Confidence Intervals, Conditional Expectation, LLSE, Markov Chains 12

Computer Science Mentors 70

April 23 - 25, 2017

1 Confidence Intervals

1.1 Questions

1. Define i. i. d. variables $A_k \sim \text{Bern}(p)$ where $k \in [1, n]$. Assume we can declare that $\text{P}[|\frac{1}{n}\sum_k A_k - p| > 0.25] = 0.01$.

   (a) Please give a 99% confidence interval for p if given A_k.

   Solution: $[\frac{1}{n}\sum_i A_k - 0.25, \frac{1}{n}\sum_i A_k + 0.25]$

   (b) We know that the variables X_i, for i from 1 to n, are i.i.d. random variables and have variance σ^2. We also have a value (an observation) of $A_n = \frac{X_1 + \ldots + X_n}{n}$. We want to guess the mean, μ, of each X_i.

   Prove that we have 95% confidence that μ lies in the interval $\big[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}\big]$

   That is, $\text{P}\big[\mu \in \big[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}\big]\big] \geq 95\%$

   Solution: To do this, we use Chebyshev's. Because $\text{E}[A_n] = \mu$ (A_n is the average of the X_is), we bound the probability that $|A_n - \mu|$ is more than the interval size at 5%:

$$P\big[|A_n - \mu| \geq 4.5\frac{\sigma}{\sqrt{n}}\big] \leq \frac{\text{Var}(A_n)}{(4.5\frac{\sigma}{\sqrt{n}})^2} \approx \frac{\frac{\sigma^2}{n}}{\frac{20\sigma^2}{n}} = \frac{1}{20} = 5\%$$

.

Thus, the probability that μ is *in* the interval is 95%.

(c) Give the 99% confidence interval for μ.

Solution: Solution is similar to that of the 95% confidence interval.

$\big[A_n - 10\frac{\sigma}{\sqrt{n}}, A_n + 10\frac{\sigma}{\sqrt{n}}\big]$, because $P\big[|A_n - \mu| \geq 10\frac{\sigma}{\sqrt{n}}\big] \leq \frac{\text{Var}(A_n)}{(10\frac{\sigma}{\sqrt{n}})^2} \approx \frac{\frac{\sigma^2}{n}}{\frac{100\sigma^2}{n}} = \frac{1}{100} = 1\%$.

2. We have a die with 6 faces of values $1, 2, 3, 4, 5, 6$.

(a) Develop a 99% confidence interval for the value of n samples assuming n is large enough.

Solution: Consider the 99% confidence interval for the average of n rolls of a standard die, X. We can model it using the normal distribution with a mean $E[\frac{nD}{n}] = \frac{nE[D]}{n} = E[D] = 3.5$ and a variance $\text{Var}[X] = \text{Var}[\frac{nD}{n}] = (\frac{1}{n})^2\text{Var}[nD] = (\frac{1}{n})^2 n\text{Var}[D] = \frac{\text{Var}[D]}{n} = \frac{35}{12n}$ where D is the distribution of a standard dice roll. We can now find the 99% confidence interval as $|\frac{x - E[X]}{\sqrt{\text{Var}[X]}}| = Z(99\%) = 2.576$. We solve this to find $x = E[X] \pm Z(99\%)\sqrt{\text{Var}[X]} = 3.5 \pm 2.576\sqrt{\frac{35}{12n}}$

(b) Now, we say the die's values are consecutive integers, but we do not know where it starts. It is shifted over by some value k; for example, if $k = 6$, the die faces would take on the values $7, 8, 9, 10, 11, 12$. If we observe that the average of the n samples is 15.5, develop a 99% confidence interval for the value of k assuming n is large enough.

Solution: We can now say that the confidence interval for k is given by 15.5 - confidence interval for X.

2 Conditional Expectation

2.1 Introduction

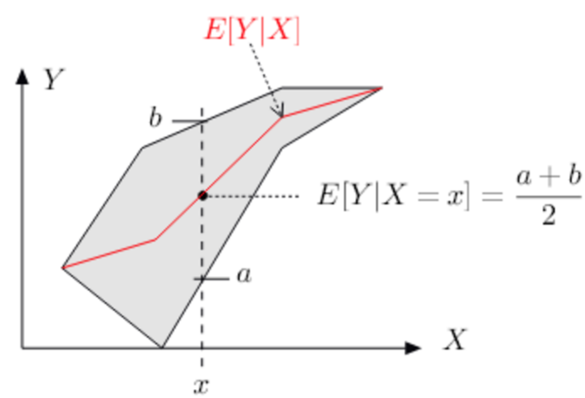
The **conditional expectation** of Y given X is defined by

$$E[Y|X=x] = \sum_y y \cdot P[Y=y|X=x] = \sum_y y \cdot \frac{P[X=x, Y=y]}{P[X=x]}$$

