

# Markov Chains

Richard Lockhart

Simon Fraser University

STAT 870 — Summer 2013



# Purposes of Today's Lecture

- Define Markov Chain, transition matrix.
- Prove Chapman-Kolmogorov equations.
- Introduce classification of states: communicating classes.
- Define hitting times; prove the Strong Markov property.
- Define initial distribution.
- Establish relation between mean return time and stationary initial distribution.
- Discuss ergodic theorem.



# Markov Chains

- **Stochastic process:** family  $\{X_i; i \in I\}$  of rvs;  $I$  the **index set**. Often  $I \subset \mathbb{R}$ , e.g.  $[0, \infty)$ ,  $[0, 1]$

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

or

$$\mathbb{N} = \{0, 1, 2, \dots\}.$$

- **Continuous time:**  $I$  is an interval
- **Discrete time:**  $I \subset \mathbb{Z}$ .
- Generally all  $X_n$  take values in **state space**  $S$ . In following  $S$  is a finite or countable set; each  $X_n$  is discrete.
- Usually  $S$  is  $\mathbb{Z}$ ,  $\mathbb{N}$  or  $\{0, \dots, m\}$  for some finite  $m$ .



## Definition

- **Markov Chain:** stochastic process  $X_n; n \in \mathbb{N}$ . taking values in a finite or countable set  $S$  such that for every  $n$  and every event of the form

$$A = \{(X_0, \dots, X_{n-1}) \in B \subset S^n\}$$

we have

$$P(X_{n+1} = j | X_n = i, A) = P(X_1 = j | X_0 = i) \quad (1)$$

- Notation:  $\mathbf{P}$  is the (possibly infinite) array with elements

$$P_{ij} = P(X_1 = j | X_0 = i)$$

indexed by  $i, j \in S$ .

- $\mathbf{P}$  is the (one step) **transition matrix** of the Markov Chain.
- WARNING: in (1) we require the condition to hold **only** when

$$P(X_n = i, A) > 0$$

- NOTE: this chain has *stationary* transitions.



# Stochastic matrices

- Evidently the entries in  $\mathbf{P}$  are non-negative and

$$\sum_j P_{ij} = 1$$

for all  $i \in S$ .

- Any such matrix is called **stochastic**.
- We define powers of  $\mathbf{P}$  by

$$(\mathbf{P}^n)_{ij} = \sum_k (\mathbf{P}^{n-1})_{ik} P_{kj}$$

- Notice that even if  $S$  is infinite these sums converge absolutely because for all  $i$

$$\sum_j \mathbf{P}_{ij}^n = 1.$$



# Chapman-Kolmogorov Equations 1

Condition on  $X_{l+n-1}$  to compute  $P(X_{l+n} = j | X_l = i)$ .

$$\begin{aligned} P(X_{l+n} = j | X_l = i) &= \sum_k P(X_{l+n} = j, X_{l+n-1} = k | X_l = i) \\ &= \sum_k P(X_{l+n} = j | X_{l+n-1} = k, X_l = i) P(X_{l+n-1} = k | X_l = i) \\ &= \sum_k P(X_1 = j | X_0 = k) P(X_{l+n-1} = k | X_l = i) \\ &= \sum_k P(X_{l+n-1} = k | X_l = i) \mathbf{P}_{kj} \end{aligned}$$

Now condition on  $X_{l+n-2}$  to get

$$P(X_{l+n} = j | X_l = i) = \sum_{k_1 k_2} \mathbf{P}_{k_1 k_2} \mathbf{P}_{k_2 j} P(X_{l+n-2} = k_1 | X_l = i)$$



## Chapman-Kolmogorov Equations 2

Notice: sum over  $k_2$  computes  $k_1, j$  entry in matrix  $\mathbf{P}\mathbf{P} = \mathbf{P}^2$ .

$$P(X_{l+n} = j | X_l = i) = \sum_{k_1} (\mathbf{P}^2)_{k_1, j} P(X_{l+n-2} = k_1 | X_l = i)$$

We may now prove by induction on  $n$  that

$$P(X_{l+n} = j | X_l = i) = (\mathbf{P}^n)_{ij}.$$

This proves Chapman-Kolmogorov equations:

$$P(X_{l+m+n} = j | X_l = i) = \sum_k P(X_{l+m} = k | X_l = i) P(X_{l+m+n} = j | X_{l+m} = k)$$

These are simply a restatement of the identity

$$\mathbf{P}^{n+m} = \mathbf{P}^n \mathbf{P}^m.$$



## Remarks

- These probabilities depend on  $m$  and  $n$  but **not** on  $l$ .
- We say the chain has **stationary** transition probabilities.
- A more general definition of Markov chain than (1) is

$$P(X_{n+1} = j | X_n = i, A) = P(X_{n+1} = j | X_n = i).$$

Notice RHS now permitted to depend on  $n$ .

- Define  $\mathbf{P}^{n,m}$ : matrix with  $i, j$ th entry

$$P(X_m = j | X_n = i)$$

for  $m > n$ .

- Get more general form of Chapman-Kolmogorov equations:

$$\mathbf{P}^{r,s} \mathbf{P}^{s,t} = \mathbf{P}^{r,t}$$

This chain does not have stationary transitions.

- Calculations above involve sums with all terms are positive. They therefore apply even if the state space  $S$  is countably infinite.





## Extensions of the Markov Property

- Function  $f(x_0, x_1, \dots)$  defined on  $S^\infty =$  all infinite sequences of points in  $S$ .
- Example  $f$  might be

$$\sum_{k=0}^{\infty} 2^{-k} 1(x_k = 0).$$

### Theorem

Let  $B_n$  be the event

$$f(X_n, X_{n+1}, \dots) \in C$$

for suitable  $C$  in range space of  $f$ . Then

$$P(B_n | X_n = x, A) = P(B_0 | X_0 = x) \quad (2)$$

for any event  $A$  of the form

$$\{(X_0, \dots, X_{n-1}) \in D\}$$

## Extensions of the Markov Property 2

### Theorem

*With  $B_n$  as before*

$$P(AB_n|X_n = x) = P(A|X_n = x)P(B_n|X_n = x) \quad (3)$$

“Given the present the past and future are conditionally independent.”



## Markov Property: Proof of (2)

Special case:

$$B_n = \{X_n = x, X_{n+1} = x_1, \dots, X_{n+m} = x_m\}$$

LHS of (2) evaluated by repeated conditioning (cf. Chapman-Kolmogorov):

$$P_{X, X_1} P_{X_1, X_2} \cdots P_{X_{m-1}, X_m}$$

Same for RHS.

Events defined from  $X_n, \dots, X_{n+m}$ : sum over appropriate vectors  $x, x_1, \dots, x_m$ .

