2 Markov Chains

Discrete Time Index

Consider a discrete time and space stochastic process

$$\{X_n;\ n = 0, 1, 2, \ldots\}, \text{ state space } \mathbf{S} = \{0, 1, \cdots, k\}$$

so that $X_n \in \mathbf{S}$.

Example 2.1 X_n = number of heads − number of tails, after n tosses.

Example 2.2 $X_n = 0$ or 1, 0 = dry day, 1 = wet day. Here i indexes days.

Joint and Conditional Distributions

Complete probabilistic explanation of $\{X_n\}$ requires $p(x_1, x_2, \cdots, x_n)$, which is too complicated. Often $p(x)$ can be constructed sequentially

$$p(x_0, x_1, \cdots, x_n) = p(x_0)\, p(x_1|x_0)\ \cdots\ p(x_{n-1}|x_0, \cdots, x_{n-2})\, p(x_n|x_0, \cdots, x_{n-1}).$$

Example 2.1 # heads has a 1–step memory.
Example 2.2 rain may have a short memory.
note: a valid set of conditional distributions defines a stochastic process.

2.1 Markov Chains

Simplest assumption beyond independence: 1 step memory.

Definition 2.1 $\{X_n\}$ is a Markov chain if

$$\begin{aligned} \Pr(X_{n+1} = x_{n+1}|X_0 = x_0, \cdots, X_n = x_n) &= \Pr(X_{n+1} = x_{n+1}|X_n = x_n) \text{ or} \\ \Pr(X_n = j|X_0, \cdots, X_{n-1} = i) &= \Pr(X_n = j|X_{n-1} = i) = P_{ij}^{n,n+1} \end{aligned}$$

Example 2.1 $\{X_n\}$ is Markov if successive tosses are independent (ie: Pr(H) = p on each occasion.

$$P_{ij}^{(n,n+1)} = \begin{cases} p & \text{if } j = i + 1 \\ q & \text{if } j = i - 1 \\ 0 & \text{otherwise} \end{cases}$$

Example 2.2 $\{X_n\}$ could be Markov within a season. So $P_{ij}^{n,n+1}$ might be independent of n.

Definition 2.2 $P_{ij}^{(n,n+1)}$ are called the 1–step transition probabilities of a Markov chain.
note: $P_{ij}^{(n,n+1)}$ must satisfy:

1. $P_{ij}^{(n,n+1)} \geq 0\ \forall i, j \in \mathbf{S}$ and $n = 0, 1, \cdots$.

2. $\sum_{j\in\mathbf{S}} P_{ij}^{(n,n+1)} = 1 \; \forall i,j \in \mathbf{S},$ and $n = 0,1,\cdots.$

For Markov chains

$$p(x_1,\cdots,x_n|x_0) = P_{x_0x_1}^{(0,1)}\, P_{x_1x_2}^{(1,2)}\, \cdots\, P_{x_{n-1}x_n}^{(n-1,n)}$$

Definition 2.3 $P_i^{(0)} \equiv \Pr(X_0 = i),\; i \in \mathbf{S}$ is called the initial state distribution.

So

$$p(x_0,\cdots,x_n) = P_{x_0}^{(0)}\, P_{x_0x_1}^{(0,1)}\, P_{x_1x_2}^{(1,2)}\, \cdots\, P_{x_{n-1}x_n}^{(n-1,n)}.$$

Some intuitively obvious results

Lemma 2.1 $p(x_3|x_0,x_1) = p(x_3|x_1)$ More generally, $p(x_n|x_0,\cdots,x_r) = p(x_n|x_r)$.
Proof

$$\begin{aligned}
p(x_3|x_0,x_1) &= \sum_{x_2\in\mathbf{S}} p(x_2,x_3|x_0,x_1) \\
&= \sum_{x_2\in\mathbf{S}} p(x_3|x_0,x_1,x_2)\, p(x_2|x_0,x_1) \\
&= \sum_{x_2\in\mathbf{S}} p(x_3|x_1,x_2)\, p(x_2|x_1) \\
&= \sum_{x_2\in\mathbf{S}} p(x_2,x_3|x_1) = p(x_3|x_1)
\end{aligned}$$

Lemma 2.2 $p(x_0,x_2|x_1) = p(x_0|x_1)p(x_2|x_1)$ "The past and future of a Markov chain are conditionally independent of the present."
Proof

$$\begin{aligned}
p(x_0,x_2|x_1) &= \frac{p(x_0,x_1,x_2)}{p(x_1)} \\
&= \frac{p(x_2|x_0,x_1)p(x_0,x_1)}{p(x_1)} \\
&= p(x_2|x_1)p(x_0|x_1)
\end{aligned}$$

Time Homogeneous Markov Chains

Definition 2.4 If $P_{ij}^{(n,n+1)}$ are independent of $n = 0,1,\ldots$, the Markov chain is *time homogeneous* (or has stationary transition probabilities). In this case,

$$P_{ij} = \Pr(X_{n+1} = j|X_n = i).$$

We will assume Markov chains to be time homogeneous from now on.

Matrix Notation

Suppose without loss of generality the states of $\mathbf{S}$ are integers. The square matrix $P = (P_{ij})$ is called the *transition probability matrix* (TPM) of the chain.

1. $P_{\mathbf{S}\times\mathbf{S}}$.
2. $P \geq 0$ elementwise.
3. $P\mathbf{1} = \mathbf{1}$. ie: $\mathbf{1}$ is a right eigenvector of P with eigenvalue $\lambda = 1$.

Any such matrix is called stochastic.

Marginal Probabilities

$p_i^{(n)} = \Pr(X_n = i)$ so that $\{p_i^{(i)},\ i \in \mathbf{S}\}$ is the marginal probability distribution of X_n. Write $\mathbf{p}^{(n)}$ for the column vector with i^{th} element $p_i^{(n)}$.

TPM Examples

1. *coin tossing*

   Successive tosses are independent with Pr(heads on any toss) $= p$. Let

$$\begin{aligned} n &= \text{number of tosses} \\ H_n &= \text{number of heads after } n \text{ tosses} \\ T_n &= \text{number of tails after } n \text{ tosses} \\ X_0 &= 0 \\ X_n &= H_n - T_n \\ \mathbf{S} &= \{0, \pm 1, \pm 2, \ldots\} \end{aligned}$$

   The TPM has the form

$$P = \begin{array}{c} \\ \vdots \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} \cdots & -2 & -1 & 0 & 1 & 2 & \cdots \end{array} \\ \left(\begin{array}{ccccccc} & & & & & & \\ & 0 & p & 0 & 0 & 0 & \\ & q & 0 & p & 0 & 0 & \\ \cdots & 0 & q & 0 & p & 0 & \cdots \\ & 0 & 0 & q & 0 & p & \\ & 0 & 0 & 0 & q & 0 & \\ & & & & & & \end{array} \right) \end{array}$$

   Distribution of X_n?

   $X_n = 2H_n - n$
   $H_n = \frac{1}{2}(X_n + n)$.
   $H_n \sim \text{Bin}(n, p)$

$$\begin{aligned} \Pr(X_n = i) &= \Pr(H_n = \tfrac{1}{2}(n+i)) \\ &= \binom{n}{\frac{1}{2}(n+i)} p^{\frac{1}{2}(n+i)} q^{\frac{1}{2}(n-i)},\ i = -n, -n+2, \cdots, n-2, n. \end{aligned}$$

$$\Pr(X_{2n} = 0) = \binom{2n}{n} p^n q^n, \; n = 1, 2, \ldots$$

Let u denote the waiting time until $H_n = T_n$ for the first time after $n = 0$. Here $u = \inf_{n>0}\{X_n = 0\}$. What is the distribution of u? What is Eu.

2. *Rainfall: Tel–Aviv*
Here a wet day means precipitation over 1mm.

