Notes 21 : Markov chains: definitions, properties

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

References: [Dur10, Sections 6.1-4] and [Nor98, Sections 1.1-6].

We will need:

THM 21.1 (Monotone class theorem) *Let $\mathcal{H}$ be a class of bounded functions from a set S to $\mathbb{R}$ satisfying:*

1. *$\mathcal{H}$ is a vector space over $\mathbb{R}$.*
2. *The constant 1 is an element of $\mathcal{H}$.*
3. *If $(f_n)_n$ is a sequence of non-negative functions in $\mathcal{H}$ such that $f_n \uparrow f$ where f is a bounded function on S, then $f \in \mathcal{H}$.*

Then if $\mathcal{H}$ contains the indicator function of every set in some π-system $\mathcal{I}$, then $\mathcal{H}$ contains every bounded $\sigma(\mathcal{I})$-measurable function on S.

See Theorem 6.1.3 in [Dur10].

LEM 21.2 (Conditioning on an independent RV) *Let X and Y be RVs taking values in $(S_x, \mathcal{S}_x)$ and $(S_y, \mathcal{S}_y)$ respectively. Suppose $\mathcal{F}$ and Y are independent. Let $X \in \mathcal{F}$ and $\phi : S_x \times S_y \to \mathbb{R}$ be a bounded measurable function. Define $g(x) = \mathbb{E}(\phi(x, Y))$. Then,*

$$\mathbb{E}(\phi(X, Y)|\mathcal{F}) = g(X).$$

Proof: If $\phi(x, y)$ is of the form $\mathbb{1}\{(x, y) \in A \times B\}$ for sets $A, B \in \mathcal{S}$, then for any $C \in \mathcal{F}$,

$$\begin{aligned}
\mathbb{E}[\phi(X, Y); C] &= \mathbb{P}[\{X \in A, Y \in B\} \cap C] \\
&= \mathbb{P}[(\{X \in A\} \cap C) \cap \{Y \in B\}] \\
&= \mathbb{P}[\{X \in A\} \cap C]\, \mathbb{P}[Y \in B] \\
&= \mathbb{E}[\mathbb{1}\{X \in A\}\mathbb{P}[Y \in B]; C] \\
&= \mathbb{E}[g(X); C],
\end{aligned}$$

since indeed $g(x) = \mathbb{E}[\phi(x, Y)] = \mathbb{E}[\mathbb{1}\{x \in A\}\mathbb{1}\{Y \in B\}] = \mathbb{1}\{x \in A\}\mathbb{P}[Y \in B]$. Because sets of the form $A \times B$ are a π-system that contains Ω and generates the product σ-field, the monotone class theorem (together with bounded convergence) gives the result if one takes $\mathcal{H}$ to be those functions ϕ for which the equality in the statement holds. ∎

1 Definition

Markov chains are an important class of stochastic processes, with many applications. We will restrict ourselves here to the temporally-homogeneous discrete-time case. The main definition follows.

DEF 21.3 (Markov chain) *Let $(S, \mathcal{S})$ be a measurable space. A function $p : S \times \mathcal{S} \to \mathbb{R}$ is said to be a* transition kernel *if:*

1. *For each $x \in S$, $A \to p(x, A)$ is a probability measure on $(S, \mathcal{S})$.*
2. *For each $A \in \mathcal{S}$, $x \to p(x, A)$ is a measurable function.*

We say that $\{X_n\}_{n \geq 0}$ is a Markov chain (MC) *with transition kernel p if*

$$\mathbb{P}[X_{n+1} \in B \,|\, \mathcal{F}_n] = p(X_n, B) \qquad \forall B \in \mathcal{S}. \tag{1}$$

We refer to the law of X_0 as the initial distribution.

The key in (1) is that the RHS depends only on X_n—not on the full history up to time n.

We have already encountered many examples.

EX 21.4 (Random walk in $\mathbb{R}^d$) *Let $Z_1, Z_2, \ldots$ be i.i.d. $\mathbb{R}^d$-valued RVs with distribution μ. Define $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$. For some fixed $x \in \mathbb{R}^d$, let $X_0 = z_0$ and $X_n = z_0 + Z_1 + \cdots + Z_n$. Then:*

CLAIM 21.5 *The process $\{X_n\}$, as defined above, is an MC with transition kernel*

$$p(x, B) = \mu(B - x),$$

where, for any $B \in \mathcal{B}(\mathbb{R}^d)$, we define $B - x := \{y - x : y \in B\}$.

