Markov Chains

Frank Kelley

July 5, 2008

This is a beautiful but accessible theory with a wide range of applications. The reader should be familiar with basic probability theory

The book for this course is J R Norris' "Markov Chains".

A preview of the highlight of this course; for a simple symmetric random walk in 1D, the probability of returning to the origin is 1. Likewise, the same is true for a 2D "drunkard's walk". However, in 3D the probability is <1.

1 Definitions and basic properties

A Markov chain with state space I and transition matrix P is a sequence of I-valued RVs (X_n) s.t. $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1}) = \mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) P_{i_n i_{n+1}} \forall i_0, \ldots, i_n,$ $I \forall n \geq 0$ [M1] ($\mathbb{P}(E)$ denote the probability of E to distinguish this from the matrix P).

If [M1] holds then inductively we have $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_0 = i_0) P_{i_0 i_1} \ldots P_{i_{n-1} i_n} \forall i_0, \ldots i_n$ $I \forall n \geq 0$ [M2]; clearly [M2]$\Rightarrow$[M1]

We call the elements of I states.

This definition is very nice - it is clearly very simple and describes reasonable models for a huge range of phenomena, yet we shall see it is sufficient to prove a huge number of useful properties.

Set $\lambda_i = \mathbb{P}(X_0 = i)$; let $\lambda = (\lambda_i : i \in I)$ be the initial distribution of (X_n), i.e. $\lambda_i = \mathbb{P}(X_0 = i)$ [check]. By [M2], λ, P determine the probabilities of any event of the form $\{X_0 = i_0, \ldots, X_n = i_n\}$, called an elementary event. So we write $(X_n) \sim Markov(\lambda, P)$. We usually take $\lambda = \delta_i$ for some $i \in I$, i.e. we start in state i; in this case we can write $\mathbb{P}_i$ rather than $\mathbb{P}$ to indicate we are conditioning on this.

A row vector $(\lambda_i : i \in I)$ with $\lambda_i \geq 0 \forall i \in I$is a measure; its total mass is $\sum_{i \in I} \lambda_i$. If this is 1 it is called a distribution or probability measure; the initial distribution is of course a distribution.

In a slight abuse of notation, we also use λ to denote the function $\lambda(A) = \sum_{i \in A} \lambda_i$ defined on subsets $A \subset I$.

A matrix $P = (P_{ij} : i, j \in I)$ all of whose rows $(P_{ij} : j \in I)$ are distributions is called a stochastic matrix; the transition matrix is stochastic.

An example we shall be returning to frequently is $P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$.

Probability revision (not part of this course)

For a set Ω and set $\mathcal{F}$ of subsets of Ω satisfying $\Omega \in \mathcal{F}, A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ and, for A_n all $\in \mathcal{F}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$, a probability measure is a $\mathbb{P}\mathcal{F} \to [0,1]$ satisfying $\mathbb{P}(\Omega) = 1$ and for disjoint $A_n \in \mathcal{F}$, $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$.

Let I be a finite (all of its elements can be enumerated as $i_1, i_2, \dots, i_n$ for some $n \in \mathbb{N}$) or countably infinite set. The theory we will be learning is valid for countably infinite sets, since if we have $\mu_i \geq 0 \forall i \in I$ then for any two enumerations of I as $i_1, i_2, \dots$ and $j_!, j_2, \dots$, if we set $S_n = \sum_{k=1}^{n} \mu_{i_j}$ and $S = \sum_{k=1}^{\infty} \mu_{i_k}$, similarly T_n and T for the j_k, then $\forall n \exists m : \{i_1, i_2, \dots, i_n\} \subset \{j_1, j_2, \dots j_n\}$ so $S_n \leq T_m \leq T$, so $S \leq T$, similarly $T \leq S$ and $S = T$. Therefore $\sum_{i\in I} \mu_i$ is well defined even for (countably) infinite I.

For a random variable $X : \Omega \to I$ we write $\{X_i\} = \{\omega \in \Omega : X(\omega) = i\}$ and $\mathbb{P}(X = i)$ for $\mathbb{P}(\{X_i\})$.

These definitions are usually just underlying what we do; we won't normally use them directly. However, if we encounter apparent paradoxes we can return to these to resolve them.

Connection with Matrix Multiplication

If we treat λ as a row vector then for $(X_n) \sim Markov(\lambda, P)$, $\mathbb{P}(X_1 = j) = \sum_{i\in I} \mathbb{P}(X_0 = i, X_1 = j) = \sum_{i\in I} \lambda_i P_{ij} = (\lambda P)_j$. Taking $\lambda = \delta_i$ we have $\mathbb{P}_i(X_2 = j)$ is similarly $\left(P^2\right)_{ij}$, and ultimately $\mathbb{P}(X_n = j) = (\lambda P^n)_j$ and $\mathbb{P}_i(X_n = j) = (\delta_i P^n)_{ij}$ or $p_{ij}^{(n)}$. Of course the best way to find these is diagonalising the matrix, or rather simply find the eigenvalues $\lambda_1, \lambda_2, \dots$, and then we know that e.g the first element of P^n is $A\lambda_1^n + B\lambda_2^n \dots$, and can find $A, B, \dots$ by simultaneous equations from the first $|I|$ matricies calculated by hand [The lecturer is lame and can't diagonalise properly].

Thm (Markov Property)

Let (X_n) for $n \geq 0$ be $Markov(\lambda, P)$. For each $n \geq 0$ and $i \in I$, conditioning on the "present" $X_n = i$ the past $(X_0, \dots, X_n)$and future $(X_{n+1}, \dots)$ are indep with the latter $\sim Markov(\delta_i, P)$.

