Markov chains: definition and basic properties

Janko Gravner

MAT 235B

February 22, 2024

Markov chains

Definition. Assume $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ is a filtration. Assume that X_n are random variables with values in the *state space* $(S, \mathcal{S})$, and that $X_0, X_1, \ldots$ is an adapted process (that is, X_n is $\mathcal{F}_n$-measurable). If

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = P(X_{n+1} \in B \mid X_n), \tag{1}$$

for all $B \in \mathcal{S}$, then X_n is called a (discrete-time) Markov chain.

Normally, $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$.

Note: if you compute the LHS of (1) and is a function of X_n only, you are done, by the tower property.

Markov chains

We will restrict to:

- time-homogeneous MC: $P(X_{n+1} \in B \mid X_n)$ depends only on X_n (and not on n); and
- countable state-space, $\mathcal{S} = 2^S$, with *transition probabilities* $p(i,j) = P(X_{n+1} = j \mid X_n = i)$.

Example. Random walk $S_n = X_0 + \xi_1 + \cdots + \xi_n$, where ξ_i are i.i.d., $\mathbb{Z}^d$-valued. So $S = \mathbb{Z}^d$. Then $p(i,j) = P(\xi_1 = j - i)$:

$$\begin{aligned} &P(S_{n+1} = j \mid S_n = i) \\ &= P(X_0 + \xi_1 + \cdots + \xi_n + \xi_{n+1} = j \mid X_0 + \xi_1 + \cdots + \xi_n = i) \\ &= P(\xi_{n+1} = j - i) = P(\xi_1 = j - i) \end{aligned}$$

Example. Branching process. $S = \{0, 1, 2, \ldots\}$. If $\xi_1, \xi_2, \ldots$ are i.i.d. with offspring distribution, then $p(i,j) = P(\xi_1 + \cdots + \xi_i = j)$ for $i > 0$, and $p(0,0) = 1$.

Example. Birth-death chain. $S = \{0, 1, 2, \ldots\}$. Here, we assume that $p(i,j) = 0$ if $|i - j| > 1$.

Example. Voter model. Let V be a finite set, and let E be a collection of undirected edges, that is, $E \subset \{\{x, y\} \subset V : x \neq y\}$. The state of the system at time $n = 0, 1, 2, \ldots$ is determined by a function $\zeta_n : V \to \{0, 1\}$ (where 0s and 1s are opinions), so $S = \{0, 1\}^V$. At each time n a random ordered pair (x, y) of neighbors is chosen and then x assumes the opinion of y to get ζ_{n+1}. This is a MC.

However,

$$X_n = |\{\zeta_n = 1\}| = \text{number of opinions 1}$$

is not a MC! For example, in the cycle of 4 points, with two 0 opinions and two 1 opinions, the probability that X_n transitions from 2 to 3 can be $1/2$ or $1/4$, depending on whether the two 1s are separate or together.

In general a function of a MC is not necessarily a MC.

Example. M/G/1 queue. Customers arrive at a single (1) counter as a Poisson process with rate λ (M). Each customer requires an independent serving time with d.f. F (G). Let X_n be the number of customers at the time nth customer starts being served. Probability that k customers arrive during the service time is

$$a_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} \, dF(t)$$

and then

$$p(0, k-1) = a_k, \; k \geq 2, \quad p(0,0) = a_0 + a_1,$$
$$p(j, j-1+k) = a_k, \; j \geq 1, k \geq 0$$

Markov chains

Transition probabilities satisfy

$$(2) \qquad p(i,j) \geq 0, \quad \sum_{j \in S} p(i,j) = 1 \text{ for every } i \in S$$

and they, together with the initial distribution $\mu = \mu_{X_0}$,

$$P(X_0 = i) = \mu(\{i\}) = \mu(i),$$

determine all probabilities:

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \mu(i_0)p(i_0, i_1) \cdots p(i_{n-1}, i_n).$$

We sometimes give μ as a subscript when computing probabilities and expectations. If p is interpreted as a matrix and μ as a row vector,

$$P_\mu(X_n = y) = (\mu p^n)(y),$$

the yth coordinate of the row vector μp^n.

Proposition

If $p(\cdot, \cdot)$ *satisfying (2), and* μ*, are given, there exists a MC* $X_0, X_1, \dots$ *such that* $\mu_{X_0} = \mu$ *and* $p(i,j) = P(X_{n+1} = j \mid X_n = i)$*.*

Proof.

