Notes 23 : Markov chains: asymptotic behavior

Math 733-734: Theory of Probability *Lecturer: Sebastien Roch*

References: [Dur10, Sections 6.6] and [Nor98, Sections 1.8].

Recall:

THM 23.1 (Strong Markov property) *Let $\{X_n\}$ be an MC on S with transition kernel p. Let T be a stopping time. For each $n \geq 0$, let h_n be a bounded measurable function from $S^{\mathbb{Z}_+}$ to $\mathbb{R}$. Then, on $\{T < +\infty\}$,*

$$\mathbb{E}_\mu[h_T(X_T, X_{T+1}, \ldots) \,|\, \mathcal{F}_T] = \phi_T(X_T),$$

where $\phi_n(x) = \mathbb{E}_x[h_n(X_n, X_{n+1}, \ldots)]$.

THM 23.2 (Distribution at time n) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. Then for all $n \geq 0$ and $j \in S$*

$$\mathbb{P}_\mu[X_n = j] = \sum_{i \in S} \mu(i) p^n(i, j),$$

where p^n is the n-th matrix power of p, i.e.,

$$p^n(i, j) = \sum_{k_1, \ldots, k_{n-1}} p(i, k_1)\, p(k_1, k_2) \cdots p(k_{n-1}, j).$$

Let $\{X_n\}$ be an MC on a countable set S with transition probability p. For $x, y \in S$, let $T_x^+ = \inf\{n > 0 \,:\, X_n = x\}$, $\rho_{xy} = \mathbb{P}_x[T_y^+ < +\infty]$, and $N(x) = \sum_{n \geq 1} \mathbb{1}\{X_n = x\}$. Define $T_x = \inf\{n \geq 0 \,:\, X_n = x\} = T_x^+ \mathbb{1}\{X_0 = x\}$. If $\mathbb{P}_x[T_y < +\infty] > 0$, we write $x \to y$. If $x \to y$ and $y \to x$, we write $x \leftrightarrow y$ and say that x *communicates with* y.

LEM 23.3 *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. Then, for distinct states $x \neq y \in S$, the following are equivalent:*

(a) $\rho_{xy} > 0$

(b) $p^n(x, y) > 0$ *for some* $n \geq 1$

(c) $\exists i_0 = x, i_1, \ldots, i_n = y \in S$ *such that* $p(i_{r-1}, i_r) > 0$ *for all* $r = 1, \ldots, n$

DEF 23.4 (Recurrence) *A state $x \in S$ is* recurrent *if $\rho_{xx} = 1$. Otherwise it is* transient.

LEM 23.5 (Recurrence is contagious) *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$.*

DEF 23.6 (Irreducibility) *A subset $C \subseteq S$ is* irreducible *if for all $x, y \in C$, we have $x \leftrightarrow y$. An MC on S is irreducible if the full space S is irreducible.*

DEF 23.7 (Stationary measure) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. A measure μ on S is* stationary *if*

$$\sum_{i \in S} \mu(i)\, p(i,j) = \mu(j).$$

If in addition μ is a probability measure, then we say that μ is a stationary distribution.

For $x, y \in S$, let

$$\gamma_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{T_x^+ - 1} \mathbb{1}\{X_n = y\}\right] = \sum_{n=0}^{+\infty} \mathbb{P}_x\left[X_n = y, n \le T_x^+ - 1\right]. \tag{1}$$

THM 23.8 (Existence of stationary measure) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. Let x be recurrent. Then γ_x is a stationary measure. In addition: $\rho_{xy} = 0$ implies $\gamma_x(y) = 0$; while $\rho_{xy} > 0$ implies $0 < \gamma_x(y) < +\infty$.*

DEF 23.9 (Positive recurrence) *A recurrent state $x \in S$ is* positive recurrent *if $\mathbb{E}_x[T_x^+] < +\infty$. Otherwise it is* null recurrent.