Properties of Conditional Expectation

$$E(a|Y) = a$$
$$E(aX+bZ|Y) = a \cdot E(X|Y) + b \cdot E(Z|Y)$$
$$E(X|Y) \geq 0 \text{ if } X \geq 0$$
$$E(X|Y) = E(X) \text{ if } X, Y \text{ independent}$$
$$E(E(X|Y)) = E(X)$$

Here is a picture that shows that conditioning creates a new random variable with a new distribution, taken from Figure 9 of note 26.



2.2 Questions

1. Consider the random variables Y and X with the following probabilities

   This table gives the probability distribution for $P[X \cap Y]$

		X		
		0	1	2
Y	0	0	.1	.2
	1	.1	.2	.1
	2	.2	.1	0

Find:

(a) $\mathrm{E}(Y|X=0)$

Solution:

$$\begin{aligned}\mathrm{E}(Y|X=0) &= \mathrm{P}[Y=0|X=0]\cdot 0+\mathrm{P}[Y=1|X=0]\cdot 1+\mathrm{P}[Y=2|X=0]\cdot 2\\ &= \frac{0}{0+.1+.2}\cdot 0+\frac{.1}{0+.1+.2}\cdot 1+\frac{.2}{0+.1+.2}\cdot 2\\ &= \frac{20}{12}=\frac{5}{3}\end{aligned}$$

(b) $\mathrm{E}(Y|X=1)$

Solution:

$$\begin{aligned}\mathrm{E}(Y|X=1) &= \mathrm{P}[Y=0|X=1]\cdot 0+\mathrm{P}[Y=1|X=1]\cdot 1+\mathrm{P}[Y=2|X=1]\cdot 2\\ &= \frac{0.1}{0.1+0.2+0.1}\cdot 0+\frac{0.2}{0.1+0.2+0.1}\cdot 1+\frac{0.1}{0.1+0.2+0.1}\cdot 2\\ &= 0+\frac{0.2}{0.4}+\frac{0.1\cdot 2}{0.4}\\ &= 0.5+0.5=1\end{aligned}$$

(c) $\mathrm{E}(Y|X=2)$

Solution:

$$\begin{aligned}\mathrm{E}(Y|X=2) &= \mathrm{P}[Y=0|X=2]\cdot 0+\mathrm{P}[Y=1|X=2]\cdot 1+\mathrm{P}[Y=2|X=2]\cdot 2\\ &= \frac{0.2}{0.2+0.1+0}\cdot 0+\frac{0.1}{0.2+0.1+0}\cdot 1+\frac{0}{0.2+0.1+0}\cdot 2\\ &= 0+\frac{0.1}{0.3}+0=\frac{1}{3}\end{aligned}$$

(d) $\mathrm{E}(Y)$

Solution: These events are disjoint, so to find $E[Y]$, we can just sum up individual probabilities (note that sum of all the probabilities should be 1).

$$\begin{aligned} E(Y) &= E(Y|X=0) \cdot P[X=0] + E(Y|X=1) \cdot P[X=1] + E(Y|X=2) \cdot P[X=2] \\ &= \frac{20}{12} \cdot (0 + 0.1 + 0.2) + 1 \cdot (0.1 + 0.2 + 0.1) + \frac{1}{3} \cdot (0.2 + 0.1 + 0) \\ &= \frac{20}{12} \cdot \frac{3}{10} + 0.4 + \frac{0.3}{3} = 1 \end{aligned}$$

3 Linear Least Squares Estimator

Theorem: Consider two random variables, X, Y with a given distribution $P[X = x, Y = y]$. Then

$$L[Y|X] = E(Y) + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - E(X))$$

3.1 Questions

1. Assume that

$$Y = \alpha X + Z$$

where X and Z are independent and $E(X) = E(Z) = 0$. Find $L[X|Y]$.

