Lecture 17: 5/28/2009

Announcements: Ps9 out. Ps8, solutions out
FINAL EXAM: Monday June 8, 10:30-12:30 106 Olson, Open Book, 1 double sided sheet of paper for notes.
Review Session: Fri, 6/5: 12:10-2 146 Olson

Eg 4: Biased coin

Now, what if the coin is biased?
Say that the coin lands heads with probability .51, tails with probability .49.
each flip independent of the rest.

You flip an unfair coin 100 times. The coin lands heads
a fraction p=0.51 of the time:

S = {0,1}^100
P(x) = p^{#1(x)} (1-p)^{#0(x)}   where #1(x) = the number of 1-bits
of 0-bits in the string x.

What's the Probability of 51 heads now?

C(100,5)(.51^5)(.49^50) ≈ 0.0796050538

Makes sense -- 51 heads should now be the most likely number,
and things should fall off from there. Before, 50 was the most likely number.

Eg 5: Dice. .

----

o Pair of dice, what's the chance of rolling an "8"?

Event E = {(2,6),(3,5),(4,4),(5,3),(6,2)}

P(E) = 5/36 \approx 14%

```
    Be careful: P(E) = |E|/|S| *if* we are assuming the *uniform*
    distribution.

o  What's the chance of rolling an 8 if I tell you
   "both dice were even".
   Probability (Roll an 8 | both dice even)

   Method 1:   Imagine the new probability space:

    (2,2), (2,4), (2,6),  (4,2), (4,4), (4,6),  (6,2), (6,4), (6,6)
                   ***            ****           ***

   So probability is 3/8 \approx 33%


   Method 2: A little more "mechanically"

   A = "rolled an 8"
   B = "both dice even"

   P(A | B) = P(A\cap B)/P(B)
            = (3/36)  /  (9/36) = 5/9
                ^
         look back at the five points -- 3 of the five had both even


------------------------------------------------------------------------
Eg 6:  An urn contains  30 white balls and 30 black balls.
       You pull out 5 balls (no replacement).
       a. What's the chance they all have the same color?
       b. What's the chance if I tell you that the first one was white?
       c. What's the chance if I tell you that the first two were white?
------------------------------------------------------------------------


a) P(monochromatic) = P(allwhite) + P(allblack)
                    = 2 * C(30,5) / C(60,5)
```

```
                         = 285,012 / 5 461 512
                         \approx 0.521    (5.2%)

b) first one red -- unchanged, no information

   Symbolically,

     P(monochromatic |firstwhite)

      P(monochromatic and firstwhite)      C(30,5) / C(60,5)
      ---------------------------------- =  -----------------
              P(firstwhite)                        1/2


c) P(monochromatic | firsttwowhite)

  P(monochromatic and firsttwowhite)    C(30,5) / C(60,5)
  ----------------------------------  =  -----------------
\approx 0.1062
            P(firsttwowhite)                (1/2)(29/59)
(10.6%)

-----------------------------------------------
```

Eg 7: . Medical Testing

:

Suppose we have some rare disease X that 1/10,000 people have. We can test for disease X, and if the person has it the test always confirms that fact (no false negatives). If the person does not have the disease, the test is correct 99% of the time, but 1% of the time it reports that the healthy person has disease X (and the only way to confirm for sure if they have X or not is with expensive, risky surgery).

Thus is we test 100,000 people, roughly 10 will have disease X , and the test correctly confirms it. However, of the remaining 99,990 healthy people, about 1% will get a false positive, so about 99,990/100 or about 1000. So to cure 10 people we subject about 1000 people to substantial harm from from the misdiagnosis.

Above is informal, for us, what is the sample spade? Simple model: pairs {X, no X} x {Test-yes, Test-no} = {(X,Test-yes), (X, Test-no), (no-X, test-yes), (no-X, test-no)}

With P(X, Test-yes) = 1/10,000; P(X-Test-no) = 0; P(no-X, test-yes)= 9,999/10,000 * .01; P(no-X, test-no) = 9,999/10,000 * .99

The experiment is to draw a sample according to this distribution, and we then see only the test result. The question is: given test-yes, what is the probability no-X?

= P(no-x, test-yes)/P(test-yes)=
(9,999/10,000 * .01)/ (1/10,000 + 9,999/10,000 * .01) approx= .99

Eg 8: Birthday phenomenon
----

n=23 people gather in a room.
What' the chance that some two have the same birthday?
Assume nobody born 2/29, all other birthdays equiprobable.
S = [1..365]^23

P(same birthday) = 1 - (all birthday's different)
= 1 - (1-1/365)(1-364) ... (1-(n-1)/365)
= 1 - prod_{i=1}^{i=22} (1-1/i) = 0.507

More detailed analysis (Not done in class):

Let c(n,q) = probability of at least one collision in the experiment of throwing q balls into n bins. Then

c(n,q) \approx 1-e^{-q^2/2n}

Solving for c(n,q)=1/2:   q = \sqrt(ln 2) \sqrt(n)
≈1.1774 sqrt(n)

0.3 q(q-1)/n <= c(n,q) <= 0.5q^2/n

if q <= sqrt(2n)

Eg 5.1 An application: Cryptographic Hash Functions
(didn't do this in class)

SHA1: \bits^* -> \bits^160

About how long will it take to find a collision
if compute one new point every used?

roughly 2^80 msec,
   10^37 years   (1 year is about pi * 10^7 seconds)

Eg 9: Monty Hall Problem (keep-or-switch game)

```
 ========     ========     ========
 |      |     |      |     |      |
 | bad  |     | bad  |     | good |
 |      |     |      |     |      |
 ========     ========     ========
    1            2             3
```

