Recap: Stationary Distribution

- π is said to be a stationary distribution for the Markov chain with transition probabilities P if

$$\pi(y) = \sum_{x \in S} \pi(x) P(x, y), \ \ \forall y \in S$$

- When π is stationary distribution,
$\pi_0 = \pi \ \ \Rightarrow \ \ \pi_n = \pi, \ \forall n$
- If $\pi_n = \pi, \ \forall n$ then π is a stationary distribution
- For a finite chain: $P^T \pi = \pi$
- A stationary distribution always exists for a finite chain

Recap

- $N_n(y)$ – number of visits to y till n
- $G_n(x,y) = E_x[N_n(y)] = \sum_{m=1}^{n} P^m(x,y)$
  – expected number of visits to y till n
- $m_y = E_y[T_y]$ – mean return time to y

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{I_{[T_y<\infty]}}{m_y}, \quad w.p.1$$

$$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m(x,y) = \frac{\rho_{xy}}{m_y}$$

Recap: positive and null recurrent states

- y is positive recurrent if $m_y < \infty$
- y is null recurrent if $m_y = \infty$
- If x is positive recurrent and x leads to y, then y is positive recurrent
- In a closed irreducible set of recurrent states either all states are positive recurrent or all states are null recurrent
- A finite closed set has to have at least one positive recurrent state
- A finite chain cannot have null recurrent states

Recap: Existence of stationary distribution

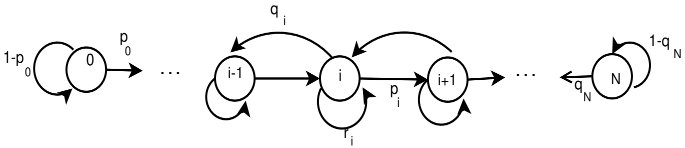
- In any stationary distribution π, $\pi(y) = 0$ if y is transient or null recurrent
- An irreducible transient or null recurrent chain does not have a stationary distribution
- An irreducible positive recurrent chain has a unique stationary distribution: $\pi(y) = \frac{1}{m_y}$
- An irreducible chain has a stationary distribution iff it is positive recurrent
- For a non-irreducible chain, for each closed irreducible set of positive recurrent states, there is a unique stationary distribution concentrated on that set.
- All stationary distributions of the chain are convex combinations of these

Recap: Periodic chains

- The period of a state x is $d_x = \gcd\{n \geq 1 \ : \ P^n(x,x) > 0\}$
- If x and y lead to each other, $d_x = d_y$
- In an irreducible chain, all states have the same period
- An irreducible chain is called aperiodic if the period is 1
- For an irreducible aperiodic positive recurrent chain, π_n converges to π, the unique stationary distribution, irrespective of what π_0 is.
- Also, for an irreducible, aperiodic, positive recurrent chain, $P^n(x,y)$ converges to $\frac{1}{m_y}$

Recall: Birth-Death chains - stationary distributions

- The following is a finite irreducible birth-death chain



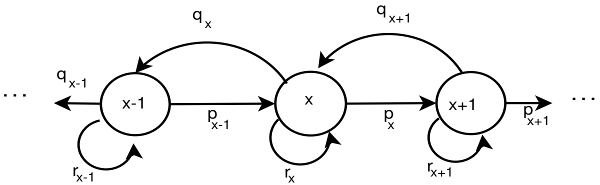
- The stationary distribution is given by

$$\pi(0) = \frac{1}{\sum_{j=0}^{N} \eta_j} \;\; \text{and} \;\; \pi(n) = \eta_n \, \pi(0), \; n = 1, \cdots, N$$

  where $\eta_n = \frac{p_0 p_1 \cdots p_{n-1}}{q_1 q_2 \cdots q_n}, \; n = 1, 2, \cdots, N$.
- This is applicable to an infinite chain also
  However, we need $\sum_{j=0}^{N} \eta_j < \infty$ for the stationary distribution to exist.