General case: monotone class techniques.

To prove (3) write, using (2):

$$\begin{aligned} P(AB_n | X_n = x) &= P(B_n | X_n = x, A) P(A | X_n = x) \\ &= P(B_n | X_n = x) P(A | X_n = x) \end{aligned}$$



# Classification of States

If an entry  $\mathbf{P}_{ij}$  is 0 it is not possible to go from state  $i$  to state  $j$  in one step. It may be possible to make the transition in some larger number of steps, however. We say  $i$  **leads to**  $j$  (or  $j$  is accessible from  $i$ ) if there is an integer  $n \geq 0$  such that

$$P(X_n = j | X_0 = i) > 0.$$

We use the notation  $i \rightsquigarrow j$ . Define  $\mathbf{P}^0$  to be identity matrix  $\mathbf{I}$ . Then  $i \rightsquigarrow j$  if there is an  $n \geq 0$  for which  $(\mathbf{P}^n)_{ij} > 0$ .

States  $i$  and  $j$  **communicate** if  $i \rightsquigarrow j$  and  $j \rightsquigarrow i$ .

Write  $i \leftrightarrow j$  if  $i$  and  $j$  communicate.

Communication is an equivalence relation: reflexive, symmetric, transitive relation on states of  $S$ .



# Equivalence Classes

More precisely:

**Reflexive:** for all  $i$  we have  $i \leftrightarrow i$ .

**Symmetric:** if  $i \leftrightarrow j$  then  $j \leftrightarrow i$ .

**Transitive:** if  $i \leftrightarrow j$  and  $j \leftrightarrow k$  then  $i \leftrightarrow k$ .

Proof:

Reflexive: follows from inclusion of  $n = 0$  in definition of leads to.

Symmetry is obvious.

Transitivity: suffices to check that  $i \rightsquigarrow j$  and  $j \rightsquigarrow k$  imply that  $i \rightsquigarrow k$ . But if  $(\mathbf{P}^m)_{ij} > 0$  and  $(\mathbf{P}^n)_{jk} > 0$  then

$$\begin{aligned}(\mathbf{P}^{m+n})_{ik} &= \sum_l (\mathbf{P}^m)_{il} (\mathbf{P}^n)_{lk} \\ &\geq (\mathbf{P}^m)_{ij} (\mathbf{P}^n)_{jk} \\ &> 0\end{aligned}$$



# Equivalence Classes

Any equivalence relation on a set partitions the set into **equivalence classes**; two elements are in the same equivalence class if and only if they are equivalent.

Communication partitions  $S$  into equivalence classes called **communicating classes**.



## Example

Example:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$



## Example

Find communicating classes: start with say state 1, see where it leads.

- $1 \rightsquigarrow 2$ ,  $1 \rightsquigarrow 3$  and  $1 \rightsquigarrow 4$  in row 1.
- Row 4:  $4 \rightsquigarrow 1$ . So: (transitivity) 1, 2, 3 and 4 all in the same communicating class.
- Claim: none of these leads to 5, 6, 7 or 8.
- Suppose  $i \in \{1, 2, 3, 4\}$  and  $j \in \{5, 6, 7, 8\}$ .
- Then  $(\mathbf{P}^n)_{ij}$  is sum of products of  $\mathbf{P}_{kl}$ .
- Cannot be positive unless there is a sequence  $i_0 = i, i_1, \dots, i_n = j$  with  $\mathbf{P}_{i_{k-1}, i_k} > 0$  for  $k = 1, \dots, n$ .
- Consider first  $k$  for which  $i_k \in \{5, 6, 7, 8\}$
- Then  $i_{k-1} \in \{1, 2, 3, 4\}$  and so  $\mathbf{P}_{i_{k-1}, i_k} = 0$ .





## Example

So:  $\{1, 2, 3, 4\}$  is a communicating class.

- $5 \rightsquigarrow 1$ ,  $5 \rightsquigarrow 2$ ,  $5 \rightsquigarrow 3$  and  $5 \rightsquigarrow 4$ .
- None of these lead to any of  $\{5, 6, 7, 8\}$  so  $\{5\}$  must be communicating class.
- Similarly  $\{6\}$  and  $\{7, 8\}$  are communicating classes.

Note: states 5 and 6 have special property. Each time you are in either state you run a risk of going to one of the states 1, 2, 3 or 4. Eventually you will make such a transition and then never return to state 5 or 6. States 5 and 6 are **transient**.



# Hitting Times

- To make this precise define hitting times:

$$T_k = \min\{n > 0 : X_n = k\}$$

- We define

$$f_k = P(T_k < \infty | X_0 = k)$$

- State  $k$  is **transient** if  $f_k < 1$  and **recurrent** if  $f_k = 1$ .



# Geometric Distribution

Let  $N_k$  be number of times chain is ever in state  $k$ .

Claims:

- 1 If  $f_k < 1$  then  $N_k$  has a Geometric distribution:

$$P(N_k = r | X_0 = k) = f_k^{r-1}(1 - f_k)$$

for  $r = 1, 2, \dots$

- 2 If  $f_k = 1$  then

$$P(N_k = \infty | X_0 = k) = 1$$



## Stopping Times

**Def'n:** A **Stopping time** for the Markov chain is a random variable  $T$  taking values in  $\{0, 1, \dots\} \cup \{\infty\}$  such that for each finite  $k$  there is a function  $f_k$  such that

$$1(T = k) = f_k(X_0, \dots, X_k)$$

Notice that  $T_k$  in theorem is a stopping time.

Standard shorthand notation: by

$$P^x(A)$$

we mean

$$P(A|X_0 = x).$$

Similarly we define

$$E^x(Y) = E(Y|X_0 = x).$$



# Strong Markov Property

Goal: explain and prove

$$E(f(X_T, \dots) | X_T, \dots, X_0) = E^{X_T}(f(X_0, \dots))$$

Simpler claim:

$$P(X_{T+1} = j | X_T = i) = \mathbf{P}_{ij} = P^i(X_1 = j).$$

Explanation: given what happens up to and including a random stopping time  $T$ , future behaviour of chain is like that of chain started from  $X_T$ .



