

# Stochastic Processes

MATH 447

Nicholas Hayek

*Taught by Prof. Louigi Addario-Berry*

## CONTENTS

<b>I</b>	<b>Markov Chains</b>	<b>3</b>
	<i>Time-Homogeneous Markov Chains</i>	
	<i>Multi-Step Transition Probabilities</i>	
	<i>Long Term Behavior</i>	
	<i>Classification of States</i>	
	<b>Index of Definitions</b>	

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

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

**Fig. 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.

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

**Fig. 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.

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

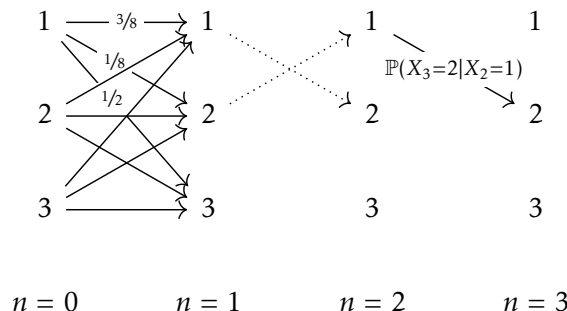
**Fig. 1.3** Choose  $k$  independent random points in the square  $[0, \sqrt{k}]^2$ . On average, then, there is 1 point within any unit square  $U \subseteq [0, \sqrt{k}]^2$ .

DEF 1.1 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:

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

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 1.4

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

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 1.5

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

**Eg. 1.4** Recall the game Snakes and Ladders. A  $6 \times 6$  grid is indexed  $1, \dots, 36$ . 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

$$\mathbb{P}(X_3 = a | X_2 = b) = 1/9$$

$$\mathbb{P}(X_3 = a | X_2 = b, X_1 = c) = 1/8$$

$$\mathbb{P}(X_3 = a | X_2 = b, X_1 = a) = 0$$

so we do not satisfy the Markov property.

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

- What is the average number of rounds to finish? 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?

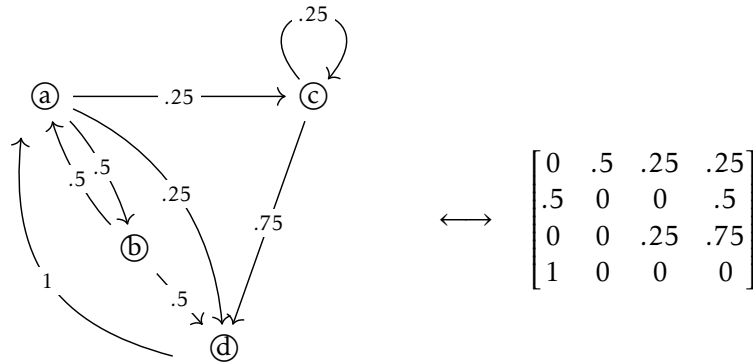
DEF 1.6 A matrix  $P = (p_{u,v})_{(u,v) \in V^2}$  is called a *stochastic matrix* if every row sums to 1, i.e.

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

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:



**Fig. 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  $w(k, k+1) = \alpha$ ,  $w(k-1, k) = \frac{\alpha}{2}$ .

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

### 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 PROOF.

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

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.

### Theorem 1.1 Markov Property

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

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

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

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

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  $\alpha$  as a subscript of our state probabilities. For instance, DEF 1.7

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

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

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

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  $S$  to be the set of tuples  $(u_0, \dots, u_{n-1})$  that guarantee  $E$ . It follows that

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

Similarly, we have

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

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  $P^k$ .

PROOF.

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

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

### 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

**DEF 1.8** Recall from probability the *law of large numbers*: if  $Y_n : n \geq 1$  are IID with common mean  $\mu$ , then  $\frac{S_n}{n} \rightarrow \mu$  in probability, where  $S_n = \sum_{i=1}^n Y_i$ , i.e.  $\forall \varepsilon > 0$ ,

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

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:

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

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  $\pi$ , then  $X_1$  also has distribution  $\pi$ . 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  $P$ .

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

DEF 1.9

DEF 1.10

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

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

If  $\lambda$  is a limiting distribution for  $P$ , then  $\lambda$  is stationary for  $P$ .

PROP 1.6



PROOF.

Fix any initial distribution  $\alpha$ , 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 1.11 A stochastic matrix  $P$  is called *regular* if  $\exists n \geq 1$  such that all entries of  $P^n$  are positive.

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

### Theorem 1.3 Fundamental Theorem of Markov Chains

Every regular stochastic matrix  $P$  has a limiting distribution  $\pi$ .

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  $\mathbb{P}_u(X_n = v) \rightarrow \pi_v \forall u, v \in V$ .*

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

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  $\pi$  is a distribution).

### Classification of States

DEF 1.12 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 1.13 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 an equivalence relation. Hence, the state space  $V$  can be written as a disjoint union of mutually-communicating states, called a

DEF 1.14 *communication class*. Note that, in the directed graph generated by  $P$ , these correspond to the strongly connected components.

DEF 1.15 We say that  $P$  is *irreducible* if there is only one communication class.

Clearly, if  $P$  is regular, then it is irreducible

PROP 1.7  $u \rightarrow v \iff \mathbb{P}_u(T_{\{v\}} < \infty) > 0$ .

DEF 1.16 The *period* of a state  $u \in V$  is

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

DEF 1.17 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.

PROP 1.8 If  $u \leftrightarrow v$ , then  $d(u) = d(v)$ .

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

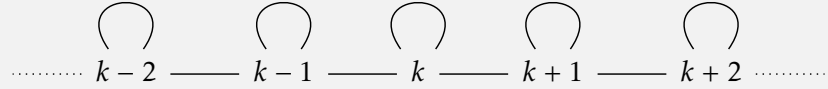
PROOF.

#### Theorem 1.4 theorem name

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

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

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).

( $\impliedby$ )  $\square$

PROOF.

We now have the tools to prove [Thm 1.3](#), the fundamental theorem of Markov chains.

**Statement:** every regular stochastic matrix  $P$  has a limiting distribution.

Let  $P = (p_{u,v})_{u,v \in V}$  be a stochastic  $\square$

PF. OF 1.3

## INDEX OF DEFINITIONS

accessible <a href="#">1.12</a>	Markov property <a href="#">1.3</a>
aperiodic <a href="#">1.17</a>	
communicate <a href="#">1.13</a>	period <a href="#">1.16</a>
communication class <a href="#">1.14</a>	regular <a href="#">1.11</a>
initial distribution <a href="#">1.7</a>	state space <a href="#">1.2</a>
irreducible <a href="#">1.15</a>	stationary <a href="#">1.9</a>
law of large numbers <a href="#">1.8</a>	stochastic matrix <a href="#">1.6</a>
limiting distribution <a href="#">1.10</a>	
Markov chain <a href="#">1.1</a>	time-homogeneous <a href="#">1.4</a>
	transition matrix <a href="#">1.5</a>