Tel–Aviv rainfall data

	wet today	dry today	total
wet yesterday	687	351	1038
dry yesterday	350	1049	1399
	1037	1400	2437

Clearly, rainfall of consecutive days is not independent. Is a Markov chain a possible model?

$$\widehat{P} = \begin{pmatrix} .662 & .338 \\ .250 & .750 \end{pmatrix}$$

If the Markov assumption is correct, then

$$\Pr(X_n = 1, X_{n+1} = 1, X_{n+2} = 0 | X_{n-1} = 1) = p(110|1) = p(1|1)p(1|0)p(0|1).$$

To asses the model we can

(a) Test the Markov assumption

$$p(x_{n+1}|x_0, \cdots, x_n) = p(x_{n+1}|x_n)$$

(b) Test time homogeneity

$$P_{ij}^{(m)} \equiv P_{ij}$$

(c) Compare observed and model based distributions of wet and dry runs.

(d) Assess reliability of forecasts.

3. *Queuing System*
Customers arrive at a single service facility and wait in line. At the start of each unit time interval, service is available to a single customer and takes at most 1 unit of time.

$$\begin{aligned} X_n &= \text{number of customers in system as } n^{th} \text{ interval begins} \\ Y_n &= \text{number of arrivals at the } n^{th} \text{ unit time interval} \end{aligned}$$

$$X_{n+1} = \begin{cases} X_n + Y_n - 1 & \text{if } X_n \geq 1 \\ Y_n & \text{if } X_n = 0. \end{cases}$$

Suppose Y_n's are iid with $\Pr(Y_n = k) = a_k$. Then $\{X_n\}$ is a Markov chain with TPM

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \\ a_0 & a_1 & a_2 & a_3 & \\ \cdot & a_0 & a_1 & a_2 & \cdots \\ \cdot & \cdot & a_0 & a_1 & \\ \cdot & \cdot & \vdots & a_0 & \end{pmatrix} \end{array}$$

Questions of interest about this system include:

- What is the limiting distribution of X_n?
- What is the distribution of the waiting time?
- What is the distribution of the length of idle periods?

2.2 Chapman–Kolmogorov Equations

Definition 2.5 The n–stem transition probability matrix $P^{(n)}$ is defined by

$$P_{ij}^{(n)} = \Pr(X_{n+m} = j | X_m = i),\ n \geq 0,\ i, j \in \mathbf{S}.\square$$

Forward equations

$$P_{ij}^{(n)} = \sum_{k \in \mathbf{S}} P_{ik}^{(n-1)} P_{kj}$$

Backward equations

$$P_{ij}^{(n)} = \sum_{k \in \mathbf{S}} P_{ik} P_{kj}^{(n-1)}$$

More generally this gives the equations

$$P_{ij}^{(n+m)} = \sum_{k \in \mathbf{S}} P_{ik}^{(n)} P_{kj}^{(m)}.$$

Proof

$$\begin{aligned}
P_{ij}^{(n+m)} &= \Pr(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in \mathbf{S}} \Pr(X_{n+m} = j, X_n = k | X_0 = i) \\
&= \sum_{k \in \mathbf{S}} \Pr(X_{n+m} = j | X_n = k, X_0 = i) \Pr(X_n = k | X_0 = i) \\
&= \sum_{k \in \mathbf{S}} P_{ik}^{(n)} P_{kj}^{(m)}
\end{aligned}$$

In terms of TPM's, this says

$$P^{(n+m)} = P^{(n)} P^{(m)}$$

This means

$$P^{(2)} = PP = P^2$$

and by induction

$$P^{(n)} = P^n.$$

How do we find P^n. If $\mathbf{S}$ is small, compute P^n on the computer.

```
matpower_function(P,pow){
 # a function that computes P^pow
 if(pow==1) return(P)
```

```
 else return( P %*% matpower(P,pow-1) )
}
     P                P^2              P^5                  P^20
0.2  0.0  0.8   0.76 0.08 0.16   0.367 0.073 0.560   0.487 0.086 0.427
0.0  0.5  0.5   0.45 0.30 0.25   0.410 0.103 0.487   0.485 0.086 0.429
0.9  0.1  0.0   0.18 0.05 0.77   0.630 0.097 0.272   0.481 0.086 0.434
```

In terms of $\mathbf{p}^{(n)}$, the marginal probability vector for X_n, the Chapman–Kolomogorov equations give

$$\begin{aligned}
\text{forward equation} \qquad & {\mathbf{p}^{(n)}}^T = {\mathbf{p}^{(n-1)}}^T P \\
\text{backward equation} \qquad & {\mathbf{p}^{(n)}}^T = {\mathbf{p}^{(1)}}^T P^{n-1} \\
\text{general equation} \qquad & {\mathbf{p}^{(n+m)}}^T = {\mathbf{p}^{(n)}}^T P^m
\end{aligned}$$

Theorem 2.1 Derivation of P^n
Let $\mathbf{S} = \{1, \cdots, S\}$, $P_{\mathbf{S}\times\mathbf{S}}$, with S finite. Assume P has distinct eigenvalues $\lambda_1, \cdots, \lambda_S$ with right eigenvectors $\mathbf{u}_1, \cdots, \mathbf{u}_S$ (ie: $P\mathbf{u}_k = \lambda_k \mathbf{u}_k$). Then

$$P^n = \sum_{k=1}^{S} \lambda_k^n A_k \text{ where } A_1, \cdots, A_S \text{ are fixed } S \times S \text{ matricies .}$$

Example 2.3 binary Markov chain $\mathbf{S} = \{0, 1\}$, $S = 2$

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \; 0 < \alpha, \beta < 1.$$

$\det(P - \lambda I) = 0$ gives $\lambda^2 - (2 - \alpha - \beta)\lambda + 1 - \alpha - \beta = 0$.
Since we know $P\mathbf{1} = \mathbf{1}$, we have $(\lambda - 1)(\lambda - (1 - \alpha - \beta)) = 0$.
Since $1 - \alpha - \beta < 1$, the eigenvalues λ_1 and λ_0 are distinct. The above theorem gives

$$P^n = A + (1 - \alpha - \beta)^n B, \; A \text{ and } B \text{ fixed } \forall n.$$

To determine A and B, we use the values $n = 0, 1$. This gives

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \to \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}$$

since $|1 - \alpha - \beta| < 1$. This implies that the effect of $\mathbf{p}^{(0)}$ does not affect $\lim_{n\to\infty} \mathbf{p}^{(n)}$,

$${\mathbf{p}^{(n)}}^T = {\mathbf{p}^{(0)}}^T P^n \to \pi^T.$$

proof:
Let

$$U_{S\times S} = \left(\begin{array}{c|c|c} \mathbf{u}_1 & \cdots & \mathbf{u}_S \end{array} \right), \text{ and } \Lambda_{S\times S} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_S \end{pmatrix}.$$

Then

$$P\mathbf{u}_k = \lambda_k \mathbf{u}_k \Rightarrow PU = U\Lambda \Rightarrow P = U\Lambda U^{-1},$$

so

$$P^n = (U\Lambda U^{-1})\cdots(U\Lambda U^{-1}) = U\Lambda^n U^{-1}.$$

Now

$$U^{-1} = \begin{pmatrix} \mathbf{v}_1^T \\ \hline \vdots \\ \hline \mathbf{v}_S^T \end{pmatrix}$$

where $\mathbf{v}_1, \cdots, \mathbf{v}_S$ are left eigenvectors (ie: $\mathbf{v}_k^T P = \lambda_k \mathbf{v}_k^T$). Therefore

$$\begin{aligned} P^n &= \left(\begin{array}{c|c|c} \mathbf{u}_1 & \cdots & \mathbf{u}_S \end{array} \right) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_S \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \hline \vdots \\ \hline \mathbf{v}_S^T \end{pmatrix} \\ &= \sum_{k=1}^{S} \lambda_k^n \mathbf{u}_k \mathbf{v}_k^T . \square \end{aligned}$$

To solve for P^n by hand, solve $\det(P - \lambda I) = 0$ for λ_k's. Use the fact that $\lambda_1 = 1$ and the above theorem (**note:** sometimes $n = \infty$ is useful).
Corallary If $|\lambda_k| \leq 1 \; \forall \; k$ then $P^n \to A$.

