- Initial distribution : Prob of being in each state at beginning. π_0
- Transition probabilities : (2 states : 2×2 , 5 states : 5×5)

Markov property ↓

$$P = \begin{bmatrix} 1\to1 & 1\to2 & 1\to3 \\ 2\to1 & 2\to2 & 2\to3 \\ 3\to1 & 3\to2 & 3\to3 \end{bmatrix}, \quad 1\to2 \text{ means } P(X_{n+1}=2 \mid X_n=1, \cancel{\text{and all prev states } X_{n-1} \sim X_0})$$

- each row sums up to 1
- if each column also sums up to 1, the chain is "doubly stochastic", more later.

$$P(X_0=i_0, X_1=i_1 \cdots X_n=i_n) = P(X_0=i_0)P(X_1=i_1 \mid X_0=i_0) \cdot P(X_2=i_2 \mid X_1=i_1) \cdots P(X_n=i_n \mid X_{n-1}=i_{n-1})$$

$$= \pi_0(i_0) P_{i_0,i_1} \cdot P_{i_1,i_2} \cdots P_{i_{n-1},i_n}$$

$$P(X_n=i_n) = \sum_{i_0 \sim i_{n-1}} \pi_0(i_0) \cdot P_{i_0,i_1} \cdot P_{i_1,i_2} \cdots P_{i_{n-1},i_n}$$

$$= \pi_0 P^n (i_n)$$

$\Rightarrow \pi_n = \pi_0 P^n$ (apply the matrix n times)

π_n is a row vector, of prob in each state

notation: $\pi_n(i)$ state i, or, $P(X_n=i)$. (ex. $P(X_0=i)=\pi_0(i)$ $\forall i$ in state space)

at step n

π with no subscript : the invariant distribution for P.

Doesn't matter which step, n.

- Distribution π is invariant for P if $\pi=\pi P$. In this case, $\pi_n=\pi_0$ for all $n \geq 0$. Can solve for π by setting up the "$\pi=\pi P$" balance equations.

special case : If $P=I$, $\pi=\pi P$ $\forall \pi$, so any distribution is invariant (distribution at all steps = the initial distribution)

(MC is) Irreducible : can (non-zero prob) go from every state to every other state. state transition diagram is a directed graph w/ single connected component.

(long term) Fraction of time (spent) in state $\boxed{i}$ is :

$$\lim_{n\to\infty} \frac{1}{n} \underbrace{\sum_{m=0}^{n-1} 1\cdot\{X_m = i\}}$$

↳ counts # steps among 0 ~ n-1 such that $X_m = i$

For finite, irreducible chains :

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} 1\cdot\{X_m = i\} = \pi(i)$$

This means the invariant distribution exists and is unique for all irreducible chains

For irreducible chains :

period of state i = $\gcd(n>0 \mid P(X_n=i \mid X_0=i) > 0) = d(i)$

If the period equals 1 ($d(i)=1$), the state is aperiodic.

ex. (gcd(1,2,3,4,5), or gcd(4,5)).

- Markov chain is aperiodic if all states are aperiodic ($d(i)=1$ for all i)
- $\boxed{\text{Every state in irreducible chain has the same period}}$
- ⇒ In an irreducible chain, if one state is aperiodic, the chain is aperiodic

↳ Proof :

```
    ___ r __
 /            \    __ t _
 i             j /        |
 \____ s __  /  \____/
```

Consider an arbitrary pair of states i and j. Since the chain is irreducible, there exists a length r walk from i to j, and a length s walk from j to i.
So, if you start at i, go to j, and come back to i, how far have you walked? r + s
This is divisible by d(i), the period of i.

If, instead, you start at i, go to j, take the length - t path, get back to j, and finally return to i, how far have you walked? r + t + s
This is divisible by the period of i as well because you started at i and returned to i.

Since both r + s and r + t + s are divisible by d(i), t is divisible by d(i).

What is t? It is a multiple of the period of j
So d(j) is divisible by d(i).
Similarly, d(i) is also divisible by d(j) if we switch the positions of i and j.

Therefore, d(i) = d(j), for all i, j.

