time of first entry of an adapted process (Y_n) in a mesurable set, keeping in mind that there are two slightly different natural versions of such a time: the (first) hitting time

$$T_A^\flat(\omega) \;=\; T_A^{Y,\flat}(\omega) \;:=\; \inf\{n = 0, 1, \ldots :\; Y_n(\omega) \in A\} \;, \tag{1.26}$$

and the (first) return time

$$T_A(\omega) \;=\; T_A^{Y}(\omega) \;:=\; \inf\{n = 1, 2, \ldots :\; Y_n(\omega) \in A\} \;, \tag{1.27}$$

where we adopt the convention that the infimum of the empty set is ∞: of course $T_A^{Y,\flat} \leq T_A^Y$.

Also the successive entries do A are stopping times: successive entries to a measurable set is an important sequence of stopping time for what we are going to do and will be treated when it will come up.

THEOREM 3.1 (Strong Markov Property). *X is a $\mathtt{p}$-MC on the canonical space. For every positive (or bounded) random variable Y, for every choice of distribution μ of X_0 and for every stopping time τ we have*

$$\mathbb{E}_\mu\left[Y \circ \theta_\tau \mathbf{1}_{\{\tau<\infty\}} \middle| \mathcal{F}_\tau\right] \;=\; \mathbb{E}_{X_\tau}\left[Y\right] \mathbf{1}_{\{\tau<\infty\}} \,, \tag{1.28}$$

$\mathbb{P}_\mu$-a.s..

PROOF. It is clear (is it? Do verify it as an exercise) that the right-hand side in (1.28) is $\mathcal{F}_\tau$- measurable. Therefore, by definition of conditional expectation, since Y is either positive or bounded it suffices to check that for every $n = 0, 1, \ldots$ and every $A \in \mathcal{F}_n$ we have

$$\mathbb{E}_\mu\left[Y \circ \theta_\tau \mathbf{1}_{\{\tau<\infty\}} \mathbf{1}_A\right] \;=\; \mathbb{E}_\mu\left[\mathbb{E}_{X_\tau}\left[Y\right] \mathbf{1}_{\{\tau<\infty\}} \mathbf{1}_A\right] \;. \tag{1.29}$$

Now the point is simply to remark that $\{\tau < \infty\} = \sqcup_{n=0,1,\ldots}\{\tau = n\}$ so that σ-additivity tells us that (1.29) is equivalent to verifying that

$$\mathbb{E}_\mu\left[Y \circ \theta_\tau \mathbf{1}_{\{\tau=n\}}\mathbf{1}_A\right] = \mathbb{E}_\mu\left[\mathbb{E}_{X_\tau}\left[Y\right]\mathbf{1}_{\{\tau=n\}}\mathbf{1}_A\right] . \tag{1.30}$$

for every n. In turn (1.30) is equivalent to

$$\mathbb{E}_\mu\left[Y \circ \theta_n \mathbf{1}_{\{\tau=n\}\cap A}\right] = \mathbb{E}_\mu\left[\mathbb{E}_{X_n}\left[Y\right]\mathbf{1}_{\{\tau=n\}\cap A}\right] . \tag{1.31}$$

Since $\{\tau = n\} \cap A \in \mathcal{F}_n$ we can rewrite (1.31) as

$$\mathbb{E}_\mu\left[\mathbb{E}\left[Y \circ \theta_n \middle| \mathcal{F}_n\right]\mathbf{1}_{\{\tau=n\}\cap A}\right] = \mathbb{E}_\mu\left[\mathbb{E}_{X_n}\left[Y\right]\mathbf{1}_{\{\tau=n\}\cap A}\right] , \tag{1.32}$$

but this equality holds as a direct consequence of the (simple) Markov property (1.21). $\square$

4. Useful tool: martingales and harmonic functions

A measurable function f that is bounded below and satisfies $pf \leq f$ is called superharmonic, or $\mathtt{p}$-superhamonic if we need to be more explicit. If we have only $pf(x) \leq f(x)$ for $x \in A \in \mathcal{E}$ we say that f is superharmonic on A. Moreover we say that a function is harmonic (on A) if $pf = f$ (on A).

Here is a first result that links superharmonic functions to supermartingales.

PROPOSITION 4.1. *X is a MC with probability kernel $\mathtt{p}$ and f is superharmonic and bounded. Then $\left(f\left(X_n\right)\right)_{n=0,1,\ldots}$ is a supermartingale.*

We leave the proof to the reader and prove the following more advanced result:

PROPOSITION 4.2. *X is a MC with probability kernel $\mathtt{p}$ and assume that $X_0 = x$ is not random. Then $f \geq 0$ is superharmonic on A if and only if $\left(f\left(X_{n\wedge T^\flat_{A^\complement}}\right)\right)_{n=0,1,\ldots}$ is a non negative supermartingale for every $x \in E$.*

PROOF. We set $Y_n := f\left(X_{n\wedge T^\flat_{A^\complement}}\right)$.

If f is superharmonic on A we have

$$\begin{aligned}
\mathbb{E}\left[Y_{n+1}\middle|\mathcal{F}_n\right] &= \mathbb{E}\left[Y_{n+1}\left(\mathbf{1}_{\left\{T^\flat_{A^\complement}\leq n\right\}} + \mathbf{1}_{\left\{T^\flat_{A^\complement}> n\right\}}\right)\middle|\mathcal{F}_n\right] \\
&= f\left(X_{T^\flat_{A^\complement}}\right)\mathbf{1}_{\left\{T^\flat_{A^\complement}\leq n\right\}} + \mathbb{E}\left[f(X_{n+1})|\mathcal{F}_n\right]\mathbf{1}_{\left\{T^\flat_{A^\complement}> n\right\}} \\
&= f\left(X_{T^\flat_{A^\complement}}\right)\mathbf{1}_{\left\{T^\flat_{A^\complement}\leq n\right\}} + \mathtt{p}f(X_n)\mathbf{1}_{\left\{T^\flat_{A^\complement}> n\right\}} \\
&\leq f\left(X_{T^\flat_{A^\complement}}\right)\mathbf{1}_{\left\{T^\flat_{A^\complement}\leq n\right\}} + f(X_n)\mathbf{1}_{\left\{T^\flat_{A^\complement}> n\right\}} = f\left(X_{n\wedge T^\flat_{A^\complement}}\right) = Y_n .
\end{aligned} \tag{1.33}$$

Since $f \geq 0$ the bound we just established implies that $\|Y_n\|_1 \leq |f(x)| < \infty$ for every n. We have therefore established that Y is a supermartingale, for every deterministic initial condition.

On the other hand, let us assume that Y is non negative and it is a supermartingale for every choice of $x_0 = x$. So $Y_0 = f(X_0) \geq 0$ yields $f(x) \geq 0$ for every x. Moreover for $X_0 = x \in A$, which implies $T^\flat_{A^\complement} \geq 1$, the supermartingale property directly yields

$$f(x) \;\geq\; \mathbb{E}\left[f\left(X_{1\wedge T^\flat_{A^\complement}}\right)\right] \;=\; \mathbb{E}\left[f\left(X_1\right)\right] \;=\; \mathtt{p}f(x)\,, \tag{1.34}$$

and, of course, $f(x) = f\left(X_{n\wedge T^\flat_{A^\complement}}\right)$ if $x \notin A$. So the proof is complete. □

Proposition 4.3. *Choose a probability kernel* $\mathtt{p}$. *For every* $A \in \mathcal{E}$

(1) $x \mapsto \mathbb{P}_x\left(T^\flat_A < \infty\right)$ *is superharmonic and it is harmonic in* $A^\complement$;

(2) $x \mapsto \mathbb{P}_x\left(T_A < \infty\right)$ *is superharmonic.*

Proof. For (1) we set $f(x) = \mathbb{P}_x\left(T^\flat_A < \infty\right)$ and note that $pf(x) = \mathbb{E}_x[f(X_1)] = \mathbb{E}_x[\mathbb{P}_{X_1}(T^\flat_A < \infty)]$. But the Markov property tells us that $\mathbb{P}(T_A < \infty|\mathcal{F}_1) = \mathbb{P}_{X_1}(T^\flat_A < \infty)$, because $T_A = 1 + T^\flat_A \circ \theta$, so $\mathtt{p}f(x) = \mathbb{P}_x(T_A < \infty)$. Since $T^\flat_A \leq T_A$ we have that $\mathtt{p}f(x) \leq \mathbb{P}_x(T^\flat_A < \infty)$. Moreover if $x \in A^\complement$ then $T^\flat_A = T_A$ and the proof is complete.

For (2) we set $g(x) = \mathbb{P}_x(T_A < \infty)$. Like before $\mathtt{p}g(x) = \mathbb{E}_x[\mathbb{P}_{X_1}(T_A < \infty)]$, but this time the Markov property yields $\mathbb{P}_{X_1}(T_A < \infty) = \mathbb{P}(T_A \circ \theta < \infty|\mathcal{F}_1)$, so $\mathtt{p}g(x) = \mathbb{P}_x(T_A \circ \theta < \infty)$ and $\{T_A \circ \theta < \infty\} \subset \{T_A < \infty\}$, which completes the proof. □

5. The potential kernel

Given a $\mathtt{p}$-MC we introduce the total number of visits N_A of the chain to a set $A \in \mathcal{E}$:

$$N_A \;:=\; \sum_{k=0}^{\infty} \mathbf{1}_A(X_k)\,. \tag{1.35}$$

We introduce also the potential kernel $U : E \times \mathcal{E} \to \{0, 1, \ldots\} \cup \{\infty\}$ by setting

$$U(x, A) \;:=\; \mathbb{E}_x\left[N_A\right] \;=\; \sum_{k=0}^{\infty} \mathtt{p}_k(x, A)\,. \tag{1.36}$$

For every x, $U(x, \cdot)$ is a (positive) measure, because it is a countable sum of measures: we will soon see that $U(x, \cdot)$ may be a finite measure, but it may be that $U(x, A) = \infty$ for every $A \neq \emptyset$ without this being a pathological case.

In case $A = \{y\}$ we will use the short-cut notation $U(x, y)$ for $U(x, \{y\})$. In the same way we write T_x for $T_{\{x\}}$ and $T^\flat_x$ for $T^\flat_{\{x\}}$.

PROPOSITION 5.1. *For every $x \in E$ and $A \in \mathcal{E}$ we have*

$$U(x,A) \,\le\, \mathbb{P}_x\left(T_A < \infty\right) \sup_{y \in A} U(y,A)\,. \tag{1.37}$$

For $A = \{y\}$ we have

$$U(x,y) \,\le\, \mathbb{P}_x\left(T_y < \infty\right) U(y,y)\,. \tag{1.38}$$

PROOF. This uses the Strong Markov Property, by noting first that $N_A = \sum_{n=T_A^\flat}^\infty \mathbf{1}_A(X_n)$ if $T_A^\flat < \infty$ and $N_A = 0$ otherwise. We have

$$\begin{aligned} U(x,A) \,&=\, \mathbb{E}_x\left[\sum_{n=T_A^\flat}^\infty \mathbf{1}_A(X_n)\mathbf{1}_{\{T_A^\flat<\infty\}}\right] \,=\, \sum_{n=0}^\infty \mathbb{E}_x\left[\mathbf{1}_A(X_n\circ\theta_{T_A^\flat})\mathbf{1}_{\{T_A^\flat<\infty\}}\right] \\ &= \sum_{n=0}^\infty \mathbb{E}_x\left[\mathbf{1}_{\{T_A^\flat<\infty\}}\mathbb{E}_{X_{T_A^\flat}}\left[\mathbf{1}_A(X_n)\right]\right] \,\le\, \mathbb{P}_x\left(T_A<\infty\right)\sup_{y\in A} U(y,A)\,. \end{aligned} \tag{1.39}$$

□

For the next result we introduce the successive visits to A: we set $T_A^{(1)} := T_A$ and, by recurrence, we set

$$T_A^{(n+1)} \,:=\, \inf\left\{k > T_A^{(n)} \,:\, X_k \in A\right\}\,, \tag{1.40}$$

and by this we mean that $T_A^{(n+1)}(\omega) = \infty$ if $T_A^{(n)}(\omega) = \infty$. $T_A^{(n)}$ is a stopping time. Moreover one readily checks that on the event $\{T_A^{(n)} < \infty\}$ we have

$$T_A^{(n+1)} \,=\, T_A^{(n)} + T_A \circ \theta_{T_A^{(n)}}\,. \tag{1.41}$$

We also use the concept of stochastic domination: given two real valued variables X and Y, X (stochastically) dominates Y if $\mathbb{P}(X > x) \ge \mathbb{P}(Y > x)$ for every $x \in \mathbb{R}$.

PROPOSITION 5.2. $\mathtt{p}$ *is a probability kernel and $A \in \mathcal{E}$.*

(1) Assume there exists $\delta \in [0,1)$ such that $\sup_{x\in A}\mathbb{P}_x\left(T_A < \infty\right) \le \delta$. Then $\sup_{x\in A}\mathbb{P}_x\left(T_A^{(k)} < \infty\right) \le \delta^k$ for every $k = 1, 2, \ldots$ and, for every $X_0 \in A$, N_A is stochastically dominated by a geometric random variable of parameter $1-p$. In particular $\sup_{x\in A} U(x,A) \le 1/(1-\delta)$.

(2) If instead $\mathbb{P}_x\left(T_A < \infty\right) = 1$ for every $x \in A$, then $\mathbb{P}_x\left(T_A^{(k)} < \infty\right) = 1$ for every $x \in A$ and every k. In this case we also have $\mathbb{P}_x\left(N_A = \infty\right) = 1$ for every $x \in A$.

Proposition 5.2 becomes more elegant (and really a dichotomy!) if A contains just one element:

PROPOSITION 5.3. $\mathtt{p}$ *is a probability kernel. For every* $x \in E$*:*

(1) If $\mathbb{P}_x\left(T_x < \infty\right) =: \delta \in [0,1)$ *then* $\mathbb{P}_x\left(T_x^{(k)} < \infty\right) = \delta^k$ *for every* $k = 1, 2, \ldots$*, hence, with* $X_0 = x$*,* N_A *is a geometric random variable (time of first success) of parameter* $1-p$*. In particular* $U(x,x) = 1/(1-\delta) < \infty$*.*

(2) If instead $\mathbb{P}_x\left(T_x < \infty\right) = 1$ *then* $\mathbb{P}_x\left(T_A^{(k)} < \infty\right) = 1$ *for every* k *and* $\mathbb{P}_x\left(N_x = \infty\right) = 1$ *and* $U(x,x) = \infty$*.*

We give a proof of Proposition 5.2 and leave the details of Proposition 5.3 as an exercise.

PROOF OF PROPOSITION 5.2. We have

$$\mathbb{P}_x\left(T_A^{(k+1)} < \infty\right) = \mathbb{P}_x\left(T_A^{(k)} < \infty, T_A^{(k+1)} < \infty\right) = \mathbb{P}_x\left(T_A^{(k)} < \infty, T_A \circ \theta_{T_A^{(k)}} < \infty\right), \tag{1.42}$$

so by the Strong Markov Property

$$\mathbb{P}_x\left(T_A^{(k+1)} < \infty\right) = \mathbb{E}_x\left[\mathbf{1}_{\{T_A^{(k)} < \infty\}} \mathbb{P}_{X_{T_A^{(k)}}}\left(T_A < \infty\right)\right], \tag{1.43}$$

so in case (1) we readily obtain

$$\mathbb{P}_x\left(T_A^{(k+1)} < \infty\right) \le \delta\, \mathbb{P}_x\left(T_A^{(k)} < \infty\right), \tag{1.44}$$

which yields $\mathbb{P}_x\left(T_x^{(k)} < \infty\right) = \delta^k$ for every x. For what concerns N_A we remark that, for $X_0 = x \in A$, $N_A = 1 + \sum_{k=1}^\infty \mathbf{1}_{\{T_A^{(k)} < \infty\}}$. So $\mathbb{P}_x(N_A > 0) = 1$ and, for $k = 1, 2, \ldots$, $\mathbb{P}_x(N_A > k) = \mathbb{P}_x(T_A^{(k)} < \infty) \le \delta^k$ and this establishes the claimed stochastic domination. Since $U(x,A) = \int_0^\infty \mathbb{P}_x(N_A > x)\,\mathrm{d}x = 1 + \sum_{k=1}^\infty \mathbb{P}_x(T_A^{(k)} < \infty)$, the bound on $\sup_{x \in A} U(x,A)$ follows.

For what concerns (2) we go back to (1.43) and we readily see that in this case $\mathbb{P}_x\left(T_A^{(k)} < \infty\right) = 1$ for every k. Therefore $\mathbb{P}_x\left(\cap_k \left\{T_A^{(k)} < \infty\right\}\right) = 1$ too, therefore $\mathbb{P}_x(N_A = \infty) = 1$. □

We now introduce the important notation

$$\rho_{x,y} := \mathbb{P}_x(T_y < \infty). \tag{1.45}$$

So we have seen (Proposition 5.3) that $\rho_{x,x} = 1$ if and only if $\mathbb{P}_x(N_x = \infty) = 1$. And if $\rho_{x,x} < 1$ then the chain that starts from x visits a.s. x only a finite number of times.

It is therefore natural to call recurrent a state x with $\rho_{x,x} = 1$ and transient if $\rho_{x,x} < 1$ and we will use this language.

Note that $\rho(x,y) > 0$ is equivalent to the existence of a value of k such that $\mathtt{p}_k(x,\{y\}) > 0$ (and it is also equivalent to $\mathtt{p}_\star(x,\{y\}) > 0$, $\mathtt{p}_\star$ given in (1.23)). We will say that y is accessible from x if $\rho_{x,y} > 0$. We will also say that y is accessible if it accessible from every state $x \in E$.

The next result is of particular interest in the case in which E is countable.

PROPOSITION 5.4. *If* $\rho_{x,x} = 1$ *and* $\rho_{x,y} > 0$ *then* $\rho_{y,y} = \rho_{x,y} = \rho_{y,x} = 1$.

PROOF. By Proposition 5.3 we know that $\rho_{x,x} = 1$ is equivalent to $\mathbb{P}_x(N_x = \infty) = 1$ (and $U(x,x) = \infty$). So by the Strong Markov Property

$$0 = \mathbb{P}_x\left(N_x < \infty\right) \ge \mathbb{P}_x\left(T_y < \infty, T_x \circ \theta_{T_y} = \infty\right) = \mathbb{P}_x\left(T_y < \infty\right)\mathbb{P}_y\left(T_x = \infty\right), \tag{1.46}$$

that is $\rho_{x,y}\mathbb{P}_y\left(T_x = \infty\right) = 0$. But $\rho_{x,y} > 0$ by hypothesis, so $\mathbb{P}_y\left(T_x = \infty\right) = 0$, i.e. $\rho_{y,x} = \mathbb{P}_y\left(T_x < \infty\right) = 1$. Since both $\rho_{y,x} > 0$ and $\rho_{x,y} > 0$ we know that there exist $n_{x,y}$ and $n_{y,x}$ such that $\mathtt{p}_{n_{x,y}}(x, \{y\}) > 0$ and $\mathtt{p}_{n_{y,x}}(y, \{x\}) > 0$. So for every n

$$\mathtt{p}_{n_{y,x}+n+n_{x,y}}(y, \{y\}) \ge \mathtt{p}_{n_{y,x}}(y, \{x\})\mathtt{p}_n(x, \{x\})\mathtt{p}_{n_{x,y}}(y, \{x\}), \tag{1.47}$$

and therefore

$$\begin{aligned} U(y,y) &\ge \sum_{n=0}^{\infty} \mathtt{p}_{n_{y,x}}(y, \{x\})\mathtt{p}_n(x, \{x\})\mathtt{p}_{n_{x,y}}(y, \{x\}) \\ &= \mathtt{p}_{n_{y,x}}(y, \{x\})U(x,x)\mathtt{p}_{n_{x,y}}(y, \{x\}) = \infty. \end{aligned} \tag{1.48}$$

Therefore, by Proposition 5.3, $\rho_{y,y} = 1$. So we have proven that $\rho_{y,x} = \rho_{y,y} = 1$. By exchanging x and y this implies also that $\rho_{x,y} = 1$ and the proof is complete. □

REMARK 5.5. *In what follows the concept of* accessibility *is playing an important role, but up to now we just spoke of accessibility of a state. More generally, we say that* $A \in \mathcal{E}$ *is accessible from* x *if* $\mathbb{P}_x(T_A < \infty) > 0$. *This of course means that the exists* k *such that* $\mathtt{p}_k(x, A) > 0$. *Specializing to* $A = \{y\}$, *we recover that* y *is accessible from* x *if and only if* $\rho_{x,y} > 0$. *We stress that this does not mean that* y *is certainly visited starting from* x. *But if* x *is recurrent, i.e.* $\rho_{x,x} = 1$, *then Proposition 5.4 tells us that, if* y *is accessible from* x, y *is visited with probability one (and* y *is recurrent too). Note also that for many natural MC no state is accessible from any other point: consider for example a random walk* X, *on* $\mathbb{R}$, $X_{n+1} = X_n + \xi_{n+1}$, $\{\xi_j\}$ *IID* $\mathcal{N}(0,1)$ *random variables. Then* $\mathtt{p}(x, \mathrm{d}y) = g(y-x)\,\mathrm{d}y$, *with* g *the density of a variable* $\mathcal{N}(0,1)$. *Hence* $\rho_{x,y} = 0$ *for every* x, y. *On the other hand* $\mathtt{p}(x, A) > 0$ *for every* A *of positive Lebesgue measure. So every positive Lebesgue measure set is accessible from any state* x.

We now give a result that is going to be useful later on. It is rather intuitive, but the proof is not straightforward and provides a nice example of interplay of MC and martingales.

PROPOSITION 5.6. *Consider a* $\mathtt{p}$*-MC and* $A, B \in \mathcal{E}$ *such that* $\inf_{x\in A}\mathbb{P}_x(T_B < \infty) > 0$. *Then, for every probability* μ *on* $(E, \mathcal{E})$, *we have that* $\mathbb{P}_\mu(N_A = \infty, N_B < \infty) = 0$.