Note that sometimes looking at the TPM isn't always best if the Markov chain has some structure to take advantage of. Here's such an example.

Example: Successive maxima of iid discrete RVs

Let $z_1, z_2, \ldots$ be a sequence of iid, non-negative integer valued random variables such that

$$\Pr\{z_n = i\} = a_i.$$

Define $x_n = \max\{z_1, \ldots, z_n\}$ with $x_0 = 0$.
For example z_n is the max tide level in year n; or peak electricity demand in week n.
$\{x_n\}$ is a Markov chain.
Define M as the time when $z_1, \ldots$ first attains the value r (ie. $M = \min\{n : x_n = r\}$), with $\mu = E(M)$.
Let $A_k = \Pr\{z_n \leq k\} = a_0 + a_1 + \cdots + a_k$.
$\{x_n\}$ is a Markov chain with TPM:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} a_0 = A_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & A_1 & a_2 & a_3 & \cdots \\ 0 & 0 & A_2 & a_3 & \cdots \\ 0 & 0 & 0 & A_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix} \end{array}$$

Find μ with a 1st step analysis

$$\text{Define } I = \begin{cases} 0 & \text{if 1st record } < r \;\; (\text{if } z_1 < r) \\ 1 & \text{if 1st record } \geq r \;\; (\text{if } z_1 \geq r) \end{cases}$$

$$\begin{aligned} E(M) = E\left[E(M|I)\right] &= \Pr\{I=0\} \cdot E(M|I=0) + \Pr\{I=1\} \cdot E(M|I=1) \\ &= \Pr\{z_1 < r\} \cdot (\mu + 1) + \Pr\{z \geq r\} \cdot 1 \\ &= 1 + \Pr\{z_1 < r\} \cdot \mu \end{aligned}$$

$$\Rightarrow \mu = \frac{1}{\Pr\{z_1 \geq r\}} = \frac{1}{a_r + a_{r+1} + \cdots} = \frac{1}{A_{r-1}}$$

2.3 Long Run Behavior of Markov Chains

P^n often tends to a constant (usually very quickly). In the best case $P_{ij}^{(n)} \to a_j$ as $n \to \infty$ so that

$$P^n \to \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix} \text{ as } n \to \infty. \tag{2.1}$$

In this case π is called the limiting distribution of P: $\mathbf{p}^{(n)} \to \pi$ for any $\mathbf{p}^{(0)}$

Examples:

two state Markov Chain

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \quad P^n \to \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$$

As long as $0 < \alpha, \beta < 1$. $\square$

a deterministic Markov Chain

This chain flip-flops between state 0 and state 1.

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This is an example of a *periodic* chain. Here the eigenvalues of P are ± 1, so Theorem 2.1 applies, but doesn't give a limit to P^n. $\square$

Regular TPM

Definition 2.6 *$P_{S\times S}$ is regular if there exists m such that $P^m > 0$ (ie. $P_{ij} > 0$ for all i and j).*

Lemma 2.3 *$P^m > 0 \Rightarrow P^{m+n} > 0$ for any $n = 0, 1, \ldots$.*
Proof *If $P^m > 0$ then $P^{m+1} = PP^m > 0$.* $\Box$

So if (2.1) holds and $\pi > 0$ then P is regular. What about the converse?

Theorem 2.2 *Suppose $P_{S\times S}$ is regular with $S = \{1, \ldots, s\}$. Then*

1. *$P^n \to A$, as $n \to \infty$.*
2. $$A = \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}, \text{ where } \pi \text{ is a probability vector.}$$
3. *$\pi > 0$*

Proof: To be supplied later

Issues:
Is $P_{S\times S}$ regular? (ie. is there an n such that $P^n > 0$)

– or –

How many m's must we try before concluding P is not regular?

- If $P_{kk} > 0$ for some k, then P is regular if and only if $P^{2s-2} > 0$.

  Why? Suppose P is regular and that $P_{kk} > 0$ for a particular k. Then it must be possible to reach k from any i in $s-1$ steps or less. Also, any j can be reached from k in $s-1$ steps or less. So there exists m and n such that $P_{ik} > 0$ and P_{kj} with $m, n \le s-1$.

  $$\Rightarrow P^{(}_{ij}2s-2) \le P^{(}_{ik}m)P^{(}_{kk}2s-m-n-2)P^{(}_{kj}n) > 0.$$

- (Taylor and Karlin p122) For any P

  $$P \text{ is regular } \Leftrightarrow P^{S^2} > 0.$$

  Why? Basic idea: $P^{(\ell)}_{kk} > 0$ for some k and $\ell \le s$. With a bit of number theory the number of steps $\ell(2s-m-n-2)$ can be bounded by s^2.

How do we find π?

Theorem 2.3 *If* $P^n \to A = \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}$, *then* π *is the unique probability vector such that* $\pi^T P = \pi^T$.

Proof:

$P^n P = P^{n+1}$.
Letting $n \to \infty$ gives

$$\begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix} P = \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}$$

Therefore $\pi^T P = \pi^T$.

To show uniqueness suppose $x^T P = x^T$ with $x^T \mathbf{1} = 1$.
$x^T P^n = x^T \Rightarrow x^T A = x^T$. Now

$$x^T A = \begin{pmatrix} \pi_1 & \cdots & \pi_s \\ \vdots & & \vdots \\ \pi_1 & \cdots & \pi_s \end{pmatrix} = \begin{pmatrix} \pi_1 & \cdots & \pi_s \end{pmatrix}$$

since $x^T \mathbf{1} = 1$.$\square$

note: theorem doesn't require $\pi > 0$, only $\pi \geq 0$.

Corallary: If π exists, then $\mathbf{p}^{(n)} \to \pi$ as $n \to \infty$ for any $\mathbf{p}^{(0)}$.

Proof:

$$\mathbf{p}^{(n)\,T} = \mathbf{p}^{(0)\,T} P^n \to \mathbf{p}^{(0)\,T} A = \pi^T, \text{ for all } \mathbf{p}^{(0)}.$$

Example: Two state Markov chain.

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}, \; 0 < \alpha, \beta < 1$$

The equations $\pi^T P = \pi^T$ and $\pi^T \mathbf{1} = 1$ give $\pi_0 = \frac{\beta}{\alpha+\beta}$ and $\pi_1 = \frac{\alpha}{\alpha+\beta}$.

note: $s < \infty \Rightarrow P^n \not\to 0$ as $n \to \infty$ since for any row i there is some j such that $P_{ij}^{(n)} \geq \frac{1}{s}$.
If $s = \infty$, then it's possible that $P^n \to 0$. For example $\{X_n\} = \#$ heads - $\#$ tails.