Solution:

$$\begin{aligned} \text{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\ &= E(X \cdot (\alpha X + Z)) = \alpha E(X^2) \end{aligned}$$

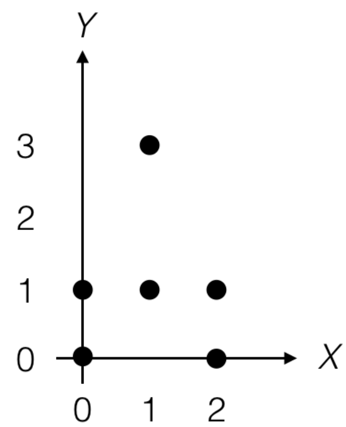
$$\begin{aligned} \text{Var}(Y) &= \alpha^2 \text{Var}(X) + \text{Var}(Z) \\ &= \alpha^2 E(X^2) + E(Z^2) \end{aligned}$$

Therefore,

$$L[X|Y] = \frac{\alpha E(X^2)}{\alpha^2 E(X^2) + E(Z^2)} \cdot Y$$

2. The figure below shows the six equally likely values of the random pair (X, Y). Specify the functions of:

   - $L[Y \mid X]$
   - $E(X \mid Y)$
   - $L[X \mid Y]$
   - $E(Y \mid X)$



Solution: Let's calculate some useful properties of the distribution first and then see how we can use them to calculate the estimates.

$$
\begin{aligned}
|\Omega| = 6 &\implies P[\text{one point}] = \frac{1}{6} \\
\mathrm{E}(X) &= 0\left(\frac{2}{6}\right) + 1\left(\frac{2}{6}\right) + 2\left(\frac{2}{6}\right) \\
&= 1 \\
\mathrm{E}(Y) &= 0\left(\frac{2}{6}\right) + 1\left(\frac{3}{6}\right) + 30\left(\frac{1}{6}\right) \\
&= 1 \\
\mathrm{E}(XY) &= 0\left(\frac{3}{6}\right) + 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) \\
&= 1 \\
\mathrm{Cov}(X, Y) = 0 &\implies \mathrm{L}[Y|X] = E(Y)
\end{aligned}
$$

- $L[Y \mid X]$: Using the LLSE formula: $\mathrm{L}[Y \mid X] = \mathrm{E}[Y] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathrm{E}[Y]) = \mathrm{E}[Y]$. Therefore $\boxed{L[Y \mid X] = 1}$

- $E[X \mid Y]$: Notice the symmetry across $X = 1$. For all values of y, $E[X|Y = y]$ is the same; therefore $\boxed{E[X|Y] = E[X] = 1}$.
- $L[X \mid Y]$: The MMSE estimator for X given Y is a linear function, therefore $\boxed{\mathrm{L}[X \mid Y] = \mathrm{E}[X \mid Y] = 1}$
- $E[Y \mid X]$ For this one we can't make use of symmetry or directly apply what we calculated above. We must go back to the definition of conditional expectation. We can calculate $\mathrm{E}[Y \mid X = x]$ for every point x, and that entirely defines the expression:

$$\mathrm{E}(Y \mid X = x) = \begin{cases} \frac{1}{2} & \text{if } x = 0 \\ 2 & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 2 \end{cases}$$

The above equation is sufficient, but we can go further by realizing that these points are part of a flipped absolute value function centered around $x = 1$: $\boxed{\mathrm{E}[Y \mid X] = \frac{-3}{2}|X - 1| + 2}$. Indeed, this is not linear, which is why $\mathrm{L}[Y \mid X] \neq \mathrm{E}[Y \mid X]$.

4 Markov Chains

P is a **transition probability matrix** if:

1. All of the entries are non-negative.
2. The sum of entries in each row is 1.

A **Markov chain** is defined by four things: ($\mathscr{X}$, π_0, P, $\{X_n\}_{n=0}^{\infty}$)

$\mathscr{X}$	Set of states
π_0	Initial probability distribution
P	Transition probability matrix
$\{X_n\}_{n=0}^{\infty}$	Sequence of random variables where:

$$\mathrm{P}[X_0 = i] = \pi_0(i), i \in \mathscr{X}$$
$$\mathrm{P}[X_{n+1} = j | X_n = i, X_{n-1}, \ldots, X_0] = \mathrm{P}(i, j), \forall n \geq 0, \forall i, j \in \mathscr{X}$$

A Markov chain is **irreducible** if we can go from any state to any other state, possibly in multiple steps.