Put otherwise, Proposition 5.6 says that if $N_A(\omega) = \infty$, then $N_B(\omega) = \infty$ $\mathbb{P}_\mu(\mathrm{d}\omega)$-a.s..

Proof. Let us set $\delta := \inf_{x\in A} \mathbb{P}_x(T_B < \infty)$ and $\delta > 0$ by hypothesis. A direct coin séquence of the definition of δ and N_A is that $\{N_A = \infty\} \subset \{\sum_n \mathbb{P}_{X_n}(T_B < \infty) = \infty\}$. We claim that

$$\lim_{n\to\infty} \mathbb{P}_{X_n(\omega)}(T_B < \infty) \stackrel{\mathbb{P}_\mu(\mathrm{d}\omega)\text{-a.s.}}{=} \mathbf{1}_{\{N_B(\omega)=\infty\}}\,. \tag{1.49}$$

This suffices to conclude because if $\omega \in \{\sum_n \mathbb{P}_{X_n}(T_B < \infty) = \infty\}$ then we have that $\liminf_{n\to\infty} \mathbb{P}_{X_n(\omega)}(T_B < \infty) \geq \delta$, hence, by (1.49), $\lim_{n\to\infty} \mathbb{P}_{X_n(\omega)}(T_B < \infty) = 1$ and $N_B(\omega) -= \infty$ (both facts $\mathbb{P}_\mu(\mathrm{d}\omega)$-a.s.). We have therefor established that, if (1.49) holds, then $N_A(\omega) = \infty$ implies $N_B(\omega) = \infty$ $\mathbb{P}_\mu(\mathrm{d}\omega)$-a.s..

So the proof of Proposition 5.6 boils down to showing (1.49).

In order to establish this we use that $x \mapsto \mathbb{P}_x(T_B < \infty)$ is superharmonic (see Proposition 4.3(2)). It is of course also a bounded non negative function, so, by Proposition 4.1, $\{\mathbb{P}_{X_n}(T_B < \infty)\}_{n=0,1,\ldots}$ is a non negative supermartingale. Therefore [**8**, Ch. 11] $\lim_n \mathbb{P}_{X_n}(T_B < \infty)$ exists a.s. and in $\mathbb{L}^1$. Therefore for every non negative integer m and every event $F \in \mathcal{F}_n$ we have

$$\begin{aligned} \mathbb{E}_\mu\left[\mathbf{1}_F \lim_n \mathbb{P}_{X_n}(T_B < \infty)\right] &= \lim_n \mathbb{E}_\mu\left[\mathbf{1}_F \mathbb{P}_{X_n}(T_B < \infty)\right] \\ &= \lim_n \mathbb{E}_\mu\left[\mathbf{1}_F \mathbb{P}_\mu\left(T_B \circ \theta_n < \infty \middle| \mathcal{F}_n\right)\right] \\ &= \lim_n \mathbb{P}_\mu\left(F \cap \{T_B \circ \theta_n < \infty\}\right). \end{aligned} \tag{1.50}$$

Now remark that

$$\{T_B \circ \theta_n < \infty\} = \bigcup_{j>n} \{X_j \in B\} \stackrel{n\to\infty}{\searrow} \{N_B = \infty\}. \tag{1.51}$$

Therefore, by the Lebesgue Dominated Convergence Theorem, (1.50) and (1.51) yield

$$\mathbb{E}_\mu\left[\mathbf{1}_F \lim_n \mathbb{P}_{X_n}(T_B < \infty)\right] = \mathbb{P}_\mu\left(F \cap \{N_B = \infty\}\right), \tag{1.52}$$

for every $F \in \mathcal{F}_m$. Since this holds for every m, (1.52) holds for every $F \in \cup_m \mathcal{F}_m$ and therefore for every $F \in \mathcal{F}_\infty$. Therefore (1.49) is proven and the proof of Proposition 5.6 is complete. □

6. Invariant measures

Given a probability kernel $\mathtt{p}$ we say that a non zero σ-finite (positive) measure μ on $(E, \mathcal{E})$ is $\mathtt{p}$-invariant if $\mu\mathtt{p} = \mu$, that is if $\int_E \mu(\mathrm{d}x)\mathtt{p}(x, A) = \mu(A)$ for every $A \in \mathcal{E}$. We recall that μ is σ-finite if there exists a sequence of events (A_j) with $\mu(A_j) < \infty$ for every j and $E = \cup_j A_j$. For example, the Lebesgue measure on $\mathbb{R}$ is σ-finite because the Lebesgue measure of $(-n, n)$ is $2n < \infty$.

If $\mu(E) < \infty$ we can normalize μ, that is we can redefine μ as $\mu/\mu(E)$, so μ becomes a probability. If there exists a $\mathtt{p}$-stationary probability μ we can consider the stationary $\mathtt{p}$-MC X by choosing the law of X_0 to be μ: in this case the law of X_n does not depend on n (and this is the very concept of *stationary process*).

Note that when a MC has an invariant probability, the question of uniqueness is well posed (and important). On the other hand, if a MC has an invariant measure μ which is not normalizable, then, for every $c > 0$, $c\mu$ is an invariant measure too. And there is no *canonical* way of choosing c. So the question of uniqueness in this case makes sense only up to a multiplicative constant. In order to avoid repetitions given $\mathtt{p}$ we will say that μ is the *essentially unique* $\mathtt{p}$-invariant measure if μ' $\mathtt{p}$-invariant implies that there exists $c > 0$ such that $\mu' = c\mu$.

It is interesting to remark that if μ is a $\mathtt{p}$-invariant measure, then it is also $\mathtt{p}_k$-invariant (this is trivial). On the other hand (less trivial), if μ is a $\mathtt{p}_k$-invariant measure for a $k > 1$, then there exists a $\mathtt{p}$-invariant measure. In fact if we set $\mu' = \sum_{j=0}^{k-1} \mu \mathtt{p}_j / k$, then $\mu' \mathtt{p} = \sum_{j=1}^{k} \mu \mathtt{p}_j / k = \sum_{j=0}^{k-1} \mu \mathtt{p}_j / k = \mu'$.

We will not go much into the structure of the space of invariant probability measures. We just make the remark that *convex superpositions* of $\mathtt{p}$-invariant probabilities are still $\mathtt{p}$-invariant probabilities. By this we mean that if μ and μ' are two $\mathtt{p}$-invariant probabilities, $q\mu + (1-q)\mu'$ is a $\mathtt{p}$-invariant probability for $q \in [0,1]$.

We say that μ is $\mathtt{p}$-reversible if μ is σ-finite and if the measure $\mu(\mathrm{d}x)\mathtt{p}(x, \mathrm{d}y)$ on $(E \times E, \mathcal{E} \otimes \mathcal{E})$ is symmetric. More explicitly:

$$\int_{E \times E} h(x,y) \mu(\mathrm{d}x) \mathtt{p}(x, \mathrm{d}y) \,=\, \int_{E \times E} h(y,x) \mu(\mathrm{d}x) \mathtt{p}(x, \mathrm{d}y)\,, \tag{1.53}$$

for every bounded and measurable $h : E \times E \to \mathbb{R}$.

Proposition 6.1. *μ is $\mathtt{p}$-reversible. Then*

(1) μ is $\mathtt{p}$-invariant;

(2) if in addition μ is a probability, then if X is a $\mathtt{p}$-MC with X_0 of law μ we have that for every n the law of $(X_0, X_1, \ldots, X_n)$ coincides with the law of $(X_n, X_{n-1}, \ldots, X_0)$.

Proof. For what concerns (1) we write for every $A \in \mathcal{E}$

$$\mu \mathtt{p}(A) \,=\, \int_{E \times E} \mathbf{1}_E(x) \mathbf{1}_A(y) \mu(\mathrm{d}x) \mathtt{p}(x, \mathrm{d}y) \,=\, \int_{E \times E} \mathbf{1}_A(x) \mathbf{1}_E(y) \mu(\mathrm{d}x) \mathtt{p}(x, \mathrm{d}y)\,, \tag{1.54}$$

where the second equality is due to reversibility. Now we use that $\int_E \mathtt{p}(x, \mathrm{d}y) = 1$ for every x so, by Fubini-Tonelli Theorem, we obtain $\mu \mathtt{p}(A) = \mu(A)$.

For (2) we remark that

$$\begin{aligned}
\mathbb{P}_\mu\left(X_0 \in A_0, X_1 \in A_1\right) &= \int_{E\times E} \mathbf{1}_{A_0}(x_0)\mathbf{1}_{A_1}(x_1)\mu(\,\mathrm{d}x_0)\mathtt{p}(x_0,\,\mathrm{d}x_1) \\
&= \int_{E\times E} \mathbf{1}_{A_1}(x_0)\mathbf{1}_{A_0}(x_1)\mu(\,\mathrm{d}x_0)\mathtt{p}(x_0,\,\mathrm{d}x_1) \\
&= \int_{E\times E} \mathbf{1}_{A_1}(x_1)\mathbf{1}_{A_0}(x_0)\mu(\,\mathrm{d}x_1)\mathtt{p}(x_1,\,\mathrm{d}x_0) \\
&= \mathbb{P}_\mu\left(X_0 \in A_1, X_1 \in A_0\right) ,
\end{aligned} \tag{1.55}$$

where in the second step we used reversibility and the third step is just a change of variables. This verifies the claim for $n = 2$. We leave to the reader the details for the case of $n > 2$. □

7. The special case of Markov chains on countable state spaces

When E is countable we choose $\mathcal{E} = \mathcal{P}(E)$ (the set of all subsets of E) and we recall the notation $Q(x, y) := \mathtt{p}(x, \{y\})$, see (1.12).

In this case it is natural to introduce an *equivalence relation* between states: $x \sim y$ (x *and* y *communicate*) if $\mathbb{P}_x(T_y^\flat < \infty)$ and $\mathbb{P}_y(T_x^\flat < \infty)$. Note $x \sim y$ if and only if $\rho_{x,y} > 0$ and $\rho_{y,x} > 0$ for $x \neq y$ and $x \sim x$ for every x. Moreover we recall that $\rho_{x,y} > 0$ is equivalent to the existence of k such that $Q^k(x, y) > 0$, which in turn is equivalent to the existence of $x =: x_0, x_1, x_2, \ldots, x_k := y$ such that $Q(x_{j-1}, x_j) > 0$ for $j = 1, 2, \ldots, k$.

The kernel $\mathtt{p}$, or the associated MC, is said *irreducible* if the only equivalence class is E, that is if all states communicate.

The $\sim$-equivalence classes partition E into sets that may be closed or open: an equivalence class A is closed if $Q(x, y) = 0$ for $x \in A$ and $y \notin A$. Otherwise A is said to be open. By Proposition 5.3 we readily see that either $\rho_{x,x} = 1$ or $\rho_{x,x} < 1$ for every x in a given class: we therefore generalize the terminology introduced for states right after (1.45) by saying that an equivalence class is recurrent if it contains recurrent state (and hence all states in the class are recurrent), and we say that the class is transient otherwise. So an open class is transient, but note that a closed class needs not to be recurrent: the simple random walk on $\mathbb{Z}$ is irreducible if the probability of jumping to the right is not 0 or 1 and, as we will soon recall in more detail, it is recurrent if and only if it is symmetric. Hence the asymmetric simple random walk is transient, but $\mathbb{Z}$ is (trivially) a closed class.

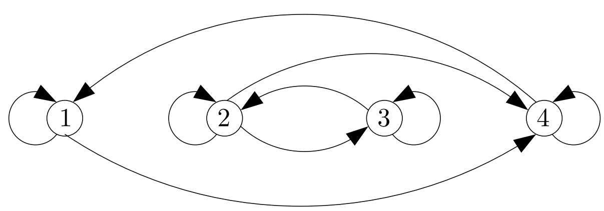
And, always by Proposition 5.3, if the class is open then x belongs to an open class (and therefore we say that the class is transient). In fact, Proposition 5.3 is also telling us that if x and y belong to the same class, then either $\rho_{x,x} = \rho_{y,y} = 1$ or both $\rho_{x,x} < 1$ and $\rho_{y,y} < 1$: as we will now see with examples, a closed class needs not to be (made of) recurrent (states). In order to make this more concrete let us develop an elementary exercise.

EXERCISE 7.1. *Le us consider a MC on $E = \{1, 2, 3, 4\}$ with*

$$Q = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 3/4 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix} . \tag{1.56}$$

Identify the $\sim$-equivalence classes and determine whether they are transient or recurrent. Find all the invariant measures for this MC.

SOLUTION. It is often useful to represent graphically the stochastic matrix Q: and the arrows show the allowed one step transitions. Hence there are two equiv-



alence classes, $\{1, 4\}$ and $\{2, 3\}$. The state 2 *leaks* towards state 4, hence $\{2, 3\}$ is open (hence transient). The class $\{1, 4\}$ instead is closed (hence recurrent. This class structures becomes more apparent at the matrix level if we exchange the label between the states 2 and 4. With this new numbering the stochastic matrix becomes

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 1/4 & 1/2 & 1/4 \end{pmatrix} , \tag{1.57}$$

which is made of the two 2×2 matrices

$$R := \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \text{and} \quad T := \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/4 \end{pmatrix} , \tag{1.58}$$

on the diagonal. Note that R is stochastic, but T is only substochastic, in the sense that $T_{2,1} + T_{2,2} < 1$. This is the mark of the open character of the class $\{2, 3\}$ in the original numbering. It is straightforward to find all the invariant measures that (in the original numbering) are simply $(c, 0, 0, c)$, $c > 0$. Hence the only invariant measure can be normalized to be a probability: $(1/2, 0, 0, 1/2)$. This is of course a direct linear algebra exercise, but (as we detail just below) an invariant probability gives probability zero to transient states. On the other hand, when E is finite any measure can be normalized. Moreover the chain restricted to the states $\{1, 4\}$ is irreducible, hence there is only one invariant measure for this chain. These considerations, plus the symmetry between the states 1 and 4 allow to find all the invariant measures with no computations. □

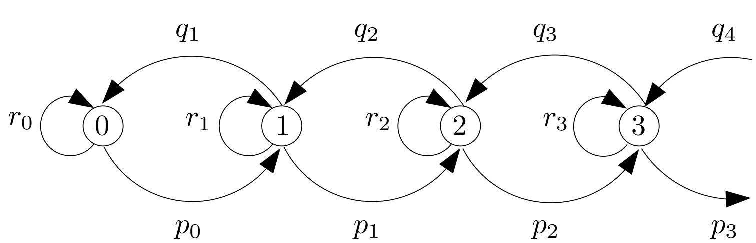
We remark that if μ is a $\mathtt{p}$-invariant probability, then $\mu(x) = \mu(\{x\}) = 0$ for every x in a transient class. This is simply because $\mathbb{E}_\mu[N_x] = \sum_{n=0}^\infty \mu(x)$ which is equal to ∞ if $\mu(x) > 0$. But this is impossible because $\mathbb{E}_\mu[N_x] = \sum_y \mu(y)\mathbb{E}_y[N_x] = \mu(x)\mathbb{E}_x[N_x] + \sum_{y \neq x} \mu(y)\rho_{y,x}\mathbb{E}_x[N_x] \leq \mathbb{E}_x[N_x] \sum_y \mu(y)$, so $U(x,x) = \mathbb{E}_x[N_x] = \infty$, i.e. x is recurrent. We insist that this is true only for invariant probabilities: invariant measures supported by transient classes may exist. And about this let us remark that if μ is an invariant measure with $\mu(x) > 0$, then $\mu(y) > 0$ for every x in the same equivalence class: $\mu(y) = \sum_x \mu(x)Q^k(x,y)$ for every k and, since $x \sim y$, there exists k such that $Q^k(x,y) > 0$, so $\mu(y) \geq \mu(x)Q^k(x,y) > 0$.

Finally, we say that a $\mathtt{p}$-MC is $\mathtt{p}$-irreducible if E is the only equivalence class (i.e., if all sites communicate). In this case all states are either recurrent or transient. In this case we simply say that the MC is recurrent, respectively transient, if there exists $x \in E$ which is recurrent (respectively, transient). We will show in Theorem 2.1 that a recurrent MC admits an invariant measure that is essentially unique: if this invariant measure is normalizable (hence, if it can be chosen to be a probability), then the invariant probability is unique. A consequence of Theorem 2.1(iv) is that an irreducible MC that admits an invariant probability is recurrent: but the existence of an invariant probability is not necessary for recurrence, so an irreducible MC that has an invariant probability (hence a unique invariant probability) is said *positive recurrent*. On the other hand, an irreducible MC that admits no invariant probability is said *null recurrent*. This terminology is extended to recurrent classes.

7.1. Birth and death chain. $E = \mathbb{N} \cup \{0\}$ and Q is defined by

$$Q(j, j+1) \,=\, p_j\,, \quad Q(j, j-1) \,=\, q_j\,, \quad Q(j,j) \,=\, r_j, \tag{1.59}$$

with $p_j + q_j + r_j = 1$ for every j. We assume that $p_j > 0$ for every $j \in E$, $q_j > 0$ for every $j \in E \setminus \{0\}$ and $q_0 = 0$.



It is straightforward to argue that this MC is irreducible: from every state the chain moves to both nearest-neighbors with positive probability (the graphical representation may help).

We introduce the function $\varphi : E \mapsto [0, \infty)$ defined by $\varphi(0) = 0$, $\varphi(1) = 1$ and by imposing that

$$(Q\varphi)(k) = \varphi(k) \quad \text{for } k = 1, 2, \ldots. \tag{1.60}$$

This yields $(\varphi(k+1)-\varphi(k)) = (q_k/p_k)(\varphi(k)-\varphi(k-1))$ for $k \geq 1$ and therefore for $n \geq 2$

$$\varphi(n) \;=\; 1 + \sum_{m=1}^{n-1} \prod_{j=1}^{m} \frac{q_j}{p_j}\,. \tag{1.61}$$

Note that φ is increasing, so $\lim_{n\to\infty} \varphi(n) =: \varphi(\infty)$ exists and takes value in $(1,\infty]$. Just for this example we write τ_a for $T_a^\flat$, that is $\tau_a := \inf\{n = 0,1,\ldots : X_n = a\}$. Remark that, by (1.60), φ is Q-harmonic on $\mathbb{N}$ (not on the whole of E because $Q\varphi(0) = p_0 > 0 = \varphi(0)$) so $(\varphi(X_{n\wedge\tau_0}))$ is a martingale. Remark also that, for every $L \geq x$, we have $\mathbb{E}_x(\tau_L) < \infty$: this follows from a standard estimates because $\inf_{x<L} \mathbb{P}_x(\tau_L \leq L) > 0$ (see for example [**8**, Ch. 10, Sec. 11]). Therefore by the Optional Stopping Theorem [**8**, Ch. 10, Sec. 10] we have that, for $a \leq x \leq b$, $\varphi(x) = \mathbb{E}_x[\varphi(X_{\tau_a\wedge\tau_b}]$ from which we readily extract

$$\mathbb{P}_x(\tau_b > \tau_a) \;=\; \frac{\varphi(b)-\varphi(x)}{\varphi(b)-\varphi(a)} \;\Longrightarrow\; \mathbb{P}_x(\tau_b > \tau_0) \;=\; 1 - \frac{\varphi(x)}{\varphi(b)}\,. \tag{1.62}$$

The event $\{\tau_b > \tau_0\}$ becomes larger when b grows, so

$$\lim_{b\to\infty} \mathbb{P}_x(\tau_b > \tau_0) \;=\; \mathbb{P}_x\left(\cup_{b\geq x}\{\tau_b > \tau_0\}\right) \;=\; 1 - \frac{\varphi(x)}{\varphi(\infty)}\,. \tag{1.63}$$

It is rather intuitive that this formula is saying that the MC is recurrent if $\varphi(\infty) = \infty$ (and it is recurrent if $\varphi(\infty) < \infty$. Below we give details for this, but it is better to get convinced about it (and try to prove it independently of what is written below).

If 0 is recurrent (hence any other x is) we have in particular $\rho_{x,0} = \mathbb{P}_x(\tau_0 < \infty) = 1$ for every x. Let us choose a value of $x \neq 0$ and note that for $b > x$ we have that $\tau_b \geq b - x$ under $\mathbb{P}_x$, so $\mathbb{P}_x(\tau_b > \tau_0) \geq \mathbb{P}_x(b - x > \tau_0)$ and $\lim_{b\to\infty} \mathbb{P}_x(b-x > \tau_0) = 0$ because $\mathbb{P}_x(\tau_0 < \infty) = 1$. Therefore $\lim_{b\to\infty} \mathbb{P}_x(\tau_b > \tau_0) = 1$ so, by (1.63), $\varphi(\infty) = \infty$.

On the other hand, if 0 is transient, i.e. $U(0,0) < \infty$, we have $U(x,0) < \infty$ which implies $\rho_{x,0} < 1$ (otherwise from x the chain goes a.s. to 0 and a.s. to x again because it makes just steps of length one and it cannot stay on a bounded set for an infinite time (see the estimate just before the applying the Optional Stopping Theorem), and we can repeat the argument indefinitely. Remark now that $\mathbb{P}_x(\tau_b < \tau_0) \geq \mathbb{P}_x(\tau_b < \tau_0,\, \tau_0 = \infty) = 1 - \rho_{x,0}$ for $b \geq x$, because $1 - \rho_{x,0} = \mathbb{P}_x(\tau_0 = \infty)$ and we have used again that this MC cannot stay in a bounded set for an infinite time and that it jumps to nearest neighbors. Therefore $\lim_{b\to\infty} \mathbb{P}_x(\tau_b < \tau_0) \geq 1 - \rho_{x,0} > 0$ and, by (1.63), $\varphi(\infty) < \infty$.

We have therefore proven that the birth and death MC is recurrent if and only if $\varphi(\infty) = \infty$.

