Probability
Stationarity and Periodicity

Daniel Yu

October 22, 2024

Contents

1 Stationary Distributions

Note. What is the long-term behavior of an absorbing markov chain? It gets absorbed by one of the absorbing states but which one? Thus, the question is: What is the probability that I get absorbed by any particular absorbing state in the markov chain?

Theorem 1. $P[\text{ends at a} \mid X_0 = i] = (NR)_{i,a}$ for any transient state i and absorbing state a. $N = (Q - I)^{-1}$ from the matrix

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.$$

Proof. Let X_n denote the state of the chain at step n before absorption. We want to compute $P[\text{absorbed at } a \mid X_0 = i]$ for an initial transient state i and absorbing state a.

By the law of total probability, we have:

$$P[\text{absorbed at } a \mid X_0 = i] = \sum_{n=0}^{\infty} \sum_j P[\text{absorbed at } a \mid X_n = j, X_0 = i] \cdot P[X_n = j \mid X_0 = i].$$

Since absorption is an absorbing event and the chain can be "reset" upon reaching any transient state j, we focus on the probability of reaching a from j in one step after reaching j from i after n steps. For any transient j, by the time-homogeneity of Markov chains, we get:

$$P[\text{absorbed at } a \mid X_n = j, X_0 = i] = P[X_1 = a \mid X_0 = j].$$

Therefore,

$$\begin{aligned} P[\text{absorbed at } a \mid X_0 = i] &= \sum_{n=0}^{\infty} \sum_j P[X_1 = a \mid X_0 = j] \cdot P[X_n = j \mid X_0 = i] \\ &= \sum_{n=0}^{\infty} \sum_j R_{j,a} \cdot (Q^n)_{i,j} \\ &= \sum_j R_{j,a} \sum_{n=0}^{\infty} (Q^n)_{i,j}. \end{aligned}$$

The inner sum $\sum_{n=0}^{\infty} (Q^n)_{i,j} = (I + Q + Q^2 + \ldots)_{i,j}$ represents the expected number of times the chain is in state j given it starts at i, which is the (i,j)-entry of the fundamental matrix $N = (I - Q)^{-1}$. Thus,

$$\begin{aligned} P[\text{absorbed at } a \mid X_0 = i] &= \sum_j R_{j,a} \cdot N_{i,j} \\ &= (NR)_{i,a}. \end{aligned}$$

This completes the proof. □

Note. What is the long-term effect of non-absorbing time-homogenous markov chains?

Definition 1. A pair of states $i \neq j$ are intercommunicating. if starting at i, I can make it to j and starting at j, I can make it to i. Formally $i \equiv j$ (equivalence relation) if :

$$\exists n \text{ such that} P[X_n = j \mid X_0 = i] > 0, \text{ i.e.} (P^n)_{i,j} > 0$$
$$\exists m \text{ such that} P[X_m = i \mid X_0 = j] > 0, \text{ i.e.} (P^m)_{j,i} > 0.$$

Definition 2. Define $C_i = \{$ all states $j : i \equiv j\}$. Trivially, each vertex (state) is an equivalent to itself, so the equivalence class of C_i must be nonempty.

Theorem 2. We can partition the state space Ω into **intercommunicating classes** $\{C_i\}$ (i.e can partition the vertices of the graph into subgraphs) . Note that this forms **strongly connected components**

Definition 3. A markov chain is **irreducible** if it has precisely one intercommunicating class (i.e. $\forall i, j \in \Omega, \exists n$ such that $P[X_n = j \mid X_0 = i] > 0, (P^n)_{i,j} > 0$)

Definition 4. A markov chain is **regular** if $\exists N$, a single choice, such that $\forall i, j, (P^N)_{i,j} > 0$ which implies that $P[X_N = j \mid X_0 = i] > 0$.

