Great Theoretical Ideas in Computer Science

Infinite Sample Spaces and Random Walks

Lecture 11, September 30, 2008

See handout for probability review and some new stuff

Use linearity of expectation

Suppose we have m people each with a uniformly chosen birthday from 1 to 366

\[ X = \text{number of pairs of people with the same birthday} \]

\[ E[X] = ? \]
Step Right Up…

You pick a number \( n \in [1..6] \). You roll 3 dice. If any match \( n \), you win $1. Else you pay me $1. Want to play?

Hmm… let’s see

Analysis

\( A_i = \) event that \( i \)-th die matches

\( X_i = \) indicator RV for \( A_i \)

Expected number of dice that match:

\[
E[X_1 + X_2 + X_3] = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}
\]

But this is not the same as \( \Pr(\text{at least one die matches}) \)

\( X = \) number of pairs of people with the same birthday

\( X_{jk} = 1 \) if person \( j \) and person \( k \) have the same birthday; else 0

\[
E[X_{jk}] = \frac{1}{366} \cdot 1 + \left(1 - \frac{1}{366}\right) \cdot 0 = \frac{1}{366}
\]

\[
E[X] = E\left[ \sum_{j \leq k \leq m} X_{jk} \right] = \sum_{j \leq k \leq m} E[X_{jk}] = m(m-1)/2 \times 1/366
\]
Analysis

\[ \Pr(\text{at least one die matches}) = 1 - \Pr(\text{none match}) = 1 - (\frac{5}{6})^3 = 0.416 \]

An easy question

\[ \sum_{i=0}^{\infty} \frac{1}{2^i} =? \]
Answer: 2

But it never actually gets to 2. Is that a problem?

A related question

Suppose I flip a coin of bias \( p \), stopping when I first get heads.

What’s the chance that I:

• Flip exactly once?  
  Ans: \( p \)
• Flip exactly two times?  
  Ans: \( (1-p)p \)
• Flip exactly \( k \) times?  
  Ans: \( (1-p)^{k-1}p \)
• Eventually stop?  
  Ans: 1. (assuming \( p > 0 \))

No, we really mean the limit of the partial sums. In this case, the partial sum is \( 2 - (\frac{1}{2})^n \) which goes to 2.
A related question

\[
\Pr(\text{flip once}) + \Pr(\text{flip 2 times}) + \\
\Pr(\text{flip 3 times}) + \ldots = 1:
\]
\[
p + (1-p)p + (1-p)^2p + (1-p)^3p + \ldots = 1.
\]
Or, using \(q = 1-p\),
\[
\sum_{i=0}^{\infty} q^i = \frac{1}{1-q} = \frac{1}{p}
\]

Reason about expectations too!

Suppose \(A\) is a node in this tree
\[
\Pr(x|A) = \text{product of edges on path from } A \text{ to } x.
\]
\[
E[X] = \sum_x \Pr(x)X(x).
\]
\[
E[X|A] = \sum_{x \in A} \Pr(x|A)X(x).
\]
I.e., it is as if we started the game at \(A\).

Pictorial view

Sample space \(S = \text{leaves in this tree.}\)
\[
\Pr(x) = \text{product of edges on path to } x.
\]
If \(p>0\), \(\Pr(\text{not halting by time } n)\) goes to zero.

Expected number of heads

Flip bias-\(p\) coin until heads.

What is expected number of flips?
Expected number of heads

Let $X = \#$ flips.

$B = \text{event “1st flip is heads”}$

$E[X] = E[X|B] \times \Pr(B) + E[X|\text{not } B] \times \Pr(\text{not } B)$

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

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

$\Rightarrow E[X] = \frac{1}{p}.$

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

$\Pr(\text{haven't halted by time } t)$ goes to 0 as $t$ gets large

General picture

Let sample space $S$ be leaves of a choice tree.

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

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

If $\lim_n \Pr(S_n)=1,$ can define:

$\Pr(A) = \lim_n \Pr(A_n).$

Setting that doesn’t fit our model

Event: “Flip coin until $\#\text{heads} > 2^{\#\text{tails}}.$”

There’s a reasonable chance this will never stop...
How to walk home drunk

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

- Shizuo Kakutani

Abstraction of Student Life

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

Example questions: "What is the probability of reaching goal on string Work,Eat,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

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 $n$

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

($\delta_i$ is RV for change in your money at time $i$)

$E[\delta_i] = 0$

So, $E[X_t] = k$

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 $t$

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 $n$?