2.4 Stationary Distributions

Definition 2.7 *Any probability vector* $\boldsymbol{\rho}$ *such that* $\boldsymbol{\rho}^T P = \boldsymbol{\rho}^T$ *is called a* stationary distribution *for* P*.*

Remarks:

- P need not be regular nor s finite in the definition.
- If $\mathbf{p}^{(0)} = \boldsymbol{\rho}$, then $\mathbf{p}^{(n)} = \boldsymbol{\rho}$ for all $n = 1, 2, \ldots$. Say "P maintains $\boldsymbol{\rho}$." Starting at $\mathbf{p}^{(0)} = \boldsymbol{\rho}$ can be very helpful.
- If $\mathbf{p}^{(0)} = \boldsymbol{\rho}$, then all finite dimensional distributions are time-invariant – for all non-negative subscripts

  $$(X_{n_1}, X_{n_2}, \ldots, X_{n_r}) \stackrel{d}{=} (X_{m+n_1}, X_{m+n_2}, \ldots, X_{m+n_r}).$$

  This condition is called *strict stationarity.*

  note: Quite generally $\{Y_n \ : \ n = 0, 1, \ldots\}$ is called *weakly stationary* if

  $$\begin{aligned} E(Y_0) &= E(Y_n), \quad n = 1, 2, \ldots \\ \mathrm{Var}(Y_0) &= \mathrm{Var}(Y_n), \quad n = 1, 2, \ldots \\ \mathrm{Cov}(Y_0, Y_n) &= \mathrm{Cov}(Y_m, Y_{m+n}), \quad m, n = 1, 2, \ldots \end{aligned}$$

  note: For a Markov chain, $\mathbf{p}^{(0)}$ and P determine its distribution. So knowing $(X_{n_1}, X_{n_2}) \stackrel{d}{=} (X_{m+n_1}, X_{m+n_2})$ for all m, n_1, n_2 is enough to give strict stationarity.

- If P has limit distribution $\boldsymbol{\pi}$ (ie. $P^n \to \begin{pmatrix} \pi^T \\ \vdots \\ \pi^T \end{pmatrix}$), and a stationary distribution $\boldsymbol{\rho}$ exists, then $\boldsymbol{\rho}$ is unique and $\boldsymbol{\rho} = \pi$. If $P^n \to 0$, there is no stationary distribution.
- However, one or more stationary distributions may exist, even if $\lim P^n$ has unequal rows or doesn't exist.

  Example: Successive \$1 bets; stop when you've won \$2 or have lost \$2 ($\Pr\{\text{win}\} = p$).

  $$P = \begin{array}{c} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

  Here $\boldsymbol{\rho}_1^T = (1, 0, 0, 0, 0)$ and $\boldsymbol{\rho}_2 = (0, 0, 0, 0, 1)$ are stationary distributions – so is a mixture $\alpha\boldsymbol{\rho}_1 + (1-\alpha)\boldsymbol{\rho}_2, \ 0 \le \alpha \le 1$.

It can be shown (see Taylor and Karlin) that

$$P^n \to \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ u_1 & 0 & 0 & 0 & 1-u_1 \\ u_2 & 0 & 0 & 0 & 1-u_2 \\ u_3 & 0 & 0 & 0 & 1-u_3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ where } u_j = \frac{(q/p)^{j+2} - (q/p)^4}{1-(q/p)^4}, \; j = -2, \ldots, 2.$$

Here u_j is determined by the difference equation $u_j = pu_{j+1} + qu_{j-1}, \; j = -1, \ldots, 1$ and the boundary conditions $u_{-2} = 1$, $u_2 = 0$. □

Example:

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Here $\boldsymbol{\rho}^T = (\frac{1}{2}, \frac{1}{2})$ is a stationary distribution. So if $X_0 \sim \text{Bin}(1, \frac{1}{2})$, then $X_{2n} \stackrel{d}{=} X_0$ and $X_{2n+1} \stackrel{d}{=} 1 - X_0 \stackrel{d}{=} X_0$□.

2.5 Classification of States

It turns out that we can make rather general statements about Markov chains (even when $s = \infty$). Consider a Markov chain with TPM P and write

$$\begin{aligned} T_j &= \text{time of first entry into state } j \text{ after time } 0 \\ f_{ij}^{(n)} &= \Pr\{T_j = n | X_0 = i\} = \Pr\{X_1 \neq j, \ldots, X_{n-1} \neq j, X_n = j | X_0 = i\} \\ &= \Pr\{\text{chain first enters } j \text{ from } i \text{ on } n^{\text{th}} \text{ step}\} \\ f_{ij} &= \sum_{n=1}^{\infty} f_{ij}^{(n)} = \Pr\{T_j < \infty | X_0 = i\} \\ \mu_{ij} &= E(T_j | X_0 = i) = \sum_{n=1}^{\infty} n f_{ij}^{(n)} = \sum_{n=1}^{\infty} \Pr\{T_j \geq n | X_0 = i\} \end{aligned}$$

Some Definitions:

i state j is called *recurrent* (persistent) if $f_{jj} = 1$.

ii state j is called *transient* if $f_{jj} < 1$.

iii state j is called *positive* if $f_{jj} = 1$ and $\mu_{jj} < \infty$.

iv state j is called *null* if $f_{jj} = 1$ and $\mu_{jj} = \infty$.

v state j is called *periodic* with period d if $f_{jj}^{(n)} = 0$ for any n not divisible by d, with d the greatest such integer.

Example: $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has period $d = 2$.

vi state j is called *aperiodic* if $d = 1$.

vii state j is called *ergodic* if it is positive and aperiodic.

viii state j is called *absorbing* if $P_{jj} = 1$.

Closure and Irreducibility

If j can be reached from i (ie. $\exists n \geq 0 : \ P_{ij}^{(n)} > 0$), we write $i \to j$; if $i \to j$ and $j \to i$ we write $i \leftrightarrow j$. Here i and j communicate.

Lemma 2.4 $\leftrightarrow$ *is an equivalence relation:*

- $j \leftrightarrow i$
- $i \leftrightarrow j \Rightarrow j \leftrightarrow i$.
- $i \leftrightarrow j$ *and* $j \leftrightarrow k \Rightarrow i \leftrightarrow k$.

A set C of states is *closed* if no $j \notin C$ can be reached from any $i \in C$.

If $i \leftrightarrow j$ for all $i, j \in C$, then C is an *irreducible closed* set.

So, $\leftrightarrow$ induces a partitioning of S into closed sets $C_1, C_2, \ldots$ and a possibly empty set of residual states C_0.

Any TPM P can be written in *reduced form*:

$$P = \begin{array}{c} C_1 \\ C_2 \\ \vdots \\ C_0 \end{array} \begin{pmatrix} P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ R_1 & R_2 & \cdots & \Omega \end{pmatrix}$$

- Residual states in C_0 are transient.
- $P_1, P_2, \ldots$ are stochastic matricies; Ω isn't.
- $P_1, P_2, \ldots$ (sub-Markov chains) may be analized separately.
- States in $C_1, C_2, \ldots$ need not be recurrent.

Example:

$$P = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{pmatrix} .1 & .2 & 0 & 0 & .7 & 0 & 0 \\ 0 & .4 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & .2 & 0 & 0 & .8 & 0 \\ .1 & .1 & .1 & .3 & 0 & .4 & 0 \\ .4 & .2 & 0 & 0 & .4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & .1 & 0 & .2 & .5 & 0 & .2 \end{pmatrix}$$

Now figure out the partition induced by $\leftrightarrow$:

state i	$\{j:\ P_{ij}>0\}$	$\{\text{additional } j:\ P_{ij}^{(n)}>0\}$
1	1,2,5	–
2	2,5	1
3	3,6	–
4	1,2,3,4,6	5
5	1,2,5	–
6	3	6
7	2,4,5,7	1,3,6

TPM P in reduced form is

$$P = \begin{array}{cc} & 1 \\ C_1 & 2 \\ & 5 \\ \hline & 3 \\ C_2 & 6 \\ \hline & 4 \\ C_0 & 7 \end{array} \left(\begin{array}{ccc|cc|cc} .1 & .2 & .7 & 0 & 0 & 0 & 0 \\ 0 & .4 & .6 & 0 & 0 & 0 & 0 \\ .4 & .2 & .4 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & .2 & .8 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline .1 & .1 & 0 & .1 & .4 & .3 & 0 \\ 0 & .1 & .5 & 0 & 0 & .2 & .2 \end{array} \right)$$

Here states 4 and 7 are transient. A later theorem states that all states in a closed set are the same: transient; or d-periodic; or aperiodic; or ....□

A Markov chain is *irreducible* if there is only one equivalence class (ie. all states communicate).