Note. Not all markov chains that are irreducible are regular. For example, consider $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ which is irreducible but not regular. There is no way to make all four entries strictly positive for one choice of N in P^N

Theorem 3. if $i \equiv j$ then both are recurrent or both transient

Proposition 1. If $\mid \Omega \mid < \infty$ then at least one state is recurrent.

Proof. Assume by contradiction that all states are transient. Then $V(i) < \infty$ $\forall i \in \Omega$. But $\sum_{i \in \Omega} V(i) = \infty$ since we can continue the markov chain forever. This is a contradiction, so at least one $V(i) = \infty$ and thus recurrent. $\square$

Corollary 1. given the theorem and proposition, then all states in a irreducible markov chain are recurrent.

Note.

1. Given an irreducible Markov chain $\{X_n\}$, what proportion of time is spent in state i? Does the answer to the above matter depending on where I start?
2. LEt $\{X_n\}_{n=0,1,2,\ldots}$ be the sequence of RV. Does X_n converge in distribution? i.e.

$$\lim_{n \to \infty} P[X_n = j \mid X_0 = i] \to c?$$

Does it depend on $X_0 = i$?

Definition 5. A row vector π is **stationary** if $\pi P = \pi$ and $\pi_i \geq 0$. Notice that this is equivalent to saying π is a row-eigenvector of P with $\lambda = 1$. So,

$$\pi P^n = (\pi P) \times P \ldots \times P = \pi.$$

Remark. $\sum_{i=\Omega} \pi_i = 1$ blc probability

Note. Example

Suppose π is stationary vector and $P[X_0 = j] = \pi_j$:

$$\begin{aligned}
P[X_i = j] &= \sum_{i \in \Omega} P[X_i = j \mid X_0 = i] \cdot P[X_0 = i] \\
&= \sum_{i \in \Omega} P_{i,j} \pi_j \\
&= (\pi P)_j \\
&= \pi_j .
\end{aligned}$$

Note. Only for irreducible markov chain
So if we have $X_0 \sim \pi$ (i.e. $P[X_0 = i] = \pi_i$, then the distribution of $X_n \sim \pi$). What if $X_0 \not\sim \pi$?

Intuition

$P[X_n = j \mid X_0 = i] = (P^n)_{i,j}$

We can diagonalize the matrix P, if it is diagonalizable, as: $P = Q \Lambda Q^{-1}$
where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a diagonal matrix with the eigenvalues of P
$= (Q \Lambda^n Q^{-1})_{i,j}$
Given that $\lambda_1 = 1$ (1 is always an eigenvalue of a Markov transition matrix)
and for an irreducible Markov chain,
the remaining eigenvalues satisfy $|\lambda_i| < 1$ for $i \geq 2$,
as $n \to \infty, \Lambda^n = \mathrm{diag}(1, \lambda_2^n, \lambda_3^n, \ldots, \lambda_k^n) \to \mathrm{diag}(1, 0, 0, \ldots, 0)$,
$= (Q \ \mathrm{diag}(1, 0, 0, \ldots, 0) Q^{-1})_{i,j}$
$\approx \pi_j .$

Theorem 4. Every finite-state markov chain has at least one stationary vector. If the finite state markov is irreducible this vector is unique.

Lemma 1. Let S be a row-stochastic matrix with strictly positive entries.

$$d = \min_{i,j} S_{i,j} > 0.$$

Let v be a column vector, $M_0 = max\{v_i\}$, $m_0 = min\{v_i\}$.
Let $M_1 = max\{(Sv)_i\}$, $m_1 = min\{(Sv)_i\}$ for all entries in Sv. Then,

$$M_1 - m_1 \leq (1 - 2d)(M_0 - m_0).$$

and $M_1 < m_0$ and $m_1 \gg m_0$.