Recap: Birth-Death chains

- Consider a finite or infinite birth-death chain



- Define

$$U(y) = P_y[T_a < T_b], \ a < y < b, \quad U(a) = 1, \ U(b) = 0$$

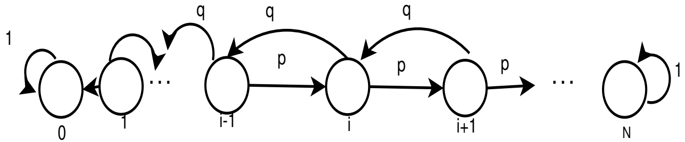
- Then,

$$U(y) = \frac{\sum_{x=y}^{b-1} \gamma_x}{\sum_{x=a}^{b-1} \gamma_x}, \quad \gamma_x = \frac{q_x q_{x-1} \cdots q_{a+1}}{p_x p_{x-1} \cdots p_{a+1}}$$

$$P_y[T_b < T_a] = \frac{\sum_{x=a}^{y-1} \gamma_x}{\sum_{x=a}^{b-1} \gamma_x}$$

- Suppose this is a Gambler's ruin chain:

$$p_x = p, q_x = q, \ \forall x$$

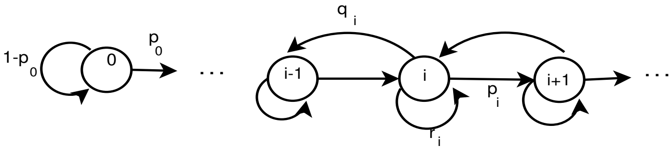


- Then, $\gamma_x = \left(\frac{q}{p}\right)^x$
- Hence, for a Gambler's ruin chain we get, e.g.,

$$P_i[T_N < T_0] = \frac{\sum_{x=0}^{i-1} \gamma_x}{\sum_{x=0}^{N-1} \gamma_x} = \frac{\left(\frac{q}{p}\right)^i - 1}{\left(\frac{q}{p}\right)^N - 1}$$

- This is the probability of gambler being successful

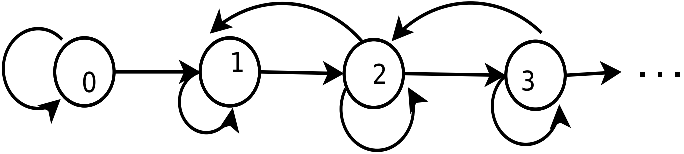
- Consider the following chain over $\{0, 1, \cdots\}$



- This is an infinite irreducible birth-death chain
- We want to know whether the chain is transient or recurrent etc.
- We can use the earlier analysis for this too.

$$\begin{aligned} P_1[T_0 < T_n] &= \frac{\sum_{x=1}^{n-1} \gamma_x}{\sum_{x=0}^{n-1} \gamma_x}, \ \forall n > 1 \\ &= \frac{\sum_{x=0}^{n-1} \gamma_x - \gamma_0}{\sum_{x=0}^{n-1} \gamma_x} = 1 - \frac{1}{\sum_{x=0}^{n-1} \gamma_x} \end{aligned}$$

- Consider this chain started in state 1.



$$[T_0 < T_n] \subset [T_0 < T_{n+1}], \ \ n = 2, 3, \cdots$$

since the chain cannot hit $n+1$ without hitting n.

- Also, $1 \leq T_2 < T_3 < \cdots < T_n$
- Hence $[T_0 < \infty]$ is same as $[T_0 < T_n, \ \text{ for some } n]$

- Consider this chain started in state 1.

$$[T_0 < T_n] \subset [T_0 < T_{n+1}], \;\; n = 2, 3, \cdots$$

since the chain cannot hit $n+1$ without hitting n.