Let us try to identify the invariant measures of this chain. This is an interesting exercise because it shows that finding the invariant measures can be a daunting task. In fact $\mu Q = \mu$ amounts to finding positive solutions μ of

$$\mu(j-1)p_{j-1} + \mu(j+1)q_{j+1} \;=\; \mu(i)(1-r_j) \quad \text{for every } j = 1,2,\ldots\,, \tag{1.64}$$

and $\mu(0) = \mu(1)q_1/p_0$. After some elementary (but not a priori obvious) steps we see that this is equivalent to finding positive solutions μ of

$$q_j\mu(j-1)\left(\frac{p_{j-1}}{q_j} - \frac{\mu(j)}{\mu(j-1)}\right) + p_j\mu(j+1)\left(\frac{q_{j+1}}{p_j} - \frac{\mu(j)}{\mu(j+1)}\right) = 0 \quad \text{for every } j\,. \tag{1.65}$$

At this stage we realize that one solution is

$$\mu(j) = \prod_{k=1}^{j}(p_{k-1}/q_k)\,, \quad \text{for } j = 1, 2, \ldots\,, \tag{1.66}$$

and $\mu(0) = \mu(1)q_1/p_0$, because with this choice both termes between parentheses in (1.65) are zero.

In reality it is not at clear that we have found all the solutions: in fact we solved the easier problem of finding the reversible measures! In fact reversibility is equivalent to requiring $\mu(j-1)Q(j-1, j) = \mu(j)Q(j, j-1)$, i.e. $\mu(j-1)p_{j-1} = \mu(j)q_j$ that readily leads to (1.66).

However this is not too bad, because, in Theorem 2.1, we will show that if a MC is irreducible and recurrent then the invariant measure is unique (up to a multiplicative factor). So, by exploiting Theorem 2.1, we have obtain the following rather complete result:

Proposition 7.2. *The birth and death MC is recurrent if and only if $\varphi(\infty) = \infty$. In this case ν is essentially unique. Moreover it is positive recurrent if and only if $\sum_x \prod_{k=1}^{x}(p_{k-1}/q_k) < \infty$.*

Note that this result implies in particular that the symmetric simple random walk is null recurrent, as well as the well known fact that an asymmetric simple random walk is transient. In fact the symmetric simple random walk ($E = \mathbb{Z}$ and $Q(j, j+1) = Q(j, j-1) = 1/2$ for every j) is *equivalent* to the birth and death MC with $q_j = p_j = 1/2$ for $j = 1, 2\ldots$ and $p_0 = 1$. To be precise the birth and death chain is the absolute value of the symmetric simple random walk. So $\varphi(\infty) = \infty$ (recurrence!), but the invariant measure μ in (1.66) is explicitly given by $\mu(j) = 2$ for $j = 1, 2, \ldots$ and $\mu(0) = 1$, hence $\mu(E) = \infty$ and the MC is null recurrent.

7.1.1. *A pathologic, but interesting, birth and death chain.* Let us briefly consider the case in which all the conditions of the birth and death MC are satisfies, except for $p_0 > 0$.

Assume $p_0 = 0$, hence $r_0 = 1$. In this case we loose irreducibility, in fact 0 traps the MC. So $E_0 := \{0\}$ represents a class on its own (of course a recurrent class): note that E_0 is accessible (from any state in the system). And $E_1 := \mathbb{N}$ is another class and it is clearly open. So, *one way or the other*, E_1 is transient. We say *one way or the other* because E_1 can be transient in different ways, in the sense that the MC could escape to infinity with positive probability, or it could fall into the trap E_0 with probability one.

In order to look at these different scenarios we consider the simplified set up of $p_j = p$ and $q_j = q$ for $j = 1, 2, \ldots$ and $p_0 = 1 - r_0 = 0$. If $p > 1/2$ the process is a random walk with increment with increment expectation $2p-1>0$, excepts when it hits 0 (and it is trapped). The law of large numbers suffices to conclude that $\mathbb{P}_x(\tau_0=\infty)>0$ for every $x>0$. But we can make this estimate quantitative exploiting the function φ in (1.61) which in this case becomes

$$\varphi(n) \,=\, \frac{1-(q/p)^n}{1-(q/p)}\,, \tag{1.67}$$

so that $\mathbb{P}_x(\tau_0<\infty)=\lim_{b\to\infty}\mathbb{P}_x(\tau_b>\tau_0)=1-\varphi(x)/\varphi(\infty)=(q/p)^x$. So the MC that starts from x is eventually trapped in 0 with probability $(q/p)^x$ and walks off to $+\infty$ with probability $1-(q/p)^x$.

If $q\ge p$ instead one can argue (we leave the details to the reader) that the MC is eventually trapped in 0 with probability 1.

Let us look at the invariant measure: it is clear that $\mu(x)=\delta_0(x)$ (Kronecker delta) is an invariant probability for every choice of p. But more than that is true: $\mu Q(0)=\mu(0)$ means $\mu(0)+\mu(1)q=\mu(0)$, that is $\mu(1)=0$. Then $\mu Q(1)=\mu(1)$ means $\mu(2)=\mu(1)/q=0$. And $\mu Q(j)=\mu(j)$ for $j\ge 2$ yields $(\mu(j+1)-\mu(j))q=(\mu(j)-\mu(j-1))p$, so $\mu(j+1)-\mu(j)=0$ for every $j\ge 2$. This means that $\mu(x)=\delta_0(x)$ is the unique invariant measure (up to a multiplicative factor) and it is the unique invariant probability.

We will see that such a statement follows from the general theory (Theorem 2.1) because 0 is recurrent and it is accessible. While this model of non irreducible birth and death chain is artificial it illustrates well a phenomenon that is not artificial at all (see Sec. 7.3).

7.2. The simple random walk on $\mathbb{Z}$. Here we are again with the simple random walk on $E=\mathbb{Z}$, that is $X_{n+1}=X_n+\xi_{n+1}$, $\mathbb{P}(\xi=1)=1-\mathbb{P}(\xi=1)=p$. This MC is of course irreducible. The quickest way to see that it is transient for $p\neq=1/2$ is to use the law of large numbers: $\lim_n X_n/n=2p-1$ a.s., so every state is visited at most a finite number of times. On the other hand there are several ways to see that if $p=1/2$ the simple random walk (in this case, symmetric) is recurrent: an elementary (but computationally possibly a bit demanding) way to see it is to use the binomial formula for $Q^{2n}(x,x)$ and apply Stirling asymptotic formula to see that $Q^{2n}(x,x)=Q^{2n}(0,0)\sim 1/(\pi\sqrt{n})$, and of course $Q^{2n+1}(x,x)=0$. so $U(x,x)=\sum_n Q^{2n}(0,0)=\infty$.

Let us find the invariant measures: $\mu Q=\mu$ spells out

$$\mu(j-1)p+\mu(j+1)(1-p) \,=\, \mu(j) \;\text{ for every } j\in\mathbb{Z}\,. \tag{1.68}$$

This means that $(\mu(j+1)-\mu(j))/(\mu(j)-\mu(j-1))=p/(1-p)$.

Therefore if $p=1/2$, $\mu(j)=\mu(0)+jc$ for every choice of c. But the only choice that yields $\mu(j)\ge 0$ for every j is $c=0$. So the only invariant measure is the uniform measure on $\mathbb{Z}$: $\mu(j)=\mu(0)>0$ for every j.

On the other hand, if $p\neq 1/2$ the general solution is $\mu(x)=C_0+C_1(p/(1-p))^x$, which is a positive measure if $C_0\ge 0$, $C_1\ge 0$ and $C_0\wedge C_1>0$. So we see that we have a two dimensional family of measures: the invariant measure is a linear combination of the uniform measure and a measure that grows exponentially at $+\infty$ or at $-\infty$, according to whether $p > 1/2$ or $p < 1/2$. Therefore we see that when the simple random walk is transient, essential uniqueness of the invariant measure is lost.

One should not draw quick conclusions from this example: in this case we have see that we loose uniqueness on the invariant measure if the MC is transient, but in Section 7.4 we will present an example of a transient MC with no invariant measure.

We will see (again, Theorem 2.1) that what we can exclude for a transient MC is the existence of an invariant probability: we will in fact show that if an invariant probability exists, the MC is positive recurrent.

7.3. Branching process (Bienaymé-Galton-Watson process). The BGW process (Z_n) is a MC on $E := \mathbb{N} \cup \{0\}$ defined starting from the IID family $\xi := (\xi_{n,j})_{(n,j)\in\mathbb{N}^2}$, with $\mathbb{P}(\xi_{1,1} \in E) = 1$. We use the notation $p_j := \mathbb{P}(\xi_{1,1} = j)$ and we assume $p_j < 1$ for every j to avoid trivialities. The chain can be introduced by iteration once Z_0 independent of ξ is given (unless otherwise said, we choose $Z_0 = 1$) via

$$Z_{n+1} = \begin{cases} \xi_{n+1,1} + \xi_{n+1,2} + \ldots + \xi_{n+1,Z_n} & \text{if } Z_n > 0, \\ 0 & \text{if } Z_n = 0. \end{cases} \tag{1.69}$$

We assume hat $\mu = \mathbb{E}[\xi_{1,1}] = \sum_j jp_i = \mu \in (0,\infty)$ and that $p_1 < 1$ (to avoid trivialities).

It is useful to establish that (Z_n/μ^n) is a (non-negative) martingale with respect to the natural filtration of the MC (this is left as exercise). Hence $\lim_n Z_n/\mu^n$ exists a.s. and we denote the limit (non-negative) random variable by H. Understanding whether $H \equiv 0$ or not is very helpful in understanding the BGW chain.

EXERCISE 7.3. *Show that if $\mu > 1$ and $\mathbb{E}[\xi_{1,1}^2] < \infty$ then the martingale (Z_n/μ^n) is UI (Uniformly Integrable, see [**8**, Ch. 13]), hence in this case $H \not\equiv 0$.*

SOLUTION. Set $\sigma^2 := \mathrm{var}(\xi_{1,1})$, $\sigma^2 > 0$ because $\xi_{1,1}$ is non trivial, and $M_n := Z_n/\mu^n$. We have that $\mathbb{E}[Z_{n+1}^2|Z_n] = \mu^2 Z_n^2 + \sigma^2 Z_n$ so $\mathbb{E}[Z_{n+1}^2] = \mu^2\mathbb{E}[Z_n^2] + \sigma^2\mu^n$ and $\mathbb{E}[M_{n+1}^2] = \mathbb{E}[M_n^2] + \sigma^2\mu^{-n-2}$. Therefore $\sup_n \mathbb{E}[M_n^2] < \infty$ which implies that (M_n) is UI. A UI martingale converges a.s. and in $\mathbb{L}^1$. So $\mathbb{E}[\lim_n Z_n/\mu^n] = \mathbb{E}[H] = 1$ which means that $H \not\equiv 0$. $\square$

For $s \in (0,1]$ we introduce also $\varphi(s) = \mathbb{E}[s^{\xi_{1,1}}]$. Note that (*Exercise*) $\varphi(\cdot)$ is convex, increasing and smooth. Since $\varphi(0) = \lim_{s\searrow 0} \varphi(s) = p_0$ and $\varphi(1) = 1$, there exists only one solution in $[0,1)$ to the the fixed point equation $s = \varphi(s)$. Call this solution ϱ.

PROPOSITION 7.4. *0 is a recurrent state for the chain (Z_n): note that 0 is accessible (from any other state, i.e. $\rho_{n,0} > 0$ for every n) if and only if $p_0 > 0$. All other states n are transient. Moreover*

(1) if $\mu \le 1$ then $\sum_n \mathbf{1}_{Z_n>0} < \infty$ a.s. (hence $H \equiv 0$);

(2) if $\mu > 1$ then $\mathbb{P}(H = 0) = \rho$, hence $\mathbf{1}_{H=0} \sum_n \mathbf{1}_{Z_n>0} < \infty$ a.s., and if $\mathbb{E}[\xi_{1,1}^2] < \infty$ on the event $\{H > 0\}$ we have $Z_n \sim H\mu^n$ a.s. (here $\sim$ is aymptotic equivalence).

If $p_0 > 0$ he only invariant measure can be normalized and it is δ_0. If $p_0 = 0$ then there exists no invariant measure.

Proof. If $p_0 = 0$ then 0 is not accessible, so we can even choose $E = \mathbb{N}$, and $\mu > 1$. If $p_0 > 0$ instead 0 becomes accessible.

If $\mu < 1$, recalling that (Z_n/μ^n) converges a.s. we see that $Z_n = O(\mu^n) \to 0$ a.s., hence $Z_n = 0$ for n sufficiently large because the process (Z_n) is integer valued. This implies directly also that $H \equiv 0$. In the *critical* $\mu = 1$ case we can also conclude that $\lim_n Z_n = 0$, hence that , hence $Z_n = 0$ for n sufficiently large and $H \equiv 0$, because $\lim_n Z_n =: Z_\infty$ exists a.s. and this (a priori random) limit can only be 0 because otherwise $Z_n(\omega) = Z_\infty(\omega)$ for n sufficiently large and $Z_\infty \not\equiv 0$ implies that $\xi \equiv 0$, which is impossible. So for $\mu \leq 1$ (critical and subcritical case) the process hits (and is absorbes by) 0 in a finite time.

Let us turn to $\mu > 1$. In this case $\mathbb{E}[Z_n] = \mu^n$ grows exponentially. But this does not mean that (Z_n) grows exponentially, but recall that we know that $Z_n(\omega) \sim H(\omega)\mu^n$ where this asymptotic equivalence is properly stated only if $H(\omega) > 0$. And we did establish that H is not trivial if the ξ variables are in $\mathbb{L}^2$.

What happens for $\mu > 1$ is that the process may not grow because it gets to 0 and stays there (*extinction*). In order to compute the probability of extinction let us remark that for $n \in \mathbb{N}$ and $s \in (0, 1]$

$$\mathbb{E}\left[s^{Z_n}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{Z_n} \middle| Z_{n-1}\right]\right] = \mathbb{E}\left[\varphi(s)^{Z_{n-1}}\right], \tag{1.70}$$

so

$$\mathbb{E}\left[s^{Z_n}\right] = \varphi \circ \ldots \circ \varphi(s) =: \varphi^{\circ n}(s). \tag{1.71}$$

It is elementary to see that $\lim_n \varphi^{\circ n}(s) = \varrho$ for every $s < 1$.

Remark now that if $H(\omega) = 0$, then $Z_n = 0$ for n large, hence (by Dominated Convergence) $\lim_n \mathbb{E}[s^{Z_n}; H = 0] = \mathbb{P}(H = 0)$ for every $s \in (0, 1)$. On the other hand $\lim_n \mathbb{E}[s^{Z_n}; H > 0] = 0$ for every $s \in (0, 1)$ (again by Dominated convergence, because $Z_n \to \infty$ when $H > 0$). Hence $\mathbb{P}(H = 0) = \varrho$.

From the MC viewpoint we remark that we have proven that for $\mu > 1$ the chain may get absorbed by 0 or may go infinity.

For what concerns the invariant measures we remark that we must have in particular $\mu Q(0) = \mu(0)$, that is $\mu(0) + \sum_{j=1}^\infty \mu(j)Q(j, 0) = \mu(0)$ because $Q(0, 0) = 1$. Hence if if $p_0 > 0$, then $Q(j, 0) > 0$ for every j, and we can conclude that $\mu(j) = 0$ for every $j > 0$.

On the other hand, if $p_0 = 0$ we choose $E = \mathbb{N}$. The invariance condition is $\sum_{j=1}^\infty \mu(j)Q(j, k) = \mu(k)$ for every $k \in \mathbb{N}$. But $Q(j, k) = 0$ if $k < j$, so the invariance condition may be written as $\sum_{j=1}^k \mu(j)Q(j, k) = \mu(k)$. In particular $\mu(1)Q(1, 1) = \mu(1)$, that is $\mu(1) = 0$ because $Q(1, 1) < 0$. We can the iterate this argument: if $\mu(1) = \mu(2) = \ldots = \mu(k-1) = 0$, then $\mu(k)Q(k, k) = \mu(k)$ and $\mu(k) = 0$ because $Q(k, k) < 1$. $\square$

7.4. The discrete renewal process. A discrete renewal process is just a random walk with positive increments: $S_0 := 0$ and $S_{n+1} = S_n + \xi_{n+1}$ with $\{\xi_j\}_{j\in\mathbb{N}}$ IID with $\mathbb{P}(\xi_1 \in \mathbb{N} \cup \{\infty\}) = 1$. If $S_0 \neq 0$ (and we choose to work with $S_0 \geq 0$) we will say that S is a renewal sequence with delay.

While the random walk viewpoint is natural, it is as natural and at times helpful for the intuition to consider the renewal process as a *point process*. By this we mean that we look at the sequence S (with or without delay) as a random subset of $\mathbb{N} \cup \{0\} \cup \{\infty\}$: $\eta := \{S_0, S_1, \ldots\}$ is the random subset. Note that we do not exclude the case in which $\mathbb{P}(\xi = \infty) > 0$: in this case the renewal set contains only a finite number of points and ∞. So $\{k \in \eta\}$ is the event {there exists $j = 0, 1, \ldots$ such that $S_j = k$}. The points in η are called renewal epochs, the ξ variables are the inter-arrival variables and η is the renewal set.

Here are two relevant questions for us:

(1) What is the probability that n is a renewal epoch? That is, what is the value of $u(n) := \mathbb{P}(n \in \eta)$? Does the limit of the sequence $(u(n))_{n=0,1,\ldots}$, called *renewal sequence*, exist and what is its value?
(2) Consider the intersection of two independent renewal sets with same inter-arrival law but different delay. Under which conditions is this random set almost surely not empty?

The first question is a classical important question. The second one solves a technical step that is going to be important in the general theory of Markov Chains. But the discrete renewal process provides also a very nice example of *fully solvable* Markov chain.

In order to study η we introduce a MC called *Backward Recurrence Time* process: $A_n := n - \sup\{S_k : S_k \leq n\}$, that is A_n is the time elapsed since the last renewal epoch if we are at time n (see Fig. 1). In particular, $\eta = \{n : A_n\}$ is the zero level set of the random function $n \mapsto A_n$, when $\mathbb{P}(\xi < \infty) = 1$, otherwise $\eta = \{n : A_n\} \cup \{\infty\}$.

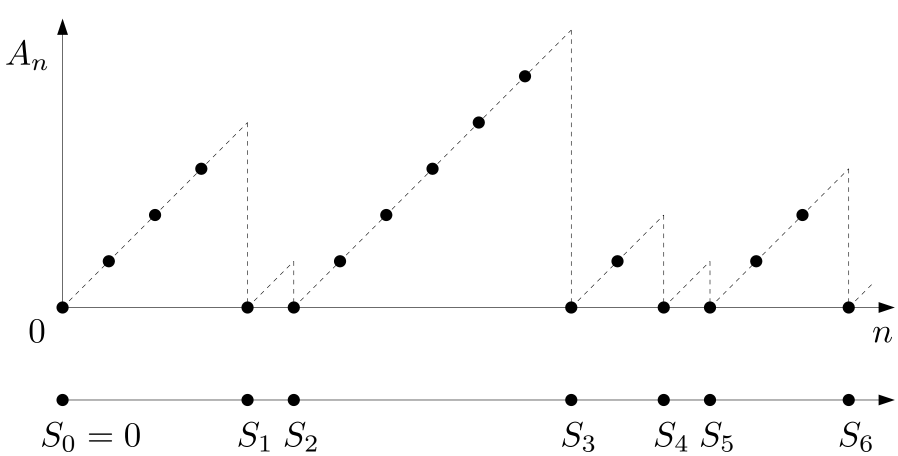


FIGURE 1. A trajectory of a renewal process without delay ($S_0 = 0$) and, above, the corresponding backward renewal time (A_n).

Note that, given A_n we know that A_{n+1} is either A_n+1 or 0. We invite the reader to provide a proof of the (a priori not obvious) fact that (A_n) is a MC with $E=\mathbb{N}\cup\{0\}$ and for $j=0,1,\ldots$

$$Q(j,j+1)\;=\;\mathbb{P}\left(\xi>j+1|\xi>j\right)\;=\;\frac{\mathbb{P}\left(\xi>j+1\right)}{\mathbb{P}\left(\xi>j\right)}\;=:\;p_j\,,\,, \tag{1.72}$$

with $\xi\sim\xi_1$. Of course $Q(j,0)=1-p_j$.

Let us forget for a while the ξ variables and let us just consider the Q-MC with a general sequence of (p_j) that are simply numbers in $[0,1]$. Note however that if $p_j=0$ then the chain cannot climb above j: by (1.72) we see that if $p_j=0$ then $p_k=0$ for every $k>j$. So we make this hypothesis on the (p_j) sequence. In principle, if $p_j=0$ for $j\geq K$ (and $p_j<1$ otherwise) we could choose as state space $E=\{0,1,\ldots,K\}$, but we will not do so: we can imagine to start with a delay A_0 larger than K, the $A_1=0$ and the chain will stay in $\{0,1,\ldots,K\}$ for all times in $\mathbb{N}$. The states in $\{K+1,K+2,\ldots\}$ are just transient and $\{0,1,\ldots,K\}$ is the (closed hence) recurrent class for the MC.

It is straightforward to see that the chain is irreducible if $p_j>0$ for every j and if $p_j<1$ for infinitely many j's.

Let us try to compute $\rho_{0,0}$ to decide if the chain is recurrent or transient: if the chain is not irreducible then the result holds for the class to which 0 belongs, so for $\{0,1,\ldots,K\}$ with the notation we used just above. Note that if we keep in the game the ξ variables we see directly that $\rho_{0,0}=\mathbb{P}(\xi<\infty)$ and, as it is clear from the origin of the process, 0 is recurrent if and only if the inter-arrivals are a.s. finite. But let us compute using the (p_j) sequence:

$$\mathbb{P}_0\left(T_0=j+1\right)\;=\;p_0p_1\cdots p_{j-1}(1-p_j)\qquad\text{for}\;\;j=0,1,2,\ldots\,, \tag{1.73}$$

where of course we mean $\mathbb{P}(T_0=1)=1-p_0$. So if we introduce the non increasing sequence $\Pi_j:=p_0p_1\cdots p_{j-1}$ for $j=1,2,\ldots$ and $\Pi_0:=1$ we see that $\sum_{j=0}^n\mathbb{P}_0(T_0=j+1)=\Pi_0-\Pi_{n+1}$ for $n=0,1,\ldots$. Therefore

$$\rho_{0,0}\;=\;\mathbb{P}_0\left(T_0<\infty\right)\;=\;\Pi_0-\Pi_\infty\;=\;1-\Pi_\infty\,, \tag{1.74}$$

and the chain is recurrent if and only if $\Pi_\infty=0$. And in fact from (1.72) we see that $\Pi_\infty=\mathbb{P}(\xi=\infty)$.