Proof.

$$(Sv)_i = \sum_{k \in \Omega} S_{i,k} \cdot v_k$$

to maximize this, let's assume $V_k = M_0$ for all k except one, $V_{i^*} = m_0$ (worst case)

$$\leq \left(\sum_{k \neq i^*} S_{i,k} \right) M_0 + S_{_{i,i^*}} m_0.$$

Since S is positive and row stochastic matrix, the worst case scenario is $\sum_{k \neq i^*} S_i, k = 1-d, S_{i,i^*} = d$. Thus, $\forall i$,

$$(Sv)_i \leq (1 - d)M_0 + dm_0.$$

To bound m_1, assume $v_k = m_0$ for all $k \neq i^*$, and $v_{i^*} = M_0$. Then,

$$(Sv)_i \geq \sum_{k \neq i^*} S_{i,k} m_0 + S_{i,i^*} M_0.$$

Since $\sum_{k \neq i^*} S_{i,k} = 1 - S_{i,i^*} = 1 - d, S_{i,i^*} = d$,

$$(Sv)_i \geq (1 - d)m_0 + dM_0.$$

Combining the upper and lower bounds, we get:

$$M_1 - m_1 \leq [(1 - d)M_0 + dm_0] - [(1 - d)m_0 + dM_0].$$

Simplifying the expression:

$$M_1 - m_1 \leq (1 - 2d)(M_0 - m_0).$$

which implies that the range of the vector v shrinks by at least a factor of $(1 - 2d)$ after each application of S. Since $0 < d < 1/2$, repeated application of S after k steps leads to convergence toward a stationary vector where $M_k = m_k$ i.e. max = min, proving contraction of the range. Therefore, the stationary distribution is unique. $\square$

Corollary 2. Let P be the transition matrix of a regular finite state markov chain. If $Pv = v$ for some column vecetor v, then v is constant. In particular, $Null(P - I) = \vec{1}$.

Proof. $\exists N$ such that P^N is strictly positive and row-stochastic. Let M_0 and m_0 be the maximum and minimum of v, and let M_1 and m_1 be the maximum and minimum of $P^N v$. By the lemma, we have

$$M_1 - m_1 \leq (1 - 2d)(M_0 - m_0).$$

Since $P^N v = v$, it follows that $M_1 = M_0$ and $m_1 = m_0$. Therefore, from the inequality we conclude that

$$M_0 - m_0 = 0,$$

because $1 - 2d < 1$ and the only way for this inequality to hold is if $M_0 = m_0$.

Thus, the maximum and minimum values of v are the same, implying that v is a constant vector i.e. $v = \begin{pmatrix} x \\ \vdots \\ x \end{pmatrix}$. Since for any v', $Pv' = v'$ where v' must be the constant vector, then in fact $v' = cv$ and unique, the null space of $P - I$ is one-dimensional. Therefore, $Null(P - I) = \vec{1}$. $\square$

Lemma 2. $dim(\pi) = 1$ for $\pi P = \pi$. If there exists a stationary vector for P_1, the transition matrix of a regular finite state markov chain, then the vector is unique.

Theorem 5. Ergodic Theorem
The ergodic theorem for finite state markov chain states the following. Let P be a the transition matrix. Then $\lim_{n\to\infty} P^n$ exists, has constant columns, and each row is made up of the unique stationary vectors of P. Furthermore, for any intial probability distribution v, $\lim_{n\to\infty} P^n v = \pi$ the unique stationary distribution.

Proof. Let v_1 be the first column of P and let N be such taht P^N is strictly positive. Denote,

$$\begin{aligned} M_0 &= max\{v_1\} \\ m_0 &= max\{v_1\}. \end{aligned}$$

Let

$$\begin{aligned} M_k &= max\{P^{Nk}v_1\} \\ m_k &= min\{P^{Nk}v_1\}. \end{aligned}$$

By the Lemma 1, $\{M_k\}$ is non-increasing sequence and $\{m_k\}$ is a non-descreasing sequence. We also know that $(M_k - m_k) \leq (1-2d)(M_0 - m_0)$. A limit must exists and we know that the convergence:

$$\lim_{k\to\infty} m_k = \lim_{k\to\infty} M_k = w_1 \in [0,1].$$

and thus,

$$\lim_{k\to\infty} \left(P^{Nk}v_1\right) = \begin{pmatrix} w_1 \\ \vdots \\ w_1 \end{pmatrix}.$$