THM 23.10 *Let $\{X_n\}$ be an irreducible MC on a countable set S. Then the following statements are equivalent:*

(i) every state is positive recurrent;

(ii) some state is positive recurrent;

(iii) there exists a stationary distribution.

Moreover, when any of the conditions above holds, the unique stationary distribution is given by $\pi(x) = \frac{1}{\mathbb{E}_x[T_x^+]}$.

We will also need:

THM 23.11 (Strong law of large numbers) *Let $X_1, X_2, \ldots$ be IID with $\mathbb{E}|X_1| < +\infty$. Let $S_n = \sum_{k \leq n} X_k$ and $\mu = \mathbb{E}[X_1]$. Then*

$$\frac{S_n}{n} \to \mu, \quad \textit{a.s.}$$

THM 23.12 (SLLN: Infinite mean case) *Let $X_1, X_2, \ldots$ be IID with $\mathbb{E}[X_1^+] = +\infty$ and $\mathbb{E}[X_1^-] < +\infty$. Then*

$$\frac{S_n}{n} \to +\infty, \quad \textit{a.s.}$$

1 Convergence to equilibrium

Throughout, we assume that S is countable. We also restrict ourselves to the irreducible, positive recurrent case, where a unique stationary distribution is known to exist by the theorems above. (Observe that we have already proved that, when y is transient, then $\mathbb{E}_x[N(y)] = \sum_{n \geq 0} p^n(x, y) < +\infty$ so that $p^n(x, y) \to 0$.)

Even when a stationary distribution π exists, there is no guarantee in general that $p^n(x, y) \to \pi(y)$. For instance:

EX 23.13 (Periodic behavior) *Let $S = \{1, 2\}$ and*

$$P = \begin{pmatrix} p(1,1) & p(1,2) \\ p(2,1) & p(2,2) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Note that P^2 is the identity I and, as a result, that $P^m = P$ for odd m and $P^m = I$ for even m. Because, by THM 23.2,

$$P^m = \begin{pmatrix} p^m(1,1) & p^m(1,2) \\ p^m(2,1) & p^m(2,2) \end{pmatrix},$$

we have established that $p^m(1,1)$ does not converge as $m \to +\infty$. This despite the fact that a stationary distribution exists, as can be checked from noting that $\pi P = \pi$ where $\pi = (1/2, 1/2)$ (as a row vector).

To exclude the effect seen in the previous example, we introduce a definition (for a more through treatment of periodicity, see [Dur10, Chapter 6]).

DEF 23.14 (Aperiodicity) *An MC $\{X_n\}$ on a countable set S with transition probability p is* aperiodic *if, for all $x \in S$, we have $p^n(x, x) > 0$ for all n large enough.*

LEM 23.15 (Criterion for aperiodicity) *For an irreducible chain $\{X_n\}$ to be aperiodic, it suffices that there exists a state $x \in S$ and an integer K such that $p^K(x,x) > 0$ and $p^{K+1}(x,x) > 0$. In particular, this is immediate if $p(x,x) > 0$ for some x.*

Proof: Let x as in the statement. By irreducibility, for any $y \neq x$, there are L and M such that $p^L(y,x) > 0$, $p^M(x,y) > 0$ and hence

$$p^{L+n+M}(y,y) \geq p^L(y,x)p^n(x,x)p^M(x,y)$$

and it suffices to show that $p^n(x,x) > 0$ for n sufficiently large.

We also need the following simple observations: for m, m' such that $p^m(x,x) > 0$ and $p^{m'}(x,x) > 0$, we have also $p^{km}(x,x) \geq (p^m(x,x))^k > 0$ and $p^{m+m'}(x,x) \geq p^m(x,x)p^{m'}(x,x) > 0$.

Now take any $n \geq K^2$ and write $n - K^2 = mK + r$ where $0 \leq r < K$. Then

$$n = K^2 + mK + r = r(K+1) + (K+m-r)K,$$

so $p^n(x,x) > 0$ by the observations above.