- Also, $1 \leq T_2 < T_3 < \cdots < T_n$
- Hence $[T_0 < \infty]$ is same as $[T_0 < T_n, \;\text{for some } n]$

$$\begin{aligned} P_1[T_0 < T_n, \;\text{for some } n] &= P_1\left(\cup_{n>1}\, [T_0 < T_n]\right) \\ &= P_1\left(\lim_{n\to\infty} [T_0 < T_n]\right) \\ &= \lim_{n\to\infty} P_1\left([T_0 < T_n]\right) \\ &= \lim_{n\to\infty}\; 1 - \frac{1}{\sum_{x=0}^{n-1} \gamma_x} \\ \Rightarrow \;\; P_1[T_0 < \infty] &= 1 - \frac{1}{\sum_{x=0}^{\infty} \gamma_x} \end{aligned}$$

- **Theorem**: The chain is recurrent iff $\sum_{x=0}^{\infty} \gamma_x = \infty$

Proof

- Supoose chain is recurrent. Since it is irreducible,

$$P_1[T_0 < \infty] = 1 \;\Rightarrow\; \sum_{x=0}^{\infty} \gamma_x = \infty$$

- Suppose $\sum_{x=0}^{\infty} \gamma_x = \infty \Rightarrow \; P_1[T_0 < \infty] = 1$

$$\begin{aligned} P_0[T_0 < \infty] &= P(0,0) \; + \; P(0,1) \; P_1[T_0 < \infty] \\ &= P(0,0) + P(0,1) = 1 \end{aligned}$$

- Implies state 0 is recurrent and hence the chain is recurrent because it is irreducible.
- Note that we have used the fact that the chain is infinite only to the right.

- The chain is transient if $\sum_{x=0}^{\infty} \gamma_x < \infty$
- Let $p_x = p, q_x = q \Rightarrow \quad \gamma_x = \left(\frac{q}{p}\right)^x$

$$\text{Transient if } \sum_{x=0}^{\infty} \left(\frac{q}{p}\right)^x < \infty \quad \Leftrightarrow \quad q < p$$

$$\text{Recurrent if } \sum_{x=0}^{\infty} \left(\frac{q}{p}\right)^x = \infty \quad \Leftrightarrow \quad q \geq p$$

- Intuitively clear
- This chain with $q < p$ is an example of an irreducible chain that is wholly transient

- We know the chain is recurrent if $\sum_{x=0}^{\infty} \left(\frac{q}{p}\right)^x = \infty$
- When will this chain be positive recurrent?
- We know that an irreducible chain is positive recurrent if and only if it has a stationary distribution.
- We can check if it has a stationary distribution
- The equations that we derived earlier hold for this infinite case also.

- We derived earlier the equations that a stationary distribution of this chain (if it exists) has to satisfy

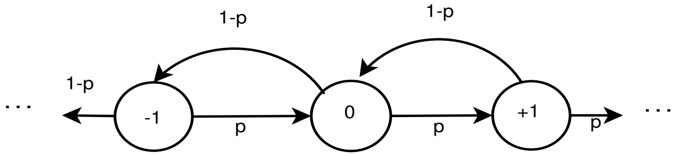
$$\pi(n) = \eta_n \, \pi(0), \quad \text{where} \quad \eta_n = \frac{p_0 p_1 \cdots p_{n-1}}{q_1 q_2 \cdots q_n}, \; n = 1, 2, \cdots ,$$

- Setting $\eta_0 = 1$, we get $\pi(0) \sum_{j=0}^{\infty} \eta_j = 1$
- Hence stationary distribution exists iff $\sum_{j=0}^{\infty} \eta_j < \infty$
- Let $p_x = p, q_x = q$

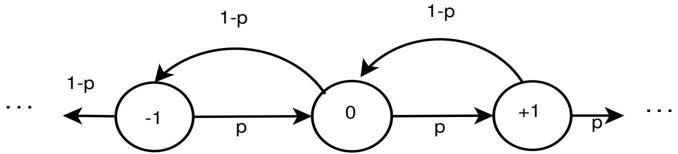
$$\sum_{j=0}^{\infty} \eta_j = \sum_{j=0}^{\infty} \left(\frac{p}{q}\right)^j < \infty \quad \Leftrightarrow \quad p < q$$

- Thus in this special case, the chain is
  - transient if $p > q$; recurrent if $p \leq q$
  - positive recurrent if $p < q$
  - null recurrent if $p = q$

- This analysis can handle chains which are infinite in one direction
- Consider the following random walk chain



- The state space here is $\{\cdots, -1, 0, +1, \cdots\}$
- The chain is irreducible and periodic with period 2
- $P^{2n+1}(0,0) = 0$ and $P^{2n}(0,0) = {}^{2n}C_n p^n (1-p)^n$.
- State 0 is recurrent if $\sum_n P^{2n}(0,0) = \infty$
- We can show that the chain is transient if $p \neq 0.5$ and is recurrent if $p = 0.5$.