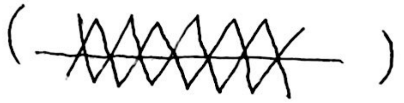
Convergence :

irreducible : The unique invariant distri π always exists. $(\pi = \pi P)$

---

periodic :

- Some periodic chains don't ever converge to π.
- Some do converge to π, depending on the <u>initial distribution</u>.
- So, periodic chains are not <u>guaranteed</u> to converge (But may converge !)
- The fraction of time of being in each state always converges to π.



aperiodic : π_n always converges to π as $n \to \infty$, for <u>any</u> π_0.

In other words : $P(X_n = i) \to \pi(i)$ $\forall$ states i as $n \to \infty$.

---

reducible : period is not defined ~~(we are not interested)~~ states may not have same period. same states can't return to themselves ...

---

whether a chain is periodic (periodicity), and whether a chain is reducible (reducibility) are <u>not</u> dependent on the initial distribution! These are properties of the transition matrix/graph!

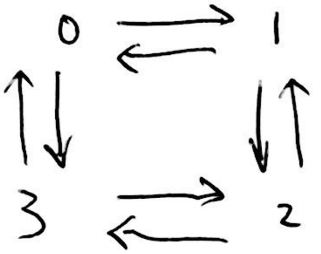
"proof": irreducible → all states have the same period

a state is aperiodic if period = 1. otherwise periodic

$$P(\enclose{circle}{X_n} = i \mid \enclose{circle}{X_0} = i) = \sum_{x_1 \sim x_{n-1}} P(X_1 = x_1 \mid \enclose{circle}{X_0} = i)\, P(X_2 = x_2 \mid X_1 = x_1) \cdots \cdots P(\enclose{circle}{X_n} = i \mid X_{n-1} = x_{n-1})$$

Σ sum over all ~~Paths~~ possibilities

Non-trivial (not necessarily $= \pi$) initial distributions that converge for periodic chains



$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

Continued from the bottom of the page:
If there doesn't exist such eigenvectors, then the chain is aperiodic, and any initial distribution will converge.
If all eigenvalues have modulus / magnitude 1, then the only initial distribution that converges is the invariant distribution, which converges trivially.
For all other periodic chains, there exists a non-trivial set of initial distributions that converge.

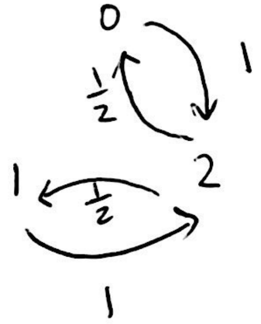
eigenvalues: $0, \pm 1$

for eigenvalue $= -1$, eigenvector $= [-x, x, -x, x]$

$\pi_0 \cdot V = 0$, so $\pi_0(0) + \pi_0(2) - \pi_0(1) - \pi_0(3) = 0$

example π_0: $[\underbrace{0.25, \overbrace{0.3}^{}}_{0.5}, \underbrace{0.25, 0.2}_{}]$

$$[\underbrace{0.25,\ 0.3}_{0.5},\ 0.25,\ 0.2] \quad [0.25,\ \underbrace{0.3,\ 0.25,\ 0.2}_{0.5}]$$



$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

irreducible, periodic w/ period 2

for eigenvalue -1, $V = [-x, -x, x]$

$\pi_0 \cdot V = 0 \Rightarrow \pi_0(0) + \pi_0(1) - \pi_0(2) = 0$

example π_0: $[\frac{1}{4}, \frac{1}{4}, \frac{1}{2}]$

For a periodic Markov chain to converge, the initial distribution π_0 has to be orthogonal to the eigenvector(s) with eigenvalue -1 (or modulus -1, for complex eigenvalue).

$$\pi_0 P^n = \sum_k \pi_0 \underset{\substack{\uparrow \\ k^{th} \text{ eigenvector (right)}}}{V_k} \overset{\text{left eigenvector}}{l_k} \lambda_k \leftarrow \text{eigenvalue}$$

so $\pi_0 V_k$ needs to be 0 when $\lambda_k \neq 1$ but $|\lambda_k| = 1$ (ex. -1) for the sum to not oscillate.