Now consider when v is the jth column vector of P, dentoed as v_j:

$$\left(P^{NK}v_i\right) \to \begin{pmatrix} w_j \\ \vdots \\ w_j \end{pmatrix}.$$

So,

$$\lim_{n\to\infty} P^{n+1} = \lim_{n\to\infty} P^n(v_1, v_2, \ldots, v_{|\Omega|}) = \begin{pmatrix} w_1 & w_2 & \cdots & w_{|\Omega|} \\ w_1 & w_2 & \cdots & w_{|\Omega|} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_{|\Omega|} \end{pmatrix}.$$

, all $w_i \geq 0$ and $\sum_{i\in\Omega} w_i = 1$ (since P^n is still the transition matrix for markov chain). Only thing left is to show that $w = wP$ (and thus $\pi = w$). Consider,

$$\begin{aligned} \lim_{n\to\infty} P^{n+1} &= \lim_{n\to\infty} P^n \cdot P \\ &= \begin{pmatrix} w_1 & w_2 & \cdots & w_{|\Omega|} \\ w_1 & w_2 & \cdots & w_{|\Omega|} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_{|\Omega|} \end{pmatrix} \cdot P \\ &= \begin{pmatrix} w_1 & w_2 & \cdots & w_{|\Omega|} \\ w_1 & w_2 & \cdots & w_{|\Omega|} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_{|\Omega|} \end{pmatrix}. \end{aligned}$$

□

Consider any row, we get $\vec{w} = \vec{w} \cdot P$ and $\pi = w$

Remark. It follows quite clearly that $\lim_{n\to\infty}(P^n)_{i,j} = \lim_{n\to\infty} P[X_n = j \mid X_0 = i] = \pi_j$ the stationary distribution at j. If I suppose $X_0 = i$, then $X_n \to \pi$ in distribution

