Intro to Markov Chains

- Markov property and Chapman-Kolmogorov Eqns
- Classification of states
- Existence & uniqueness of stationary distribution
- Finite chains & Perron-Frobenius
- Reversibility
- The Ergodic Theorem for HMC
- Foster-Lyapunov condition

- Stochastic Process - Collection of r.v. $(X_t; t \in T)$, $X_t \in \mathcal{X}$, on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and indexed by a time parameter t.
  - $T = \mathbb{N}_0$, $\mathcal{X}$ = discrete $\rightarrow$ discrete-time, discrete-space process. Eg - random walk, branching process
  - $T = \mathbb{R}_+$, $\mathcal{X}$ = discrete $\rightarrow$ continuous-time, discrete-space process
    Eg - Poisson process, queueing models, epidemics
  - $T = \mathbb{R}_+$, $\mathcal{X}$ = continuous $\rightarrow$ continuous-time, continuous-space process
    Eg - Brownian motion

- $\boxed{\text{Markov chain}}$ - Stochastic process $(X_n; n \in \mathbb{N}_0)$ on discrete space $\mathcal{X}$ obeying $\forall n \in \mathbb{N}_0, (x_0, x_1, \ldots, x_{n-1}, x) \in \mathcal{X}^{n+1}$

$$\mathbb{P}[X_n = x \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] = \mathbb{P}[X_n = x \mid X_{n-1} = x_{n-1}]$$

- If in addition, $\mathbb{P}[X_n = x \mid X_{n-1} = y] = \mathbb{P}[X_m = x \mid X_{m-1} = y]$ for all $n, m \in \mathbb{N}_0$, then the Markov Chain is said to be time-homogeneous (or homogeneous Markov Chain or HMC)

- HMC (X_n) has associated transition probability matrix $P = \{P_{ij}\}_{i,j \in X}$, where

$$P_{ij} = \mathbb{P}[X_{n+1} = i \mid X_n = j]$$

- Properties of $P \equiv$ i) $P_{ij} \geq 0 \;\forall i,j \in X^2$
  ii) $\sum_{j \in X} P_{ij} = 1 \;\forall i \in X$

Any matrix with these properties is a stochastic matrix (note though that X may be finite or countably infinite)

- We want to study X_n starting from some $X_0 \in X$

Some notation (all vectors are <u>column</u> vectors)

- $\Pi_n = (\Pi_n(i))_{i \in X}$, $\sum_{i \in X} \Pi_n(i) = 1 \equiv$ Distribution of X_n

$\Pi_0 \equiv$ Starting distribution of chain

- $P_{ij}(m) = \mathbb{P}[X_{n+m} = j \mid X_n = i] \equiv$ m-step transition matrix

$\Rightarrow$ By definition, $\underline{\Pi_n^T = \Pi_0^T P(n)}$, $\Pi_{n+m}^T = \Pi_n^T P(n)$

(Chapman-Kolmogorov Eqns) For an HMC, we have $P(n) = P^n \;\forall n \in \mathbb{N}_0$, and hence

$$\underline{\Pi_{n+m}^T = \Pi_n^T P^m} \quad \forall n,m \in \mathbb{N}_0$$

- The Chapman-Kolmogorov eqns give a linear algebraic view of an HMC. An alternate probabilistic view is to define it in terms of a recurrence relation

(Recurrence View of HMC) - Let $(Z_n; n \in \mathbb{N})$ be an iid sequence of random variables in some space F, and let $\mathcal{X}$ be a countable space. Given any function $f: \mathcal{X} \times F \longrightarrow \mathcal{X}$, and $X_0 \in \mathcal{X}$, the recurrence relation

$$X_{n+1} = f(X_n, Z_{n+1}) \quad , n \in \mathbb{N}$$

defines a HMC $(X_n; n \in \mathbb{N}_0)$.

Eg (Simple random walk) - $(X_n; n \in \mathbb{N}_0)$ on $\mathcal{X} = \mathbb{Z}$ is called a simple random walk if $X_0 \sim \pi_0$, and

- (Matrix view) Let $P = (P_{ij})$ where $P_{i,i+1} = p$, $P_{i,i-1} = 1-p$ and $P_{ij} = 0$ if $j \notin \{i-1, i+1\}$. Then $X_n \sim \pi_n$ with $\pi_n^T = \pi_0^T P^n \; \forall n$
- (Recurrence view) Let $Z_n = \begin{cases} 1 & \text{wp } p \\ -1 & \text{wp } 1-p \end{cases}$. Then $X_{n+1} = X_n + Z_{n+1}$

(The RW is said to be symmetric if $p = 1/2$)

• Any stochastic matrix $P \equiv$ fn $f(X_n, Z_{n+1})$ with $Z_{n+1} \sim U[0,1]$

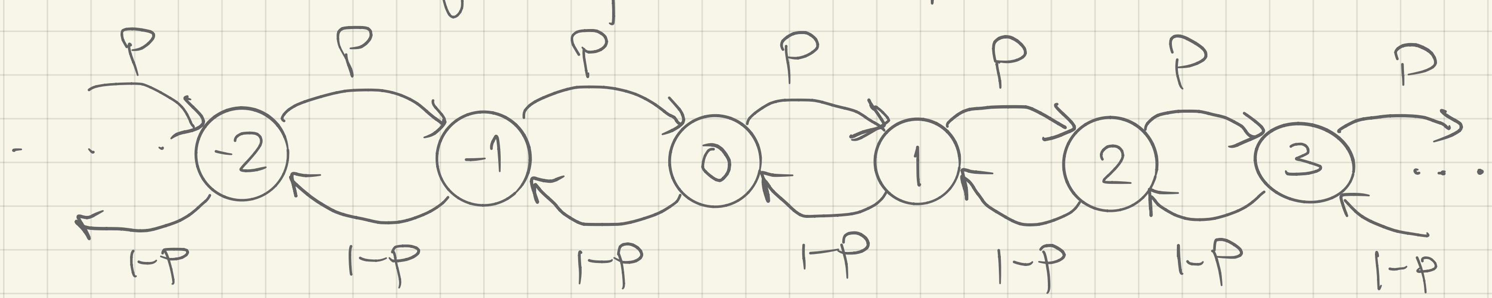
(If $X_n = i$, then choose $X_{n+1} = j$ if $\sum_{k=0}^{j-1} P_{ik} \le Z_{n+1} < \sum_{k=0}^{j} P_{ik}$)

Any $f(X_n, Z_{n+1})$ for any $Z_{n+1} \in F \equiv$ stoch matrix P

(set $P_{ij} = \mathbb{P}[f(X_n, Z_{n+1}) = j \mid X_n = i]$)

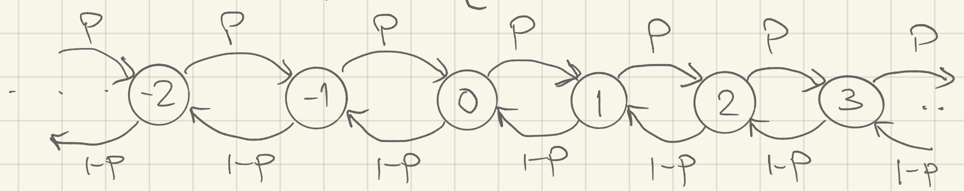
Finally, any MC can also be viewed as a random walk on an edge-weighted directed graph (Random Walk View of HMC) - Consider an edge-weighted directed graph $G(V, E, W)$ with $V = \mathcal{X}$, $(i,j) \in E$ if $P_{ij} > 0$, and $W_{ij} = P_{ij}$. Then HMC $(X_n, n \in \mathbb{N})$ corresponds to a random walk on G, where the walk transitions from node i to a neighboring node j with probability W_{ij}. The graph $G(V, E, W)$ is called a transition diagram.