Pf

We want to show $\mathbb{P}\left(X_{n+1}=i_{n+1},\ldots,X_{n+m}=i_{n+m}\mid X_0=i_0,\ldots,X_{n-1}=i_{n-1},X_n=i\right)=\prod_{r=0}^{m-1}P_{j_rj_{r+1}}$ for $j_0=i,j_r=i_{n+r}\forall r\in[1,m]$. We prove this by simply substituting [M2], which also gives that the RHS is the definition of a Markov distribution as required.

2 Class Structure

Consider the diagram of our example (graph with verticies representing states and directed edges labelled with the transition probabilities existing if these are >0) with just the arrows, no numbers; it divides naturally into classes where from any state in a class we can reach any other. Write $i\sim j$ if $\exists n:\mathbb{P}_i\left(X_n=j\right)>0$; "$i$ leads to j" or "j is accessible from i". Equivalent [for $i\neq j$] that $i\sim j$, $\exists n\geq 1,i_1,\ldots i_n$ with $P_{i_ki_{k+1}}>0\forall 0\leq k\leq n$ (considering $i_0=i,i_n=j$), or $\left(P^n\right)_{ij}>0$ for some $n\geq 0$. If $i\sim j$ and $j\sim i$ we write $i\leftrightsquigarrow j$; "i communicates with j", and this is an equiv rel so it partitions I into "communicating classes"; we call a class C "open" if $\exists i\in C,j\notin C$ with $i\sim j$, "you can escape", otherwise C is closed. If I is a class we say P is irreducible; from every state we can reach every other state.

3 Hitting times and absorbtion probabilities

For a Markov chain and $A\subset A$ we define an rv $H^A:\Omega\to\mathbb{N}\cup\{0,\infty\}$ by $H^A(\omega)=\inf_{n\geq 0}\{X_n(\omega)\}$; note $\inf\{\}=\infty$.

Let $h_i^A=\mathbb{P}_i\left(H^A<\infty\right)$ be the prob we hit A (called the absorbtion probability if A is a closed class); $k_i^A=\mathbb{E}_i\left(H^A\right)=\sum_{n<\infty}n\mathbb{P}_i\left(H^A=n\right)+\infty\mathbb{P}_i\left(H_A=\infty\right)$ taking the last term to be 0 if the probability is 0 is the mean time to hit A. $h^A=\left(h_i^A\right)$ and sim k^A are i-vectors over the state space I.

h^A is the minimal nonnegative solution to $h_i^A=1\forall i\in A,h_i^A=\sum_{j\in I}p_{ij}h_j^A\forall i\notin A$; finding the minimal solution is generally the hard part rather than solving. For $i\in A$, $X_0=i\Rightarrow H_A=0<\infty$ so $h_i^A=1$, and for $i\notin A$ $H^A\geq 1h_i^A=\mathbb{P}_i\left(H^A<\infty\right)=\sum_j\mathbb{P}\left(H^A<\infty\mid X_1=j\right)\mathbb{P}_i\left(X_1=j\right)=\sum_j\mathbb{P}_j\left(H^A<\infty\right)p_{ij}$ and so by the Markov property $h_i^A=\mathbb{P}_i\left(H^A<\infty\right)=\sum_j\mathbb{P}\left(H^A<\infty\mid X_1=j\right)\mathbb{P}_i\left(X_1=j\right)=\sum_j\mathbb{P}_j\left(H^A<\infty\right)p_{ij}$, so h solves the given eqns; h is nonnegative as it is a probability (or rather a vector therof). Now if x is any nonnegative sol to the given eqns, $h_i^A=x_i=1$ for $i\in A$; for $i\notin A$, $x_i=\sum_{j\in I}p_{ij}x_j=\sum_{j\in A}p_{ij}+\sum_{j\notin A}p_{ij}x_j$; substituting repeatedly for x_j and rearranging we have $x_i=\mathbb{P}_i\left(X_1\in A\right)+\mathbb{P}_i\left(X_1\notin A,X_2\in A\right)+\cdots+\mathbb{P}_i\left(X_1\notin A,\ldots,X_{n-1}\notin A,X_n\in A\right)+\sum_{j_1\notin A}\cdots\sum_{j_n\notin A}p_{ij_1}p_{j_1j_2}\ldots p_{j_{n-1}j_n}x_{j_n}$; x is non-negative so this last term is ≥ 0 and so $\mathbb{P}_i\left(H^A<n\right)\leq x_i\forall i,n$ so $x_i\geq\lim_{n\to\infty}\mathbb{P}_i\left(H^A\leq n\right)=\mathbb{P}_i\left(H^A\leq\infty\right)=h_i$.

Example: Gambler's Ruin

States $0, 1, \dots$; probability p of moving to the next state, $q = 1 - p$ to the previous state from any non-zero state, with $0 < p < 1$; think of as a casino where we win £1 each turn with probability p, lose £1 with prob q; what is the probability we eventually go broke?. Set h_i to be $\mathbb{P}_i$ (hit 0); this is then the minimal non-negative sol to $h_0 = 1, h_i = ph_{i+1} + qh_{i-1}$ for $i \geq 1$. For $p \neq q$ the general sol is $h_i = A + B\left(\frac{q}{p}\right)^i$; for $p < q$ we must have $B = 0$ as $h_i \leq 1 \forall i$, so $h_i = 1 \forall i$; for $p > q$ the sols are of the form $\left(\frac{q}{p}\right)^i + A\left(1 - \left(\frac{q}{p}\right)^i\right)$; the minimal non-negative one of these is given by $A = 0$ so $h_i = \left(\frac{q}{p}\right)^i$; if $q = p$ we have $h_i = A + B_i = 1 \forall i$. CPS