Proof: *We use LEM 21.2 with $\mathcal{F} = \mathcal{F}_n$, $X = X_n$, $Y = Z_{n+1}$, $\phi(x, y) = \mathbb{1}\{x+y \in B\}$ and*

$$g(x) = \mathbb{E}[\phi(x, Y)] = \mathbb{P}[x + Z_{n+1} \in B] = \mathbb{P}[Z_{n+1} \in B - x] = \mu(B - x),$$

by definition of μ. Note that Z_{n+1} is indeed independent of $\mathcal{F}_n$. Note also that $p(x, B) = \mu(B - x) = \mathbb{E}[\phi(x, Y)]$ is measurable for any x by Fubini's theorem. Then the conclusion of LEM 21.2 reads

$$\mathbb{E}[\mathbb{1}\{X_n + Z_{n+1} \in B\} \,|\, \mathcal{F}_n] = \mu(B - X_n),$$

or put differently

$$\mathbb{P}[X_{n+1} \in B \,|\, \mathcal{F}_n] = p(X_n, B),$$

as desired. ■

EX 21.6 (Countable-space MCs) *Let S be countable with $\mathcal{S} = 2^S$, let μ be a probability measure on S, and let $p : S \times S \to [0, 1]$ such that*

$$\sum_{j \in S} p(i, j) = 1, \qquad \forall i \in S.$$

That is, for each i, $p(i, \cdot)$ is a probability distribution on S. We extend the notation to sets as follows: $p(x, B) = \sum_{j \in B} p(x, j)$ (which is measurable as a function of x since its values are discrete). We construct a MC on S whose transition kernel is p. (Alternatively, we could use Kolmogorov's extension theorem. See Theorem 6.1.1 in [Dur10].) Consider the array

$$\{Z(n, i) \,:\, n \geq 0, i \in S\},$$

where the entries are independent and for each n and i

$$Z(n, i) \sim p(i, \cdot).$$

Then pick $X_0 \sim \mu$ and define by induction

$$X_n = Z(n, X_{n-1}).$$

We also let $\mathcal{F}_n = \sigma(X_0, Z(1, \cdot), \ldots, Z(n, \cdot))$.

CLAIM 21.7 *The process $\{X_n\}$, as defined above, is an MC with transition kernel p.*

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

$$\mathbb{P}[X_{n+1} \in B \,|\, \mathcal{F}_n] = \mathbb{P}[Z(n+1, X_n) \in B \,|\, X_0, Z(1, \cdot), \ldots, Z(n, \cdot)].$$

To compute the RHS, we use again LEM 21.2. Take $X = X_n$, $Y(\cdot) = Z(n+1,\cdot)$, $\phi(x,y) = \mathbb{1}\{y(x) \in B\}$, $\mathcal{F} = \mathcal{F}_n$, and

$$g(x) = \mathbb{E}[\phi(x,Y)] = \mathbb{P}[Z(n+1,x) \in B] = p(x,B).$$

Then by LEM 21.2,

$$\begin{aligned}\mathbb{P}[Z(n+1,X_n) \in B \,|\, X_0, Z(1,\cdot),\ldots,Z(n,\cdot)] &= \mathbb{E}[\phi(X,Y) \,|\, \mathcal{F}] \\ &= g(X) \\ &= p(X_n,B),\end{aligned}$$

as desired. ■

The last example includes many important special cases we have seen in previous lectures. We refer to the $p(i,j)$s as *transition probabilities*.