The second claim also follows from the observations above. ■

Our main convergence result is:

THM 23.16 (Convergence to equilibirum) *Let $\{X_n\}$ be an MC on countable set S with transition probability p. Assume it is irreducible, aperiodic and has stationary distribution π. Then for all $x, z \in S$*

$$p^n(x,z) \to \pi(z),$$

as $n \to +\infty$.

The proof is based on a technique called coupling. Before giving the proof, we begin with some background.

1.1 Coupling

A formal definition of coupling follows. Recall that for measurable spaces $(S_1, \mathcal{S}_1)$ $(S_2, \mathcal{S}_2)$, we can consider the product space $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ where

$$S_1 \times S_2 := \{(s_1, s_2) : s_1 \in S_1, s_2 \in S_2\}$$

is the Cartesian product of S_1 and S_2, and $\mathcal{S}_1 \times \mathcal{S}_2$ is the smallest σ-field $S_1 \times S_2$ containing the rectangles $A_1 \times A_2$ for all $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$.

DEF 23.17 (Coupling) *Let μ and ν be probability measures on the same measurable space $(S, \mathcal{S})$. A* coupling *of μ and ν is a probability measure γ on the product space $(S \times S, \mathcal{S} \times \mathcal{S})$ such that the* marginals *of γ coincide with μ and ν, i.e.,*

$$\gamma(A \times S) = \mu(A) \quad \textit{and} \quad \gamma(S \times A) = \nu(A), \qquad \forall A \in \mathcal{S}.$$

Here is an example.

EX 23.18 (Coupling of Bernoulli variables) *Let X and Y be Bernoulli random variables with parameters $0 \leq q < r \leq 1$ respectively. That is, $\mathbb{P}[X = 0] = 1 - q$ and $\mathbb{P}[X = 1] = q$, and similarly for Y. Here $S = \{0, 1\}$ and $\mathcal{S} = 2^S$.*

- *(Independent coupling) One coupling of X and Y is (X', Y') where $X' \stackrel{\mathrm{d}}{=} X$ and $Y' \stackrel{\mathrm{d}}{=} Y$ are* independent. *Its law is*

$$\Big(\mathbb{P}[(X', Y') = (i, j)]\Big)_{i,j \in \{0,1\}} = \begin{pmatrix} (1-q)(1-r) & (1-q)r \\ q(1-r) & qr \end{pmatrix}.$$

- *(Monotone coupling) Another possibility is to pick U uniformly at random in $[0, 1]$, and set $X'' = \mathbb{1}_{\{U \leq q\}}$ and $Y'' = \mathbb{1}_{\{U \leq r\}}$. The law of coupling (X'', Y'') is Then (X'', Y'') is a coupling of X and Y with law*

$$\Big(\mathbb{P}[(X'', Y'') = (i, j)]\Big)_{i,j \in \{0,1\}} = \begin{pmatrix} 1-r & r-q \\ 0 & q \end{pmatrix}.$$

One use of coupling is to quantify the "distance" between two measures. Let μ and ν be probability measures on $(S, \mathcal{S})$. The total variation distance between them is

$$\|\mu - \nu\|_{\mathrm{TV}} := \sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)|.$$

LEM 23.19 (Coupling inequality) *Let μ and ν be probability measures on $(S, \mathcal{S})$. For any coupling γ of μ and ν,*

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \mathbb{P}[X \neq Y],$$

where $(X, Y) \sim \gamma$.