Periodicity has to do with the period of occurrence of a state. If a state s has period 2, the Markov chain can be in s at every other time point. If a state has period 1, it's aperiodic; otherwise, it's periodic.
More quantitatively, define value $d(i)$ for each state i as:

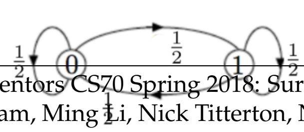
$$d(i) := g.c.d\{n > 0 | P^n(i,i) = \mathrm{P}[X_n = i | X_0 = i] > 0\}, i \in \mathscr{X}$$

If $d(i) = 1$, then the Markov chain is **aperiodic**. If $d(i) \neq 1$, then the Markov chain is periodic and its **period** is $d(i)$.

A distribution π is **invariant** for the transition probability P if it satisfies the following **balance equations**

$$\pi \cdot P = \pi.$$

Theorem 24.3: A finite irreducible Markov chain has a unique invariant distribution.
Theorem 24.4: All irreducible and aperiodic Markov chains converge to the unique invariant distribution. If a Markov chain is finite and reducible, the amount of time spent in each state approaches the invariant distribution as n grows large
Equations that model what will happen at the next step are called **first step equations**



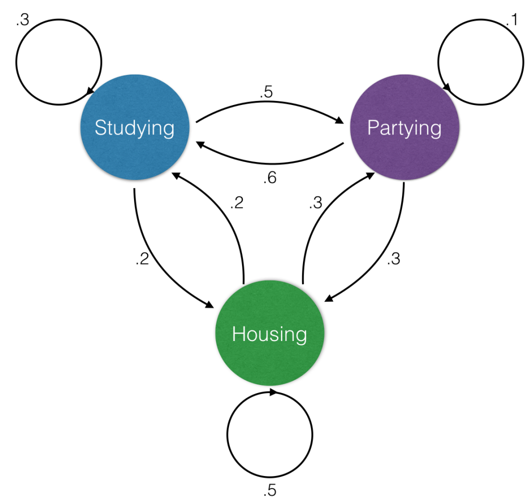
Denote $\beta(i,j)$ as the expected amount of time it would take to move from i to j. $\beta(0,1) = 1 + \frac{1}{2} \cdot \beta(0,1)$ $\beta(1,1) = 0$

4.1 Questions

1. **Life of Alex**

   Alex is enjoying college life. She spends a day either studying, partying, or looking for housing for the next year. If she is studying, the chances of her studying the next day are 30%, the chances of her partying the next day are 50%, and the chances of her looking for housing the next day are 20%. If she is partying, the chances of her partying the next day are 10%, the chances of her studying the next day are 60%, and the chances of her looking for housing the next day are 30%. If she is looking for housing, the chances of her looking for housing the next day are 50%, the chances of her partying the next day are 30% and the chances of her studying the next day are 20%.

   (a) Draw a Markov chain to visualize Alexs life.

   Solution:

   (b) Write out a matrix to represent this Markov chain.

   Solution:

$$\begin{bmatrix} .3 & .5 & .2 \\ .6 & .1 & .3 \\ .2 & .3 & .5 \end{bmatrix}$$

   (c) If Alex studies on Monday, what is the chance that she is partying on Friday? (Don't do the math, just write out the expression that you would use to find it.)

Solution: If P is the matrix above, then it is $[1, 0, 0] \cdot P^4$

(d) What percentage of her time should Alex expect to use looking for housing?

> **Solution:** Solve the following system of equations: (first step equations)
>
> $$S = .3S + .6P + .2H$$
> $$P = .5S + .1P + .3H$$
> $$H = .2S + .3P + .5H$$
> $$S + P + H = 1$$

(e) If Alex parties on Monday, what is the chance of Alex partying again before studying?

> **Solution:** Set up the following equations:
>
> $$H1 = 0$$
> $$H2 = .6(H1) + .1(1) + .3(H3)$$
> $$H3 = .2(H1) + .3(1) + .5(H3)$$
>
> Solving for $H2$, we get 0.28

2. **Stanford Cinema**
   You have a database of an infinite number of movies. Each movie has a rating that is uniformly distributed in 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 independent of all other movies. You want to find two movies such that the sum of their ratings is greater than 7.5 (7.5 is not included).

   a) A Stanford student chooses two movies each time and calculates the sum of their ratings. If is less than or equal to 7.5, the student throws away these two movies and chooses two other movies. The student stops when he/she finds two movies such that the sum of their ratings is greater than 7.5. What is the expected number of movies that this student needs to choose from the database?