EX 21.8 (Branching process) *Let $\{q_i\}_{i\geq 0}$ be a probability distribution on non-negative integers and let $\{Z_m\}$ be i.i.d. with distribution $\{q_i\}_{i\geq 0}$. Then, the MC $\{X_n\}$ on $S = \{0,1,\ldots\}$ with transition probability*

$$p(i,j) = \mathbb{P}\left[\sum_{m=1}^{i} Z_m = j\right],$$

is a branching process *with offspring distribution $\{q_i\}_{i\geq 0}$.*

EX 21.9 (Birth-death chain) *An MC on $S = \{0,1,\ldots\}$ with the restriction that $p(i,j) = 0$ if $|i-j| > 1$ is called a* birth-death chain*. The standard notation is*

$$p(i,i+1) = p_i, \quad p(i,i-1) = q_i, \quad p(i,i) = r_i,$$

where $q_0 = 0$.

2 The Markov property and some formulas

An advantage of MCs is that they admit simple formulas. We first extend the defining property of MCs. We will often indicate the initial distribution μ with a subscript: $\mathbb{P}_\mu$, $\mathbb{E}_\mu$. We will also use the notation $\mathbb{P}_x$, $\mathbb{E}_x$ for the case where the initial distribution is a point mass at $x \in S$.

LEM 21.10 (One-step expectation) *Let $\{X_n\}$ be an MC on S with transition kernel p and initial distribution μ. Then for any bounded measurable function $f : S \to \mathbb{R}$*

$$\mathbb{E}[f(X_{n+1}) \,|\, \mathcal{F}_n] = \int f(y)\, p(X_n, \mathrm{d}y).$$

Proof: Let $\mathcal{H}$ be the set of bounded functions for which the identity holds. The result follows from the definition of an MC and the monotone class theorem. ∎

LEM 21.11 (Finite-dimensional distributions) *Let $\{X_n\}$ be an MC on S with transition kernel p and initial distribution μ. Then, for any bounded measurable function $f_k : S \to \mathbb{R}$, $k = 0, \ldots, m$,*

$$\mathbb{E}\left[\prod_{k=0}^{m} f_k(X_{n+k}) \,\middle|\, \mathcal{F}_n\right]$$
$$= f_0(X_n) \int f_1(x_{n+1}) p(X_n, \mathrm{d}x_{n+1}) \cdots \int f_m(x_{n+m}) p(x_{n+m-1}, \mathrm{d}x_{n+m}).$$

In particular

$$\mathbb{E}\left[\prod_{k=0}^{m} f_k(X_k)\right]$$
$$= \int f_0(x_0)\mu(\mathrm{d}x_0) \int f_1(x_1) p(x_0, \mathrm{d}x_1) \cdots \int f_n(x_n) p(x_{n-1}, \mathrm{d}x_n).$$

Proof: By the previous lemma and the tower property

$$\mathbb{E}\left[\prod_{k=0}^{m} f_k(X_{n+k}) \,\middle|\, \mathcal{F}_n\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\prod_{k=0}^{m} f_k(X_{n+k}) \,\middle|\, \mathcal{F}_{n+m-1}\right] \,\middle|\, \mathcal{F}_n\right]$$
$$= \mathbb{E}\left[\prod_{k=0}^{m-1} f_k(X_{n+k})\, \mathbb{E}\left[f_m(X_{n+m}) \,|\, \mathcal{F}_{n+m-1}\right] \,\middle|\, \mathcal{F}_n\right]$$
$$= \mathbb{E}\left[\prod_{k=0}^{m-1} f_k(X_{n+k}) \int f_m(x_{n+m})\, p(X_{n+m-1}, \mathrm{d}x_{n+m}) \,\middle|\, \mathcal{F}_n\right].$$

We proceed by induction and note that the last step, a function of X_n only, is $\mathcal{F}_n$-measurable.

The second formula in the statement follows from the first one by taking $n = 0$ and taking an expectation over X_0. ∎

The previous result “extends to infinity,” in what is usually referred to as the *Markov property*.

