Notes 22 : Markov chains: stationary measures

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

References: [Dur10, Sections 6.5] and [Nor98, Sections 1.7].

Recall:

DEF 22.1 (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.

EX 22.2 (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). Previously, we showed how to construct an MC on S whose transition kernel is p.

THM 22.3 (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 22.4 (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 $N(x) = \sum_{n \geq 1} \mathbb{1}\{X_n = x\}$.

LEM 22.5 *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 22.6 (Recurrence) *A state $x \in S$ is* recurrent *if $\rho_{xx} = 1$. Otherwise it is* transient.

THM 22.7 *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 22.8 *Let $\{X_n\}$ be an MC on a countable set S with transition probability p. If y is transient then, for any x, $\mathbb{E}_x[N(y)] < +\infty$.*

Define

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

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 22.9 (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.*

LEM 22.10 (Recurrence is contagious) *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$.*

THM 22.11 *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.*

1 Stationary measures

Throughout we assume that S is countable.

The notion of stationary measure provides a more quantitative picture of the limit behavior of an MC. We first define it and discuss issues of existence and uniqueness. The connection to asymptotics is developed in the next section.

1.1 Definition

First the main definition:

DEF 22.12 (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.

The following observation explains the name.

LEM 22.13 *If μ is a stationary distribution, then for all $n \geq 0$*

$$\mathbb{P}_\mu[X_n = j] = \mu(j).$$

Proof: By THM 22.4,

$$\begin{aligned}
\mathbb{P}_\mu[X_n = j] &= \sum_{i \in S} \mu(i) p^n(i,j) \\
&= \sum_{i \in S} \mu(i) \sum_{k_1,\ldots,k_{n-1}} p(i,k_1)\, p(k_1,k_2) \cdots p(k_{n-1},j) \\
&= \sum_{k_1,\ldots,k_{n-1}} \sum_{i \in S} \mu(i) p(i,k_1)\, p(k_1,k_2) \cdots p(k_{n-1},j) \\
&= \sum_{k_1,\ldots,k_{n-1}} \mu(k_1)\, p(k_1,k_2) \cdots p(k_{n-1},j),
\end{aligned}$$

and the result follows by repeated application of the definition of stationarity. ■

EX 22.14 (Branching processes) *Let $\{X_n\}$ be a branching process with offspring distribution $\{q_i\}_{i \geq 0}$. Then $\mu(i) = \mathbb{1}\{i = 0\}$ is a stationary distribution. Indeed,*

$$\sum_{i=0}^{+\infty} \mathbb{1}\{i = 0\}\, p(i,j) = p(0,j) = \mathbb{1}\{j = 0\},$$

as 0 is an absorbing state.

EX 22.15 (Biased RW) *Recall that biased simple RW on $\mathbb{Z}$ with parameter $1/2 < q < 1$ is the process $\{S_n\}_{n\geq 0}$ with $S_0 = 0$ and $S_n = \sum_{k\leq n} X_k$ where the X_ks are iid in $\{-1,+1\}$ s.t. $\mathbb{P}[X_1 = 1] = q$. Let $r = 1 - q$. This process is an MC on $S = \mathbb{Z}$ with transition probability*

$$p(i,j) = \begin{cases} q, & \textit{if } j = i+1 \\ r, & \textit{if } j = i-1 \\ 0, & \textit{o.w.} \end{cases}$$

Then

$$\mu(i) = \left(\frac{q}{r}\right)^i,$$

is a stationary measure. Indeed,

$$\sum_{i\in\mathbb{Z}} \left(\frac{q}{r}\right)^i p(i,j) = q\left(\frac{q}{r}\right)^{j-1} + r\left(\frac{q}{r}\right)^{j+1} = (r+q)\times\left(\frac{q}{r}\right)^j = \left(\frac{q}{r}\right)^j,$$

as claimed. Note, however, that μ is not a stationary distribution; in fact,

$$\sum_{i\in S} \mu(i) = +\infty,$$

so it cannot be "normalized" into one.