# Proof of Strong Markov Property

Notation:  $A_k = \{X_k = i, T = k\}$

Notice:  $A_k = \{X_T = i, T = k\}$ :

$$\begin{aligned}P(X_{T+1} = j | X_T = i) &= \frac{P(X_{T+1} = j, X_T = i)}{P(X_T = i)} \\&= \frac{\sum_k P(X_{T+1} = j, X_T = i, T = k)}{\sum_k P(X_T = i, T = k)} \\&= \frac{\sum_k P(X_{k+1} = j, A_k)}{\sum_k P(A_k)} \\&= \frac{\sum_k P(X_{k+1} = j | A_k) P(A_k)}{\sum_k P(A_k)} \\&= \frac{\sum_k P(X_1 = j | X_0 = i) P(A_k)}{\sum_k P(A_k)} \\&= \mathbf{P}_{i,j}\end{aligned}$$



# More Proof of Strong Markov Property

- Notice use of fact that  $T = k$  is event defined in terms of  $X_0, \dots, X_k$ .
- Technical problems with proof:
  - ▶ It might be that  $P(T = \infty) > 0$ . What are  $X_T$  and  $X_{T+1}$  on the event  $T = \infty$ .
  - ▶ Answer: condition also on  $T < \infty$ .
  - ▶ Prove formula only for stopping times where  $\{T < \infty\} \cap \{X_T = i\}$  has positive probability.
- We will now fix up these technical details.



## Proof of Strong Markov Property continued

Suppose  $f(x_0, x_1, \dots)$  is a (measurable) function on  $S^{\mathbb{N}}$ . Put

$$Y_n = f(X_n, X_{n+1}, \dots).$$

Assume  $E(|Y_0| | X_0 = x) < \infty$  for all  $x$ . Claim:

$$E(Y_n | X_n, A) = E^{X_n}(Y_0) \tag{4}$$

whenever  $A$  is any event defined in terms of  $X_0, \dots, X_n$ .

**1** Family of  $f$  for which claim holds includes all indicators; see extension of Markov Property in previous lecture.

**2** family of  $f$  for which claim is true is vector space (so if  $f, g$  in family then so is  $af + bg$  for any constants  $a$  and  $b$ ).





## Proof of Strong Markov Property continued

- So family of  $f$  for which claim is true includes all simple functions.
- family of  $f$  for which claim true is closed under monotone increasing limits (of non-negative  $f_n$ ) by the Monotone Convergence theorem.
- So claim true for every non-negative integrable  $f$ .
- Claim follows for integrable  $f$  by linearity.

Aside on “measurable”: what sorts of events can be defined in terms of a family  $\{Y_i : i \in I\}$ ?



# Strong Markov Property Commentary

- Natural: any event of form  $(Y_{i_1}, \dots, Y_{i_k}) \in C$  is “defined in terms of the family” for any finite set  $i_1, \dots, i_k$  and any (Borel) set  $C$  in  $S^k$ .
- For countable  $S$ : each singleton  $(s_1, \dots, s_k) \in S^k$  Borel. So every subset of  $S^k$  Borel.
- Natural: if you can define each of a sequence of events  $A_n$  in terms of the  $Y$ s then the definition “there exists an  $n$  such that (definition of  $A_n$ )...” defines  $\cup A_n$ .
- Natural: if  $A$  is definable in terms of the  $Y$ s then  $A^c$  can be defined from the  $Y$ s by just inserting the phrase “It is not true that” in front of the definition of  $A$ .
- So family of events definable in terms of the family  $\{Y_i : i \in I\}$  is a  $\sigma$ -field which includes every event of the form  $(Y_{i_1}, \dots, Y_{i_k}) \in C$ .
- We call the smallest such  $\sigma$ -field,  $\mathcal{F}(\{Y_i : i \in I\})$ , the  $\sigma$ -field generated by the family  $\{Y_i : i \in I\}$ .



## Use of Strong Markov Property

- Toss coin till I get a head. What is the expected number of tosses?
- Define state to be 0 if toss is tail and 1 if toss is heads.
- Define  $X_0 = 0$ .
- Let  $N = \min\{n > 0 : X_n = 1\}$ . Want

$$E(N) = E^0(N)$$

- Note: if  $X_1 = 1$  then  $N = 1$ . If  $X_1 = 0$  then  $N = 1 + \min\{n > 0 : X_{n+1} = 1\}$ .
- In symbols:

$$N = \min\{n > 0 : X_n = 1\} = f(X_1, X_2, \dots)$$

and

$$N = 1 + 1(X_1 = 0)f(X_2, X_3, \dots)$$



## Use of Strong Markov Property

Take expected values starting from 0:

$$E^0(N) = 1 + E^0\{1(X_1 = 0)f(X_2, X_3, \dots)\}$$

Condition on  $X_1$  and get

$$E^0(N) = 1 + E^0[E\{1(X_1 = 0)f(X_2, \dots)|X_1\}]$$

But

$$\begin{aligned} E\{1(X_1 = 0)f(X_2, X_3, \dots)|X_1\} &= 1(X_1 = 0)E^{X_1}\{f(X_1, X_2, \dots)\} \\ &= 1(X_1 = 0)E^0\{f(X_1, X_2, \dots)\} \\ &= 1(X_1 = 0)E^0(N) \end{aligned}$$



# Use of Strong Markov Property

- Hence

$$E^0(N) = 1 + pE^0\{N\}$$

where  $p$  is the probability of tails.

- Solve for  $E(N)$  to get

$$E(N) = \frac{1}{1-p}$$

- This is the formula for expected value of the sort of geometric which starts at 1 and has  $p$  being the probability of failure.



# Initial Distributions

- Meaning of unconditional expected values?
- Markov property specifies only cond'l probs; no way to deduce marginal distributions.
- For every dstbn  $\pi$  on  $S$  and transition matrix  $\mathbf{P}$  there is a stochastic process  $X_0, X_1, \dots$  with

$$P(X_0 = k) = \pi_k$$

and which is a Markov Chain with transition matrix  $\mathbf{P}$ .

- Note Strong Markov Property proof used only conditional expectations.
- Notation:  $\pi$  a probability on  $S$ .  $E^\pi$  and  $P^\pi$  are expected values and probabilities for chain with initial distribution  $\pi$ .



## Summary of easy properties

- For any sequence of states  $i_0, \dots, i_k$

$$P(X_0 = i_0, \dots, X_k = i_k) = \pi_{i_0} \mathbf{P}_{i_0 i_1} \cdots \mathbf{P}_{i_{k-1} i_k}$$