Proof: For any $A \in \mathcal{S}$,

$$\begin{aligned}
\mu(A) - \nu(A) &= \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \\
&= \mathbb{P}[X \in A,\ X = Y] + \mathbb{P}[X \in A,\ X \neq Y] \\
&\qquad - \mathbb{P}[Y \in A,\ X = Y] - \mathbb{P}[Y \in A,\ X \neq Y] \\
&= \mathbb{P}[X \in A,\ X \neq Y] - \mathbb{P}[Y \in A,\ X \neq Y] \\
&\leq \mathbb{P}[X \neq Y],
\end{aligned}$$

and, similarly, $\nu(A) - \mu(A) \leq \mathbb{P}[X \neq Y]$. Hence

$$|\mu(A) - \nu(A)| \leq \mathbb{P}[X \neq Y].$$

■

A *coupling of Markov chains* with transition probability p is a Markov chain $\{(X_n, Y_n)\}$ on $S \times S$ such that both $\{X_n\}$ and $\{Y_n\}$ are Markov chains with transition probability p. For our purposes, the following special type of coupling will suffice.

DEF 23.20 (Markovian coupling) *A* Markovian coupling *of a transition probability p is a Markov chain $\{(X_n, Y_n)\}$ on $S \times S$ with transition probability q satisfying:*

- (Markovian coupling) *For all $x, y, x', y' \in S$,*

$$\sum_{z'} q((x, y), (x', z')) = p(x, x'),$$

$$\sum_{z'} q((x, y), (z', y')) = p(y, y').$$

We say that a Markovian coupling is coalescing *if further:*

- (Coalescing) *For all $z \in S$,*

$$x' \neq y' \implies q((z, z), (x', y')) = 0.$$

Note that not every coupling of Markov chains is itself Markovian.

Let $\{(X_n, Y_n)\}$ be a coalescing Markovian coupling of p. By the coalescing condition, if $X_m = Y_m$ then $X_n = Y_n$ for all $n \geq m$. That is, once $\{X_n\}$ and $\{Y_n\}$ meet, they remain equal. Let τ_{meet} be the *coalescence time* (also called coupling time), i.e.,

$$\tau_{\mathrm{meet}} = \inf\{n \geq 0 \,:\, X_n = Y_n\}.$$

By the coupling inequality, for any distributions μ_x and μ_y,

$$\begin{aligned}\left\| \sum_{z \in S} \mu_x(z) p^n(z, \cdot) - \sum_{z \in S} \mu_y(z) p^n(z, \cdot) \right\|_{\mathrm{TV}} &\leq \mathbb{P}_{\mu_x \times \mu_y}[X_n \neq Y_n] \\ &= \mathbb{P}_{\mu_x \times \mu_y}[\tau_{\mathrm{meet}} > n], \qquad (2)\end{aligned}$$

where $\times$ indicates the product measure.

1.2 Proof of convergence theorem

We are now ready to go back to THM 23.16.
Proof:(of THM 23.16) By definition of the total variation distance, it suffices to prove that

$$\|p^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \to 0.$$

By the definition of stationarity, this is equivalent to

$$\left\| p^n(x,\cdot) - \sum_{z\in S} \pi(z)\, p^n(z,\cdot) \right\|_{\mathrm{TV}} \to 0. \tag{3}$$

We use the coupling inequality. Let $\{(X_n, Y_n)\}$ be a coalescing Markovian coupling of p with transition probability q defined as:

$$q((x,y),(x',y')) = \begin{cases} p(x,x')\, p(y,y') & \text{if } x \neq y, \\ p(x,x') & \text{if } x = y \text{ and } x' = y', \\ 0 & \text{o.w.} \end{cases}$$

In words, $\{X_n\}$ and $\{Y_n\}$ are independent with transition probability p until they meet, at which point they remain equal from then on. Assume that $X_0 = x$ and that $Y_0 \sim \pi$, that is, the initial distribution of $\{(X_n, Y_n)\}$ is $\delta_x \times \pi$ (where δ_x is the unit mass at x). By (2), to show (3), it then suffices to show

$$\mathbb{P}_{\delta_x \times \pi}[\tau_{\mathrm{meet}} > n] \to 0. \tag{4}$$

This is implied by $\mathbb{P}_{\delta_x \times \pi}[\tau_{\mathrm{meet}} < +\infty] = 1$, for which it suffices in turn to prove

$$\mathbb{P}_{(x,y)}[\tau_{\mathrm{meet}} < +\infty] = 1, \qquad \forall y, \tag{5}$$

since

$$\mathbb{P}_{\delta_x \times \pi}[\tau_{\mathrm{meet}} < +\infty] = \sum_{y\in S} \pi(y)\, \mathbb{P}_{(x,y)}[\tau_{\mathrm{meet}} < +\infty].$$