Example: Birth and Death chain

As above but with the probabilities from state i being p_i to the next state and $q_i = 1 - p_i$ to the previous one. Let X_n be the population size, $h_i = \mathbb{P}_i$ (hit 0) so have $h_0 = 1, h_i = p_i h_{i+1} + q_i h_{i-i}$ for $i \geq 1$. We cannot solve this as a recurrence since p, q are not constant; consider $u_i = h_{i-1} - h_i$ (taking $h_{-1} = 1$) which is clearly ≥ 0, then $p_i u_{i+1} = q_i u_i$ for $i \geq 1$, so $u_{i+1} = \frac{q_i}{p_i} u_i = \cdots = \frac{q_i \times \cdots \times q_1}{p_i \times \cdots \times p_1} u_1$; we let the fraction be γ_i and set $\gamma_0 = 1$. We have $u_1 + \cdots + u_i = h_0 - h_i$ so $h_i = 1 - A\left(\gamma_0 + \cdots + \gamma_{i-1}\right)$; by h_i A is u_1.

In the case $\sum \gamma_i = \infty$ $0 \leq h_i \leq 0 \Rightarrow A = 0$ and $h_i = 1 \forall i$

In the case $\sum_i \gamma_i < \infty$ the sol h_i decreases as A increases, so we want the largest A with $h_i \geq 0$; therefore we want $0 = \lim_{i \to \infty} h_i = 1 - A\left(\sum_{i=0}^{\infty} \gamma_i\right)$ and so $h_i = \frac{\sum_{j=i}^{\infty} \gamma_j}{\sum_{j=0}^{\infty} \gamma_i}$; in particular the population survives with probability > 0; we can check our results by comparison with the above special case.

4 Strong Markov Property

Def a random time τ is an rv taking vals $\infty, 0, 1, 2 \dots$; τ is a stopping time for the Markov chain (X_n) if $\forall n \geq 0$ the event $\{\tau = n\}$ is determined entirely by the rvs $X_0, X_1, \dots, X_n$ (ie is a function of them); informally "we know whether to stop"; if we want to stop at time τ we can.

Conditional on $(X_0 = i_0, \dots, X_n = i_n)$ the event $\tau = n$ is indep of $(X_{n+1}, X_{n+2}, \dots)$.

Examples

The first passage time to state i $T_i = \min\{n : n \geq 1, X_n = i\}$ if such an n exists, ∞ otherwise. $(T_i = n) = (X_1 \neq i, \dots, X_{n-1} \neq i, X_n = i)$ so T_i is a stopping time; note that for T_i finite, $X_{T_i} = i$.

The last time in state i, $L_i = \sup\{n : n \geq 0, X_n = i\}$ is not a stopping time.

Thm (Strong Markov Property)

For X_n $Markov\left(\lambda,P\right)$ and τ a stopping time; conditional on $\tau<\infty$ and $X_\tau=i$, $(X_0,X_1,\ldots,X_{\tau-1})$ and $(X_{\tau+1},X_{\tau+2}\ldots)$ are indep with $(X_\tau,X_{\tau+1},X_{\tau+2}\ldots)\sim Markov\left(\delta_i,P\right)$

Remarks

This is the markov property with n replaced by τ; the point is that τ gives no information about $X_{\tau+1},X_{\tau+2},\ldots$ other than that contained in X_τ. The proof of this is non-examinable.

We can use this to easily solve the gambler's ruin problem above: $h_1=ph_2+q$ but $h_2=h_1\times h_1$ as it is the probability we hit 1 $\times$ the probability we hit 0 starting from 1; the time when we first hit 1 is a stopping time. So we have $h_1=ph_1^2+q$ meaning $h_1=1$ or $\frac{q}{p}$; we find which by minimality.

5 Recurrence and transience

Let $V_i=\sum_{n=0}^{\infty}I_{X_n=i}$ the no of visits to state i, $T_i=\inf\left(n\geq 1:X_n=i\right)$ return time to state i; $f_i=\mathbb{P}_i\left(T_i<\infty\right)$, $m_i=\mathbb{E}_i\left(T_i\right)$ mean return time to state i.

Prop

$\forall k\geq 0$ $\mathbb{P}_i\left(V_i\geq k+1\right)=\left(f_i\right)^k$ - use strong markov after kth visit.

We say a state i is recurrent if $f_i=1$, otherwise i is transient. A recurrent state i is positive recurrent if $m_i<\infty$, otherwise it is null recurrent. So $\mathbb{P}_i\left(V_i=\infty\right)=1$ if i is recurrent, 0 otherwise, from that it is $\left(f_i\right)^k$.

T

A state i is recurrent iff $\sum_{n=0}^{\infty}\left(P^n\right)_{ii}=\infty$, by $\mathbb{E}_i\left(V_i\right)=\mathbb{E}_i\left(\sum_{n=1}^{\infty}I_{X_n=i}\right)=\sum_{n=1}^{\infty}\mathbb{E}_iI_{X_n=i}=\sum_{n=0}^{\infty}\mathbb{P}_i\left(X_n=i\right)$.

There is a recurrence/transience dichotomy: either $f_i=1,\mathbb{P}_i\left(V_i=\infty\right)=1,\sum_{n=0}^{\infty}\left(P^n\right)_{ii}=\infty$ and i is recurrent, or $f_i<1,\mathbb{P}_i\left(V_i=\infty\right)=0,\sum_{n=0}^{\infty}\left(P^n\right)_{ii}<\infty$ and i is transient.