First Step Equations :

States ABCDE , $\beta(i)$ = Avg # steps until it reaches state E, Starting from state i

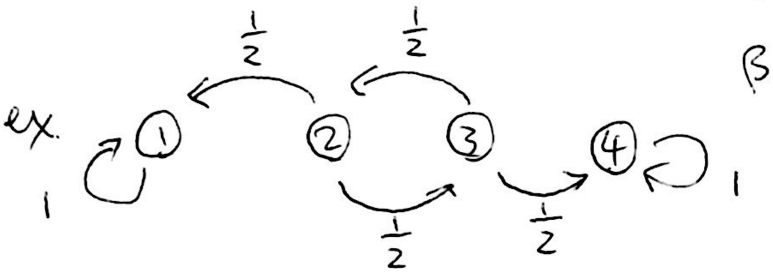
for states ABCDE, FSE's are $\beta(A) = 1 + \sim$

$\beta(B) = 1 + \sim$

$\vdots$

$\beta(E) = 0$

ex.



Starting at ②, expected time to absorption ?

$\beta(1) = 0$

$\beta(2) = 1 + \frac{1}{2}\beta(1) + \frac{1}{2}\beta(3) = 1 + \frac{1}{2}\beta(3)$

$\beta(3) = 1 + \frac{1}{2}\beta(2) + \frac{1}{2}\beta(4) = 1 + \frac{1}{2}\beta(2)$

$\beta(4) = 0$

$\Rightarrow \beta(3) = 2$

$\beta(2) = 1 + \frac{1}{2}\beta(3) = 2$

- Doesn't really matter what the goal is (2 absorbing states vs 1), write down the 1st hop(s) for each state and let the 1st hop(s) 'figure it out'.

Prob of event A (collection of States) before B :

$\alpha(i)$ = Starting at State i

- 3 types / cases of FSE'S

$\alpha(i) = 1 \quad \forall i \in A$

$\alpha(i) = 0 \quad \forall i \in B$

$$\alpha(i) = \sum_j \underbrace{P(i,j)}_{\text{getting to 1st hop}} \overbrace{\alpha(j)}^{\text{1st hop}} \quad \forall i \notin A \cup B$$

Doubly Stochastic Chains

X_n : Sum of n independent rolls of a die

$K \geq 2$

$\lim_{n\to\infty} P(X_n \text{ is divisible by } k) = ?$

K states : $0, 1, 2 \dots k-1$

From state i, can move to $i+1 \pmod{k}$, $i+2 \pmod{k} \dots i+6 \pmod{k}$

each of the 6 transitions : $\frac{1}{6}$ prob

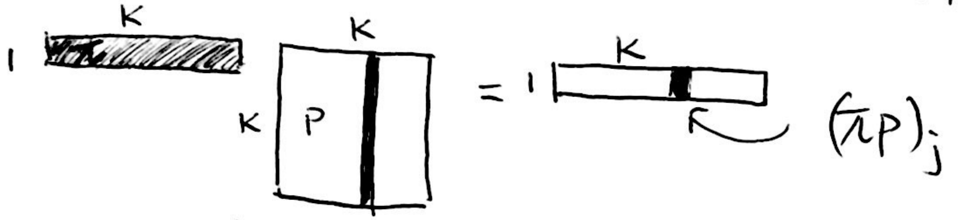
Def : Y_n = state of the chain after n steps

(Yn can only be one of K states)

$Y_n = X_n \pmod{k}$, so X_n is divisible by K iff $Y_n = 0$

---

consider the uniform distribution $\pi = (\frac{1}{K}, \frac{1}{K} \cdots \frac{1}{K}) \in [0,1]^K$



$$(\pi P)_j = \sum_{i=0}^{K-1} \pi_i P_{ij} = \frac{1}{K} \sum_{i=0}^{K-1} P_{ij} = \pi(j) \cdot (1) = \pi(j)$$

diff values for i, same j

i can be :

$j-1 \pmod{k}$	$i \to j : \frac{1}{6}$
$j-2 \pmod{k}$	$i \to j : \frac{1}{6}$
$\vdots$	$\vdots$
$j-6 \pmod{k}$	$\frac{1}{6}$

$\pi P = \pi$, so $\pi = (\frac{1}{K}, \frac{1}{K} \cdots \frac{1}{K})$ is the invariant distribution.

$$\lim_{n\to\infty} P(X_n \text{ divisible by } K) = \lim_{n\to\infty} P(Y_n = 0) = \pi(0) = \boxed{\frac{1}{K}}$$

generalize: finite, irreducible, aperiodic MC w/ doubly stochastic trans matrix has uniform invariant distribution.