It remains to prove the claim:

CLAIM 23.21 (Coupling in finite time) *For any x, y, $\mathbb{P}_{(x,y)}[\tau_{\mathrm{meet}} < +\infty] = 1$.*

Proof: We consider a second Markovian coupling (this time not coalescing). Let $\{Y'_n\}$ be an independent copy of $\{Y_n\}$ started at y and let

$$\tau'_{\mathrm{meet}} = \inf\{n \geq 0 \,:\, X_n = Y'_n\}.$$

By construction τ_{meet} and τ'_{meet} are identically distributed, hence $\mathbb{P}_{(x,y)}[\tau_{\text{meet}} < +\infty] = \mathbb{P}_{(x,y)}[\tau'_{\text{meet}} < +\infty]$. Fix a state $a \in S$ and let

$$\tau'_a = \inf\{n \geq 0 \,:\, X_n = Y'_n = a\}.$$

Observe that $\tau'_{\text{meet}} \leq \tau'_a$, hence $\mathbb{P}_{(x,y)}[\tau'_{\text{meet}} < +\infty] \geq \mathbb{P}_{(x,y)}[\tau'_a < +\infty]$. As a result of the previous two observations, it suffices to show

$$\mathbb{P}_{(x,y)}[\tau'_a < +\infty] = 1. \tag{6}$$

This follows immediately from LEM 23.5 and the next lemma (where we really only need irreducibility and recurrence in the conclusion).

LEM 23.22 *Let $\{(X_n, Y'_n)\}$ be a Markovian coupling of p where $\{X_n\}$ and $\{Y'_n\}$ are independent. Assume p is irreducible, aperiodic and positive recurrent. Then $\{(X_n, Y'_n)\}$ is irreducible, aperiodic and positive recurrent.*

Proof: Let r be the transition probability of $\{(X_n, Y'_n)\}$.

1. $\{(X_n, Y'_n)\}$ **is irreducible and aperiodic**. By irreducibility, for all x_1, x_2, y_1, y_2, there is K and L such that $p^K(x_1, x_2) > 0$ and $p^L(y_1, y_2) > 0$. By aperiodicity, we also have $p^n(x_2, x_2) > 0$ and $p^n(y_2, y_2) > 0$ for all $n \geq n_0$ for some n_0. Hence, for all $n \geq n_0$, by Chapman-Kolmogorov

$$\begin{aligned}
&r^{K+L+n}((x_1, y_1), (x_2, y_2)) \\
&= p^{K+L+n}(x_1, x_2)\, p^{K+L+n}(y_1, y_2) \\
&\geq p^K(x_1, x_2)\, p^{L+n}(x_2, x_2)\, p^L(y_1, y_2)\, p^{K+n}(y_2, y_2) \\
&> 0.
\end{aligned}$$

   One such n suffices to establish irreducibility. Moreover, since we can take $x_1 = x_2$ and $y_1 = y_2$ (and $K = L = 0$), we also have aperiodicity.

2. $\{(X_n, Y'_n)\}$ **is positive recurrent**. The probability measure $\pi \times \pi$ is stationary for r. Indeed, note

$$\begin{aligned}
&\sum_{x_1, y_1} \pi(x_1)\, \pi(y_1)\, r((x_1, y_1), (x_2, y_2)) \\
&= \sum_{x_1, y_1} \pi(x_1)\, \pi(y_1)\, p(x_1, x_2)\, p(y_1, y_2) \\
&= \sum_{x_1} \pi(x_1)\, p(x_1, x_2) \sum_{y_1} \pi(y_1)\, p(y_1, y_2) \\
&= \pi(x_2)\, \pi(y_2).
\end{aligned}$$

   By THM 23.10, $\{(X_n, Y'_n)\}$ is therefore positive recurrent.