> **Solution:** Each time when the Stanford student chooses two movies, there are $11^2 = 121$ different possible pairs of ratings. By simple counting, we know that there are 15 pairs whose sum is greater than 7.5. Therefore, the probability that in a single trial, the Stanford student gets two movies such that the sum of their ratings is greater than 7.5 is $\frac{15}{121}$. Then the number of times that the student needs to pick movies is geometrically distributed with mean $\frac{121}{15}$. Then the expected number of movies that the student needs to choose is $\frac{242}{15} \approx 16.13$.

b) A Berkeley student chooses movies from the database one by one and keeps the movie with the highest rating. The student stops when he/she finds the sum of the ratings of the last movie that he/she has chosen and the movie with the highest rating among all the previous movies is greater than 7.5. What is the expected number of movies that the student will have to choose?

Solution: We use a Markov chain to represent to process that the Berkeley student gets two desired movies. There are 11 possible ratings:

$$S = \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$$

We divide the set of ratings into 2 subsets, $L = \{0, 0.5, 1, 1.5, 2, 2.5\}$ and $H = \{3, 3.5, 4, 4.5, 5\}$. Since the goal of the student is to get two movies that the sum of their rating is greater than 7.5, the movies whose ratings are in the set L have no contribution to this goal. Then we can use 7 states $\{s_L, s_3, s_{3.5}, s_4, s_{4.5}, s_5, s_E\}$ to represent the progress to get two movies such that the sum of their ratings is greater than 7.5. The state s_L denotes the cases when highest movie rating that the Berkeley student has got is in L. The states $s_i, 3 \leq i \leq 5$, denote the cases when the highest movie rating that the student has got is i. The state s_E denotes the case when the student has got two movies such the sum of their ratings is greater than 7.5 and the choosing process ends. We can see that the process is a Markov chain with probability transition matrix P as follows:

$$P = \frac{1}{11}\begin{array}{cc} \begin{bmatrix} 6 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 7 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 8 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 8 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 7 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 6 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 11 \end{bmatrix} & \begin{array}{c} s_L \\ s_3 \\ s_{3.5} \\ s_4 \\ s_{4.5} \\ s_5 \\ s_E \end{array} \\ \begin{array}{cccccc} s_L & s_3 & s_4 & s_{4.5} & s_5 & s_E \end{array} & \end{array}$$

Let F_s be the expected time to get to state s_E, starting from state s, $F = \begin{bmatrix} F_{s_L} & F_{s_3} & F_{s_{3.5}} & F_{s_4} & F_{s_{4.5}} & F_{s_5} \end{bmatrix}^T$ P' be the sub-matrix of P consisting of the first 6 columns and rows of P (which has 7), and U be an all-one column vector with length 6. We have the first step equations

$$F = P'F + U$$

Solving these linear equations, we get
$F = \begin{bmatrix} 6.0164 & 5.5764 & 4.8889 & 3.666667 & 2.75 & 2.2 \end{bmatrix}^T$
The initial state is s_L with probability $\frac{6}{11}$ and is $s_i, 3 \le i \le 5$ with probability $\frac{1}{11}$ respectively. Then we know that the expected number of movies that the student needs to choose is

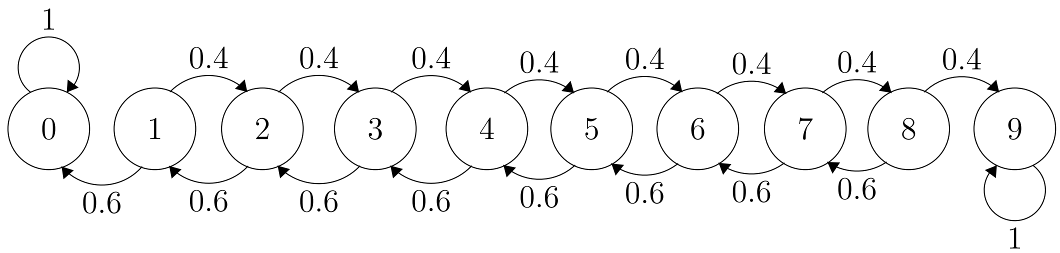
$$1 + \frac{6}{11}F_{s_L} + \frac{1}{11}(F_{s_3} + F_{s_{3.5}} + F_{s_4} + F_{s_{4.5}} + F_{s_5}) = 6.02$$

This shows that the Berkeley student is smarter than the Stanford student.

3. **Bet On It**
Smith is in jail and has 3 dollars; he can get out on bail if he has 8 dollars. A guard agrees to make a series of bets with him. If Smith bets A dollars, he wins A dollars with probability 0.4 and loses A dollars with probability 0.6.

a) Find the probability that he wins 8 dollars before losing all of his money if he bets 1 dollar each time.

