Activity: Discrete Random Variables
Topic: Random Babies Continued

Example 1: Random Babies (cont.)
Recall your analysis of distributing four babies to their mothers at random. You found that there are 24 equally likely outcomes in the sample space of this process. You also calculated the number of matches for each of the outcomes, as follows:

1234: *4*	1243: *2*	1324: *2*	1342: *1*	1423: *1*	1432: *2*
2134: *2*	2143: *0*	2314: *1*	2341: *0*	2413: *0*	2431: *1*
3124: *1*	3142: *0*	3214: *2*	3241: *1*	3412: *0*	3421: *0*
4123: *0*	4132: *1*	4213: *1*	4231: *2*	4312: *0*	4321: *0*

A *random variable* is a function that assigns a real number to each outcome in a sample space. A *discrete random variable* is one whose set of possible values is either finite or countably infinite. [A set is countably infinite if you can identify a first element, a second element, and so on.] The *probability distribution* of a discrete random variable is given by its set of possible values along with their associated probabilities. Random variables are typically denoted by capital letters, while their possible values are usually symbolized by lower case letters.

(a) List the possible values for the random variable X = number of matches in the first row of the table below. Then report the probabilities in the second row of the table:

x					
P(X=x)					

Considered as a function of the possible values x, the probability distribution of a discrete random variable X is called a *probability mass function* (pmf). This function can be displayed with a *line graph*.

(b) Construct a line graph to display the probability mass function for the random variable X = number of matches. [*Hints*: Put the possible values along the horizontal axis and the probabilities along the vertical axis, using vertical lines with heights equal to the probability above each possible value.]

A random variable can also be characterized by its cumulative distribution function (cdf). This function has the entire set of real numbers as its domain, and it outputs the probability that the random variable will take a value less than or equal to its input. Commonly denoted by a capital letter, this cdf can be expressed as F(x)=P(X≤x).

(c) Fill in the following values of the cdf F(x) for the r.v. X = number of matches. [*Hint*: For each input, ask yourself for the probability that there are that many matches or fewer.]

x	-2.3	0	1	1.4	4	6.8375
F(x) =P(X≤x)						

(d) Sketch this cdf F(x) as a function of x, not only for the values of x in the table above but for all real numbers x.

With discrete random variables, the graph of the cumulative distribution function F(x) is a step function. The steps occur at the possible values of X, and the heights of the steps are their probabilities. The cdf must equal (or at least approach) 0 as x gets infinitely small and must equal (or at least approach) 1 as x gets infinitely large. It must be a non-decreasing function.

Example 2 for Homework if You Do Not Finish in Class: Solitaire
A teacher of statistics tried out the solitaire game on her computer. She won 74 times in 444 attempts. Suppose that she decides to start playing again until she wins for the first time. Let the random variable X = number of games that she loses before her **next** win.

(a) What are the possible values of X? Can you identify a first possible value, a second, and so on? Is X a discrete random variable? Explain.

Use 74/444 (1/6) as the probability that she wins a game, and assume that plays of the game are independent.

(b) What has to happen for X to equal 0? What is the probability of this?

(c) For X to equal 1, what has to happen in her first game? What about her second game? What is the probability that both of these vents will occur and so X=1?

(d) Calculate P(X=2). [Hint: What has to happen for her first win to come in her third game?]

(e) Derive a general expression for the probability mass function P(X=k) in terms of k. [*Hint*: Think about what has to happen on the first k games and then what has to happen in game k+1.]

You have just derived the probability mass function for the *geometric distribution*. It applies to situations that consist of trails for which:

- Each trial has two possible outcomes (typically referred to as "success" and "failure").
- The probability of "success" remains constant on each trial (call it p).
- The trials are independent.
- The random variable of interest (call it X) is the number of trials to get the first success.

The probability distribution of a geometric random variable with parameters n and p is given by $P(X = n) = (1-p)^{n-1}p$ for $n = 1,2,\ldots$. It can be shown that the E(X)= $\frac{1}{p}$ and V(X)= $\frac{1-p}{p^2}$.