That concludes the proof of the lemma. ■ ■ ■

2 Law of large numbers for MCs

Our second asymptotic result is a law of large numbers for countable MCs. This time, we do not need aperiodicity.

THM 23.23 (Law of large numbers for MCs) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. Assume it is irreducible and has stationary distribution π. Let $f : S \to \mathbb{R}$ be a function such that $\sum_{z \in S} |f(z)|\pi(z) < +\infty$. Then for any initial distribution μ, we have*

$$\frac{1}{n}\sum_{m=1}^{n} f(X_n) \to \sum_{z\in S} f(z)\pi(z),$$

almost surely as $n \to +\infty$.

We first prove the result in the special case where $f(x) = \mathbb{1}\{x = y\}$, in which case the sum on the LHS above counts the frequency of visits to y and the limit on the RHS is simply $\pi(y)$. The proof relies on the strong Markov property to break up the sample path into i.i.d. excursions from y back to it. That reduces the problem to an application of the standard Strong Law of Large Numbers.

2.1 Excursions

Recall that, for $y \in S$, we let $T_y^0 = 0$ and, for $k \geq 1$, we let

$$T_y^k = \inf\{n > T_y^{k-1} \,:\, X_n = y\},$$

be the time of the k-th return to y. For $k \geq 1$, we also let $\Delta_y^k = T_y^k - T_y^{k-1}$ be the k-th inter-visit time.

LEM 23.24 *Assume that y is recurrent. Under the initial distribution δ_y, the vectors (of random length)*

$$V^k := \left(\Delta_y^k, X_{T_y^{k-1}}, \ldots, X_{T_y^k - 1}\right), \qquad k \geq 1,$$

are i.i.d.

Proof: By recurrence of y, we have $T_y^k < +\infty$ almost surely for all $k \geq 0$. Hence the vectors V^k have finite length a.s. Because each component of V^k is in a countable set, the set of all possible values V^k can take is then itself countable. Let us denote this set by $\mathcal{V}$. Fix $v \in \mathcal{V}$. We use the strong Markov property with $h_n(X_0, X_1, \ldots) = h(X_0, X_1, \ldots) := \mathbb{1}\{V^1 = v\}$ for all n. Then

$$\phi_n(y) = \phi(y) := \mathbb{E}_y[h(X_0, X_1, \ldots)] = \mathbb{P}_y[V^1 = v].$$

And THM 23.1 implies that, for all $k \geq 1$, on $\{T_y^{k-1} < +\infty\}$,

$$\begin{aligned}\mathbb{P}_y[V^k = v \,|\, \mathcal{F}_{T_y^{k-1}}] &= \mathbb{E}_y[h(X_{T_y^{k-1}}, X_{T_y^{k-1}+1}, \ldots) \,|\, \mathcal{F}_{T_y^{k-1}}] \\ &= \mathbb{P}_y[V^1 = v],\end{aligned}$$

almost surely, where we used that $X_{T_y^{k-1}} = y$ by definition. Since this is true for every $v \in \mathcal{V}$, that implies that V^k is independent of $\mathcal{F}_{T_y^{k-1}}$ and therefore of $V^1, \ldots, V^{k-1}$. It also implies that the laws of the V^k's are identical. ■

Using the strong Markov property again, we get the following generalization:

THM 23.25 (Excursions) *Assume that y is recurrent. Under any initial distribution μ such that $\mathbb{P}_\mu[T_y^1 < +\infty] > 0$, conditioned on the event $\{T_y^1 < +\infty\}$, the vectors*

$$V^k := \left(\Delta_y^k, X_{T_y^{k-1}}, \ldots, X_{T_y^k - 1}\right), \qquad k \geq 2,$$

are i.i.d.