T

Suppose i recurrent and $i\sim j$, then:

1. $\mathbb{P}_j\left(H_i<\infty\right)=1$. Note $\{V_i<\infty\}\supset\left(\{H_j<\infty\}\cap\{X_n\neq i\forall n\geq H_j\}\right)$. Now H_j is a stopping time, and given $H_j<\infty$ we have $X_{H_j}=j$, so by strong Markov $\mathbb{P}_i\left(V_i<\infty\right)\geq\mathbb{P}_i\left(H_j<\infty\right)\mathbb{P}_j\left(H_i=\infty\right)$; the LHS is 0 by recurrence of i. So $2.\Rightarrow\mathbb{P}_i\left(H_j<\infty\right)>0\Rightarrow\mathbb{P}_j\left(H_i=\infty\right)=0$.

2. $\mathbb{P}_i(H_j < \infty) = 1$. Let $T_i^{(0)} = 0, T_i^{(1)} = T_i$ and generally $T_i^{(k)}$ = time of kth return to i. We have $\mathbb{P}_i\left(T_i^{(k)} < \infty\right) = 1$ as i recurrent, so $X_{T_i^{(k)}} = i$. Now define the events $A_k = \left\{X_n = j : T_i^{(k-i)} \leq n < T_i^{(k)}\right\}$ for $k \geq 1$. By strong Markov on $T_i^{(k-1)}$ $\mathbb{P}_i(A_k) = \mathbb{P}_i(H_j < T_i)$; say this is p. The A_k are indep so $\mathbb{P}\left(\bigcup_k A_k\right)$ is either 0 for $p = 0$ or 1 for $p > 0$, but this is $\mathbb{P}_i(H_j < \infty)$ which is > 0 since $i \sim j$ so must be 1.

3. j is recurrent; this follows flom the above two and strong Markov; $\mathbb{P}_j(T_j < \infty) \geq \mathbb{P}_i(H_j < \infty)\mathbb{P}_j(H_i < \infty) = 1$. Note that this implies recurrence (and so transience) are class properties; combined with 1. above this means every recurent class is closed.

T

Every finite closed class C is recurrent; take any inital dist on C, then $\sum_{i\in C} V_i = \infty$ as C closed, so $\sum_{i\in C}\mathbb{P}(V_i = \infty) \geq \mathbb{P}\left(\bigcup_{i\in C}\{V_i = \infty\}\right) = 1$ so for some i, $0 < \mathbb{P}(V_i = \infty) = \mathbb{P}(H_i < \infty)\mathbb{P}_i(V_i = \infty)$; $\mathbb{P}_i(V_i = \infty)$ is always 0 or 1 su must be 1 and i is recurrent and C is recurrent.

Finite state spaces are "easy"; infinite state spaces are more interesting as we can have closed classes but be uncertain whether they are recurrent.

6 Recurrence and Transience of random walks

Simple Symmetric random walk on $\mathbb{Z}$

$I = \mathbb{Z}, P_{ii+1} = P_{ii-1} = \frac{1}{2}$. Let $h = \mathbb{P}_1(\text{hit } 0)$ which is $\mathbb{P}_{-1}(\text{hit } 0)$ by symmetry; by the homogeneity (probs the same $\forall i$) and strong Markov, $\mathbb{P}_2(\text{hit } 0) = h^2$ ($\mathbb{P}_2(\text{hit } 1) \times \mathbb{P}_1(\text{hit } 0)$; then $h = \frac{1}{2} + \frac{1}{2}h^2$ which gives $(h-1)^2 = 0$ so $h = 1$ and the walk is recurrent.

Simple biased random walk on $\mathbb{Z}$

$I = \mathbb{Z}, P_{ii+1} = p, P_{ii-1} = q, p + q = 1$. wlog take $q < p$, then $\mathbb{P}_0(T_0 < \infty) = ph_+ + qh_-$ where $h_\pm = \mathbb{P}_{\pm 1}(\text{hit } 0)$, but by an earlier example $h_+ = \frac{q}{p}, h_- = 1$ so $\mathbb{P}_0(T_0 < \infty) = 2q < 1$; the walk is transient, we are not certain to return to 0. The reader should now consider the simple symmetric random walk on the plane, which can be reduced to the above by projection onto the axes $x = \pm y$.

2D simple symmetric random walk

If $(X_n), (Y_n)$ indep 1D simple symmetric random walks, $\left(\frac{X_n+Y_n}{2}, \frac{X_n-Y_n}{2}\right)$ is clearly a 2D simple symmetric random walk. To return we must have each of the 1D random walks return, i.e. $\left(P^{2m}\right)_{(0,0)(0,0)} = \left(\left(\begin{array}{c} 2m \\ m \end{array}\right)\left(\frac{1}{2}\right)^{2m}\right)^2$ since

$\begin{pmatrix} 2m \\ m \end{pmatrix} \left(\frac{1}{2}\right)^{2m}$ is the probability of a 1D simple symmetric random walk returning to the origin after $2m$ steps (we must step left m times and right m times). This tends to $\left(\frac{1}{\sqrt{\pi m}}\right)$ by Stirling's formula so $\left(P^{2m}\right)_{(0,0)(0,0)} \sim \frac{1}{\pi m}$ and so as $\sum_m \frac{1}{\pi m} = \infty$, $\sum_m \left(P^{2m}\right)_{(0,0)(0,0)} = \infty$ and the random walk is recurrent.

It is not so easy to express a 3D random walk in terms of 1D random walk as each point has $6 \neq 2^n$ neighbours.