Let us turn to the invariant measures. Luckily the computation turns out to be particularly simple: for $j=1,2,\ldots$ we have $\mu Q(j)=\mu(j)$ yields $\mu(j)=\mu(j-1)p_{j-1}$, so $\mu(j)=\mu(0)p_0p_1\cdots p_{j-1}$. Moreover $\mu Q(0)=\mu(0)$ yields

$$\sum_{j=0}^\infty\mu(j)(1-p_j)\;=\;\mu(0)\,, \tag{1.75}$$

that is $\sum_{j=0}^\infty p_0\cdots p_{j-1}(1-p_j)\;=\;1$ which simply means that $\mathbb{P}_0(T_0<\infty)=1$.

The conclusion is rather remarkable: there exists an invariant measure if and only if the chain is recurrent! This gives another example of a MC without invariant measures (we have see that this is the case also for the BGW process when $p_0=0$).

Let us investigate when μ can be normalized. Of course the condition is:

$$\sum_{j=0}^{\infty} \mu(j) \;=\; \mu(0)\left(1+\sum_{j=1}^{\infty} p_0 p_1 \cdots p_{j-1}\right) \;<\; \infty. \tag{1.76}$$

This becomes more enlightening if we go back to (1.72). In fact (1.76) is equivalent to $1+\sum_{j=1}^{\infty} \mathbb{P}(\xi > j) = \mathbb{E}[\xi]$. So there exists a (unique) probability if and only if the inter-arrival variable is in $\mathbb{L}^1$. And the invariant probability is for $j = 0, 1, \ldots$

$$\pi(j) \;=\; \frac{\mathbb{P}(\xi > j)}{\mathbb{E}[\xi]}. \tag{1.77}$$

We are going to pick up again the analysis of the renewal process in Section 6.2 of Chapter 2.

CHAPTER 2

Markov Chains with accessible recurrent states

1. Accessible recurrent states

This chapter is devoted to a class of MC that a priori may look artificial: we consider MC on a state space $(E, \mathcal{E})$ that is general for which there exists $x \in E$ such that

(1) $\rho_{x,x} = 1$, that is x is recurrent;
(2) $\rho_{y,x} > 0$ for every $y \in E$, that is x is accessible (from every state).

In this case for conciseness we will simply say that x is accessible and recurrent. If only the second condition is satisfied, then x is simply accessible. Note that if x is recurrent and accessible, then if there is another state y that is recurrent, then, by Proposition 5.4, y is accessible too (in fact, $\rho_{x,x} = \rho_{x,y} = \rho_{y,x} = \rho_{y,y} = 1$). So, if x is recurrent and accessible, we can simply say that the MC is recurrent. We will see that this generalizes also to finite and null recurrence.

This class of MC of course contain plenty of natural models when E is countable. We are now going to give an interesting example of a MC with an accessible (possibly recurrent) state. But we want to stress that in the next chapter we will show that several MC that do not have a accessible states can be related to an auxiliary MC that has an accessible (possibly recurrent) state and that results proven for the auxiliary MC can be transferred to the original MC. This of course enhances substantially the importance of the content of this chapter.

1.1. Reflected walks: the Lindley process. The Lindley process with increments ξ is the MC with state space $E = [0, \infty)$ (in general $\mathcal{E} = \mathcal{B}(E)$) defined in the standard random dynamical system way by

$$X_{n+1} = (X_n + \xi_{n+1})_+ \,, \tag{2.1}$$

with the ξ variables taking values in $\mathbb{R}$ and of course $x_+ = \max(x, 0)$. Therefore the kernel of the Lindley MC is

$$\mathtt{p}(x, [0, y]) = \mathbb{P}(\xi_1 + x \le y) \,, \tag{2.2}$$

for every $y \ge 0$. In particular

$$\mathtt{p}(x, \{0\}) = \mathbb{P}(\xi_1 \le -x) \,. \tag{2.3}$$

Such a process is a random walk reflected at the origin: In fact, to the Lindley process we just introduced one naturally associates the random walk S on $\mathbb{R}$, defined by $S_{n+1} = S_n + \xi_{n+1}$. Note that if $S_0 = X_0 = x \ge 0$, then $S_n = X_n$ for $n < \inf\{n : S_n < 0\}$, see Fig. 1.

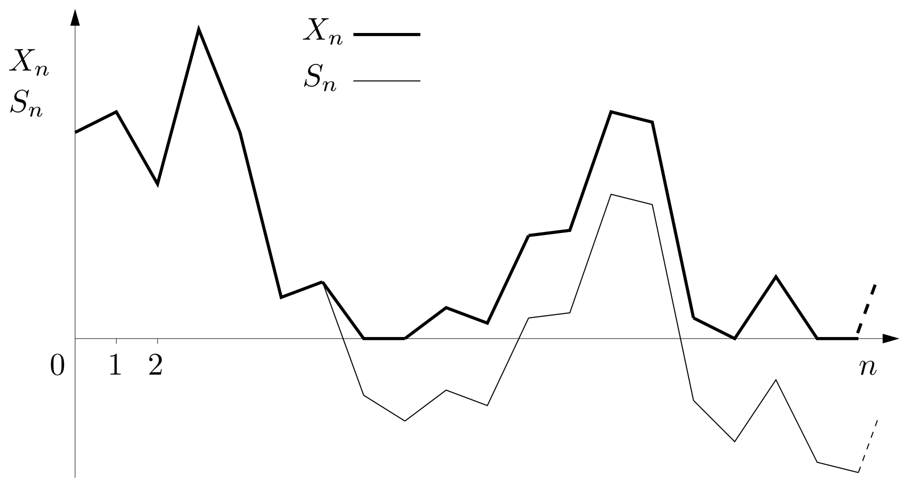


FIGURE 1. A trajectory of a Lindley process (X_n), thick line, and of the underlying random walk (S_n), thin line. The two processes coincide for the first six steps, so the thick line hides the thin line.

It is clear that 0 is accessible as soon as $\mathbb{P}(\xi_1 < 0) > 0$: we are assuming this to avoid trivialities. Whether 0 is recurrent or not is, in general, less evident, and we are going to address this issue. But it is for example easy to establish that if $\xi_1 \in \mathbb{L}^1$ and if $\mathbb{E}[\xi_1] < 0$ then 0 is recurrent (in fact, positive recurrent, where, like in the case of E countable, positive recurrence means that $\mathbb{E}_0[T_0] < \infty$).

2. Invariant measures and accessibles recurrent states

If x is accessible for the kernel $\mathtt{p}$ we introduce the measure λ_x by setting

$$\lambda_x(A) = \mathbb{E}_x\left[\sum_{j=1}^{T_x} \mathbf{1}_A(X_j)\right], \tag{2.4}$$

for every $A \in \mathcal{E}$ (and X is a $\mathtt{p}$-MC). Note that if $\mathbb{P}_x(T_x < \infty) = 1$, i.e. if x is recurrent, we have

$$\lambda_x(A) = \mathbb{E}_x\left[\sum_{j=0}^{T_x-1} \mathbf{1}_A(X_j)\right], \tag{2.5}$$

just because $X_j = x$ both at $j = 0$ and $j = T_x$.

THEOREM 2.1. *Choose a probability kernel* $\mathtt{p}$ *which admits an accessible state* x. *Then*

(1) *If x is recurrent then λ_x is $\mathtt{p}$-invariant.*
(2) *If λ_x is $\mathtt{p}$-invariant then x is recurrent.*
(3) *If x is recurrent and μ is a $\mathtt{p}$-invariant measure, then $\mu(\{x\}) < \infty$ and $\mu = \mu(\{x\})\lambda_x$.*

(4) If x is recurrent then $\mathbb{E}_x[T_x] < \infty$ if and only if there exists a unique p-invariant probability π. In this case $\pi = \lambda_x/\mathbb{E}_x[T_x]$.

We are going to say that x is positive recurrent if we are in the framework of point (4), that is if $\mathbb{E}_x[T_x] < \infty$. Of course we say that x is null recurrent if it is recurrent but $\mathbb{E}_x[T_x] = \infty$. Note that Theorem 2.1 (3) is an essential uniqueness statement.

Proof. For (1) let us start by showing that $\lambda_x \mathsf{p} = \lambda_x$. For this we rewrite (2.5) as

$$\lambda_x(A) = \mathbb{E}_x\left[\sum_{j=0}^{\infty} \mathbf{1}_A(X_j)\mathbf{1}_{\{T_x>j\}}\right] = \sum_{j=0}^{\infty} \mathbb{P}_x\left(X_j \in A,\, T_x > j\right) =: \sum_{j=0}^{\infty} \widetilde{\mathsf{p}}_j(x, A)\,. \tag{2.6}$$

Note that $\widetilde{\mathsf{p}}_j$ is a kernel, in the sense that $x \mapsto \widetilde{\mathsf{p}}_n(x, A)$ is measurable for every A and $\widetilde{\mathsf{p}}_j(x, \cdot)$ is a measure for every x. Of course $\widetilde{\mathsf{p}}_j(x, E) \le 1$ so it is not a probability kernel. In particular, one sees that if $h : E \to \mathbb{R}$ is measurable (and positive or bounded), $\int_E \widetilde{\mathsf{p}}_j(x,\, \mathrm{d}y)h(y) = \mathbb{E}_x[h(X_j); T_x > j]$.

Let us consider first the case $x \notin A$ and observe that

$$\begin{aligned}\int_E \widetilde{\mathsf{p}}_j(x,\, \mathrm{d}y)\mathsf{p}(y, A) &= \mathbb{E}_x\left[\mathsf{p}(X_j, A);\, T_x > j\right] \\ &= \mathbb{E}_x\left[\mathbb{P}_x\left(X_{j+1} \in A \middle| \mathcal{F}_j\right);\, T_x > j\right] \\ &= \mathbb{P}_x\left(T_x > j,\, X_{j+1} \in A\right) \\ &= \mathbb{P}_x\left(T_x > j+1,\, X_{j+1} \in A\right) = \widetilde{\mathsf{p}}_{j+1}(x, A)\,,\end{aligned} \tag{2.7}$$

where the second step is the (simple) Markov Property, in the third step we used $\{T_x > j\} \in \mathcal{F}_j$ to carry the indicator function inside the conditional expectation, that then becomes an expectation, and in the fourth step we used $x \notin A$. Therefore if $x \notin A$ we have

$$\int_E \lambda_x(\,\mathrm{d}y)\mathsf{p}(y, A) = \sum_{j=0}^{\infty} \widetilde{\mathsf{p}}_{j+1}(x, A) = \sum_{j=1}^{\infty} \widetilde{\mathsf{p}}_j(x, A) \overset{x\notin A}{=} \sum_{j=0}^{\infty} \widetilde{\mathsf{p}}_j(x, A) = \lambda_x(A)\,, \tag{2.8}$$

where in the last step we have used (2.6).

Let us consider now the case $A = \{x\}$: in this case we proceed exactly like before, except for the last step

$$\begin{aligned}\int_E \widetilde{\mathsf{p}}_j(x,\, \mathrm{d}y)\mathsf{p}(y, \{x\}) &= \mathbb{E}_x\left[\mathsf{p}(X_j, \{x\}); T_x > j\right] \\ &= \mathbb{E}_x\left[\mathbb{P}_x\left(X_{j+1} = x \middle| \mathcal{F}_j\right);\, T_x > j\right] \\ &= \mathbb{P}_x\left(T_x > j,\, X_{j+1} = x\right) = \mathbb{P}_x\left(T_x = j+1\right)\,.\end{aligned} \tag{2.9}$$

Recalling (2.6) this implies that

$$\int_E \lambda_x(\,\mathrm{d}y)\mathsf{p}(y, \{x\}) = \sum_{j=0}^{\infty} \mathbb{P}_x\left(T_x = j+1\right) = \mathbb{P}_x(T_x < \infty)\,(=\,1)\,. \tag{2.10}$$

But it is straightforward to see also that $\lambda_x(\{x\}) = \mathbb{P}_x(T_x < \infty)$, hence we have proven that $\lambda_x \mathtt{p}(A) = \lambda_x(A)$ for every $A \in \mathcal{E}$.

In order to conclude that λ_x is invariant we are left to show that is σ-finite. For this we use that $\lambda_x(\{x\}) = 1 < \infty$. Since $\lambda_x \mathtt{p}_j = \lambda_x$ we have also $\lambda_x \sum_{j=1}^{\infty} 2^{-j} \mathtt{p}_j = \lambda_x \mathtt{p}_\star = \lambda_x$, where $\mathtt{p}_\star$ has been introduced in (1.23). Since we have assumed that x is accessible, $\mathtt{p}_\star(y, \{x\}) > 0$ for every $y \in E$, which implies that $E = \cup_{n \in \mathbb{N}} \{y \in E : p_\star(y, \{x\}) \geq 1/n\} =: \cup_{n \in \mathbb{N}} E_n$. Then

$$\infty \,>\, \lambda_x(\{x\}) \,=\, \int_E \lambda_x(\,\mathrm{d}y) \mathtt{p}_\star(y, \{x\}) \,\geq\, \int_{E_n} \lambda_x(\,\mathrm{d}y) \mathtt{p}_\star(y, \{x\}) \,\geq\, \frac{1}{n} \lambda_x\left(E_n\right) . \tag{2.11}$$

So $\lambda_x\left(E_n\right) < \infty$ for every n and λ_x is σ-finite.

Therefore λ_x is invariant and (1) is proven.

For what concerns (2), we have to show that x is recurrent. So let us assume that x is transient, i.e. $\mathbf{P}_x(T_x = \infty) > 0$, and let us show that this is incompatible with the hypothesis that λ_x is $\mathtt{p}$-invariant. Since $q_x := \mathbf{P}_x(T_x = \infty) > 0$ we set

$$\lambda_x^{\sim}(A) \,:=\, \mathbb{E}_x \left[\sum_{j=0}^{T_x - 1} \mathbf{1}_A(X_j) \right] , \tag{2.12}$$

and by direct inspection we see that $\lambda_x^{\sim} = \lambda_x + q_x \delta_x$. Moreover (2.8) and (2.10) imply that $\lambda_x^{\sim} \mathtt{p} = \lambda_x$. Therefore $\lambda_x^{\sim} \mathtt{p} = \lambda_x + q_x \delta_x \mathtt{p}$ by $\mathtt{p}$-invariance of λ_x. Since $\delta_x \mathtt{p}(A) = \mathtt{p}(x, A)$ is a positive measure, we obtain a contradiction. Hence x is recurrent and (2) is proven.

For what concerns (3), let us start by observing that, since μ is invariant, it is σ-finite. So $\mu(\{x\}) < \infty$. On the other hand $\mu(\{x\}) > 0$ because $\mu(\{x\}) = 0$ implies $\mu \equiv 0$ because x is accessible (this can be seen by arguing like in the beginning of (2.11): $0 = \mu(\{x\}) = \int_E \mu(\,\mathrm{d}y) \mathtt{p}_*(y, \{x\})$ implies $\mu \equiv 0$ because $\mathtt{p}_*(y, \{x\}) > 0$ for every y).

The idea now is based on the following scheme: for every $A \in \mathcal{E}$

$$\begin{aligned} \mu(A) \,&=\, \mu \mathtt{p}(A) \,=\, \mu(\{x\}) \mathtt{p}(x, A) + \int_{E \setminus \{x\}} \mu(\,\mathrm{d}y) \mathtt{p}(y, A) \\ &=\, \mu(\{x\}) \mathbb{P}_x(X_1 \in A) + \mathbb{P}_\mu\left(X_0 \neq x, X_1 \in A\right) \end{aligned} \tag{2.13}$$

and then (restarting from the first line of (2.13))

$$\begin{aligned} \mu(A) \,&=\, \mu(\{x\}) \mathtt{p}(x, A) + \int_{E \setminus \{x\}} \int_E \mu(\,\mathrm{d}y') \mathtt{p}(y', \,\mathrm{d}y) \mathtt{p}(y, A) \\ &=\, \mu(\{x\}) \mathtt{p}(x, A) + \mu(\{x\}) \int_{E \setminus \{x\}} \mathtt{p}(x, \,\mathrm{d}y) \mathtt{p}(y, A) \\ &\qquad + \int_{E \setminus \{x\}} \int_{E \setminus \{x\}} \mu(\,\mathrm{d}y') \mathtt{p}(y', \,\mathrm{d}y) \mathtt{p}(y, A) \\ &=\, \mu(\{x\}) \mathbb{P}_x(X_1 \in A) + \mu(\{x\}) \mathbb{P}_x(X_1 \neq x, X_2 \in A) \\ &\qquad + \mathbb{P}_\mu\left(X_0 \neq x, X_1 \neq x, X_2 \in A\right) , \end{aligned} \tag{2.14}$$

and so on to obtain that for every n

$$\mu(A) \,=\, \mu(\{x\}) \sum_{j=1}^{n} \mathbb{P}_x(X_k \neq x \text{ for } k=1,\ldots,j-1,\, X_j \in A) \\ + \mathbb{P}_\mu\left(X_k \neq x \text{ for } k=0,\ldots,n-1,\, X_n \in A\right), \quad (2.15)$$

from which we infer

$$\begin{aligned} \mu(A) \,&\geq\, \mu\left(\{x\}\right) \sum_{j=1}^{\infty} \mathbb{P}_x(X_k \neq x \text{ for } k=1,\ldots,j-1,\, X_j \in A) \\ &=\, \mu\left(\{x\}\right) \sum_{j=1}^{\infty} \mathbb{P}_x(T_x \geq j,\, X_j \in A) \\ &=\, \mu\left(\{x\}\right) \mathbb{E}_x\left[\sum_{j=1}^{T_x} \mathbf{1}_A(X_j)\right] \,=\, \mu\left(\{x\}\right) \lambda_x(A)\,. \end{aligned} \quad (2.16)$$

We are now going to exploit accessibility of x (and invariance of μ, by hypothesis, and of λ_x, by point (1)) to transform inequality (2.15) into an equality: note that

$$\mu(\{x\}) \overset{\mu\mathtt{p}=\mu}{=} \int_E \mu(\,\mathrm{d}y)\mathtt{p}_\star(y,\{x\}) \overset{(2.15)}{\geq} \mu(\{x\}) \int_E \lambda_x(\,\mathrm{d}y)\mathtt{p}_\star(y,\{x\}) \overset{\lambda_x\mathtt{p}=\lambda_x}{=} \mu(\{x\})\,, \quad (2.17)$$

where in the last step we used also $\lambda_x(\{x\})=1$. This implies that $\mu=\mu(\{x\})\lambda_x$. Suppose in fact that $\mu \neq \mu(\{x\})\lambda_x$, by (2.15) we know that $\nu := \mu - \mu(\{x\})\lambda_x$ is a (positive) measure for which $\int_E h(y)\nu(\,\mathrm{d}y)=0$, with $h(y) := \mathtt{p}_\star(y,\{x\}) > 0$ for every y, which implies $\nu(h \geq 1/n)=0$ for every n, hence $\nu(E)=0$, in contradiction with the assumption $\mu \neq \mu(\{x\})\lambda_x$. This completes the proof of point (3).

For point (4) we observe that, since x is recurrent then λ_x is $\mathtt{p}$-invariant (by point (1)) and we have essential uniqueness (by point (3)). We remark that $\lambda_x(E) = \mathbb{E}_x[T_x]$. So if $\mathbb{E}_x[T_x] < \infty$ then $\lambda_x/\mathbb{E}_x[T_x]$ is the unique $\mathtt{p}$-invariant probability. On the other hand if $\lambda_x(E)=\mathbb{E}_x[T_x]=\infty$, any other $\mathtt{p}$-invariant measure is proportional to λ_x, hence there does not exist a $\mathtt{p}$-invariant probability. $\square$

We know that the existence of a invariant measure μ with $\mu(E)=\infty$ and $\mu(\{x\})>0$ does not imply that x is recurrent (just think of a transient irreducible MC on $\mathbb{Z}$, like the simple random walk with non zero expectation of the increment: see Section 7.2). On the other hand the following result is at time useful:

Proposition 2.2. *If μ is a $\mathtt{p}$-invariant probability with $\mu(\{x\})>0$, then x is recurrent.*

Proof. By the Strong Markov Property we know that $\mathbb{E}_y[N_x] \leq \mathbb{E}[N_x]$ for every $y \in E$. Therefore, since μ is a probability, we have

$$\int_E \mu(\,\mathrm{d}y)\mathbb{E}_y[N_x] \,=\, \mathbb{E}_\mu[N_x] \,\leq\, \mathbb{E}_x[N_x]\,. \quad (2.18)$$

But $\mathbb{E}_\mu[N_x] = \sum_{n=0}^\infty \mathbb{P}_\mu(X_n = x) = \sum_{n=0}^\infty \mu(\{x\}) = \infty$. Therefore x is recurrent. □

3. Excursions based on a recurrent state and Ratio Limit Theorems

In this section we assume that x is recurrent for $\mathtt{p}$. We recall that $T_x^{(j)}$ is the time of j^{th} return to x, so and $T_x^{(1)} = T_x$. We put by definition $T_x^{(0)} := 0$. A very efficient tool is the decomposition of a trajectory of the MC into *excursions* from x to x: an excursion from x to x is a trajectory of variable length that starts with x and ends with x, and no visit to x in the middle. We can see an excursion as a random variable taking values in a suitable probability space, but we will take a more implicit approach: a (measurable) function of an excursion in fact is just a $\mathcal{F}_{T_x}$- measurable random variable.