THM 21.12 (Markov property) *Let $\{X_n\}$ be an MC on S with transition kernel p. Let h be a bounded measurable function from $S^{\mathbb{Z}_+}$ to $\mathbb{R}$. Then*

$$\mathbb{E}_\mu[h(X_n, X_{n+1}, \ldots) \,|\, \mathcal{F}_n] = \phi(X_n), \tag{2}$$

where

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

Again, note that the RHS in (2) is a function of X_n only. In fact, the statement says something stronger: given the history up to time n, the process "reboots" at X_n.
Proof: Let $g_k(x) = \mathbb{1}\{x \in A_k\}$ for $k = 0, \ldots, m$. Sets of the form $A_0 \times \cdots \times A_m$ (for arbitrary m) form a π-system generating the infinite product σ-field. Hence, by the monotone class theorem (where $\mathcal{H}$ is the class of bounded measurable functions h satisfying (2)), it suffices to show that for all $B \in \mathcal{F}_n$

$$\mathbb{E}_\mu\left[\prod_{k=0}^{m} g_k(X_{n+k}); B\right] = \mathbb{E}_\mu\left[\phi(X_n); B\right]$$

where

$$\phi(x) = \mathbb{E}_x\left[\prod_{k=0}^{m} g_k(X_k)\right].$$

This is precisely the content of LEM 21.11 above. ■

For indicators, LEM 21.11 immediately gives:

THM 21.13 *Let $\{X_n\}$ be an MC on S with transition kernel p. For all $B_0, \ldots, B_n \in \mathcal{S}$*

$$\mathbb{P}_\mu[X_0 \in B_0, \ldots, X_n \in B_n] = \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(x_0, \mathrm{d}x_1) \cdots \int_{B_n} p(x_{n-1}, \mathrm{d}x_n).$$

In the countable case, THM 21.13 immediately gives:

THM 21.14 (Probability of a sample path) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. Then for all $i_0, i_1, \ldots, i_n \in S$*

$$\mathbb{P}_\mu[X_0 = i_0, \ldots, X_n = i_n] = \mu(i_0) \prod_{m=1}^{n} p(i_{m-1}, i_m).$$

Also, the distribution at time n is a matrix product.

THM 21.15 (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.,

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

By induction, $\sum_{j \in S} p^n(i,j) = 1$. Another consequence of the previous theorem is:

THM 21.16 (Chapman-Kolmogorov) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. Then for all $n, m \geq 0$ and $i, k \in S$*

$$\mathbb{P}_i[X_{n+m} = k] = \sum_j \mathbb{P}_i[X_n = j]\, \mathbb{P}_j[X_m = k].$$

3 Strong Markov property

The Markov property extends to stopping times.

First recall:

DEF 21.17 *A random variable $T : \Omega \to \overline{\mathbb{Z}}_+ \equiv \{0, 1, \ldots, +\infty\}$ is called a* stopping time *if*

$$\{T = n\} \in \mathcal{F}_n,\ \forall n \in \overline{\mathbb{Z}}_+.$$

EX 21.18 *Let $\{A_n\}$ be an adapted process and $B \in \mathcal{B}$. Then*

$$T = \inf\{n \geq 0 \,:\, A_n \in B\},$$

is a stopping time.

DEF 21.19 ($\mathcal{F}_T$) *Let T be a stopping time. Denote by $\mathcal{F}_T$ the set of all events F such that $\forall n \in \overline{\mathbb{Z}}_+$*

$$F \cap \{T = n\} \in \mathcal{F}_n.$$

Our main result is the following.

THM 21.20 (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

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

Proof: The proof uses a standard trick: summing over the possible values of T. Let $A \in \mathcal{F}_T$. Then

$$\begin{aligned}
&\mathbb{E}_\mu[h_T(X_T, X_{T+1}, \ldots); A \cap \{T < +\infty\}] \\
&\quad = \sum_{n=0}^{+\infty} \mathbb{E}_\mu[h_n(X_n, X_{n+1}, \ldots); A \cap \{T = n\}] \\
&\quad = \sum_{n=0}^{+\infty} \mathbb{E}_\mu[\phi_n(X_n); A \cap \{T = n\}] \\
&\quad = \mathbb{E}_\mu[\phi_T(X_T); A \cap \{T < +\infty\}],
\end{aligned}$$

where we used THM 21.12 and $A \cap \{T = n\} \in \mathcal{F}_n$. ■

4 Recurrence

In this section, we assume that S is countable.

We generalize the notion of recurrence (which we earlier introduced in the context of RWs) to MCs. Let $\{X_n\}$ be an MC on a countable set S with transition probability p. For $x \in S$, let $T_x^0 = 0$ and for $k \geq 1$ let

$$T_x^k = \inf\{n > T_x^{k-1} \,:\, X_n = x\},$$

be the time of the k-th return to x.

