Chapter 17: Probability Models

Bernoulli Trials:

A situation is called a ***Bernoulli trial*** if it meets the following criteria:

- There are only two possible outcomes (categorized as SUCCESS or FAILURE for each trial
- The probability of success, denoted *p,* is the same for each trial
- The trials are INDEPENDENT

  (Note that the independence assumption is violated whenever we sample without replacement, but is overridden by the 10% condition. As long as we don't sample more than 10% of the population, the probabilities don't change enough to matter.)

1. A new sales gimmick has 30% of the M&M's covered with speckles. These "groovy" candies are mixed randomly with the normal candies as they are put into the bags for distribution and sale. You buy a bag and remove candies one at a time looking for the speckles.

   a. Is this situation a Bernoulli trial? Explain.
   YES, SUCCESS=SPECKLES, FAILURE=NO SPECKLES, p=30%, TRIALS ARE INDEPENDENT (TECHNICALLY NOT INDEPENDENT, BUT CLOSE ENOUGH ACCORDING TO THE 10% RULE)

   b. What's the probability that the first speckled candy is the fourth one we draw from the bag?
   $P(X=4)=(0.7)^3(0.3)$

   c. What's the probability that the first speckled candy is the tenth one?
   $P(X=10)=(0.7)^9(0.3)$

   d. Write a general formula.
   $P(X=k)=(1-p)^{k-1}(p)$

   e. What's the probability we find the first speckled one among the first three we look at?
   $1-P(X=0)=1-(0.7)^3$

   f. How many do we expect to have to check, on average, to find a speckled one?
   $1/p=3.3333333$

Geometric Distributions:

Suppose the random variable *X* = the number of trials required to obtain the first success. Then *X* is a GEOMETRIC RANDOM VARIABLE if:

1. There are only two outcomes: SUCCESS or FAILURE
2. The probability of success *p* is CONSTANT for each observation.
3. The *n* observations are INDEPENDENT
4. The variable of interest is the NUMBER OF OBSERVATIONS UNTIL THE FIRST SUCCESS

Because *n* is not fixed there could be an infinite number of X values. However, the probability that *X* is a very large number is more and more unlikely. Therefore the probability histogram for a geometric distribution is always SKEWED RIGHT.

If *X* is a geometric random variable, it is said to have a GEOMTRIC DISTRIBUTION and is denoted as G(p).

The expected value (mean) of a geometric random variable is 1/p.

The standard deviation of a geometric random variable is $\frac{\sqrt{1-p}}{p}$.

The probability that X is equal to *x* is given by the following formula: $P(X=x)=(1-p)^{x-1}p$

2. Refer back to the M&M's distribution in problem 1.

   a. What's the probability that we'll find two speckled ones in a handful of five candies?
   (I let them try to reason this out. Some will say $(.3)^2(.7)^3$ and we discuss why that strategy doesn't work.)

   b. List all possible combinations of exactly two speckled M&M's in a handful of five candies.
   SSPPP SPSPP SPPSP SPPPS PSSPP PSPSP PSPPS PPSSP PPSPS PPPSS

Binomial Distributions:

Suppose the random variable *X* = the number of successes in *n* observations.

Then *X* is a BINOMIAL RANDOM VARIABLE if:

1. There are only two outcomes: SUCCESS or FAILURE
2. The probability of success *p* is CONSTANT for each observation.
3. The *n* observations are INDEPENDENT
4. There is a FIXED NUMBER *n* of observations.

If *X* is a binomial random variable, it is said to have a BONOMIAL DISTRIBUTION and is denoted as B(n,p).

The expected value (mean) of a binomial random variable is np.

The standard deviation of a binomial random variable is $\sqrt{npq}$.

The PROBABILITY DISTRIBUTION FUNCTION (or PDF) assigns a probability to each value of *X*.

The CUMULATIVE DISTRIBUTION FUNCTION (or CDF) calculates the sum of the probabilities up to *X*.

Example: Suppose each child born to Jay and Kay has probability 0.25 of having blood type O. If Jay and Kay have 5 children, what is the probability that exactly 2 of them have type O blood?