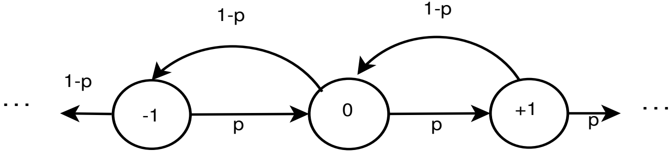


- We use Stirling approximation: $n! \sim n^{n+0.5}e^{-n}\sqrt{2\pi}$.

$$
\begin{aligned}
P^{2n}(0,0) &\sim {}^{2n}C_n p^n(1-p)^n = \frac{(2n)!}{n!n!}p^n(1-p)^n \\
&= \frac{(2n)^{2n+0.5}e^{-2n}\sqrt{2\pi}}{n^{n+0.5}n^{n+0.5}e^{-n}e^{-n}(2\pi)}(p(1-p))^n \\
&= \frac{2^{2n}\sqrt{2n}}{n\sqrt{2\pi}}(p(1-p))^n \\
&= \frac{(4p(1-p))^n}{\sqrt{\pi n}}
\end{aligned}
$$

- We got
$$P^{2n}(0,0) \sim \frac{(4p(1-P))^n}{\sqrt{\pi n}}$$
- When $a_n \sim b_n$, we have $\sum_n a_n = \infty$ iff $\sum_n b_n = \infty$
- So, for state 0 to be recurrent, we need $\sum_n \frac{(4p(1-P))^n}{\sqrt{\pi n}} = \infty$.
- If $p \neq 0.5$, then $4p(1-p) < 1$ and $\sum_n \frac{\alpha^n}{\sqrt{n}} < \infty$ if $\alpha < 1$.
- Hence, if $p \neq 0.5$, then, $\sum_n P^{2n}(0,0) < \infty$
- If $p = 0.5$ then $4p(1-p) = 1$ and hence $\sum_n P^{2n}(0,0) = \infty$.
- The chain is recurrent only when $p = 0.5$

- The one dimensional random walk chain is



- This chain is transient if $p \neq 0.5$ and is recurrent if $p = 0.5$.

- Let $\{X_n,\ n \geq 0\}$ be an irreducible markov chain on a finite state space S with stationary distribution π.
- Let $r : S \to \Re$ be a bounded function.
- Suppose we want $E[r(X)]$ with respect to the stationary distribution π ($E[r(X)] = \sum_{j \in S} r(j)\pi(j)$)
- Let $N_n(j)$ be as earlier. Then

$$\frac{1}{n}\sum_{m=1}^{n} r(X_m) = \frac{1}{n}\sum_{j \in S} N_n(j) r(j)$$

$$\Rightarrow \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} r(X_m) = \sum_{j \in S} r(j) \lim_{n\to\infty} \frac{N_n(j)}{n} = \sum_{j \in S} r(j)\pi(j)$$

- This is known as the ergodic theorem for Markov Chains

MCMC Sampling

- Consider a distribution over (finite) S: $\pi(x) = \frac{b(x)}{Z}$
- Since this is a distribution, $Z = \sum_{x \in S} b(x)$
- We assume, we can efficiently calculate $b(x)$ for any x but computation of Z is intractable or computationally expensive
  E.g., the Boltzmann distribution: $b(x) = e^{-E(x)/KT}$
- We want $E[g(X)]$ w.r.t. distribution π (for any g)

$$E[g(X)] = \sum_x g(x)\ \pi(x) \approx \frac{1}{n} \sum_{i=1}^{n} g(X_i), \quad X_1, \cdots X_n \sim \pi$$

  This is the Monte Carlo method for expectations.
- One way to generate samples is to design an ergodic markov chain with stationary distribution π
  – Markov Chain Monte Carlo sampling

- Suppose $\{X_n\}$ is a an irreducible, aperiodic positive recurrent Markov chain with stationary dist $\pi(x) = \frac{b(x)}{Z}$
- Then we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} g(X_m) = \sum_{x} g(x)\pi(x)$$

- hence, if we can design a Markov chain with a given stationary distribution, we can use that to calculate the expectation.