- For any event  $A$ :

$$\mathbf{P}^\pi(A) = \sum_k \pi_k \mathbf{P}^k(A)$$

- For any bounded rv  $Y = f(X_0, \dots)$

$$\mathbf{E}^\pi(Y) = \sum_k \pi_k \mathbf{E}^k(Y)$$



## Recurrence and Transience

- Now consider a transient state  $k$ , that is, a state for which

$$f_k = P^k(T_k < \infty) < 1$$

- Note that  $T_k = \min\{n > 0 : X_n = k\}$  is a stopping time.
- Let  $N_k$  be the number of visits to state  $k$ . That is

$$N_k = \sum_{n=0}^{\infty} 1(X_n = k)$$

- Notice that if we define the function

$$f(x_0, x_1, \dots) = \sum_{n=0}^{\infty} 1(x_n = k)$$

then

$$N_k = f(X_0, X_1, \dots)$$





## Recurrence and Transience 2

Notice, also, that on the event  $T_k < \infty$

$$N_k = 1 + f(X_{T_k}, X_{T_k+1}, \dots)$$

and on the event  $T_k = \infty$  we have

$$N_k = 1$$



## Proof

In short:

$$N_k = 1 + f(X_{T_k}, X_{T_k+1}, \dots)1(T_k < \infty)$$

Hence

$$\begin{aligned} \mathbf{P}^k(N_k = r) &= \mathbf{E}^k \{P(N_k = r | \mathcal{F}_T)\} \\ &= \mathbf{E}^k [P\{1 + f(X_{T_k}, X_{T_k+1}, \dots)1(T_k < \infty) = r | \mathcal{F}_T\}] \\ &= \mathbf{E}^k \left[ 1(T_k < \infty) P^{X_{T_k}} \{f(X_0, X_1, \dots) = r - 1\} \right] \\ &= \mathbf{E}^k \left\{ 1(T_k < \infty) P^k(N_k = r - 1) \right\} \\ &= \mathbf{E}^k \{1(T_k < \infty)\} P^k(N_k = r - 1) \\ &= f_k P^k(N_k = r - 1) \end{aligned}$$

It is easily verified by induction, then, that

$$\mathbf{P}^k(N_k = r) = f_k^{r-1} P^k(N_k = 1)$$



## Proof

- But  $N_k = 1$  if and only if  $T_k = \infty$  so

$$\mathbf{P}^k(N_k = r) = f_k^{r-1}(1 - f_k)$$

so  $N_k$  has (chain starts from  $k$ ) Geometric dist'n, mean  $1/(1 - f_k)$ .

- Argument also shows that if  $f_k = 1$  then

$$P^k(N_k = 1) = P^k(N_k = 2) = \dots$$

which can only happen if all these probabilities are 0. Thus if  $f_k = 1$

$$P(N_k = \infty) = 1$$

- Since  $N_k = \sum_{n=0}^{\infty} 1(X_n = k)$

$$\mathbf{E}^k(N_k) = \sum_{n=0}^{\infty} (\mathbf{P}^n)_{kk}$$

So state  $k$  is transient if and only if

$$\sum_{n=0}^{\infty} (\mathbf{P}^n)_{kk} = 1/(1 - f_k) < \infty.$$



# Class properties

## Theorem

*Recurrence (or transience) is a class property. That is, if  $i$  and  $j$  are in the same communicating class then  $i$  is recurrent (respectively transient) if and only if  $j$  is recurrent (respectively transient).*

## Proof:

Suppose  $i$  is recurrent and  $i \leftrightarrow j$ . There are integers  $m$  and  $n$  such that

$$(\mathbf{P}^m)_{ji} > 0 \quad \text{and} \quad (\mathbf{P}^n)_{ij} > 0$$



## Recurrence is a class property

Then

$$\begin{aligned}\sum_k (\mathbf{P}^k)_{jj} &\geq \sum_{k \geq 0} (\mathbf{P}^{m+k+n})_{jj} \geq \sum_{k \geq 0} (\mathbf{P}^m)_{ji} (\mathbf{P}^k)_{ii} (\mathbf{P}^n)_{ij} \\ &= (\mathbf{P}^m)_{ji} \left\{ \sum_{k \geq 0} (\mathbf{P}^k)_{ii} \right\} (\mathbf{P}^n)_{ij}\end{aligned}$$

The middle term is infinite and the two outside terms positive so

$$\sum_k (\mathbf{P}^k)_{jj} = \infty$$

which shows  $j$  is recurrent.



# Existence of Recurrent States

## Theorem

*A finite state space chain has at least one recurrent state*

## Proof.

If all states were transient we would have for each  $k$   $P(N_k < \infty) = 1$ . This would mean  $P(\forall k N_k < \infty) = 1$ . But for any  $\omega$  there must be at least one  $k$  for which  $N_k = \infty$  (the total of a finite list of finite numbers is finite). □

Infinite state space chain may have all states transient: the chain  $X_n$  satisfying  $X_{n+1} = X_n + 1$  on the integers has all states transient.



# Coin Tossing

More interesting example:

- Toss a coin repeatedly.
- Let  $X_n$  be  $X_0$  plus the number of heads minus the number of tails in the first  $n$  tosses.
- Let  $p$  denote the probability of heads on an individual trial.
- $X_n - X_0$  is a sum of  $n$  iid random variables  $Y_i$  where  $P(Y_i = 1) = p$  and  $P(Y_i = -1) = 1 - p$ .
- SLLN shows  $X_n/n$  converges almost surely to  $2p - 1$ .
- If  $p \neq 1/2$  this is not 0.



## Coin Tossing Example Continued

- In order for  $X_n/n$  to have a positive limit we must have  $X_n \rightarrow \infty$  almost surely.
- So all states are visited only finitely many times.
- That is, all states are transient.
- Similarly for  $p < 1/2$   $X_n \rightarrow -\infty$  almost surely and all states are transient.





## Coin Tossing

Now look at  $p = 1/2$ . The law of large numbers argument no longer shows anything. I will show that all states are recurrent.

**Proof:** We evaluate  $\sum_n (\mathbf{P}^n)_{00}$  and show the sum is infinite. If  $n$  is odd then  $(\mathbf{P}^n)_{00} = 0$  so we evaluate

$$\sum_m (\mathbf{P}^{2m})_{00}$$

Now

$$(\mathbf{P}^{2m})_{00} = \binom{2m}{m} 2^{-2m}$$



# Coin Tossing

According to Stirling's approximation

$$\lim_{m \rightarrow \infty} \frac{m!}{m^{m+1/2} e^{-m} \sqrt{2\pi}} = 1$$

Hence

$$\lim_{m \rightarrow \infty} \sqrt{m} (\mathbf{P}^{2m})_{00} = \frac{1}{\sqrt{\pi}}$$

Since

$$\sum \frac{1}{\sqrt{m}} = \infty$$

we are done.



