Markov Chains

CS70 Summer 2016 - Lecture 6B

Grace Dinh 26 July 2016

UC Berkeley

Agenda

Quiz is out! Due: Friday at noon.

What are Markov Chains? State machine and matrix representations.

Hitting Time



Suppose we flip a coin until we get a three heads in a row. How many coin flips should we expect to do?

Suppose we flip a coin until we get a three heads in a row. How many coin flips should we expect to do?

Drunkard on an arbitrary graph (remember HW?). When does the drunkard come home?

Suppose we flip a coin until we get a three heads in a row. How many coin flips should we expect to do?

Drunkard on an arbitrary graph (remember HW?). When does the drunkard come home?

Try solving directly? Problem: conditioning gets really messy.

Suppose we flip a coin until we get a three heads in a row. How many coin flips should we expect to do?

Drunkard on an arbitrary graph (remember HW?). When does the drunkard come home?

Try solving directly? Problem: conditioning gets really messy.

Need some way to express **state**.

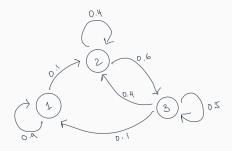
Suppose we flip a coin until we get a three heads in a row. How many coin flips should we expect to do?

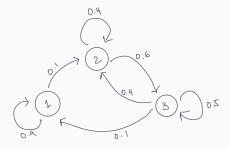
Drunkard on an arbitrary graph (remember HW?). When does the drunkard come home?

Try solving directly? Problem: conditioning gets really messy.

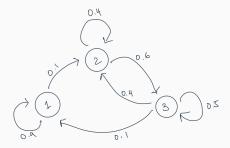
Need some way to express **state**.

Solution: Markov chains!



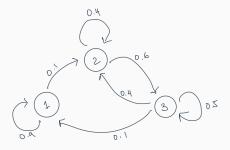


A finite Markov chain consists of states, transition probabilities between states, and an initial distribution.



A finite Markov chain consists of states, transition probabilities between states, and an initial distribution.

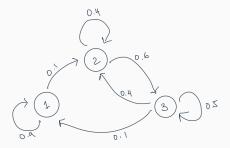
State: where are you now?



A finite Markov chain consists of states, transition probabilities between states, and an initial distribution.

State: where are you now?

Transition probability: From where you are, where do you go next?

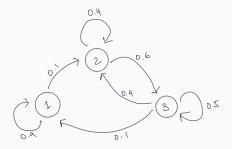


A finite Markov chain consists of states, transition probabilities between states, and an initial distribution.

State: where are you now?

Transition probability: From where you are, where do you go next?

Initial distribution: how do you start?



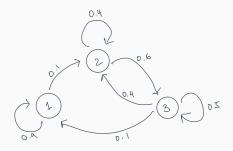
A finite Markov chain consists of states, transition probabilities between states, and an initial distribution.

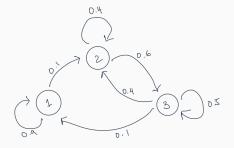
State: where are you now?

Transition probability: From where you are, where do you go next?

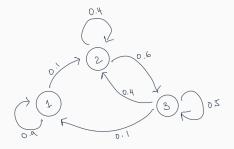
Initial distribution: how do you start?

Markov chains are **memoryless** - they don't remember anything other than what state they are.



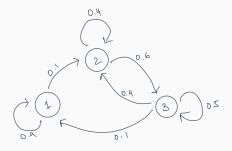


A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

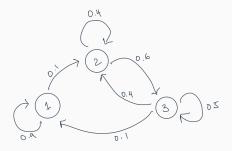


A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

A initial probability distribution π_0 on $\mathcal X$:



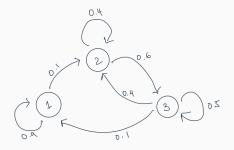
- A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$
- A initial probability distribution π_0 on $\mathcal{X}:\pi_0(i)\geq 0, \sum_i \pi_0(i)=1$



A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

A initial probability distribution π_0 on $\mathcal{X}:\pi_0(i)\geq 0, \sum_i \pi_0(i)=1$

Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

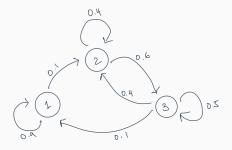


A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

A initial probability distribution π_0 on $\mathcal{X}: \pi_0(i) \geq 0, \sum_i \pi_0(i) = 1$

Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

•
$$P(i,j) \geq 0, \forall i,j;$$

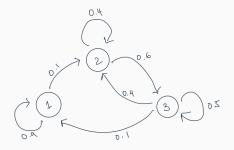


A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

A initial probability distribution π_0 on $\mathcal{X}: \pi_0(i) \geq 0, \sum_i \pi_0(i) = 1$

Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

•
$$P(i,j) \ge 0, \forall i,j; \sum_{j} P(i,j) = 1, \forall i$$



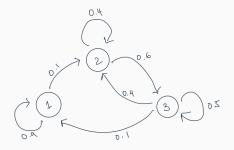
A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

A initial probability distribution π_0 on $\mathcal{X}:\pi_0(i)\geq 0, \sum_i \pi_0(i)=1$

Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

•
$$P(i,j) \ge 0, \forall i,j; \sum_{j} P(i,j) = 1, \forall i$$

 $\{X_n, n \ge 0\}$ is defined so that:



A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

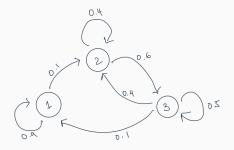
A initial probability distribution π_0 on $\mathcal{X}:\pi_0(i)\geq 0, \sum_i \pi_0(i)=1$

Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

•
$$P(i,j) \ge 0, \forall i,j; \sum_{j} P(i,j) = 1, \forall i$$

 $\{X_n, n \ge 0\}$ is defined so that:

•
$$Pr[X_0 = i] = \pi_0(i), i \in \mathcal{X}$$



A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

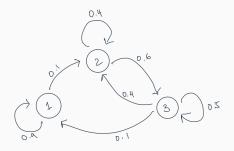
A initial probability distribution π_0 on $\mathcal{X}:\pi_0(i)\geq 0, \sum_i \pi_0(i)=1$

Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

•
$$P(i,j) \ge 0, \forall i,j; \sum_{j} P(i,j) = 1, \forall i$$

 $\{X_n, n \ge 0\}$ is defined so that:

•
$$Pr[X_0 = i] = \pi_0(i), i \in \mathcal{X}$$
 (initial distribution)



A finite set of states: $\mathcal{X} = \{1, 2, \dots, K\}$

A initial probability distribution π_0 on $\mathcal{X}: \pi_0(i) \geq 0, \sum_i \pi_0(i) = 1$

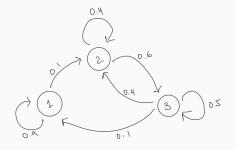
Transition probabilities: P(i,j) for $i,j \in \mathcal{X}$

•
$$P(i,j) \ge 0, \forall i,j; \sum_{j} P(i,j) = 1, \forall i$$

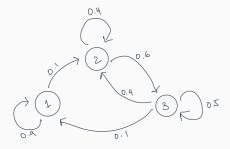
 $\{X_n, n \ge 0\}$ is defined so that:

•
$$Pr[X_0 = i] = \pi_0(i), i \in \mathcal{X}$$
 (initial distribution)

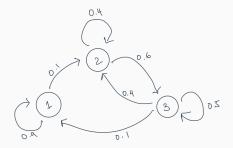
•
$$Pr[X_{n+1} = j \mid X_0, ..., X_n = i] = P(i, j), i, j \in \mathcal{X}.$$



At each timestep t we are in some state $X_t \in \mathcal{X}$.

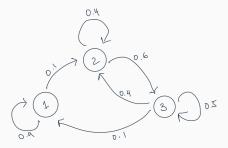


At each timestep t we are in some state $X_t \in \mathcal{X}$. (random variable.)