EX 22.16 (Not every chain has a stationary measure) *Let $\{q_i\}_{i\geq 1}$ with $r_i = 1 - q_i$ such that*

$$\sum_{i\geq 1} r_i < 1.$$

Let $\{X_n\}$ be an MC on $S = \mathbb{Z}_+$ with $p(0,1) = q_0 = 1$ and, for $i \geq 1$,

$$p(i,i+1) = q_i \qquad \textit{and} \qquad p(i,0) = r_i.$$

If μ were a non-zero stationary measure it would satisfy

$$\begin{aligned} \mu(0) &= \sum_{j\geq 1} \mu(j)\, r_j \\ \mu(i) &= \mu(i-1)\, q_{i-1}, \qquad \forall i \geq 1. \end{aligned}$$

By induction on i,

$$\mu(i) = \mu(0) \prod_{1\leq j\leq i-1} q_j, \tag{2}$$

and plugging back above gives

$$\mu(0) = \mu(0) \left\{ \sum_{j\geq 1} (1-q_j) \prod_{1\leq k\leq j-1} q_k \right\}.$$

If $q_j = 1$ *for all* $j \geq 1$ *for instance, then the above reads* $\mu(0) = 0$ *and* $\mu(i) = \mu(0) = 0$ *for all* $i \geq 1$*. Hence, there is no stationary measure. More generally, if* $0 < \mu(0) < +\infty$

$$\mu(0) \leq \mu(0) \sum_{j\geq 1} (1-q_j) < \mu(0),$$

a contradiction. On the other hand, if $\mu(0) = 0$*, then* $\mu(i) = 0$ *for all* i *by* (2).

1.2 Existence

Our main existence result follows. For $x, y \in S$, let

$$\gamma_x(y) = \mathbb{E}_x \left[\sum_{n=0}^{T_x^+ - 1} \mathbb{1}\{X_n = y\} \right] = \sum_{n=0}^{+\infty} \mathbb{P}_x \left[X_n = y, n \leq T_x^+ - 1 \right], \qquad (3)$$

that is, $\gamma_x(y)$ is the expected number of visits to y before the first return to x, when started at (and including) x. Necessarily $\gamma_x(x) = 1$. Note also that in general $\gamma_x(y)$ could be 0 or $+\infty$.

THM 22.17 (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* $0 < \gamma_x(y) < +\infty$*.*

On the other hand, when all states are transient, EX 22.15 and 22.16 show that a (non-zero) stationary measure may or may not exist.

Proof:(of THM 22.17) We use what is called the *cycle trick*. Since $T_x^+ < +\infty$

a.s. and $X_0 = X_{T_x^+} = x$, we can "shift" the summation in (3) by one to get

$$\begin{aligned}
\gamma_x(y) &= \mathbb{E}_x\left[\sum_{n=1}^{T_x^+} \mathbb{1}\{X_n = y\}\right] \\
&= \mathbb{E}_x\left[\sum_{n=1}^{+\infty} \mathbb{1}\{X_n = y, n \leq T_x^+\}\right] \\
&= \sum_{n=1}^{+\infty} \mathbb{P}_x\left[X_n = y, n \leq T_x^+\right] \\
&= \sum_{n=1}^{+\infty}\sum_{z\in S} \mathbb{P}_x\left[X_{n-1} = z, X_n = y, n \leq T_x^+\right].
\end{aligned}$$

Note that $X_{n-1} = z, n \leq T_x^+ \in \mathcal{F}_{n-1}$ and $\mathbb{E}_z[\mathbb{1}\{X_1 = x\}] = p(z,x)$. Hence, by the Markov property at time $n-1$,

$$\begin{aligned}
&\mathbb{P}_x\left[X_{n-1} = z, X_n = y, n \leq T_x^+\right] \\
&\quad = \mathbb{E}_x[\mathbb{1}\{X_n = x\}; \{X_{n-1} = z, n \leq T_x^+ \in \mathcal{F}_{n-1}\}] \\
&\quad = p(z,x)\, \mathbb{P}_x\left[X_{n-1} = z, n \leq T_x^+\right].
\end{aligned}$$