You choose a random door
Should you switch?

```
   loc of good prize  my guess
S = {1, 2, 3} x  {1, 2, 3}
```

WIN = get good prize

Results for winning if you switch:

(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
Lose	Win	Win	Win	Lose	Win	Win	Win	Lose

Win: 6/9 = 2/3

Or just choose door 1

S = {1, 2, 3}

1	2	3
lose	win	win

```
The following examples were not covered in class, listed as
extra ones (won't do them)
----------------------------------------
Eg 5:  Same parity game
----------------------------------------
Alice randomly, uniformly chooses two distinct numbers between
1 and 10. What is the probability they have the
same parity?

Is it exactly 1/2?
Should actually be less than a 1/2, because distinct

S = { (a,b) \in {1..10}^2: a\ne b}
|S| = 90

E = {(a,b) \in {1..10}^2: a mod 2 = b mod 2}

|E|= |evenAevenB| + |oddAoddB|
      5 * 4       +    5 * 4

40/90 = 4/9 \approx 44%



----------------------------------------
Eg 6:  Bigger/ smaller game
----------------------------------------

Alice uniformly chooses two distinct numbers between
1 and 10, announces the that FIRST.
Bob guesses if the second is SMALLER or LARGER.
How should Bob play optimally and, if he does so,
what is his chance to win?


As usual, start by figuring out the sample space
S = {(i,j) \in {1..10}^2: i\ne j}

   1  2  3  4  5  6  7  8  9  10

- If Alice announces 1,2,3,4,5  guess SMALLER
- If Alice announces 6,7,8,9,10 guess LARGER

P(Win) = P(Win| AliceAnswers1) P(AliceAnswers1) +
          P(Win| AliceAnswers2) P(AliceAnswers2) +
          ..
          P(Win| AliceAnswers10) P(AliceAnswers10)


       = (1/10) (P(Win | AliceAnswers1)  + ... + P(Win |
AliceAnswers10)
       = (1/10) (9/9 + 8/9 + 7/9 + 6/9 + 5/9 +
                 9/9 + 8/9 + 7/9 + 6/9 + 5/9)
       = (1/10) (7*10/9)             numbers clearly average 7
       = 70/90
       = 7/9 \approx 78%
```

Eg 7: Expected value

Alice rolls a die.
What's do you expect the square of her roll to be?

could be 1 .... could be a 36!

Definition: a RV is a function from X: S -> \R

```
Definition: E[X] = \sum X(s) P(s)
                     s
```

```
So, in this problem,
   E[X] = 1(1/6) + 2^2(1/6) + 3^2(1/6) + ... + 6^2(1/6)
        = 1/6(1+4+9+15+25+36)
        = 91/6
        \approx 15.2
```

Exercise: Repeat, supposing she rolls a *pair* of dice:

Eg 8: Subway

When Pablo leaves his office late at night, he wanders to the subway and takes the first train North or South:

```
                             Girlfriend's home
                                 /|\
                                  |
                                  |
                                  |
         Laboratory -------> subway stop
                                  |
                                  |
                                  |
                                 \|/
                             Student's home
```

There are trains every 10 mins, both N and S. During the last 31 days, Pablo only has gone home 3 times, and this seems to be about typical Explain what is going on and compute Pablo's average wait time for a triain?

Example:

```
Northbound    Southbound

              11:00
                      1 min
    11:01



                      9 min




              11:10
                      1 min
    11:11



                      9 min
              11:20
    11:21
```

Let X = Wait time

```
    (1/10) (0.5 min)  + (9/10) ( 4.5 mins)
   = 0.05 + 4.05 mins
   = 4.1 mins
```

How should the trains be staggered to minimize Pablo's wait time?

```
               11:00
     11:05
               11:10
     11:05
               11:20
```

Average wait time will be 2.5 mins

```
----------------------------------------------------------
Eg 9:  Birthday analysis -- again  (we didn't get to this)
----------------------------------------------------------
```

Select q random points, with replacement, from universe of N points

```
Let C_i = event that point i collides with a previous one.
    D_i = event that no collision up to time i
```

```
P(collision) = 1 - P(D_q)
             = 1 - P(D_q | D_{q-1}) P(D_{q-1})
             = 1 - P(D_q | D_{q-1}) P(D_{q-1})
             = 1 - P(D_q | D_{q-1}) P(D_{q-1} | D_{q-2})
P(D_{q-2})
             = ...
                    q-1
             = 1 - \prod    P(D_{i+1}|D_i)
                   i = 1

                    q-1
             = 1 - \prod    (1- i/N)
```

```
                  i = 1
```

Now, let's approximate 1-x by exp(-x)   (1-x <= e^-x)

```
    1+x \approx exp(x) when x\approx 0


              .      q-1
              = 1 - \prod    exp(-i/N)
                    i = 1

              = 1 - exp(-1/N - 2/N - 3/N  - ... - (q-1)/N)
              = 1 - exp(-q(q-1)/2N)
              .
              = 1 - exp(-q^2/2N)
```

So: about how large should q be for this to be 1/2?

```
          0.5 = 1 - exp(-q^2/2N}
          0.5 = exp(-q^2/2N}

        - ln 2 = -q^2/2N

           (2 ln 2) N  = q^2

                 q = \sqrt(2 ln 2) sqrt(N)
                   = 1.177 sqrt(N)
```

Application: SHA1 has 160-bit outputs. About how long to find a collision by
trying successive points? Assume SHA1 behaves as a random function would.

```
   1.177 * 2^80 tries
```

If each try take 1 usec, so can do 10^6 tries/second:

```
    1.177 * 2^80 / 10^6 / 3.1*10^7  = 1    1.177 * 2^80 / 10^6 / 3.1*10^7  >  10^10 years
```