LEM 21.21 *T_x^k is a stopping time.*

Proof: Arguing by induction on k,

$$\{T_x^k = n\} = \bigcup_{m=1}^{n-1} \Big[\{T_x^{k-1} = m\} \cap \{X_{m+1}, \ldots, X_{n-1} \neq x\} \cap \{X_n = x\}\Big] \in \mathcal{F}_n.$$

■

We sometimes use the notation $T_x^+ = T_x^1$. Let

$$\rho_{xy} = \mathbb{P}_x[T_y^+ < +\infty],$$

that is, ρ_{xy} is the probability of ever reaching y when started at x.

DEF 21.22 (Recurrence) *A state $x \in S$ is* recurrent *if $\rho_{xx} = 1$. Otherwise it is* transient.

Let

$$N(x) = \sum_{n \geq 1} \mathbb{1}\{X_n = x\},$$

be the number of visits to x (after time 0). To give further insights into the definition above, we prove:

THM 21.23 *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. If y is recurrent, then*

$$\mathbb{P}_y[X_n = y \text{ i.o.}] = 1.$$

THM 21.24 *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. If y is transient, then for any x*

$$\mathbb{E}_x[N(y)] < +\infty.$$

Combining the last two theorems implies in particular that

$$x \text{ is recurrent if and only if } \mathbb{E}_x[N(x)] = \sum_{n \geq 1} p^n(x, x) = +\infty, \tag{3}$$

where the first equality above comes from writing $N(x)$ as a sum of indicators. (Note that the first statement above is in fact stronger.) THM 21.23 and 21.24 are a consequence of the strong Markov property through the following formula. In words, for $\{T_y^k < +\infty\}$ to hold, we first need to visit y once and then come back to it $k-1$ times. Each of these visits is "independent conditioned on the previous one" by the strong Markov property.

LEM 21.25 (Probability of k-th return) *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. For all $x, y \in S$,*

$$\mathbb{P}_x[T_y^k < +\infty] = \rho_{xy}\rho_{yy}^{k-1}.$$

Proof: We argue by induction on k. For $k = 1$, the result holds by definition. Assume the result holds for $k \geq 1$ and let $h(z_0, z_1, \ldots)$ be 1 if $z_m = y$ for some $m \geq 1$. By the strong Markov property applied to T_y^{k-1}. Then on $\{T_y^{k-1} < +\infty\}$

$$\mathbb{P}_x \left[T_y^k < +\infty \,\middle|\, \mathcal{F}_{T_y^{k-1}} \right] = \mathbb{E}_x \left[h(X_{T_y^{k-1}}, X_{T_y^{k-1}+1}, \ldots) \,\middle|\, \mathcal{F}_{T_y^{k-1}} \right] = \phi(y),$$

where

$$\phi(y) = \mathbb{E}_y[h(X_0, X_1, \ldots)] = \mathbb{P}_y[T_y < +\infty].$$

As a consequence

$$\begin{aligned}
\mathbb{P}_x[T_y^k < +\infty] &= \mathbb{E}_x[\mathbb{1}\{T_y^k < +\infty\}; \{T_y^{k-1} < +\infty\}] \\
&= \mathbb{E}_x[\mathbb{P}_y[T_y < +\infty]; \{T_y^{k-1} < +\infty\}] \\
&= \mathbb{P}_y[T_y < +\infty]\, \mathbb{P}_x[T_y^{k-1} < +\infty] \\
&= \rho_{yy} \times \rho_{xy}\rho_{yy}^{k-2},
\end{aligned}$$

by induction, as desired. ■

We can now prove the theorems above.