$E[X_t] = k$

$E[X_t] = E[X_t | X_t = 0] \times Pr(X_t = 0)$
$E[X_t | X_t = n] \times Pr(X_t = n)$
$E[X_t | \text{neither}] \times Pr(\text{neither})$

$k = n \times Pr(X_t = n)$
$+ (\text{something}_t) \times Pr(\text{neither})$

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

$\text{p}_x = \text{Pr}(\text{reach green first starting from x})$

$\text{p}_\text{green} = 1$, $\text{p}_\text{red} = 0$

And for the rest $\text{p}_x = \text{Average}_{y \in \text{Nbr}(x)}(\text{p}_y)$

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 $n$.

0 \quad k \quad n

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

\text{voltage}(k) = k/n = \Pr[\text{hitting } n \text{ before } 0 \text{ 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?

Will this work? Is $\Pr[\text{reach home}] = 1$?

When will I get home? What is $E[\text{time to reach home}]$?
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 $kn$
$= (1 – p_1)(1 – p_2) \cdots (1 – p_k) \to 0$ as $k \to \infty$

Cover Times
Cover time (from $u$)
$C_u = E[ \text{time to visit all vertices} | \text{start at } u ]$
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 $G$

**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

$$\Pr[ X > 2 E[X] ] \leq \frac{1}{2}$$

$$\Pr[ X > k E[X] ] \leq \frac{1}{k}$$

Andrei A. Markov
Markov’s Inequality

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

\[
A = E[X] = E[X | X > 2A] \Pr[X > 2A] + E[X | X \leq 2A] \Pr[X \leq 2A]
\]

\[
\geq E[X | X > 2A] \Pr[X > 2A]
\]

(since $X$ is non-neg)

Also, $E[X | X > 2A] > 2A$

\[
\Rightarrow A \geq 2A \times \Pr[X > 2A]
\]

\[
\Rightarrow \frac{1}{2} \geq \Pr[X > 2A]
\]

\[
\Pr[X > k \times \text{expectation}] \leq 1/k
\]

An Averaging Argument

Suppose I start at $u$

\[
E[\text{time to hit all vertices} | \text{start at } u] \leq C(G)
\]

Hence, by Markov’s Inequality:

\[
\Pr[\text{time to hit all vertices} > 2C(G) | \text{start at } u] \leq \frac{1}{2}
\]

So Let’s Walk Some More!

\[
\Pr[\text{time to hit all vertices} > 2C(G) | \text{start at } u] \leq \frac{1}{2}
\]

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

\[
\Pr[\text{haven’t hit all vertices in } 2C(G) \text{ more time} | \text{start at } v] \leq \frac{1}{2}
\]

Chance that you failed both times \(\leq \frac{1}{4} = (\frac{1}{2})^2\)

Hence,

\[
\Pr[\text{haven’t hit everyone in time } k \times 2C(G)] \leq (\frac{1}{2})^k
\]

Actually, we get home pretty fast…

Chance that we don’t hit home by \((2k)2m(n-1)\) steps is \((\frac{1}{2})^k\)
Hence, if we know that Expected Cover Time $C(G) < 2m(n-1)$ then

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

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

- Shizuo Kakutani

Random walks on infinite graphs

Random Walk On a Line

Flip an unbiased coin and go left/right
Let $X_t$ be the position at time $t$

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

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

$= \left( \frac{t}{(t+i)/2} \right) /2^t$
Random Walk On a Line

\[ \Pr[X_{2t} = 0] = \left(\frac{2t}{t}\right) \frac{1}{2^{2t}} = \Theta(1/\sqrt{t}) \]

Stirling’s approx

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

\[ Z_{2n} = \text{number of visits to origin in } 2n \text{ steps} \]

\[ E[Z_{2n}] = E[\sum_{t=1}^{n} Y_{2t}] \leq \Theta(\frac{1}{\sqrt{1 + \frac{1}{2} + \ldots + \frac{1}{n}}}) = \Theta(\sqrt{n}) \]

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...

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 ] = Θ(1/√t) × Θ(1/√t)
= Θ(1/t)

E[ # of visits to origin by time n ]
= Θ(1/1 + 1/2 + 1/3 + ... + 1/n ) = Θ(log n)
But In 3D

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

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

Hence \( \text{Pr[ never return to origin ] } > 1/K \)

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