

Stochastic Processes

MATH 447

Nicholas Hayek

Taught by Prof. Louigi Addario-Berry

CONTENTS

I Markov Chains	3
Introduction	3
Time-Homogeneous Markov Chains	4
<i>Multi-Step Transition Probabilities</i>	
Long Term Behavior	8
<i>Periodicity of States</i>	
<i>Finding Stationary Distributions</i>	
<i>Transience and Recurrence</i>	
<i>Canonical Decompositions</i>	
<i>Proof of Fundamental Theorem of Markov Chains</i>	
<i>Hitting Times and Absorbing States</i>	

Index of Definitions

We assume working knowledge of probability and no knowledge of measure theory (though a grasp of analysis is essential). See these [MATH 356 notes](#), also taught by Louigi!

I Markov Chains

Conditional expectations will be important in this course. Recall $\mathbb{E}[X|Y = y_0]$, where X, Y are random variables. If Y is continuous, writing $\mathbb{E}[X|Y = y_0] = \frac{\mathbb{P}(X, Y=y_0)}{\mathbb{P}(Y=y_0)}$, will not work. Instead, we consider the slice of the joint density function $f(x, y)$ at $y = y_0$. The result is a one dimensional function $g(x)$ which may not have probability 1. Hence, we divide by $\int g(x) dx$ to make it into a density function:

$$\mathbb{E}[X|Y = y_0] = \int_{\mathbb{R}} \frac{f(x, y_0)}{\int_{\mathbb{R}} f(x, y_0) dx} x dx$$

DEF 1.1 We frequently write $f_{X|Y}(x) = f(x, y)/\int_{\mathbb{R}} f(x, y) dx$, and call this the *conditional density* of X given Y . For fixed y , then, $\mathbb{E}[X|Y = y] = \mathbb{E}[Z]$, where $Z \sim f_{X|Y}$.

INTRODUCTION

Before providing definitions, we give some examples of stochastic processes:

Eg. 1.1 A simple random walk: $S_{i+1} = S_i + X_i$, where $X_i \sim \text{Ber}(p)$ and $S_0 = 0$. We might ask: does S_i ever return to 0, i.e.

$$\mathbb{P}(\exists i > 0 : S_i = 0)$$

Eg. 1.2 A branching process: as in asexual reproduction, we have an initial node. Each node n has a number of children X_n , where $\frac{X_n}{2} \sim \text{Ber}(p)$. We denote Z_i to be the number of individuals in the i -th generation. We might ask: does Z_i ever have no children, i.e.

$$\mathbb{P}(\exists i > 0 : Z_i = 0)$$

Eg. 1.3 Choose k independent random points in the square $[0, \sqrt{k}]^2$. On average, then, there is 1 point within any unit square $U \subseteq [0, \sqrt{k}]^2$.

DEF 1.2 Given a finite or countable set V , a *Markov chain* with *state space* V is a sequence $X_n : n \geq 0$ of random variables, with $X_n \in V$, such that:

$$\mathbb{P}(\underbrace{X_{n+1} = v_{n+1}}_{\text{future}} | \underbrace{X_0 = v_0, \dots, X_{n-1} = v_{n-1}}_{\text{past}}, \underbrace{X_n = v_n}_{\text{present}}) = \mathbb{P}(X_{n+1} = v_{n+1} | X_n = v_n)$$

In other words, the future only depends on the past via the present. This is called the

Markov property.

Sometimes we allow Markov chains to be only finitely large (i.e. $0 \leq n \leq m$). For instance, we limit ourselves to one weekend of gambling in Las Vegas. A graphical example would look something like:



By repeated Bayes' Law, we observe

PROP 1.1

$$\begin{aligned} & \mathbb{P}(X_1 = v_1, \dots, X_n = v_n | X_0 = v_0) \\ &= \mathbb{P}(X_1 = v_1 | X_0 = v_0) \cdot \mathbb{P}(X_2 = v_2 | X_0 = v_0, X_1 = v_1) \cdots \mathbb{P}(X_n = v_n | X_0 = v_0, \dots, X_{n-1} = v_{n-1}) \\ &= \mathbb{P}(X_1 = v_1 | X_0 = v_0) \cdot \mathbb{P}(X_2 = v_2 | X_1 = v_1) \cdots \mathbb{P}(X_n = v_n | X_{n-1} = v_{n-1}) \quad \text{by Markov property} \end{aligned}$$

TIME-HOMOGENEOUS MARKOV CHAINS

We often write
THMC

We say that a Markov chain is *time-homogeneous* if, for all $u, v \in V$ and $n \geq 0$

DEF 1.5

$$\mathbb{P}(X_{n+1} = v | X_n = u) = \mathbb{P}(X_1 = v | X_0 = u)$$

In other words, the chain's behavior is described entirely by $\mathbb{P}(X_1 = v | X_0 = u)$ for each $(v, u) \in V \times V$. In this case, we can describe such probabilities in a *transition matrix* P :

DEF 1.6

$$P = (p_{u,v})_{(u,v) \in V^2} = (\mathbb{P}(X_1 = v | X_0 = u))_{(u,v) \in V^2}$$

Eg. 1.4 Recall the game Snakes and Ladders. A 6×6 grid is indexed 1, ..., 36. Players start at the 1 cell. The game ends when a player reaches the 36 cell. A die roll dictates how many spots one advances. There are some directed edges between cells (increasing: "ladders", decreasing: "snakes"). One must follow these edges when one lands at its tail. Suppose a ladder exists from 11 to 27. Then

$$\mathbb{P}(X_{11} = 27 | X_{10} = 6, X_9 = 3) = \frac{1}{6} = \mathbb{P}(X_{11} = 27 | X_{10} = 6) = \mathbb{P}(X_2 = 27 | X_1 = 6)$$

We see that Snakes and Ladders is naturally modeled as a time-homogeneous Markov chain.