Let X = THE NUMBER OF CHILDREN WITH TYPE 0 BLOOD

1. There are only two outcomes: success (TYPE 0) or failure (NOT TYPE 0)
2. The probability of success (TYPE 0) is CONSTANT for each of the 5 observations.
3. Each of the 5 observations is INDEPENDENT, since one child's blood type will not influence the next child's blood type.
4. There is a fixed number of observations: 5.

So X is a BINOMIAL RANDOM VARIABLE.

The following table shows the probability distribution function (PDF) for the binomial random variable, X.

x	0	1	2	3	4	5
$P(X=x)$	0.2373	0.3955	0.2637	0.0879	0.0146	0.001

P(X = 0) = P(FFFFF) = $(0.75)^5$ = 0.2373

Binompdf (5,0.25,0) = 0.2373
Binompdf (5,0.25,1) = 0.3955
Binompdf (5,0.25,2) = 0.2637
Binompdf (5,0.25,3) = 0.0879
Binompdf (5,0.25,4) = 0.0146
Binompdf (5,0.25,5) = 0.0010

Construct a histogram of the p.d.f. using the window X_1 [0, 6] and $Y_{0.1}$ [0, 1].

The following table shows the cumulative distribution function (__________) for the binomial random variable, X.

x	0	1	2	3	4	5
$P(X=x)$	0.2373	0.3955	0.2637	0.0879	0.0146	0.001
$P(X \leq x)$	0.2373	0.6328	0.8965	0.9844	0.999	1

Binomcdf (5,0.25,0) = 0.2373
Binomcdf (5,0.25,1) = 0.6328
Binomcdf (5,0.25,2) = 0. 8965
Binomcdf (5,0.25,3) = 0. 9844
Binomcdf (5,0.25,4) = 0. 999
Binomcdf (5,0.25,5) = 1

Construct a histogram of the c.d.f. using the window X_1 [0, 6] and $Y_{0.1}$ [0, 1].

1. Suppose I have a group of 4 students and I want to choose 1 of them as a volunteer. In how many ways can I choose 1 out of 4 students?

   Call this "______4 CHOOSE 1_______________**." There are** _____4_____ **ways.**

2. Suppose I have a group of 4 students and I want to choose 2 of them as volunteers. In how many ways can I choose 2 out of 4 students?

   Call this "________4 CHOOSE 2_____________**." There are** ____6______ **ways.**

3. Suppose I have a group of 5 students and I want to choose 1 of them as a volunteer. In how many ways can I choose 1 out of 5 students?

   Call this "________5 CHOOSE 1_____________**." There are** _____5_____ **ways.**

4. Suppose I have a group of 5 students and I want to choose 3 of them as a volunteer. In how many ways can I choose 3 out of 5 students?

   Call this "_________5 CHOOSE 3____________**." There are** _____10_____ **ways.**

5. Suppose I have a group of 20 students and I want to choose 4 of them as a volunteer. In how many ways can I choose 4 out of 20 students?

SSSSFFFFFFFFFFFFFFFF
FSSSSFFFFFFFFFFFFFFF
FSFFFFSFFFSFFFFFFFSF
… and so on…

Call this "________20 CHOOSE 4_____________**." There are** ____4845______ **ways.**

There is a mathematical way to count the total number of ways to arrange *k* out of *n* objects. This is called **"**_______N CHOOSE K___________**"** or the _BINOMIAL COEFFICIENT_.

The binomial coefficient is the number of ways to arrange *k* successes in *n* observations.

It is written ____ $\binom{n}{k}$ ______ and is called "_______N CHOOSE K_____________."

The value of "*n* choose *k*" is given by the formula: $\frac{n!}{k!(n-k)!}$

Example: "5 choose 2" $\frac{5!}{2!(5-2)!} = \frac{5!}{2!3!} = \frac{5 \bullet 4 \bullet 3!}{2!3!} = \frac{20}{2} = 10$

So there are ____10______ ways to arrange 2 out of 5 objects.

Think of this as flipping a coin 5 times and getting 2 heads. In how many ways can that happen? It can happen in _____10_____ ways.

If X is a binomial random variable with parameters *n* and *p*, then

$$P(X = x) = \binom{n}{k} p^k (1-p)^{n-k}$$