2.2 Asymptotic frequencies

We are now ready to prove the law of large numbers in the special case where $f(x) = \mathbb{1}\{x = y\}$. Let

$$N_n(y) = \sum_{m=1}^{n} \mathbb{1}\{X_m = y\},$$

be the number of visits to y by time n (not counting time 0). Again, we first consider the case where $\mu = \delta_y$.

LEM 23.26 *Assume that y is recurrent. Under the initial distribution δ_y,*

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y[T_y^+]},$$

almost surely as $n \to +\infty$.

Recall that when a stationary distribution π exists and is unique, then $\pi(y) = \frac{1}{\mathbb{E}_y[T_y^+]}$.

Proof:(of LEM 23.26) Writing, for $k \geq 1$,

$$T_y^k = \sum_{\ell=1}^{k} \Delta_y^\ell,$$

as a sum of i.i.d. RVs by LEM 23.24, the strong law of large numbers (THM 23.11 and 23.12) then implies that

$$\frac{T_y^n}{n} \to \mathbb{E}_y[T_y^1], \tag{7}$$

almost surely as $n \to +\infty$. This is not quite what we want. To relate T_y^k and $N_n(y)$ we note that by definition

$$T_y^{N_n(y)} \leq n < T_y^{N_n(y)+1},$$

and we use the sandwiching inequalities

$$\frac{T_y^{N_n(y)}}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{T_y^{N_n(y)+1}}{N_n(y)+1}\frac{N_n(y)+1}{N_n(y)}.$$

By the recurrence of y, we have $N_n(y) \to +\infty$ through the non-negative integers as $n \to +\infty$, almost surely. By (7), we get

$$\frac{n}{N_n(y)} \to \mathbb{E}_y[T_y^1],$$

almost surely as $n \to +\infty$. Taking inverses concludes the proof. ■

Using THM 23.25 (and noting that on $\{T_y^+ < +\infty\}$ we have $T_y^+/n \to 0$ a.s.), we get the following generalization:

THM 23.27 (Asymptotic frequencies) *Assume that y is recurrent. Under any initial distribution μ,*

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y[T_y^+]}\mathbb{1}\{T_y^+ < +\infty\},$$

almost surely as $n \to +\infty$.

2.3 Proof of law of large numbers for MCs

Proof:(of THM 23.23) Fix $y \in S$. Recall the definition of γ_y from (1). Let

$$W_f^k := \sum_{m=T_y^{k-1}}^{T_y^k - 1} f(X_m).$$

Because W_f^k is a function of V^k, THM 23.25 along with irreducibility and recurrence implies:

LEM 23.28 *Under any initial distribution* μ*, the RVs* $\{W_f^k, k \geq 2\}$ *are i.i.d. Moreover* $\mathbb{E}_\mu|W_f^k| < +\infty$.

Proof: To prove the second claim, we note that

$$\mathbb{E}_\mu|W_f^k| \leq \mathbb{E}_\mu[W_{|f|}^k].$$

For $v = (\delta, x_0, \ldots, x_{\delta-1}) \in \mathcal{V}$, define

$$|F|(v) := \sum_{m=0}^{\delta-1} |f(x_m)|.$$

THM 23.25 also implies that

$$\mathbb{E}_\mu[W_{|f|}^k] = \sum_{v=(\delta,x_0,\ldots,x_{\delta-1})\in\mathcal{V}} \mathbb{P}_y[V^1 = v]\,|F|(v).$$