Eg. 1.5 Sampling without replacement is *not* a Markov chain. If we sample from

$|X| = 10$, we have

$$\begin{aligned}\mathbb{P}(X_3 = a | X_2 = b) &= 1/9 \\ \mathbb{P}(X_3 = a | X_2 = b, X_1 = c) &= 1/8 \\ \mathbb{P}(X_3 = a | X_2 = b, X_1 = a) &= 0\end{aligned}$$

so we do not satisfy the Markov property.

Eg. 1.6 Returning to the Snakes and Ladders example, consider $S \subseteq V$. Let $T_S = \inf\{n \geq 0 : X_n \in S\}$, which we call the "*hitting time*" of S . We may ask...

- DEF 1.7
- What is the average number of rounds to finite? We can write this as $\mathbb{E}[T_{\{36\}} | X_0 = 1]$.
 - What is the probability of landing on 18 or 19 before the game ends? We can write this as $\mathbb{P}(T_{\{18,19\}} < T_{\{36\}} | X_0 = 1)$.
 - What is the average number of visits to 6 before the game ends? We can write this as

$$\mathbb{E}[\#\{n \in [T_{\{36\}}] : X_n = 6\} | X_0 = 1]$$

- What is the expected proportion of time spent on state 5 before the game ends?
- If we allow two players, what is the probability that player 1 wins? Is this still a Markov chain?

DEF 1.8 A matrix $P = (p_{u,v})_{(u,v) \in V^2}$ is called a *stochastic matrix* if every row sums to 1, i.e.

$$\forall u \in V, \sum_{v \in V} p_{u,v} = 1$$

Note that any stochastic matrix is the transition matrix of some time-homogeneous Markov chain with state space V and transition probabilities

$$\mathbb{P}(X_{n+1} = v | X_n = u) = \mathbb{P}(X_1 = v | X_0 = u) = p_{u,v}$$

A directed graph, together with its stochastic matrix, can visualize any THMC:



Eg. 1.7 Random walks on an undirected weighted graph, where edge weights dictate the proportional probability of transitioning between two states, are a special class of THMCs. In particular, given a graph $G = (V, E)$ with weights $w(e) > 0 : e \in E$, we set

$$p_{u,v} = \frac{w(\{u, v\})}{\sum_{z \in N(u)} w(\{u, z\})}$$

If there are no edges $u \leftrightarrow v$, we write $p_{u,v} = 0$.

Not every THMC can be represented by a random walk on an undirected weighted graph. In particular, see the directed graph listed above, or any transition matrix which is not symmetric.

As a concrete example, we can consider a random walk on the number line \mathbb{Z} , where, if $w(k, k+1) = \alpha$, $w(k-1, k) = \frac{\alpha}{2}$.

$$\dots \frac{1}{16} -3 \frac{1}{8} -2 \frac{1}{4} -1 \frac{1}{2} 0 \frac{1}{2} 1 \frac{1}{2} 2 \frac{1}{4} 3 \frac{1}{8} \dots$$

Multi-Step Transition Probabilities

Given a THMC $X = X_n : n \geq 0$ with a transition matrix P , we write

$$\begin{aligned} \mathbb{P}(X_2 = v | X_0 = u) &= \sum_{w \in V} \mathbb{P}(X_2 = v, X_1 = w | X_0 = u) \\ &= \sum_{w \in V} \mathbb{P}(X_1 = w | X_0 = u) \mathbb{P}(X_2 = v | X_1 = w, X_0 = u) \quad \text{by Markov property} \\ &= \sum_{w \in V} p_{u,w} p_{w,v} = (P^2)_{u,v} \quad \text{or write } P_{u,v}^2 \end{aligned}$$

Hence, to determine a two-step transition probability, and by extension an n -step transition probability from u to w , we consider $P_{u,v}^n$. PROP 1.2

See [Prop 1.1](#) to expand probabilities, using Bayes', as needed. We get that

$$\begin{aligned} \mathbb{P}(X_n = v | X_0 = u) &= \sum_{v_1, \dots, v_{n-1} \in V} \mathbb{P}(X_1 = v_1, \dots, X_{n-1} = v_{n-1}, X_n = v | X_0 = u) \\ &= \sum_{v_1, \dots, v_{n-1} \in V} p_{u,v_1} p(v_1, v_2) \cdots p(v_{n-1}, v) = (P^n)_{u,v} \quad \square \end{aligned}$$

Thus, if P is a stochastic matrix, then so is P^n . PROP 1.3

$$\sum_{v \in V} P_{u,v}^n = \sum_{v \in V} \mathbb{P}(X_n = v | X_0 = u) = 1. \quad \square$$

PROOF.

PROOF.

Theorem 1.1 Markov Property

If $X_n : n \geq 0$ is a THMC with state space V , then for all $u_0, \dots, u_{n-1}, u, v \in V$,

$$\mathbb{P}(X_{n+m} = v | X_0 = u_0, \dots, X_{n-1} = u_{n-1}, X_n = u) = \mathbb{P}(X_{n+m} = v | X_n = u) = P_{u,v}^m$$

PROOF.

One shows this by combining the Markov property with [Prop 1.2](#) via induction. \square

Somewhat nonsensically, we *also* call this the Markov property. When talking about THMCs, this will be the default notion.

DEF 1.9 We say that a Markov chain has an *initial distribution* $\alpha = (\alpha_v : v \in V)$ if $\mathbb{P}(X_0 = v) = \alpha_v$ for each $v \in V$. If this is the case, we often write α as a subscript of our state probabilities. For instance,

$$\mathbb{P}_\alpha(X_n = v) = \sum_{u \in V} \mathbb{P}_\alpha(X_0 = u, X_n = v) = \sum_{u \in V} \mathbb{P}_\alpha(X_0 = u) \mathbb{P}_\alpha(X_n = v | X_0 = u) = \sum_{u \in V} \alpha_u P_{u,v}^n$$

PROP 1.4 For any event E depending only on X_0, \dots, X_n , with $\mathbb{P}(X_n = u, E) > 0$, we have

$$\mathbb{P}(X_{n+m} = v | X_n = u, E) = P_{u,v}^m$$

PROOF.