"coupling"

- 2 independent copies of the same chain, X and Y
- The first one, X, starts with any initial distribution
- The second starts with π.
- There is a non-zero probability that they will meet. Meaning as $n \to \infty$, they will meet at some point.

This is actually enough (proof below) to show that

$$\pi_n \to \pi \text{ as } n \to \infty \qquad \text{(convergence !)}$$

---

2 chains X, Y.

T: when they meet

X_n: the chain with a rand/any starting distribution

Y_n: starts with π

define Y'_n as follows:

$$Y'_n = \begin{cases} Y_n \leftarrow \boxed{\text{the } \pi \text{ one}} & n < T \\ X_n & n \geq T \end{cases}$$

As defined, $\forall n$, $Y'_n \sim \pi$

$$\begin{aligned}\pi_n(A) - \pi(A) &= \boxed{P(X_n = A)} - P(Y'_n = A) \\ &= \boxed{\cancel{P(X_n = A, T \leq n)} + P(X_n = A, T > n)} - \\ &\quad \cancel{P(Y'_n = A, T \leq n)} - P(Y'_n = A, T > n) \\ &= P(X_n = A, T > n) - P(Y'_n = A, T > n) \leq P(T > n)\end{aligned}$$

$P(T>n)$ is P(T, the time it takes for X, Y to meet, is greater than n)

$$P(T > n) \to 0 \text{ as } n \to \infty,$$

and since $\pi_n(A) - \pi(A) \leq P(T > n) \to 0$, "they will meet" is sufficient to prove/argue that $\pi_n \to \pi$ as $n \to \infty$

The Markov property

example: run the chain m steps, obtaining states $X_0, X_1 \cdots X_m$

Reverse order: $X_m, X_{m-1} \cdots X_1, X_0$

$\longleftarrow$

$$0 \quad 1 \quad 2 \cdots k, \enclose{circle}{k+1}, \boxed{k+2 \cdots m}$$

Show: given k+1, k is indpt of k+2 ~ m.

$$\underline{P(k \mid k+1 \sim m)} = \frac{P(k \sim m)}{P(k+1 \sim m)} = \frac{P(k, k+1) \cdot P(k+2 \sim m \mid k, k+1)}{P(k+1)\, P(k+2 \sim m \mid k+1)}$$

$$= \frac{P(k, k+1)}{P(k+1)} = \underline{P(k \mid k+1)}$$

What are the transition probs Q_{ij}?

$$Q_{ij} = \enclose{circle}{P(X_k = j \mid X_{k+1} = i)} = \frac{P(X_k = j)\, P(X_{k+1} = i \mid X_k = j)}{P(X_{k+1} = i)}$$

want to use P, so how to write this as $P(X_{k+1} = i \mid X_k = j)$? Bayes.

$$= \frac{?\ P_{ji}}{?}$$

plug in π, (use it as initial distribution), so $\boxed{\frac{\pi(j) P_{ji}}{\pi(i)}}$

Def: time reversible: $\pi(i) P_{ij} = \pi(j) P_{ji}$

so $Q_{ij} = \frac{\pi(j) P_{ji}}{\pi(i)} = \frac{\cancel{\pi(i)} P_{ij}}{\cancel{\pi(i)}} = P_{ij}$

This means states follow the same transition probs whether viewed in forward or reverse order.

Gambler's Ruin

2 players

each round : a player wins \$1 w/ prob $\frac{1}{2}$
loses \$1 w/ prob $\frac{1}{2}$

State at time t : \$ won by player 1 (can be positive or neg).

initial State : 0

---

player 1 cannot lose more than l_1 dollors

player 2 cannot lose more than l_2 \$.

game ends when $-l_1$ or l_2 is reached (one of them is "ruined")

So. this is a MC with 2 absorbing / recurrent states.

P(player 1 wins l_2 before ~~los~~ losing l_1 dollars)?

Define P_i^t : prob that chain is at state i after t steps.

For $-l_1 < i < l_2$, $\lim_{t \to \infty} P_i^t = 0$, (i is transient)

Define q : prob that chain is absorbed into state l_2 (So that game ends w/ player 1 winning l_2) then $\lim_{t \to \infty} p_{l_2}^t = q$

$1 - q$: (prob that chain is absorbed into state $-l_1$

In each round / step, expected gain of player 1 is 0

- expected gain of player 1 after t steps is 0 by induction.