Solution: The Markov chain $(X_n, n = 0, 1, \dots)$ representing the evolution of Smith's money has diagram



Let $\phi(i)$ be the probability that the chain reaches state 8 before reaching state 0, starting from state i. In other words, if S_j is the first $n \leq 0$ such that $X_n = j$,

$$P_i(S_8 < S_0) = P(S_8 < S_0 | X_0 = i) \tag{1}$$

Using first-step analysis (viz. the Markov property at time $n = 1$), we have

$$\begin{aligned} &\phi(i) = 0.4\phi(i+1) + 0.6\phi(i-1),\ i = 1,2,3,4,5,6,7 \\ &\phi(0) = 0 \\ &\phi(8) = 1. \end{aligned}$$

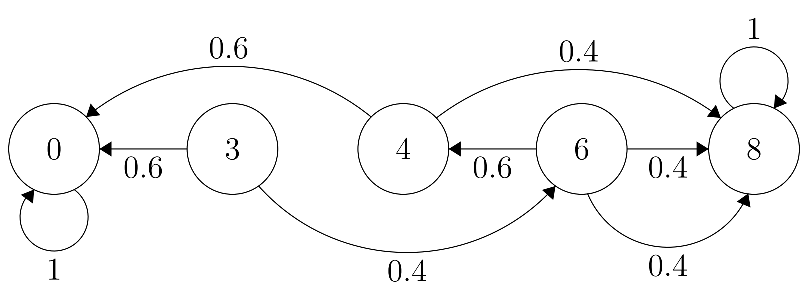
We solve this system of linear equations and find

$$\begin{aligned} \phi &= (\phi(1), \phi(2), \phi(3), \phi(4), \phi(5), \phi(6), \phi(7)) \\ &= (0.0203, 0.0508, 0.0964, 0.1649, 0.2677, 0.4219, 0.6531, 1). \end{aligned}$$

E.g. the probability that the chain reaches state 8 before reaching state 0, starting from state 3 is the third component of the vector and is equal to 0.0964. Note that $\phi(i)$ is increasing in i, which was expected.

b) Find the probability that he wins 8 dollars before losing all of his money if he bets, each time, as much as possible but not more than necessary to bring his fortune up to 8 dollars

Solution: Now the chain is



and the equations are:

$$\phi(3) = 0.4\phi(6)$$
$$\phi(6) = 0.4\phi(8) + 0.6\phi(4)$$
$$\phi(4) = 0.4\phi(8)$$
$$\phi(0) = 0$$
$$\phi(8) = 1$$

We solve and find

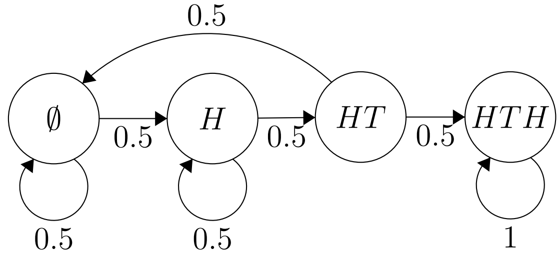
$$\phi(3) = 0.256, \phi(4) = 0.4, \phi(6) = 0.64$$

c) Which strategy gives Smith the better chance of getting out of jail?

Solution: By comparing the third components of the vector ϕ we find that the bold strategy gives Smith a better chance to get our of jail.

4. **Tossing Coins**
A fair coin is tossed repeatedly and independently. Find the expected number of tosses till the pattern HTH appears.

Solution: Call HTH our target. Consider a chain that starts from a state called nothing ($\emptyset$) and is eventually absorbed at HTH. If we first toss H then we move to state H because this is the first letter of our target. If we toss a T then we move back to $\emptyset$ having expended 1 unit of time. Being in state H we either move to a new state HT if we bring T and we are 1 step closer to the target or, if we bring H, we move back to H: we have expended 1 unit of time, but the new H can be the beginning of a target. When in state HT we either move to HTH and we are done or, if T occurs then we move to $\emptyset$. The transition diagram is



Let $\phi(i)$ be the expected number of steps to reach HTH starting from i. We have

$$\phi(HT) = 1 + \frac{1}{2}\phi(\emptyset)$$
$$\phi(H) = 1 + \frac{1}{2}\phi(H) + \frac{1}{2}\phi(HT)$$
$$\phi(\emptyset) = 1 + \frac{1}{2}\phi(\emptyset) + \frac{1}{2}\phi(H)$$

We solve and find $\phi(\emptyset) = 10$.