Proof:(Proof of THM 21.23) If $\rho_{yy} = 1$, then by monotonicity

$$\begin{aligned}
1 &= \rho_{yy}^k \\
&= \mathbb{P}_y \left[T_y^k < +\infty \right] \\
&\downarrow \mathbb{P}_y \left[\bigcap_{k \geq 1} \{T_y^k < +\infty\} \right].
\end{aligned}$$

■

Proof:(Proof of THM 21.24) By LEM 21.25,

$$\begin{aligned}
\mathbb{E}_x[N(y)] &= \sum_{k \geq 1} \mathbb{P}_x[N(y) \geq k] \\
&= \sum_{k \geq 1} \mathbb{P}_x[T_y^k < +\infty] \\
&= \sum_{k \geq 1} \rho_{xy}\rho_{yy}^{k-1} \\
&= \frac{\rho_{xy}}{1 - \rho_{yy}}.
\end{aligned}$$

When $\rho_{yy} < 1$, the latter is $< +\infty$. ■

5 Class structure

Considerations about ρ_{xy} lead to a natural decomposition of the space, which in turn helps identify recurrent states. We begin with a combinatorial interpretation of the condition $\rho_{xy} > 0$. Recall the definition of p^n from THM 21.15.

LEM 21.26 *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* $r = 1, \ldots, n$

Proof: Note that, for $x \neq 1$ and $n \geq 1$,

$$p^n(x, y) \leq \mathbb{P}_x[T_y^+ < +\infty] = \sum_{m \geq 1} \mathbb{P}_x[T_y^+ = m] \leq \sum_{m \geq 1} p^m(x, y),$$

shows that "(a) is equivalent to (b)." Moreover

$$p^n(x, y) = \sum_{i_1, \ldots, i_{n-1}} p(x, i_1)\, p(i_1, i_2) \cdots p(i_{n-1}, y),$$

shows that "(b) is equivalent to (c)." ■

Define

$$T_x = \inf\{n \geq 0 \,:\, X_n = x\}.$$

Note that the infimum above starts at $n = 0$, unlike that in T_x^+. In particular, $T_x = T_x^+$ when started at $y \neq x$. 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.

DEF 21.27 (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.*

Clearly $x \leftrightarrow x$ (since $\mathbb{P}_x[T_x < +\infty] = 1$) and $x \leftrightarrow y$ implies $y \leftrightarrow x$ (by definition). Moreover, it follows from LEM 21.26 that $x \to y$ and $y \to z$ implies $x \to z$. In particular, the relation $\leftrightarrow$ is also transitive. As a result, it is an equivalence relation and its equivalence classes define a partition of S. (Recall that the equivalence class of x is $[x] = \{y \,:\, x \leftrightarrow y\}$.)

DEF 21.28 (Communicating classes) *The equivalence classes of the relation $\leftrightarrow$ are called* communicating classes.

Hence the communicating classes are also the maximal irreducible sets.

The following fact implies that recurrence is a property of a communicating class (i.e. a so-called *class property*).

LEM 21.29 (Recurrence is contagious) *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$.*

Proof: Since $\rho_{xy} > 0$, LEM 21.26 says that $\exists i_0 = x, i_1, \ldots, i_K = y \in S$ such that $p(i_{r-1}, i_r) > 0$ for all $r = 1, \ldots, K$. Take the smallest K such that this is the case. Observe that necessarily $i_1, \ldots, i_{K-1} \neq x$ (since otherwise we could have gotten a shorter sequence by starting at the last time x appears in this list—a contradiction). By the Markov property, it is "obvious" that

$$\mathbb{P}_x[T_x^+ = +\infty] \geq (1 - \rho_{yx}) \prod_{r=1}^{K} p(i_{r-1}, i_r),$$

since the product on the RHS is the probability of following the path $i_0, \ldots, i_K$ (which does not visit x after time 0) and from there the probability of not visiting x is $1 - \rho_{yx}$. (Formally, let $h(z_0, z_1, \ldots)$ be 1 if $z_k \neq x$ for all $k \geq 1$ and 0 otherwise. Let also $\phi(z) = \mathbb{E}_z[h(X_0, X_1, \ldots)] = \mathbb{P}_z[T_x^+ = +\infty]$. Then, by the Markov property,