At each timestep t we are in some state $X_t \in \mathcal{X}$. (random variable.) Where do we go next?

$$Pr[X_{t+1} = j | X_t = i] = P_{i,j}$$

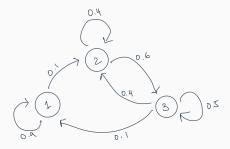


At each timestep t we are in some state $X_t \in \mathcal{X}$. (random variable.)

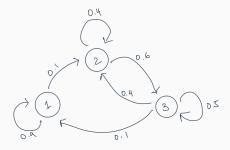
Where do we go next?

$$\Pr[X_{t+1} = j | X_t = i] = P_{i,j}$$

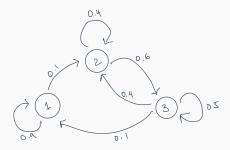
Probability depends on the previous state, but is independent of how it got to the previous state. (It's not independent of states before the previous state - but any dependence is captured in the previous state.)



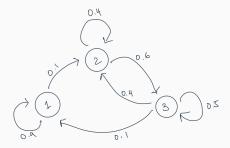
At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5.



At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ?

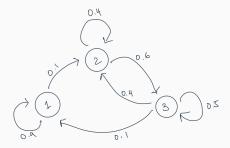


At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ? Probability that it goes to 1?



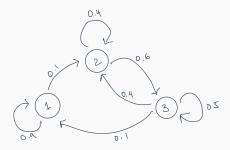
At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ? Probability that it goes to 1?

$$Pr[X_{t+1} = 1] =$$



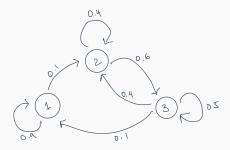
At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ? Probability that it goes to 1?

$$Pr[X_{t+1} = 1] = \sum_{i} Pr[X_{t+1} = 1 | X_t = i] Pr[X_t = i]$$



At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ? Probability that it goes to 1?

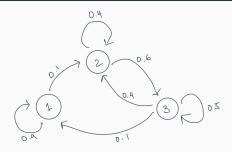
$$\Pr[X_{t+1} = 1] = \sum_{i} \Pr[X_{t+1} = 1 | X_t = i] \Pr[X_t = i] = \sum_{i} P_{i,1} \Pr[X_t = i]$$



At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ? Probability that it goes to 1?

$$Pr[X_{t+1} = 1] = \sum_{i} Pr[X_{t+1} = 1 | X_t = i] Pr[X_t = i] = \sum_{i} P_{i,1} Pr[X_t = i]$$

$$= 0.9 * 0.2 + 0 * 0.3 + 0.1 * 0.5$$



At some point we might have a distribution for X_t - say, it's 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Distribution for X_{t+1} ? Probability that it goes to 1?

$$Pr[X_{t+1} = 1] = \sum_{i} Pr[X_{t+1} = 1 | X_t = i] Pr[X_t = i] = \sum_{i} P_{i,1} Pr[X_t = i]$$

$$= 0.9 * 0.2 + 0 * 0.3 + 0.1 * 0.5 = 0.23$$

Rest of distribution for X_{t+1} can be found similarly.

Very quick linear algebra intro:

Very quick linear algebra intro:

Matrices: two-dimensional collection of numbers. $n \times m$ matrix has n rows, m columns. Element at ith row, jth column denoted A_{ij} .

Very quick linear algebra intro:

Matrices: two-dimensional collection of numbers. $n \times m$ matrix has n rows, m columns. Element at ith row, jth column denoted A_{ij} .

Very quick linear algebra intro:

Matrices: two-dimensional collection of numbers. $n \times m$ matrix has n rows, m columns. Element at ith row, jth column denoted A_{ij} .

Vector: one-dimensional collection of numbers. We deal with **row vectors** - $n \times 1$ matrices.

Very quick linear algebra intro:

Matrices: two-dimensional collection of numbers. $n \times m$ matrix has n rows, m columns. Element at ith row, jth column denoted A_{ij} .

