Computational Statistics
Assignment 2

by James D. Wilson (University of San Francisco)

Directions: For all questions in this assignment, fully answer any question that is asked. Late assignments will automatically have 10 points deducted for each day that they are late.

Quantitative Questions

1. You are told that the cdf of a random variable X is given by

$$F_X(x) = \begin{cases} 0 & \text{for } x < .5 \\ \frac{1}{4} & \text{for } .5 \leqslant x < 1 \\ \frac{3}{8} & \text{for } 1 \leqslant x < 1.75 \\ \frac{1}{2} & \text{for } 1.75 \leqslant x < 3 \\ 1 & \text{for } 3 \leqslant x < \infty. \end{cases}$$

   Find the values this random variable takes and the corresponding probabilities.

2. For a particular computer component, the lifetime of the component in hours is known to have density

$$f(x) = 1, \qquad 0 < x < 1$$

   Let the random variable $Y = e^X$. What is the pdf of Y?

3. Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

   (a) What is the value of C?
   (b) Find $P(X > 1)$.

4. A stick of length 1 is split at a point U that is uniformly distributed over $(0, 1)$. Determine the expected length of the piece that contains the point p, $0 < p < 1$.

5. The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

   Compute $P(X < Y)$.

6. Let $X \sim U(0, 1)$ and $Y \sim \text{Exp}(1)$ be independent. Find the cdf and pdf of $Z = X + Y$.

7. If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, calculate the conditional distribution of X given that $X + Y = n$.

8. Two fair dice are rolled. Find the joint probability mass function of X and Y when X is the largest value obtained on any die and Y is the sum of the values.

Computational Questions

1. Suppose that X_1 and X_2 are independent and identically distributed exponential random variables with *the same* parameter λ. You are interested in knowing the distribution of $Y = X_1 + X_2$.

   (a) Without knowing how to directly calculate the distribution of Y, one could simply simulate from the sum. Using `R`, simulate 10000 draws from Y for $\lambda = 1$. Plot a histogram for the sample and calculate the mean and variance for each.

   (b) Using the convolution formula, calculate the analytical form of probability density function for Y.

   (c) Simulate 10000 draws from the distribution you calculated in (b) and draw a histogram of the sample. Now, use a Kolmogorov Smirnoff test to test the statistical difference between this sample and the sample you obtained in (a). What is the result of your test?

2. A bus travels between the two cities A and B, which are 100 miles apart. If the bus has a breakdown, the distance from the breakdown to city A has a uniform distribution over (0, 100). There is a bus service station in city A, in B, and in the center of the route between A and B. It is suggested that it would be more efficient to have the three stations located 25, 50, and 75 miles, respectively, from A. Conduct a simulation study that compares these two competing strategies and discuss which option you think is advisable.