Plugging this back above and changing the order of summation gives

$$\begin{aligned}
\gamma_x(y) &= \sum_{z\in S} p(z,y) \sum_{n=1}^{+\infty} \mathbb{P}_x\left[X_{n-1} = z, n \leq T_x^+\right] \\
&= \sum_{z\in S} p(z,y) \sum_{m=0}^{+\infty} \mathbb{P}_x\left[X_m = z, m \leq T_x^+ - 1\right] \\
&= \sum_{z\in S} p(z,y)\, \gamma_x(z).
\end{aligned}$$

For the second statement, note that $\rho_{xy} = 0$ implies by LEM 22.5 that $p^n(x,y) = 0$ for all $n \geq 0$, so that

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

On the other hand, by LEM 22.10, $\rho_{xy} > 0$ implies that $\rho_{yx} > 0$ which in turn implies by LEM 22.5 that there are K and L such that $p^K(x,y) > 0$ and $p^L(y,x) > 0$. Moreover, by induction, for all $n \geq 0$

$$\gamma_x(w) = \sum_{z\in S} \gamma_x(z)\, p^n(z,w).$$

Hence,

$$\gamma_x(y) \geq \gamma_x(x)\, p^K(x,y) > 0$$

and

$$1 = \gamma_x(x) \geq \gamma_x(y)\, p^L(y,x),$$

so that $\gamma_x(y) \leq 1/p^L(y,x) < +\infty$. That concludes the proof. ■

1.3 Uniqueness

In general, even if a stationary measure exists, it may not be unique. Observe first that the stationary measure condition is linear. Hence:

LEM 22.18 *If μ_1 and μ_2 are stationary measures, so is $\nu = \alpha\mu_1 + \beta\mu_2$ for any α, β such that ν is non-negative (and not identically zero).*

In particular, any (positive) constant multiple of a stationary measure is a stationary measure. But this is not the only source of non-uniqueness in general, as the next examples show.

EX 22.19 (Biased RW: continued) *Going back to EX 22.15, we claim that $\nu(i) = 1$ for all i is also a stationary measure (distinct from the measure μ defined there and not a constant multiple of it). Indeed,*

$$\sum_{i\in\mathbb{Z}} \nu(i)\, p(i,j) = 1 \times p(j-1,j) + 1 \times p(j+1,j) = q + r = 1 = \nu(j),$$

as claimed.

EX 22.20 (A seven-state chain) *Going back to [Dur10, Example 6.4.1], for each recurrent communicating class, it is straightforward to construct a stationary distribution supported exclusively on it (by solving the linear system).*

The latter example hints at a more general theory, which we will not develop here.

In the irreducible, recurrent case, our main uniqueness result is the following. Recall the definition of γ_x from (3).

THM 22.21 (Uniqueness of stationary measure) *Let $\{X_n\}$ be an MC on a countable set S. If μ is a stationary measure with $\mu(x) = 1$ for some $x \in S$, then $\mu \geq \gamma_x$. If moreover the chain is irreducible and recurrent, then $\mu = \gamma_x$.*

Proof: Let $y \in S$. To relate $\mu(y)$ to $\gamma_x(y)$, we repeatedly apply the condition satisfied by a stationary measure and "pull out" the terms in the sum defining $\gamma_x(y)$. Note that

$$\begin{aligned}
\mu(y) &= \sum_{z_0 \in S} \mu(z_0)\, p(z_0, y) \\
&= p(x, y) + \sum_{z_0 \neq x} \mu(z_0)\, p(z_0, y) \\
&= \mathbb{P}_x[X_1 = y, T_x^+ \geq 2] + \sum_{z_0 \neq x} \mu(z_0)\, p(z_0, y),
\end{aligned}$$

where on the second line we used $\mu(x) = 1$. Continuing on,

$$\begin{aligned}
\sum_{z_0 \neq x} \mu(z_0)\, p(z_0, y) &= \sum_{z_0 \neq x} \left(\sum_{z_1 \in S} \mu(z_1)\, p(z_1, z_0) \right) p(z_0, y) \\
&= \sum_{z_0 \neq x} p(x, z_0)\, p(z_0, y) + \sum_{z_0, z_1 \neq x} \mu(z_1)\, p(z_1, z_0)\, p(z_0, y) \\
&= \mathbb{P}_x[X_2 = y, T_x^+ \geq 3] + \sum_{z_0, z_1 \neq x} \mu(z_1)\, p(z_1, z_0)\, p(z_0, y).
\end{aligned}$$

Inductively, we get that

$$\mu(y) \geq \sum_{m=1}^{n} \mathbb{P}_x[X_n = y, T_x^+ \geq n+1] \uparrow \gamma_x(y),$$

as n tends to $+\infty$ by definition of γ_x. That proves the first statement.