3D simple symmetric RW

$\left(P^{2n}\right)_{\vec{0}\vec{0}} = \sum_{i,j,k\geq 0:i+j+k=n} \begin{pmatrix} 2n \\ iijjkk \end{pmatrix}$(this is a multinomial coefficient rather than a binomial one with a product on the bottom) $\left(\frac{1}{6}\right)^{2n}$; we write it as $\begin{pmatrix} 2n \\ n \end{pmatrix} \left(\frac{1}{2}\right)^{2n} \sum_{i,j,k\geq 0:i+j+k=n} \begin{pmatrix} n \\ ijk \end{pmatrix}^2 \left(\frac{1}{3}\right)^{2n} \leq \begin{pmatrix} 2n \\ n \end{pmatrix} \left(\frac{1}{2}\right)^{2n} \left(\frac{1}{3}\right)^n \max_{i,j,k\geq 0:i+j+k=n} \begin{pmatrix} n \\ ijk \end{pmatrix} \left(\sum_{i,j,k}\right.$ this last bracketed term is 1 (consider the expansion of $(1+1+1)^n$) so this is $\begin{pmatrix} 2n \\ n \end{pmatrix} \left(\frac{1}{2}\right)^{2n} \left(\frac{1}{3}\right)^n \max_{i,j,k\geq 0:i+j+k=n} \begin{pmatrix} n \\ ijk \end{pmatrix}$. Now for $n = 3m$, $\begin{pmatrix} n \\ ijk \end{pmatrix} = \frac{n!}{i!j!k!} \leq \begin{pmatrix} n \\ mmm \end{pmatrix}$, and a bound on $\left(P^{6n}\right)_{\vec{0}\vec{0}}$ for $3m$ bounds all $\left(P^{2n}\right)_{\vec{0}\vec{0}}$ by at most a constant times this bound, since $\left(P^{6n}\right)_{\vec{0}\vec{0}} \geq \left(\frac{1}{6}\right)^2 \left(P^{6n-2}\right)_{\vec{0}\vec{0}} \geq \left(\frac{1}{6}\right)^4 \left(P^{6n-4}\right)_{\vec{0}\vec{0}}$, as we can certainly return to 0 at time $6n$ by returning to 0 at time $6n-2$, then moving to $(1,0,0)$ and back, and similarly. But we now have $\left(P^{6n}\right)_{\vec{0}\vec{0}} \leq \begin{pmatrix} 2n \\ n \end{pmatrix} \left(\frac{1}{2}\right)^{2n} \left(\frac{1}{3}\right)^n \begin{pmatrix} n \\ mmm \end{pmatrix} \sim \frac{c}{n^{\frac{3}{2}}}$ for some constant c as $n \to \infty$ so $\sum_n \left(P^n\right)_{\vec{0}\vec{0}}$ (terms are clearly 0 for n odd) $< \infty$ and the random walk is transient; this means it $\to \infty$ as otherwise it returns to some ϵ-ball about the origin infinitely often and it is recurrent.

7 Invariant Distributions

We say a dist or measure λ is invariant if $\lambda = \lambda P$; λ is a left evector of P with eval 1. Proofs of the properties of these are quite technical so we shall cover some examples first.

As notation we let $T_i = \inf\{n \geq 1 : X_n = i\}$; we call this the return time even if we do not start in i. $V^i_j = V_j(T_i)$ the number of visits to j before we first return to i, $m_i = \mathbb{E}_i(T_i)$ and $Y^i_j = \mathbb{E}_i\left(V^i_j\right)$. Under suitable conds $m_i = \frac{1}{\pi_i}, Y^i_j = \frac{\pi_j}{\pi_k}$ for π an invariant dist (or $Y^i_j = \frac{\lambda_j}{\lambda_j}$ for λ an invariant measure), $\left(P^n\right)_{ij} \to \pi_j$ and $\mathbb{E}\left(\frac{V_j(n)}{n}\right) \to \pi_j$ as $n \to \infty$. Note this means the 3D simple symmetric RW has no invariant dist.

As well as directly calculating π by forming equations for each of its cpts, we can try the detailled balance (DB) eqns $\pi_i P_{ij} = \pi_j P_{ji}$. If π solves these then

$\pi P = \pi$ as summing over j, $\sum_j \pi_j P_{ij} = \pi_i \sum_j P_{ij} = \pi_i$; however, sometimes there is no solution to these eqns, even if an invariant dist exists, e.g. for $P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$.

If we know an invariant dist is unique and the chain has some symmetry we can use this to find it. Note, however, there does not have to exist an invariant dist by e.g. the infinite chain on $\mathbb{N}_0$ $p_{ii+1} = 1$, $p_{ij} = 0$ otherwise, or even an invariant measure, by e.g. the success run chain on $\mathbb{N}_0$ $p_{ii+1} = p_i, p_{i0} = q_i = 1 - p_i$ if we choose p_i st $p_i < 1$ for infinitely many i and $r = \lim_{i \to \infty} r_i > 0$ where $r = p_0 p_1 \dots p_{i-1}$. Invariant dists are not necessarily unique, e.g. all dists on $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ are invariant, but even for an irreducible chain we have the chain on $\mathbb{Z}$ given by $p_{ii+1} = p, p_{ii-1} = q = 1 - p$ with $p \neq q$, then $\lambda_i = 1$ and $\mu_i = \left(\frac{p}{q}\right)^i$ are both invariant measures.

