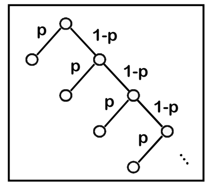
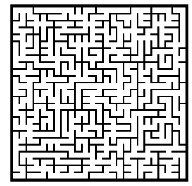
15-251

Great Theoretical Ideas in Computer Science

Infinite Sample spaces and Random Walks

Lecture 12 (October 4, 2007)





Probability Refresher

What's a Random Variable? $X: S \to \mathbb{R}$

A Random Variable is a real-valued function on a sample space S

E[X+Y] = E[X] + E[Y]

$$E[X] = \sum_{e \in S} \Pr[e]\, X(e)$$

Probability Refresher

What does this mean: E[X | A]?

$$E[X|A] = \sum_{e \in S} \Pr[e|A] \cdot X(e)$$

$$= \sum_{e \in A} \Pr[e|A]\, X(e)$$

Probability Refresher

What does this mean: E[X | A]?

Is this true:

Pr[A] = Pr[A | B] Pr[B] + Pr[A | $\overline{B}$] Pr[$\overline{B}$]

Yes!

$$\frac{\Pr[A \cap B]}{\Pr[B]} \cdot \Pr[B] + \frac{\Pr[A \cap \overline{B}]}{\Pr[\overline{B}]} \cdot \Pr[\overline{B}]$$

$$\Pr[A \cap B] + \Pr[A \cap \overline{B}] = \Pr[A]$$

Probability Refresher

What does this mean: E[X | A]?

Is this true:

Pr[A] = Pr[A | B] Pr[B] + Pr[A | $\overline{B}$] Pr[$\overline{B}$]

Yes!

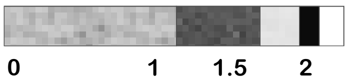
Similarly:

E[X] = E[X | A] Pr[A] + E[X | $\overline{A}$] Pr[$\overline{A}$]

An easy question

What is $\sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i$?

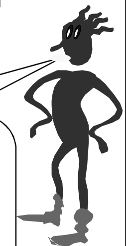
Answer: 2



No, by $\sum_{i=0}^{\infty} f(i)$, we really mean $\lim_{n\to\infty} \sum_{i=0}^{n} f(i)$.

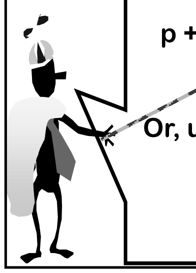
[if this is undefined, so is the sum]

In this case, the partial sum is $2-(1/2)^n$ which goes to 2.



A related question

A related question



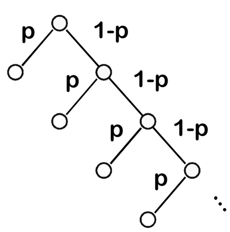
Pr(flip once) + Pr(flip 2 times) + Pr(flip 3 times) + ... = 1:

$$p + (1-p)p + (1-p)^2p + (1-p)^3p + \ldots = 1.$$

Or, using $q = 1-p$,

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$$

Pictorial view

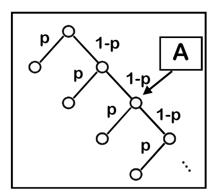


Sample space S = leaves in this tree.

Pr(x) = product of edges on path to x.

If p>0, Pr(not halting by time n) $\to$ 0 as $n\to\infty$.

Reason about expectations too!



Suppose A is a node in this tree

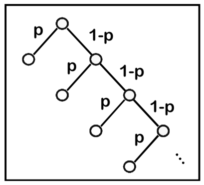
Pr(x|A)=product of edges on path from A to x.

$E[X] = \sum_x Pr(x)X(x)$.

$E[X|A] = \sum_{x\in A} Pr(x|A)X(x)$.

I.e., it is as if we started the game at A.

Expected number of heads



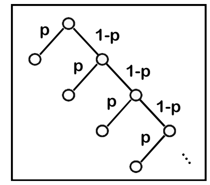
Flip bias-p coin until heads.

What is expected number of flips?

$$\sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \cdot p$$

$\underbrace{(1-p)^{i-1} \cdot p}$ Pr[i flips to see a head]

Expected number of heads



Let X = # flips.

B = event "1st flip is heads"

$E[X] = E[X|B] \times Pr(B) + E[X|\neg B] \times Pr(\neg B)$

$= 1 \times p + (1 + E[X]) \times (1-p).$

Solving: $p \times E[X] = p + (1-p)$

$\Rightarrow E[X] = 1/p.$

Infinite Probability spaces

Notice we are using infinite probability spaces here, but we really only defined things for finite spaces so far.