Vector: one-dimensional collection of numbers. We deal with **row vectors** - $n \times 1$ matrices.

$$\begin{bmatrix} 5 & 9 & 3 & 0 \end{bmatrix}$$

Matrix Multplication

For $n \times m$ matrix A and $m \times p$ matrix B:

$$(AB)_{ij} = \sum_{k} A_{ik} B_{kj}$$

Matrix Multplication

For $n \times m$ matrix A and $m \times p$ matrix B:

$$(AB)_{ij} = \sum_{k} A_{ik} B_{kj}$$

Or for vector x:

$$(xA)_i = \sum_k x_k A_{ki}$$

$$\begin{bmatrix} 5 & 9 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 6 & 7 & 2 \\ 6 & 5 & 6 & 3 \\ 8 & 6 & 2 & 2 \\ 2 & 5 & 3 & 8 \end{bmatrix} = \begin{bmatrix} 1*5+6*9+8*3+2*0 \\ 6*5+5*9+6*3+5*0 \\ 7*5+6*9+2*3+3*0 \\ 2*5+3*9+2*3+8*0 \end{bmatrix}^{T}$$

Matrix Multplication

For $n \times m$ matrix A and $m \times p$ matrix B:

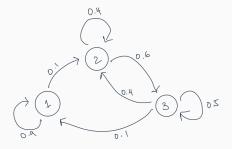
$$(AB)_{ij} = \sum_{k} A_{ik} B_{kj}$$

Or for vector x:

$$(xA)_i = \sum_k x_k A_{ki}$$

$$\begin{bmatrix} 5 & 9 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 6 & 7 & 2 \\ 6 & 5 & 6 & 3 \\ 8 & 6 & 2 & 2 \\ 2 & 5 & 3 & 8 \end{bmatrix} = \begin{bmatrix} 1*5+6*9+8*3+2*0 \\ 6*5+5*9+6*3+5*0 \\ 7*5+6*9+2*3+3*0 \\ 2*5+3*9+2*3+8*0 \end{bmatrix}^{T}$$

Matrix Markov

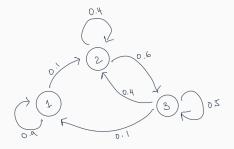


Markov chains have a very nice translation to matrices! Transition probabilities form an *transition matrix* P whose i, jth entry is $P_{i,j}$.

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Probabilities from a state sum to 1...

Matrix Markov

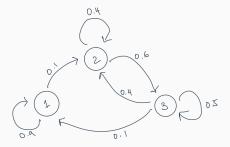


Markov chains have a very nice translation to matrices! Transition probabilities form an *transition matrix* P whose i, jth entry is $P_{i,j}$.

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Probabilities from a state sum to 1...rows sum to 1...

Matrix Markov



Markov chains have a very nice translation to matrices! Transition probabilities form an *transition matrix* P whose i, jth entry is $P_{i,j}$.

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Probabilities from a state sum to 1...rows sum to 1... (right) stochastic matrix

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are vectors. Suppose that X_t is distributed 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Write distribution as vector!

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are vectors. Suppose that X_t is distributed 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Write distribution as vector!

$$\pi_t = \begin{bmatrix} 0.2 & 0.3 & 0.5 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are vectors. Suppose that X_t is distributed 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Write distribution as vector!

$$\pi_t = \begin{bmatrix} 0.2 & 0.3 & 0.5 \end{bmatrix}$$

What's the product of π_t and P?

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are vectors. Suppose that X_t is distributed 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Write distribution as vector!

$$\pi_t = \begin{bmatrix} 0.2 & 0.3 & 0.5 \end{bmatrix}$$

What's the product of π_t and P?

$$\begin{bmatrix} 0.2*0.9 + 0.3*0 + 0.5*0.1 \\ 0.2*0.1 + 0.3*0.4 + 0.5*0.4 \\ 0.2*0 + 0.3*0.6 + 0.5*0.5 \end{bmatrix}^{T}$$

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are vectors. Suppose that X_t is distributed 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Write distribution as vector!

$$\pi_t = \begin{bmatrix} 0.2 & 0.3 & 0.5 \end{bmatrix}$$

What's the product of π_t and P?