7.1 Solidarity Property

P irredicible, $0 \leq \lambda_i \leq \infty \forall i$ and $\lambda = \lambda P \Rightarrow \lambda \equiv 0$ or $0 < \lambda_i < \infty \forall i$ or $\lambda \equiv \infty$

Pf

$\lambda = \lambda P = \lambda P^2 = \dots = \lambda P^n \forall n$; given (i, j), $\exists n : (P^n)_{ij} > 0$ since P irredicible, so $\lambda_i = \sum_k \lambda_k (P^n)_{kj} \geq \lambda_i (P^n)_{ij}$ so $\lambda_j < \infty \Rightarrow \lambda_i < \infty$ and $\lambda_i > 0 \Rightarrow \lambda_j > 0$

7.2 Prop

Recall $\lambda_j^i = \mathbb{E}_i\left(V_j^i\right) = \mathbb{E}_i\left(\sum_{n=0}^{T_i - 1} I_{X_n = j}\right)$ the expected no. of visits to j between visits to i. $\gamma^k = \left(\gamma_j^k : j \in I\right)$ is the minimal non-neg sol to $\lambda_j = (\lambda P)_j$ for $j \neq k$ and $\lambda_k = 1$

Pf

For $j \neq k$, $\gamma_j^k = \mathbb{E}_k\left(\sum_{n=1}^{T_k} I_{X_n = j}\right)$ $(\star)$ since we are in state $k \neq j$ at both time 0 and T_k; this is $\mathbb{E}_k \sum_{n=1}^{k} I_{X_n = j, n \leq T_k} = \sum_{n=1}^{\infty} \mathbb{E}_k I_{X_n = j, n \leq T_k} = \sum_{n=1}^{\infty} \mathbb{P}_k\left(X_n = j, n \leq T_k\right)$ which crucially is $\sum_{n=1}^{\infty} \sum_{i \in I} \mathbb{P}_k\left(n \leq T_k, X_{n-1} = i, X_n = j\right)$; T_k being $\geq n$ depends only on $X_1, \dots, X_{n-1}$ so by [strong] Markov this is $\sum_{n=1}^{\infty} \sum_{i \in I} \mathbb{P}_k\left(n \leq T_k, X_{n-1} = i\right) P_{ij} = \sum_{i \in I} \sum_{n=1}^{\infty} \mathbb{P}_k\left(n \leq T_k, X_{n-1} = i\right) P_{ij}$ which letting $m = n - 1$ is $\sum_{i \in I} \sum_{m=0}^{\infty} \mathbb{P}_k\left(m \leq T_k, X_m = i\right) P_{ij} = \sum_{i \in I} \gamma_i^k P_{ij}$, so γ^k is a sol; suppose λ also a sol, then for $j \neq k$ $\lambda_j = \sum_{i \in I} \lambda_i P_{ij} = P_{kj} + \sum_{i \neq k} \lambda_i P_{ij} = \dots = P_{kj} + \sum_{i_1 \neq k} P_{ki_1} P_{i_1 j} + \dots + \sum_{i, i_1, \dots, i_{n-1} \neq k} P_{ki_{n-1}} \dots P_{i_2 i_1} P_{i_1 j} + \sum_{i_1, \dots, i_n \neq k} \lambda_{i_n} P_{i_n i_{n-1}} \dots P_{i_2 i_1} P_{i_1 j}$; the second last term here is $\mathbb{P}_k\left(X_n = j, T_k \geq n\right)$ and sim the prev terms so if $\lambda \geq 0$ then for $k \neq j$, $\lambda_j \geq \sum_{m=0}^{n} \mathbb{P}_k\left(X_m = j, T_k \geq m\right)$ by ommiting the final term, so letting $n \to \infty$ $\lambda_j \geq \sum_{m=0}^{\infty} \mathbb{P}_k\left(X_m = j, T_k \geq m\right) = \gamma_j^k$ and we are done.

7.3 Thm

Suppose P irreducible and recurrent, then γ^k is the unique invariant measure λ with $\lambda_k = 1$

Pf

Since P is recurrent $\star$ remains true for $j = k$ so $\gamma^k = \gamma^k P$ by the same argument as in 7.2; since P irreducible and $\gamma_k^k = 1$ we have $\gamma_j^k < \infty \forall j$ by 7.1, thus γ^k is an invariant measure; by 7.2 if λ an invariant measure w/ $\lambda_k = 1$ then $\lambda \geq \gamma^k$ so $\lambda - \gamma^k \geq 0$ and this is an invariant measure [since it is a lin comb of such] , but its k cpt is 0 so by 7.1 it is $\equiv 0$.

7.4 Thm

For P irreducible the following are equiv:

1. Every state is positive recurrent (i.e. $m_i = \mathbb{E}_i (T_i) < \infty \forall i$)
2. Some state i is positive recurrent
3. $\exists$ an invariant dist π

and under these conds $m_i = \frac{1}{\pi_i} \forall i$.

Pf

$1 \Rightarrow 2$ trivially; for $2 \Rightarrow 3$ set $\pi_j = \frac{\gamma_j^i}{\pi_i}$ and apply 7.3: $T_i = \sum_{j\in I} \sum_{n=1}^{T_i} I_{X_n = j}$ so $m_i = \mathbb{E}\left(\sum_{j\in I} \sum_{n=1}^{T_i} I_{X_n=j}\right) = \sum_{j\in I} \gamma_j^i$. For $3 \Rightarrow 1$ fix k and set $\lambda_j = \frac{\pi_j}{\pi_k}$, then $\lambda \geq \gamma^k$ by 7.2 so $m_k = \sum_{j\in I} \gamma_j^k \leq \sum_{j\in I} \lambda_j = \frac{1}{\pi_k} < \infty$ so m_k finite and k recurrent, so P is recurrent; by 7.3 $\lambda = \gamma^k$ and so $m_k = \frac{1}{\pi_k}$.