$$\lim_{n\to\infty} P[X_n = j \mid X_0 = i] = \pi_j.$$

Geometrically, this is what is happening

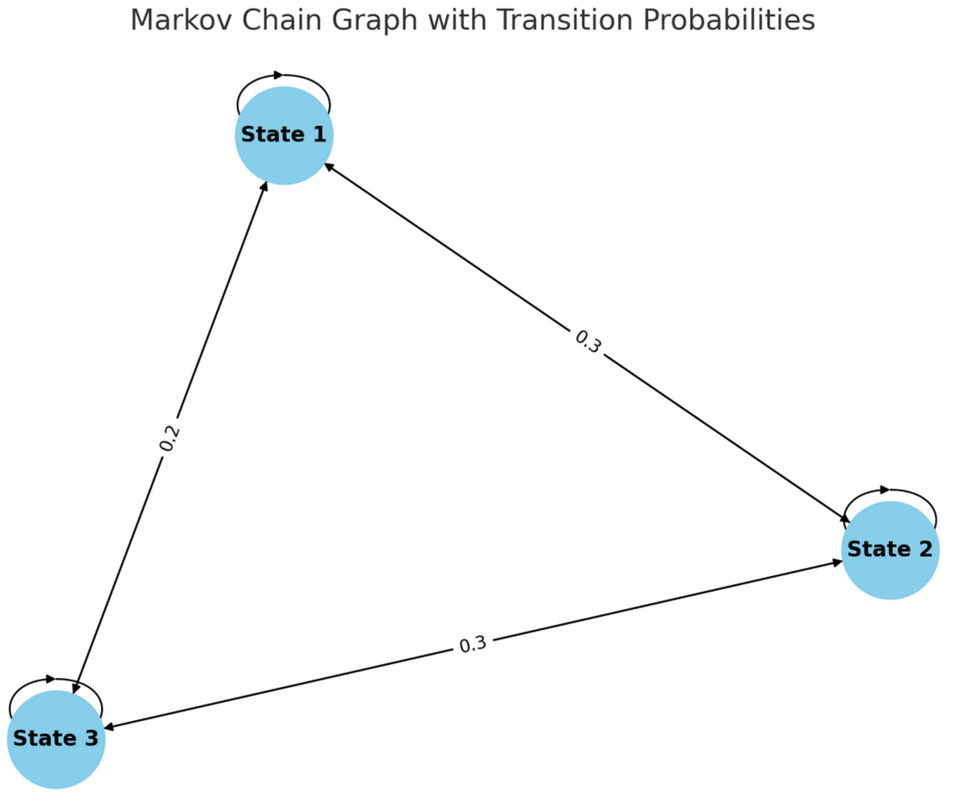

Figure 1: Markov Chain Example

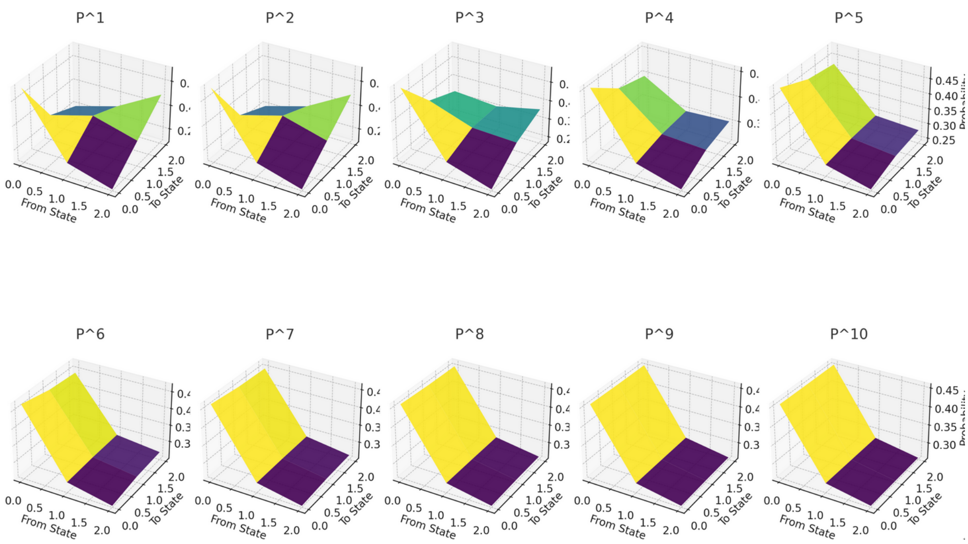

Figure 2: Transition Matrix P converging to Stationary Distribution

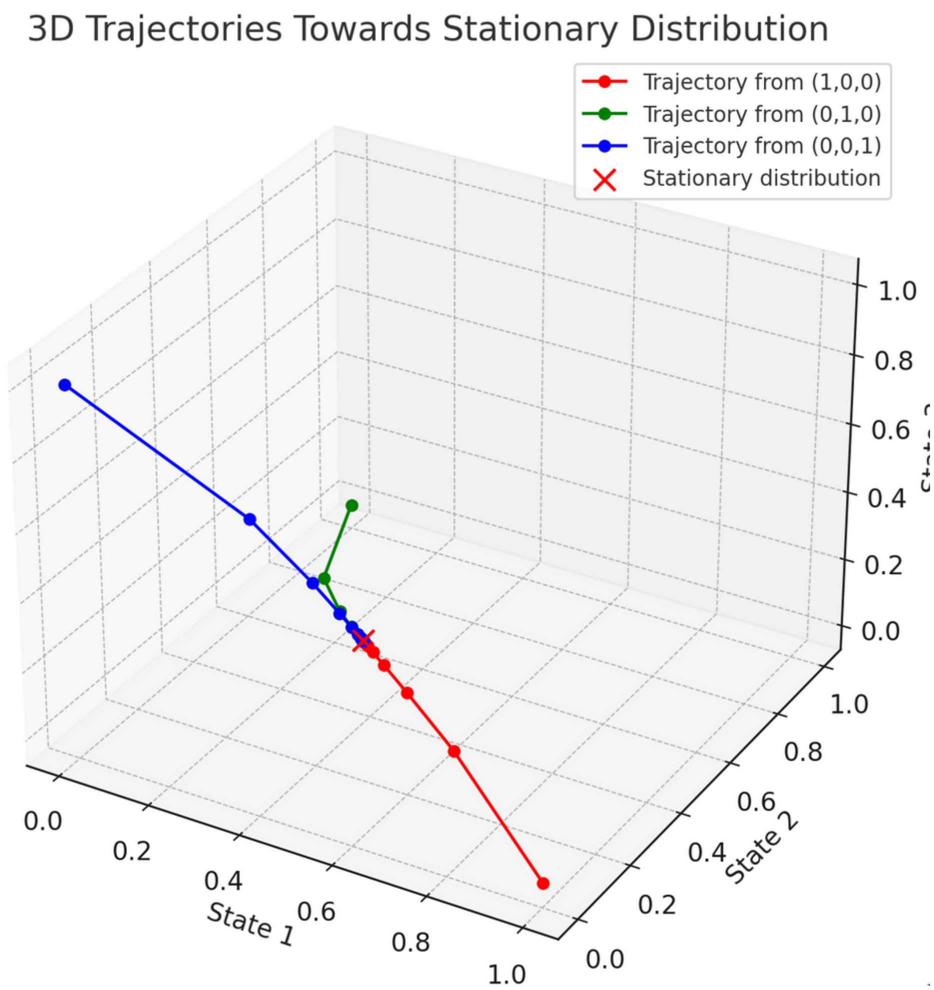

Figure 3: Trajectory of Convergence from starting vectors

Theorem 6. For a state $i \in \Omega$, let $F_n(j) = \frac{1}{n}\{\text{number of visits to } j \text{ up until time n}\}$. Let V_k be the indicator variable $V_k = 1\,|_{X_k=j}$. Then, $F_n(j) = \frac{1}{n}\sum_{k=1}^{n} V^k(j)$. We claim that

$$F_n(j) \to \pi_j.$$

in probabiltiy regardless of intial distribution.

Proof. To prove this we first need to compute:

$$\begin{aligned} E[F_n(j) \mid X_0 = i] &= \frac{1}{n}\sum_{k=1}^{n} E[V^k(j) \mid X_0 = i] \\ &= \frac{1}{n}\sum_{k=1}^{n} P[X_k = j \mid X_0 = i] \\ &= \frac{1}{n}\sum_{k=1}^{n} (P^k)_{i,j}. \end{aligned}$$

□

2 Periodicity

Recall

Definition 6. A markov chain is **irreducible** if it has precisely one intercommunicating class (i.e. $\forall i, j \in \Omega, \exists n$ such that $P[X_n = j \mid X_0 = i] > 0, (P^n)_{i,j} > 0$).

Definition 7. A markov chain is regular if $\exists N$, a single choice, such that $\forall i, j, (P^N)_{i,j} > 0$ which implies that $P[X_N = j \mid X_0 = i] > 0$.

Note. Notice the behavior for P, $P^N_{1,1} > 0$ for N even and $P^N_{1,2} > 0$ for N odd. Can we describe this behavior?

Definition 8. Thus, we introduce periodicity.

$$S_i = \{n : P(X_n = i \mid X_0 = i) > 0\}.$$

where the period is

$$period(i) = gcd(S_i).$$

. A markov chain is **aperiodic** if $period(i) = 1 \forall i \in \Omega$

Theorem 7. irreducible + regular $\Rightarrow$ aperiodic for finite state markov chains

Remark. Aperiodic, finite state markov chains are essentially **mixers**