Using γ_y, THM 23.8 and 23.10, we can re-write this as

$$\begin{aligned}
&\sum_{v=(\delta,x_0,\ldots,x_{\delta-1})\in\mathcal{V}} \mathbb{P}_y[V^1 = v]\,|F|(v) \\
&= \sum_{v=(\delta,x_0,\ldots,x_{\delta-1})\in\mathcal{V}} \mathbb{P}_y[V^1 = v] \sum_{z\in S} |f(z)| \sum_{m=0}^{\delta-1} \mathbb{1}\{x_m = z\} \\
&= \sum_{z\in S} |f(z)| \sum_{v=(\delta,x_0,\ldots,x_{\delta-1})\in\mathcal{V}} \mathbb{P}_y[V^1 = v] \sum_{m=0}^{\delta-1} \mathbb{1}\{x_m = z\} \\
&= \sum_{z\in S} |f(z)|\,\gamma_y(z) \\
&= \mathbb{E}_y[T_y^+] \sum_{z\in S} |f(z)|\,\pi(z) < +\infty,
\end{aligned}$$

by assumption. ■

Hence, by the strong law of large numbers (and repeating the calculation in the proof of the lemma above with f rather $|f|$), we have almost surely (once n is large enough that $N_n(y) \geq 2$)

$$
\begin{aligned}
&\frac{1}{N_n(y)-1} \sum_{m=1}^{T_y^{N_n(y)}-1} f(X_m) \\
&= \frac{1}{N_n(y)-1} \sum_{m=1}^{T_y^1-1} f(X_m) + \frac{1}{N_n(y)-1} \sum_{m=T_y^1}^{T_y^{N_n(y)}-1} f(X_m) \\
&\to \mathbb{E}_y[T_y^+] \sum_{z\in S} f(z)\,\pi(z),
\end{aligned}
$$

as $n \to +\infty$, where we used that the first term on the second line converges to 0. By THM 23.27, we then get:

LEM 23.29 *We have*

$$
\frac{1}{n} \sum_{m=1}^{T_y^{N_n(y)}-1} f(X_m) \to \sum_{z\in S} f(z)\,\pi(z),
$$

almost surely, as $n \to +\infty$.

This is still not what we want because the sum above stops at $m = T_y^{N_n(y)} - 1$. To argue that we can go all the way to n without affecting the limit, we appeal to the following technical observation.

LEM 23.30 *Let $Y_1, Y_2, \ldots$ be i.i.d. RVs such that $\mathbb{E}[Y_1^+] < +\infty$. Then*

$$
\frac{1}{n} \max_{1\leq i\leq n} Y_i \to 0,
$$

almost surely.

Proof: For any $\varepsilon > 0$, by the integrability assumption

$$
\sum_{n\geq 1} \mathbb{P}[Y_n^+ > \varepsilon n] < +\infty,
$$

so by BC1 there is N (random) large enough so that $Y_m^+ \leq \varepsilon m$ for all $m \geq N$. Hence for $n \geq N$

$$
\begin{aligned}
\frac{1}{n} \max_{1\leq i\leq n} Y_i &\leq \left(\frac{1}{n} \max_{1\leq m\leq N} Y_m\right) \vee \left(\max_{N+1\leq m\leq n} \frac{\varepsilon m}{n}\right) \\
&\leq \left(\frac{1}{n} \max_{1\leq i\leq N} Y_i\right) \vee \varepsilon \\
&\to \varepsilon,
\end{aligned}
$$

as $n \to +\infty$, almost surely. Since $\varepsilon > 0$ is arbitrary, we have shown that

$$\limsup_n \frac{1}{n} \max_{1 \leq i \leq n} Y_i \leq 0,$$

almost surely. On the other hand,

$$\liminf_n \frac{1}{n} \max_{1 \leq i \leq n} Y_i \geq \liminf_n \frac{1}{n} Y_1 = 0,$$

with probability one. ■

We then note that, since $N_n(y) \leq n$,

$$\left| \frac{1}{n} \sum_{m=1}^{T_y^{N_n(y)}-1} f(X_m) - \frac{1}{n} \sum_{m=1}^{n} f(X_m) \right| \leq \frac{1}{n} \max_{2 \leq k \leq n+1} W_{|f|}^k,$$

which tends to 0 almost surely by LEM 23.30. (Note that we use a maximum over the first n excursion sums because the behavior of the "completed last excursion" before time n is in itself not straightforward to characterize.)

The proof is then concluded by LEM 23.29. ■

References

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[Nor98] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.