For any such event E , we can determine whether E occurs exactly when we know the realized values u_i of X_i for $i = 1, \dots, n-1$. Hence, we may write \mathcal{S} to be the set of tuples (u_0, \dots, u_{n-1}) that guarantee E . It follows that

$$\mathbb{P}(X_n = u, E) = \sum_{\mathbf{s} \in \mathcal{S}} \mathbb{P}(\mathbf{x} = \mathbf{s}, X_n = u)$$

Similarly, we have

$$\begin{aligned} \mathbb{P}(X_{n+m} = v, X_n = u, E) &= \sum_{\mathbf{s} \in \mathcal{S}} \mathbb{P}(X_{n+m} = v, X_n = u, \mathbf{x} = \mathbf{s}) \\ &= \sum_{\mathbf{s} \in \mathcal{S}} \mathbb{P}(X_{n+m} = v | X_n = u, \mathbf{x} = \mathbf{s}) \mathbb{P}(X_n = u, \mathbf{x} = \mathbf{s}) \\ &= P_{u,v}^m \sum_{\mathbf{s} \in \mathcal{S}} \mathbb{P}(X_n = u, \mathbf{x} = \mathbf{s}) = P_{u,v}^m \mathbb{P}(X_n = u, E) \end{aligned}$$

Divide and use Bayes, and the result follows. \square

PROP 1.5 If X is a THMC with transition matrix P , then, for all $k \geq 1$, $X_{kn} : n \geq 0$ is a THMC with transition matrix P^k .

For any $n \neq 0$, any sequence $u_0, \dots, u_{n+1} \in V$ satisfies

$$\mathbb{P}(X_{(n+1)k} = u_{n+1} | X_0 = u_0, X_k = u_1, \dots, X_{nk} = u_n) = P_{u_n, u_{n+1}}^k \quad \square$$

PROOF.

Theorem 1.2 Chapman-Kolmogorov

For any Markov chain X with state space V , any $m, n \geq 0$, and $u, v \in V$,

$$\mathbb{P}(X_{m+n} = v | X_0 = u) = \sum_{w \in V} \mathbb{P}(X_n = w | X_0 = u) \mathbb{P}(X_{m+n} = v | X_n = w)$$

If the X is time homogeneous, then this is $P_{u,v}^{n+m}$, which agrees with [Prop 1.2](#).

LONG TERM BEHAVIOR

Recall from probability the *law of large numbers*: if $Y_n : n \geq 1$ are IID with common mean μ , then $\frac{S_n}{n} \rightarrow \mu$ in probability, where $S_n = \sum_{i=1}^n Y_i$, i.e. $\forall \varepsilon > 0$, DEF 1.10

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0$$

If $Y_i \in \mathbb{Z}$ then, for $k, \ell, u_i \in \mathbb{Z}$ and $i = 1, \dots, n-1$,

$$\begin{aligned} \mathbb{P}(S_{n+1} = \ell | S_n = k, S_i = u_i \ \forall i) &= \mathbb{P}(Y_{n+1} = \ell - k | S_n = k, S_i = u_i \ \forall i) \\ &= \mathbb{P}(Y_{n+1} = \ell - k | Y_1 = u_1 - u_0, Y_2 = u_2 - u_1, \dots, Y_n = k - u_{n-1}) \\ &= \mathbb{P}(Y_{n+1} = \ell - k) = \mathbb{P}(Y_1 = \ell - k) = P_{k,\ell} \end{aligned}$$

where $S_n : n \geq 0$ has transition matrix P , noting that it may be viewed as a THMC.

From now on, we denote by $\mathbb{P}_v(E)$ the probability $\mathbb{P}(E|v)$.

Eg. 1.8 A general two-state chain, with states A and B , can be described by

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Let $q_n = \mathbb{P}_A(X_n = A) = \mathbb{P}(X_n = A | X_0 = A)$. Then

$$\begin{aligned} q_{n+1} &= \mathbb{P}_A(X_{n+1} = A, X_n = A) + \mathbb{P}_A(X_{n+1} = A, X_n = B) \\ &= \mathbb{P}_A(X_{n+1} = A | X_n = A) \mathbb{P}_A(X_n = A) + \mathbb{P}_A(X_{n+1} = A | X_n = B) \mathbb{P}_A(X_n = B) \\ &= (1-\alpha)q_n + \beta(1-q_n) = \beta + (1-\alpha-\beta)q_n \end{aligned}$$

This recurrence has a unique solution. In particular, one can find

$$q_n = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n \frac{\alpha}{\alpha+\beta}$$

It follows that $q_n \rightarrow \frac{\beta}{\alpha+\beta}$, and hence $\mathbb{P}_A(X_n = B) = 1 - q_n \rightarrow \frac{\alpha}{\alpha+\beta}$. Likewise:

$$\mathbb{P}_B(X_n = B) = \frac{\alpha}{\alpha + \beta} + (1 - \frac{\alpha}{\alpha + \beta})^n \frac{\beta}{\alpha + \beta}$$

So $\mathbb{P}_B(X_n = B) \rightarrow \frac{\alpha}{\alpha+\beta}$.

Let $\pi := (\pi_A, \pi_B) := \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ be the distribution of our initial state X_0 , Then

$$\mathbb{P}_\pi(X_1 = A) = \pi_A \mathbb{P}_A(X_1 = A) + \pi_B \mathbb{P}_B(X_1 = A) = \pi_A$$

and, similarly, $\mathbb{P}_\pi(X_1 = B) = \pi_B$. Hence, if X_0 has initial distribution π , then X_1 also has distribution π . By induction, X_n has distribution $\pi \forall n \geq 0$.

When we say $X = \text{Markov}(P)$, we mean that X is a THMC with transition matrix P .

- DEF 1.11** A probability distribution π is called *stationary* if $\pi P = \pi$. Similarly, a probability distribution λ is called a *limiting distribution* if, for each $u, v \in V$

$$(P^n)_{u,v} \rightarrow \lambda_v \text{ as } n \rightarrow \infty$$

In other words, $\mathbb{P}_u(X_n = v) \rightarrow \lambda_v$. Note that, for any initial distribution α , we have $\alpha P^n \rightarrow \lambda$, i.e. $(\alpha P^n)_v \rightarrow \lambda_v$, where λ is limiting.