## Mean return times

- Compute expected times to return.
- For  $x \in S$  let  $T_x$  denote the hitting time for  $x$ .
- Suppose  $x$  recurrent in **irreducible** chain (only one communicating class).
- Derive equations for expected values of different  $T_x$ .
- Each  $T_x$  is a certain function  $f_x$  applied to  $X_1, \dots$
- Setting  $\mu_{ij} = E^i(T_j)$  we find

$$\mu_{ij} = \sum_k E^i(T_j 1(X_1 = k))$$

- Note that if  $X_1 = x$  then  $T_x = 1$  so

$$E^i(T_j 1(X_1 = j)) = \mathbf{P}_{ij}$$



# Mean Return Times

For  $k \neq j$ , if  $X_1 = k$  then

$$T_j = 1 + f_j(X_2, X_3, \dots)$$

and, by conditioning on  $X_1 = k$  we find

$$E^i(T_j 1(X_1 = k)) = \mathbf{P}_{ik} \left\{ 1 + E^k(T_j) \right\}$$

This gives

$$\mu_{ij} = 1 + \sum_{k \neq j} \mathbf{P}_{ik} \mu_{kj} \quad (5)$$



## Technical details

- Technically, I should check that the expectations in (5) are finite.
- All the random variables involved are non-negative, however, and the equation actually makes sense even if some terms are infinite.
- (To prove this you actually study

$$T_{x,n} = \min(T_x, n)$$

deriving an identity for a fixed  $n$ , letting  $n \rightarrow \infty$  and applying the monotone convergence theorem.)



## Mean Return Times

Here is a simple example:

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

The identity (5) becomes

$$\begin{aligned} \mu_{1,1} &= 1 + \frac{\mu_{2,1}}{2} + \frac{\mu_{3,1}}{2} & \mu_{1,2} &= 1 + \frac{\mu_{3,2}}{2} & \mu_{1,3} &= 1 + \frac{\mu_{2,3}}{2} \\ \mu_{2,1} &= 1 + \frac{\mu_{3,1}}{2} & \mu_{2,2} &= 1 + \frac{\mu_{1,2}}{2} + \frac{\mu_{3,2}}{2} & \mu_{2,3} &= 1 + \frac{\mu_{1,3}}{2} \\ \mu_{3,1} &= 1 + \frac{\mu_{2,1}}{2} & \mu_{3,2} &= 1 + \frac{\mu_{1,2}}{2} & \mu_{3,3} &= 1 + \frac{\mu_{1,3}}{2} + \frac{\mu_{2,3}}{2} \end{aligned}$$

Seventh and fourth show  $\mu_{2,1} = \mu_{3,1}$ . Similar calculations give  $\mu_{ii} = 3$  and for  $i \neq j$   $\mu_{i,j} = 2$ .



## Mean Return Times

Coin tossing Markov Chain with  $p = 1/2$  shows situation can be different when  $S$  is infinite. Equations above become:

$$m_{0,0} = 1 + \frac{1}{2}m_{1,0} + \frac{1}{2}m_{-1,0}$$

$$m_{1,0} = 1 + \frac{1}{2}m_{2,0}$$

and many more.

Some observations:

Have to go through 1 to get to 0 from 2 so

$$m_{2,0} = m_{2,1} + m_{1,0}$$

Symmetry (switching H and T):

$$m_{1,0} = m_{-1,0}$$



# More Coin Tossing

Transition probabilities are **homogeneous**:

$$m_{2,1} = m_{1,0}$$

Conclusion:

$$\begin{aligned}m_{0,0} &= 1 + m_{1,0} \\ &= 1 + 1 + \frac{1}{2}m_{2,0} \\ &= 2 + m_{1,0}\end{aligned}$$

Notice that there are **no** finite solutions!





# Coin Tossing Summary

- Every state is recurrent.
- All the expected hitting times  $m_{ij}$  are infinite.
- All entries  $\mathbf{P}_{ij}^n$  converge to 0.
- Jargon: The states in this chain are null recurrent.



## Example

Model: 2 state MC for weather: 'Dry' or 'Wet'.

```
> p:= matrix(2,2,[[3/5,2/5],[1/5,4/5]]);
```

$$p := \begin{bmatrix} 3/5 & 2/5 \\ 1/5 & 4/5 \end{bmatrix}$$

```
> p2:=evalm(p*p):
```

```
> p4:=evalm(p2*p2):
```

```
> p8:=evalm(p4*p4):
```

```
> p16:=evalm(p8*p8):
```

This computes the powers (evalm understands matrix algebra).

Fact:

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$



## Example

```
> evalf(evalm(p2));  
      [.4400000000    .5600000000]  
      [                ]  
      [.2800000000    .7200000000]  
  
> evalf(evalm(p4));  
      [.3504000000    .6496000000]  
      [                ]  
      [.3248000000    .6752000000]  
  
> evalf(evalm(p8));  
      [.3337702400    .6662297600]  
      [                ]  
      [.3331148800    .6668851200]  
  
> evalf(evalm(p16));  
      [.3333336197    .6666663803]  
      [                ]  
      [.3333331902    .6666668098]
```

Where did  $1/3$  and  $2/3$  come from?



## Example

- Suppose we toss a coin  $P(H) = \alpha_D$
- Start the chain with Dry if we get heads and Wet if we get tails.
- Then

$$P(X_0 = x) = \begin{cases} \alpha_D & x = \text{Dry} \\ \alpha_W = 1 - \alpha_D & x = \text{Wet} \end{cases}$$

and

$$\begin{aligned} P(X_1 = x) &= \sum_y P(X_1 = x | X_0 = y) P(X_0 = y) \\ &= \sum_y \alpha_y P_{y,x} \end{aligned}$$



## Example

- Notice last line is a matrix multiplication of row vector  $\alpha$  by matrix  $\mathbf{P}$ .
- A special  $\alpha$ : if we put  $\alpha_D = 1/3$  and  $\alpha_W = 2/3$  then

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

So: if  $P(X_0 = D) = 1/3$  then  $P(X_1 = D) = 1/3$  and analogously for  $W$ .

- This means that  $X_0$  and  $X_1$  have the same distribution.



## Initial Distributions

**Def'n:** A probability vector  $\alpha$  is called the initial distribution for the chain if

$$P(X_0 = i) = \alpha_i$$

**Def'n:** A Markov Chain is **stationary** if

$$P(X_1 = i) = P(X_0 = i)$$

for all  $i$

- Finding stationary initial distributions.
- Consider  $\mathbf{P}$  above.
- The equation

$$\alpha\mathbf{P} = \alpha$$

is really

$$\alpha_D = 3\alpha_D/5 + \alpha_W/5$$

$$\alpha_W = 2\alpha_D/5 + 4\alpha_W/5$$



# Initial Distributions

- The first can be rearranged to

$$\alpha_W = 2\alpha_D.$$

- So can the second.
- If  $\alpha$  is probability vector then

$$\alpha_W + \alpha_D = 1$$

so we get

$$1 - \alpha_D = 2\alpha_D$$

leading to

$$\alpha_D = 1/3$$



## Initial Distributions: More Examples

$$\mathbf{P} = \begin{bmatrix} 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 2/3 & 0 & 1/3 & 0 \end{bmatrix}$$

