Stat 331, Homework 5, November 11

This homework assignment contains exercises in simulation of random variables. They are meant to give you a better feeling and understanding of some of the concepts we have introduced in class. You must do all the exercises but you **do not have to turn in your solutions!** Simply **send me an email** and tell me when you are done, and what **your answer is to problem number 6**. The due date is Tuesday, November 11. There will be a help session as usual on November 10.

The following is a brief description of how to use Matlab for simulations. If you prefer, you may use other software that can do the same.

Most mathematical software is able to simulate observations from many distributions. In Matlab this is done with the 'random' command. For example, the command 'x=random('unif',0,1,1,1000)' generates 1000 observations (i.e. observations of $X_1, X_2, ..., X_{1000}$ where the X_k are i.i.d.) from a uniform distribution on (0,1) and saves them in the vector x. Do 'help random' to see how it works and which distributions that are supported.

By using the 'hist' command, you can plot so called histograms for your observations. Histograms are obtained by dividing the range into suitably sized intervals and then plotting the number of observations for each such interval. The histogram gives an idea of the shape of the pdf or pmf the observations come from. Note that Matlab allows you to choose the number of such intervals (called 'bins'). The default is 10 but larger values may sometimes be more instructive.

The commands 'mean(x)' and 'var(x)' gives the sample mean and sample variance of the values in x (and 'std(x)' gives the sample standard deviation).

1. For each of the following distributions, simulate a large (for example 1000) number of observations, compute the sample mean and sample variance and plot a histogram. Compare the sample mean and sample variance to the true mean and variance. Note how the shape of the pdf relates to the shape of the histogram and how it changes for different parameter values.

a. Uniform distribution on (a, b) for some different choices of a and b.

b. Exponential distribution for some different choices of the parameter a.

c. Normal distribution for some different choices of μ and σ^2.

d. Binomial distribution for some different choices of n and p.

2. Matlab allows you to use vectors of parameter values. For example, the commands 'x=[1,10]' followed by 'y=random('unif',0,x,1,2)' generates an observation which is uniform on (0,1) and one which is uniform on (0,10).

a. Describe how this feature can be used to simulate observations on a bivariate normal distribution with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ.

b. Use this to simulate and plot 1000 such observations for bivariate normal distributions with $\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = 1$ and the following values of $\rho : -0.9, 0.5, 0, 0.5$ and 0.9. Note how the shape of the data set changes with the value of ρ.

3. This is to illustrate the Central Limit Theorem.

a. Simulate and plot histograms for a large number of observations from the following binomial distributions. Notice how n is fixed at 100 but how they look less and less normal as the variance in the binomial distribution decreases. **Note:** In Matlab, the mean and standard deviation are given as parameters for the normal distribution, 'norm(10,10)' thus gives observations which have variance 100.

1. $n = 100, p = 0.5$
2. $n = 100, p = 0.1$
3. $n = 100, p = 0.05$
4. $n = 100, p = 0.01$.

b. Let $X_1, X_2, ...$ be independent random variables each having an exp(1)-distribution. Simulate and plot histograms for a large number of observations of the sum $\sum_{k=1}^{n} X_k$ for the following values of n. Notice how the histograms look more and more normal as n increases.

1. $n = 1$
2. $n = 2$
3. $n = 5$
4. $n = 20$

5. $n = 50$

4. This is to illustrate unbiasedness, consistency and approximate normality of the sample mean, $\bar{X}$. Choose any distribution, generate a sample of size n and compute $\bar{X}$. Repeat this a large number of times to get a sample of observations on $\bar{X}$. In this sample, compute the sample mean and sample variance and plot a histogram. Do this for different values of n (for example, 2, 5, 20 and 100). Notice how the sample mean of $\bar{X}$-values is close to the true mean and that the sample variance of $\bar{X}$-values gets lower as n increases. Finally, note how the distribution of $\bar{X}$ looks more normal as n increases.

5. This is to illustrate confidence intervals. In the following six cases, generate 100 samples and compute a confidence interval for the mean μ with confidence level 0.95. Use the formula where the variance is known. In each case count how many of the intervals that contain the mean. Recall that the formula is exact for normal distributions. This means that in **1-3** you should get close to 95 out of 100 intervals containing the mean. In **4-6**, the intervals are approximate. Note how the approximation improves with larger sample size.

1. Normal distribution, $\mu = 1$, $\sigma^2 = 1$, sample size $n = 5$.
2. Normal distribution, $\mu = 1$, $\sigma^2 = 1$, sample size $n = 20$.
3. Normal distribution, $\mu = 1$, $\sigma^2 = 1$, sample size $n = 100$.
4. Exponential distribution, $\lambda = 1$, sample size $n = 5$.
5. Exponential distribution, $\lambda = 1$, sample size $n = 20$.
6. Exponential distribution, $\lambda = 1$, sample size $n = 100$.

6. Simulation can be a very useful method in finding means and variances that are impossible or hard to compute explicitly. As an example, consider the following problem, called the "gambler's ruin". Two players, A and B play a game where a fair coin is flipped. If it shows heads, A pays B one dollar and if it shows tails, B pays A one dollar. Suppose that A starts with a dollars and B with b dollars (both a and b are integers) and let T be the time when one of them is ruined, that is, T is the gambling round after which either A or B has no money left. There is a simple explicit formula for the mean $E[T]$, expressed in terms of a and b, but it is difficult to find it without invoking more advanced theory than we have done in this course (just try and you will see!). Instead, run simulations of the game and try to guess the expression for $E[T]$ based on your simulations.