- We can use the chain to generate samples from distribution π
  We can approximate the expectation as

$$\sum_x g(x)\pi(x) \approx \frac{1}{n}\sum_{i=1}^{n} g(X_{M+i})$$

  with M is large (so that chain is in steady state)
- When we take sample mean, $\frac{1}{n}\sum_{i=1}^{n} Z_i$, we want Z_i to be uncorrelated
- We can, for example, use

$$\sum_x g(x)\pi(x) \approx \frac{1}{n}\sum_{i=1}^{n} g(X_{M+Ki})$$

- For all these, we need to design a Markov chain with π as stationary distribution

- Let $Q = [q(i,j)]$ be the transition probability matrix of an irreducible Markov chain over S.
- Q is called the proposal distribution
- We start with arbitray X_0 and generate $X_{n+1},\ n = 0, 1, 2, \cdots$, iteratively as follows
  - If $X_n = i$, we generate Y with $Pr[Y = k] = q(i,k)$
  - Let the generated value for Y be j. Set

$$X_{n+1} = \begin{cases} j & \text{with probability } \alpha(i,j) \\ X_n & \text{with probability } 1 - \alpha(i,j) \end{cases}$$

- $\alpha(i,j)$ is called the acceptance probability
- We want to choose $\alpha(i,j)$ to make X_n an ergodic markov chain with stationary probabilities π

- The stationary distribution π satisfies (with transition probabilities P)

$$\pi(y) = \sum_x \pi(x)\ P(x,y), \ \ \forall y \in S$$

- Suppose there is a distribution $g(\cdot)$ that satisfies

$$g(y)\ P(y,x) = g(x)\ P(x,y), \ \ \forall x, y \in S$$

  This is called detailed balance
- Summing both sides above over x give

$$g(y) = \sum_x g(y)\ P(y,x) = \sum_x g(x) P(x,y), \ \ \forall y$$

- Thus if $g(\cdot)$ satisfies detailed balance, then it must be the stationary distribution
- Note that it is not necessary for a stationary distribution to satisfy detailed balance

- Any stationary distribution has to satisfy

$$\pi(y) = \sum_{x} \pi(x) \; P(x,y), \;\; \forall y \in S$$

- If I can find a π that satisfies

$$\pi(x)P(x,y) = \pi(y)P(y,x), \;\; \forall x,y \in S, \; x \neq y$$

  that would be the stationary distribution
- This is called detailed balance

- Recall our algorithm for generating $X_n, \; n = 0, 1, \cdots$
- Start with arbitrary X_0 and generate X_{n+1} from X_n
  - If $X_n = i$, we generate Y with $Pr[Y = k] = q(i, k)$
  - Let the generated value for Y be j. Set

$$X_{n+1} = \begin{cases} j & \text{with probability } \; \alpha(i,j) \\ X_n & \text{with probability } \; 1 - \alpha(i,j) \end{cases}$$

- Hence the transition probabilities for X_n are

$$\begin{aligned} P(i,j) &= q(i,j)\, \alpha(i,j), \quad i \neq j \\ P(i,i) &= q(i,i) + \sum_{j \neq i} q(i,j)\, (1 - \alpha(i,j)) \end{aligned}$$

- $\pi(i) = b(i)/Z$ is the desired stationary distribution
- So, we can try to satisfy

$$\begin{aligned} \pi(i)\; P(i,j) &= \pi(j)\; P(j,i), \;\; \forall i, j, i \neq j \\ \text{that is,} \quad b(i) q(i,j)\, \alpha(i,j) &= b(j) q(j,i)\, \alpha(j,i) \end{aligned}$$

