Stat 321 - Day 15
Geometric and Negative Binomial Distributions

Example: Solitaire (cont.)
Recall that we earlier (on the day11 handout) considered a statistics teacher who has a probability p of winning a game of Solitaire and who plays until she obtains her first win. We considered the random variable X = number of games that she loses before her next win.

(a) Explain how this random variable fails to meet the conditions for the binomial distribution.

(b) Recall or re-derive the probability mass function for X.

We previously derived the probability mass function of X to be: $P(X=k) = p(1-p)^k$ for $k=0,1,2, \ldots$ This is called a *geometric* distribution with parameter p. It arises in binomial settings where the random variable is not the number of successes in a fixed number of trials, but rather the number of failures preceding the first success. The expected value of the geometric distribution is $E(X)=(1-p)/p$, and the variance is $V(X)=(1-p)/p^2$.

Recall that the teacher in question won 74 times in 444 attempts, so we estimated her success probability to be $p = 74/444 = 1/6$.

(c) Determine the expected number of losses before her first win, and explain why this value makes sense.

(d) Determine the probability that she would lose fewer games than the expected value before her first win.

(e) Determine the fewest games that she could play and have at least a 90% chance of winning at least once.

Consider the random variable Y = number of failures before the rth success. Y is said to have a *negative binomial* distribution with parameters r and p. The probability mass function for Y is: $P(Y=k) = \binom{k+r-1}{r-1} p^{r-1}(1-p)^k$. The expected value and variance are: $E(Y) = r(1-p)/p$ and $V(Y) = r(1-p)/p^2$.

(f) Explain why each of the three terms in the probability mass function makes sense.

(g) Suppose that this teacher's plan is to play until she wins two games. Determine the probability that she requires more than ten games to accomplish this. [*Hint*: Show how you can calculate this either with a negative binomial or a binomial probability.]

Example: Random Words

Computer scientists working with password generation and encryption are interested in the number of English words of various lengths. One statistical way to estimate the number of words of a given length is to generate random strings of letters and see how many of them form legitimate English words.

(h) Open the Minitab worksheet `alphabet.mtw` (which you can find within the "Stat321" folder within the "Rossman" folder on your desktop), in which the 26 letters appear in `c1`. Generate random strings of three letters until you create a genuine English word for the first time. [You can do the sampling in Minitab as follows:

```
MTB> sample 3 c1 c2;
SUBC> replace.
```

Copy and paste these lines until a word appears. Be sure to treat the letters in the order they appear, and remember to count along as the strings go by.] Record the total number of non-words that preceded your first real word. Also record what your word is.

(i) Pool your result with your classmates, and produce a dotplot of the values for the number of non-words preceding the first real English word.

Consider the random variable X = number of non-words that preceded the first English word. Let p represent the probability that a three-letter string forms a genuine English word.

(j) What probability distribution does X have? Explain.

(k) Express the expected value of X in terms of p.

(l) In order to estimate the value of p, it seems reasonable to equate the theoretical expected value with the observed value of the sample mean. Report the value of this sample mean, then do this equating, and finally solve algebraically for p.

(m) Since the three-letter strings are chosen at random, all possible three-letter strings should be equally likely. Thus, p should equal the number of three-letter English words divided by the total number of three-letter strings. Report how many three-letter strings there are, then plug in your estimate for p and solve for the number of three-letter English words.