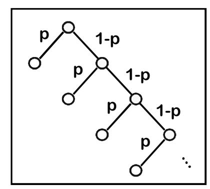
Infinite probability spaces can sometimes be weird.

Luckily, in CS we will almost always be looking at spaces that can be viewed as choice trees where

Pr(haven't halted by time t) $\rightarrow$ 0 as $t \rightarrow \infty$.

General picture

Let sample space S be leaves of a choice tree.



Let S_n = {leaves at depth $\leq$ n}.

For event A, let $A_n = A \cap S_n$.

If $\lim_{n\rightarrow\infty} Pr(S_n)=1$, can define:

$Pr(A)=\lim_{n\rightarrow\infty} Pr(A_n)$.

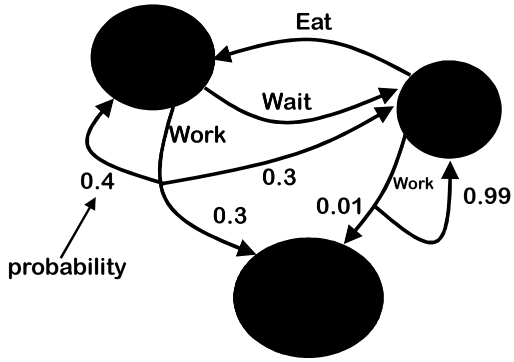
Setting that doesn't fit our model

Event: "Flip coin until #heads > 2*#tails."

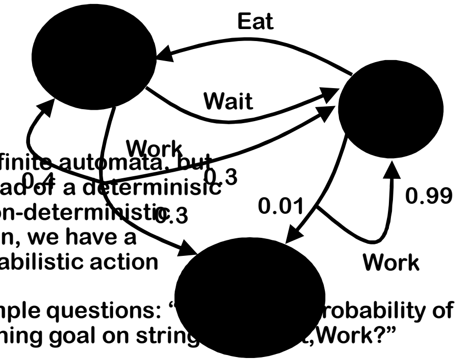
There's a reasonable chance this will never stop...

How to walk home drunk

Abstraction of Student Life



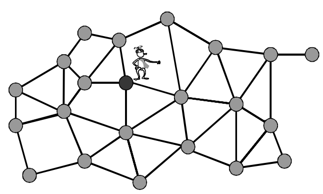
Abstraction of Student Life



Like finite automata, but instead of a deterministic or non-deterministic action, we have a probabilistic action

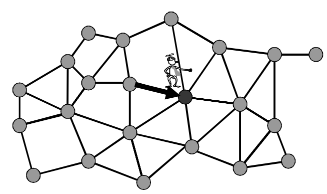
Example questions: “[redacted]robability of reaching goal on string[redacted],Work?”

Simpler: Random Walks on Graphs



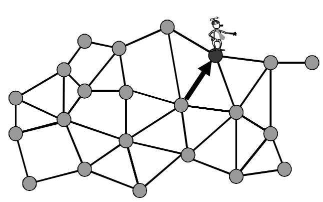
At any node, go to one of the neighbors of the node with equal probability

Simpler: Random Walks on Graphs



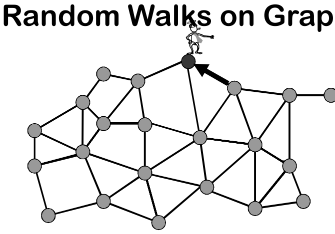
At any node, go to one of the neighbors of the node with equal probability

Simpler: Random Walks on Graphs



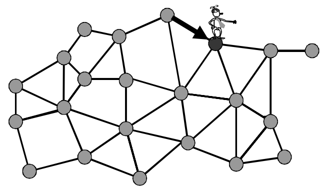
At any node, go to one of the neighbors of the node with equal probability

Simpler: Random Walks on Graphs



At any node, go to one of the neighbors of the node with equal probability

Simpler: Random Walks on Graphs

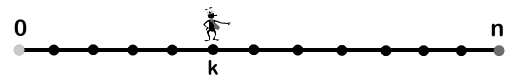


At any node, go to one of the neighbors of the node with equal probability

Random Walk on a Line

You go into a casino with \$k, and at each time step, you bet \$1 on a fair game

You leave when you are broke or have \$n



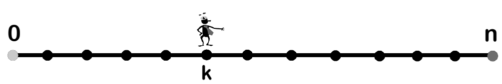
Question 1: what is your expected amount of money at time t?