- PROP 1.6** If λ is a limiting distribution for P , then λ is stationary for P .

PROOF.

Fix any initial distribution α , we have

$$\lambda = \lim_{n \rightarrow \infty} (\alpha P^n) = \lim_{n \rightarrow \infty} (\alpha P^{n-1} P) = (\lim_{n \rightarrow \infty} \alpha P^{n-1}) P = \lambda P \quad \square$$

Stationary distributions need not be unique, but limiting distributions are (as the limit $\lim_{n \rightarrow \infty} \alpha P^n$ is well-defined). In general, then, stationary distributions need not be limiting distributions.

- DEF 1.13** A stochastic matrix P is called *regular* if $\exists n \geq 1$ such that $P^n > 0$ on all entries.

Theorem 1.3 Fundamental Theorem of Markov Chains

Every finite, regular stochastic matrix P has a limiting distribution π .

When $n = 0$, $P^n = I$, which encapsulates the idea that, at timestep 0, we will be at our initial positions.

Incorporating some of the formulations above, this is equivalent to saying: *For a regular stochastic matrix, there exists a unique distribution $\pi = (\pi_v : v \in V)$ such that $\pi P = \pi$ and $\mathbb{P}_u(X_n = v) \rightarrow \pi_v \forall u, v \in V$.*

A stationary distribution always exists!

Let $\rho = \langle 1, \dots, 1 \rangle$. Then note that $P\rho = \rho$, since the sum of any row in P must be 1. Hence, P has eigenvalue 1. It follows that it has a left eigenvector, i.e. $\pi : \pi P = \pi$. This is exactly a stationary distribution, as long as we scale suitably such that π is a distribution.

In this case, there is a unique stationary distribution, and it is the unique limiting distribution.

However, then process of scaling into a distribution is non-trivial. Since π may have negative coordinates, and hence $\sum \pi_i = 0$, we must consider instead $|\pi|$, i.e. prove it is also an eigenvalue.

This is true, but requires the fact that P is stochastic

Periodicity of States

For $u, v \in V$, we say that v is *accessible* from u if $\exists n \geq 0$ such that $(P^n)_{u,v} > 0$. Equivalently, in the directed graph generated by P , there is a directed path from u to v . When v is accessible from u , we write $u \rightarrow v$.

DEF 1.14

States u and v *communicate* if $u \rightarrow v$ and $v \rightarrow u$. When u and v communicate, we write $u \leftrightarrow v$. Observe that communication is a equivalence relation. Hence, the state space V can be written as a disjoint union of mutually-communicating states, called a *communication class*. Note that, in the directed graph generated by P , these correspond to the strongly connected components.

DEF 1.15

Clearly, if P is regular, then it is irreducible

We say that P is *irreducible* if there is only one communication class.

DEF 1.17

$$u \rightarrow v \iff \mathbb{P}_u(T_{\{v\}} < \infty) > 0.$$

PROP 1.7

The *period* of a state $u \in V$ is

DEF 1.18

$$d(u) := \gcd(n > 0 : P_{u,u}^n > 0)$$

If $d(u) = 1$, we call u *aperiodic*. By extension, P is aperiodic if $d(u) = 1 \forall u \in V$, and X is aperiodic if $X = \text{Markov}(P)$ for P aperiodic.

DEF 1.19

If $u \leftrightarrow v$, then $d(u) = d(v)$.

PROP 1.8

Let $I = \{n > 0 : P_{u,u}^n > 0\}$, and similarly J for v . Hence, $d(u) = \gcd(I)$ and $d(v) = \gcd(J)$.

PROOF.

Let $a, b > 0$ such that $P_{u,v}^a > 0$ and $P_{v,u}^b > 0$. Then

$$P_{u,u}^{a+b} \geq P_{u,v}^a P_{v,u}^b > 0$$

$\implies a + b \in I$, so $d(u)|a + b$. Now, if $n \in J$, then

$$P_{u,u}^{a+b+n} \geq P_{u,v}^a P_{v,v}^n P_{v,u}^b > 0$$

$\implies a + b + n \in I$, so $d(u)|n + a + b$. But, by the previous line, $d(u)|n$. Since $n \in J$ is arbitrary, we can write $d(u)|\gcd(J) = d(v)$.

Symmetrically, we could conclude that $d(v)|d(u)$, so indeed $d(v) = d(u)$. \square

Let $I = \{n > 0 : P_{u,u}^n > 0\}$. If $\gcd(I) = 1$, then $\exists a, b \in I$ such that $\gcd(a, b) = 1$.

PROP 1.9

This is not true for any I (and thus relies not only on number theory). Let $\ell, m \in I$, with $\ell < m$. Let $k = m - \ell$. If $k = 1$, then $\gcd(\ell, m) = 1$. Otherwise, since $\gcd(I) = 1$, there is an $n \in I$ with $k \nmid n$. We then write $n = qk + r$, with $r \in [k - 1]$. Then $m' \in (q+1)m \in I$, since $P_{u,u}^{(q+1)m} \geq (P_{u,u}^m)^{q+1}$. Symmetrically, we can argue $\ell' = (q+1)\ell \in I$.

PROOF.

Similarly, $\ell^* := \ell' + n \in I$, since $P_{u,u}^{\ell'+n} \geq P_{u,u}^{\ell'} P_{u,u}^n$. We have

$$\begin{aligned} m' - \ell^* &= (q+1)m - (q+1)\ell - n = (q+1)(m-\ell) - n \\ &= (q+1)k - n = k - r \in [k-1] \end{aligned}$$

TODO... □

Theorem 1.4 Postage Stamp Lemma

If P is irreducible and aperiodic, then $\forall u, v \in V, \exists N$ such that $P_{u,v}^n > 0 \forall n \geq N$.