An irreducible Markov chain is *ergodic* if all states are positive and aperiodic.

If $j \in C_0$ then $P_{ij}^{(n)} \to 0$ as $n \to \infty$ for all $i \in S$. So for limiting behavior, it is sufficient to study irreducible Markov chains.

Theorem 2.4 *All states of an irreducible Markov chain are of the same type: all positive; all null; or all transient. And all have the same period.*

lemma: Let $R_j = \#$ visits to j after time 0. Then

1. $$\Pr\{R_j \geq r | X_0 = i\} = f_{ij} f_{jj}^{r-1}$$

2. $$E(R_j | X_0 = i) = \frac{f_{ij}}{1 - f_{jj}} = \begin{cases} 0 & \text{if } f_{ij} = 0 \\ \infty & \text{if } f_{ij} > 0 \text{ and } f_{jj} = 1 \end{cases}$$

3.

$$f_{jj} = 1 \Leftrightarrow \sum_{n=1}^{\infty} P_{jj}^{(n)} = \infty$$

Proof of lemma:

1. is just by definition.
2. is pretty simple:

$$\begin{aligned} E(R_j|X_0 = i) &= \sum_{r=1}^{\infty} \Pr\{R_j \geq r | X_0 = i\} \\ &= \sum_{r=1}^{\infty} f_{ij} f_{jj}^{r-1} \\ &= \frac{f_{ij}}{1 - f_{jj}}. \end{aligned}$$

3.

$$\begin{aligned} R_j &= \sum_{n=1}^{\infty} I_j^{(n)} \text{ where } I_j^{(n)} = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{if } X_n \neq j \end{cases} \\ E(R_j|X_0 = j) &= \sum_{n=1}^{\infty} E(I_j^{(n)}|X_0 = j) = \sum_{n=1}^{\infty} \Pr\{I_j^{(n)} = 1 | X_0 = j\} \\ &= \frac{f_{jj}}{1 - f_{jj}} = \infty \Leftrightarrow f_{jj} = 1 \ \square \end{aligned}$$

Proof of theorem (almost):

For any $i, j \in S$, there exists r, t such that $P_{ij}^{(r)} = \alpha > 0$ and $P_{ji}^{(t)} = \beta > 0$. Now a path from $i \xrightarrow{r} j \xrightarrow{t} i$ is possible in $r + t$ steps so that

$$P_{ii}^{(r+t)} \geq P_{ij}^{(r)} P_{ji}^{(t)} \text{ which implies } P_{ii}^{(r+n+t)} \geq P_{ij}^{(r)} P_{jj}^{(n)} P_{ji}^{(t)} = \alpha\beta P_{jj}^{(n)}.$$

Now

$$f_{jj} = 1 \Leftrightarrow \sum_{n=1}^{\infty} P_{jj}^{(n)} = \infty \text{ implies } \sum_{n=1}^{\infty} P_{ii}^{(r+n+t)} = \infty \Leftrightarrow f_{ii} = 1$$

Therefore all states are recurrent ($f_{jj} = 1 \forall j$) – or – all states are transient ($f_{jj} < 1 \forall j$).

Suppose i has period d_i. The above equation gives $P_{ii}^{(r+t)} > 0$ which means d_i must divide $r + t$. So unless d_i divides n, $P_{ii}^{(n+r+t)} = 0$ and $P_{jj}^{(n)} = 0$. Thus the period of state j, d_j, must be a multiple of d_i. Reversing i and j implies that d_i must be a multiple of d_j. Hence $d_i = d_j$.

Now suppose i is null. Then $P_{ii}^{(r+n+t)} \to 0 \Rightarrow P_{jj}^{(n)} \to 0 \Rightarrow j$ is null. So either: all states are null – or – all states are positive.$\square$

corollary:

- For any Markov chain in reduced form, all states in each C_i have the same type and period.
- In a finite state Markov chain there are no null states and at least one state is positive.

Reaching the stationary distribution

Convergence theorem:

Theorem 2.5 *Let P be an irreducible, aperiodic TPM with stationary distribution $\boldsymbol{\pi}$. Then*

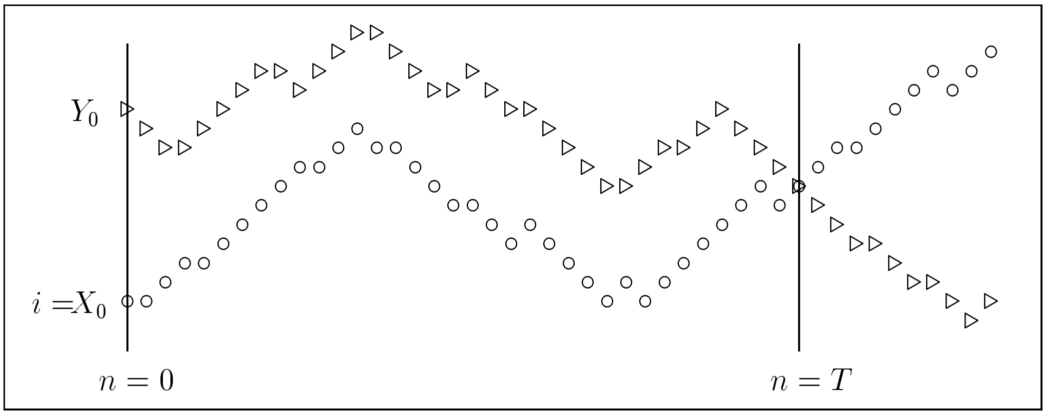
$$P_{ij}^{(n)} \to \pi_j \ \ \forall i, j \in S \ \ \textit{as} \ \ n \to \infty.$$

"Proof:"

Let $\{X_n\}$ and $\{Y_n\}$ be independent Markov chains with TPM P. Then $\{(X_n, Y_n)\}$ is a Markov chain on $S \times S$ with TPM

$$\tilde{P}: \ \ \tilde{P}_{(i_1,i_2;j_1,j_2)} = \Pr\{X_{n+1} = j_1, Y_{n+1} = j_2 | X_n = i_1, Y_n = i_2\} = P_{i_1 j_1} P_{i_2 j_2}.$$

Since P is irreducible, aperiodic with stationary distribution $\boldsymbol{\pi}$, P is ergodic. Therefore $\tilde{P}$ is ergodic with stationary distribution $\tilde{\boldsymbol{\pi}}$, where $\tilde{\pi}_j = \pi_{j_1}\pi_{j_2}$.



Define $T = \min\{n : \ X_n = Y_n\}$. Since $\tilde{P}$ is ergodic, $\Pr\{T > n\} \to 0$ as $n \to \infty$.

By the Markov property: $\Pr\{X_n = j | T \leq n\} = \Pr\{Y_n = j | T \leq n\}$.