• Transition diagram for the simple random walk

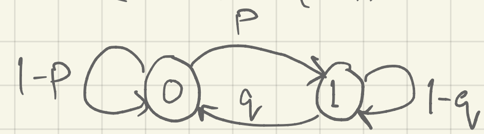


Examples of Markov Chains

- **Simple Random Walk** – $X_n = X_{n-1} + Z_n$, $Z_n \sim \begin{cases} +1 & \text{wp } 1/2 \\ -1 & \text{wp } 1/2 \end{cases}$

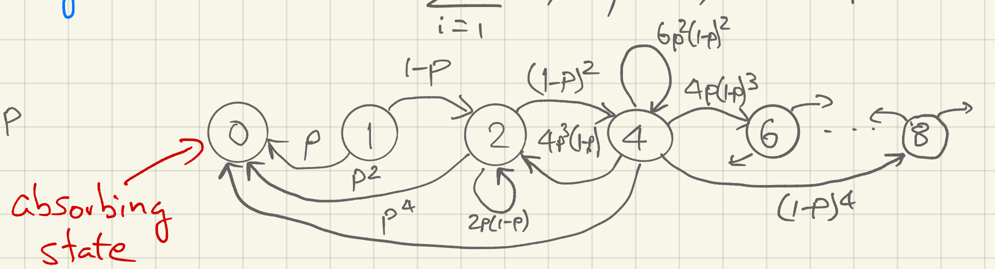


- **Markov Modulated Switch** – $X_n = (X_{n-1} + Y_n(X_{n-1})) \bmod 2$, $Y_n(x) \sim \begin{cases} Ber(p); & x=0 \\ Ber(q); & x=1 \end{cases}$

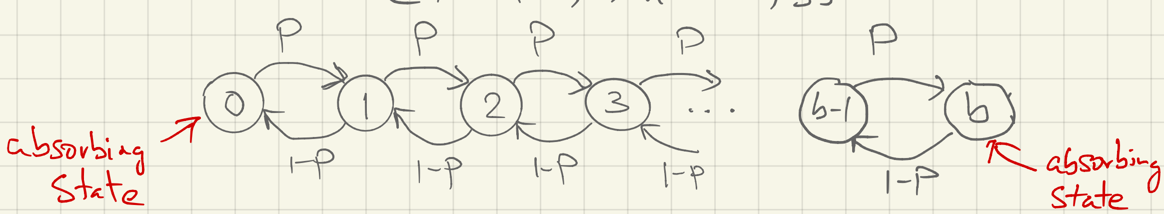


- **(Galton-Watson) Branching Process** – $X_n = \sum_{i=1}^{X_{n-1}} Z_{n,i}$, $Z_{n,i} \sim \{p_k\}_{k\in\mathbb{N}_0}$

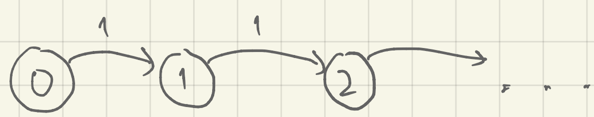
Eg - If $Z_{n,i} \sim \begin{cases} 0 & \text{wp } p \\ 2 & \text{wp } 1-p \end{cases}$



- **Gambler's Ruin** $X_n = \begin{cases} X_{n-1} + Z_n; & X_{n-1} \notin \{0,b\} \\ X_{n-1}; & X_n \in \{0,b\} \end{cases}$, $Z_n \sim \begin{matrix} 1 \text{ wp } p \\ -1 \text{ wp } 1-p \end{matrix}$



- **Deterministic Monotone Markov Chain** $X_n = X_{n-1} + 1$



(Useful for counterexamples)

- **Random Walk on G(V,E)** - Let $A = (A_{ij} = \mathbb{1}_{\{(i,j)\in E\}})$ be the adjacency matrix of G, and $D^{-1} = \mathrm{diag}(1/\deg(i))$, where $\deg(i) = \sum_j A_{ij}$. Then the RW on G is given by the transition matrix $P = D^{-1}A$

Some quantities associated with Markov chains

- Hitting Time - $\{X_n\}_{n\in\mathbb{N}_0}$ Markov chain on $\mathcal{X}$. For any set of states $B \subseteq S$, hitting time $\tau_B = \inf\{n\in\mathbb{N}_0 \mid X_n \in B\}$ (for some X_0) ($\tau_B = 0$ if $X_0 \in B$, $\tau_B \triangleq +\infty$ if $X_n \notin B \;\forall n$)

- (First) Visit Time - For any state $j \in \mathcal{X}$, its first visit time is defined as $T_j(1) = \inf\{n \in \underbrace{\mathbb{N}}_{\text{note: not } \mathbb{N}_0} \mid X_n = j\}$, and its k^{th} visit time is defined as $T_j(k) = \inf\{n > T_j(k-1) \mid X_n = j\}$

- Return Time - For any state $j \in \mathcal{X}$, its return time is defined as $\tau_{jj} = \inf\{n \in \{1,2,\ldots\} \mid X_n = j, X_0 = j\}$

- Cover Time - For any MC on $\mathcal{X}$, cover time $\tau_{cover} = \inf\{n \in \mathbb{N} \mid n \geq T_j(1) \,\forall j \in \mathcal{X}\}$

Classification of States (Probabilistic)

- A state $j \in \mathcal{X}$ is said to be
  - recurrent if $\mathbb{P}[\tau_{jj} < \infty] = 1$
    - positive recurrent if $\mathbb{E}[\tau_{jj}] < \infty$
    - null recurrent if recurrent but not positive recurrent
  - transient if $\mathbb{P}[\tau_{jj} < \infty] < 1$

We will later see conditions to determine this classification

Classification of States (topological)

The states of an HMC can also be classified by by topological properties of the transition diagram $G(V,E)$ (ie, of the unweighted graph)