Before proving this, we note that, for $a, b \geq 1$ with $\gcd(a, b) = 1$, then for any $q \geq ab$, we can write $q = ja + kb$ for integers $j, k \geq 0$.

PROOF.

Fix $u, v \in V$. Since P is aperiodic, there are $a, b \geq 1$ with $P_{u,u}^a, P_{u,u}^b > 0$ and $\gcd(a, b) = 1$, by [Prop 1.9](#). Since P is irreducible, there is some $m > 0$ with $P_{u,v}^m > 0$. Thus, let $N = m + ab$. For any $n \geq N$, let $q = n - m$. We have that $q \geq ab$, so we can find $j, k \geq 0$ with $q = ja + kb$. Then

$$P_{u,v}^n = P_{u,v}^{q+m} = P_{u,v}^{ja+kb+m} \geq P_{u,u}^j P_{u,u}^{kb} P_{u,v}^m \geq (P_{u,u}^a)^j (P_{u,u}^b)^k P_{u,v}^m$$

All are positive, so $P_{u,v}^n > 0$, as desired. □

Theorem 1.5 Characterization of Regular Markov Chains

Let $P = (p_{u,v})_{u,v \in V}$ be a stochastic matrix, where $|V| < \infty$. Then

$$P \text{ is regular} \iff P \text{ is irreducible and aperiodic}$$

PROOF.

We first note why finiteness is necessary. Consider:



with all edges having weight 1. This graph is clearly aperiodic and irreducible, but not regular.

(\implies) We start with the "easy" direction. If P is regular, then $\exists n > 0$ s.t. $P_{u,v}^n > 0$ for all $u, v \in V$. Then, for all $u, v \in V$, we have $u \rightarrow v$ and $v \rightarrow u$. Hence, P is irreducible. Now, if P is irreducible, then for all $u \in V$, there is some $v \in V$ such that $P_{v,u} > 0$ (think about this in graph theoretic terms). Then, let $n > 0$ be such that $P_{u,u}^n$ is positive. We have

$$P_{u,u}^{n+1} \geq P_{u,v}^n P_{v,u} > 0$$

So, with $I = \{m > 0 : P_{u,u}^m > 0\}$, $d(u) = \gcd(I) \leq \gcd(n, n+1) = 1$. It follows that

$d(u) = 1$, so u is aperiodic (and hence P is aperiodic).

(\Leftarrow) By Thm 1.4, for each $u, v \in V$, there exists $N : P_{u,v}^n > 0 \forall n \geq N$. Let N^* be the maximum value of N determined over all pairs $(u, v) \in V^2$. Then, for $n \geq N^*$ and all $u, v \in V$, $P_{u,v}^n > 0$. It follows that all entries of P^n are positive, and we are done. \square

Finding Stationary Distributions

Recall that $x = (x_v : v \in V)$ is a stationary distribution if $xP = x$. Let V be finite. Then, for a stationary distribution x , we have

$$\begin{aligned} x_1 p_{1,1} + \cdots + x_n p_{n,1} &= x_1 \\ x_1 p_{1,2} + \cdots + x_n p_{n,2} &= x_2 \\ &\vdots \\ x_1 p_{1,n} + \cdots + x_n p_{n,n} &= x_n \end{aligned}$$

We have n equations, n unknowns, and a homogeneous system, so there is not a unique solution. If

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

We can compute $x = \langle t, 2t, 2t \rangle$. But, noting that x is a probability distribution, and hence $5t = 1$, this yields $x = \langle 1/5, 2/5, 2/5 \rangle$. We'll consider some special cases.

UNDIRECTED GRAPHS

This is distinct from

[Example 1.7](#)

Let $G = (V, E)$ be undirected. Then we define a THMC by

$$P_{u,v} = \begin{cases} \frac{1}{\deg(u)} & \{u, v\} \in E \\ 0 & \text{o.w.} \end{cases}$$

Let $x = (\deg(v) : v \in V)$. We have

$$\begin{aligned} (xP)_v &= \sum_{u \in V} \deg(u) P_{u,v} = \sum_{u \in N(v)} \deg(u) \cdot \frac{1}{\deg(u)} \\ &= \deg(v) \end{aligned}$$

Hence, $xP = x$. Recalling that $\sum_{v \in V} \deg(v) = 2|E|$, we conclude that

$$\left(\frac{\deg(v)}{2|E|} : v \in V \right)$$

is a stationary distribution.

UNDIRECTED WEIGHTED GRAPHS

Let $G = (V, E)$ be undirected. Then, we define a THMC by

$$P_{u,v} = \begin{cases} \frac{w(\{u,v\})}{\sum_{z \in N(u)} w(\{u,z\})} & v \in N(u) \\ 0 & \text{o.w.} \end{cases}$$

This is not distinct from [Example 1.7](#)

Let $x = (\sum_{e:e \ni v} w(e) : v \in V)$. Then we can compute $xP = x$, and similar to above,

$$x = \left(\frac{\sum_{e:e \ni v} w(e)}{2 \sum_{e \in E} w(e)} : v \in V \right)$$

is a stationary distribution.

Transience and Recurrence

Recall $T_S = \inf\{n \geq 0 : X_n \in S\}$, the "hitting time" of S . We let $R_S = \inf\{n > 0 : X_n \in S\}$.

DEF 1.20

Note that if $T_S > 0$, $T_S = R_S$. Otherwise, R_S gives the first "*return time*" to the set S .

DEF 1.21

A state $v \in V$ is called *recurrent* if $\mathbb{P}_v(R_{\{v\}} < \infty) = 1$. If all states of v are recurrent, we may

DEF 1.22

P and $X = \text{Markov}(P)$ recurrent. Otherwise, we call v *transient*, and similarly extend the notion to the transition matrix and chain when all state are transient.

DEF 1.23

For a given state $v \in V$, we call $L_v = |\{n \geq 0 : X_n = v\}|$ the *local time* of v . This notion is not probabilistic: we simply consider a realized walk on the chain (or a part of the chain). Note that, if $v = X_j$ and v is recurrent, then $L_v = \infty$.

