STT 810

Homework 3

Due Monday, October 17 at 11:59:59pm

1. For the following 4 discrete distributions, calculate the expected value and standard deviation in following two/ three ways. Compare to make sure you get approximately the same result for each.

(i) create a simulation of 1,000,000 outcomes, and calculate the average and standard deviation of the outcomes;

(ii) use the formulas for expected values and variance for standard distributions (for example, for binomial, $E(X) = np$); and

iii) (for binomial distribution only) use sequences of the possible outcomes to generate the expected value and standard deviation.

(a) binomial, 12 trials, $p = 0.2$
(b) exponential, $\lambda = 0.03$
(c) Poisson, rate = 0.4/minute, t = 20 minutes
(d) Uniform, interval = [0, 6]

2. You are a manager for product development. You have a $500,000 budget to spend on research for a new product. There are two products you can research:

(i) Product 1 has a 30% chance of producing a successful prototype. If a prototype is created, it has an 80% chance of making it to market. If the product makes it to market, the amount of sales it will produce is modeled as an exponential distribution with mean value $9,000,000.

(ii) Product 2 has a 60% chance of producing a successful prototype. If a prototype is created, it has an 90% chance of making it to market. If the product makes it to market, the amount of sales it will produce is modeled as a uniform distribution between $2,000,000 and $8,000,000. However, if successful, there is a 30% chance that the company will be successfully sued for patent infringement, in which case, the company will lose $1 and $3,000,000 (according to a uniform distribution).

(a) Calculate the expected value of the alternatives and determine whether you should give the go-ahead for the development of product 1 or product 2

(b) Create a simulation for two products, with 1,000,000 runs. Confirm that the average value of the simulation is about equal to what you calculated for the alternatives in (a). For percent of the simulations do you at least make your money back for the development?

(c) For what percent of simulations do you make more money for Product 1 vs. Product 2?

(d) Suppose the company needs the revenue to pay off a $7 million loan. Which alternative is more likely to produce more than $7,000,000 in revenue?

3. Define a multivariate normal distribution with mean value = (-1, 2), var(X) = 1, var(Y) = 2, and the correlation of X and Y = -0.3. Plot this distribution, by

(a) Creating a contour plot of the pdf, and

(b) Simulate the distribution 1,000 times and plot the simulated points.

(c) Next, using a larger simulation of the same distribution (1,000,000 values), estimate

(i) E(X)

(ii) E(Y)

(iii) Var(X) and Var(Y)

(iv) E(X + Y) and Var(X + Y). Compare the result to what the results of the formulas for linear combinations of random variables.

(v) E(X|Y = 3)

4. Suppose we have an exponential probability distribution with parameter $\lambda = 1/3$, so that its expected value is 3. We will perform some simulation experiments to determine the behavior of the sample mean.

(a) Generate 2,000 simulations which each constitute 100 outcomes of this distribution. Calculate the mean values for the 2,000 experiments.

(b) Next, calculate the mean value and variance for the 2,000 sample means.

(c) Repeat (a) and (b) for N = 1,000, 10,000, 100,000, and 1,000,000 outcomes (2,000 simulations of each). Note: 1,000,000 could take several minutes.

N	Mean of Sample Means	Variance of Sample Means
100		
1,000		
10,000		
100,000		
1,000,000		

(e) Plot the log(Variance) vs. log(N). Do you get something which is close to a straight line? What is the slope? Does the relationship between N and the variance close to what you would expect with the Central Limit Theorem?

5. Take a probability density function given by $p(x) = 1/x^2$, where x is between 1 and infinity.

(a) Find the cdf. Verify that you get 0 and 1 at the minimum and “maximum” values.

(b) Find the quantile function.

(c) Use the quantile function to simulate the distribution. Recreate the experiment in (4) for this distribution (just create the table; you don’t need to plot). Do you get the same convergence behavior? What is going on?