Problem Set - Distributions

1. Using the properties of expected value and variance, show that the expected value and variance of a binomial distribution are

$$\mathrm{E}\left[X\right] = np$$
$$\mathrm{Var}\left[X\right] = np(1-p)$$

Hint: think about Bernoulli processes are combined to produce a binomial.

2. Write a Matlab function that generates a gaussian distribution given three inputs: 1) the x coordinates for the points to be returned, 2) the mean and 3) the standard deviation.
e.g. f = gaussian(xvals, mean, stdev)

Use this function to generate a gaussian from -6 to 6 with 10000 points, a mean of 0 and a standard deviation of 1. Plot the CDF of this distribution and use the CDF to estimate the probability that a random draw from this gaussian distribution would be

a. 1 standard deviation or more from the mean.
b. 2 standard deviations or more from the mean.
c. 3 standard deviations or more from the mean.

3. Generate 1000 random draws from the gaussian distribution you generated in question 2 using its CDF and uniform random numbers between 0 and 1. Verify the answers to 2a, b and c. You may find it worthwhile to write a general purpose random draw function that takes as its input a CDF and the number of points to draw.

4.a. Generate binomial distributions from a process that has a p(0) = p(1) = 0.5 with 50 trials (e.g. n in the formula = 50). Use the formula for a binomial distribution. For that distribution, come up with a measure that represents the distance between that distribution and a Gaussian distribution with the same mean and variance (Feel free to search the web of measures of the distance between the distributions if you prefer that to making up your own. Repeat with different values of p(1) from 0.05 to 0.45. How do the values of p(0) and p(1) this change the distance and why?

5. In this problem you will use the NORMRND function to generate sets of draws from a normal distribution with mean 10 and standard deviation 5. Generate a 100000 draws and convince yourself, by plotting a histogram, that you are getting a normal distribution.

a) Choose at random 5 samples from the set you generated. Repeat this to get a total of 1000 sets of 5 draws and calculate

1) the mean of the samples in each set (a total of 1000 means)

2) the estimate of the standard deviation of the samples in each set using the formula from class: $s_{n-1} = \sqrt{\sum_{i=1}^{N} \frac{\left(X_i - \bar{X}\right)^2}{n-1}}$

3) the estimate of the standard deviation of the samples in each set using $s_n = \sqrt{\sum_{i=1}^{N} \frac{\left(X_i - \bar{X}\right)^2}{n}}$

4) the estimate of the standard error of the mean computed first as $\frac{s_{n-1}}{\sqrt{n-1}}$ and then as $\frac{s_n}{\sqrt{n}}$.

What does a comparison of your results from 2) and 3) to the true standard deviation suggest?

b) Compare the values you get for the standard error of the mean to the standard deviation of the set of estimates of the mean you got from 5a1) above. Which measure of the standard error of the mean best reflects the actual variability in the means?

c) Plot the pdf and cdf of the 1000 estimates of the standard error of the mean.