- Recall $(i,j) \in E$ iff $P_{ij} > 0$. State j is said to be accessible from state i if $\exists$ directed path $i \longrightarrow j$ (in probabilistic terms, j is accessible from i iff $\mathbb{P}[\tau_j < \infty | X_0 = i] < \infty$, i.e., $\exists M > 0$ s.t. $P_{ij}(M) = (P^M)_{ij} > 0$)
- States i and j communicate if j is accessible from i, and i is accessible from j. This is denoted as $i \longleftrightarrow j$, and is an equivalence relation (ie., $i \leftrightarrow i$, $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$, and $i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k$), and it partitions $\mathcal{X}$ into disjoint equivalence classes called communicating classes
- In terms of the transition diagram, a communicating class $\Longleftrightarrow$ a strongly connected component of G

- A set $C \subseteq \mathcal{X}$ is said to be
  - closed if $\sum_{j \in C} P_{ij} = 1 \; \forall i \in C$
  - irreducible if $i \longleftrightarrow j \; \forall \, i,j \in C$ (ie., $i,j \in$ a comm^g class)
- The period of a state $i \in \mathcal{X}$ is defined as $\gcd\{n \,|\, p_{ii}(n) > 0\}$. State i is said to be aperiodic if it has period 1.

Thm (Class properties) $\forall \, i,j \in \mathcal{X}$ s.t. $i \longleftrightarrow j$

i) i and j have the same period

ii) i is transient iff j is transient

iii) i is null recurrent iff j is null recurrent

iv) i is Positive recurrent iff j is positive recurrent

Thm (Decomposition) For any MC, X can be partitioned uniquely as

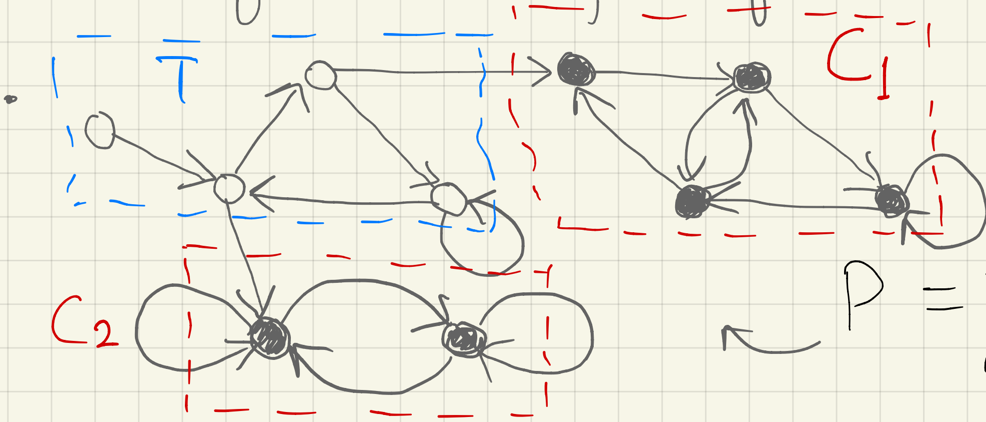
$$X = T \cup C_1 \cup C_2 \cup \dots$$

where T is the set of transient states, and C_i are irreducible, closed sets

- Every finite MC has at least one $C \equiv$ irreducible closed set

---

Pictorially we have the following



$$P = \begin{array}{c|ccc} & T & C_1 & C_2 \\ \hline T & P_{TT} & P_{TC_1} & P_{TC_2} \\ C_1 & 0 & P_{C_1} & 0 \\ C_2 & 0 & 0 & P_{C_2} \end{array}$$

- Any finite MC starting in T eventually hits some C, and then stays there
  - We will now concentrate on understanding a singe class C.

Thm - Let P be the transition matrix of an irreducible Markov chain (i.e., X has a single communicating class) with period d then $\forall\, i,j \in X$, $\exists\, m \geq 0$ and $n_0 \geq 0$ (possibly depending on i,j) s.t.

$$P_{ij}(m+nd) > 0 \quad \forall\, n \geq n_0$$

- In other words, for an irreducible MC, the matrix P^{n_0} eventually has all non-zero elements. Does it however converge?

Stationary Distribution of an HMC

- A vector π is said to be a stationary distribution of an HMC if $\pi(j) \geq 0 \;\forall j \in \mathcal{X}$, $\sum_{j \in \mathcal{X}} \pi(j) = 1$ and

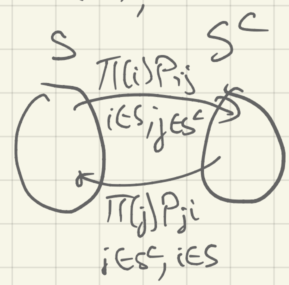
$$\pi^T = \pi^T P$$

- (Global Balance) Alternately, π can be defined by the eqns

$$\pi(i) = \sum_{j \in \mathcal{X}} \pi(j) P_{ji}$$

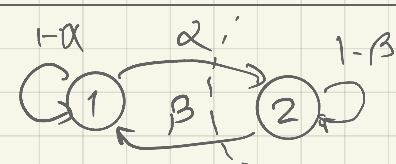
More generally, for any set $S \subseteq \mathcal{X}$ (and $S^c = \mathcal{X} \setminus S$), we have

$$\sum_{i \in S} \sum_{j \in S^c} \pi(i) P_{ij} = \sum_{j \in S^c} \sum_{i \in S} \pi(j) P_{ji}$$



- If $\pi_t = \pi \Rightarrow \pi_{t+s} = \pi \;\forall\; s \geq 0$

Eg - $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \Rightarrow \pi = \left(\frac{\beta}{\alpha+\beta} \quad \frac{\alpha}{\alpha+\beta} \right)^T$



Eg - For any MC P, its Lazy Markov chain is the one where at each step, we do nothing with prob α, else run P. Denoting its transition prob matrix as Q, we have

$$Q = \alpha I + (1-\alpha) P$$

- Let π be a stationary dist of P. Then $\pi^T Q = \pi^T$

Thus a lazy chain has the same stationary dist for any α.

- For any indexed collection of r.v.s $(X_t ; t \in \mathbb{N})$, a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ is a collection of σ-fields s.t. or $\mathbb{R}$
  $\mathcal{F}_t = \sigma(X_{t'} ; t' \leq t)$. In other words, $\mathcal{F}_t$ is made up of all the events of the form $\{X_{t'} \leq a, t' \leq t\}$.
- An event A is said to be adapted to $\mathcal{F}_t$ if $\exists$ a function ϕ s.t. $\mathbb{1}_A(\omega) = \phi(X_{t'}(\omega) ; t' \leq t)$
- For any $(X_t ; t \in \mathbb{N})$ with associated filtration $\mathcal{F}_t$, a stopping time τ is a $\mathbb{N}$-valued r.v for which $(\tau \leq t)$ is adapted to $\mathcal{F}_t$ $\forall t$
  - i.e., τ is a non-anticipative random time