Let X_0 and $D_{i,n}$ be independent with $\mu_{X_0} = \mu$ and $\mu_{D_{i,n}} = p(i, \cdot)$. Then define $X_{n+1} = D_{X_n,n}$. □

Markov property

Denote by $(S^\infty, \mathcal{S}^\infty) = (S \times S \times \cdots, \mathcal{S}^\infty)$ the *trajectory space*, where the σ-algebra $\mathcal{S}^\infty$ is generated by the sets

$$\{S \times \cdots \times S \times A \times S \times \cdots : A \subset S, \text{at any position } n \geq 1\},$$

or equivalently and more usefully by the *cylinder sets*

$$\{A_0 \times \cdots \times A_n \times S \times S \times \cdots : A_i \subset S, n \geq 0\},$$

which is a π-system that generates $\mathcal{S}^\infty$.

Define the (left) shift operator $\theta_n : S^\infty \to S^\infty$ by

$$\theta_n(x_0, x_1, \ldots) = (x_n, x_{n+1}, \ldots)$$

Markov property

Assume $X_0, X_1, \ldots$ is a sequence of r.v.'s with values in $(S, \mathcal{S})$. Then $X : (\Omega, \mathcal{F}) \to (S^\infty, \mathcal{S}^\infty)$ is given by

$$\omega \mapsto (X_0(\omega), X_1(\omega), \ldots)$$

which is measurable, as each X_n is.

Moreover, $A \in \sigma\{X_0, X_1, \ldots\}$ iff $A = \{(X_0, X_1, \ldots) \in B\}$ for some $B \in \mathcal{S}^\infty$.

Any sequence of r.v.'s induces a probability measure on $(S^\infty, \mathcal{S}^\infty)$ which is the distribution of $(X_0, X_1, \ldots)$.

Markov property

Now let $X_0, X_1, \dots$ be a time-homogeneous Markov chain with countable state space S, $\mathcal{S} = 2^S$, $\mathcal{F}_n = \sigma\{X_0, \dots, X_n\}$. Recall that the distribution of X_0 is indicated by the subscript (e.g., P_μ when $\mu_{X_0} = \mu$ or P_x when $X_0 = x \in S$).

Theorem (Markov property)

Let $\Phi : S^\infty \to \mathbb{R}$ be bounded measurable. Let $\varphi(x) = E_x[\Phi(X)]$. Then

$$E_\mu[\Phi(\theta_n(X)) \mid \mathcal{F}_n] = E_{X_n}[\Phi(X)] := \varphi(X_n).$$

Markov property

Corollary

We have

$$E_\mu[\Phi(\theta_n(X))] = E_\mu[E_{X_n}[\Phi(X)]]$$

provided $\Phi \geq 0$ *or at least one side is finite when* Φ *is replaced by* $|\Phi|$.

In words, at time n, the distribution of the rest of trajectory is the same as the distribution of trajectory started at X_n.

Example. Let S_n be a simple symmetric random walk in 2d. Assume $A \subset \mathbb{Z}^2$ and $A^c = B_1 \cup B_2$, with $B_1 \cap B_2 = \emptyset$. Let $T = \inf\{n \geq 0 : S_n \notin A\} \geq 0$.
Let $\varphi(x) = P_x(S_T \in B_1)$, $x \in \mathbb{Z}^2$. If A is finite, can we compute $\varphi(x)$?
Let $\Phi = 1_{\{S_T \in B_1\}}$, so that $\varphi(x) = E_x[\Phi(X)]$. If $x \in A$,

$$\begin{aligned} \varphi(x) &= E_x[\Phi(\theta_1(X))] \quad \text{(because } x \in A\text{)} \\ &= E_x[E_{X_1}\Phi(X)] \quad \text{(Markov property)} \\ &= E_x[\varphi(X_1)] \\ &= \frac{1}{4}\sum_{i=1}^{4} \varphi(x + e_i) \end{aligned}$$

We say that φ is *harmonic* on A. We also have boundary conditions, $\varphi\,|_{B_1} = 1$, $\varphi\,|_{B_2} = 0$.

Markov property

If A is finite, and φ is harmonic on A, φ satisfies the *maximum* principle: its maximum (and minimum) must be achieved outside A. It follows that there is a unique harmonic function with given boundary condition; the above linear equations for φ have a unique solution.

This argument is not limited to 2d random walks.

Markov property

Proof of Markov property.

Let Φ be the indicator of a cylinder set,

$$\Phi = 1_{A_0 \times \cdots \times A_k \times S \times S \times \cdots},$$

for some k and $A_i \subset S$. Take also $F \in \mathcal{F}_n$ of the form

$$F = \{X_0 \in B_0, \ldots, X_n \in B_n\},$$

for some $B_i \subset S$. Then

$$E_\mu[\Phi(\theta_n(X))1_F] = \sum \mu(i_0)p(i_0, i_1) \cdots p(i_{n+k-1}, i_{n+k}),$$

over

$$i_0 \in B_0, \ldots, i_{n-1} \in B_{n-1}, i_n \in B_n \cap A_0, i_{n+1} \in A_1, \ldots, i_{n+k} \in A_k$$

Markov property

Proof of Markov property, continued.