PROP 1.10

Let $X = \text{Markov}(P)$. For any state $v \in V$ and $k > 1$,

$$\mathbb{P}_v(L_v > k) = \mathbb{P}_v(L_v > 1)^k$$

Intuitively, if $L_v > k$ when $X_0 = v$, then $L_v > k - 1$ when $X_{i_1} = v$, where i_1 is the first time we return to v .

PROOF.

Using the law of total probability:

$$\begin{aligned} \mathbb{P}_v(L_v > k) &= \mathbb{E}[\mathbb{P}_v(L_v > k|R_v)] = \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t) \mathbb{P}_v(L_v > k|R_v = t) \\ &= \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t) \mathbb{P}_v(L_v > k|R_v = t, X_t = v) \\ &= \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t) \mathbb{P}_v(L_v > k - 1) \\ &= \mathbb{P}_v(L_v > k - 1) \sum_{t=1}^{\infty} \mathbb{P}_v(R_v = t) \\ &= \mathbb{P}_v(L_v > k - 1) \mathbb{P}_v(R_v < \infty) = \mathbb{P}_v(L_v > k - 1) \mathbb{P}_v(L_v > 1) \end{aligned}$$

As $R_v = t \iff R_v = t \wedge X_t = v$

The result follows by induction. □

PROP 1.11

$$\mathbb{P}_v(L_v = \infty) = \begin{cases} 1 & v \text{ recurrent} \\ 0 & v \text{ transient} \end{cases}$$

This follows directly from [Prop 1.10](#) + monotonicity of probability.

□

PROOF.

PROP 1.12

$$\sum_{n=0}^{\infty} P_{v,v}^n = \begin{cases} \infty & v \text{ recurrent} \\ \frac{1}{1-\mathbb{P}_v(R_{\{v\}}<\infty)} & v \text{ transient} \end{cases}$$

This follows from linearity of expectation, and the fact that, for a non-negative integer variable Z ,

$$\mathbb{E}[Z] = \sum_{k=0}^{\infty} \mathbb{P}(Z > k)$$

In particular... [TODO]

□

If $u \leftrightarrow v$, then u is transient $\iff v$ is transient.

PROP 1.13

Fix $a, b \geq 0$ with $P_{u,v}^a, P_{v,u}^b > 0$. Then

$$\begin{aligned} \sum_{n=0}^{\infty} P_{v,v}^n &\geq \sum_{n=0}^{\infty} P_{v,v}^{a+b+n} = \sum_{n=0}^{\infty} P_{v,u}^b P_{u,u}^n P_{u,v}^a \\ &= P_{v,u}^b P_{u,v}^a \sum_{n=0}^{\infty} P_{u,u}^n \end{aligned}$$

Thus, if v is transient, then $\sum_{n=0}^{\infty} P_{v,v}^n < \infty$, so it must be that $\sum_{n=0}^{\infty} P_{u,u}^n < \infty$, i.e. u is transient. The argument is identical in reverse.

PROOF.

Eg. 1.9 If $u \leftrightarrow v$ and u is recurrent, then $\mathbb{P}_u(T_{\{v\}} < \infty) = 1$.

Eg. 1.10 The following chain is completely transient:



In fact, we could replace $2/3$ by p , and $1/3$ by $1-p$. In this case, the chain is irreducible. To see that it is transient, we have

$$P_{0,0}^{2n} = \binom{2n}{n} p^n (1-p)^n$$

Then

$$\sum_{n=0}^{\infty} P_{0,0}^{2n} < \sum_{n=0}^{\infty} 2^{2n} p^n (1-p)^n = \sum_{n=0}^{\infty} (4p(1-p))^n < \infty \quad \text{if } p \neq \frac{1}{2}$$

By [Prop 1.12](#). Notice that $\sum_{n=0}^{\infty} P_{0,0}^n = \sum_{n=0}^{\infty} P_{0,0}^{2n}$, since it is only possible to return on even-length cycles.

We conclude that the chain is transient when $p \neq \frac{1}{2}$.

FACT Stirling's Formula provides

$$m! \sim \left(\frac{m}{e}\right)^m \sqrt{2\pi m}$$

in that

$$\lim_{m \rightarrow \infty} \frac{m!}{\left(\frac{m}{e}\right)^m \sqrt{2\pi m}} = 1$$

This fact implies

$$e \left(\frac{n}{e}\right)^n \leq n! \leq \frac{e(n+1)}{4} \left(\frac{n+1}{e}\right)^n$$

These facts, though out of the scope of this course, can be derived from a careful treatment of Riemann sums

Eg. 1.11 We return to the previous example, letting $p = \frac{1}{2}$. Then

$$\begin{aligned} \sum_{n=0}^{\infty} P_{0,0}^{2n} &= \sum_{n=0}^{\infty} \frac{(2n)!}{(n!)^2} [p(1-p)]^n \sim \sum_{n=0}^{\infty} \frac{\left(\frac{2n}{e}\right)^{2n} \sqrt{2\pi(2n)}}{\left[\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\right]^2} [p(1-p)]^n \\ &= \sum_{n=0}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{\pi n}} = \sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty \end{aligned}$$

We conclude that the chain is recurrent when $p = \frac{1}{2}$.

PROP 1.14 If V is finite, then there is at least one recurrent state.

PROOF.

Fix an initial distribution $\alpha = (\alpha_v : v \in V)$. Then $\mathbb{P}_{\alpha}(\sum_{v \in V} L_v = \infty) = 1$. We conclude

that there is at least one state $v \in V$ with $\mathbb{P}_\alpha(L_v = \infty) > 0$. But also:

$$\begin{aligned}\mathbb{P}_\alpha(L_v = \infty) &= \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty, T_v = n) = \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty | T_v = n) \mathbb{P}_\alpha(T_v = n) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty | T_v = n, X_n = v) \mathbb{P}_\alpha(T_v = n) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_\alpha(L_v = \infty | X_n = v) \mathbb{P}_\alpha(T_v = n) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_v(L_v = \infty) \mathbb{P}_\alpha(T_v = n)\end{aligned}$$

So $\mathbb{P}_v(L_v = \infty) > 0 \implies \mathbb{P}_v(L_v = \infty) = 1$, by [Prop 1.11](#). □

Finite, irreducible chains are recurrent.