Now, assume in addition that the chain is irreducible and recurrent. Because γ_x is then stationary by THM 22.17, the first statement and LEM 22.18 imply that $\nu = \mu - \gamma_x \geq 0$ is also stationary. We have further $\nu(x) = 0$ by assumption. Since $\{X_n\}$ is irreducible, for any $y \in S$ there is $K > 0$ such that $p^K(y, x) > 0$. But that implies (by the proof of LEM 22.13)

$$\begin{aligned}
0 &= \nu(x) \\
&= \sum_{z \in S} \nu(z)\, p^K(z, x) \\
&\geq \nu(y)\, p^K(y, x) \\
&\geq 0,
\end{aligned}$$

which is only possible if $\nu(y) = 0$, that is, $\mu(y) = \gamma_x(y)$. This is true for every $y \in S$ and that concludes the proof. ■

1.4 Stationary distributions and positive recurrence

So far, we have focused on the existence and uniqueness of a stationary *measure*. Here we address these questions for stationary distributions. A clean theory arises in this case. Let $\{X_n\}$ be a MC on a countable set S. For a fixed recurrent $x \in S$, in order for γ_x to be "normalizable" into a stationary distribution, it suffices that $\mathbb{E}_x[T_x^+] < +\infty$. Indeed, by Fubini,

$$\begin{aligned}\sum_{z\in S}\gamma_x(z) &= \sum_{z\in S}\mathbb{E}_x\left[\sum_{n=0}^{T_x^+-1}\mathbb{1}\{X_n = z\}\right]\\ &= \mathbb{E}_x\left[\sum_{n=0}^{T_x^+-1}\sum_{z\in S}\mathbb{1}\{X_n = z\}\right]\\ &= \mathbb{E}_x[T_x^+] < +\infty, \end{aligned} \tag{4}$$

and $\gamma_x/\mathbb{E}_x[T_x^+]$ is a stationary distribution (as it is stationary and sums to 1).

DEF 22.22 (Positive recurrence) *A recurrent state $x \in S$ is* positive recurrent *if $\mathbb{E}_x[T_x^+] < +\infty$. Otherwise it is* null recurrent*.*

THM 22.23 *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

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x^+]}.$$

Note that irreducibility is important here: in general the existence of a stationary distribution does not imply that all states are positive recurrent; in fact some can be transient (see EX 22.20). We can alternatively restrict the chain to a closed, irreducible set.

Proof:(of THM 22.23) Note that (i) trivially implies (ii). If (ii) holds, then (iii) holds by THM 22.17 and (4).

If (iii) holds, let π be a stationary distribution. Since π sums to 1, there is at least one $z \in S$ such that $\pi(z) > 0$. In fact, let $y \neq z$ be any other state in S; by irreducibility, there is K such that $p^K(z,y) > 0$ and LEM 22.13 implies that

$$\pi(y) = \sum_{w \in S} \pi(w)\, p^K(w,y) \geq \pi(z)\, p^K(z,y) > 0.$$

Hence π is strictly positive on S. Now fix $x \in S$ and let $\mu(w) = \pi(w)/\pi(x)$ for all $w \in S$. Note that $\mu(x) = 1$ and that μ is a stationary measure. By THM 22.21, $\mu \geq \gamma_x$. Hence

$$\mathbb{E}_x[T_x^+] = \sum_{w \in S} \gamma_x(w) \leq \sum_{w \in S} \frac{\pi(w)}{\pi(x)} = \frac{1}{\pi(x)} < +\infty. \tag{5}$$

Hence x is positive recurrent. Since x was arbitrary, we have proved that (iii) implies (i). Moreover, we have shown that the chain is recurrent and therefore $\mu = \gamma_x$ by THM 22.21. So the inequality in (5) is in fact an equality. That proves the last statement. ■

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.