Math 55 - Fall 2007 - Lecture notes # 26 - October 29 (Monday)

Begin reading Chapter 6 of Rosen
Read sections 1-5 of Lenstra's notes (see class web page for pointer)
Additional reading: CS70 notes,
inst.eecs.berkeley.edu/~cs70/sp07
starting with second half of Lecture 16

Goals for today: Introduce discrete probability theory

Long term goal: we would like to make sense of statements like:

1) "The chance of getting a "flush" (all cards the same suit) in 5-card poker is about 2 in 1000."
2) "If you flip a fair coin 50 times, and each time it comes out heads, then the chance you get a head the 51st time is still 50%."
3) "If quicksort picks a random "pivot item" at each step, then it will sort n numbers, in O(n log n) time with high probability".
4) "With this algorithm for balancing the workload among computers the probability (or chance) that a user has to wait more than 1 minute is 2%."
5) "There is a 60% chance of the Big One (large earthquake) hitting Northern California in the next 30 years."
6) In an election, if B's reported vote total is 2,912,790, G's reported vote total is 2,912,253, i.e. a difference of 537 votes, and the chance of a vote having been mistakenly counted wrong is 1 out of 1000, so that about 2900 votes may have been counted wrong, what is the probability that G actually won the election?
(recall an election in Florida in 2000).

To understand these examples and others, we must specify a
1) an experiment, whose outcome is "random"
2) the set of possible outcomes (the "sample space")
3) probability of each possible outcome

EX 1): The experiment is shuffling a deck of cards and dealing 5 of them
The set of outcomes is all possible 5 card subsets of 52 cards
ASK&WAIT: How many ways are there?
The probability of each outcome is equal (assuming we shuffle well)
ASK&WAIT: What is the probability of any particular shuffle?

EX 2): The experiment is flipping a fair coin 51 times
       The set of outcomes is all possible sequences is of 51 H's and T's
ASK&WAIT: How many outcomes are there?
       The probability of each outcome is equal (assuming the coin is fair)
ASK&WAIT: What is the probability of any particular sequence?

EX 3 through 5 are tricker, especially 5) (ask your local geologist!)
We'll return to EX 6 later.

DEF: A _sample space_ is a finite (or countable) set S
together with a function (called probability function, or just probability)
P: S --> [0,1]
such that
sum_{x in S} P(x) = 1.

S is the set of all possible outcomes of the experiment,
with P(x) equal to the probability that the outcome is x.

The most notable case is when S is finite and P(x) has the same value
for all x in S, i.e. when all events are equally likely. In this case,
we say S has a "uniform probability distribution."
What is P(x) equal to in this case? We have
1 = sum_{x in S} P(x) = P(x) |S|
for all x in S, so, in the uniform distribution case, we have
P(x) = 1/|S|
for all x in S.

EX 1 and 2 above: |S| = C(52,5) or |S| = 2^51

DEF: An _event_ E is a subset of the sample space S, and the _probability_
P(E) of an event E is given by
P(E) = sum_{x in E} P(x).
Note: the empty set has probability 0. The whole sample space S has
probability 1.

In the case of a uniform distribution, we have
P(E) = sum_{x in E} P(x)
     = sum_{x in E} 1/|S|
     = |E|/|S|.

EX: One toss of a fair coin.
    S = {H,T} P(H) = 1/2, P(T) = 1/2.

```
EX: 3 tosses of a fair coin
    S = {HHH, HHT, ..., TTT}, P(any particular outcome) = 1/8
    E = {2 heads and a tail}

ASK&WAIT: What is P(E)?

EX: 3 tosses of a biased coin. This means P(H) not equal to P(T), i.e. not 1/2
    Suppose P(H) = 1/3

ASK&WAIT: What is P(T)?

ASK&WAIT: What is P(HHH)? P(HTH)?

ASK&WAIT: What is P(E), E as above?

EX: A roll of a die.
    S = {1,2,3,4,5,6}.
    P(x) = 1/6 for all x in S.
    Let E = "the roll of the die is odd" = {1,3,5}
ASK&WAIT: What is P(E)?

EX: A roll of two dice, one red and one blue.
    S = {1,2,3,4,5,6}x{1,2,3,4,5,6},
    i.e. all pairs S = {(i,j), 1 <= i <= 6, 1 <= j <= 6 }
    P(x) = 1/36 for all x in S, since |S| = 36

ASK&WAIT: What is P(E), E = "the first die is a 6"?

ASK&WAIT: What is P(E), E = "at least one die is a 6"?

ASK&WAIT: What is P(E), E = "the dice sum to 7"

ASK&WAIT: What is P(E), E = "the dice sum to 10"

EX: A roll of two indistinguishable dice (eg both blue)
    indistinguishable means that, say (1,6) and (6,1) no longer different
    S = {(1,1),(1,2),...,(1,6),(2,2),(2,3),...,(2,6),(3,3),...,(3,6),(4,4),...,(6,6)}
    S = {(i,j), 1 <= i <= j <= 6}
ASK&WAIT: What is P(i,i)? What is P(i,j) for i not = j?
ASK&WAIT: What is P(E), E = "dice sum to 10"
```