Let X_t be a R.V. for the amount of \$\$\$ at time t

Random Walk on a Line

You go into a casino with \$k, and at each time step, you bet \$1 on a fair game

You leave when you are broke or have \$n



$X_t = k + \delta_1 + \delta_2 + \ldots + \delta_t,$

(δ_i is RV for change in your money at time i)

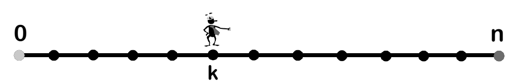
$E[\delta_i] = 0$

So, $E[X_t] = k$

Random Walk on a Line

You go into a casino with \$k, and at each time step, you bet \$1 on a fair game

You leave when you are broke or have \$n



Question 2: what is the probability that you leave with \$n?

Random Walk on a Line

Question 2: what is the probability that you leave with \$n?

$E[X_t] = k$

$E[X_t] = E[X_t | X_t = 0] \times Pr(X_t = 0)$ ← 0

$\quad + E[X_t | X_t = n] \times Pr(X_t = n)$ ← $n \cdot Pr[X_t = n]$

$\quad + E[X_t | \text{neither}] \times Pr(\text{neither})$

$k = n \times Pr(X_t = n)$

$\quad + (\text{something}_t) \times Pr(\text{neither})$

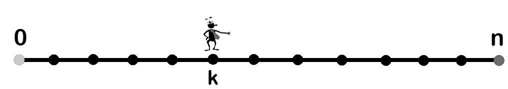
As $t \to \infty$, $Pr(\text{neither}) \to 0$, also $\text{something}_t < n$

Hence $Pr(X_t = n) \to k/n$

Another Way To Look At It

You go into a casino with \$k, and at each time step, you bet \$1 on a fair game

You leave when you are broke or have \$n

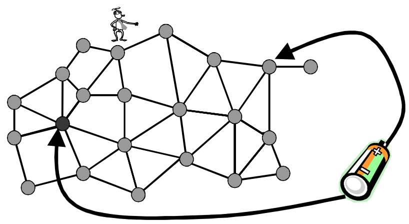


Question 2: what is the probability that you leave with \$n?

= probability that I hit green before I hit red

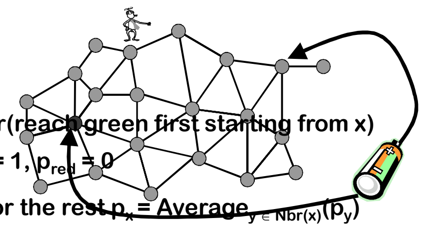
Random Walks and Electrical Networks

What is chance I reach green before red?



Same as voltage if edges are resistors and we put 1-volt battery between green and red

Random Walks and Electrical Networks



p_x = Pr(reach green first starting from x)

p_{green}= 1, p_{red} = 0

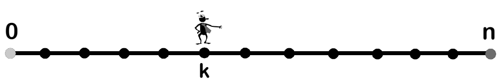
And for the rest $p_x = \text{Average}_{y \in Nbr(x)}(p_y)$

Same as equations for voltage if edges all have same resistance!

Another Way To Look At It

You go into a casino with $k, and at each time step, you bet $1 on a fair game

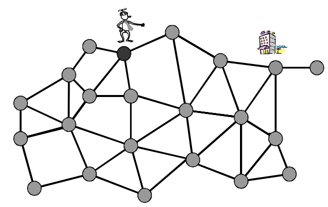
You leave when you are broke or have $n



Question 2: what is the probability that you leave with $n?

voltage(k) = k/n
= Pr[hitting n before 0 starting at k] !!!

Getting Back Home

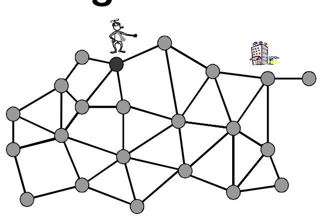


Lost in a city, you want to get back to your hotel
How should you do this?

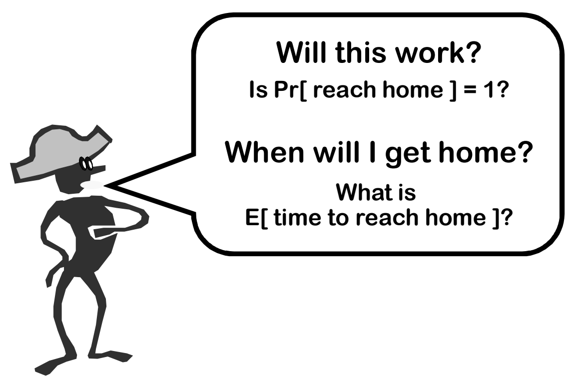
Depth First Search!