Also,

$$E_\mu[\varphi(X_n)1_F] = \sum \mu(i_0)p(i_0, i_1)\cdots p(i_{n-1}, i_n)\varphi(i_n)$$

over the same range of indices $i_0, \ldots, i_n$, as $\varphi(i_n) = 0$ unless $i_n \in A_0$, so we may indeed restrict $i_n \in B_n \cap A_0$. Then,

$$\begin{aligned}\varphi(i_n) &= P_{i_n}(X_0 \in A_0, X_1 \in A_1, \ldots, X_k \in A_k)\\ &= P_{i_n}(X_1 \in A_1, \ldots, X_k \in A_k)\\ &= \sum p(i_n, i_{n+1})\cdots p(i_{n+k-1}, i_{n+k})\end{aligned}$$

again over the same range of indices $i_{n+1}, \ldots i_{n+k}$.
It follows that

$$E_\mu[\Phi(\theta_n(X))1_F] = E_\mu[\varphi(X_n)1_F].$$

Markov property

Proof of Markov property, continued.

We have

$$E_\mu[\Phi(\theta_n(X))1_F] = E_\mu[\varphi(X_n)1_F].$$

The set of F's we have chosen is a π-system that generates $\mathcal{F}_n$, so the above is true for all $F \in \mathcal{F}_n$ by the $\pi - \lambda$ theorem. Also, by the same theorem, the above is true for $\Phi = 1_D$, $D \in \mathcal{S}^\infty$. Then, by linearity, it is true for simple functions Φ, then by MCT for all positive Φ, and then by linearity again for all bounded Φ. □

Strong Markov property

Recall that N is a stopping time if $\{N = n\} \in \mathcal{F}_n$. Define the information available at time N:

$$\mathcal{F}_N = \{A \in \mathcal{F} : A \cap \{N = n\} \in \mathcal{F}_n, \text{ for all } n < \infty\}$$

(If you stop at time n, you know at that time whether A happened or not.)

Proposition

The family $\mathcal{F}_N$ is a σ-algebra, N and $X_N 1_{\{N<\infty\}}$ are $\mathcal{F}_N$-measurable, and so is any stopping time $M \leq N$.

Proof.

HW. □

Strong Markov property

We define the random shift on $\{N < \infty\} \subset S^\infty$, $\theta_N : \{N < \infty\} \to S^\infty$, by

$$\theta_N(x) = \theta_n(x) \text{ on } \{N = n\}$$

Theorem (Strong Markov property)

Let $\Phi_n : S^\infty \to R$ be measurable and uniformly bounded, $|\Phi_n| \leq M$ for every n. Let $\varphi(x, n) = E_x[\Phi_n(X)]$. Then, on $\{N < \infty\}$,

$$E_\mu[\Phi_N(\theta_N(X)) \mid \mathcal{F}_N] = E_{X_N}[\Phi_N(X)] := \varphi(X_N, N).$$

Corollary

If $P(N < \infty) = 1$ and $\Phi \geq 0$,

$$E_\mu[\Phi(\theta_N(X))] = E_\mu[E_{X_N}[\Phi(X)]].$$

Strong Markov property

Proof.

Take $A \in \mathcal{F}_N$. Then

$$
\begin{aligned}
&E_\mu[\Phi_N(\theta_N(X))1_{A\cap\{N<\infty\}}] \\
&= \sum_{n=0}^{\infty} E_\mu[\Phi_n(\theta_n(X))1_{A\cap\{N=n\}}] \\
&= \sum_{n=0}^{\infty} E_\mu[\varphi(X_n, n)1_{A\cap\{N=n\}}] \quad (\text{MP, as } A\cap\{N=n\}\in\mathcal{F}_n) \\
&= E_\mu[\varphi(X_N, N)1_{A\cap\{N<\infty\}}].
\end{aligned}
$$

□

Strong Markov property

Example. Let S_n be a simple 1d random walk, with $p \geq 1/2$, started at 0. Let T_b be the time to reach $b > 0$. Show that

$$T_1, T_2 - T_1, T_3 - T_2, \ldots$$

are i.i.d.

It is enough to show that for bounded Borel functions f,

$$E_0[f(T_b - T_{b-1}) \mid \mathcal{F}_{T_{b-1}}] = E_0[f(T_1)].$$

Let $\Phi(X) = f(T_b)$ and $N = T_{b-1}$. Then

$$\begin{aligned}\Phi(\theta_N(X)) &= \sum_{n=1}^{\infty} \Phi(\theta_n(X)) 1_{\{T_{b-1}=n\}} \\ &= \sum_{n=1}^{\infty} f(T_b - n) 1_{\{T_{b-1}=n\}} \\ &= f(T_b - T_{b-1}).\end{aligned}$$

Also, $X_N = b - 1$ and

$$\varphi(b-1) = E_{b-1}[f(T_b)] = E_0[f(T_1)].$$