$$\begin{bmatrix} 0.2 * 0.9 + 0.3 * 0 + 0.5 * 0.1 \\ 0.2 * 0.1 + 0.3 * 0.4 + 0.5 * 0.4 \\ 0.2 * 0 + 0.3 * 0.6 + 0.5 * 0.5 \end{bmatrix}^{T} = \begin{bmatrix} 0.23 & 0.34 & 0.43 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are vectors. Suppose that X_t is distributed 1 w.p. 0.2, 2 w.p. 0.3, and 3 w.p. 0.5. Write distribution as vector!

$$\pi_t = \begin{bmatrix} 0.2 & 0.3 & 0.5 \end{bmatrix}$$

What's the product of π_t and P?

$$\begin{bmatrix} 0.2 * 0.9 + 0.3 * 0 + 0.5 * 0.1 \\ 0.2 * 0.1 + 0.3 * 0.4 + 0.5 * 0.4 \\ 0.2 * 0 + 0.3 * 0.6 + 0.5 * 0.5 \end{bmatrix}^{T} = \begin{bmatrix} 0.23 & 0.34 & 0.43 \end{bmatrix}$$

This is the distribution of X_{t+1} .

One step: $\pi_t \to \pi_t P$

One step: $\pi_t \to \pi_t P$

What if we take two steps? What's the distribution?

One step: $\pi_t \to \pi_t P$

What if we take two steps? What's the distribution?

$$\pi_t \rightarrow (\pi_t P) P$$

One step: $\pi_t \to \pi_t P$

What if we take two steps? What's the distribution?

$$\pi_t \rightarrow (\pi_t P)P = \pi_t P^2$$

```
One step: \pi_t \to \pi_t P
```

What if we take two steps? What's the distribution?

$$\pi_t \rightarrow (\pi_t P)P = \pi_t P^2$$

n steps? $\pi_t P^n$.

One step: $\pi_t \to \pi_t P$

What if we take two steps? What's the distribution?

$$\pi_t \rightarrow (\pi_t P)P = \pi_t P^2$$

n steps? $\pi_t P^n$.

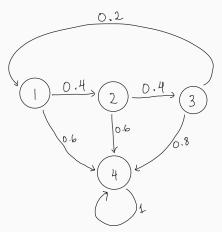
This will be very useful when we start talking about limiting distributions (next lecture).

An Example

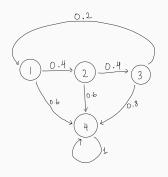
California driving test: you get 3 retakes before you have to start the application process all over again. Suppose someone passes a driving test w.p. 0.6, unless it's their final retake, in which case they're more careful and pass w.p. 0.8.

An Example

California driving test: you get 3 retakes before you have to start the application process all over again. Suppose someone passes a driving test w.p. 0.6, unless it's their final retake, in which case they're more careful and pass w.p. 0.8.



An Example

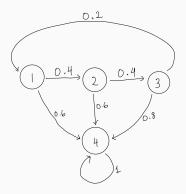


Initial distribution? $\pi_0 = [1 \ 0 \ 0 \ 0]$ Transition matrix?

$$T = \begin{bmatrix} 0 & .4 & 0 & .6 \\ 0 & 0 & .4 & .6 \\ .2 & 0 & 0 & .8 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

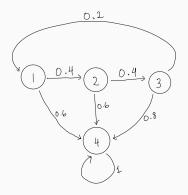
Hitting Time

Motivation



How long does it take to get a driver's license, in expectation?

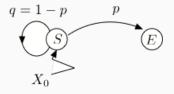
Motivation



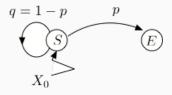
How long does it take to get a driver's license, in expectation?

Generally: given a Markov chain and an initial distribution, how many timesteps do we expect to take before reaching a particular state?

Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?

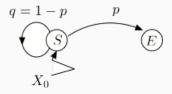


Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E, starting from S.