PROP 1.15

Since the chain is finite it has at least one recurrent state, by [Prop 1.14](#). Then all states must be recurrent, since the chain is irreducible, by [Prop 1.13](#). □

PROOF.

Canonical Decompositions

Fix a transition matrix P and list the communication classes of V as

$$D_1, D_2, \dots \quad (\text{transient}) \quad C_1, C_2, \dots \quad (\text{recurrent})$$

Note that we can split the chain up in this way by [Prop 1.13](#). Set $D = \cup_{i \geq 0} D_i$. Then the *canonical decomposition* of the chain is defined to be

$$D \sqcup C_1 \sqcup C_2 \sqcup \dots$$

We say that a communication class C is *closed* if, for any $u \in C, v \notin C$, $p_{u,v} = 0$. Intuitively, if $X_0 \in C$, or we enter C at some later time, we will never leave C .

DEF 1.24

If C is a recurrent communication class, then C is closed.

PROP 1.16

Fix $u \in C, v \notin C$. Suppose $v \mapsto u$. If $p_{u,v} > 0$, then $v \mapsto v$, so $v \in C$. Suppose $v \not\mapsto u$. Then $\mathbb{P}_u(R_u = \infty) \geq \mathbb{P}_u(X_1 = v) = p_{u,v}$. But $\mathbb{P}_u(R_u = \infty) = 0$, since u is recurrent. It follows that $p_{u,v} = 0$. □

PROOF.

The converse of [Prop 1.16](#) is not true in generality, but it is in the finite case:

Finite, closed communication classes are recurrent.

PROP 1.17

From any starting state in C , we must visit some state $u \in C$ infinitely often, as $|C| < \infty$ and $X_t \in C \forall t$. But recurrence is a class property by [Prop 1.13](#). Hence, all of C is recurrent. □

PROOF.

When our communication classes are closed, we have



Proof of Fundamental Theorem of Markov Chains

Recall [Thm 1.3](#):

Every finite, regular stochastic matrix P has a limiting distribution π .

We will prove this in two steps. First, we will find some stationary distribution. Then, we will prove that this is a limiting distribution.

Theorem 1.6 Existence Theorem

Let P be irreducible and recurrent. Let $(X_n : n \geq 0) = \text{Markov}(P)$. Fix $u \in V$, which we call a reference vertex, and, for any $v \in V$, define

$$\gamma_v = \mathbb{E}_u[|\{0 \leq n < R_u : X_n = v\}|]$$

Let $\gamma = (\gamma_v : v \in V)$. Then $\gamma P = \gamma$, and $0 < \gamma_v < \infty \forall v \in V$.

By \mathbb{E}_u , we mean the expectation, under the assumption that $X_0 = u$

PROOF.

Observe that $\gamma_u = 1$. Write

$$\gamma_v = \mathbb{E}_u \left[\sum_{n=0}^{R_u-1} \mathbb{1}_{X_n=v} \right] = \mathbb{E}_u \left[\sum_{n=1}^{R_u} \mathbb{1}_{X_n=v} \right] = \mathbb{E}_u \left[\sum_{n=1}^{\infty} \mathbb{1}_{X_n=v} \mathbb{1}_{n \leq R_u} \right]$$

We utilize a second indicator variable in order to use linearity of expectation (otherwise, our sum would index over a random variable, which is not valid). Then

$$\gamma_v = \sum_{n=1}^{\infty} \mathbb{E}_u[\mathbb{1}_{X_n=v} \mathbb{1}_{n \leq R_u}] = \sum_{n=1}^{\infty} \mathbb{P}_u(X_n = v, n \leq R_u)$$

Now, using the law of total probability,

$$\begin{aligned} \mathbb{P}_u(X_n = v, n \leq R_u) &= \sum_{w \in V} \mathbb{P}_u(X_{n-1} = w, X_n = v, n \leq R_u) \\ &= \sum_{w \in W} \mathbb{P}_u(X_n = v | X_{n-1} = w, n \leq R_u) \mathbb{P}_u(X_{n-1} = w, n \leq R_u) \\ &= \sum_{w \in V} \mathbb{P}_u(X_{n-1} = w, n \leq R_u) P_{w,v} \end{aligned}$$

So

$$\begin{aligned}
\gamma_v &= \sum_{n=1}^{\infty} \sum_{w \in W} P_{w,v} \mathbb{P}_u(X_{n-1} = w, n \leq R_u) = \sum_{w \in V} P_{w,v} \left(\sum_{n=1}^{\infty} \mathbb{P}_u(X_{n-1} = w, n-1 < R_u) \right) \\
&= \sum_{w \in V} P_{w,v} \sum_{n=0}^{\infty} \mathbb{P}_u(X_n = w, n < R_u) = \sum_{w \in V} P_{w,v} \sum_{n=0}^{\infty} \mathbb{E}_u[\mathbb{1}_{X_n=w} \mathbb{1}_{n < R_u}] \\
&= \sum_{w \in V} P_{w,v} \mathbb{E}_u \left[\sum_{n=0}^{R_u-1} \mathbb{1}_{X_n=w} \right] = \sum_{w \in V} P_{w,v} \gamma_w = (\gamma P)_v
\end{aligned}$$

$\implies \gamma P = \gamma$. Furthermore, $\gamma_v > 0$, since $u \mapsto v$. Letting $n : P_{v,u}^n > 0$, we have $\gamma_u = (\gamma P^n)_u \geq \gamma_v P_{v,u}^n$. Noting that $\gamma_u = 1$, this shows $\gamma_v < \infty$. \square

As a corollary, under the same conditions, if $\mathbb{E}_u[R_u]$ is finite, $\pi = (\pi_v : v \in V)$, where

PROP 1.18

$$\pi_v = \frac{\gamma_v}{\sum_{w \in V} \gamma_w}$$

is a stationary distribution.

