Notes 24 : Markov chains: martingale methods

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

References: [Nor98, Sections 4.1-2], and [Ebe, Sections 0.3, 1.1-2, 3.1-2], [Bre20, Sections 7.1-7.3].

1 Martingale problem

Throughout, we assume that S is countable. The following linear operator associated to transition probability p arises naturally: for any bounded measurable function $f : S \to \mathbb{R}$, define

$$\mathcal{L}f(x) := \sum_{z \in S} p(x,z)[f(z) - f(x)],$$

which we refer to as the (discrete-time) generator associated to p. To see where this is coming from, note that if $\{X_n\}$ is an MC on S with transition probability p then

$$\mathbb{E}_x[f(X_1) - f(X_0)] = \mathcal{L}f(x) \tag{1}$$

and, by the Markov property, we also have

$$\mathbb{E}[f(X_{n+1}) - f(X_n) \,|\, \mathcal{F}_n] = \mathcal{L}f(X_n). \tag{2}$$

EX 24.1 (Simple random walk on $\mathbb{Z}$) *For simple random walk on $\mathbb{Z}$,*

$$\begin{aligned} \mathcal{L}f(x) &= \sum_{z \in S} p(x,z)[f(z) - f(x)] \\ &= \frac{1}{2}\{[f(x+1) - f(x)] - [f(x) - f(x-1)]\}, \end{aligned}$$

which is a discretized second derivative.

THM 24.2 (Martingale problem) *Let S be countable, let $\{X_n\}$ be a stochastic process adapted to $\{\mathcal{F}_n\}$ and taking values in S, and let p be a transition probability on S with associated generator $\mathcal{L}$. Then the following are equivalent:*

(i) The process $\{X_n\}$ is a Markov chain with transition probability p.

(ii) For any bounded measurable function $f : S \to \mathbb{R}$, the process

$$M_n^f = f(X_n) - \sum_{m=0}^{n-1} \mathcal{L}f(X_m),$$

is a martingale with respect to $\{\mathcal{F}_n\}$.

In our setting the above theorem is an elementary observation. However, this type of martingale formulation in fact plays an important role in the modern theory of general Markov processes. We will not elaborate here. See e.g. [Ebe].

Proof:(of THM 24.2) Assume (i) holds. Because f is bounded and $p(x, \cdot)$ sums to 1, M_n^f is integrable for all n. By (2),

$$\begin{aligned}
\mathbb{E}[M_{n+1}^f - M_n^f \,|\, \mathcal{F}_n] &= \mathbb{E}\left[f(X_{n+1}) - f(X_n) - \mathcal{L}f(X_n) \,|\, \mathcal{F}_n\right] \\
&= \mathbb{E}\left[f(X_{n+1}) - f(X_n) \,|\, \mathcal{F}_n\right] - \mathcal{L}f(X_n) \\
&= 0. \qquad (3)
\end{aligned}$$

That shows (i) implies (ii).

Assume instead that (ii) holds. Fix a subset $B \subseteq S$ and let $f(x) = \mathbb{1}\{x \in B\}$. Then, rearranging (3) and using the definition $\mathcal{L}$, we get

$$\begin{aligned}
\mathbb{P}[X_{n+1} \in B \,|\, \mathcal{F}_n] &= \mathbb{E}\left[f(X_{n+1}) \,|\, \mathcal{F}_n\right] \\
&= f(X_n) + \mathcal{L}f(X_n) \\
&= f(X_n) + \sum_{z \in S} p(X_n, z)[f(z) - f(X_n)] \\
&= \sum_{z \in S} p(X_n, z) f(z) \\
&= p(X_n, B),
\end{aligned}$$

as claimed. ■

2 Potential theory

Many quantities of interest, which we have encountered previously, can be expressed in the following form. Let $D \subset S$ and

$$T_{D^c} = \inf\{n \geq 0 \,:\, X_n \in D^c\}.$$

Let also $f : D^c \to \mathbb{R}_+$ and $c : D \to \mathbb{R}_+$. Define the quantity

$$u(x) := \mathbb{E}_x\left[f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{m<T_{D^c}} c(X_m)\right]. \tag{4}$$

The first term on the RHS is a cost incurred when we exit D, while the second term is a cost incurred along the path. Observe that the function $u(x)$ may take the value $+\infty$; the expectation is well-defined by the nonnegativity of the terms.