Eg - First visit to x is a stopping time
Last visit to x is not a stopping time

- $\boxed{\text{Thm}}$ (Strong Markov Property) For any HMC with transition matrix P, and any stopping time P
  i) Given $X_\tau = i$, Process before and after τ are independent
  ii) Given $X_\tau = i$, process after τ is an HMC with $\hat{X}_0 = i$, transition matrix P

$\boxed{\text{Thm}}$ (Existence and Uniqueness of π for irreducible chains)

If $\mathcal{X}$ comprises of a single $\boxed{\text{irreducible, positive recurrent}}$ class then there the equation $x^T P = x^T$ has a unique positive soln upto multiplicative constants. Moreover, the unique stationary distr obeys $\pi(x) = \dfrac{1}{\mathbb{E}[\tau_{xx}]}$

Pf - We will show this by constructing a 'soln' $\tilde{\pi}$

- Consider any $z \in \mathcal{X}$. Define $\mathbb{E}_z[\cdot] \triangleq \mathbb{E}[\cdot \mid X_0 = z]$

  Let $\tilde{\pi}(y) = \mathbb{E}_z[\#$ of visits to y before returning to $z]$

$$= \lim_{T\to\infty} \mathbb{E}_z\Big[\sum_{t=0}^{T} \mathbb{1}\{X_t = y, \underbrace{\tau_{zz}}_{\inf\{t\geq 1 \mid X_t = z, X_0 = z\}} > t\}\Big]$$

$$= \sum_{t=0}^{\infty} \mathbb{P}_z[X_t = y, \tau_{zz} > t]$$

- Since chain is positive recurrent, we have $\mathbb{E}[\tau_{zz}] < \infty \ \forall z$

$$\Rightarrow \tilde{\pi}(y) \leq \sum_{t=0}^{\infty} \mathbb{P}_z[\tau_{zz} > t] = \mathbb{E}[\tau_{zz}] < \infty$$

- Now to check $\tilde{\pi}$ is a stationary dist, consider

$$\sum_{x\in\mathcal{X}} \tilde{\pi}(x) P_{xy} = \sum_{x\in\mathcal{X}} \Big[\sum_{t=0}^{\infty} \mathbb{P}_z[X_t = x, \tau_{zz} \geq t+1]\Big] P_{xy} \qquad (*)$$

for some $y \in \mathcal{X}$

- Let $\mathcal{F}_t = \sigma(X_0, X_1, \dots, X_t)$. We have
$$\{\tau_{zz} \geq t+1\} = \{\tau_{zz} > t\} \in \mathcal{F}_t$$
$$\Rightarrow \mathbb{P}_z[X_t = x, X_{t+1} = y, \tau_{zz} \geq t+1] = \mathbb{P}_z[X_t = x, \tau_{zz} \geq t+1] P_{xy}$$

- By Tonelli's thm, we can interchange $\sum$ in ⊛

$$\Rightarrow \sum_{x \in X} \tilde{\pi}(x) P_{xy} = \sum_{t=0}^{\infty} \sum_{x \in X} \mathbb{P}_z[X_t = x, X_{t+1} = y, \tau_{zz} \geq t+1]$$
$$= \sum_{t=1}^{\infty} \mathbb{P}_z[X_t = y, \tau_{zz} \geq t] \quad \text{(By Markov property)}$$
$$= \tilde{\pi}(y) - \underbrace{\mathbb{P}_z[X_0 = y, \tau_{zz} > 0]}_{\delta_1} + \underbrace{\sum_{t=1}^{\infty} \mathbb{P}_z[X_t = y, \tau_{zz} = t]}_{\delta_2}$$

Now if $y \neq z$, then $X_0 = X_{\tau_{zz}} = t$ and $\delta_1 = \delta_2 = 0$. If $y = z$, then $X_0 = X_{\tau_{zz}} = z \Rightarrow \delta_1 = \delta_2 = 1$

Thus we have $\sum_{x \in X} \tilde{\pi}(x) P_{xy} = \tilde{\pi}(y) \quad \forall y \in X$

- Finally, to make $\tilde{\pi}$ a probability measure, we can set $\pi(x) = \dfrac{\tilde{\pi}(x)}{\mathbb{E}[\tau_{zz}]}$. In particular, we have $\pi(x) = 1/\mathbb{E}[\tau_{xx}] > 0$ since $\mathbb{E}[\tau_{xx}] < \infty \; \forall x$

- Now we want to show that $\pi(x) = 1/\mathbb{E}[\tau_{xx}]$ is unique

For this, let $\hat{\pi}$ be another stationary dist. We know that if $X_0 \sim \hat{\pi}$, then $X_t \sim \hat{\pi}$ $\forall\, t \geq 0$

- Now suppose $X_0 \sim \hat{\pi}$. For any $x \in \mathcal{X}$, we have

$$\hat{\pi}(x)\,\mathbb{E}[\tau_{xx}] = \mathbb{P}[X_0 = x] \sum_{t=1}^{\infty} \mathbb{P}[\tau_{xx} \geq t]$$

$$= \sum_{t=1}^{\infty} \mathbb{P}[T_x(2) \geq t \mid X_0 = x]\,\mathbb{P}[X_0 = x]$$

second visit to x

$$= \sum_{t=1}^{\infty} \mathbb{P}[T_x(2) \geq t,\, X_0 = x]$$

- Define $a_n = \mathbb{P}[X_t \neq x \text{ for } 0 \leq t \leq n]$, $a_0 = \mathbb{P}[X_0 \neq x]$
  - Note that $\{X_t \neq x \text{ for } 0 \leq t \leq n\} \subseteq \{X_t \neq x \text{ for } 0 \leq t \leq n-1\}$

    $\Rightarrow a_n \leq a_{n-1} \leq a_{n-2} \leq \dots$
  - Moreover if $X_t \sim \hat{\pi}$ $\forall t$, then we also have

    $\mathbb{P}[X_t \neq x \text{ for } 0 \leq t \leq n] = \mathbb{P}[X_t \neq x \text{ for } 1 \leq t \leq n+1]$

- Now consider $b_n = \mathbb{P}[T_x(2) \geq n, X_0 = x]$, $b_1 = \mathbb{P}[T_x(2) \geq 1, X_0 = x] = \mathbb{P}[X_0 = x]$

Then we have $\hat{\pi}(x)\,\mathbb{E}[\tau_{xx}] = \sum_{n=1}^{\infty} b_n = \mathbb{P}[X_0 = x] + \sum_{n=2}^{\infty} b_n$