The key point is that excursions are independent, and even IID if we work with $\mathbb{P}_x$. The key result to establish this is the following consequence of the Strong Markov Property:

Proposition 3.1. *If x is a recurrent state for $\mathtt{p}$ and if, for every $k \in \mathbb{N}$, $Z_0, \ldots, Z_k$ are $\mathcal{F}_{T_x}$-measurable non-negative or bounded random variables, then for every choice of X_0 such that $\mathbb{P}(T_x < \infty) = 1$*

$$\mathbb{E}\left[\prod_{j=0}^k Z_j \circ \theta_{T_x^{(j)}}\right] = \mathbb{E}[Z_0] \prod_{j=1}^k \mathbb{E}_x[Z_j] . \tag{2.19}$$

Proof. We proceed by induction. For $k = 1$ by the Strong Markov Property

$$\mathbb{E}[Z_0 (Z_1 \circ \theta_{T_x})] = \mathbb{E}\left[Z_0 \mathbb{E}_{X_{T_x}}[Z_1]\right] , \tag{2.20}$$

and, since $\mathbb{P}(T_x < \infty) = 1$, we replace X_{T_x} by x and (2.19) follows for $k = 1$.

If (2.19) holds for $k \geq 1$ and any $\mathbb{P}$ such that $\mathbb{P}(T_x < \infty) = 1$, then using that $\theta_{T_x^{(j)}} = \theta_{T_x^{(j-1)}} \circ \theta_{T_x}$ on $\{T_x^{(j)} < \infty\}$ and that $\mathbb{P}(T_x^{(j)} < \infty) = 1$ we see that

$$\begin{aligned}\mathbb{E}\left[\prod_{j=0}^{k+1} Z_j \circ \theta_{T_x^{(j)}}\right] &= \mathbb{E}\left[Z_0 \left(\left(\prod_{j=1}^{k+1} Z_j \circ \theta_{T_x^{(j-1)}}\right) \circ \theta_{T_x}\right)\right] \\ &= \mathbb{E}[Z_0]\, \mathbb{E}_x\left[\prod_{j=1}^{k+1} Z_j \circ \theta_{T_x^{(j-1)}}\right] = \mathbb{E}[Z_0] \prod_{j=1}^{k+1} \mathbb{E}_x[Z_j] ,\end{aligned} \tag{2.21}$$

where in the second step we used the Strong Markov Property and in the last step we used the induction hypothesis for $\mathbb{P} = \mathbb{P}_x$. □

Here is a direct consequence of Proposition 3.1. For $f : E \to \mathbb{R}$ we set

$$Z_0(x, f) := \sum_{k=1}^{T_x} f(X_k) , \tag{2.22}$$

and for $j \in \mathbb{N}$

$$Z_j(x,f) \, := \, Z_0(x,f) \circ \theta_{T^{(j)}} \, = \sum_{k=T^{(j)}+1}^{T^{(j+1)}} f(X_k)\,. \tag{2.23}$$

Corollary 3.2. *Under the hypotheses of Proposition 3.1, for every f measurable we have that $(Z_j(x,f))_{j=0,1,\ldots}$ is a sequence of independent random variables and $(Z_j(x,f))_{j=1,2,\ldots}$ is an IID sequence. In particular, if $\mathbb{P} = \mathbb{P}_x$, then $(Z_j(x,f))_{j=0,1,\ldots}$ is IID.*

The major consequence of Corollary 3.2 (and of the Law of Large Numbers for IID sequences is the following result that makes invariant measures very relevant for the asymptotic behavior of MCs with one recurrent state (and not only when the invariant measure can be normalized):

Theorem 3.3 (Ratio Limit Theorem). *Let $\mathtt{p}$ be a Markov kernel admitting an accessible recurrent state and call λ the (essentially unique) invariant measure. Then for every choice of the law of X_0 such that $\mathbb{P}(T_x < \infty) = 1$ we have that for every f, g such that $\int_E |f| \, \mathrm{d}\lambda < \infty$, $\int_E |g| \, \mathrm{d}\lambda < \infty$ and $\int_E g \, \mathrm{d}\lambda \neq 0$*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n g(X_k)} \, = \, \frac{\int_E f \, \mathrm{d}\lambda}{\int_E g \, \mathrm{d}\lambda} \right) \, = \, 1\,. \tag{2.24}$$

Proof. By Theorem 2.1 (1) and (3) we know that $\lambda = \lambda(\{x\})\lambda_x$ so we can replace λ with λ_x. Then Theorem 3.3 is a direct consequence of

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n \mathbf{1}_{\{x\}}(X_k)} \, = \, \int_E f \, \mathrm{d}\lambda_x \right) \, = \, 1\,, \tag{2.25}$$

for f non negative and measurable.

Let us first show (2.25) with $\mathbb{P} = \mathbb{P}_x$. In this case $(Z_j(x,f))_{j\in\mathbb{N}}$ is an IID sequence (Corollary 3.2). Note that $\mathbb{E}_x[Z_j(x,f)] = \mathbb{E}_x[\sum_{k=1}^{T_x} f(X_k)] = \int_E f \, \mathrm{d}\lambda_x$ for every j. Moreover

$$\sum_{k=1}^{T_x^{(n)}} f(X_k) \, = \, \sum_{j=0}^{n-1} Z_j(x,f)\,, \tag{2.26}$$

so Kolmogorov Law of Large Numbers yields that

$$\mathbb{P}_x\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{T_x(n)} f(X_k) \, = \, \int_E f \, \mathrm{d}\lambda_x \right) \, = \, 1\,. \tag{2.27}$$

This of course remains true if n is replaced by any $\mathbb{N}$ valued sequence (even random) (L_n) such that $\mathbb{P}_x(\lim_n L_n = \infty) = 1$. We choose

$$L_n \, = \, \sum_{k=1}^n \mathbf{1}_{\{x\}}(X_k)\,, \tag{2.28}$$

and $\lim_n L_n = \infty$ λ_x-a.s. because x is recurrent. Note that $T_x^{(L_n)} \le n < T_x^{(L_n+1)}$. Next we remark that

$$\frac{\sum_{k=1}^{T_x^{(L_n)}} f(X_k)}{L_n} \le \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} \mathbf{1}_{\{x\}}(X_k)} \le \frac{L_{n+1}}{L_n} \frac{\sum_{k=1}^{T_x(L_n+1)} f(X_k)}{L_n+1}. \tag{2.29}$$

By taking $n \to \infty$ we obtain (2.25) with $\mathbb{P} = \mathbb{P}_x$.

The general case is a straightforward exercise that uses the Strong Markov Property, or even directly the Law of Large Numbers when the first random variable in the sequence has a different law with respect to the other variables in the sequence. Of course we need that this first variable is a.s. finite, for which we use that $\mathbb{P}(T_x < \infty) = 1$. □

A direct corollary to the Ratio Limit Theorem in the positive recurrent case is the a.s. Ergodic Theorem for MC (of course, the a.s. Ergodic Theorem, also known as Birkhoff Ergodic Theorem, holds in much greater generality):

Corollary 3.4 (Almost Sure Ergodic Theorem for Markov Chains). *Let* $\mathtt{p}$ *be a Markov kernel admitting an accessible recurrent state and assume that* x *is positive recurrent (i.e.* $\lambda_x(E) < \infty$ *and we set* $\pi := \lambda_x/\lambda_x(E)$*). Then for every choice of the law of* X_0 *such that* $\mathbb{P}(T_x < \infty) = 1$ *and for every* f *measurable and positive or such that* $\int_E f \,\mathrm{d}\pi < \infty$ *we have*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \int_E f \,\mathrm{d}\pi \right) = 1. \tag{2.30}$$

Proof. Apply Theorem 3.3 with $g \equiv 1$. □

Also in he null recurrent case we have an interesting consequence of the Ratio Limit Theorem:

Corollary 3.5. *Let* $\mathtt{p}$ *be a Markov kernel admitting an accessible recurrent state and assume that* x *is null recurrent (i.e.* $\lambda_x(E) = \infty$*). Then if* f *is such that* $\int_E |f| \,\mathrm{d}\lambda_x < \infty$ *we have*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = 0 \right) = 1. \tag{2.31}$$

Proof. In this case one is tempted to choose again $g \equiv 1$ in Theorem 3.3. But in this case $\int_E g \,\mathrm{d}\lambda_x = \infty$. To get around this problem we use the fact that λ_x is σ-finite so for every $\varepsilon > 0$ we can find an event F_ε such that $1/\varepsilon \le \lambda_x(F_\varepsilon) < \infty$. Therefore Theorem 3.3 with $g = \mathbf{1}_{F_\varepsilon}$ yields that

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} g(X_k)} = \frac{\int_E f \,\mathrm{d}\lambda_x}{\int_E g \,\mathrm{d}\lambda_x} \right) = 1. \tag{2.32}$$

Since $|\int_E f \,\mathrm{d}\lambda_x| / \int_E g \,\mathrm{d}\lambda_x \leq \varepsilon \int_E |f| \,\mathrm{d}\lambda_x$ we conclude. □

4. The Ergodic Theorem

The Almost Sure Ergodic Theorem for Markov Chains (Corollary 3.4) directly yields that, in presence of a finite recurrent accessible state x, for every initial condition and every $A \in \mathcal{E}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}(X_k \in A) \,=\, \pi(A)\,. \tag{2.33}$$

This result cannot be improved to $\lim_{n\to\infty} \mathbb{P}(X_n \in A) = \pi(A)$ in full generality: consider in fact that irreducible MC with $E = \{1,2\}$ and $Q(1,2) = Q(2,1) = 1$ (hence $Q(1,1) = Q(2,2) = 0$). In this case the only $\mathbb{P}_1(X_n = 1) = Q^n(1,1) = 0$ if n is odd and $Q^n(1,1) = 2$ if n is even. So $(\mathbb{P}_1(X_n = 1))_{n\in\mathbb{N}}$ does not converge.

In fact the only problem that we have to watch out for is the *periodic behavior* that is evident in the simple example we just developed.

Given $\mathtt{p}$, we say that the period t_x of a state x is the Greatest Common Divisor (GCD) of $E_x := \{n = 1, 2, \ldots :\, \mathtt{p}_n(x, \{x\}\}$, with $GCD(\emptyset) = \infty$. If $t_x = 1$ then we say that x is aperiodic. There are two relevant facts about periods that we collect in the next statement:

PROPOSITION 4.1. *Choose a Markov kernel* $\mathtt{p}$. *We have that*

(1) if x is accessible, then there exists n_0 such that $\mathtt{p}_{nt_x}(x, \{x\}) > 0$ for every $n > n_0$;

(2) if x and y are accessible, then $t_x = t_y$.

PROOF. ADD □

The Ergodic Theorem is stated in terms of the *total variation* between probability measures: if μ and ν are two probabilities on $(E, \mathcal{E})$

$$\mathrm{d}_{\mathtt{TV}}(\mu, \nu) \,:=\, \sup_{A\in\mathcal{E}} \left(\mu(A) - \nu(A)\right)\,. \tag{2.34}$$

Various facts about the total variation distance are discussed in Secton 7 Here we point out that convergence in total variation is stronger that weak convergence of probabilities (but the two types of convergence coincide if E is countable).

THEOREM 4.2. *Choose a Markov kernel* $\mathtt{p}$ *that admits an accessible, aperiodic, positive recurrent state. Denote by* π *the unique* $\mathtt{p}$*-invariant probability. Then for every probability* μ *on* $(E, \mathcal{E})$ *such that* $\mathbb{P}_\mu(T_x < \infty) = 1$ *we have*

$$\lim_{n\to\infty} \mathrm{d}_{\mathtt{TV}}(\mu \mathtt{p}_n, \nu) \,=\, 0\,. \tag{2.35}$$

PROOF. Let us start by remarking that

$$\begin{aligned} \mathrm{d}_{\mathtt{TV}}(\mu \mathtt{p}_n, \pi) \,=\, \sup_A \left(\mathbb{P}_\mu(X_n \in A) - \pi(A)\right) \,=\, \sup_A \left|\mathbb{P}_\mu(X_n \in A) - \pi(A)\right| \\ \le \int_E \sup_A \left|\mathbb{P}_y(X_n \in A) - \pi(A)\right| \mu(\,\mathrm{d}y)\,, \end{aligned} \tag{2.36}$$

So, by Dominated Convergence, it suffices to show that

$$\lim_n \sup_A \left|\mathbb{P}_y(X_n \in A) - \pi(A)\right| \,=\, 0\,, \tag{2.37}$$

$\mu(\,\mathrm{d}y)$-a.s..

We give a proof of (2.37) based on *independent coupling*, in the sense that we consider the process $(X_n^{(1)}, X_n^{(2)})_{n=0,1,\ldots}$ with $(X^{(1)})_{n=0,1,\ldots}$ a $\mathtt{p}$-MC with $X^{(1)})_0 = y$ and $(X^{(2)})_{n=0,1,\ldots}$ a $\mathtt{p}$-MC with $X^{(2)})_0$ with law π. Moreover $(X^{(1)})_{n=0,1,\ldots}$ and $(X^{(2)})_{n=0,1,\ldots}$ are independent.

In fact, $(X_n^{(1)}, X_n^{(2)})_{n=0,1,\ldots}$ is a MC with state space $E \times E$ equipped with the standard product σ-algebra: we leave the proof of this fact to the reader, but we point out that the kernel $\mathtt{p}^{(2)}$ of this new Markov chain from the state $(y,z) \in E \times E$ to an event of the form $A \times B$ is $\mathtt{p}^{(2)}((y,z), A \times B) = \mathtt{p}(y,A)\mathtt{p}(z,B)$. Let us remark that $\pi \times \pi$ is $\mathtt{p}^{(2)}$-invariant. Moreover, if x is an accessible recurrent aperiodic state for $\mathtt{p}$, then (x,x) is an accessible recurrent state for $\mathtt{p}^{(2)}$. In fact, it is accessible because for every (y,z) there exists $n_{y,z}$ such that $\mathtt{p}^{(2)}_{n_{y,z}+n}((y,z), \{x\} \times \{x\}) = \mathtt{p}_{n_{y,z}+n}(y, \{x\})\mathtt{p}_{n_{y,z}+n}(z, \{x\}) > 0$ for every $n \ge 0$: this follows from the accessibility of x and aperiodicity (this is the only, but crucial, point in which aperiodicity is used!), see Proposition 4.1. We now use that $\pi \times \pi$ is an invariant probability, so Proposition 2.2 tells us that x is (positive) recurrent and Theorem 2.1, parts (1) and (3), imply that $\pi \times \pi$ is the unique $\mathtt{p}^{(2)}$-invariant measure. And part (4) of the same theorem implies that $\mathbb{E}_{(x,x)}[T_{(x,x)}] < \infty$ (this is not really used in the proof). We use the simplified notation T for the stopping time $T^\flat_{(x,x)}$ (we work with the natural filtration of $(X_n^{(1)}, X_n^{(2)})_{n=0,1,\ldots}$):

$$T \,:=\, \inf\left\{n = 0, 1, \ldots :\, X_n^{(1)} = X_n^{(2)} = x\right\}\,. \tag{2.38}$$

We have

$$\begin{aligned} \left|\mathbb{P}_y(X_n \in A) - \pi(A)\right| \,&=\, \left|\mathbb{E}_{\delta_y \times \pi}\left[\mathbf{1}_A\left(X_n^{(1)}\right) - \mathbf{1}_A\left(X_n^{(2)}\right)\right]\right| \\ &=\, \left|\mathbb{E}_{\delta_y \times \pi}\left[\mathbf{1}_A\left(X_n^{(1)}\right) - \mathbf{1}_A\left(X_n^{(2)}\right); T > n\right]\right| \\ &\le\, \mathbb{P}_{\delta_y \times \pi}\left(T > n\right)\,, \end{aligned} \tag{2.39}$$

where in the second step we have used that

$$\mathbb{E}_{\delta_y \times \pi}\left[\mathbf{1}_A\left(X_n^{(1)}\right) - \mathbf{1}_A\left(X_n^{(2)}\right); T \le n\right] \,=\, 0\,, \tag{2.40}$$

which is a consequence of the Strong Markov Property that implies that for $j \le n$ and on the event $\{T = j\}$ we have

$$\mathbb{E}_{\delta_y \times \pi}\left[\mathbf{1}_A\left(X_n^{(1)}\right) - \mathbf{1}_A\left(X_n^{(2)}\right) \Big| \mathcal{F}_T\right] \,=\, \mathtt{p}_{n-j}(x,A) - \mathtt{p}_{n-j}(x,A) \,=\, 0\,. \tag{2.41}$$

So we are left with showing that $\lim_n \mathbb{P}_{\delta_y \times \pi}\left(T > n\right) = 0$, $\mu(\,\mathrm{d}y)$-a.s..

REMARK 4.3. *It is rather intuitive (and we claim) that* $\mathbb{P}_\pi(T_x < \infty) = 1$. *Here is a proof: call* $C := \{y : \mathbb{P}_y(T_x < \infty) = 1\}$ *and* $x \in C$. *Then for* $y \in C$ *we have* $0 = \mathbb{P}_y(T_x = \infty) = \mathbb{P}_y(X_1 \neq x, T_x \circ \theta_1 = \infty) \geq \mathbb{P}_y(X_1 \notin C, T_x \circ \theta_1 = \infty) = \mathbb{E}_y[\mathbb{P}_{X_1}(T_x = \infty); X_1 \notin C]$. *Since* $\mathbb{P}_{X_1}(T_x = \infty) > 0$ *for* $X_1 \notin C$ *we see that* $\mathbb{P}_y(X_1 \notin C) = 0$. *This means that* C *is* closed *(the process cannot get out of it). Hence*

$$\lambda_x(C) = \mathbb{E}_x[\sum_{k=1}^{T_x} \mathbf{1}_C(X_k)] = \mathbb{E}_x[T_x]\,, \tag{2.42}$$

which completes the proof of the claim.

In view of Remark 4.3, it suffices to show that $\lim_n \mathbb{P}_{(y,z)}(T > n) = 0$, $\mu \times \pi(\mathrm{d}(y,z))$-a.s., i.e. $\mathbb{P}_{(y,z)}(T < \infty) = 1$, $\mu \times \pi(\mathrm{d}(y,z))$-a.s..

REMARK 4.4. *If* E *is countable and given the assumption of accessibility of* x, *one sees that the support of* π, *i.e.* $E_0 := \{x : \pi(x) > 0\}$, *is closed fo the Markov chain (and* $E \setminus E_0$ *is transient). Hence it suffices to show that* $\mathbb{P}_{(y,z)}(T < \infty) = 1$ *for every* $y, z \in E_0$. *Moreover the* $\mathtt{p}$*-MC is irreducible on this set (Proposition 5.4). Moreover aperiodicity implies that* $\mathtt{p}_n(x, \{x\}) > 0$ *for* n *sufficiently large (in fact, for every* $x \in E_0$*). Therefore* $\mathtt{p}_n(y, \{x\})\mathtt{p}_n(z, \{x\}) > 0$ *for* n *sufficiently large too, for every* y *and* z *in* E_0. *This means that* $\mathbb{P}_{(y,z)}(T < \infty) = 1$ *for every* $y, z \in E_0$ *and the proof of Theorem 4.2 is complete. If* E *is not countable this argument is no longer available and we attack the problem using renewal theory.*

The last step of the proof that can be restated in terms of renewals (and intersection renewals).

For this we introduce the sequence $(\tau_j^{(1)})_{j=1,2,\dots}$ of the times of successive visits to x by $(X_n^{(1)})_{n=0,1,\dots}$: so $\tau_j^{(1)} = T_x^{(j)}$ for the MC $(X_n^{(1)})$. In the same way we call $(\tau_j^{(2)})_{j=1,2,\dots}$ of the times of successive visits to x by $(X_n^{(2)})_{n=0,1,\dots}$. Note that this are a.s. infinite sequences of finite numbers because we have assumed $\mathbb{P}_\mu(T_x < \infty) = 1$ and we have proven $\mathbb{P}_\pi(T_x < \infty) = 1$. Moreover, by the Strong Markov Property, we know that

$$\left(\tau_{j+1}^{(1)} - \tau_j^{(1)}\right)_{j=1,2,\dots} \quad \text{and} \quad \left(\tau_{j+1}^{(2)} - \tau_j^{(2)}\right)_{j=1,2,\dots}\,, \tag{2.43}$$

are two IID sequences. More than that, these two sequence are equal in law because they are just IID sequences of random variables that have the same law as T_x for the $\mathtt{p}$-MC with $X_0 = x$. By the hypothesis of positive recurrence we have that $\mathbb{E}_x[T_x] < \infty$. Note that this means that $(\tau_j^{(1)})_{j=1,2,\dots}$ and $(\tau_j^{(2)})_{j=1,2,\dots}$ are two independent random walks with positive increments that are in $\mathbb{L}^1$. In different language, $(\tau_j^{(1)})$ and $(\tau_j^{(2)})$ are two independent delayed positive persistent renewals with the same inter-arrival law. Moreover, they are aperiodic because the original Markov chain is aperiodic, see Proposition 6.3.

Therefore $\mathbb{P}_{(y,z)}(T < \infty) = 1$ is a direct consequence of Proposition 6.5 and the proof of Theorem 4.2 is complete. $\square$

5. The Lindley process

The Lindley MC has been introduced in Sec. 1.1, to which we refer for the notations (ξ_n), (X_n) and (S_n).

An extremely precise understanding of the Lindley MC has been developed (see for example [**1**] and references therein) and we will just touch a few aspects of all this activity. Let us start by observing that, by construction, $X_n \geq S_n$ for every n. So, if $\lim_n S_n = \infty$ a.s., this happens in particular when $\xi_1 \in \mathbb{L}^1$ and $\mathbb{E}[\xi_1] > 0$ by the Law of Large Numbers, then $X_n \to \infty$ and $N_0 < \infty$ a.s.: so 0 is transient (recall that we assume that $\mathbb{P}(\xi_1 < 0) > 0$, so 0 is accessible).

There is an interesting direct link between the Lindley process with $W_0 = 0$ and the random walk (S_n) with $S_0 = 0$. In fact, it is straightforward to see that the return times $(T_0^{(j)})_{j=0,1,\ldots}$ of 0 by (W_n) are the descending ladder times of (S_n): the descending ladder times are defined by setting $\tau_0 = 0$ and, for $k \geq 0$, $\tau_{k+1} := \inf\{n > \tau_k : S_n \leq S_{\tau_k}\}$. So the ladder times are the times in which the walk hits a new minimum. The identity we just claimed is actually *pathwise*: that is $T_0^{(j)}(\omega) = \tau_j(\omega)$ for every ω.