EX 24.3 (Some special cases) *Here are some important special cases:*

- *For two disjoint subsets A, Z of S the probability*

  $$u(x) := \mathbb{P}_x[T_A < T_Z],$$

  of hitting A before Z as a function of the starting point $x \in S$ is obtained by taking $D := (A \cup Z)^c$, $f \equiv 1$ (respectively $\equiv 0$) on A (respectively Z), and $c \equiv 0$ on V. The further special case $Z = \emptyset$ leads to the exit probability *from A*

  $$u(x) := \mathbb{P}_x[T_A < +\infty].$$

  On the other hand, if A and Z form a disjoint partition of D^c, we get the exit law *from D*

  $$u(x) := \mathbb{P}_x[X_{T_{D^c}} \in A; T_{D^c} < +\infty].$$

- *The* average occupation time *of $A \subseteq D$ before exiting D*

  $$u(x) := \mathbb{E}_x\left[\sum_{0\le t<T_{D^c}} \mathbb{1}_{\{X_t\in A\}}\right],$$

  is obtained by taking $f \equiv 0$ and $c \equiv 1$ (respectively $\equiv 0$) on A (respectively on A^c). The Green function of the chain stopped at T_{D^c}

  $$u(x) := \mathbb{E}_x\left[\sum_{0\le t<T_{D^c}} \mathbb{1}_{\{X_t=y\}}\right],$$

  is obtained by taking $f \equiv 0$ and $c \equiv 1$. Another special case is $A = D$ where we get the mean exit time *from A*

  $$u(x) := \mathbb{E}_x\left[T_{A^c}\right].$$

The function u turns out to satisfy a certain discrete version of a Dirichlet problem. In undergraduate courses, this is usually called "first-step analysis." A more general statement can be found, e.g., in [Ebe, Theorem 1.3].

THM 24.4 (First-step analysis) *Let p be a transition matrix on a finite or countable spate space S. Let D be a proper subset of S. Let $f : D^c \to \mathbb{R}_+$ and $k : D \to \mathbb{R}_+$ be bounded functions. Then the function $u \geq 0$, as defined in* (4)*, satisfies the system of equations*

$$\begin{cases} u(x) = \sum_y p(x,y)u(y) + c(x) & \textit{for } x \in D, \\ u(x) = f(x) & \textit{for } x \in D^c. \end{cases} \tag{5}$$

Proof: For $x \in D^c$, by definition $u(x) = f(x)$ since $T_{D^c} = 0$. Fix $x \in D$. By TOWER and the Markov property,

$$\begin{aligned} u(x) &= \mathbb{E}_x\left[f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{m < T_{D^c}} c(X_m)\right] \\ &= \mathbb{E}_x\left[\mathbb{E}\left[f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{m < T_{D^c}} c(X_m) \,\middle|\, \mathcal{F}_1\right]\right] \\ &= \mathbb{E}_x\left[c(x) + \mathbb{E}\left[f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{1 \leq m < T_{D^c}} c(X_m) \,\middle|\, \mathcal{F}_1\right]\right] \\ &= \mathbb{E}_x\left[c(x) + u(X_1)\right], \end{aligned}$$

which gives the claim. ■

If further u is finite, then the system of equations (5) can be re-written as

$$\begin{cases} \mathcal{L}u = -c & \text{on } D, \\ u = f & \text{on } D^c. \end{cases} \tag{6}$$

This is the case for instance if D is a finite subset and p is irreducible. Indeed, as the next lemma (of independent interest) shows, the stopping time T_{D^c} then has a finite expectation. Because h is bounded, it follows that

$$\begin{aligned} u(x) &:= \mathbb{E}_x\left[f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{m < T_{D^c}} c(X_m)\right] \\ &\leq \sup_{x \in D^c} f(x) + \sup_{x \in D} c(x) \sup_{x \in D} \mathbb{E}_x\left[T_{D^c}\right] \\ &< +\infty, \end{aligned}$$

uniformly in x. So the system (6) is well defined. Using (1) and rearranging (5) gives the claim. It remains to prove that T_{D^c} has a finite expectation

LEM 24.5 *Let (X_t) be a finite, irreducible Markov chain with state space V and initial distribution μ. For $A \subseteq V$, there is $\beta_1 > 0$ and $0 < \beta_2 < 1$ depending on A such that*

$$\mathbb{P}_\mu[T_A > t] \leq \beta_1 \beta_2^t.$$

In particular, $\mathbb{E}_\mu[T_A] < +\infty$ for any μ, A.