Moreover $b_n = \mathbb{P}[X_t \neq x \ \forall 1 \leq t \leq n-1, X_0 = x] \quad \forall n \geq 2$

$$= \mathbb{P}[X_t \neq x \ \forall 1 \leq t \leq n-1] - \mathbb{P}[X_t \neq x \ \forall 0 \leq t \leq n-1]$$

$$= \mathbb{P}[X_t \neq x \ \forall 0 \leq t \leq n-2] - \mathbb{P}[X_t \neq x \ \forall 0 \leq t \leq n-1]$$

$$= a_{n-2} - a_{n-1}$$

where the last line uses that $X_t \sim \hat{\pi} \ \forall t$

- Thus $\hat{\pi}(x) \mathbb{E}[\tau_{xx}] = \mathbb{P}[X_0 = x] + \sum_{n=2}^{\infty} (a_{n-2} - a_{n-1})$

$$= \mathbb{P}[X_0 = x] + \mathbb{P}[X_0 \neq x] - \lim_{n \to \infty} a_n$$

Also $\lim_{n \to \infty} a_n = \lim_{n \to \infty} \mathbb{P}[X_t \neq x \ \forall 0 \leq t \leq n] = 1 - \mathbb{P}[\tau_{xx} < \infty] = 0$

as the MC is positive recurrent $\forall x \in \mathcal{X}$

$\Rightarrow \hat{\pi}(x) \mathbb{E}[\tau_{xx}] = 1 \quad \forall x \in \mathcal{X}, \ \hat{\pi}$ stationary

Thus $\pi(x) = 1 / \mathbb{E}[\tau_{xx}]$ is the unique stationary dist

$\blacksquare$

Thus, for an irreducible, positive recurrant MC, we have that $\pi^T P = \pi^T$ has a unique solution s.t. $\pi(x) > 0 \ \forall x \in \mathcal{X}$, and $\sum_{x \in \mathcal{X}} \pi(x) = 1$. Moreover π satisfies $\pi(x) = 1 / \mathbb{E}[\tau_{xx}]$

Some useful facts + roadmap

i) How do we check if a MC is positive recurrent?
(irreducibility is easier to check)

- Directly check $\mathbb{E}[\tau_{x,x}] < \infty$ for some $x \in \mathcal{X}$
- Finite-state, irreducible chains (via Perron-Frobenius Thm)
- Foster-Lyapunov criterion - 'Potential fn argument'

ii) What does π look like? When is it easy to compute?

Eg (Doubly Stochastic Matrix) If P is $n \times n$, irreducible, and $\sum_{x \in \mathcal{X}} P_{xy} = 1$ (i.e., each column sum is 1), then $\pi = \left(\frac{1}{n}\ \frac{1}{n} \cdots \frac{1}{n}\right)^T$

Pf - Check $\pi^T P = \pi^T$. By uniqueness of π, we are done!

- A more useful condition - reversibility

iii) When does $\pi_n \longrightarrow \pi$ for any starting state π_0

- Convergence thm

iv) What can we say about time-averages of functions of an MC? - MC Ergodic thm

v) How fast is this convergence? How can we quantify it in terms of the MC properties?

- Mixing times of MCs

- # Finite MC and Perron-Fröbenius

- Finding π for an MC involves solving $\pi^T P = \pi^T$. Now for $\mathcal{X}$ finite (so say $P = n \times n$), this is now essentially same as computing a left eigenvector with eigenvalue 1. Our previous thm says this always exists and is unique if MC is irreducible and positive recurrent. We next see this specialized to finite P

- First, we note that existence and uniqueness of π does not imply convergence.

Eg - Let $\mathcal{X} = \{1,2\}$ and $P_{12} = P_{21} = 1$. Let $\pi_0 = \begin{pmatrix}1\\0\end{pmatrix}$

$\Rightarrow$ $\pi_t = \begin{pmatrix}1\\0\end{pmatrix}$ if t is even, and $\pi_t = \begin{pmatrix}0\\1\end{pmatrix}$ if t is odd.

Clearly $\pi_t \not\to \pi$ (even though $\pi_t^T = P^t \pi_0^T$, and π is unique)

- The problem in the example is that the MC is periodic. Its easy to see that this will always lead to non convergence. What if MC is aperiodic?

- Defn - A non-negative square matrix A is said to be primitive iff $\exists k$ s.t $A^k > 0$.

- P primitive $\iff$ P is irreducible and aperiodic

- For any matrix A, its characteristic polynomial $f_A(\lambda)$ is defined as $f_A(\lambda) = \det(A - \lambda I)$.
  - The eigenvalues $(\lambda_1, \lambda_2, \dots, \lambda_n)$ of A are the roots (possibly complex) of $f_A(\lambda)$
  - For any e-value λ_i of A
    - Its algebraic multiplicity $\mu_A(\lambda_i)$ is defined as $\mu_A(\lambda_i) =$ largest integer k s.t $(\lambda - \lambda_i)^k$ divides $f_A(\lambda)$
    - Its right e-vectors $E_i^R = \{v \mid (A - \lambda_i I)v = 0\}$
    - Its left e-vectors $E_i^L = \{v \mid v(A - \lambda_i I) = 0\}$
    - Its geometric multiplicity $\gamma_A(\lambda_i) \triangleq$ dimension of E_i^R (i.e. # of linearly independent right e-vectors)
    - $1 \leq \gamma_A(\lambda_i) \leq \mu_A(\lambda_i) \leq n$

$\boxed{\text{Thm}}$ (Perron-Frobenius) Let A be a non-negative primitive $n \times n$ matrix. Then $\exists$ real e-value λ_1 s.t.

i) $\lambda_1 \in \mathbb{R}$

ii) $\mu_A(\lambda_1) = \gamma_A(\lambda_1) = 1$

iii) $\lambda_1 > 0$ and $\lambda_1 > |\lambda_j|$ $\forall$ e-values j

iv) $\exists$ left and right e-vectors corresponding to λ_1 s.t $u_1^T v_1 = 1$

- Corollary – If P is the transition matrix of an irreducible MC

i) $\lambda_1 = 1$, $\underbrace{|\lambda_2|}_{} \triangleq \max_{j\neq 1}\{|\lambda_j|\} \leq 1$

SLEM = second largest e-value modulus

ii) If P is aperiodic (ie, primitive), then $|\lambda|_2 < 1$

(If P has period d, then $\lambda_1 = \omega^0, \lambda_2 = \omega^1, \ldots, \lambda_d = \omega^{d-1}$, where $\omega = e^{2\pi i/d}$ are the complex roots of 1)

iii) We can choose $v_1 = \mathbb{1}$, $u_1 = \pi$ and hence

$$P^t = \mathbb{1}\pi^T + O\left(t^{m_2-1}|\lambda_2|^t\right)$$

where $m_2 = \mu_A(\lambda_2)$

- Thus $\pi_0^T P^t = \pi^T + \underbrace{\sum_{j=2}^n \lambda_j^t \pi_0^T (v_j u_j^T)}_{=O\left(t^{m_2-1}|\lambda_2|^t\right)}$

Eg- $\mathcal{X} = \{1,2\}$, $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$

$\Rightarrow f_P(\lambda) = (1-\alpha-\lambda)(1-\beta-\lambda) - \alpha\beta$, $\lambda_1 = 1$, $\lambda_2 = 1-\alpha-\beta$

Also $\pi = \frac{1}{\alpha+\beta}(\beta \;\; \alpha)^T$, and we have

$$P^n = \underbrace{\frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}}_{\mathbb{1}^T\pi} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\underbrace{\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}}_{v_2^T u_2}$$