So

$$\begin{aligned} \Pr\{X_n = j\} &= \Pr\{X_n = j, T \leq n\} + \Pr\{X_n = j, T > n\} \\ &\leq \Pr\{X_n = j\} + \Pr\{X_n = j, T > n\}. \end{aligned}$$

Similarly,

$$\Pr\{Y_n = j\} \leq \Pr\{X_n = j\} + \Pr\{Y_n = j, T > n\}.$$

These last two equations give:

$$|\Pr\{X_n = j\} - \Pr\{Y_n = j\}| \leq \Pr\{X_n = j, T > n\} + \Pr\{Y_n = j, T > n\}.$$

Summing over all states in S gives

$$\sum_{j \in S} |\Pr\{X_n = j\} - \Pr\{Y_n = j\}| \leq 2\Pr\{T > n\} \to 0 \ \ \text{as} \ \ n \to \infty.$$

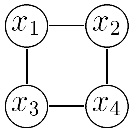
In particular, if $X_0 = i$ and $Y_0 \sim \boldsymbol{\pi}$ then

$$\sum_{j \in S} |P_{ij}^{(n)} - \pi_j| \to 0 \ \ \text{as} \ \ n \to \infty. \ \square$$

Example: 2×2 Ising Model

The Ising model

$$\pi(x) \propto \exp\{\beta \sum_{i \sim j} I[x_i = x_j]\}, \; x \in S = \{0,1\}^4$$



where β is a constant proportional to the inverse of temperature and the set $\{i \sim j\}$ indicates the four adjacent pairs on the above lattice: (1,2); (1,3); (2,4); and (3,4). Consider the following TPM which has this Ising model as its stationary distribution:

- select one of the components i uniformly at random and replace x_i by a draw from its conditional distribution given the remaining components x_{-i}.

The conditional distribution for x_i can be written

$$\pi(x_i | x_{-i}) \propto \exp\{\beta u_i(x)\}$$

where $u_i(x)$ $(= 0, 1,$ or $2)$ is the number of neighboring values that match x_i.

The resulting TPM for this updating rule is given below. The rows are not normalized, so to get the actual probabilities, each row must be rescaled so that the elements add to 1.

	$4e^{2\beta}$	e^0	e^0	e^0	e^0											
	$e^{2\beta}$	θ				e^β	e^β	e^0								
	$e^{2\beta}$		θ			e^β			e^0	e^β						
	$e^{2\beta}$			θ			e^β		e^0		e^β					
	$e^{2\beta}$				θ			e^0		e^β	e^β					
		e^β	e^β			$4e^\beta$						e^β	e^β			
		e^β		e^β			$4e^\beta$					e^β		e^β		
		$e^{2\beta}$			$e^{2\beta}$			$4e^0$					$e^{2\beta}$	$e^{2\beta}$		
			$e^{2\beta}$	$e^{2\beta}$					$4e^0$			$e^{2\beta}$			$e^{2\beta}$	
			e^β		e^β					$4e^\beta$			e^β		e^β	
				e^β	e^β						$4e^\beta$			e^β	e^β	
						e^β	e^β		e^0			θ				$e^{2\beta}$
						e^β		e^0		e^β			θ			$e^{2\beta}$
							e^β	e^0			e^β			θ		$e^{2\beta}$
									e^0	e^β	e^β				θ	$e^{2\beta}$
												e^0	e^0	e^0	e^0	$4e^{2\beta}$

where $\theta = e^0 + 2e^\beta + e^{2\beta}$.

comments:

For implementing this scheme, there is no need for determining the TPM – it is given implicitly in the updating rule.

It's easy to check that any state can be reached from any other state (in at most 4 time steps). Hence the TPM is irreducible – in fact, the TPM is regular – $P_{ij}^{(4)} > 0$.

Regularity gives aperodicity. Alternatively one could argue that irreducibility with $P_{ii} > 0$ for all $i \in S$, make the TPM aperiodic (why?).

How can we be sure this TPM maintains $\pi(x)$? You could check directly: $\boldsymbol{\pi}^T P \stackrel{?}{=} \boldsymbol{\pi}$ since $|S|$ is only 16. But generally, this can be quite difficult.

A second approach is to check for "detailed balance." This is a more stringent condition than $\boldsymbol{\pi}^T P = \boldsymbol{\pi}$, but much easier to check. We'll talk about detailed balance, also called time reversibility, in the next section.

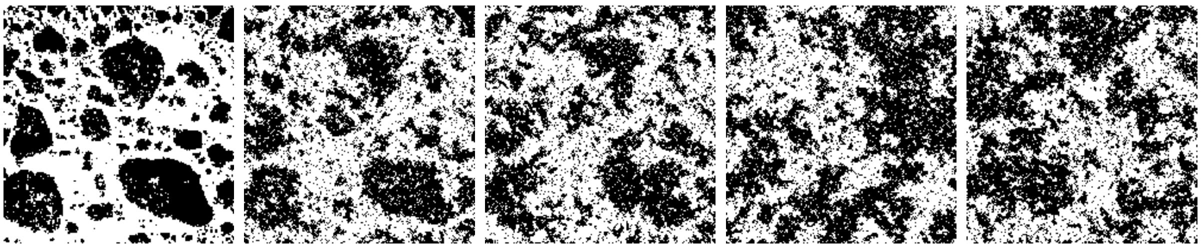
One could argue that this particular rule/TPM maintains $\pi(x)$ as follows. Suppose $x \sim \pi(x)$. Then if I replace x_1 by a draw x_1^* from its conditional distribution $\pi(x_1|x_2, x_3, x_4)$ then the new state $x^* = (x_1^*, x_2, x_3, x_4)$ must also be a draw from the joint distribution $\pi(x)$.

a draw from $\pi(x)$**:**

So one could obtain a draw from $\pi(x)$ by starting with an arbitrary initial configuration x^0, then repeatedly update according to the above rule giving $x^1, x^2, \ldots$. Eventually, the chain will lose the "memory" of its starting value so that x^n can be taken to be a draw from $\pi(x)$.

How many iterations – or how big should n be? For this 2×2 example, around 40 will be fine. For large lattices at the critical temperature, it can take 1000s of updates before the chain loses the influence of the initial state. There are some techniques for ensuring that effects from the initial state have vanished (see Propp and Wilson (1996) and Green and Murdoch (1999)), but in general, this is still somewhat of an art.

Example: 194×200 Ising model



The figures above show the progression of a Markov chain $x^0, x^5, x^{10}, x^{250}, x^{1000}$ where each time step consists of Metropolis updates of each of the 194×200 sites as described in the first chapter. Can we take x^{1000} to be a draw from $\pi(x)$?

Behavior of averages

Theorem 2.6 *Let $\{X_n\}$ be an ergodic Markov chain with stationary distribution $\boldsymbol{\pi}$. Then*

$$\frac{1}{n}\sum_{m=1}^{n} I[X_m = j] \overset{a.s.}{\to} \pi_j = \frac{1}{\mu_{jj}} \quad as \quad n \to \infty$$

for any initial distribution $\mathbf{p}^{(0)}$.

notes:

- This is a generalization of the strong law of large numbers (SLLN) which applies to independent random variables.
- We can construct the empirical distribution $\widehat{\boldsymbol{\pi}}_n^T = (\widehat{\pi}_1, \ldots, \widehat{\pi}_s)$ from a partial realization of the Markov chain $X_1, \ldots, X_n$ by

  $$\widehat{\pi}_j = \frac{1}{n}\sum_{m=1}^{n} I[X_m = j].$$

  It can be shown that $\sup_{j \in S} |\widehat{\pi}_j - \pi_j| \to 0$ as $n \to \infty$, even for continuous and infinite state spaces (see Tierney (1995)).

  This says the empirical distribution of an ergodic Markov chain converges to the stationary distribution $\widehat{\boldsymbol{\pi}}_n \to \boldsymbol{\pi}$. Therefore functionals will converge: $T(\widehat{\boldsymbol{\pi}}_n) \to T(\boldsymbol{\pi})$. So, for a realization of a Markov chain $x_1, x_2, \ldots$, the average, median, and α-percentile all converge to their respective values under the stationary distribution $\boldsymbol{\pi}$.