EX: A single poker hand, gotten by shuffling a deck of 52 card and taking 5.
    S has C(52,5) = 2,598,960 elements, which I will not list here.
    P(x) = 1/2,598,960 for all x in S.

ASK&WAIT: What is P(E), E = "royal flush" = "A,K,Q,J,10" of the same suit?

ASK&WAIT: What is P(E), E = "the hand has four of a kind"

ASK&WAIT: What is P(E), E = "the hand contains a full house"

ASK&WAIT: What is P(E), E = "the hand contain a flush"

EX: Balls and bins. Suppose you take 20 distinguishable balls (tasks)
and throw them into 10 distinguishable bins (computers) so that
each ball has an equal chance of landing in each bin. (This is
a common way of distributing work among multiple computers, eg
web requests coming into a company)

ASK&WAIT: What is S? |S|? P(any particular outcome)?

What is P(E), E = {each bin has at most 4 balls}
                = {each computer has at most 4 requests}?
We will eventually answer this...

ASK&WAIT: What is P({each bin has exactly 2 balls})?
          i.e. that the balls are perfectly evenly distributed?

EX: Strategy for a TV game show, where you have to pick one of 3 doors,
and you win whatever is behind the door.
First a prize is placed behind one of three doors, each with equal
probability. You are then allowed first to choose one door.
Then, one of the other two doors is revealed (behind which, of course,
no prize appears).
Finally, you are allowed the option of switching to another door.
You will win whatever is behind the door you select.
Should you switch to the third door, stay where you are,
or does it not matter?

Answer: you should switch, always! Why?
Let's figure out the sample space describing the situation
up to the moment you have to choose whether to switch:
   S = {(i,j,k) where

i=1,2 or 3 indicates the door where the prize is,
j=1,2 or 3 indicates the door you originally choose, and
k=1,2 or 3 indicates the door opened on the show}
So i and j and take any values from {1,2,3} with equal probability.
But k is restricted: if i not = j, then k must be chosen not
to equal either i or j, so its value is determined. But if i=j,
then k can equal either of the other 2 values with equal probability.

So here is the sample space with probabilities shown below
each outcome in parentheses.
For example i=2,j=1 means k=3, and has probability (1/3)*(1/3)=1/9.
For example i=2,j=2 means k=1 or 3, and i=2,j=2,k=3 has
probability (1/3)*(1/3)*(1/2)=1/18

	i=1	i=2	i=3
j=1	k=2 or k=3 (1/18) (1/18)	k=3 (1/9)	k=2 (1/9)
j=2	k=3 (1/9)	k=1 or k=3 (1/18) (1/18)	k=1 (1/9)
j=3	k=2 (1/9)	k=1 (1/9)	k=1 or k=2 (1/18) (1/18)

Now suppose that your strategy is not to switch doors;
what is the probably of the event E = {you win!}?
ASK&WAIT? Can you indicate which parts of the sample space is in E? What is P(E)?
Now suppose that your strategy is to switch doors;
what is the probably of the event E = {you win!}?
ASK&WAIT? Can you indicate which parts of the sample space is in E? What is P(E)?
ASK&WAIT: What is the best strategy, switch or not?

Now we go on to techniques that make it easier to compute
the probabilities of certain events.