Reversibility & Detailed Balance

- Given MC P with stationary dist π, define new matrix Q as $\pi(i) q_{ij} = \pi(j) p_{ji} \ \forall i,j \in X$

Claim - Q is a stochastic matrix and $\pi^T Q = \pi^T$

Pf - $q_{ij} = \frac{\pi(j) p_{ji}}{\pi(i)} \geq 0 \ \forall ij$

Also $\sum_{j \in X} q_{ij} = \frac{1}{\pi(i)} \sum_{j \in X} \pi(j) p_{ji} = \frac{\pi(i)}{\pi(i)} = 1$

Finally $(\pi^T Q)_j = \sum_{i \in X} \pi(i) \cdot q_{ij} = \sum_{i \in X} \pi(j) p_{ji} = \pi(j)$

$\Rightarrow \pi^T Q = \pi^T$

- Q is the distribution of the 'time-reversed' chain. In particular, an MC P is said to be reversible iff $Q = P$.

- The equations $\pi(i) p_{ij} = \pi(j) q_{ji} \ \forall i,j$ are called the detailed balance equations. They are particularly useful as they give a surprising way to compute π!

$\boxed{\text{Thm}}$ (Kelly's Lemma) Let P be a stochastic matrix on $\mathcal{X}$. Suppose we are given π distrib on $\mathcal{X}$, and matrix Q s.t.

i) Q is stochastic, i.e., $\sum_{j \in \mathcal{X}} q_{ij} = 1$

ii) Detailed balance holds, i.e., $\pi(i) q_{ij} = \pi(j) p_{ji} \ \forall i, j$

Then π is a stationary matrix of P

Pf - For any $i \in \mathcal{X}$ we have

$$\sum_{j \in \mathcal{X}} \pi(j) p_{ji} = \sum_{j \in \mathcal{X}} \pi(i) q_{ij}$$

$$= \pi(i) \sum_{j \in \mathcal{X}} q_{ij} = \pi(i)$$

Thus π satisfies global balance $\Rightarrow \pi^T P = \pi$

Corollary - For any MC P, if $\exists$ distribution π s.t.

$$\pi(i) p_{ij} = \pi(j) p_{ji} \quad \forall i, j$$

Then P is reversible and π is a stationary distribution of P

The Markov Chain Ergodic Theorem

- We now want to look at 'long-run averages' along sample paths of a MC, ie., $\frac{1}{T}\sum_{t=1}^{T} g(X_t)$.
  - If X_t were iid, this is equal to $\mathbb{E}[g(X_1)]$. Can we do something similar for MCs? The ergodic thm asserts that if the MC is irreducible and positive recurrent, then in the limit $T \uparrow \infty$, we can equate the long-run time average with $\mathbb{E}_\pi[g(X)]$, the space average under the stationary distribution

---

Proposition (Convergence of Canonical Measures) Let $(X_n, n \in \mathbb{N})$ be an irreducible recurrent (could be null) HMC, and let for any state $z \in \mathcal{X}$, define the canonical measure η_z as

$$\eta_z(x) = \mathbb{E}_z\left[\sum_{t \geq 1} \mathbb{1}\{X_t = x\}\, \mathbb{1}\{t \leq T_z(2)\}\right] \quad \forall x \in \mathcal{X}$$

where $T_z(2)$ is the second visit time to z. For any $t \geq 0$, define $\nu_z(t) = \sum_{k=0}^{t} \mathbb{1}\{X_k = z\}$, and consider any fn f s.t. $\sum_{x \in \mathcal{X}} |f(x)|\eta_z(x) < \infty$. Then for any starting distr Π_0

$$\lim_{T \uparrow \infty} \frac{1}{\nu_z(T)} \sum_{t=1}^{T} f(X_t) = \sum_{x \in \mathcal{X}} f(x)\, \eta_z(x) \quad \text{a.s}$$

Pf of Prop: Let $T_z(1), T_z(2), \ldots$ be the successive returns to state z, and define $U_k = \sum_{t=T_z(k)+1}^{T_z(k+1)} f(X_t)$. By the strong Markov property, $\{U_k\}$ is an iid sequence.

- Now if $f \geq 0$, we have (by strong Markov)

$$\mathbb{E}[U_k] = \mathbb{E}_z\left[\sum_{t=1}^{T_z(1)} f(X_t)\right]$$

$$= \mathbb{E}_z\left[\sum_{t=1}^{T_z(1)} \sum_{x \in \mathcal{X}} f(x) \mathbb{1}_{\{X_t = x\}}\right]$$

$$= \sum_{x \in \mathcal{X}} f(x) \mathbb{E}_z\left[\sum_{t=1}^{T_z(1)} \mathbb{1}_{\{X_t = x\}}\right] \quad \text{(Tonelli's Thm, Since } f \geq 0\text{)}$$

$$= \sum_{x \in \mathcal{X}} f(x) n_z(x) < \infty \quad \text{by assumption}$$

- By the SLLN, we have $\lim_{N \uparrow \infty} \frac{1}{N} \sum_{k=1}^{N} U_k = \sum_{x \in \mathcal{X}} f(x) n_z(x)$ a.s.

$$\Rightarrow \lim_{N \uparrow \infty} \frac{1}{N} \sum_{t=T_z(1)+1}^{T_z(N+1)} f(X_t) = \sum_{x \in \mathcal{X}} f(x) n_z(x) \quad \text{a.s.}$$

- Now since $T_z(\nu_z(T)) \leq T < T_z(\nu_z(T)+1)$, we have

$$\frac{\sum_{t=1}^{T_z(\nu_z(T))} f(X_t)}{\nu_z(T)} \leq \frac{\sum_{t=1}^{T} f(X_t)}{\nu_z(T)} \leq \frac{\sum_{t=1}^{T_z(\nu_z(T)+1)} f(X_t)}{\nu_z(T)}$$

Since chain is recurrent, $\lim_{T \uparrow \infty} \nu(T) = \infty$ and thus all three terms above converge to $\sum_{x \in \mathcal{X}} f(x) n_z(x)$ as $T \uparrow \infty$.

- For general f, write $f = f^+ - f^-$, where $f^+ = \max(0, f)$, $f^- = \max(0, -f)$. Since $\sum |f(x)| n_z(x) < \infty \Rightarrow$ each term is well defined. ▨

Thm (Markov chain Ergodic Thm) Let $(X_n; n \in \mathbb{N})$ be an irreducible, positive recurrent Markov chain with stationary distribution π. For any $f: \mathcal{X} \to \mathbb{R}$ s.t. $\sum_{x\in\mathcal{X}} |f(x)| \pi(x) < \infty$, and any initial distr $X_0 \sim \pi_0$

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} f(X_t) = \sum_{x\in\mathcal{X}} f(x)\pi(x) \quad \text{a.s.}$$

Pf - Apply the convergence result for canonical measures to $f(x) = 1$. Since MC is positive recurrent, we have $\sum_{x\in\mathcal{X}} f(x) n_z(x) = \sum_{x\in\mathcal{X}} n_z(x) = \mathbb{E}[\tau_{zz}] < \infty$.