One can actually show that (S_n) has only three possible behaviors [**1**, p. 224, Th. 2.4]:

(1) either $S_n \to \infty$ a.s.;
(2) or $S_n \to -\infty$ a.s.;
(3) or $\limsup_n S_n = +\infty$ and $\liminf_n S_n = -\infty$ a.s.;

and $\mathbb{P}(\tau_1 < \infty) = 1$ if and only if we are in cases (2) or (3). Therefore ((2) or (3)) is a necessary and sufficient condition for recurrence of the Lindley process. One can actually show that (W_n) is null recurrent if and only if (S_n) is in case (3), and this happens if and only if $\xi \in \mathbb{L}^1$ and $\mathbb{E}\xi_1 = 0$ or $\xi \sim -\xi$. In Proposition 5.2 we will show a part of these results: namely, that $S_n \to -\infty$ implies that (W_n) is positive recurrent.

We are going to develop in detail for the Lindley MC a certain remarquable identity (and we will develop some consequences):

PROPOSITION 5.1. *If $h_\xi(x) = (x + \xi)_+$ we have that for every $x \geq 0$*

$$h_{\xi_1} \circ h_{\xi_2} \circ \ldots \circ h_{\xi_n}(x) \;=\; \max(0, \xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \ldots, \xi_{n-1}, \xi_1 + \ldots, \xi_n + x)\,. \tag{2.44}$$

Proposition 5.1 acquires its interest if we recall (1.7)

$$X_n \;=\; h_{\xi_n} \circ h_{\xi_{n-1}} \circ \ldots \circ h_{\xi_1}(x)\,, \tag{2.45}$$

and for the the Lindley MC $h_\xi(x) = (x + \xi)_+$, so (X_n) is a Lindley MC with $X_0 = x$. And even if X_n does not coincide at all with the quantity in (2.44), nevertheless these two random variables have the same law

$$X_n \;\sim\; Y_n \;:=\; h_{\xi_1} \circ h_{\xi_2} \circ \ldots \circ h_{\xi_n}(x)\,, \tag{2.46}$$

because $(\xi_1, \xi_2, \ldots, \xi_n) \sim (\xi_n, \xi_{n-1}, \ldots, \xi_1)$.

Let us however stress that $(X_n) \not\sim (Y_n)$: they are not at all the same process. In particular, if $x = 0$ then (Y_n) is non decreasing, while of course (X_n) oscillates up and down! See Fig. 1.

Proof of Proposition 5.1. Let us deal the case $x = 0$ (the generalization is immediate). The proof follows by induction because $h_{\xi_1}(0) = (\xi_1)_+$ which is the first step of the induction. The iterative step uses that for every $x \geq 0$ and $y \in \mathbb{R}$

$$(x + y_+)_+ \;=\; \max\left(0, x, x+y\right) \,, \tag{2.47}$$

which can be established by considering separately the cases $y \geq 0$ and $y < 0$. □

Here is a remarquable consequence of Proposition 5.1.

Proposition 5.2. *If $S_n \to -\infty$ a.s. then the Lindley MC (X_n) converges in total variation toward a limit law that is the unique invariant probability and the process is positive recurrent.*

Proof. If $S_n \to -\infty$ then for every x we have $\lim_n Y_n := Y_\infty < \infty$ exists! And

$$Y_\infty \;:=\; \max\left(0, \xi_1, \xi_1 + \xi_2, \ldots\right) \,. \tag{2.48}$$

Note that Y_∞ does not depend on x.

Since $X_n \sim Y_n$ we readily obtain that, for every $X_0 = x \geq 0$, (X_n) converges in law to Y_∞. Let us show that the law of Y_∞ is an invariant probability for (X_n): for this we introduce a variable ξ_0 such that $(\xi_j)_{j=0,1,\ldots}$ is IID. We have $\lim_n h_{\xi_0}(Y_n) = h_{\xi_0}(Y_\infty)$. But $h_{\xi_0}(Y_n) \sim h_{\xi_{n+1}}(X_n) = X_{n+1}$, which converges in law to Y_∞. This means that $h_{\xi_0}(Y_\infty) \sim Y_\infty$, hence the law of Y_∞ is an invariant probability.

It is straightforward to see that $\mathbb{P}(Y_\infty = 0) > 0$ by exploiting that its law is invariant and that $\mathbb{P}(\xi_1 < 0) > 0$. Therefore by Proposition 2.2 we have that 0 is recurrent and, by Theorem 2.1, we obtain uniqueness of the invariant measure that is a probability.

We are left with improving the convergence in law to total variation. For this we apply Theorem 4.2: the only property that we are left to verify is aperiodicity. But $\mathtt{p}(0, \{0\}) = \mathbb{P}(\xi_1 \leq 0) > 0$ hance 0 is aperiodic and the proof is complete. □

In the previous proof we exploited the general theory to infer uniqueness of the invariant probability. We could have extracted uniqueness (among probability measures) from the fact that we know the weak convergence of the process and that the limit does not depend on the initial condition.

Proposition 5.3. *(X_n) is a $(\mathtt{p}, E)$-MC and if (X_n) converges in law to X_∞ for every initial condition $X_0 = x \in E$. If there exists an invariant probability π, then $X_\infty \sim \pi$ (hence π is the unique invariant probability).*

Proof. The hypotheses can be restated by saying that the n-step Markov kernel $\mathtt{p}_n(x,\cdot)$ computed at x, converges weakly to the law ν of X_∞ and this holds for every $x \in E$. Therefore for every $h \in C_b^0$ and every n (and passing to $n \to \infty$ in the last step we have

$$\begin{aligned}\int_E h(y)\pi(\,\mathrm{d}y) \,=\, \int_E h(y)\pi\mathtt{p}_n(\,\mathrm{d}y) \,=\, \int_E h(y)\int_E \pi(\,\mathrm{d}x)\mathtt{p}_n(x,\,\mathrm{d}y) \\ = \int_E \left(\int_E h(y)\mathtt{p}_n(x,\,\mathrm{d}y)\right)\pi(\,\mathrm{d}x) \longrightarrow \int_E h(y)\nu(\,\mathrm{d}y)\,, \quad (2.49)\end{aligned}$$

hence $\pi = \nu$. □

6. Back to Markov chains with countable state space

6.1. Random walks on $\mathbb{Z}$. We focus on the MC defined by the iteration $X_{n+1} = X_n + \xi_{n+1}$ with $X_0 \in \mathbb{Z}$ and the IID sequence (ξ_n) is make of random variables of law P_ξ, and of course $P_\xi(\mathbb{Z}) = 1$, so we write $P_\xi(x)$ for $P_\xi(\{x\})$. We assume that ξ is not trivial, that is $P_\xi(x) < 1$ for every x. Note that for this MC $\mathtt{p}(x,\{y\}) = Q(x,y) = P_\xi(y-x)$. This is particular implies that $U(x,y)$ just depends on $y-x$, so either all states are recurrent, or they are all transient. Moreover it implies also that the uniform measure μ on $\mathbb{Z}$ ($\mu(x) = 1$ for every $x \in \mathbb{Z}$) is invariant.

The issue of essential uniqueness of this invariant measure, at least in the case in which x is positive recurrent, depends on whether or not x is accessible. Note that if ξ takes values only on even (respectively, odd) sites, then even sites communicate only with even (respectively, odd) sites. So, in general, $E = \mathbb{Z}$ can be decomposed into equivalence classes and the uniform measure over the equivalence class is going to be invariant. And this invariant measure is going to be the (essentially) unique invariant measure that is supported on a given equivalence class. In any case, we can conclude that that no state is positive recurrent for for random walks on $\mathbb{Z}$.

Let us look more carefully at the irreducibility issue. We have the following result (that can be easily generalized also to dimension larger than one):

Proposition 6.1. *A random walk on $\mathbb{Z}$ with increment variable ξ for which a state (hence all) is recurrent is irreducible if and only if the subgroup generated by the support of the law of ξ is $\mathbb{Z}$.*

Proof. Since $Q(x,y) = P_\xi(y-x)$ for every x and y, it suffices to consider the cas $X_0 = 0$. Let us call G subgroup generated by the support of P_ξ. It is straightforward that $\mathbb{P}_0(X_n \in G$ for every $n) = 1$. Therefore if $G \neq \mathbb{Z}$, the MC is not irreducible. On the other hand, if $G = \mathbb{Z}$ then we consider the set $H_0 := \{y : U(0,y) > 0\}$ and we remark that H_0 is a subgroup of $\mathbb{Z}$. In fact

(1) $0 \in H_0$ because $U(0,0) > 0$ by definition;
(2) if $x, y \in H_0$ then there exist n_x and n_y such that $Q^{n_x}(0,x) > 0$ and $Q^{n_y}(0,y) > 0$, hence

$$Q^{n_x+n_y}(0,x+y) \geq Q^{n_x}(0,x)Q^{n_y}(x,x+y) = Q^{n_x}(0,x)Q^{n_y}(0,y) > 0\,, \qquad (2.50)$$

so $x+y \in\in H_0$;

(3) $x \in H_0$ means $U(0,x) > 0$, which implies (Proposition 5.4) $U(x,0) > 0$ because 0 (or x) is recurrent. But in this case $U(x,0) = U(0,-x)$, hence $-x \in H_0$.

Since we have assumed that $G = \mathbb{Z}$, we have that $H_0 = \mathbb{Z}$, so any state is accessible from 0 and, since 0 is recurrent (again: Proposition 5.4) we obtain that all states communicate. □

PROPOSITION 6.2. *Consider a random walk on $\mathbb{Z}$ with increment variable ξ such that $\xi \in \mathbb{L}^1$: $x \in \mathbb{Z}$ is recurrent if and only if $\mathbb{E}[\xi] = 0$.*

The *only if* part of Proposition 6.2 is straightforward and holds also in higher dimension. The *if* part is much less trivial and $d = 1$ is crucially used.

PROOF. The *only if* part is a direct consequence of Kolmogorov Law of Large Numbers. Let us focus on the *if* part and let us choose $x = 0$ without loss of generality.

We proceed by contradiction and assume that $U(0,0) < \infty$. We start by recalling that $U(0,x) \le U(x,x) = U(0,0)$. Therefore $\sum_{x=-n}^{n} U(0,x) \le (2n+1)U(0,0)$, i.e. for every $n \in \mathbb{N}$ we have

$$\sum_{x=-n}^{n} U(0,x) \le C \quad \text{with } C := 3\,U(0,0)\,. \tag{2.51}$$

On the other hand $\lim_n X_n/n = 0$ a.s. by the Law of Large Numbers. Hence for every $\varepsilon > 0$ we can find n_ε such that for $n \ge n_\varepsilon$ we have

$$\mathbb{P}\left(|X_n| \le \varepsilon n\right) = \sum_{x:\,|x|\le \varepsilon n} Q^n(0,x) > \frac{1}{2}\,. \tag{2.52}$$

By elementary arguments we have therefore that for every choice of $n_1 \ge m \ge n_0$ we have

$$\sum_{x:\,|x|\le \varepsilon n_1} Q^m(0,x) \ge \sum_{x:\,|x|\le \varepsilon m} Q^m(0,x) > \frac{1}{2}\,. \tag{2.53}$$

By summing over $m = n_0, n_0+1, \ldots, n_1$ we arrive at

$$\sum_{m=n_0}^{n_1} \sum_{x:\,|x|\le \varepsilon n_1} Q^m(0,x) > \frac{n_1 - n_0}{2}\,, \tag{2.54}$$

which directly entails that $\sum_{|x|\le \varepsilon n_1} U(0,x) > (n_1 - n_0)/2 \sim n_1/2$ for $n_1 \to \infty$. Let us make now the definite choice of $\varepsilon := 1/(3C)$: by (2.51) we have that $\sum_{|x|\le \varepsilon n_1} U(0,x) \le C\varepsilon n_1 = n_1/3$. Therefore for n_1 sufficiently large a contradiction emerges and this denies the assumption $U(0,0) < \infty$. Therefore $U(0,0) = \infty$ and the proof is complete. □

6.2. Back to discrete renewals. We pick up again the analysis from Section 7.4 of Chapter 1. Let us first show that the following result that is an exercise in establishing aperiodicity of a chain that has a direct applications to the remainder of this subsection.

PROPOSITION 6.3. *Assume that the backward recurrence time MC (A_n) is irreducible. (A_n) is aperiodic if and only if the support of ξ is aperiodic (i.e. if $\{n \in \mathbb{N} : \mathbb{P}(\xi = n) > 0\} \subset d\,\mathbb{N}$ only for $d = 1$).*

PROOF. We observe that $E_0 := \{n : Q^n(0,0) > 0\} = \{n : \mathbb{P}(\sum_{j=1}^n \xi_j) > 0\} = \{\sum_{j=1}^k x_j : k \in \mathbb{N}$ and x_j belongs to the support of ξ for every $j\}$. Moreover E_0 is clearly stable under addition. Set $d :=$GCD(E_0): Bézout Theorem implies that there exists n_0 such that $nd \in E_0$ for every $n \geq n_0$. So if the support of ξ is aperiodic we have $d = 1$, hence 0 is aperiodic for (A_n). On the other hand if 0 is aperiodic for (A_n), $d = 1$ by definition. □

THEOREM 6.4 (Discrete Renewal Theorem). *Consider η an aperiodic renewal with inter-arrival variable ξ. Then*

$$\lim_{n\to\infty} \mathbb{P}\left(n \in \eta\right) \,=\, \frac{1}{\mathbb{E}[\xi]} \in (1, \infty]\,. \tag{2.55}$$

PROOF. We prove the result without delay (the case with delay follows directly from the result without delay). We have that $\mathbb{P}\left(n \in \eta\right) = \mathbb{P}_0(A_n = 0)$ and, by Theorem 4.2 (and Remark 4.4), if the renewal is positive persistent we have $\lim_n \mathbb{P}_0(A_n = 0) = \pi(\{0\})$, where π is the invariant probability in (1.77). So the proof is complete in the positive persistent case.

In the transient case the result is a direct consequence of the Dominated Convergence Theorem, because $\mathbf{1}_\eta(n) = 0$ for n sufficiently large.

ADD: case null recurrent.

□

We recall that the renewal set $\eta = \{n \in \mathbb{N} : A_n = 0\}$ if (A_n) is recurrent. In this case, η is an infinite set. When (A_n) is recurrent (respectively, transient) we say that the renewal is persistent (respectively, terminating). We say that η is positive persistent if (A_n) is positive recurrent, i.e. if the inter-arrival variable is in $\mathbb{L}^1$.

PROPOSITION 6.5. *Consider two independent positive persistent delayed renewals $\eta = \{\eta_0, \eta_1, \ldots\}$ ans $\eta' = \{\eta'_0, \eta'_1, \ldots\}$ that share the same inter-arrival law. We assume in addition that the inter-arrival law is aperiodic. Then $\eta \cap \eta' \neq \emptyset$ a.s..*

PROOF. It suffices to consider the case which η has no delay (i.e., $\eta_0 = 0$). Moreover since η' is a.s. an infinite set, it suffices to show that for every infinite subset $H \subset \mathbb{N}$ we have

$$\mathbb{P}(\eta \cap H \neq \emptyset) = 1\,. \tag{2.56}$$

In fact, we are going to show that $\eta \cap H$ is a.s. an infinite set.

It is practical at this stage to look at the renewal process as a random walk (S_n) with positive increments, hence living on the non negative integers: $\eta = \{S_0, S_1, \ldots\}$. The fact that η is positive persistent means that $\mathbb{E}[S_1] = \mathbb{E}[S_{j+1} - S_j] < \infty$. With standard notation $T_n = T_n^S$ is the first hitting time (notice that in this case $T_n^\flat = T_n$) and $\{n \in \eta\} = \{T_n < \infty\}$. Therefore $u(n) := \mathbb{P}(T_n < \infty)$ is the renewal function and $\lim_n u(n) = 1/\mathbb{E}[S_1] =: 2\delta > 0$ by the Discrete Renewal Theorem (Theorem 6.4). Therefore there exists k such that $u(n) \geq \delta$ for every $n \geq k$. In turn, since H is an infinite subset of $\mathbb{N} \cup \{0\}$, for every i we can choose $j \in H$ in such a way that $j \geq i + k$, so

$$\delta \leq \mathbb{P}\left(T_{j-i} < \infty\right)\,. \tag{2.57}$$

If we now consider the renewal with delay i, that is $S_0 = i$ instead of $S_0 = 0$ as we did up to now, so the law of S is denoted by $\mathbb{P}_i$ (and $\mathbb{P} = \mathbb{P}_0$), we have $\mathbb{P}\left(T_{j-i} < \infty\right) = \mathbb{P}_i\left(T_j < \infty\right)$ and, since $j \in H$, $\mathbb{P}_i\left(T_j < \infty\right) \leq \mathbb{P}_i\left(T_H < \infty\right)$. Therefore

$$\inf_{i=0,1,\ldots} \mathbb{P}_i\left(T_H < \infty\right) \geq \delta\,. \tag{2.58}$$

Therefore by Proposition 5.6 applied to the MC (S_n) (even with an arbitrary law of S_0, but for us $S_0 = 0$ suffices) with $A = \mathbb{N} \cup \{0\}$, so $N_A = \infty$ because τ is persistent, and $B = H$, we obtain that $N_H = \infty$ a.s., that is (S_n) visits H infinitely often. The proof is therefore complete. $\square$

7. Complement: the total variation distance

We have introduced the distance in total variation between probability measures in (2.34). Let us point out that

$$\mathrm{d}_{\mathtt{TV}}(\mu, \nu) = \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|\,, \tag{2.59}$$

because A can be replaced by $A^{\complement}$. From (2.59) its is clear that $\mathrm{d}_{\mathtt{TV}}(\cdot, \cdot)$ is a distance.

Here is an equivalent expression for the total variation distance:

$$\mathrm{d}_{\mathtt{TV}}(\mu, \nu) = \frac{1}{2} \sup_{f\colon \|f\|_\infty \leq 1} \left(\int_E f\,\mathrm{d}\mu - \int_E f\,\mathrm{d}\nu \right)\,. \tag{2.60}$$

That the right-hand side dominates $\mathrm{d}_{\mathtt{TV}}(\mu, \nu)$ is seen by choosing $f = \mathbf{1}_A - \mathbf{1}_{A^{\complement}} = 2\mathbf{1}_A - 1$. The other bound may be established by remarking that $\int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\nu = \int f\,\mathrm{d}(\mu - \nu)_+ - \int f\,\mathrm{d}(\mu - \nu)_-$ where we have written the signed measure $\mu - \nu$ as difference of the two positive measures $(\mu - \nu)_\pm$ with disjoint support. Call A_+, respectively A_-, the support of $(\mu - \nu)_+$, respectively of $(\mu - \nu)_-$. For $\|f\|_\infty \leq 1$

we therefore have

$$\begin{aligned}\int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\nu \,&\le\, \int \mathbf{1}_{A_+} \,\mathrm{d}(\mu-\nu)_+ - \int \mathbf{1}_{A_-} \,\mathrm{d}(\mu-\nu)_- \\ &=\, \int \left(\mathbf{1}_{A_+} - \mathbf{1}_{A_-}\right) \,\mathrm{d}(\mu-\nu) \\ &=\, \int \left(2\,\mathbf{1}_{A_+} - 1\right) \,\mathrm{d}(\mu-\nu) \,=\, 2\left(\mu\left(A_+\right) - \nu\left(A_+\right)\right) .\end{aligned} \tag{2.61}$$

Therefore (2.60) is proven.

Note that (2.60) directly implies that convergence in total variation is stronger than weak convergence. In fact (μ_n) converges to μ weakly if $\int_E f \,\mathrm{d}\mu_n \to \int_E f \mu$ for every f bounded and continuous: the convergence in total variation demands that this convergence holds for every bounded function, and uniformly in the class of functions bounded by a a fixed constant.

The next result says that if μ and ν are both absolutely continuous with respect to a measure, then the distance in total variation is $1/2$ times the $\mathbb{L}^1$ norm of the difference of the densities. This makes clear that in general convergence in total variation is *strictly stronger* than the standard (weak) convergence of probabilities. In fact, if a sequence of probabilities that admit a density converges in total variations, then the limit admits a density. So, for example, if $X_n \sim \mathcal{N}(, 1/n)$, then of course (X_n) converges in law toward 0, i.e. the law of X_n converges weakly toward δ_0. But the law of (X_n) does not converge to δ_0 in total variation.

PROPOSITION 7.1. *If the probabilities μ and ν (on $(E, \mathcal{E})$) are absolutely continuous with respect to the σ-finite measure λ, hence $\mu = f_\mu \lambda$ and $\nu = f_\nu \lambda$ with f_μ and f_ν non negative measurable functions, then*

$$\mathrm{d}_{\mathtt{TV}}(\mu, \nu) \,=\, \frac{1}{2} \| f_\mu - f_\nu \|_{\mathbb{L}^1(\lambda)} \,. \tag{2.62}$$

PROOF. By (2.60)

$$\mathrm{d}_{\mathtt{TV}}(\mu, \nu) \,=\, \frac{1}{2} \sup_{f :\, \|f\|_\infty \le 1} \left(\int_E f\, f_\mu \,\mathrm{d}\lambda - \int_E f\, f_\nu \,\mathrm{d}\lambda \right) \,\le\, \frac{1}{2} \int_E |f_\mu - f_\nu| \,\mathrm{d}\lambda \,, \tag{2.63}$$

The other bound is obtained by choosing $f = \mathbf{1}_{f_\mu > f_\nu} - \mathbf{1}_{f_\mu < f_\nu}$ in (2.60). □

On the other hand, if E is countable then weak converge and convergence in total variation coincide. Here is the precise statement:

PROPOSITION 7.2. *If E is countable and endowed with the discrete topology (then all subset of E is a Borel set, consistent with our choice of $\mathcal{E}$) and (μ_n) is a sequence of probabilities on $(E, \mathcal{E})$ then (μ_n) converges in total variation if and only if it converges weakly. This convergence take place if and only if $\lim_n \mu_n(x) =: \mu(x)$ for every x and $\sum_x \mu(x) = 1$.*

PROOF. Let us start by remarking that in this countable context we have

$$\mathrm{d}_{\mathtt{TV}}(\mu,\nu) \,=\, \frac{1}{2}\sum_{x\in E}|\mu(x)-\nu(x)|\,, \tag{2.64}$$

because, by Proposition 7.1, $x \mapsto \mu(x)$ and $x \mapsto \nu(x)$ may be seen as densities respectively of μ and ν with respect to the counting measure (defined by $\lambda(x) := 1$ for every x).