Requires a good memory and a piece of chalk

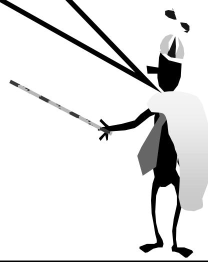
Getting Back Home



How about walking randomly?



Pr[will reach home] = 1



We Will Eventually Get Home

Look at the first n steps

There is a non-zero chance p_1 that we get home

Also, $p_1 \geq (1/n)^n$

Suppose we fail

Then, wherever we are, there is a chance $p_2 \geq (1/n)^n$ that we hit home in the next n steps from there

Probability of failing to reach home by time kn

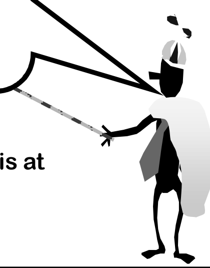
$= (1 - p_1)(1 - p_2) \ldots (1 - p_k) \rightarrow 0$ as $k \rightarrow \infty$

Furthermore:

If the graph has n nodes and m edges, then

E[time to visit all nodes] $\leq 2m \times (n-1)$

E[time to reach home] is at most this



Cover Times

Cover time (from u)

C_u = E [time to visit all vertices | start at u]

Cover time of the graph

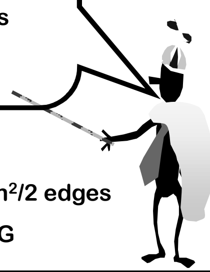
$C(G) = \max_u \{ C_u \}$

(worst case expected time to see all vertices)

Cover Time Theorem

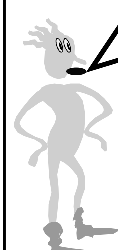
If the graph G has n nodes and m edges, then the cover time of G is

$C(G) \leq 2m (n - 1)$



Any graph on n vertices has < $n^2/2$ edges

Hence $C(G) < n^3$ for all graphs G



Actually, we get home pretty fast…

Chance that we don't hit home by $(2k)2m(n-1)$ steps is $(½)^k$

A Simple Calculation

True or False:

If the average income of people is \$100 then more than 50% of the people can be earning more than \$200 each

False! else the average would be higher!!!

Markov's Inequality

If X is a non-negative r.v. with mean E[X], then

$\Pr[\, X > 2\, E[X] \,] \le ½$

$\Pr[\, X > k\, E[X] \,] \le 1/k$

Andrei A. Markov

Markov's Inequality

Non-neg random variable X has expectation A = E[X]

$A = E[X] = E[X \mid X > 2A]\Pr[X > 2A] + E[X \mid X \le 2A]\Pr[X \le 2A]$

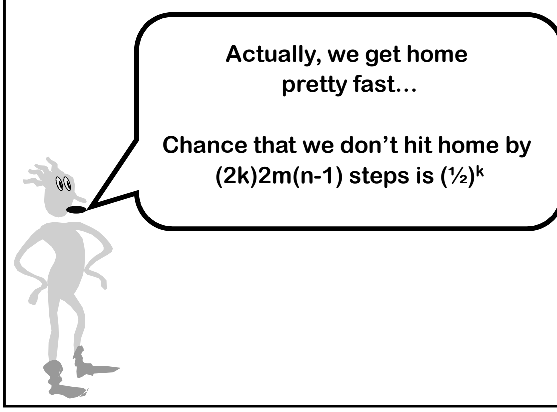
$\ge E[X \mid X > 2A]\Pr[X > 2A]$ (since X is non-neg)

Also, $E[X \mid X > 2A] > 2A$

$\Rightarrow A \ge 2A \times \Pr[X > 2A]$

$\Rightarrow ½ \ge \Pr[X > 2A]$

$\Pr[\, X > k \times \text{expectation} \,] \le 1/k$



An Averaging Argument

Suppose I start at u

E[time to hit all vertices | start at u] ≤ C(G)

Hence, by Markov's Inequality:
Pr[time to hit all vertices > 2C(G) | start at u] ≤ ½

So Let's Walk Some Mo!