Thus $\lim_{T\uparrow\infty} \frac{1}{\nu_z(T)} \sum_{t=1}^{T} f(X_t) = \lim_{T\uparrow\infty} \frac{T}{\nu_z(T)} = \sum_{x\in\mathcal{X}} n_z(x)$

Now for any f, if $\sum_{x\in\mathcal{X}} |f(x)| \pi(x) < \infty \Rightarrow \sum_{x\in\mathcal{X}} |f(x)| n_z(x) < \infty$ as well, since $\pi(x) \propto n_z(x)$ for any z. Thus we have

$$\lim_{T\uparrow\infty} \frac{\sum_{t=1}^{T} f(X_t)}{T} = \lim_{T\uparrow\infty} \left(\frac{\nu_z(T)}{T}\right)\left(\frac{\sum_{t=1}^{T} f(X_t)}{\nu_z(T)}\right)$$

$$= \frac{\sum_{x\in\mathcal{X}} f(x) n_z(x)}{\sum_{x\in\mathcal{X}} n_z(x)}$$

From before, we know that for a positive recurrent, irreducible MC, we have $\frac{n_z(x)}{\sum_{x\in\mathcal{X}} n_z(x)} = \pi(x) \; \forall x, z$. This completes the proof. ∎

Testing for Positive Recurrence - Lyapunov Functions

- We finally present a way to test for positive recurrence. The main idea is to map all states to a 1-dimensional potential function, which we can then analyze as a birth-death chain.

Thm (Foster-Lyapunov Condition) Given irreducible MC P on countable state-space $\mathcal{X}$, suppose $\exists$ function $h : \mathcal{X} \to \mathbb{R}$ s.t.

i) $h(i) \geq 0 \;\forall\, i \in \mathcal{X}$ ← Lyapunov Function

ii) $\sum_{k\in\mathcal{X}} P(i,k)\,h(k) < \infty \;\forall\, i \in \mathcal{X}$ ← $E[h(X_{n+1}) \mid X_n = i] < \infty \;\forall i$

iii) For some $\varepsilon > 0$ and finite set F, we have
$$\sum_{k\in\mathcal{X}} P(i,k)\,h(k) < h(i) - \varepsilon \;\forall\, i \in \mathcal{X}\setminus F$$
← $E[h(X_{n+1}) \mid X_n = i] < h(i) - \varepsilon \;\forall\, i \notin F$

Then the MC is positive recurrent.

Pf - Let τ = return time to F, $Y_t = h(X_t)\,\mathbb{1}_{\{t<\tau\}}$

- By prop (iii), we have $E[h(X_{t+1}) \mid X_t = i] \leq h(i) - \varepsilon \;\forall\, i \notin F$

prop (ii) implies $E[h(X_{t+1}) \mid X_t = i] < \infty \;\forall i \in \mathcal{X}$

$\Rightarrow \forall\, x \notin F$, we have

$$\begin{aligned}
E_x[Y_{t+1} \mid X_0^t] &= E_x[Y_{t+1}\mathbb{1}_{\{t<\tau\}} \mid \mathcal{F}_t] + \underbrace{E_x[Y_{t+1}\mathbb{1}_{\{t>\tau\}} \mid \mathcal{F}_t]}_{\geq 0} \\
&\leq E_x[h(X_{t+1})\mathbb{1}_{\{t<\tau\}} \mid \mathcal{F}_t] \quad \leftarrow \sigma(X_0, X_1, \dots, X_t) \\
&= \mathbb{1}_{\{t<\tau\}}\, E_x[h(X_{t+1}) \mid \mathcal{F}_t] \\
&\leq \mathbb{1}_{\{t<\tau\}}\, h(X_t) - \varepsilon\,\mathbb{1}_{\{t<\tau\}}
\end{aligned}$$

where the last $\leq$ follows from the fact that $X_t \notin F$ if $t < \tau$

Thus we have $\mathbb{E}_x[Y_{t+1}] \leq \mathbb{E}_x[Y_t] - \varepsilon \mathbb{P}_x[\tau > t]$

- Now since Y_t is non-negative, we iterate to get

$$0 \leq \mathbb{E}_x[Y_{t+1}] \leq \mathbb{E}_x[Y_0] - \varepsilon \sum_{k=0}^{t} \mathbb{P}_x[\tau > k]$$

Also $Y_0 = h(x)$ since $x \notin F$, and $\sum_{k=0}^{\infty} \mathbb{P}_x[\tau > k] = \mathbb{E}_x[\tau]$

$\Rightarrow \mathbb{E}_x[\tau] \leq \varepsilon^{-1} h(x)$

- For $y \in F$, we have $\mathbb{E}_y[\tau] = 1 + \sum_{x \notin F} P(y,x) \mathbb{E}_x[\tau]$

$\Rightarrow \mathbb{E}_y[\tau] \leq 1 + \varepsilon^{-1} \sum_{x \notin F} P(y,x) h(x) < \infty$ by (iii)

- Thus return time to F starting anywhere in F has finite expectation.

Now let $\tau_1, \tau_2, \tau_3, \ldots$ be the return times to F. By the strong Markov property, $Z_1 = X_{\tau_1}, Z_2 = X_{\tau_2}, \ldots$ form a HMC on state space F. Now X_t irreducible means Z_t is also irreducible, and since F is finite $\Rightarrow Z_t$ is positive recurrent, with $\mathbb{E}[\tilde{\tau}_{xx}] < \infty \; \forall x \in F$ under Z_t MC

- In the original MC, $\mathbb{E}[\tau_{xx}] = \mathbb{E}_x\left[\sum_{k=0}^{\infty} S_k \mathbb{1}\{\tilde{\tau}_{xx} > k\}\right]$, where $S_k = \tau_{k+1} - \tau_k \;\forall k \geq 1$.