Set  $\alpha\mathbf{P} = \alpha$  and get

$$\alpha_1 = \alpha_2/3 + 2\alpha_4/3$$

$$\alpha_2 = \alpha_1/3 + 2\alpha_3/3$$

$$\alpha_3 = 2\alpha_2/3 + \alpha_4/3$$

$$\alpha_4 = 2\alpha_1/3 + \alpha_3/3$$

$$1 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$$

First plus third gives

$$\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4$$

so both sums  $1/2$ . Continue algebra to get unique solution

$$(1/4, 1/4, 1/4, 1/4).$$





# Initial Distributions

```
p:=matrix([[0,1/3,0,2/3],[1/3,0,2/3,0],  
          [0,2/3,0,1/3],[2/3,0,1/3,0]]);
```

```
      [ 0      1/3      0      2/3 ]  
      [                               ]  
      [1/3     0      2/3     0 ]  
p := [                               ]  
      [ 0      2/3     0      1/3 ]  
      [                               ]  
      [2/3     0      1/3     0 ]
```

```
> p2:=evalm(p*p);  
      [5/9     0      4/9     0 ]  
      [                               ]  
      [ 0      5/9     0      4/9 ]  
p2:= [                               ]  
      [4/9     0      5/9     0 ]  
      [                               ]  
      [ 0      4/9     0      5/9 ]
```



# Initial Distributions

```
> p4:=evalm(p2*p2):  
> p8:=evalm(p4*p4):  
> p16:=evalm(p8*p8):  
> p17:=evalm(p8*p8*p):
```



# Initial Distributions

```
> evalf(evalm(p16));  
[.5000000116 , 0 , .4999999884 , 0]  
[  
[0 , .5000000116 , 0 , .4999999884]  
[  
[.4999999884 , 0 , .5000000116 , 0]  
[  
[0 , .4999999884 , 0 , .5000000116]  
  
> evalf(evalm(p17));  
[0 , .4999999961 , 0 , .5000000039]  
[  
[.4999999961 , 0 , .5000000039 , 0]  
[  
[0 , .5000000039 , 0 , .4999999961]  
[  
[.5000000039 , 0 , .4999999961 , 0]
```



## Initial Distributions

```
> evalf(evalm((p16+p17)/2));  
[.2500, .2500, .2500, .2500]  
[  
[.2500, .2500, .2500, .2500]  
[  
[.2500, .2500, .2500, .2500]  
[  
[.2500, .2500, .2500, .2500]
```

$\mathbf{P}^n$  doesn't converge but  $(\mathbf{P}^n + \mathbf{P}^{n+1})/2$  does. Next example:

$$\mathbf{P} = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$



## Initial Distributions

Solve  $\alpha\mathbf{P} = \alpha$ :

$$\alpha_1 = \frac{2}{5}\alpha_1 + \frac{1}{5}\alpha_2$$

$$\alpha_2 = \frac{3}{5}\alpha_1 + \frac{4}{5}\alpha_2$$

$$\alpha_3 = \frac{2}{5}\alpha_3 + \frac{1}{5}\alpha_4$$

$$\alpha_4 = \frac{3}{5}\alpha_3 + \frac{4}{5}\alpha_4$$

$$1 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$$

Second and fourth equations redundant. Get

$$\alpha_2 = 3\alpha_1 \quad 3\alpha_3 = \alpha_4 \quad 1 = 4\alpha_1 + 4\alpha_3$$

Pick any  $\alpha_1$  in  $[0, 1/4]$ ; put  $\alpha_3 = 1/4 - \alpha_1$ .

$$\alpha = (\alpha_1, 3\alpha_1, 1/4 - \alpha_1, 3(1/4 - \alpha_1))$$

solves  $\alpha\mathbf{P} = \alpha$ . So solution is not unique.



# Initial Distributions

```
> p:=matrix([[2/5,3/5,0,0],[1/5,4/5,0,0],
             [0,0,2/5,3/5],[0,0,1/5,4/5]]);
             [2/5   3/5   0   0 ]
             [
             [1/5   4/5   0   0 ]
p := [
       [ 0   0   2/5   3/5]
       [
       [ 0   0   1/5   4/5]

> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
```



# Initial Distributions

```
> evalf(evalm(p8*p8));  
  [.2500000000 , .7500000000 , 0 , 0]  
  [  
  [.2500000000 , .7500000000 , 0 , 0]  
  [  
  [0 , 0 , .2500000000 , .7500000000]  
  [  
  [0 , 0 , .2500000000 , .7500000000]
```



## Initial Distributions

Notice that rows converge but to two different vectors:

$$\alpha^{(1)} = (1/4, 3/4, 0, 0)$$

and

$$\alpha^{(2)} = (0, 0, 1/4, 3/4)$$

Solutions of  $\alpha\mathbf{P} = \alpha$  revisited? Check that

$$\alpha^{(1)}\mathbf{P} = \alpha^{(1)}$$

and

$$\alpha^{(2)}\mathbf{P} = \alpha^{(2)}$$

If  $\alpha = \lambda\alpha^{(1)} + (1 - \lambda)\alpha^{(2)}$  ( $0 \leq \lambda \leq 1$ ) then

$$\alpha\mathbf{P} = \alpha$$

so again solution is not unique.





## Initial Distributions: Last example

```
> p:=matrix([[2/5,3/5,0],[1/5,4/5,0],  
             [1/2,0,1/2]]);
```

```
           [2/5   3/5   0 ]  
           [      ]  
p := [1/5   4/5   0 ]  
      [      ]  
      [1/2   0   1/2]
```

```
> p2:=evalm(p*p):
```

```
> p4:=evalm(p2*p2):
```

```
> p8:=evalm(p4*p4):
```

```
> evalf(evalm(p8*p8));
```

```
[.2500000000 .7500000000      0      ]  
[      ]  
[.2500000000 .7500000000      0      ]  
[      ]  
[.2500152588 .7499694824 .00001525878906]
```



## Interpretation of examples

- For some  $\mathbf{P}$  all rows converge to some  $\alpha$ . In this case this  $\alpha$  is a stationary initial distribution.
- For some  $\mathbf{P}$  the locations of zeros flip flop.  $\mathbf{P}^n$  does not converge.  
Observation: average

$$\frac{\mathbf{P} + \mathbf{P}^2 + \dots + \mathbf{P}^n}{n}$$

*does converge.*

- For some  $\mathbf{P}$  some rows converge to one  $\alpha$  and some to another. In this case the solution of  $\alpha\mathbf{P} = \alpha$  is not unique.

Basic distinguishing features: pattern of 0s in matrix  $\mathbf{P}$ .