If P is irreducible and I finite then as per chapter 5 P is recurrent and so an invariant measure λ exists, then $\sum_{i\in I} \lambda_i < \infty$ since I is finite so we can normalize λ to get an invariant dist.

Summary: existance and uniqueness of invariant measures and dists

Without irreducibility we can find Markov chains with no or many invariant dists, as above; assuming irreducibility, a finite state sp $\Rightarrow$ a unique invariant dist $\Leftrightarrow$ positive recurrence; for a general state sp recurrence implies $\exists$ a "unique" invariant measure up to scalar multiplication (since we found the unique invariant measure with $\lambda_k = 1$). For transient chains even with irreducibility we can find examples with no or many invariant measures.

If P is irreducible and an invariant dist exists then it is unique.

As an example, we can apply the above graph example to a knight moving randomly on a chessboard; the corners have valence $v_C = 2$, the squares next to them 3, the rest of the edge of the board and the corners of the "next square in" have valence 4 and so forth; the average return time for a corner square is $m_C = \frac{1}{\pi_C} = \frac{\sum v_i}{v_C}$ which we can find to be 168.

8 Conv to equilibrium

Main T

We shall not prove this as yet; if P is irreducible and aperiodic and has invariant dist π then for any initial dist $\mathbb{P}(X_n = j) \to \pi_j$ as $n \to \infty$.

We def a state i is aperiodic if $\exists n_1, \ldots, n_k \geq 1$ w/ no common factor s.t. $(P^{n_1})_{ii}, \ldots, (P^{n_k})_{ii}$ all > 0; if $n_1, \ldots, n_k \geq 1$ with no common factor then $\exists N : n \geq N \Rightarrow n = a_1 n_1 + \cdots + a_k n_k$ some $a_1, \ldots, a_n \in \mathbb{N}_0$ (of course the a_i depend on n).

Lemma

For P irreducible w/ an aperiodic state i, $\forall j, k \in I$ $(P^n)_{jk} > 0$ and all states are aperiodic (periodicity is a class property), since can find $r, s \geq 0$ w/ $(P^r)_{ji}, (P^s)_{ik} > 0$, then use aperiodicity of i.

The lecturer here performed a card trick; get n people to each think of a number from 1 to 10, then turn through the cards in a deck; each person counts their number of cards, then changes their number to the number on the card which appears and repeats (treating court cards as 6s, iirc); we find that when the pack has been exhausted everyone's number is the same. This works by coupling; observe that once two people reach the same state by chance (i.e. finish counting to their current number on the same card) they will then remain "locked together" forever, so it is likely all people will be in the same state by the time the pack is finished.

T

This is the most subtle T in the course: for P irreducible and aperiodic w/ invariant dist π, for any initial dist, $\mathbb{P}(X_n = j) \to \pi_j$ as $n \to \infty$, $\forall j$.

Pf (non-examinable)

Suppose $(X_n)_{n\geq 0} \sim Markov(\lambda, P)$, $(Y_n)_{n \geq 0}$ indep $\sim Markov(\pi, P)$. For a fixed state b set $T = \inf\{n \geq 1 : X_n = Y_n = b\}$. We first show $\mathbb{P}(T < \infty) = 1$; let $(W_n)_{n\geq 0}$ be the Markov chain (X_n, Y_n); it has trans mat $\tilde{P}$ with $\left(\tilde{P}^n\right)_{(i,j)(k,l)}$ and initial dist $\mu_{(i,k)} = \lambda_i \pi_k$. It has an invariant dist $\tilde{\pi}_{(i,k)} = \pi_i \pi_k$ as we can verify from the eqns defning an invariant dist; by aperiodicity from the previous L $\left(\tilde{P}^n\right)_{(i,j)(k,l)} > 0$ for sufficiently large n [the lecturer claimed how large n must be depends on (i, j, k, l) but this is not actually necessary], so $\tilde{P}$ is irreducible; by 7.4 it is (positive) recurrent so $\mathbb{P}(T < \infty) = 1$.

Now [conditioning on $T < \infty$] set $Z_n = X_n$ for $n \leq T$, Y_n for $n > T$. (T is a stopping time for W so) By strong Markov $(X_{T+n}, Y_{T+n})_{n \geq 0}$ is $Markov\left(\delta_{(b,b)}, \tilde{P}\right)$ and indep of $(X_0, Y_0), \ldots, (X_T, Y_T)$; by symmetry so is (Y_{T+n}, X_{T+n}). So $W_n' = (Z_n, Z_n')$ is $Markov\left(\mu, \tilde{P}\right)$ (where $Z_n' = Y_n$ for $n \leq T, X_n$ for $n > T$); in particular $(Z_n)_{n \geq 0}$ is $Markov(\lambda, P)$.

Now $\mathbb{P}(Z_n = j) = \mathbb{P}(X_n = j, n \leq T) + \mathbb{P}(Y_n = j, n > T)$ so $|\mathbb{P}(X_n = j) - \pi_j| = |\mathbb{P}(Z_n = j) - \mathbb{P}(Y_n = j)|$ since Z_n and X_n have the same dist. But this is $|\mathbb{P}(X_n = j, n \leq T) - \mathbb{P}(Y_n = j, n \leq T)| \leq \mathbb{P}(n \leq T) \to 0$ as $n \to \infty$ since $\mathbb{P}(T < \infty) = 1$.

