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 10-1: Solitaire (cont.)**
Suppose that I play Solitaire until I achieve my first win. My probability of winning any one game is .2, 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) Explain why $X$ does not have a binomial distribution.

b) What are the possible values of $X$?

c) Determine $\Pr(X = 2)$.

d) Determine $\Pr(X = 5)$.

e) Determine $\Pr(X = x)$ for any possible value $x$.

- $X \sim \text{Geometric}(p)$
  - 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: $p(x) = p(1-p)^{x-1}, x = 1, 2, 3, \ldots$
  - $E(X) = 1/p; \ Var(X) = (1-p)/p^2$

f) 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.

g) What are the possible values of $Y$?
h) Determine $\Pr(Y = 2)$.

i) Determine $\Pr(Y = 3)$.

j) Determine $\Pr(Y = 4)$ and $\Pr(Y = 5)$.

k) Determine $\Pr(Y = y)$ for any possible value $y$.

l) Now suppose that I play until I win $k$ times rather than just twice, where $k > 2$. Let $W$ be the number of games required. Determine the pmf of $W$.

- $W \sim$ 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: $p(w) = \binom{w-1}{k-1} p^k (1-p)^{w-k}$, $w = k, k+1, k+2, \ldots$
  - $E(W) = k/p$; $\text{Var}(W) = k(1-p)/p^2$
  - Geometric is special case of negative binomial with $k = 1$

**Example 10-2: Women Senators (cont.)**
The 2016 U.S. Senate consists of 20 women and 80 men. Suppose that 4 Senators are selected at random. Consider the random variable $X = \text{number of women selected}$.

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

b) Determine the pmf of $X$. 
• X ~ Hypergeometric(N, r, n)
  o Counts number of successes in random sample of n trials selected without replacement (so the selections are not independent)
  o pmf: \( p(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \), \( x = \max(0, n - (N - r)), \ldots, \min(n, r) \)
  o \( E(X) = n \frac{r}{N} \); \( \text{Var}(X) = n \frac{r}{N} \left( 1 - \frac{r}{N} \right) \frac{N-n}{N-1} \)
    - Note similarities to binomial, with \( p = r/N \)
    - \( \frac{N-n}{N-1} \) is called the finite population correction factor

c) Determine and interpret \( E(X) \). Then calculate \( \text{Var}(X) \), and \( \text{SD}(X) \).

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.

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

e) 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 (with sample size of \( n = 4 \))? Explain.

f) Compare the probability distributions of the random variables \( X_A = \) number of women in a random sample of 4 people from Population A with 80 men and 20 women, \( X_B = \) number of women in a random sample of 4 people from Population B with 8000 men and 2000 women, and the appropriate binomial random variable.
X \sim \text{Poisson}(\mu)

- pdf: \( p(x) = \frac{e^{-\mu} \mu^x}{x!} \), \( x = 0, 1, 2, \ldots \)
- \( \mathbb{E}(X) = \mu \); \( \text{Var}(X) = \mu \)
- 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
- Calculations in R:
  - pdf: \( \text{dpois}(x, \mu) \)
  -cdf: \( \text{ppois}(x, \mu) \)
  -inverse cdf: \( \text{qpois}(\text{prob}, \mu) \)
  -simulation: \( \text{rpois}(\text{numreps}, \mu) \)

**Example 10-3: 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 \( \mu = 0.35 \).

a) Write out the pdf of this random variable. Use calculus to verify that it sums to 1.

b) Use R to produce a graph of this pdf.

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

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

e) Determine the probability that a randomly selected page has at least two typographical errors. *(Hint: Look for the simplest way to determine this. Calculate this by hand and with R.)*
f) Calculate the difference between the previous two answers, and indicate what this is the probability of.

Now suppose that you take a random sample of 25 pages.

g) Determine the expected value of the number of these pages that have at least one typographical error. \([\text{Hint: First define a new random variable and identify its probability distribution.}]\)

h) Determine the probability that at least half of these 25 pages contain at least one typographical error. Show how to calculate this as exactly and efficiently as possible in R.

Now suppose that you sample pages, one at a time, until you find one with at least one typographical error.

i) Determine the expected value of the number of pages that you will need to sample. \([\text{Hint: Once again first define a new random variable and identify its probability distribution.}]\)

j) Determine the probability that you will need to sample more than 5 pages in order to find one with at least one typographical error.