- We want to satisfy

$$b(i)q(i,j)\ \alpha(i,j) = b(j)q(j,i)\ \alpha(j,i)$$

- Choose

$$\alpha(i,j) = \min\left(\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)}, 1\right) = \min\left(\frac{b(j)q(j,i)}{b(i)q(i,j)}, 1\right)$$

- Note that one of $\alpha(i,j),\ \alpha(j,i)$ is 1

$$\begin{aligned} \text{suppose}\ \ \alpha(i,j) &= \frac{\pi(j)q(j,i)}{\pi(i)q(i,j)} < 1 \\ \Rightarrow\ \ \pi(i)\ q(i,j)\ \alpha(i,j) &= \pi(j)\ q(j,i) \\ &= \pi(j)\ q(j,i)\ \alpha(j,i) \end{aligned}$$

- Note that $\pi(i)$ above can be replaced by $b(i)$

Metropolis-Hastings Algorithm

- Start with arbitrary X_0 and generate X_{n+1} from X_n
  - If $X_n = i$, we generate Y with $Pr[Y = k] = q(i,k)$
  - Let the generated value for Y be j. Set

$$X_{n+1} = \begin{cases} j & \text{with probability } \alpha(i,j) \\ X_n & \text{with probability } 1 - \alpha(i,j) \end{cases}$$

  Where $Q = [q(i,j)]$ is the transition probabilities of an irreducible chain and

$$\alpha(i,j) = \min\left(\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)}, 1\right)$$

- Then $\{X_n\}$ would be an irreducible, aperiodic chain with stationary distribution π.
- Q is called the proposal chain and $\alpha(i,j)$ is called acceptance probabilities

- Consider Boltzmann distribution: $b(x) = e^{-E(x)/KT}$
- Take proposal to be uniform: from any state, we go to all other states with equal probabilities
- Then,

$$\alpha(x, y) = \min\left(\frac{b(y)}{b(x)}, 1\right) = \min\left(e^{-(E(y)-E(x))/KT}, 1\right)$$

- In state x you generate a random new state y.
  If $E(y) \leq E(x)$ you always go there;
  if $E(y) > E(x)$, accept with probability $e^{-(E(y)-E(x))/KT}$
- An interesting way to simulate Boltzmann distribution
- We could have chosen Q to be 'uniform over neighbours'

- Suppose $E : S \rightarrow \Re$ is some function.
- We want to find $x \in S$ where E is *globally* minimized.
- A gradient descent type method tries to find a locally minimizing direction and hence gives only a 'local' minimum.
- The Metropolis-Hastings algorithm gives another view point on how such optimization problems can be handled.
- We can think of E as the energy function in a Boltzmann distribution

- Let $b(x) = e^{-E(x)/T}$ where T is a parameter called 'temparature'
- $\{X_n\}$ be Markov chain with stationary dist $\pi(x) = \frac{b(x)}{Z}$
- We can find relative occupation of different states by the chain by collecting statistics during steady state
- We know

$$\frac{\pi(x_1)}{\pi(x_2)} = \frac{b(x_1)}{b(x_2)} = e^{-(E(x_1)-E(x_2))/T}$$

- We spend more time in global minimum
  We can increase the relative fraction of time spent in global minimum by decreasing T (There is a price to pay!)
- Gives rise to interesting optimization technique called simulated annealing

- In most applications of MCMC, $x \in S$ is a vector.
- One normally changes one component at a time. That is how neighbours can be defined
- A special case of proposal distribution is the conditional distribution.
- Suppose $X = (X_1, \cdots, X_N)$. To propose a value for X_i, we use $f_{X_i|X_{-i}}$ $(= f_{X_i|X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_N})$
- Here the conditional distribution is calculated using the target π as the joint distribution.
- With such a proposal distribution, one can show that $\alpha(i,j)$ is always 1
- This is known as Gibbs sampling