STAT 515
Homework #9, due Friday, Apr. 6 at 2:30pm
This homework must be submitted electronically to ANGEL. I strongly encourage the use of LaTeX.

Please make every assignment easier to grade by neatly organizing your writeup and clearly labeling your final answers when appropriate. Try using LaTeX!

1. Customers arrive at a post office at a Poisson rate of 8 per hour. There is a single person serving customers, and service times are exponentially distributed (and independent) with mean 5 minutes. Suppose that an arriving customer will decide to wait in line if and only if there are three or fewer people already in the post office (which means two or fewer people in line). Of interest is $E(L)$, the expected total number of potential customers lost, during a single eight-hour day.

   (a) Simulate this eight-hour process 10,000 times, then find the sample mean

   $$\hat{\mu}_1 = \frac{1}{10^4} \sum_{i=1}^{10^4} L_i,$$

   where L_i is the number lost in the ith simulation.

   (b) Prove that $E(L) = 8E(T)$, where T is the total amount of time spent in the state "four people in the post office". (Hint: Use conditioning.)

   (c) For the simulation in part (a), calculate

   $$\hat{\mu}_2 = \frac{8}{10^4} \sum_{i=1}^{10^4} T_i,$$

   where T_i is the total time spent in the four-people state in the ith simulation.

   (d) Both μ_1 and μ_2 are estimators of the same quantity, so which one is better? In a statistical sense, "better" often means "smaller variance". Use the conditional variance formula to prove that

   $$\text{Var } \hat{\mu}_2 < \text{Var } \hat{\mu}_1.$$

   (e) Calculate two separate 95% confidence intervals for μ based on $\hat{\mu}_1$ and $\hat{\mu}_2$. Do your results agree with the finding of part (d)?

2. Describe two different algorithms for simulating from a beta$(3, 2)$ density, one based on an inversion method and the other based on a rejection method. Try simulating the same number of variables (something more than a million) using each method. Does one method appear to be more efficient than the other? Explain every step. If you are using R, you might find the timing function `system.time` useful if you want to compare the algorithms based on their total time.

   As a check on your simulated values, make sure that the sample mean and variance of your variables are close to the theoretical mean and variance of $3/5$ and $1/25$.

3. Suppose we want to conduct a simple hypothesis test to see whether the correlation ρ between X and Y is greater than zero. Let us assume that X and Y come from a bivariate normal distribution, and a sample of 30 points $(X_1, Y_1), \ldots, (X_{30}, Y_{30})$ gives a sample correlation of $\hat{\rho} = 0.3$. We wish to find the p-value for this sample correlation.

   To find the p-value, we need to know the null distribution of the sample correlation. In this case, the exact null distribution (i.e., the distribution of $\hat{\rho}$ when $\rho = 0$) is known, but it is quite complicated. Instead, we shall approximate the p-value using the fact (not proven here) that the distribution of $\hat{\rho}$ depends only on ρ and the sample size n.

   The p-value in this problem is defined to be $P(\hat{\rho} > 0.3 \mid \rho = 0)$. Use 10^5 Monte Carlo samples of size 30 to obtain a 95% confidence interval for the p-value.

4. Suppose (X_1, X_2) are bivariate normal with $EX_1 = EX_2 = 0$ and covariance matrix

   $$\Sigma = \begin{bmatrix} 4 & 4 \\ 4 & 9 \end{bmatrix}.$$

   Use a Monte Carlo method to estimate $P(\max\{|X_1|, |X_2|\} < 1)$ to such a precision that a 99% confidence interval for the true value has a width of no more than $1/1000$.