Coupling is a very powerful technique

9 Time Reversal

Thm

Let P irreducible w/ invariant distn π; suppose $(X_n)_{0 \leq n \leq N} \sim Markov(\pi, P)$ and set $Y_n = X_{N-n}$, then $(Y_n)_{0 \leq n \leq N} \sim Markov\left(\pi, \hat{P}\right)$ where $\hat{P}$ given by $\pi_j \widehat{P}_{ji} = \pi_j P_{ij} \forall i, j$; also $\hat{P}$ is irreducible w/ invariant distn π.

Pf

For $n = 0, 1, \ldots, N$ $\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_n = i_n) = \mathbb{P}(X_{N-n} = i_n, \ldots, X_N = i_0) = \pi_{i_n} P_{i_n i_{n-1}} P_{i_{n-1} i_{n-2}} \ldots P_{i_1 i_0} = \pi_{i_0} \hat{P}_{i_0 i_1} \ldots \hat{P}_{i_{n-1} i_n}$; next $X_{N-1} \sim \pi$ and $Y_1 \sim \pi\hat{P}$ but these are the same so $\pi = \pi\hat{P}$; irreducibility by $i \sim j$ under P iff $j \sim i$ under $\hat{P}$.

If the detailled balance conds are satisied i.e. $\pi_j P_{ji} = \pi_i P_{ij}$ then $\hat{P} = P$; a Markov chain in the form of a line of states must satisfy DB if an invariant exists; more generally this is true if the graph of the state space is a tree (i.e. has no cycles).

10 Ergodic T

Ergodic generally means of or about averages over time.

T (Strong law of large numbers) (without proof)

For $Y_1, Y_2, \ldots$ non-neg i.i.d. rvs w/ $\mathbb{E}Y_1 = \mu$, $\mathbb{P}\left(\frac{Y_1+Y_2+\cdots+Y_n}{n} \to \mu \text{ as } n \to \infty\right) = 1$. Any reader confused by this statement should return to their basic definitions; any point ω in the state space Ω defines the value of all the Y_i, so it is then a simple fact whether or not $\frac{Y_1+Y_2+\cdots+Y_n}{n} \to \mu$. Of course there may exist sets of values for Y_i for which this is not the case, e.g. if the Y_i are coin tosses (i.e.

1 w/ prob $\frac{1}{2}$, 0 otherwise) $Y_i = 1 \forall i$ has $\frac{Y_1+Y_2+\cdots+Y_n}{n} = 1 \nrightarrow \frac{1}{2} = \mu$ - but the statement is that the probability mass of all such states is 0; all the probability is concentrated on the states where this does $\to \mu$.

Contrast this w/ the WLLN; that considers static times [before taking a limit over them, as now], wheras this is an average over time.

$V_i(n) = \sum_{m=0}^{n-1} I_{X_m=i}$, the no. of visits to state i before time n.

Thm (Ergodic T for Markov Chains)

Let P be irredicible and λ any initial dist, then $\mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1$; the long run proportion of time spent in state i is $\frac{1}{m_i}$. Often $\frac{1}{m_i} = \pi_i$ where π is an invariant dist (in fact, this is the case precisely when $0 < m_i < \infty$ - this result is true for many chains without invariant dists), which gives a usefil way to calculate $\frac{1}{m_i}$.

If P is transient then $\mathbb{P}_i(V_i = \infty) = 0$ and conditioning on $V_i < \infty$, $\frac{V_i(n)}{n} \to 0$ as $n \to \infty$; $0 = \frac{1}{m_i}$ so we are done. Otherwise P is recurrent; $\mathbb{P}(H_i < \infty) = 1$; note that $\frac{V_i(n)}{n}$ converges iff $\frac{1}{n}\sum_{m=0}^{n-1} I_{X_{H_i+m}=i}$ does and with the same limit, so by strong Markov at H_i it suffices to consider $\lambda = \delta_i$.

Set $T^{(0)} = 0$, let $T^{(k)}$be the time of the kth return to state i. Set $S^{(k)} = T^{(k)} - T^{(k-1)}$ so $T^{(k)} = S^{(1)} + \cdots + S^{(k)}$; by strong Markov the $S^{(i)}$ are i.i.d. rvs w/ expectation $\mathbb{E}_i(T_i) = m_i$, so by SLLN $\mathbb{P}_i\left(\frac{S^{(1)}+\cdots+S^{(k)}}{k} \to m_i \text{ as } k \to \infty\right) = 1$; we have $S^{(1)} + \cdots + S^{(V_i(n)-1)} = T^{(V_i(n)-1)}$ the time of the last visit to i before n so $\leq n-1$, $S^{(1)} + \cdots + S^{(V_i(n))} = T^{(V_i(n))}$ the time of the first visit to i after time n so $\geq n$; then $\frac{S^{(1)}+\cdots+S^{(V_i(n)-1)}}{V_i(n)} < \frac{n}{V_i(n)} \leq \frac{S^{(1)}+\cdots+S^{(V_i(n))}}{V_i(n)}$; since P is recurrent $\mathbb{P}(V_i(n) \to \infty \text{ as } n \to \infty) = 1$, so $\mathbb{P}\left(\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty\right) = 1$, but for any state where $\frac{n}{V_i(n)} \to m_i$, $\frac{V_i(n)}{n} \to \frac{1}{m_i}$ (as $n \to \infty$), so $\mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1$.

This completes the content of this course. Related part II courses are applied probability, the continuation of this course to cnts time, and probability and measure which looks at the mathematical foundations for this course; parts of this course are applied in the courses Stochastic Financial Models and Mathematical Biology.