THEOREM: Let E be an event in a sample space S. The probability of
the event S-E, the complement of E in S, is given by 1-P(E).

PROOF: 1 = sum_{x in S} P(x)
 = sum_{x in E} P(x) + sum_{x in S-E} P(x)
 = P(E) + P(S-E).
So P(S-E) = 1 - P(E).

THEOREM: Let E and F be events in a sample space S. Then
P(E union F) = P(E) + P(F) - P(E intersect F).

PROOF: Similar to the proof of inclusion-exclusion, which may
be stated as
sum_{x in E union F} 1 = sum_{x in E} 1 + sum_{x in F} 1
- sum_{x in E intersect F} 1.
Just replace 1 in the sums above by P(x).

EX: What is the probability that a randomly chosen integer between 1 and 100
    is divisible by 5 or 7?

THEOREM: Let E1, E2,... En be pairwise disjoint event in a sample space S.
         Then P(E1 U E2 U ... U En) = P(E1) + P(E2) + ... + P(En)
ASK&WAIT: What is the proof?

Goals of next section: Continue discrete probability theory
                       Conditional probability
                       Independence
                       Bernoulli trials

Now we start discussing conditional probability.
Here is an example that we would like to understand:

  A pharmaceutical company is marketing a new test for a certain
  medical condition. According to clinical trials, the test has
  the following properties:
    1. When applied to an affected person, the test comes up
       positive in 90% of cases, and negative in 10% ("False negatives")
    2. When applied to a healthy person, the test comes up
       negative in 80% of cases and positive in 20% ("False positives")
  Suppose that 5% of the US population has the condition.
  In other words, a random person has a 5% chance of being affected.
  When a random person is tested and comes up positive, what is the
  probability that the person actually has the condition?

This is an example of conditional probability: what is the
probability of event A (person is affected) given that we know
event B occurs (the person tests positive). We write this P(A|B),
the probability of A given B.

Def: P(A|B) = P(A inter B)/P(B)

Justification: Let S be the original sample space, and P() the
original probability function on S. Since we know B occurs,
we have a new sample space, namely B subset S. What is the
new probability function? If x in B, then P(x|B) must satisfy
  1 = sum_{x in B} P(x|B), so
the obvious choice is P(x|B) = P(x)/P(B).
So if A subset B is any event in the new sample space B,
then P(A|B) = sum_{x in A} P(x|B) = sum_{x in A} P(x)/P(B)
            = P(A)/P(B)
What if A is not a subset of B? If x in A but x not in B,
then clearly P(x|B) = 0; if B occurs then x cannot occur.
Thus we finally get P(A|B) = P(A inter B)/P(B).

Let N = US population.
Returning to medical testing, the population consists of 4 groups:
  1) TP (true positives) |TP|=90% of 5% of N = (9/200)*N, P(TP)=9/200
  2) FP (false positives) |FP|=20% of 95% of N = (19/100)*N, P(FP)=19/100
  3) TN (true negatives) |TN|=80% of 95% of N = (76/100)*N, P(TN)=76/100
  4) FN (false negatives) |FN|=10% of 5% of N = (1/200)*N, P(FN)=1/200
Now let A = {person is affected} = TP U FN
        B = {person tests positive} = TP U FP
        A inter B = TP
  and finally P(A|B) = P(TP)/P(TP U FP)
                     = (9/200)/(9/200 + 19/100) = 9/47 ~ .19
So if a random person tests positive, there is only a 19% chance
that they really have it.

ASK&WAIT: What is P(B|A) = P(person tests positive | person is affected)?

ASK&WAIT: What is P(test correct when given to random person)?

ASK&WAIT: Let a "phony test" simply declare everyone healthy
          what is P(phony test correct when given to a random person)?

Ex: Suppose we toss 3 balls into 3 bins
ASK&WAIT: What is P(first bin empty)?
ASK&WAIT: What is P(second bin empty | first bin empty)?

Ex: Roll two fair dice, what is P(rolling a 6 | sum of dice is 10)?