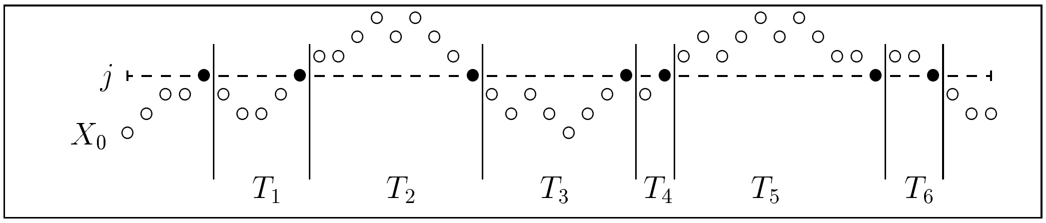
  Example:

  Below are three Markov chains with $S = \{0, 1\}$, each with the same stationary distribution. Which converges faster to $\frac{1}{2}$: $\bar{X}_n$? $\bar{Y}_n$? $\bar{Z}_n$? $\bar{U}_n$ or $\bar{V}_n$?

  - $X_1, X_2, \ldots \quad \mathbf{p}^{(0)^T} = (\frac{1}{2}, \frac{1}{2}) \quad P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$
  - $Y_1, Y_2, \ldots \quad \mathbf{p}^{(0)^T} = (\frac{1}{2}, \frac{1}{2}) \quad P = \begin{pmatrix} .8 & .2 \\ .2 & .8 \end{pmatrix}$
  - $Z_1, Z_2, \ldots \quad \mathbf{p}^{(0)^T} = (\frac{1}{2}, \frac{1}{2}) \quad P = \begin{pmatrix} .2 & .8 \\ .8 & .2 \end{pmatrix}$
  - $U_1, U_2, \ldots \quad \mathbf{p}^{(0)^T} = (\frac{1}{2}, \frac{1}{2}) \quad P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
  - $V_1, V_2, \ldots \quad \mathbf{p}^{(0)^T} = (\frac{1}{2}, \frac{1}{2}) \quad P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Proof:

Let $\{X_n\}$ be a Markov chain with $X_0 \sim \boldsymbol{\pi}$. Let T_0 be the waiting time until the chain's first arrival to state j and let $T_1, T_2, \ldots$ be the waiting times between returns to state j as shown below. Note $T_1, T_2, \ldots, T_{N_n}, \ldots$ are i.i.d. random variables with mean $\mu_{jj} < \infty$.



Now

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I[X_m = j] &= \lim_{n\to\infty} \frac{N_n}{T_1 + \cdots + T_{N_n}} \\
&= \lim_{n\to\infty} \frac{1}{\frac{1}{N_n}(T_1 + \cdots + T_{N_n})} \\
&= \frac{1}{\mu_{jj}}
\end{aligned}
$$

The last equality holds since $T_1, T_2, \ldots$ are i.i.d. with mean $\mu_{jj} < \infty$ and $N_n \to \infty$ since all states are persistent.
Now

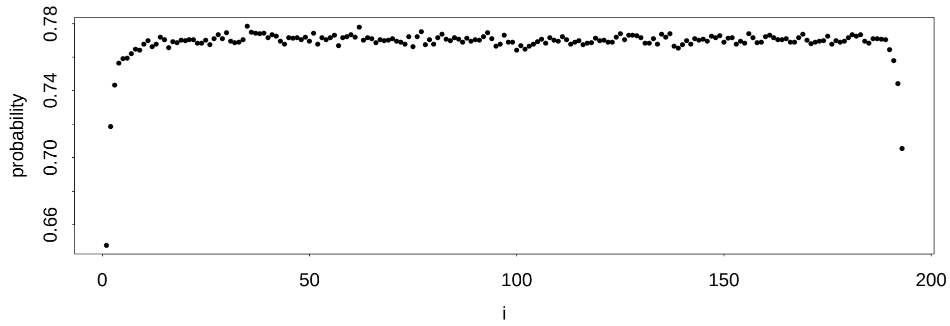
$$
\begin{aligned}
\frac{1}{\mu_{jj}} &= E\left(\frac{1}{\mu_{jj}}\right) = E\left(\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{\infty} I[X_m = j]\right) \\
&= \lim_{n\to\infty} E\left(\frac{1}{n} \sum_{m=1}^{\infty} I[X_m = j]\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{\infty} \pi_j = \pi_j \ \Box
\end{aligned}
$$

Using Markov chains to explore $\boldsymbol{\pi}$

One can use an ergodic Markov chain to learn about the stationary distribution $\boldsymbol{\pi}$. One approach is to construct many runs – using a different random number seed for each run – and collecting the final x_n at the end of each run. An alternative is to make a single "long" run, collecting all the realizations once the chain has reached the stationary distribution.

Example: nearest neighbor correlations in the Ising model.

For the 194×200 ising model, we can compute $\Pr\{X_{1,1} = X_{1,2}\}$ as well as $\Pr\{X_{100,100} = X_{100,101}\}$. For the infinite lattice Ising model, the process is stationary, so the two quantities are equal. However boundaries lead to slightly different dependence structure near the edge as compared to the center of the lattice. The figure below shows a Monte carlo estimate of $\Pr\{X_{i,i} = X_{i,i+1}\}$ for $i = 1, \ldots, 194$.



The probabilities were estimated using 40,000 sweeps of the metropolis algorithm, after discarding the first 1,000.

Theorem 2.7 Perron-Frobenius *If P is the $S \times S$ TPM of finite, irreducible Markov chain with period d, then*

1. $\lambda_1 = 1$ *is an eigenvalue of P.*
2. *The d complex roots of 1, $\lambda_1 = \omega^0, \lambda_2 = \omega^1, \ldots, \lambda_d = \omega^{d-1}$ ($\omega = e^{2\pi i/d}$), are eigenvalues of P.*
3. *The remaining eigenvalues of P satisfy $|\lambda_{d+1}|, \ldots, |\lambda_S| < 1$.*
4. $\mu_{ii} = \frac{1}{\pi_i}$.

Time Reversibility in Markov Chains

Definition 2.8 *A discrete time stochastic process $\{X_n\}$, $n \in Z$ is called* time reversible *if*

$$(X_{n_1}, X_{n_2}, \ldots, X_{n_r}) \stackrel{d}{=} (X_{m-n_1}, X_{m-n_2}, \ldots, X_{m-n_r}),\ \forall n_1, \ldots, n_r, m \in Z,\ r = 1, 2, \ldots$$

example: Any random walk is time reversible.

Lemma 2.5 *If $\{X_n\}$ is time reversible, then $\{X_n\}$ is strictly stationary.*

Proof:

$$\begin{aligned}(X_{n_1}, X_{n_2}, \ldots, X_{n_r}) &\stackrel{d}{=} (X_{-n_1}, X_{-n_2}, \ldots, X_{-n_r}) \\ &\stackrel{d}{=} (X_{m+n_1}, X_{m+n_2}, \ldots, X_{m+n_r})\end{aligned}$$

Both $\stackrel{d}{=}$'s are from the definition of time reversibility. □

Time reversibility extends a process to all of the integers. For strict stationarity we now require all $\mathbf{p}^{(n)} = \boldsymbol{\pi}$, not just $\mathbf{p}^{(0)} = \boldsymbol{\pi}$. **An example?**

An equivalent condition for strict stationarity:

1. $\mathbf{p}^{(n)} = \boldsymbol{\pi}$ where $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ for some n

2. and that P is non-singular.

In this case $\boldsymbol{\pi}^T = \mathbf{p}^{(n)^T} = \mathbf{p}^{(n-1)^T} P$.

From now on we'll assume $\{X_n\}$ is an ergodic Markov chain with limit distribution $\boldsymbol{\pi} > 0$ (uniquely defined by $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ and $\boldsymbol{\pi}^T \mathbf{1} = \mathbf{1}$).