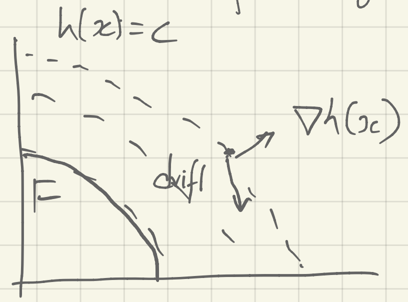
Since F is finite, $\mathbb{E}[S_k \mid X_{\tau_k} = \ell] = \mathbb{E}_\ell[\tau] \leq \left(\max_{\ell \in F} \mathbb{E}_\ell[\tau]\right)$

$$\Rightarrow \mathbb{E}[\tau_{xx}] = \sum_{k=0}^{\infty} \sum_{\ell \in F} \mathbb{E}_x[S_k \mid X_{\tau_k} = \ell] \, \mathbb{E}_x\left[\mathbb{1}\{X_{\tau_k} = \ell\} \mathbb{1}\{\tilde{\tau}_{xx} > k\}\right]$$

$$\leq \left(\max_{\ell \in F} \mathbb{E}_\ell[\tau]\right) \sum_{k=0}^{\infty} \mathbb{P}_x[\tilde{\tau}_{xx} > k] < \infty$$

∎

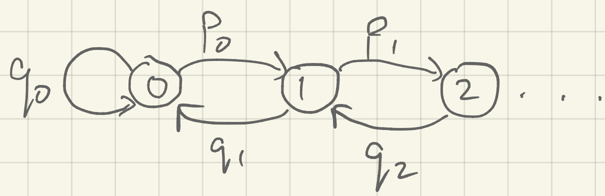
Intuition for designing h



Suppose $h: \mathbb{R}^n_+ \to \mathbb{R}_+$ is differentiable

$$\mathbb{E}[\Delta h(X_t)] = \mathbb{E}\left[h(X_{t+1}) - h(X_t) \mid X_t = x\right]$$
$$= \mathbb{E}\left[(X_{t+1} - X_t)^T \nabla h(X_t) \mid X_t = x\right]$$
$$= \underbrace{\mathbb{E}\left[(X_{t+1} - X_t) \mid X_t = x\right]^T}_{\text{drift of } X_t} \nabla h(x) < -\varepsilon$$

Eg (Birth-death chain)



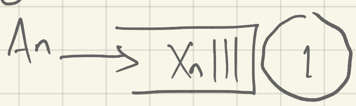
$p_i + q_i = 1 \ \forall i \in \mathbb{N}$

- Let $h(x) = x$

$$\Rightarrow \mathbb{E}\left[h(X_{n+1}) \mid X_n = x\right] = p_x \cdot (x+1) + q_x(x-1)$$
$$= h(x) + p_x - q_x < \infty \quad \forall x \in \mathcal{X}$$

- Now suppose $p_x - q_x < -\varepsilon$ for all except finite x, then by Foster-Lyapunov, we have that the MC is positive recurrent.

Eg (Discrete-time queue) $\quad X_{n+1} = (X_n - 1)^+ + A_n$



- If A_n is iid $\Rightarrow$ it is a MC. Also it is irreducible under mild conditions on A_n

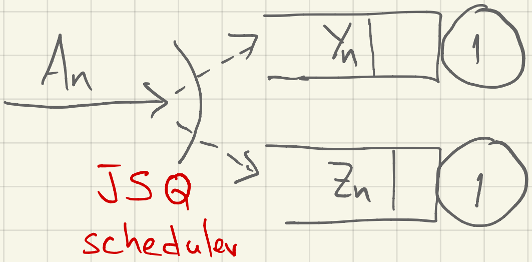
- Let $h(x) = x$
- $\mathbb{E}\left[h(X_{n+1}) \mid X_n = x\right] = (x-1)^+ + \mathbb{E}[A_n]$

$$= \begin{cases} h(x) - 1 + \mathbb{E}[A_n] ; & \forall x \geq 1 \\ \mathbb{E}[A_n] ; & x = 0 \end{cases}$$

Clearly this is finite if $\mathbb{E}[A_n] < \infty$.

Moreover, if $\mathbb{E}[A_n] - 1 < -\varepsilon$ (ie, $\mathbb{E}[A_n] < 1 - \varepsilon$), then we can use Foster-Lyapunov to say that MC is positive recurrent.

Eg (Join-the-shortest queue) - Switch routing in 2 server system

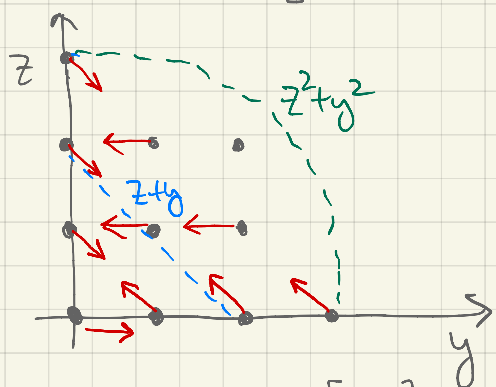


$X_n = (Y_n, Z_n)^T$

lexicographic tie breaking

$$X_{n+1} = \begin{pmatrix} Y_n + A_n \mathbb{1}\{Y_n \le Z_n\} - \mathbb{1}\{Y_n > 0\} \\ Z_n + A_n \mathbb{1}\{Y_n < Z_n\} - \mathbb{1}\{Z_n > 0\} \end{pmatrix}$$

- Intuitively, we need $\mathbb{E}[A_n] < 2$. Is this sufficient
- Let $\mathbb{E}[A_n] = \lambda = 2 - \varepsilon$, $\mathrm{Var}(A_n) = \sigma^2$



Now using $h(y,z) = y+z$ can not work (As $\mathbb{E}[\text{drift}]$ at boundary does not point inwards)

· Let $h(y,z) = y^2 + z^2$

Define $\Delta h(y,z) = \mathbb{E}\left[Y_{n+1}^2 + Z_{n+1}^2 - (Y_n^2 + Z_n^2) \mid (Y_n, Z_n) = (y,z)\right]$. When is $\Delta h(y,z) < -\delta$?

(i) $y \geq z > 0$

$$\begin{aligned} \Delta h(y,z) &= (y-1)^2 - y^2 + \mathbb{E}[(z-1+A_n)^2] - z^2 \\ &= -(2y-1) - (2z-1) + 2(z-1)\lambda + \sigma^2 \\ &= 2(z(1-\varepsilon) - y) + \sigma^2 - 2(1-\varepsilon) \leq -2y\varepsilon - 2(1-\varepsilon) + \sigma^2 \\ &< -\delta \quad \text{if } y > \left\lceil \frac{\sigma^2 + \delta - 2(1-\varepsilon)}{2\varepsilon} \right\rceil \leftarrow \alpha \end{aligned}$$

(ii) $y > z = 0$

$$\Delta h(y,z) = -(2y-1) + \sigma^2 < -\delta \quad \text{if } y > \left\lceil \frac{\delta + \sigma^2 + 1}{2} \right\rceil \leftarrow \beta$$

(iii) $z > y > 0$ (Symmetric to (i))

$$\Delta h(y,z) \leq -\delta \quad \text{if } z > \left\lceil \frac{\sigma^2 + \delta - 2(1-\varepsilon)}{2\varepsilon} \right\rceil$$

(iv) $z > y = 0$ (Symmetric to (ii))

$$\Delta h(y,z) \leq -\delta \quad \text{if } z > \left\lceil \frac{\delta + \sigma^2 + 1}{2} \right\rceil$$

Thus $\forall\ (y,z)$ s.t $y > \max(\alpha, \beta)$, $z > \max(\alpha, \beta)$, we have $\Delta h(y,z) < -\delta$