# The ergodic theorem

- Consider a finite state space chain.
- If  $x$  is a vector then the  $i$ th entry in  $\mathbf{P}x$  is

$$\sum_j \mathbf{P}_{ij} x_j$$

- Rows of  $\mathbf{P}$  probability vectors, so a weighted average of the entries in  $x$ .
- If weights strictly between 0, 1 and largest and smallest entries in  $x$  not same then  $\sum_j \mathbf{P}_{ij} x_j$  strictly between largest and smallest entries in  $x$ .



# Ergodic Theorem

In fact

$$\begin{aligned}\sum_j \mathbf{P}_{ij} x_j - \min(x_k) &= \sum_j \mathbf{P}_{ij} \{x_j - \min(x_k)\} \\ &\geq \min_j \{p_{ij}\} (\max\{x_k\} - \min\{x_k\})\end{aligned}$$

and

$$\max\{x_j\} - \sum_j \mathbf{P}_{ij} x_j \geq \min_j \{p_{ij}\} (\max\{x_k\} - \min\{x_k\})$$



# Ergodic Theorem

- Now multiply  $\mathbf{P}^r$  by  $\mathbf{P}^m$ .
- $ij$ th entry in  $\mathbf{P}^{r+m}$  is a weighted average of the  $j$ th column of  $\mathbf{P}^m$ .
- $i$ th entry in the  $j$ th column of  $\mathbf{P}^{r+m}$  must be strictly between the minimum and maximum entries of the  $j$ th column of  $\mathbf{P}^m$ .
- In fact, fix a  $j$ .
- $\bar{x}_m =$  maximum entry in column  $j$  of  $\mathbf{P}^m$
- $\underline{x}_m$  the minimum entry.
- Suppose all entries of  $\mathbf{P}^r$  are positive.



# Ergodic Theorem

Let  $\delta > 0$  be the smallest entry in  $\mathbf{P}^r$ . Our argument above shows that

$$\bar{x}_{m+r} \leq \bar{x}_m - \delta(\bar{x}_m - \underline{x}_m)$$

and

$$\underline{x}_{m+r} \geq \underline{x}_m + \delta(\bar{x}_m - \underline{x}_m)$$

Putting these together gives

$$(\bar{x}_{m+r} - \underline{x}_{m+r}) \leq (1 - 2\delta)(\bar{x}_m - \underline{x}_m)$$

In summary the column maximum decreases, the column minimum increases and the gap between the two decreases exponentially along the sequence  $m, m + r, m + 2r, \dots$



# Ergodic Theorem

This idea can be used to prove

## Theorem

Suppose  $X_n$  finite state space Markov Chain with stationary transition matrix  $\mathbf{P}$ . Assume that there is a power  $r$  such that all entries in  $\mathbf{P}^r$  are positive. Then  $\mathbf{P}^k$  has all entries positive for all  $k \geq r$  and  $\mathbf{P}^n$  converges, as  $n \rightarrow \infty$  to a matrix  $\mathbf{P}^\infty$ . Moreover,

$$(\mathbf{P}^\infty)_{ij} = \pi_j$$

where  $\pi$  is the unique row vector satisfying

$$\pi = \pi \mathbf{P}$$

whose entries sum to 1.



# Proof of Ergodic Theorem

- First for  $k > r$

$$(\mathbf{P}^k)_{ij} = \sum_{\ell} (\mathbf{P}^{k-r})_{i\ell} (\mathbf{P}^r)_{\ell j}$$

- For each  $i$  there is an  $\ell$  for which  $(\mathbf{P}^{k-r})_{i\ell} > 0$ .
- Since  $(\mathbf{P}^r)_{\ell j} > 0$  we see  $(\mathbf{P}^k)_{ij} > 0$ .
- The argument before the proposition shows that

$$\lim_{j \rightarrow \infty} \mathbf{P}^{m+jk}$$

exists for each  $m$  and  $k \geq r$ .





# Proof of Ergodic Theorem

- This proves  $\mathbf{P}^n$  has a limit which we call  $\mathbf{P}^\infty$ .
- Since  $\mathbf{P}^{n-1}$  also converges to  $\mathbf{P}^\infty$  we find

$$\mathbf{P}^\infty = \mathbf{P}^\infty \mathbf{P}$$

- Hence each row of  $\mathbf{P}^\infty$  is a solution of  $x\mathbf{P} = x$ .
- The argument before the statement of the proposition shows all rows of  $\mathbf{P}^\infty$  are equal.
- Let  $\pi$  be this common row.



# Proof of Ergodic Theorem

- Now if  $\alpha$  is any vector whose entries sum to 1 then  $\alpha \mathbf{P}^n$  converges to

$$\alpha \mathbf{P}^\infty = \pi$$

- If  $\alpha$  is any solution of  $x = x\mathbf{P}$  we have by induction  $\alpha \mathbf{P}^n = \alpha$  so  $\alpha \mathbf{P}^\infty = \alpha$  so  $\alpha = \pi$ .
- That is exactly one vector whose entries sum to 1 satisfies  $x = x\mathbf{P}$ . •

Note conditions:

- There is an  $r$  for which all entries in  $\mathbf{P}^r$  are positive.
- The chain has a finite state space.



## Finite state space case: $\mathbf{P}^n$ need not have limit

Example:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

- Note  $\mathbf{P}^{2n}$  is the identity while  $\mathbf{P}^{2n+1} = \mathbf{P}$ .
- Note, too, that

$$\frac{\mathbf{P}^0 + \dots + \mathbf{P}^n}{n+1} \rightarrow \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

- Consider the equations  $\pi = \pi\mathbf{P}$  with  $\pi_1 + \pi_2 = 1$ .
- We get

$$\pi_1 = \frac{1}{2}\pi_1 + \frac{1}{2}(1 - \pi_1) = \frac{1}{2}$$

so that the solution to  $\pi = \pi\mathbf{P}$  is again unique.



## Periodic Chains

**Def'n:** The period  $d$  of a state  $i$  is the greatest common divisor of

$$\{n : (\mathbf{P}^n)_{ii} > 0\}$$

### Lemma

*If  $i \leftrightarrow j$  then  $i$  and  $j$  have the same period.*

**Def'n:** A state is **aperiodic** if its period is 1.

I do the case  $d = 1$ . Fix  $i$ . Let

$$G = \{k : (\mathbf{P}^k)_{ii} > 0\}$$

If  $k_1, k_2 \in G$  then  $k_1 + k_2 \in G$ .

This (and aperiodic) implies (number theory argument) that there is an  $r$  such that  $k \geq r$  implies  $k \in G$ .

Now find  $m$  and  $n$  so that

$$(\mathbf{P}^m)_{ij} > 0 \text{ and } (\mathbf{P}^n)_{ji} > 0$$



## Periodic Chains

For  $k > r + m + n$  we see  $(\mathbf{P}^k)_{jj} > 0$  so the gcd of the set of  $k$  such that  $(\mathbf{P}^k)_{jj} > 0$  is 1.