Ex: Roll two fair coins, what is P(second is head | first is head)

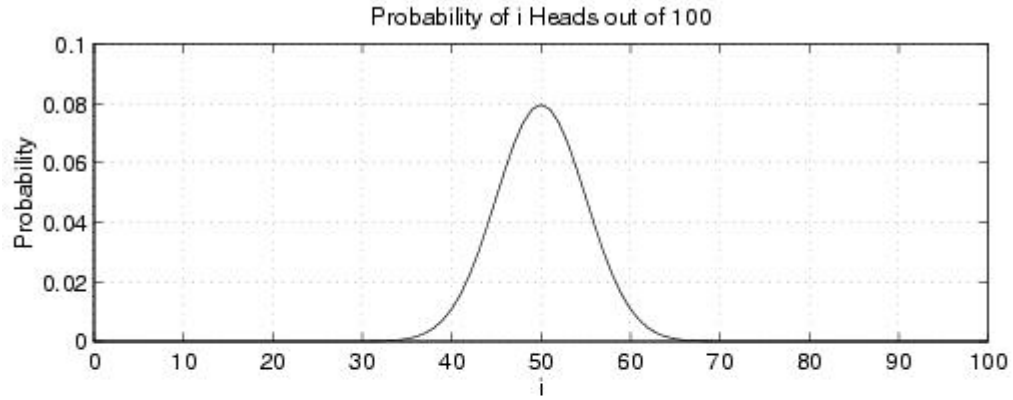
```
Def: Two events A and B are independent if P(A inter B) = P(A)*P(B)

EX:  flip two coins, A = {HH, TH}, B = {HH, HT}, A inter B = {HH}
     P(A) = 1/2 = P(B), P(A inter B) = 1/4

Prop: If A and B are independent, then P(A|B) = P(A) and P(B|A) = P(B)
Proof: P(A|B) = P(A inter B)/P(B) = P(A)*P(B)/P(B) = P(A)
       P(B|A) = P(A inter B)/P(A) = P(A)*P(B)/P(A) = P(B)

ASK&WAIT: Throw 3 balls into 3 bins, are
          A = {first bin empty} and B = {second bin empty} independent?
ASK&WAIT: Throw 2 dice, are
          A = {rolling a 6} and B ={sum=10} independent?
ASK&WAIT: Throw 2 dice, are
          A = {sum even}, B = {first die even} independent?

Def: Events A1, A2, ... , An are mutually independent if
     for every i and every subset J of {1,2,...,n} - {i} then
     P(Ai | inter_{j in J} Aj) = Pr(Ai)
     i.e. Ai does not depend on any combination of the other events

Thm: P(B inter A) = P(B)*P(A|B)
Proof: follows from definition of P(A|B)

Thm: P(A1 inter A2 inter ... inter An) =
     P(A1) * P(A2|A1) * P(A3|A1 inter A2) * P(A4| A1 inter A2 inter A3)
          * ... * P(An | A1 inter A2 inter ... inter An-1 )
Proof: induction on n:
       Base case: n=1: P(A1)=P(A1)
       Induction step: Assume
          P(A1 inter ... inter An-1)
                = P(A1) * ... * P(An-1 | A1 inter ... inter An-2)
       Then P(A1 inter ... inter An)
          = P(A1 inter ... inter An-1) * P(An | A1 inter ... inter An-1)
          = P(A1) * ... * P(An-1 | A1 inter ... inter An-2) *
            P(An | A1 inter ... inter An-1)     (by induction, as desired)

Corollary: Suppose A1, A2, ... , An are mutually independent. Then
           P(A1 inter A2 inter ... inter An) = P(A1)*P(A2)*...*P(An)
  Proof: in above proof, each
           P(Ai | A1 inter ... inter Ai-1) = P(Ai) by mutual independence
```