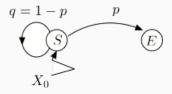
Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E, starting from S. Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

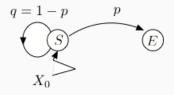
Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E, starting from S. Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



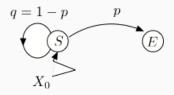
Let $\beta(S)$ be the average time until E, starting from S. Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

Hence,

$$p\beta(S)=1,$$

Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E, starting from S. Then,

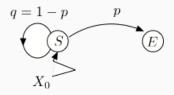
$$\beta(S) = 1 + q\beta(S) + p0.$$

Hence,

$$p\beta(S) = 1$$
, so that $\beta(S) = 1/p$.

A Simple Example

Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E, starting from S. Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

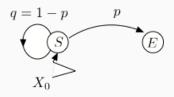
Hence,

$$p\beta(S) = 1$$
, so that $\beta(S) = 1/p$.

Note: Time until E is G(p).

A Simple Example

Let's flip a coin with Pr[H] = p until we get H. How many flips, on average?



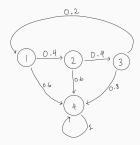
Let $\beta(S)$ be the average time until E, starting from S. Then,

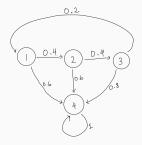
$$\beta(S) = 1 + q\beta(S) + p0.$$

Hence,

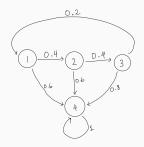
$$p\beta(S) = 1$$
, so that $\beta(S) = 1/p$.

Note: Time until E is G(p). We have rediscovered that the mean of G(p) is 1/p.





Let $\beta(S)$ denote expected time to get a driver's license from S.

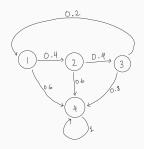


Let $\beta(S)$ denote expected time to get a driver's license from S.

$$\beta(1) = 1 + 0.6 * 0 + 0.4 * \beta(2)$$

$$\beta(2) = 1 + 0.6 * 0 + 0.4 * \beta(3)$$

$$\beta(3) = 1 + 0.8 * 0 + 0.2 * \beta(1)$$



Let $\beta(S)$ denote expected time to get a driver's license from S.

$$\beta(1) = 1 + 0.6 * 0 + 0.4 * \beta(2)$$

$$\beta(2) = 1 + 0.6 * 0 + 0.4 * \beta(3)$$

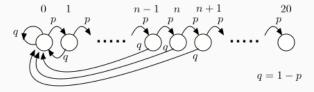
$$\beta(3) = 1 + 0.8 * 0 + 0.2 * \beta(1)$$

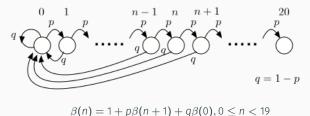
Solves to $\beta(1) \approx 1.61$.

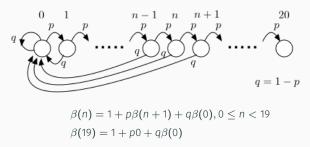
A driving test consists of 20 maneuvers that must be done properly.

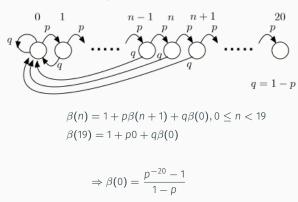
A driving test consists of 20 maneuvers that must be done properly. The examinee succeeds w.p. p=0.9 for each maneuver.

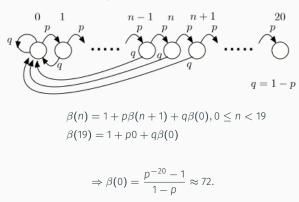
A driving test consists of 20 maneuvers that must be done properly. The examinee succeeds w.p. p=0.9 for each maneuver. Otherwise, he fails the driving test and has to start all over again.



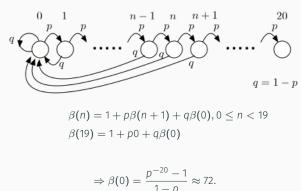








A driving test consists of 20 maneuvers that must be done properly. The examinee succeeds w.p. p=0.9 for each maneuver. Otherwise, he fails the driving test and has to start all over again. How many maneuvers does it take to pass the test?



See Lecture Note 24 for algebra.

Gig: Random names, random headlines