$$\begin{aligned}
&\mathbb{P}_x[T_x^+ = +\infty] \\
&\geq \mathbb{P}_x[X_1 = i_1, \ldots, X_{K-1} = i_{K-1}, X_K = y, T_x^+ = +\infty] \\
&= \mathbb{E}_x[h(X_K, X_{K+1}, \ldots); \{X_1 = i_1, \ldots, X_{K-1} = i_{K-1}, X_K = y\}] \\
&= \mathbb{E}_x[\phi(y); \{X_1 = i_1, \ldots, X_{K-1} = i_{K-1}, X_K = y\}] \\
&= \mathbb{P}_y[T_x^+ = +\infty]\mathbb{P}_x[X_1 = i_1, \ldots, X_{K-1} = i_{K-1}, X_K = y] \\
&= (1 - \rho_{yx}) \prod_{r=1}^{K} p(i_{r-1}, i_r),
\end{aligned}$$

where we used that $\{X_1 = i_1, \ldots, X_{K-1} = i_{K-1}, X_K = y\} \in \mathcal{F}_K$.) The latter would be > 0 if $\rho_{yx} < 1$, a contradiction. This proves that $\rho_{yx} = 1$.

In addition, the lemma tells us that $\rho_{yx} = 1$ implies that $p^L(y, x) > 0$ for some $L \geq 0$. Applying Chapman-Kolmogorov (twice),

$$\sum_{n \geq 1} p^{L+n+K}(y, y) \geq \sum_{n \geq 1} p^L(y, x) p^n(x, x) p^K(x, y) = +\infty,$$

by (3), which also proves that y is recurrent. That concludes the proof. ■

THM 21.30 *If $\{X_n\}$ is an MC on a countable set S and $C \subseteq S$ is irreducible, then either all states in C are recurrent or all are transient.*

Proof: Follows from LEM 21.29. ■

One still needs to determine whether a class is recurrent. The following gives a useful criterion. First, a definition:

DEF 21.31 *A subset $C \subseteq S$ is* closed *of for all $x \in C$ and $y \notin C$, it holds that $\rho_{xy} = 0$.*

In other words, a subset C is closed if $\{X_n\}$ cannot "get out of it." Note that, by LEM 21.26, $\rho_{xy} = 0$ implies $p^n(x, y) = 0$ for all $n \geq 1$.

THM 21.32 *If $\{X_n\}$ is an MC on a countable set S and $C \subseteq S$ is recurrent communicating class, then C is closed.*

Proof: By LEM 21.29, if $x \in C$ and $y \notin C$ but $\rho_{xy} > 0$, then it would hold that $\rho_{yx} = 1$, a contradiction to the fact that C is a communicating class. ■

EX 21.33 (Branching processes: recurrence) *Let $\{X_n\}$ be a branching process with offspring distribution $\{q_i\}_{i\geq 0}$. Assume that $q_0 > 0$. That implies that $\rho_{k0} \geq q_0^k > 0$ for all $k > 0$, as there is a positive probability that no children is produced in any generation. On the other hand, $\rho_{0k} = 0$ for all $k > 0$, as 0 is a so-called* absorbing state*, i.e., $p(0, 0) = 1$. As a result 0 is a recurrent communicating class. On the other hand, all other states are transient. Indeed, for each $k > 0$, by the above, the communicating class of k is not closed.*

Things are simpler in the finite case.

THM 21.34 *Let $\{X_n\}$ be an MC on a countable set S. If $C \subseteq S$ is finite and closed, then it contains a recurrent state.*

Proof: We argue by contradiction. Suppose all $y \in C$ are transient. Then by THM 21.24, for any $x \in C$

$$+\infty > \sum_{y\in C} \mathbb{E}_x[N(y)] = \sum_{y\in C}\sum_{n\geq 1} p^n(x, y) = \sum_{n\geq 1}\sum_{y\in C} p^n(x, y) = \sum_{n\geq 1} 1,$$

by the closedness of C (and the facts, encountered previously, that $\sum_{y\in S} p^n(x, y) = 1$ and $p^n(x, y) = 0$ for all $y \notin C$). That is a contradiction. ■

THM 21.35 (Decomposition theorem) *If $\{X_n\}$ is an MC on a finite set S, then all closed communicating classes are recurrent. All other communicating classes are transient.*

Proof: Follows from THM 21.30, 21.32 and 21.34. ■

EX 21.36 (A seven-state chain) *See [Dur10, Example 6.4.1].*

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.