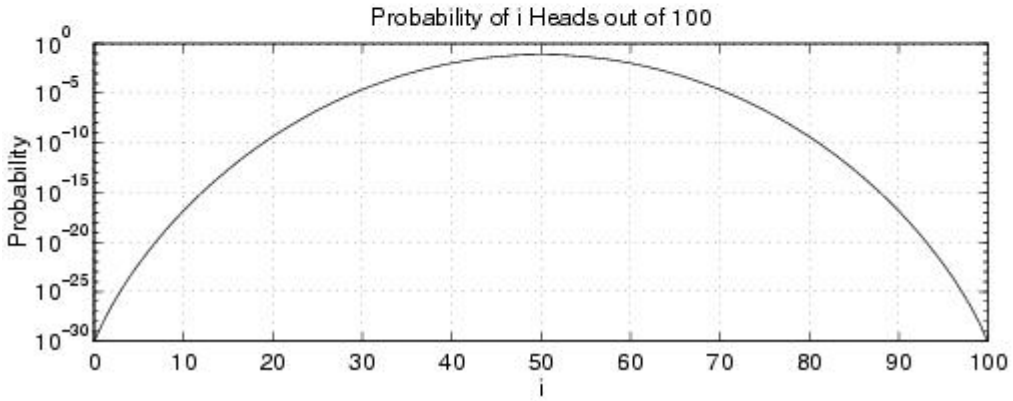
```
EX: Toss a fair coin 3 times. Let A={HHH}, A1={Hxx}, A2={xHx}, A3={xxH}
    A = A1 inter A2 inter A3
    P(A) = P(A1) * P(A2|A1) * P(A3|A1 inter A2)
         = P(A1) * P(A2)    * P(A3)
         = 1/2   *  1/2     * 1/2
         = 1/8 as expected
EX: Toss a biased coin 3 times, with P(H) = p
ASK&WAIT: what is P(A)?

Def: a Bernoulli trial is a (sequence) of (independent, identical)
     experiments, each of which has two outcomes

EX: Suppose we flip a fair coin 100 times. What is P(50 Heads)?
    sample space S = {all sequences of 100 H's and T's},
    each with P(x)=1/2^100 because
        P(HTH...) = P(1st = H)*P(2nd = T)*P(3rd = H)* ... = 1/2^100
    (or because it's a uniform distribution over 2^100 possibilities)
    E = {all sequences with 50 heads, 50 tails}
ASK&WAIT: What is |E|? P(E)?
ASK&WAIT: Let E(i) = {i Heads out of n flips} What is |E(i)|? P(E(i))?
    Note that E(i) and E(j) are disjoint, and
    S = E(0) U E(1) U ... E(n), so P(S) = P(E(0)) + ... + P(E(n)) = 1
    Check this: sum_{i=0 to n} P(E(i)) = sum_{i=0 to n} C(n,i)/2^n
                   = 2^(-n) * sum_{i=0 to n} C(n,i)
                   = 2^(-n) * (1+1)^n  ... by the Binomial Theorem
                   = 1 as desired

EX; Now flip a biased coin, with P(H) = p and P(T) = 1-p, 100 times
    The sample space is the same as above.
    But not all P(x) are the same
ASK&WAIT: What is P(50 Hs followed by 50 Ts)?
ASK&WAIT: What is P(50 Hs and 50 Ts, in some fixed order)?
ASK&WAIT: What is P(50 Hs and 50 Ts, in any order)?
    Now flip a biased coin n times
ASK&WAIT: What is P(i Hs and n-i Ts, in any order)?
ASK&WAIT: What is sum_{i=0 to n} P(i Hs and n-i Ts, in any order)?

Theorem: If you flip a biased coin n times, with P(H) = p,
         the probability of getting i Heads is C(n,i)*p^i*(1-p)^(n-i)

What does P(getting i heads out of n flips) look like as a function of i?
```

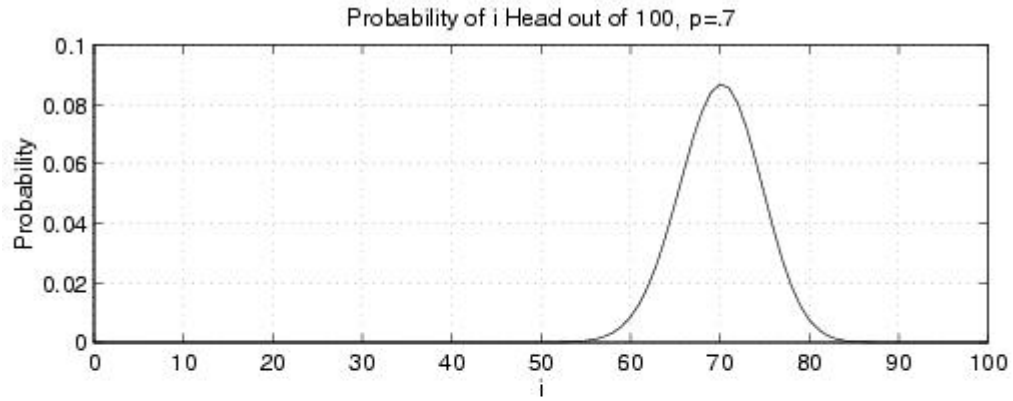
Let's look for n=100, p = .5 , and for n=100, p=.7
Comments on the plots:

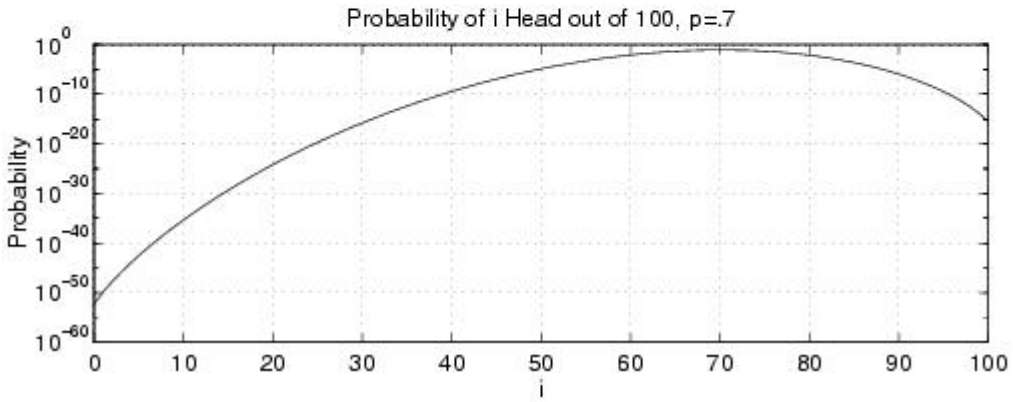
when p=.5, the probability is largest at i=50 (equal numbers of heads and tails), and quickly gets smaller for larger or smaller i. It gets so small that it is easier to look at a logarithmic scale (second plot), where the probability of getting 30 Hs and 70 Ts (or 70 Hs and 30 Ts), is about 10^(-5), and the probability of getting 10 Hs (or 10Ts) is down to 10^(-17).

when p = .7, then most noticeable feature of the 3rd plot is that it look very much like the first plot, except slid over to have its peak at 70 Hs instead of 50 Hs. This makes sense because with P(H) = .7, one expects close to 70 Hs out of 100. We will return later to explain the remarkable resemblance of these two plots when we discuss the Central Limit Theorem.


Probability of i Heads out of 100
Probability
i


Probability of i Heads out of 100
Probability
i


Probability of i Head out of 100, p=.7
Probability
i


Probability of i Head out of 100, p=.7
Probability
i

EX: Probability of {a flush in poker} (5 cards of same suit)
    P(flush) = 4*P(A), A = {flush in hearts}
    A = A1 inter A2 inter ... inter A5
    Ai = {ith card is a heart}
    By Theorem: P(A) = P(A1) * P(A2|A1) * P(A3|A1 inter A2) * ...
ASK&WAIT: What is P(flush)?

EX: You go to a casino, which advertises the following game:
    You pick a number from 1 to 6. Then they role 3 die, and
    you win if your number comes up at least once.
ASK&WAIT: The casino claims that your chance of winning is 50%,
          since it is 1/6 for each die, each die is independent,
          so the probability is 3*(1/6)=1/2. Is this argument reasonable?
    Let's figure out the real probability of winning at this game.
    Let Ai = {your number comes up on die i}, and A = A1 U A2 U A3.
    We want P(A). The casino said P(A) = P(A1) + P(A2) + P(A3) = 3*(1/6)=1/2
    But this is only true if the Ai are disjoint, which they are not
    (your number can come up twice). So we need inclusion/exclusion:

    Recall: P(A1 U A2) = P(A1) U P(A2) - P(A1 inter A2)
ASK&WAIT: what is P(A1 U A2 U A3)?
ASK&WAIT: What is P(Ai inter Aj)?
ASK&WAIT: What is P(A1 inter A2 inter A3)?
ASK&WAIT: What is P(A) = P(winning)? Should you play an even bet?