PROOF.

Observe

$$\sum_{w \in V} \gamma_w = \sum_{w \in V} \mathbb{E}_i \left[\sum_{n=0}^{R_i-1} \mathbb{1}_{X_n=w} \right] = \mathbb{E}_u \left[\sum_{n=0}^{R_i-1} \sum_{w \in V} \mathbb{1}_{X_n=w} \right] = \mathbb{E}_u[R_u]$$

So we require that $\mathbb{E}_u[R_u] < \infty$, then a stationary distribution exists. \square

An an exercise, if u is recurrent and $u \mapsto v$, then $\mathbb{P}_u(T_v < \infty) = 1 \ \forall v \in V$. π , as defined above, is a limiting distribution.

PROP 1.19

Consider two independent copies $X_n, Y_n = \text{Markov}(P)$. Then (X_n, Y_n) is a Markov chain with transition matrix $Q = (q_{(v,w),(x,y)})_{(v,w),(x,y) \in V \times V}$. In particular,

$$q_{(v,w),(x,y)} = p_{v,x} p_{w,y}$$

Fix some state $u \in V$. Let α be the initial distribution that has $X_0 = u$ and $Y_0 \sim \pi$.

$$\alpha_{(x,y)} = \begin{cases} \pi(y) & x = u \\ 0 & \text{o.w.} \end{cases}$$

Remark that, as π is stationary, $Y_n \sim \pi \ \forall n \geq 0$. For any $u \in V$, we'd like that $P_{u,v}^n \rightarrow \pi(v)$.

Let $M = \inf\{n \geq 0 : X_n = Y_n\}$ be the first meeting time of the X_n and Y_n chains. If P is

PROOF.

finite and regular, then Q is finite and regular. Then

$$\mathbb{P}_\alpha(M < \infty) \geq \mathbb{P}_\alpha(T_{v,v} < \infty) = \sum_{(x,y) \in V \times V} \alpha(x,y) \mathbb{P}_{(x,y)}(T_{v,v} < \infty) = 1$$

It follows that $\lim_{n \rightarrow \infty} \mathbb{P}_\alpha(M > n) = 0$. We claim that $\mathbb{P}_\alpha(X_n = v, M \leq n) = \mathbb{P}_\alpha(Y_n = v, M \leq n) \forall n \geq 0$. Assuming this, then

$$\begin{aligned} \mathbb{P}_u(X_n = v) &= \mathbb{P}_\alpha(X_n = v) = \mathbb{P}_\alpha(X_n = v, M \leq n) + \mathbb{P}_\alpha(X_n = v, M > n) \\ \mathbb{P}_\pi(Y_n = v) &= \mathbb{P}_\alpha(Y_n = v) = \mathbb{P}_\alpha(Y_n = v, M \leq n) + \mathbb{P}_\alpha(Y_n = v, M > n) \\ \implies \mathbb{P}_u(X_n = v) - \pi(v) &= \mathbb{P}_\alpha(X_n = v, M > n) - \mathbb{P}_\alpha(Y_n = v, M > n) \\ \implies |\mathbb{P}_u(X_n = v) - \pi(v)| &= |\mathbb{P}_\alpha(X_n = v, M > n) - \mathbb{P}_\alpha(Y_n = v, M > n)| \leq \mathbb{P}_\alpha(M > n) \end{aligned}$$

But $P_\alpha(M > n) \rightarrow 0$, so it must be that $\mathbb{P}_u(X_n = v) \rightarrow \pi(v)$.

We still must show the claim, however.

$$\begin{aligned} \mathbb{P}_\alpha(X_n = v, M \leq n) &= \sum_{k=0}^n \sum_{w \in V} \mathbb{P}_\alpha(X_n = v, M = k, X_k = w) \\ &= \sum_{k=0}^n \sum_{w \in V} \mathbb{P}_\alpha(M = k, X_k = w) \mathbb{P}_\alpha(X_n = v | X_k = w, M = k) \\ &\stackrel{\text{MP}}{=} \sum_{k=0}^n \sum_{w \in V} \mathbb{P}_\alpha(M = k, X_k = w) P_{w,v}^{n-k} \end{aligned}$$

and now we reverse!

$$\begin{aligned} &= \sum_{k=0}^n \sum_{w \in V} \mathbb{P}_\alpha(M = k, Y_k = w) P_{w,v}^{n-k} \\ &\stackrel{\text{MP}}{=} \sum_{k=0}^n \sum_{w \in V} \mathbb{P}_\alpha(M = k, Y_k = w) \mathbb{P}_\alpha(X_n = v | Y_k = w, M = k) \\ &= \mathbb{P}_\alpha(Y_n = v, M \leq n) \quad \square \end{aligned}$$

Hitting Times and Absorbing States

DEF 1.26

We call a state $v \in V$ *absorbing* if $\mathbb{P}_v(X_1 = v) = 1$.

Eg. 1.12 Consider a game where we continually bet \$1 with winning probability p , until we earn $\$k$ or run out of money. Modeling the state space as our balance, we have absorbing states at $\$0$ and $\$k$.

Observe that each absorbing state forms a recurrent communication class of size 1. As a consequence, we have a variant of the canonical form, where $A \subseteq V$ are absorbing states, and we write $V = A \sqcup (V \setminus A)$. Schematically,

$$\left[\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right]$$

INDEX OF DEFINITIONS

absorbing	1.26	Markov chain	1.2
accessible	1.14	Markov property	1.4
aperiodic	1.19	period	1.18
canonical decomposition	1.24	recurrent	1.21
closed	1.25	regular	1.13
communicate	1.15	return time	1.20
communication class	1.16	state space	1.3
conditional density	1.1	stationary	1.11
hitting time	1.7	stochastic matrix	1.8
initial distribution	1.9	time-homogeneous	1.5
irreducible	1.17	transient	1.22
law of large numbers	1.10	transition matrix	1.6
limiting distribution	1.12		
local time	1.23		