According to the American Lung Association (http://www.lunusa.org) 27.6% of American adult men are smokers, as are 22.1% of American adult women. If we assume that 52% of American adults are women, what is the probability that a randomly chosen American adult will be a smoker? Suppose all you know about an adult is that they smoke. What is the probability that they are male? female?

Suppose you wish to test a population for a certain characteristic. For instance, you may wish to test high school athletes for the use of steroids. Unfortunately, tests may be imperfect. They may produce false positives or false negatives. In our case a false positive corresponds to a "clean" athlete who tests positive for steroid use. (What would a false negative be?)

We'll need two definitions: the true negative rate of a test is called its specificity. The true positive rate is called its sensitivity. Suppose you are administering a test with specificity 98% and sensitivity 90%. You already know from other sources that the prevalence of steroid use among student athletes is 1%.

You test an athlete and find that the test indicates steroid use. What is the probability that the athlete has, in fact, used steroids?

Independence: When conditioning provides no information on probabilities two events are said to be *independent*. That is, A is said to be independent of B if

$$P(A|B) = P(A)$$

Show that if A is independent of B then B is independent of A and we may just speak of the independence of two events.

Show that if A and B are independent then $P(A \cap B) = P(A)P(B)$.

When there are more than two events under consideration we need to make a distinction between pairwise independence (weaker) and mutual independence (stronger). Define a collection of sets $A_1, \ldots, A_n$ to be mutually independent if, for any collection of these events $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ we have

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \ldots \cdot P(A_{i_k})$$

Note that pairwise independence does not guarantee mutual independence.

Example: Toss a fair coin 6 times. What is the probability of obtaining 3 heads?

Example: There are 32 people in a room. What is the probability that at least two of them have the same birthday?

Bayes' Theorem: We often search for the *cause* of an event. In the case where the sample space is partitioned into events B_1, B_2, $\ldots$, B_n we may calculate the probability that event B_j is the cause of event A. That is, calculate $P(B_j|A)$.

The Theorem:

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

proof

Example: A bolt factory produces bolts on three production lines: A, B, and C. Line A accounts for 25% of the bolts produced, line B for 35%, and line C for 40%. On line A 5% of the bolts produced are defective, on line B 4% are defective, and on line C 2% are defective. You have a defective bolt. What is the probability that the bolt came from line A? From line B? From line C?

Famous Example Suppose you are rolling dice with a friend. She offers you \$ 1 if no 6 occurs on the next 4 rolls of a fair die. You must pay her \$ 1 if you obtain at least one 6. Is this a fair bet?

The following *MATLAB* code produced the result below.

```
%mere.m
numThrows = 1000000;
numTrials = 4;
gotOne = 0;
for i=1:numThrows
   data = unidrnd(6,1,numTrials);
   for j = 1:numTrials
      if ( data(j) == 6 )
         gotOne = gotOne + 1;
         break
      end
   end
end
gotOne/numThrows
```

RESULT = 0.517549

Discrete Random Variables and Their Distributions

Example: You will toss a fair coin until it comes up Heads. What is the probability that Heads will show up for the first time on the

1. First toss?

2. Second toss?

3. Third toss?

4. Fourth toss?

5. i^{th} toss?

It is often useful to assign numerical values to the outcomes in a sample space. For example, given the two sequences arising from coin tosses

$$H\ T\ H\ H\ T\ T\ H\ H\ T\ H\ T\ H$$

and

$$H\ H\ H\ H\ H\ T\ T\ H\ T\ T\ T\ H$$

we might only be concerned with the number of heads (if, for instance, heads pay \$5). In this case we would consider the two sequences to be identical and only care about the probability of k heads on a 12 toss sequence.

<u>Definition</u>
Given a sample space S, we define a random variable, X, to be a function which maps elements of the sample space, s, to real numbers, $X(s)$.

Notice that a random variable *induces* probabilities on subsets of the real line. For example, consider the toss of a fair coin with 4 tosses. What is the sample space?

Now define a function X on this sample space as

$$X(s) = \textit{number heads} \; - \; \textit{number tails}.$$

What are possible values for $X(s)$? What are the probabilities associated with these values?

<u>Definition</u>
Given a random variable X, define X to be discrete if X may assume a finite or a countably infinite number of values $x = X(s)$. Also, define a function $p(x) = P(X = x)$. $p(x)$ is called the probability function of X or the probability mass function of X.

We have the following:

1. $p(x) \; \geq \; 0 \; \forall \; x$

2. $\sum_x p(x) = 1$