**Homework, Simulating Univariate Distributions**

For all of the following exercises, set the seed to 54321.

1. The geometric distribution can be defined as X, the number of Bernoulli trials needed to get one success. If p is the probablility of success, then





* 1. Use a DATA STEP to generate a random sample of 1,000 observations from the geometric distribution for each of p=.2,.4,.6.
  2. Use a PROC SQL step to create a data set geomnum that contains the number of occurrences for each combination of p and x (the random geometric).
  3. Use PROC SGPLOT to graph the distribution of the number of trials needed for each probability. All of the plots should appear on the same graph. Use a series plot and include markers on the plot.

1. 1. Define macro variables, a and b and obs with values -2 and 2 and 1000 respectively.

Write a data step program that uses these macro variables to generate 1,000 observations from a uniform(a,b) distribution.

* 1. Use a PROC means step to examine the mean, minimum and maximum values of the sample generated and check that these are consistent with being from a uniform(-2,2) distribution.
  2. Include a modified version of your DATA step in a macro definition, unif\_a\_b, that generates the data step SAS code to generate a specified number of observations from a uniform(a,b) distribution (a<b). Use keyword parameters with default values. The definition should be as follows:

**%macro** unif\_a\_b(varnm=x,a=**0**,b=**1**,seed=**54321**,obs=**100**,outdata=unif\_a\_b);

* 1. Use the macro to generate the datastep necessary to obtain a 1000 random observations from a uniform(-1,3). The 1000 observations should be in a data set called tmp.
  2. Use a PROC MEANS step to verify that the sample generated is consistent with being a sample from a uniform(-1,3)

1. The negative binomial distribution describes the number of failures until the kth success in a series of Bernoulli trials. The random variable has distribution function.



* 1. Write an IML step to generate 10,000 trials from a negative binomial, p=.3, k=3. Create a data set, negbin, in the iml step containing the 10,000 random negative binomial observations.
  2. Use proc sql to create a dataset, counts, that contains the number of occurrences of each distinct value.
  3. Analyze the results using sgplot to plot the distribution of the number of failures before the 4th success. Use a series plot and include markers for the observed numbers in the plot.