Define G^t : Gain of player 1 after t steps. So $E(G^t) = 0 \; \forall t$

$0 = E(G^t)$ can be written as $\sum_{i=-l_1}^{l_2} i \, P_i^t = 0$

$\lim_{t\to\infty} E(G^t) = l_2 q - l_1(1-q) = 0$

Solve: $l_2 q - l_1(1-q) = 0$ $\quad q = \frac{l_1}{l_1+l_2}$

recall: q is the prob that player 1 wins l_2

⇒ Fact: Prob of winning is proportional to the amount of money a player is willing to lose.

---

example: $0 \sim n-1$, $n-1$ sheep, 1 wolf, in a circle

each step: wolf moves left or right w/ prob $\frac{1}{2}$, and eats the sheep there ...

which sheep is most likely eaten last? (what's the 'best' position for a sheep to be in this circle?)

P_i : i^{th} sheep (point) is eaten last (reached last)

– Have reached $i-1$ and $i+1$, and all other points already

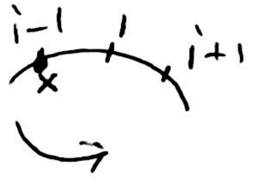
case 1: $i-1$ is visited before $i+1$

Meaning at $i-1$, it hasn't visited $i+1$ or i yet

$P(i+1 \text{ is visited before } i) = ?$

Gambler's ruin:

[player 1 has \$1, player 2 has \$$n-2$]



$P(i+1 \text{ is visited before } i) = P(\text{player 1 wins}) = \frac{1}{1+(n-2)} = \frac{1}{n-1}$

case 2: $i+1$ is visited before $i-1$. Identical

$P_i = P(i-1 \text{ visited before } i+1)\cdot\frac{1}{n-1} + P(i+1 \text{ visited before } i-1)\cdot\frac{1}{n-1} = \frac{1}{n-1}$

All sheep are equally likely to be eaten last.

Hmm; can only observe evidence at a state, not the actual state

have:
- transition matrix
- observation matrix

$$P(X_t | e_1 \sim e_t) = P(X_t | e_1 \sim e_{t-1}, e_t)$$

note that $P(a | b, c) = \alpha P(a, b | c)$

$$\text{so} = \alpha P(X_t, e_t | e_1 \sim e_{t-1})$$

$$= \alpha \sum_{X_{t-1}} P(X_{t-1}, X_t, e_t | e_1 \sim e_{t-1})$$

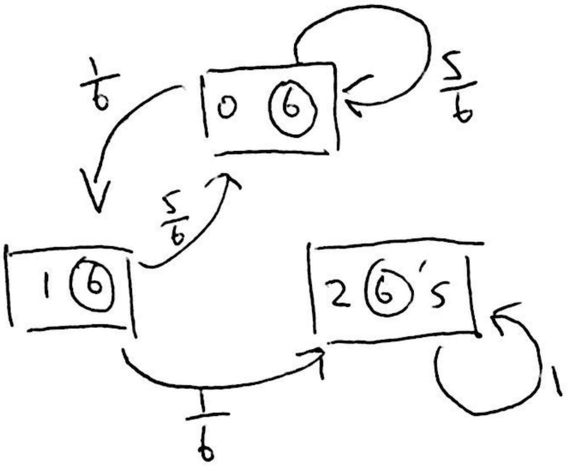
$$= \alpha \sum_{X_{t-1}} P(X_{t-1} | e_1 \sim e_{t-1}) \overset{P(X_t | X_{t-1}, e_1 \sim e_{t-1})}{P(X_t | X_{t-1})} \overset{P(e_t | X_t, X_{t-1}, e_1 \sim e_{t-1})}{P(e_t | X_t)}$$

$$= \alpha P(e_t | X_t) \sum_{X_{t-1}} P(X_t | X_{t-1}) P(X_{t-1} | e_1 \sim e_{t-1})$$

$$P(X_t | e_1 \sim e_t) = \alpha \underbrace{\overbrace{P(e_t | X_t)}^{\text{observation matrix}}}_{\text{update}} \underbrace{\sum_{X_{t-1}} \overbrace{P(X_t | X_{t-1})}^{\text{transition matrix}} P(X_{t-1} | e_1 \sim e_{t-1})}_{\text{predict}}$$

