MPCS 58020 2017: Homework 2

Due date: Monday April 17, 2017 5:30pm (before class)

Solve the following problems and show your work, or if specified in the question write a program to solve the problem. Create a README file describing how the solutions are organized, requirements for running the programs, and how to run them. The README can be short and sweet, this is not a project on github, but always defining a README is a good habit and will make things easier for grading. Upload your solution directory to your private dropbox directory (you will receive an email invite) with name `hw2`. For non programming problems, scanned handwritten work is fine.

Programming Guidelines

You may use any programming language, but not high level functionality specific to the homework problem. In particular, you may use basic linear algebra routines, but not specialized statistics features. You may use uniform random number generation on $(0, 1)$ and integer generation within a range, but not more complicated distributions. If you are unsure if a feature is allowed, ask on slack #general. You may of course use any features you want to debug and check your results, but the main solution cannot use the them. If you are unsure on what language to use, I recommend Python+numpy, Julia, Matlab, or Octave.

For problems 6-8, your program should support running with no arguments and use enough realizations (trials) to get a good estimate of the result without an unreasonably long exeuction time (this can be determined by trial and error). It should take an optional argument specifying the number of simulations. A command line program that prints the results is preferred, but a main script (e.g. in Matlab) is also fine.

Probability

For questions 1-5, your solutions should be exact and not a simulation, except where specified for problem of 4. Consider using simulation to check your answers, but don't include that with your solution set.

1. You ask your neighbor to feed your beloved 17-year-old dog, Charlie, while you are on vacation. If he isn't fed, there is an 80% probability that Charlie will die. If he is fed, there is still a 15% probability that he'll die. You are 90 percent certain that your neighbor will remember to feed Charlie.
   (a) What is the probability that Charlie will be alive when you return?
   (b) If Charlie is dead, what is the probability your neighbor forgot to feed him?

2. The continuous random variable X has the following probability density function, with fixed parameters x_m and α, both of which are greater than zero.

$$f(x) = \begin{cases} 0 & \text{if } x < x_m \\ \dfrac{\alpha x_m^\alpha}{x^{\alpha+1}} & \text{if } x \geq x_m \end{cases}$$

   (a) What is the cumulative distribution function, $F(x)$?
   (b) What is $E[X]$?
   (c) What is $Var(X)$?

3. The lifetime in hours of a certain kind of radio tube is a random variable having a probability density function given by:

$$f(x) = \begin{cases} 0 & \text{if } x \leq 100 \\ \frac{100}{x^2} & \text{if } x > 100 \end{cases}$$

   In a set of 5 radio tubes, what is the probability that at least one tube will need to be replaced within the first 150 hours of operation?

4. The random variables X and Y have a joint PDF specified by:

$$f(x,y) = 2e^{-(x+2y)} \qquad 0 < x < \infty,\, 0 < y < \infty$$

What is $P\{X < Y\}$?. Calculate the answer analytically, and then write a program that simulates drawing from $f(x,y)$ to estimate the probability. Compare your results.

5. If the random variables X and Y have a joint PDF $f(x,y)$ such that:

$$P\{X \in C, Y \in D\} = \int_D \int_C f(x,y)\,dx\,dy$$

Then the *marginal PDFs* $f_X(x)$ and $f_Y(y)$ are the univariate PDFs given by:

$$f_X(x) = \int_D f(x,y)\,dy$$
$$f_Y(y) = \int_C f(x,y)\,dx$$

And it follows that:

$$E[X] = \int_C x f_X(x)\,dx$$
$$E[Y] = \int_D y f_Y(y)\,dy$$

Consider the following PDF:

$$f(x,y) = \frac{9}{10}xy^2 + \frac{1}{5} \qquad 0 < x < 2, \quad 0 < y < 1$$

What is $\mathrm{Cov}(X,Y)$?

Simulation

6. ***Random permutations*** *can be generated using a variation of the discrete inverse transform method, as described in Ross, Example 4b.*

A deck of 50 cards are labeled with the numbers 1,2,...,50. The cards are shuffled and then turned over one card at a time. Say that a "hit" occurs whenever card labeled when the ith card to be turned over is labeled with the number i. Let the random variable X be the total number of hits after all cards have been turned over.

(a) Without a simulation, derive the expected value and variance of X.

(b) Compose and run a simulation to estimate the expected value and variance of X.

7. ***Exponential random variables*** *have the PDF* $f(x) = \lambda e^{-\lambda x}$ *and the CDF* $F(x) = 1 - e^{-\lambda x}\,dx$ *over the interval* $(0, \infty)$.

In many applications, the exponential distribution can describe a continuous quantity that may take on any positive value, but for which larger values are increasingly unlikely. For example, the time it takes for a radioactive particle to to decay is an exponential random variable. Ross, Example 5b, describes how to use the inverse transform method to simulate exponential random variables.

A casualty insurance company has 1000 policyholders, each of whom will independently present a claim in the next month with probability 0.05. Assuming that the amounts of the claims made are independent exponential random variables with mean \$800, use a simulation to estimate the probability that the sum of these claims will exceed \$50,000.

8. In class, we used the rejection method to generate random numbers with PDF:

$$f(x) = 3x^2 \quad 0 < x1$$

by generating numbers from:

$$g(x) = 1 \quad 0 < x < 1$$

. We can make a more efficient implementation by using a different PDF in place of $g(x)$.

(a) Use the rejection method to generate random numbers from $f(x)$ by generating random numbers from:

$$h(x) = 2x \quad 0 < x < 1$$

.

(b) Attempt to find your own h(x) that overall give superior performance.

(c) Compare the efficiency (average time and iterations per random number accepted; and associated variances for both) of the three implementations (with $g(x)$ and $h(x)$).

(d) Derive a performance model for the expected number of iterations per random number accepted (and the associated variance) for both implementations. How does your performance model compare to your results?

Define a main program that runs both methods, and prints out the mean and variance of the generated numbers, the mean and variance of the number of iterations per random number accepted, and the mean and variance of the run time in seconds.

Include a copy of the output of your main program as plaintext in `problem8_output.txt`.

Time Series

These problems cover Chapter 2.2 and 2.3 of Shumway and Stoffer, *Time Series Analysis and its Applications* (the 3rd ed, with 603 pages and a yellow cover) and Lecture 3.

- Problem 2.1, page 78. Use either of the data files from the class dropbox in the `datasets` directory: `jj_tables.txt` or `jj_series.txt`. The series contains the same values as the table, listed in chronological order by (year,quarter).