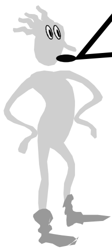
Pr [time to hit all vertices > 2C(G) | start at u] ≤ ½

Suppose at time 2C(G), I'm at some node with more nodes still to visit

Pr [haven't hit all vertices in 2C(G) more time | start at v] ≤ ½

Chance that you failed both times ≤ ¼ = $(½)^2$

Hence,
Pr[havent hit everyone in time k × 2C(G)] ≤ $(½)^k$



Hence, if we know that

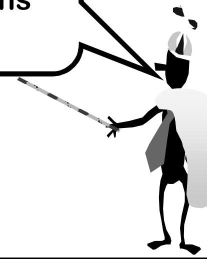
Expected Cover Time

$C(G) < 2m(n-1)$

then

Pr[home by time 4k m(n-1)] $\geq 1 - (\frac{1}{2})^k$

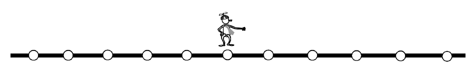
Random walks on infinite graphs



Drunk man will find way home, but drunk bird may get lost forever

- Shizuo Kakutani

Random Walk On a Line



Flip an unbiased coin and go left/right

Let X_t be the position at time t

$\Pr[X_t = i] = \Pr[\#\text{heads} - \#\text{tails} = i]$

$= \Pr[\#\text{heads} - (t - \#\text{heads}) = i]$

$= \binom{t}{(t+i)/2} / 2^t$

Random Walk On a Line

$\Pr[X_{2t} = 0] = \binom{2t}{t} / 2^{2t} \leq \Theta(1/\sqrt{t})$ Sterling's approx

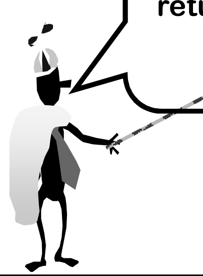
Y_{2t} = indicator for $(X_{2t} = 0) \Rightarrow E[Y_{2t}] = \Theta(1/\sqrt{t})$

Z_{2n} = number of visits to origin in 2n steps

$E[Z_{2n}] = E[\sum_{t=1\ldots n} Y_{2t}]$

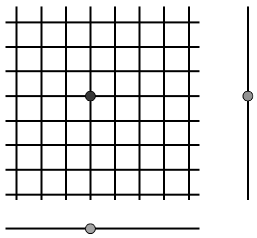
$\leq \Theta(1/\sqrt{1} + 1/\sqrt{2} + \ldots + 1/\sqrt{n}) = \Theta(\sqrt{n})$

In n steps, you expect to return to the origin $\Theta(\sqrt{n})$ times!



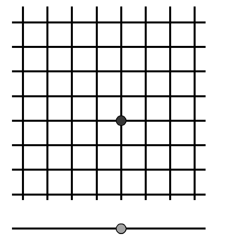
How About a 2-d Grid?

Let us simplify our 2-d random walk:
move in both the x-direction and y-direction…



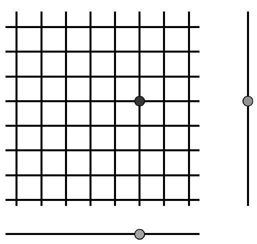
How About a 2-d Grid?

Let us simplify our 2-d random walk:
move in both the x-direction and y-direction…



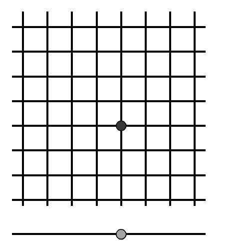
How About a 2-d Grid?

Let us simplify our 2-d random walk:
move in both the x-direction and y-direction…



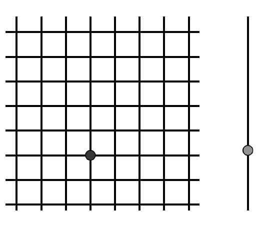
How About a 2-d Grid?

Let us simplify our 2-d random walk:
move in both the x-direction and y-direction…



How About a 2-d Grid?

Let us simplify our 2-d random walk:
move in both the x-direction and y-direction…



In The 2-d Walk

Returning to the origin in the grid
⇔ both "line" random walks return to their origins

Pr[visit origin at time t] $= \Theta(1/\sqrt{t}) \times \Theta(1/\sqrt{t})$
$= \Theta(1/t)$

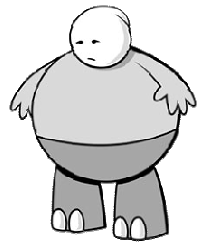
E[# of visits to origin by time n]
$= \Theta(1/1 + 1/2 + 1/3 + \ldots + 1/n) = \Theta(\log n)$

But In 3D

Pr[visit origin at time t] = $\Theta(1/\sqrt{t})^3 = \Theta(1/t^{3/2})$

$\lim_{n\to\infty}$ E[# of visits by time n] < K (constant)

Hence Pr[never return to origin] > 1/K



Here's What You Need to Know…

Conditional expectation

Flipping coins with bias p
Expected number of flips before a heads

Random Walk on a Line

Cover Time of a Graph

Markov's Inequality