We already know that convergence in total variation implies weak convergence (in full generality). Let us show the converse statement in the countable set up: assume that (μ_n) converges weakly to μ. Then, by using the test function $\mathbf{1}_{\{x\}}$, $\lim_n \mu_n(x) = \mu(x)$ for every x. For every $\varepsilon > 0$ there exists a finite subset K of E with $\mu(K) \geq 1-\varepsilon/2$. Hence there exists n_0 such that $\mu_n(K) \geq 1-\varepsilon$ for every $n \geq n_0$. But by (2.64)

$$\begin{aligned}\mathrm{d}_{\mathtt{TV}}(\mu_n,\mu) \,&\leq\, \frac{1}{2}\sum_{x\in K}|\mu_n(x)-\mu(x)|+\frac{1}{2}\left(\mu_n\left(K^{\complement}\right)+\mu\left(K^{\complement}\right)\right)\\ &\stackrel{n\geq n_0}{\leq}\, \frac{1}{2}\sum_{x\in K}|\mu_n(x)-\mu(x)|+\varepsilon\,,\end{aligned} \tag{2.65}$$

and by passing to the limit $n\to\infty$ we see that (μ_n) converges toi μ in total variation.

Note that the convergence (weak) convergence of $(\mu_n))$ to μ embodies the fact that μ is a probability. But the existence of the limit $\lim_n \mu_n(x) =: \mu(x)$ does not implies that μ is a probability, i.e. $\sum_x \mu(x) = 1$. But the argument we just gave shows that the existence of the limit $\lim_n \mu_n(x) =: \mu(x)$ and $\sum_x \mu(x) = 1$ yields convergence (in total variation, hence also weak convergence). □

Finally wide point out the following *coupling viewpoint* on convergence in total variation. We exploit the fact that a probability on E may be viewed as the law of an E valued random variable.

PROPOSITION 7.3. *We have that*

$$\mathrm{d}_{\mathtt{TV}}(\mu,\nu) \,=\, \inf_{X\sim\mu,Y\sim\nu}\mathbb{P}(X\neq Y)\,, \tag{2.66}$$

where X and Y are random variables defined on the same probability space.

PROOF. The upper bound is easy: from (2.34) we have that for every choice of $X\sim\mu$ and $Y\sim\nu$ on the same probability space

$$\mathrm{d}_{\mathtt{TV}}(\mu,\nu)=\sup_A\mathbb{E}\left[\mathbf{1}_A(X)-\mathbf{1}_A(Y)\right]=\sup_A\mathbb{E}\left[\mathbf{1}_A(X)-\mathbf{1}_A(Y);\,X\neq Y\right]\,\leq\mathbb{P}(X\neq Y). \tag{2.67}$$

For the lower bound we refer to ADD. □

CHAPTER 3

General Markov Chains

1. Harris Markov chains

A lot of interesting MC have no recurrent accessible state. In fact, in plenty of cases (think of a random walk with increment law that gives measure zero to every state, for example when the increment law has a density with respect to Lebesgue) no state is accessible (leave alone being recurrent). T. Harris ideas in this context is: can we modify the MC so that the new MC has an accessible recurrent state and such that we can relate results for the new MC to the original MC?

This is possible for a class of MC that we call Harris MC. To be precise, in general to an Harris MC we can associate a new MC with an accessible state. This state may not be recurrent: proving recurrence requires more work, but it is not difficult to give conditions that are sufficient for recurrence.

We say that a $\mathtt{p}$-MC is Harris if there exist A and B in $\mathcal{E}$, $\varepsilon > 0$ and a probability ρ on $(E, \mathcal{E})$ with $\rho(B) = 1$ such that

(1) A is accessible;
(2) if $x \in A$ et $C \subset B$, $C \in \mathcal{E}$, we have $\mathtt{p}(x, C) \geq \varepsilon \rho(C)$.

Let us give immediately some examples that are not particularly interesting, but they start giving a gist of what Harris chains are.

- If E is countable and x is accessible we can take $A = \{x\}$ and $B = \{y\}$, any y such that $Q(x, y) =: \varepsilon > 0$. End ρ is the Knornecker delta on y. In particular any irreducible MC with E countable is Harris.
- A random walk on $\mathbb{R}$, $X_{n+1} = X_n + \xi_{n+1}$, with $\xi \sim \mathcal{U}(-1, 1)$. In this case we can choose $A = B = [-1/2, 1/2]$, ρ the uniform measure on B and $\varepsilon = 1/2$. It is straightforward to show that A is accessible.
- A chain is said *atomic* if it possesses an *atom*, that is a measurable set A such that $\mathtt{p}(x, \mathrm{d}y) = \nu(\mathrm{d}y)$ for every x. If this atom is accessible then the chain is Harris with A the atom, $B = E$, $\rho = \nu$ and $\varepsilon = 1$.

Consider now a $\mathtt{p}$-MC that is Harris. We build the auxiliary chain that is going to be on the enlarged state space $E^\alpha := E \cup \alpha$, with α a singleton, that is α contains only one element, that (with abuse of notation) we call α. The σ-algebra of measurable subsets of E^α is $\mathcal{E}^\alpha := \{C, C \cup \alpha : C \in \mathcal{E}\}$. The probability kernel $\mathtt{p}^\alpha$

of the auxiliary chain is introduced first by defining $\mathtt{p}^\alpha(x,C)$ for $x\in E$

$$\mathtt{p}^\alpha(x,C) := \begin{cases} \mathtt{p}(x,C) & \text{if } x\in E\setminus A \text{ and } C\in\mathcal{E}, \\ \varepsilon & \text{if } x\in A \text{ and } C=\alpha, \\ \mathtt{p}(x,C)-\varepsilon\rho(C) & \text{if } x\in A \text{ and } C\in\mathcal{E}, \end{cases} \tag{3.1}$$

and the first line implies that $\mathtt{p}^\alpha(x,\alpha)=0$ for $x\in E\setminus A$ because $\mathtt{p}^\alpha$ is a probability kernel, and then by completing the definition with

$$\mathtt{p}^\alpha(\alpha,C) := \int_E \mathtt{p}^\alpha(y,C)\rho(\,\mathrm{d}y)\,, \tag{3.2}$$

where of course we can replace E with B as region of integration.

Remark 1.1. $\mathtt{p}^\alpha$ *is a probability kernel because* $\mathtt{p}^\alpha(x,E^\alpha)=1$ *for* $x\in E\setminus A$ *as we have already seen. For* $x\in A$ *we have* $\mathtt{p}^\alpha(x,E^\alpha)=\mathtt{p}(x,E)-\varepsilon\rho(E)+\varepsilon=1$ *and of course* $\mathtt{p}^\alpha(\alpha,E^\alpha)=\int_E \mathtt{p}^\alpha(y,E^\alpha)\rho(\,\mathrm{d}y)=1$. *We remark also that*

$$\mathtt{p}^\alpha(\alpha,\alpha) = \varepsilon\rho(A) = \varepsilon\rho(A\cap B)\,, \tag{3.3}$$

so in general the chain cannot go from α *to* α *in one step. But* α *is aperiodic for the* $\mathtt{p}^\alpha$*-MC if* $\rho(A\cap B)>0$ *(of course this is only a sufficient condition for aperiodicity).*

So α is an accessible state for $\mathtt{p}^\alpha$-MC. If we are able to show (and we will develop examples in this direction) that α is recurrent then we know that there exists an (essentially unique) $\mathtt{p}^\alpha$-invariant measure. So we can apply the Ratio Limit Theorem. And if the invariant measure is normalizable we can also apply the Ergodic Theorem (with convergence in total variation distance). Of course these results are interesting if we can export them to the $\mathtt{p}$-MC! But this is the case as we show now.

We start by introducing the elementary probability kernel $\mathtt{v}: E^\alpha\times\mathcal{E}^\alpha\to[0,1]$ by setting $\mathtt{v}(x,\{x\})=1$ if $x\in E$ and $\mathtt{v}(\alpha,C)=\rho(C)$ if $C\in\mathcal{E}$. Note that these requirements identify $\mathtt{v}$ because they imply that $\mathtt{v}(x,\alpha)=0$ for every $x\in E^\alpha$. So this kernel sends states in E^α into E: if the state is in E the kernel does nothing, if the state is α then it is sent inside E, in fact in B, according to the law ρ.

Lemma 1.2. $\mathtt{v}\mathtt{p}^\alpha=\mathtt{p}^\alpha$ *as kernels on* $E^\alpha\times\mathcal{E}^\alpha$ *and* $\mathtt{p}^\alpha\mathtt{v}=\mathtt{p}$ *as kernels on* $E\times\mathcal{E}$.

Proof. For the first identity we take $C\in\mathcal{E}^\alpha$ and observe that

$$\mathtt{v}\mathtt{p}^\alpha(\alpha,C) = \int_{E^\alpha}\mathtt{v}(\alpha,\,\mathrm{d}y)\mathtt{p}^\alpha(y,C) = \int_E \rho(\,\mathrm{d}y)\mathtt{p}^\alpha(y,C) \stackrel{(3.2)}{=} \mathtt{p}^\alpha(\alpha,C)\,. \tag{3.4}$$

If $x\in E$ instead

$$\mathtt{v}\mathtt{p}^\alpha(x,C) = \int_E \mathtt{v}(x,\,\mathrm{d}y)\mathtt{p}^\alpha(y,C) = \mathtt{p}^\alpha(x,C)\,, \tag{3.5}$$

so the first identity is established.

For the second one we compute for $x \in E$ and $C \in \mathcal{E}$

$$\begin{aligned} \mathtt{p}^\alpha \mathtt{v}(x, C) \,&=\, \int_{E^\alpha} \mathtt{p}^\alpha(x, \,\mathrm{d}y)\mathtt{v}(y, C) \,=\, \mathtt{p}^\alpha(x, \alpha)\mathtt{v}(\alpha, C) + \int_E \mathtt{p}^\alpha(x, \,\mathrm{d}y)\mathtt{v}(y, C) \\ &= \varepsilon \mathbf{1}_A(x)\rho(C) + \mathbf{1}_A(x) \int_C (\mathtt{p}(x, \,\mathrm{d}y) - \varepsilon\rho(\,\mathrm{d}y)) + \mathbf{1}_{A^\complement}(x) \int_C \mathtt{p}(x, \,\mathrm{d}y) \,=\, \mathtt{p}(x, C)\,, \end{aligned} \tag{3.6}$$

and the proof is complete. □

We are now building an inhomogeneous MC, that is a MC in which the kernel depends on the *time*. In fact, the process we build is only *very mildly* inhomogeneous: we just use a different probability kernel for even and odd times: for even times we apply the kernel $\mathtt{v}$ (that takes a process from the state E^α and puts it into E by *doing nothing* if the process is already in E and by taking α to a random point distributed according to ρ otherwise) and for odd times we apply the kernel $\mathtt{p}^\alpha$, that moves the process from E to E^α. The point is to realize that this process for even times is a $\mathtt{p}^\alpha$-MC (on E^α) and for odd times it is our original $\mathtt{p}$-MC on E.

We formalize this discussion (still keep a bit at an informal level, but little work is needed to make everything proper, except that one would have to introduce heavier notations). The process $(Y_n)_{n=0,1,\ldots}$ is a sequence of E^α valued random variables and if we call μ the law of Y_0, for every $m = 2, 4, 6, \ldots$ and every choice of events $A_0, \ldots, A_m$ in $\mathcal{E}^\alpha$

$$\begin{aligned} &\mathbb{P}\left(Y_j \in A_j,\, j = 1, \ldots, m\right) \,= \\ &\quad \int_{A_0}\int_{A_1} \cdots \int_{A_m} \mu(dy_0)\mathtt{v}(y_0, \,\mathrm{d}y_1)\mathtt{p}^\alpha(y_1, \,\mathrm{d}y_2) \cdots \mathtt{v}(y_{m-2}, \,\mathrm{d}y_{m-1})\mathtt{p}^\alpha(y_{m-1}, \,\mathrm{d}y_m)\,, \end{aligned} \tag{3.7}$$

and by choosing $A_m = E^\alpha$ we obtain the corresponding expression for m odd.

Lemma 1.3. *$(Y_{2n})_{n=0,1,\ldots}$ is a $\mathtt{p}^\alpha$-MC. Moreover $\mathbb{P}(Y_{2n+1} \in E) = 1$ for every $n = 0, 1, \ldots$ and $(Y_{2n+1})_{n=0,1,\ldots}$ is a $\mathtt{p}$-MC.*

Proof. This is just a matter of observing that the kernel from time $2n$ to time $2n+2$ is $\mathtt{v}\mathtt{p}^\alpha$ which coincides with $\mathtt{p}^\alpha$ by Lemma 1.2. On the other hand, the kernel from time $2n+1$ to time $2n+3$ is $\mathtt{p}^\alpha\mathtt{v}$ which coincides with $\mathtt{p}$, always because of Lemma 1.2. □

A priori one may worry whether with this procedure we cover all $\mathtt{p}$-MC, i.e. is it true that, given a law μ on E, we can find a law for Y_0 so that Y_1 is distributed according to μ? The answer is yes because it suffices to choose Y_0 according to μ^α which is the probability on E^α that assigns measure zero to α and such that $\mu^\alpha(C) = \mu(C)$ for every $C \in \mathcal{E}$.

Then it is not difficult to see that for every $f : E \to \mathbb{R}$ (measurable and bounded or positive)

$$\mathbb{E}_\mu \left[f(X_n)\right] \,=\, \mathbb{E}_{\mu^\alpha} \left[f(Y_{2n+1})\right] \,=\, \mathbb{E}_{\mu^\alpha} \left[\mathtt{v}f(Y_{2n})\right] \,, \tag{3.8}$$

where (X_n) is a $\mathtt{p}$-MC. Note that $\mathtt{v}f : E^\alpha \to \mathbb{R}$, in fact $\mathtt{v}f(x) = f(x)$ if $x \in E$ and $\mathtt{v}f(\alpha) = \int f \, \mathrm{d}\rho$.

2. Contractive Markov chains

Let us put ourselves in the Random Dynamical Systems context (or formalism), i.e. in the context of Proposition 1.4. When the function h that defines the random dynamical system has *good contractive properties* we can directly control the time asymptotic behavior of the MC and establish existence and uniqueness of the invariant probability. For this we need to assume that the state space E is metric: d denotes the distance between states. Here is the result:

Proposition 2.1. *Let us fix a measurable $h : E \times E' \to E$ (recall the notation $h(x, \xi) = h_\xi(x)$). Let us assume that*

1. *there exists a measurable function $K : E' \to [0, \infty)$ such that*
$$\mathrm{d}\left(h_\xi(x), h_\xi(y)\right) \,\le\, K(\xi)\,\mathrm{d}(x, y) \tag{3.9}$$
for every $x, y \in E$ and every $\xi \in E'$, and $\mathbb{E}[\log_+ K(\xi_1)] < \infty$ as well as $\mathbb{E}[\log K(\xi_1)] < 0$;
2. *there exists $y \in E$ such that $\mathbb{E}[\log_+ \mathrm{d}(y, h_{\xi_1}(y)] < \infty$.*

Then for every initial condition $X_0 = x \in E$ the MC (X_n) converges a.s. to an E valued random variable X_∞ that does not depend on x. Moreover the law of X_∞ is the unique invariant probability of the MC.

We give a full proof of Proposition 2.1 only for a particular model: the Random Coefficient Autoregressive Markov Chain. The arguments we develop in this restricted context should allow the interested reader to develop a full proof of Proposition 2.1.

2.1. Random Coefficient Autoregressive MC. The basic Random Coefficient Autoregressive (RCA) MC is defined iteratively, once the initial condition $X_0 = x \in \mathbb{R}$ is given, by

$$X_{n+1} \,=\, K_{n+1} X_n + C_{n+1} \,=:\, f_{K_{n+1}, J_{n+1}}(X_n) \,, \tag{3.10}$$

with $((K_n, J_n))_{n\in\mathbb{N}}$ and IID sequence of random variables taking values in $\mathbb{R}^2$. Note that in this case

$$\left|f_{K,J}(x) - f_{K,J}(y)\right| \,\le\, |K|\, |x - y| \,, \tag{3.11}$$

so this process has good contractive properties if $\log |K| \in \mathbb{L}^1$ and $\mathbb{E} \log |K| < 0$. In order to make this clear and explicit let us consider the particular case in which $J_n = K_n$ for every n, so

$$X_{n+1} \,=\, K_{n+1} \left(1 + X_n\right) \,=:\, f_{K_{n+1}}(X_n) \,, \tag{3.12}$$

and let us simplify things a bit by choosing the K variables non negative and non trivial (recall that we assume $\log K \in \mathbb{L}^1$ and $\mathbb{E} \log K < 0$). In this case it is not difficult to see that if $x < 0$, the stopping time $\inf\{n : X_n \geq 0\}$ is a.s. finite, so the negative semi-axis is transient, but we will work in any case with $E = \mathbb{R}$ and what we just claimed will come out of the analysis we will develop.

By direct inspection it is not difficult to see that

$$\begin{aligned} X_n \;&=\; f_{K_n} \circ \ldots \circ f_{K_1}(x) \;= \\ & K_n + K_n K_{n-1} + K_n K_{n-1} K_{n-2} + \ldots + K_n K_{n-1} \cdots K_2 + K_n K_{n-1} \cdots K_1 (1 + x)\,. \end{aligned} \tag{3.13}$$

If we *reverse the engine* of this MC we obtain

$$\begin{aligned} Y_n \;&:=\; f_{K_1} \circ \ldots \circ f_{K_n}(x) \;= \\ & K_1 + K_1 K_2 + K_1 K_2 K_3 + \ldots + K_1 K_2 \cdots K_{n-1} + K_1 K_2 \cdots K_n (1 + x)\,, \end{aligned} \tag{3.14}$$

which is an increasing process if $x \geq 0$. But even if $x < 0$, it is not difficult to see that $\lim_n Y_n := Y_\infty$

$$Y_\infty \;:=\; \sum_{n=1}^{\infty} \prod_{j=1}^{n} K_j\,, \tag{3.15}$$

exists and it is a.s. finite. In fact, by the law of large numbers $(1/n) \log K_1 K_2 \cdots K_n \longrightarrow \mathbb{E} \log K < 0$, so for every $\beta \in (\exp(\mathbb{E} \log K), 1)$ there exists $C(\omega)$ (C is an a.s. finite random variable) such that for every $n \in \mathbb{N}$

$$K_1(\omega) K_2(\omega) \cdots K_n(\omega) \;\leq\; C(\omega) \beta^n\,. \tag{3.16}$$

This suffices to show that $(Y_n(\omega))$ is a.s. a Cauchy sequence and therefore the limit $\lim_n Y_n =: Y_\infty$ exists a.s.. Note that (3.16) yields also that the limit is independent of the value of x. Note moreover that Y_∞ is supported on $[0, \infty)$. And now by exploiting that $X_n \sim Y_n$ for every n we can complete (*Exercise*) the proof of

PROPOSITION 2.2. *The random coefficient autoregressive process defines by* (3.12) *has a unique invariant probability* ν *and for every initial condition* $x \in \mathbb{R}$ *we have* (X_n) *converges in law to* Y_∞.

This result can be easily generalized almost (verbatim) to the case in which $J = J_1$ is random with $\mathbb{E}[(\log |J|)_+] < \infty$ and to the case in which the K and J variables assume also negative values. Of course the invariant probability will no longer be supported on the positive semi-axis.

REMARK 2.3. *The invariant probability is more interesting than it looks at first. Note, for example, that even if* $K(= J) \geq 0$ *is a bounded random variable,* Y_∞ *may not even be in* $\mathbb{L}^1$. *In fact*

$$\mathbb{E}[Y_\infty] \;=\; \sum_{n=1}^{\infty} (\mathbb{E}[K])^n\,, \tag{3.17}$$

so $\mathbb{E}[Y_\infty] < \infty$ *if and only if* $\mathbb{E}[K] < 1$. *With some work (but not too much) it is also possible to see that an analogous result holds for all moments:* $\mathbb{E}[Y_\infty^k] < \infty$ *if and only if* $\mathbb{E}[K^k] < 1$. *Therefore* Y_∞ *has a heavy tail unless* $\mathbb{E}[K^k] < 1$ *for every* k, *which requires* $\mathbb{P}(K < 1) = 1$ *(i.e., that the process is contractive for every* ω, *not just in a probabilistic sense).*

Sticking for simplicity to the case (3.12), let us address the question of whether this process is a Harris MC. This requires conditions on law of K. We will not try to look for optimal conditions and we start by observing that if K is a continuous random variable – we denote by f its density – then measure $\mathtt{p}(x,\cdot)$, $\mathtt{p}$ is the transition kernel, has density $y \mapsto f(y/(1+x))/(1+x)$. In particular for $x = 1$ the map is $y \mapsto f(y/2)/2$ Therefore if $f(1/2) > 0$, by continuity we can find $\varepsilon_0 > 0$ and $\delta > 0$ such that $f(y/(1+x))/(1+x) \geq \varepsilon_0$ uniformly in $x, y \in [1-\delta, 1+\delta]$. Therefore we can choose $A = B = [1-\delta, 1+\delta]$, ρ the uniform mesure on B, and $\varepsilon = 2\delta\varepsilon_0$ and satisfy the second of the Harris requirement.