Lemma 2.6 *A reversed Markov chain is also a Markov chain. However the reversed TPM is time invariant only if the chain is strictly stationary.*

Example:

Let $\{\ldots, X_n, X_{n+1}, X_{n+2}, \ldots\}$ be a Markov chain with TPM P. If we think of the "reversed" process $\{\ldots, X_{n+2}, X_{n+1}, X_n, \ldots\}$, we have

$$\begin{aligned}
\Pr\{X_n = i | X_{n+1} = j, X_{n+2} = k, \ldots\} &= \Pr\{X_n = i | X_{n+1} = j\} \\
&= \frac{\Pr\{X_n = i, X_{n+1} = j\}}{\Pr\{X_{n+1} = j\}} \\
&= \frac{\Pr\{X_n = i\} P_{ij}}{\Pr\{X_{n+1} = j\}}
\end{aligned}$$

This reversed process is Markov since the probability doesn't depend on k, but it is not time homogeneous since the transition does depend on n – unless $\mathbf{p}^{(n)} = \boldsymbol{\pi}$. In this case

$$\begin{aligned}
\Pr\{X_n = i | X_{n+1} = j, X_{n+2} = k, \ldots\} &= \Pr\{X_n = i | X_{n+1} = j\} \\
&= \frac{\pi_i P_{ij}}{\pi_j} \\
&= Q_{ji}
\end{aligned}$$

Here Q is the TPM of the reverse chain.□

Theorem 2.8 *The Markov chain $\{X_n\}$ is time reversible if and only if $\{X_n\}$ is strictly stationary and*

$$\pi_i P_{ij} = \pi_j P_{ji} \text{ for all } i, j \in S. \qquad (*)$$

note: This makes sense from the previous example since $\pi_i P_{ij} = \pi_j P_{ji} \Rightarrow P = Q$.

note: $\pi_i P_{ij} = \pi_j P_{ji} \Rightarrow \boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ which follows by summing over i. The condition (*) is often called *detailed balance* since it is more general than the "global balance" conditions: $\sum_{i \in S} \pi_i P_{ij} = \pi_j$, $\forall j \in S$.

Proof:

($\Rightarrow$) From previous lemma, time reversibility $\Rightarrow$ strict stationarity (with stationary distribution $\boldsymbol{\pi}$) and $(X_0, X_1) \stackrel{d}{=} (X_1, X_0)$. Now

$$\begin{aligned}(X_0, X_1) \stackrel{d}{=} (X_1, X_0) &\Rightarrow \Pr\{X_0 = i, X_1 = j\} = \Pr\{X_0 = j, X_1 = i\} \\ &\Rightarrow \Pr\{X_0 = i\}\Pr\{X_1 = j | X_0 = i\} = \Pr\{X_0 = j\}\Pr\{X_1 = i | X_0 = j\} \\ &\Rightarrow \pi_i P_{ij} = \pi_j P_{ji}\end{aligned}$$

($\Leftarrow$) Need to show that $(X_0, X_1, \ldots, X_n) \stackrel{d}{=} (X_n, X_{n-1}, \ldots, X_0)$ since strict stationarity does the rest. Without loss of generality, consider $n = 3$.

$$\begin{aligned}\Pr\{X_3 = i, X_2 = j, X_1 = k, X_0 = \ell\} &= \pi_\ell P_{\ell k} P_{kj} P_{ji} \\ &= \pi_k P_{k\ell} P_{kj} P_{ji} \\ &= P_{k\ell}\pi_k P_{kj} P_{ji} = P_{k\ell}\pi_j P_{jk} P_{ji} \\ &= P_{k\ell} P_{jk} \pi_j P_{ji} = P_{k\ell} P_{jk} \pi_i P_{ij} \\ &= \Pr\{X_0 = i, X_1 = j, X_2 = k, X_3 = \ell\}\end{aligned}$$

Recognizing suitable P for time reversibility

Definition 2.9 *Ergodic TPM P is called time reversible if its limit distribution $\boldsymbol{\pi}$ satisfies $\pi_i P_{ij} = \pi_j P_{ji}$, $\forall i, j \in S$.*

note: For $\{X_n\}$ to be time reversible requires that $\mathbf{p}^{(n)} = \boldsymbol{\pi}$ for $n = 0, \pm 1, \ldots$.

Theorem 2.9 *TPM P is time reversible $\Leftrightarrow$ any path from $i \to i$ has the same probability as the reverse path, for all $i \in S$:*

$$P_{ii_1} P_{i_1 i_2} \cdots P_{i_n i} = P_{ii_n} P_{i_n i_{n-1}} \cdots P_{i_1 i}, \ \forall i, i_1, \ldots, i_n \in S \text{ and } n = 1, 2, \ldots.$$

Proof:

($\Rightarrow$) P is time reversible $\Rightarrow$ $\begin{array}{rcl}\pi_i P_{ii_1} &=& \pi_{i_1} P_{i_1 i} \\ &\vdots& \\ \pi_{i_n} P_{i_n i} &=& \pi_i P_{ii_n}\end{array}$. Taking the product gives the result (note that since $\boldsymbol{\pi} > 0$, division is ok).

($\Leftarrow$) For all $i, i_1, \ldots, i_n \in S$ and $n = 1, 2, \ldots$ we have

$$P_{ii_1} P_{i_1 i_2} \cdots P_{i_n i} = P_{ii_n} P_{i_n i_{n-1}} \cdots P_{i_1 i}$$

summing over all $i_1, \ldots, i_{n-1}$ gives

$$P_{ii_n}^{(n)} P_{i_n i} = P_{ii_n} P_{i_n i}^{(n)}$$

taking the limit as $n \to \infty$ gives

$$\pi_{i_n} P_{i_n i} = P_{ii_n} \pi_i$$

giving time reversibility.

Example:

A finite state ergodic Markov chain with TPM P for which $P_{ij} = 0$ for any i, j such that $|i - j| > 1$ is time reversible with $\boldsymbol{\pi}$ satisfying $\pi_i P_{ii+1} = \pi_{i+1} P_{i+1i}$. (See next week's homework)

Example: List problem

m objects labeled $1, 2, \ldots, m$ are arranged in some order. At each time, independently, an object i is requested with probability q_i, $\sum_{i=1}^{s} q_i = 1$. The selected item is then placed one closer to the front of the list.

$$13425 \xrightarrow{\text{select } 2} 13245 \xrightarrow{\text{select } 5} 13254$$

$\{X_n\} \equiv$ order at n^{th} time $|S| = m!$

For $m = 3$ a path from $123 \longrightarrow 123$, and its reverse path are:

$$\begin{array}{rl} \text{path} & 123 \xrightarrow{3} 132 \xrightarrow{1} 132 \xrightarrow{3} 312 \xrightarrow{2} 321 \xrightarrow{2} 231 \xrightarrow{1} 213 \xrightarrow{1} 123 \\ \text{reverse path} & 123 \xrightarrow{2} 213 \xrightarrow{3} 231 \xrightarrow{3} 321 \xrightarrow{1} 312 \xrightarrow{1} 132 \xrightarrow{1} 132 \xrightarrow{2} 123 \end{array}$$

So, $\Pr\{\text{path}\} = q_1^3 q_2^2 q_3^2 = \Pr\{\text{reverse path}\}$.

If $i_1 i_2 \cdots i_m$ denotes a permutation of $1, 2, \ldots, m$

$$\pi_{i_1 i_2 \cdots i_m} q_{i_{k+1}} = \pi_{i_1 \cdots i_{k+1} i_k \cdots i_m} q_{ik}$$

which implies

$$\pi_{i_1 i_2 \cdots i_m} \propto q_{i_1}^{m} q_{i_2}^{m-1} \cdots q_{i_m}^{1}$$

For this $m = 3$ example, $\pi_{312} \propto q_3^3 q_1^2 q_2^1$.

note: This method has better efficiency than front end replacement.

example: Ising Model