Proof: For any integer m, for some distribution θ,

$$\mathbb{P}_\mu[T_A > ms \,|\, T_A > (m-1)s] = \mathbb{P}_\theta[T_A > s] \leq \max_x \mathbb{P}_x[T_A > s] =: 1 - \alpha_s.$$

Choose s large enough that, from any x, there is a path to A of length at most s of positive probability. Such an s exists by irreducibility. In particular $\alpha_s > 0$. By induction, $\mathbb{P}_\mu[T_A > ms] \leq (1-\alpha_s)^m$ or $\mathbb{P}_\mu[T_A > t] \leq (1-\alpha_s)^{\lfloor \frac{t}{s} \rfloor} \leq \beta_1 \beta_2^t$ for $\beta_1 > 0$ and $0 < \beta_2 < 1$ depending on α_s.

The result for the expectation follows from

$$\mathbb{E}_\mu[T_A] = \sum_{k \geq 0} \mathbb{P}_\mu[T_A > k] \leq \sum_t \beta_1 \beta_2^t < +\infty.$$

That concludes the proof. ■

3 Lyapounov functions

A maximum principle allows to establish uniqueness of the solution of (5). Perhaps even more useful, it also gives an effective approach to bound the function u from above. This is based on a supermartingale related to THM 24.2.

LEM 24.6 (Locally superharmonic functions) *Let $\psi : S \to \mathbb{R}_+$ satisfy*

$$\mathcal{L}\psi \leq -c \qquad \text{on } D.$$

Then the process

$$M_n^{\psi,c,D} := \psi(X_{n \wedge T_{D^c}}) + \sum_{m < n \wedge T_{D^c}} c(X_m),$$

is a non-negative superMG.

Proof: Indeed,

$$\begin{aligned}\mathbb{E}[M_{n+1}^{\psi,c,D} - M_n^{\psi,c,D} \,|\, \mathcal{F}_n] &= \mathbb{E}[\mathbb{1}\{T_{D^c} > n\}(\psi(X_{n+1}) - \psi(X_n) + c(X_n)) \,|\, \mathcal{F}_n] \\ &= \mathbb{1}\{T_{D^c} > n\}(\mathbb{E}[\psi(X_{n+1}) - \psi(X_n) \,|\, \mathcal{F}_n] + c(X_n)) \\ &= \mathbb{1}\{T_{D^c} > n\}(\mathcal{L}\psi(X_n) + c(X_n)) \\ &\leq 0,\end{aligned}$$

since $X_n \in D$ on $\{T_{D^c} > n\}$. ■

We come to the maximum principle.

THM 24.7 (Maximum principle) *Let p be a transition matrix on a finite or countable spate space S. Let D be a proper subset of S, and let $f : D^c \to \mathbb{R}_+$ and $c : D \to \mathbb{R}_+$ be bounded functions. Suppose the nonnegative function $\psi : S \to \mathbb{R}_+$ satisfies the system of inequalities*

$$\begin{cases}\mathcal{L}\psi \leq -c & \text{on } D, \\ \psi \geq f & \text{on } D^c.\end{cases} \tag{7}$$

Then

$$\psi \geq u, \quad \text{on } S, \tag{8}$$

where u is defined in (4).

Proof: The claim, i.e., (8), holds on D^c by THM 24.4 and (7).

Let $x \in D$. Consider the nonnegative supermartingale $N_n := M_n^{\psi,c,D}$ in LEM 24.6. By the convergence of nonnegative supermartingales, (N_n) converges a.s. to a finite limit with expectation $\leq \mathbb{E}_x[N_0]$. In particular, the limit $N_{T_{D^c}}$ is well-defined, nonnegative and finite, including on the event that $\{T_{D^c} = +\infty\}$. As a result,

$$\begin{aligned}N_{T_{D^c}} &= \psi(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{0 \leq m < T_{D^c}} c(X_m) \\ &\geq f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{0 \leq m < T_{D^c}} c(X_m),\end{aligned}$$

where we used (7).

Hence, by definition of u,

$$\begin{aligned}u(x) &= \mathbb{E}_x\left[f(X_{T_{D^c}})\mathbb{1}\{T_{D^c} < +\infty\} + \sum_{0 \leq n < T_{D^c}} c(X_n)\right] \\ &\leq \mathbb{E}_x\left[N_{T_{D^c}}\right] \\ &\leq \mathbb{E}_x[N_0] \\ &= \psi(x),\end{aligned}$$

where, on the last line, we used that the initial state is $x \in D$. That proves the claim. ■

Here is an important application, bounding from above the hitting time T_A to a set A in expectation.

