STAT 325 – Handout 11
Other Discrete Random Variables (3.2, 3.3, 3.5)

We have studied the binomial probability distribution, and now we study four more discrete random variables, all of which are related in some way to the binomial.

Example 11-1: Solitaire (cont.)
Suppose that I play Solitaire until I achieve my first win. My probability of winning any one game is 1/9, independently from game to game. Let the random variable X be the number of games that I play in order to win for the first time.

a) What are the possible values of X?

b) Determine Pr(X = 2).

c) Determine Pr(X = 5).

d) Determine Pr(X = *x*) for any possible value *x*.

- X ~ Geometric(*p*)
  - Counts the number of trials until the first success
  - Where those trials have two possible outcomes (S, F), are independent, and have a constant probability of success (*p*)
  - pmf: $p(x) = p(1-p)^{x-1}$, $x = 1, 2, 3, \ldots$
  - $E(X) = 1/p$; $Var(X) = (1-p)/p^2$

e) Determine and interpret E(X) in this context.

Now suppose that I play until I win twice. Let the random variable Y be the number of games required.

f) What are the possible values of Y?

g) Determine Pr(Y = 2).

h) Determine Pr(Y = 3).

i) Determine Pr(Y = 4).

j) Determine Pr(Y = *y*) for any possible value *y*.

- Y ~ Negative binomial(k, p)
  - Counts the number of trials until the k^{th} success
  - Where those trials have two possible outcomes (S, F), are independent, and have a constant probability of success (p)
  - pmf: $p(y) = \binom{y-1}{k-1} p^k (1-p)^{y-k}$, $y = k, k+1, k+2, \ldots$
  - E(Y) = k/p; Var(Y) = $k(1-p)/p^2$
  - Geometric is special case of negative binomial with $k = 1$

Example 11-2: Women Senators

The 2010 U.S. Senate consists of 17 women and 83 men. Suppose that 4 Senators are selected at random. Consider the random variable X = number of women selected.

a) Explain why X does not have a binomial distribution.

b) Determine the pmf of X.

- X ~ Hypergeometric(N, r, n)
  - Counts number of successes in random sample of n trials selected *without* replacement
  - pmf: $p(x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $x = \max(0, n - (N - r)), \ldots, \min(n, r)$

- $E(X) = n\frac{r}{N}$; $Var(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$
  - Note similarities to binomial, with $p = r/N$
  - $\left(\frac{N-n}{N-1}\right)$ is called finite population correction factor

b) Determine E(X), Var(X), and SD(X).

Now suppose that the sampling were done *with* replacement, so the same Senator could be chosen multiple times. Let Y represent the number of women selected.

c) Determine the probability distribution of Y. Then calculate E(Y), Var(Y), and SD(Y). Comment on how much these differ from when the sampling was done without replacement.

d) Suppose that a population consists of 10,000 people rather than 100. Would the difference between binomial and hypergeometric be more or less pronounced in this case? Explain.

e) Use Minitab to produce pmf graphs to investigate question d), with a population of 4000 men and 6000 women and a sample size of 10.

- X ~ Poisson(μ)
  - pmf: $p(x) = \frac{e^{-\mu}\mu^x}{x!}$, $x = 0, 1, 2, \ldots$
  - E(X) = μ; Var(X) = μ
  - Can be shown to be limit of Bin(n,p) with $\mu = np$ as $n \to \infty$ and $p \to 0$ with np held constant
  - Often used to model number of occurrences of a rare event in some fixed interval of time/space

Example 11-3: Typographical errors
Suppose that the number of typographical errors on a randomly selected page of a textbook is modeled as a Poisson distribution with parameter μ = 0.3.

a) Determine the probability that a randomly selected page has no typographical errors.

b) Determine the probability that a randomly selected page has exactly two typographical errors.

c) Determine the probability that a randomly selected page has at least one typographical error.

Now suppose that you take a random sample of 10 pages.

d) Determine the probability that none of these pages has a typographical error.

e) Determine the probability that at least half of these pages contain at least one typographical error. (*Hint*: First define a new random variable and identify its distribution.)