3 States: $\boxed{0\ ⑥}$, $\boxed{1\ ⑥}$, $\boxed{2\ ⑥\text{'s}}$

Problem: What's the expected number of rolls of a die before "66"?
Here, we derive an absorbing MC trick.



absorbing state: $\boxed{2\ ⑥\text{'s}}$

transient states (non-absorbing): $\boxed{0\ ⑥}$, $\boxed{1\ ⑥}$.

transitional Matrix:

$$P = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & \frac{5}{6} & \frac{1}{6} & 0 \\ 1 & \frac{5}{6} & 0 & \frac{1}{6} \\ 2 & 1 & 0 & 0 \end{array}$$

, where $\begin{bmatrix} 0\to0 & 0\to1 & 0\to2 \\ 1\to0 & 1\to1 & 1\to2 \\ 2\to0 & 2\to1 & 2\to2 \end{bmatrix}$

The transitional matrix for an "absorbing Markov Chain" can be rearranged into this form:

$$P = \begin{array}{c|cc} & \text{transient} & \text{absorbing} \\ \hline \text{transient} & Q & R \\ \text{absorbing} & 0 & I \end{array}$$

, where Q and R are submatrices, I is the identity matrix, "transient" = transient states, "absorbing" = absorbing states.

P^n:

Recall that P_{ij}^n is the prob of being in state j after n steps, starting at state i.

(so, $i \to j$ in n steps)

$$P^n = \begin{array}{c|cc} & \text{trans} & \text{abs} \\ \hline \text{transient} & Q^n & ? \\ \text{absorbing} & 0 & I \end{array}$$

, note that Q becomes Q^n

P ↗ $\quad$ P^n ↗

Define: $N = I + Q + Q^2 + \cdots$

s.t. $N_{ij} = P(i \to j \text{ in } 0 \text{ step}) +$
$P(i \to j \text{ in } 1 \text{ step}) +$
$P(i \to j \text{ in } 2 \text{ steps}) + \cdots$
$P(i \to j \text{ in } \infty \text{ steps})$

using $\sum_{k=0}^{\infty} Q^k = (I-Q)^{-1}$, $N = (I-Q)^{-1}$ is the "fundamental matrix" r.v. (indicator)

$X_k = 1$ if in state j after k steps, starting at i.

$= 0$ otherwise

then, $P(X_k = 1) = Q^k_{ij}$

$P(X_k = 0) = 1 - Q^k_{ij}$

$E[i \to j \text{ in } n \text{ steps}]$:

$$E(X_0 + X_1 + \cdots + X_n) = E(X_0) + E(X_1) + \cdots + E(X_n)$$
$$= Q^0_{ij} + Q^1_{ij} + Q^2_{ij} + \cdots + Q^n_{ij}$$

As $n \to \infty$

$$E(X_0 + X_1 + \cdots + X_n) = Q^0_{ij} + Q^1_{ij} + \cdots\cdots = N_{ij}$$

$X_0 + X_1 + \cdots + X_n$: Number of times in state j, given that we started in state i.

So N_{ij} is the expected number of times in transient state j, given that we started in transient state i.

If we add all entries in i^{th} row of N, we get the expected number of times in any of the transient (non-absorbing) states for a given starting state i.
This is equal to expected time before absorption!

$t = N \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, t_i : expected number of steps before chain is absorbed, given that we started in state i

---

Going back to the problem, the matrix
$$\begin{array}{cccc} & 0 & 1 & 2 \\ 0 & \frac{5}{6} & \frac{1}{6} & 0 \\ 1 & \frac{5}{6} & 0 & \frac{1}{6} \\ 2 & 1 & 0 & 0 \end{array}$$
is already in the desired form because <u>0</u> and <u>1</u> are the transient states.

$$\left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{5}{6} & 0 \end{bmatrix}\right]^{-1} = \begin{bmatrix} \frac{1}{6} & -\frac{1}{6} \\ -\frac{5}{6} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 36 & 6 \\ 30 & 6 \end{bmatrix}$$

$$\begin{array}{cccc} & 0 & 1 & 2 \\ 0 & \frac{5}{6} & \frac{1}{6} & 0 \\ 1 & \frac{5}{6} & 0 & \frac{1}{6} \\ 2 & 1 & 0 & 0 \end{array}$$

$$36 + 6 = \boxed{42}$$