The case of period  $d > 1$  can be dealt with by considering  $\mathbf{P}^d$ .

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

For this example  $\{1, 2, 3\}$  is a class of period 3 states and  $\{4, 5\}$  a class of period 2 states.

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

has a single communicating class of period 2.

A chain is **aperiodic** if all its states are aperiodic.



## Hitting Times

Start irreducible recurrent chain  $X_n$  in state  $i$ . Let  $T_j$  be first  $n > 0$  such that  $X_n = j$ . Define

$$m_{ij} = \mathbb{E}(T_j | X_0 = i)$$

First step analysis:

$$\begin{aligned} m_{ij} &= 1 \cdot P(X_1 = j | X_0 = i) \\ &\quad + \sum_{k \neq j} (1 + \mathbb{E}(T_j | X_0 = k)) P_{ik} \\ &= \sum_j P_{ij} + \sum_{k \neq j} P_{ik} m_{kj} \\ &= 1 + \sum_{k \neq j} P_{ik} m_{kj} \end{aligned}$$

Example

$$\mathbf{P} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$



## Stationary Initial Distributions: Equations

$$\begin{aligned}m_{11} &= 1 + \frac{2}{5}m_{21} & m_{12} &= 1 + \frac{3}{5}m_{12} \\m_{21} &= 1 + \frac{4}{5}m_{21} & m_{22} &= 1 + \frac{1}{5}m_{12}\end{aligned}$$

The second and third equations give immediately

$$m_{12} = \frac{5}{2} \text{ and } m_{21} = 5$$

Then plug in to the others to get

$$m_{11} = 3 \text{ and } m_{22} = \frac{3}{2}$$

Notice stationary initial distribution is

$$\left( \frac{1}{m_{11}}, \frac{1}{m_{22}} \right)$$



## Stationary Initial Distributions

Consider fraction of time spent in state  $j$ :

$$\frac{1(X_0 = j) + \cdots + 1(X_n = j)}{n + 1}$$

Imagine chain starts in chain  $i$ ; take expected value.

$$\frac{\sum_{r=1}^n \mathbf{P}_{ij}^r + 1(i = j)}{n + 1}$$

If rows of  $\mathbf{P}^r$  converge to  $\pi$  then fraction converges to  $\pi_j$ ; i.e. limiting fraction of time in state  $j$  is  $\pi_j$ .

Heuristic: start chain in  $i$ . Expect to return to  $i$  every  $m_{ii}$  time units. So are in state  $i$  about once every  $m_{ii}$  time units; i.e. limiting fraction of time in state  $i$  is  $1/m_{ii}$ .

Conclusion: for an irreducible recurrent finite state space Markov chain

$$\pi_i = \frac{1}{m_{ii}}.$$





## Stationary Initial Distributions

Real proof: Renewal theorem or variant.

Idea:  $S_1 < S_2 < \dots$  are times of visits to  $i$ . Segment  $i$ :

$$X_{S_{i-1}+1}, \dots, X_{S_i}.$$

Segments are iid by Strong Markov.

Number of visits to  $i$  by time  $S_k$  is exactly  $k$ .

Total elapsed time is  $S_k = T_1 + \dots + T_k$  where  $T_i$  are iid.

Fraction of time in state  $i$  by time  $S_k$  is

$$\frac{k}{S_k} \rightarrow \frac{1}{m_{ii}}$$

by SLLN. So if fraction converges to  $\pi_i$  must have

$$\pi_i = \frac{1}{m_{ii}}.$$



# Summary of Theoretical Results

For an irreducible aperiodic positive recurrent Markov Chain:

- 1  $\mathbf{P}^n$  converges to a stochastic matrix  $\mathbf{P}^\infty$ .
- 2 Each row of  $\mathbf{P}^\infty$  is  $\pi$  the unique stationary initial distribution.
- 3 The stationary initial distribution is given by

$$\pi_i = 1/m_i$$

where  $m_i$  is the mean return time to state  $i$  from state  $i$ .

If the state space is finite an irreducible chain is positive recurrent.



## Ergodic Theorem

Notice slight of hand: I showed

$$\frac{\mathbb{E} \left\{ \sum_{i=0}^n 1(X_i = k) \right\}}{n} \rightarrow \pi_k$$

but claimed

$$\frac{\sum_{i=0}^n 1(X_i = k)}{n} \rightarrow \pi_k$$

almost surely which is also true. This is a step in the proof of the ergodic theorem. For an irreducible positive recurrent Markov chain and any  $f$  on  $S$  such that  $\mathbb{E}^\pi(f(X_0)) < \infty$ :

$$\frac{\sum_0^n f(X_i)}{n} \rightarrow \sum \pi_j f(j)$$

almost surely. The limit works in other senses, too. You also get

$$\frac{\sum_0^n f(X_i, \dots, X_{i+k})}{n} \rightarrow \mathbb{E}^\pi \{f(X_0, \dots, X_k)\}$$

E.g. fraction of transitions from  $i$  to  $j$  goes to

$$\pi_i \mathbf{P}_{ij}$$



# Positive Recurrent Chains

For an irreducible positive recurrent chain of period  $d$ :

- 1  $\mathbf{P}^d$  has  $d$  communicating classes each of which forms an irreducible aperiodic positive recurrent chain.
- 2  $(\mathbf{P}^{n+1} + \dots + \mathbf{P}^{n+d})/d$  has a limit  $\mathbf{P}^\infty$ .
- 3 Each row of  $\mathbf{P}^\infty$  is  $\pi$  the unique stationary initial distribution.
- 4 Stationary initial distribution places probability  $1/d$  on each of the communicating classes in 1.



# Null Recurrent and Transient Chains

For an irreducible null recurrent chain:

- 1  $\mathbf{P}^n$  converges to 0 (pointwise).
- 2 there is no stationary initial distribution.

For an irreducible transient chain:

- 1  $\mathbf{P}^n$  converges to 0 (pointwise).
- 2 there is no stationary initial distribution.



# Reducible Chains

For a chain with more than 1 communicating class:

- 1 If  $\mathcal{C}$  is a recurrent class the submatrix  $\mathbf{P}_{\mathcal{C}}$  of  $\mathbf{P}$  made by picking out rows  $i$  and columns  $j$  for which  $i, j \in \mathcal{C}$  is a stochastic matrix. The corresponding entries in  $\mathbf{P}^n$  are just  $(\mathbf{P}_{\mathcal{C}})^n$  so you can apply the conclusions above.
- 2 For any transient or null recurrent class the corresponding columns in  $\mathbf{P}^n$  converge to 0.
- 3 If there are multiple positive recurrent communicating classes then the stationary initial distribution is not unique.