THM 24.8 (Bounding hitting times via Lyapounov functions) *Let p be a transition matrix on a finite or countable spate space S. Let A be a proper subset of S. Suppose the nonnegative function $\psi : S \to \mathbb{R}_+$ satisfies the system of inequalities*

$$\mathcal{L}\psi \leq -1, \quad \text{on } A^c. \tag{9}$$

Then

$$\mathbb{E}_x\left[T_A\right] \leq \psi(x),$$

for all $x \in S$.

Proof: Indeed, by (9) and nonnegativity (in particular on A), the function ψ satisfies the assumptions of THM 24.7 with $W = A^c$, $f \equiv 0$ and $c \equiv 1$. Hence, by definition of u and the claim in THM 24.7,

$$\begin{aligned}\mathbb{E}_x\left[T_A\right] &= \mathbb{E}_x\left[f(X_{T_A})\mathbb{1}\{T_A < +\infty\} + \sum_{0 \leq n < T_A} c(X_n)\right]\\ &= u(x)\\ &\leq \psi(x).\end{aligned}$$

That establishes the claim. ■

Recalling (2), condition (9) is equivalent to the following conditional expected decrease in ψ outside A:

$$\mathbb{E}[\psi(X_{n+1}) - \psi(X_n) \,|\, \mathcal{F}_n] \leq -1, \quad \text{on } \{X_n \in A^c\}. \tag{10}$$

A function satisfying an inequality of this type (and its many variants; see, e.g., [Ebe, Section 1.2]) is known as a Lyapounov function. We consider a simple example next.

EX 24.9 (A Markov chain on the nonnegative integers) *Let $(Z_n)_{n\geq 1}$ be i.i.d. integrable random variables taking values in $\mathbb{Z}$ such that $\mathbb{E}[Z_1] < 0$. Let $(X_n)_{n\geq 0}$ be the chain defined by $X_0 = x$ for some $x \in \mathbb{Z}_+$ and*

$$X_{n+1} = (X_n + Z_{t+1})^+,$$

where recall that $z^+ = \max\{0, z\}$. *Observe that, for any* $y \in \mathbb{Z}_+$, *we have*

$$\begin{aligned}\mathbb{E}_x[X_{n+1} - y \mid X_n = y] &= \mathbb{E}[(y + Z_{n+1})^+ - y] \\ &= \mathbb{E}[-y\mathbb{1}\{Z_{n+1} \leq -y\} + Z_{n+1}\mathbb{1}\{Z_{n+1} > -y\}] \\ &\leq \mathbb{E}[Z_{n+1}\mathbb{1}\{Z_{n+1} > -y\}] \\ &= \mathbb{E}[Z_1\mathbb{1}\{Z_1 > -y\}]. \end{aligned} \tag{11}$$

For all y, *the random variable* $|Z_1\mathbb{1}\{Z_1 > -y\}|$ *is bounded by* $|Z_1|$, *itself an integrable random variable. Moreover,* $Z_1\mathbb{1}\{Z_1 > -y\} \to Z_1$ *as* $y \to +\infty$ *almost surely. Hence, the dominated convergence theorem implies that*

$$\lim_{y\to+\infty} \mathbb{E}[Z_1\mathbb{1}\{Z_1 > -y\}] = \mathbb{E}[Z_1] < 0.$$

So for any $0 < \varepsilon < -\mathbb{E}[Z_1]$, *there is* $y_\varepsilon \in \mathbb{Z}_+$ *large enough that* $\mathbb{E}[Z_1\mathbb{1}\{Z_1 > -y\}] < -\varepsilon$ *for all* $y > y_\varepsilon$. *Fix* ε *as above and define*

$$A := \{0, 1, \ldots, y_\varepsilon\}.$$

We use THM 24.8 to bound T_A *in expectation. Define*

$$\psi(x) = \frac{x}{\varepsilon}, \quad \forall x \in \mathbb{Z}_+.$$

On the event $\{X_n = y\}$, *we re-write* (11) *as*

$$\begin{aligned}\mathbb{E}[\psi(X_{n+1}) - \psi(X_n) \mid \mathcal{F}_n] &\leq \frac{\mathbb{E}[Z_1\mathbb{1}\{Z_1 > -y\}]}{\varepsilon} \\ &\leq -1,\end{aligned}$$

for $y \in A^c$. *This is the same as* (10). *Hence, we can apply THM 24.8 to get*

$$\mathbb{E}_x[T_A] \leq \psi(x) = \frac{x}{\varepsilon},$$

for all $x \geq y_\varepsilon$.

References

[Bre20] Pierre Bremaud. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*, volume 31 of *Texts in Applied Mathematics*. Springer, Cham, 2020. Second edition [of 1689633].

[Ebe] Andreas Eberle. Markov Processes. 2021.

[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.