But also the first Harris requirement is fulfilled. In fact it suffices to show that for every $x \in \mathbb{R}$ we can find a value of n and $a_1, \ldots, a_n \in I_\eta := (-\eta + 1/2, \eta + 1/2)$ (we are choosing η so that $\inf_{I_\eta} f > 0$) we have that

$$a_n + a_n a_{n-1} + \ldots + \ldots + a_n a_{n-1} \cdots a_2 + a_n a_{n-1} \cdots a_1 (1+x) \in (1-\delta, 1+\delta)\,. \quad (3.18)$$

Once this is established, the result is obtained because (3.18) holds also in an open neighborhood of $(a_1, \ldots, a_n)$. The requirement (3.18) may appear difficult to establish, but it is not the case. In fact, it suffices to remark that if $x < 1$ then $x < (1+x)/2 < 1$, so even by choosing simply $a_j = 1/2$ for every j we will hit the target (of becoming larger than $1-\delta$) in a finite number of steps. Analogous reasoning for $x > 1$.

As a matter of fact, the argument we just developed shows that the chain is Harris with $A = B = [1-\delta, 1+\delta]$ and suitable choice of ε (ρ is the uniform probability on B) under the assumption that the law of K is bounded below by a measure with a density f that is continuous and $f(1/2) > 0$.

Remark 2.4. *A similar argument can be developed if* $f(x) > 0$ *for* $x \in (1/2, 1]$.

Remark 2.5. *Another interesting point is to notice that the minimum of the support of the invariant probability can be determined with precision: if we call* a, $a < 1$ *by hypothesis, the minimum of the support of the law of* K, *then by iterating from* $x = 0$ *we see that the support of the invariant probability does not go below* $a + a^2 + \ldots = a/(1-a)$. *Just a slightly more involved argument gives the bound in the other direction. On the other hand the supremum of the support of the invariant probability is* $+\infty$, *since we are assuming that the maximum of the support of* K *is larger than one.*

In the cases in which we are able to prove that the MC is Harris, we can apply the general theory and conclude that , under the assumption that $\mathbb{E}[\log K] < 0$, the only invariant measure is the invariant probability and that $p_n(x,\cdot)$ converges to the invariant probability in total variation distance for every x such that $\mathbb{P}_x(T_A < \infty) = 1$. But we know by the weak convergence result (obtained by exploiting the convergence in law result in Proposition 2.2) that $\lim_n \mathbb{E}_x[h(X_n)] = \int h \, \mathrm{d}\nu$ for every x (ν is the invariant probability). By choosing h to be a smoothed version of $\mathbf{1}_A$ we see that the chain that starts from x visits A a.s., that is $\mathbb{P}_x(T_A < \infty) = 1$ for every x.

3. Feller chains and Foster-Lyapunov criteria

In this section we assume that E is a metric space, d is the notation for the distance, and the elements of $\mathcal{E}$ are the Borel subsets of E. Moreover for the results in this section the notion of *Feller* kernel is important: we say that $\mathtt{p}$ is Feller if $\mathtt{p}f \in C^0$ (hence $\mathtt{p}f \in C^0_b$) for every $f \in C^0_b$. Note that if $\mathtt{p}$ is Feller, so are $\mathtt{p}_n$ (any n) and $\mathtt{p}_\star$ (given in (1.23)).

EXERCISE 3.1. *If the* $\mathtt{p}$*-MC* (X_n) *is defined by the random dynamical system* $X_{n+1} = h(X_n, \xi_{n+1})$ *(see Proposition 1.4) and if* $x \mapsto h(x, \xi)$ *is* C^0 *for almost every* ξ*, then* $\mathtt{p}$ *is Feller.*

SOLUTION. If $\lim_n x_n = x$ we have $\mathtt{p}f(x_n) = \mathbb{E}[f(h(x_n, \xi))] \longrightarrow \mathbb{E}[f(h(x, \xi))] = \mathtt{p}f(x)$ by Dominated Convergence, because $f \circ h_\xi$ is C^0 for almost every ξ and it is bounded. ☐

For $n \in \mathbb{N}$, μ a probability and $\mathtt{p}$ a Markov kernel we introduce the probability

$$\pi_n \,=\, \pi_n^\mu \,=\, \pi_n^{\mu,\mathtt{p}} \,:=\, \frac{1}{n} \sum_{j=0}^{n-1} \mu \mathtt{p}_k \,. \tag{3.19}$$

LEMMA 3.2. *If* $\mathtt{p}$ *is Feller then for every* μ *we have that the limit of every weakly convergent subsequence of* $(\pi_n^{\mu,\mathtt{p}})$ *is a* $\mathtt{p}$*-invariant probability.*

PROOF. This is a direct consequence of the identity

$$\pi_n \mathtt{p} \,=\, \pi_n + \frac{1}{n}\left(\mu \mathtt{p}_n - \mu\right) \,. \tag{3.20}$$

In fact, assume without loss of generality that (π_n) converges weakly to π. For $f \in C^0_b$ we have $\int_E f \, \mathrm{d}(\pi_n \mathtt{p}) = \int_E \mathtt{p}f \, \mathrm{d}\pi_n$ and by the Feller property we obtain that $(\pi_n \mathtt{p})$ converges to $\pi \mathtt{p}$. On the other hand $|\int_E f \, \mathrm{d}(\mu \mathtt{p}_n - \mu)|/n \le 2\|f\|_\infty / n$ and we obtain that $\pi \mathtt{p} = \pi$. ☐

Lemma 3.2 hides the difficulty in the convergence assumption that embodies the fact that the limit is a probability. In turn, the fact that the limit is a probability is due to the test functions for the weak convergence we consider, that is C^0_b. And it is not at all granted that $(\pi_n^{\mu,\mathtt{p}})$ has a convergent subsequence: as an example we can take the (Feller) Kernel $\mathtt{p}(x, \mathrm{d}y) = \delta_{x+1}(\mathrm{d}y)$ for which $\mathtt{p}_n f(x) = f(x+n)$. The problem is of course that, no matter what the initial condition is, π_n in this case *concentrates* (i.e., gives probability $1-\varepsilon$, any $\varepsilon > 0$) on a set that walks out to $+\infty$ as n becomes large.

In spite of the very special nature of this example, it captures the only problem that can happen. This is the content of the next result, *Prohorov Theorem*, for which we need the notion of *tightness* (we just give it for sequences): a sequence (μ_n) of probabilities on $(E, \mathcal{E})$, with E a topological space, is said *tight* if for every $\varepsilon > 0$ there exists a compact $K \subset E$ such that $\mu_n(K) \geq 1-\varepsilon$ for every n.

THEOREM 3.3 (Prohorov Theorem). *If E is a metric space and (μ_n) is tight, then every subsequence of (μ_n) contains a weakly convergent subsequence.*

For a proof of Prohorov Theorem (along with a necessary and sufficient version of it) we refer to [**3**, Ch. 1, Sec. 5]. For the simpler case of $E = \mathbb{R}$ (in this context the result is often called Helly-Bray Lemma or Helly's Theorem) see for example [**2**, p. 336].

Here is the first *Foster-Lyapunov* tool to control that the probability does not escape to infinity in time. It directly yields tightness, as a the result of a strong quantitative bound (at the expense of strong assumptions, that we will weaken in the Foster-Lyapunov arguments that follow this first one).

PROPOSITION 3.4. *Assume that there exists $V : E \to [0, \infty]$ measurable with $V(x_0) < \infty$ for a state x_0 and that satisfies*

$$\mathtt{p}V + f \;\leq\; V + b\,, \tag{3.21}$$

for a $b \in \mathbb{R}$ and for $f : E \to \mathbb{R}$ measurable, bounded below and such that $\{x \in E : f(x) \leq c\}$ is relatively compact (i.e., its closure is compact) for every c. Then $\left(\pi_n^{\delta_{x_0}}\right)$ is tight and, if $\mathtt{p}$ is Feller, there exists a $\mathtt{p}$-invatiant probability.

PROOF. Let us start with the non crucial remark the one can always choose $f \geq 0$ by an appropriate choice of b. By applying the kernel to both terms of (3.21) and by using again (3.21) we obtain

$$\mathtt{p}_2 V + \mathtt{p}f \;\leq\; \mathtt{p}V + b \;\leq\; V + 2b - f\,, \tag{3.22}$$

that is

$$\mathtt{p}_2 V + \mathtt{p}f + f \;\leq\; V + 2b\,, \tag{3.23}$$

and we can iterate this procedure to obtain

$$\mathtt{p}_n V + \sum_{k=0}^{n-1} \mathtt{p}_k f \;\leq\; \mathtt{p}V + nb\,. \tag{3.24}$$

Therefore, since $V \geq 0$, we obtain

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathtt{p}_k f(x_0) \;\leq\; \frac{1}{n} V(x_0) + b\,, \tag{3.25}$$

that yields

$$\pi_n^{\delta_{x_0}} f \;\le\; V(x_0) + b\,, \tag{3.26}$$

for every n. By considering the (relatively) compact set $K_c = \{x : \, f(x) \le c\}$ for $c > 0$ we readily see that

$$\pi_n^{\delta_{x_0}}\left(K_c^{\complement}\right) \;\le\; \frac{1}{c}\pi_n^{\delta_{x_0}} f \;\le\; \frac{V(x_0)+b}{n}\,, \tag{3.27}$$

which implies that $\left(\pi_n^{\delta_{x_0}}\right)$ is tight. The proof is completed by Prohorov Theorem and by applying Lemma 3.2. □

The next result yields existence of an invariant probability under much weaker hypotheses (but it requires conditions on the metric space, so, strictly speaking, the hypotheses are weaker only if we restrict to the metric spaces to which the result applies). This new result does not really yield explicit estimates.

PROPOSITION 3.5. *If E is a locally compact and separable metric space,* $\mathtt{p}$ *is Feller and there exists a measurable function $V : E \to [0, \infty]$ with $V(x_0) < \infty$ for a state x_0 and such that*

$$\mathtt{p}V + 1 \;\le\; V + b\mathbf{1}_K\,, \tag{3.28}$$

for a positive constant b and K a compact subset of E, then there exists a $\mathtt{p}$*-invariant probability.*

PROOF. The proof uses the weak-$*$ convergence of (finite) measures: a sequence of positive finite measures (μ_n) weak-$*$ converges if there exists a finite measure μ such that $\lim_n \int_E h \,\mathrm{d}\mu_n = \int_E h\,\mathrm{d}\mu$ for every $h \in C_c^0(E;\mathbb{R})$, that is for every h which is continuous and compactly supported. The key point here is that from every subsequence of (μ_n) one can extract a weak-$*$ convergent subsequence. In general, weak-$*$ limit measures can be null measures. When dealing with probabilities, weak-$*$ convergence readily yields that $\mu(K) \le 1$ for every weak-$*$ limit μ and every compact set K. So $\mu(E) \le 1$ and, in general, μ is a subprobability (and, again, it can be that $\mu(E) = 0$). On the other hand, if there exists a compact set K such that $\mu_n(K) \ge \varepsilon > 0$ for every n, then for every weak-$*$ limit μ we have $\mu(K) \ge \varepsilon$.

We are going to apply these facts to $\left(\pi_n^{\delta_{x_0},\mathtt{p}}\right)$, like for Propositon 3.4. And, like in the proof of the same proposition, by iterating (3.28) we obtain

$$\begin{aligned} V \;&\ge\; \mathtt{p}V + 1 - b\,\mathbf{1}_K \\ &\ge\; \mathtt{p}\,(V + 1 - b\,\mathbf{1}_K) + 1 - b\,\mathbf{1}_K \;=\; \mathtt{p}_2 V + 2 - b\,\mathtt{p}\mathbf{1}_K - b\,\mathbf{1}_K \\ &\ge \mathtt{p}_n V + n - b\sum_{j=0}^{n-1} \mathtt{p}_k \mathbf{1}_K\,, \end{aligned} \tag{3.29}$$

that is

$$V(x_0) \;\ge\; \mathtt{p}_n V(x_0) + n - b\sum_{j=0}^{n-1} \mathtt{p}_k(x_0, K)\,, \tag{3.30}$$

that we rewrite as

$$\frac{1}{n}V(x_0) \,\geq\, \frac{1}{n}\mathtt{p}_n V(x_0) + 1 - b\,\pi_n^{\delta_{x_0},\mathtt{p}}(K)\,, \tag{3.31}$$

and we use this last bound to obtain

$$b\,\pi_n^{\delta_{x_0},\mathtt{p}}(K) \,\geq\, 1 - \frac{1}{n}V(x_0)\,. \tag{3.32}$$

Therefore for any weak-$*$ limit μ of $\left(\pi_n^{\delta_{x_0},\mathtt{p}}\right)$ we have $\mu(K) \geq 1/b$.

On the other hand, by recalling (3.20), we see that (with $\pi_n = \pi_n^{\delta_{x_0},\mathtt{p}}$) for every h bounded

$$|\pi_n \mathtt{p} h - \pi_n h| \,\leq\, \frac{2}{n}\|h\|_\infty\,. \tag{3.33}$$

We now apply this bound with $h \in C_c^0$ and $h \geq 0$ so that, if the weak-$*$ limit of (π_{n_j}) is μ, we obtain that $\int_E \mathtt{p}h\,\mathrm{d}\mu \leq \int_E h\,\mathrm{d}\mu$, which readily yields that $\mu\mathtt{p}(A) \leq \mu(A)$ for every $A \in \mathcal{E}$. On the other hand, $\mu\mathtt{p}(E) = \mu(E)$ (this holds for every measure μ) and therefore $\mu\mathtt{p}(A) = \mu(A)$ for every $A \in \mathcal{E}$.

Now we recall that $\mu(K) > 0$ so μ is not null. A priori we know that μ is a subprobability, so we can normalize it to be a probability: $\pi := \mu/\mu(E)$ and $\pi\mathtt{p} = \pi$. We have therefore found a $\mathtt{p}$-invariant probability. □

3.1. A Foster-Lyapunov argument for (null) recurrence. Roughly, we would like to say that if we can find V bounded below such that $\lim_{x:\,|x|\to\infty} V(x) = \infty$ and such that $pV \leq V$ outside of a compact set, then the $\mathtt{p}$-MC is recurrent. In order to minimize introducing definitions, we give a minimal version of this result that is tailored to the two applications we give.

PROPOSITION 3.6. *Let us consider $E = [l, \infty)$, $l \in \mathbb{R}$, and a $\mathtt{p}$-MC on this space. We assume that there exists $V : E \to [0, \infty)$ which outside of a bounded set satisfies two properties:*

(1) V is (strictly) increasing and $\lim_{x\to\infty} V(x) = \infty$;

(2) $pV \leq V$.

We also assume that for every x outside of a bounded set there exists $\varepsilon > 0$ such that $\inf_{y\in B_\varepsilon(x)} \mathbb{P}_y(T_{B_\varepsilon(x)^\complement} < \infty) > 0$ with $B_\varepsilon(x) := \{z : |z - x| < \varepsilon\}$. Then there exists $r > l$ such that for every $x \in E$ we have $\mathbb{P}_x(T_{[l,r]} < \infty) = 1$.

The condition on $\mathbb{P}_y(T_{B_\varepsilon(x)^\complement} < \infty) > 0$ is a very weak requirement (see applications) in order to avoid that the process gets stuck at some x for arbitrarily large x.

The proof is given as a guided exercise.

PROOF. First of all note (*Exercise*) that the hypothesis involving $B_\varepsilon(x)$ implies that $\mathbb{P}_x(\sum_n \mathbf{1}_{B_\varepsilon(x)^\complement}(X_n) = \infty) = 1$. Moreover, we can choose $r > l$ such that the hypotheses hold for $x \geq r$. We set for conciseness $T = T_{[l,r]}$ (the hitting time of $[l, r]$) and we introduce $Y_n := V(X_{n\wedge T})$ for $n = 0, 1, \ldots$. The process $(Y_n)_{n=0,1,\ldots}$ is a super-martingale for every choice of $Y_0 = x \in E$ (*Exercise*: note that by iterating

$pV \le V + b\mathbf{1}_{[l,r]}$, that holds by hypothesis for a positive b, we obtain $Y_n \in \mathbb{L}^1$ for every n). Since $Y_n \ge 0$ we have that $(Y_n(\omega))$ converges to a limit that we call $Y_\infty(\omega) < \infty$ for every $\omega \in G$, with $\mathbb{P}_x(G) = 1$.

If there exists x such that $\mathbb{P}_x(T_{[l,r]} = \infty) > 0$ then for $\omega \in \{T_{[l,r]} = \infty\} \cap G$ we have that $X_n(\omega) > r$ for every n and therefore $Y_n(\omega) = V(X_n(\omega)) \longrightarrow Y_\infty(\omega) < \infty$. Since V diverges at infinity and since it is a bijection on the region we consider, we obtain that $(X_n(\omega))$ converges to a limit in $X_\infty(\omega) \in [r, \infty)$. But this means that $X_n(\omega) \in B_\varepsilon(X_\infty(\omega))$ for every $\varepsilon > 0$ and all n large and this is incompatible with our hypothesis on the exit probability. Therefore $\mathbb{P}_x(T_{[l,r]} = \infty) = 0$ for every x. $\square$

3.1.1. *Application to the Lindley MC with centered drift.* We consider the Lindley MC with $\mathbb{E}[\xi] = 0$ (we recall that we exclude the trivial case of $\xi \equiv 0$). We treat only the case in which there exists L such that $\mathbb{P}(\xi < -L) = 0$. This is an assumption that simplifies (a lot!) the analysis: it is possible to generalize this result at least to the case in which the variance of ξ is finite (with the very same choice of V). We choose $V(x) = \log(1 + x)$, $x \in [0, \infty) = E$. We remark that the following elementary bound holds: for $y > -1$

$$\log(1 + y) \le y - \frac{1}{2} y^2 \mathbf{1}_{y<0} . \tag{3.34}$$

Then

$$pV(x) = \mathbb{E}\left[\log\left(1 + (x + \xi)_+\right)\right] = V(x) + \mathbb{E}\left[\log\left(1 + \frac{(x + \xi)_+ - x}{1 + x}\right)\right] . \tag{3.35}$$

If $x > L$ then a.s. $(x + \xi)_+ = x$. Therefore for $x > L$

$$pV(x) - V(x) = \mathbb{E}\left[\log\left(1 + \frac{\xi}{1 + x}\right)\right] \le -\frac{1}{2(1 + x)^2} \mathbb{E}[\xi^2; \xi < 0] . \tag{3.36}$$

Therefore $pV(x) < V(x)$ for every $x > L$. Finally, in this case and outside $[0, L]$, the evolution is just a random walk and since ξ is centered and nontrivial, we have that $\mathsf{p}(y, B_y(2\varepsilon)^\complement) = \mathbb{P}(\xi \in B_0(2\varepsilon)^\complement) =: p_\varepsilon > 0$ for $y > L$ and $\varepsilon > 0$, so $\mathsf{p}(y, B_x(\varepsilon)) \ge p_\varepsilon$ for every $y \in B_x(\varepsilon)$ and $x > L + \varepsilon$. Therefore we can apply Proposition 3.6 and the set $[0, L + \varepsilon]$ is visited infinitely often by the MC. From this one easily extracts that also 0 is visited infinitely often (so 0, and the whole MC since 0 is accessible from every x, is recurrent).

3.1.2. *Application to the RCA MC with* $\mathbb{E}[\log K] = 0$. Also in this case we simplify our life by making a strong assumption on the support of $\log K$: there exists L such that $\mathbb{P}(\log K > -L) = 1$.

The first step is to work with $Z_n := \log X_n$:

$$Z_{n+1} = \log K_{n+1} + \log\left(1 + \exp(Z_n)\right) = \log K_{n+1} + Z_n + \log\left(1 + \exp(-Z_n)\right) , \tag{3.37}$$

which makes clear that Z behaves almost as a random walk when it is positive and far from the origin. On the other hand, Z has a lot of difficulty to enter the negative semi axis. In fact, since $\xi \ge -L$, if $Z_0 < -L$ then $Z_1 \ge -L$ and $Z_n \in [-L, \infty)$ for every $n \ge 1$. So we can choose $E = [-L, \infty)$.

We then choose $V(x) = \log_+(x)$ for $x > 0$ and $V(x) = 0$ for $x \in [-L, 0]$. Since $\xi \geq -L$ for $x \geq L+1$

$$pV(z) \;=\; \log z + \mathbb{E}\left[\log\left(1 + \frac{\log K}{z} + \frac{\log(1+\exp(-z))}{z}\right)\right], \tag{3.38}$$

so by (3.34)

$$\begin{aligned} pV(z) \;&=\; V(z) + \frac{\log(1+\exp(-z))}{z} - \\ &\quad \frac{1}{2z^2}\mathbb{E}\left[(\log K + \log(1+\exp(-z)))^2 ;\, \log K + \log(1+\exp(-z)) < 0\right] \\ &\overset{z\to\infty}{=} V(z) + O\left(\frac{\exp(-z)}{z}\right) - \frac{1}{2z^2}\left(\mathbb{E}\left[(\log K)^2 ;\, \log K < 0\right] + o(1)\right) . \end{aligned} \tag{3.39}$$

Therefore there exists $M > L+1$ such that $pV(z) \leq V(z)$ for every $z \geq M$. The argument to show that Z cannot visit infinitely many times a neighborhood of a point goes pretty much as for the Lindley case. Therefore we can apply Proposition 3.6 and conclude that $[-L, M]$ is a.s. visited by Z and therefore $[\exp(-L], \exp(M)]$ is a.s. visited by the RCA MC X.

Bibliography

[1] S. Asmussen, *Applied Probability and Queues*, Second edition, Springer-Verlag, New York, 2003.
[2] P. Billingsley, *Probability and Measure*, third edition, John Wiley and sons, 1995.
[3] P. Billingsley, *Convergence of Probability Measures*, second edition, Wiley Series in Probability and Statistics, 1999.
[4] R. Douc, E. Moulines, P. Priouret and P. Soulier, *Markov Chains*, Springer Ser. Oper. Res. Financ. Eng., Springer, Cham, 2018.
[5] R. Durret, *Probability: Theory and Examples*, Second edition, Duxbury Press, 1996.
[6] J.-F. Le Gall, *Intégration, Probabilités er Processus Aléatoires*, https://www.math.u-psud.fr/ jflegall/IPPA2.pdf
[7] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, second edition, Cambridge University Press, 2009.
[8] D. Williams, *Probability with Martingales*, Cambridge Mathematical textbooks, Cambridge University Press, 1991.