HOMEWORK 3

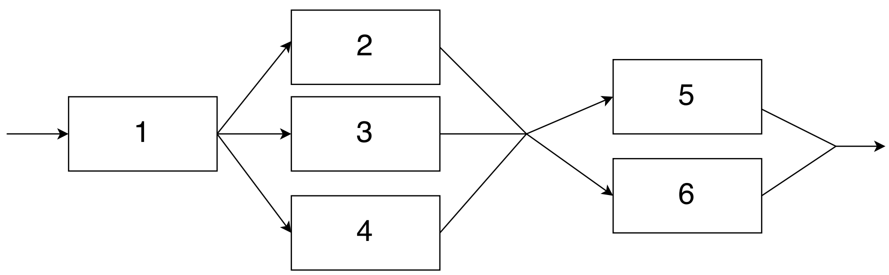
Due Wednesday, March 15th, start of class

You should be working on your homework throughout these two weeks. If you can't solve some of the problems, please come to office hours. Email is fine only for very short questions.

THEORETICAL PORTION

The theoretical problems should be **neatly** numbered, written out, and solved. Do not turn in messy work.

1. Consider a system of 6 components pictured in the following diagram:



For the system to work, **all** of the following have to be satisfied:

- Component 1 has to work.
- At least one of components 2, 3 or 4 has to work.
- At least one of components 5 or 6 has to work.

Now, suppose component 1 has an exponentially distributed lifetime with a mean of 1/2 year. Components 2, 3 and 4 each have an exponentially distributed lifetime with a mean lifetime of 1 year and components 5 and 6 have exponentially distributed lifetimes with mean lifetime of 1.5 year. Assume all components function independently.

What is the probability that the system will function uninterrupted for at least 2 years?

2. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.
   (a) Show (without integration), that $Var(\bar{X}) = \sigma^2/n$.
   (b) What is $E(\bar{X}^2)$?

3. Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and $Y \sim \mathcal{N}(\theta, \gamma)$. Assume X and Y are independent. Show all steps in solving the problems below:
   (a) What is E(X+Y)? Var(X+Y)?
   (b) What is the distribution of X+Y and how do you know?
   (c) What is the distribution of 3X+20Y+8? Include not only the type of distribution but also the values of any parameters of that distribution.

4. Let $X \sim U(a, b)$. What is $E(X^3)$? Show your steps. *Major hint:* $E(X^3) \neq (E(X))^3$*!*

5. You are given a data set of time to recovery after knee surgery for a sample of 22 elderly women.
   (a) Write an equation for the 98% confidence interval you would use for this data, assuming it is normally distributed.
   (b) You make a histogram of the data, and realize it is not normally distributed, but in fact looks like it is exponentially distributed. What could you do to improve on your original confidence interval?

6. Under specific conditions, the drying time of a certain type of paint is normally distributed with a mean value of 55 minutes, and a standard deviation of 7 minutes. Use only the values given to you to solve this problem:

$$\Phi(0.385) = 0.65, \Phi(0.428) = 0.67, \Phi(0.71) = 0.762, \Phi(1.43) = 0.924, \Phi(1.92) = 0.973.$$

$$\Phi^{-1}(0.972) = 1.917, \Phi^{-1}(0.740) = 0.643, \Phi^{-1}(0.95) = 1.645$$

$$z_{0.0017} = 2.923, z_{0.016} = 2.14, z_{0.027} = 1.93.$$

   (a) What is the probability that the paint will dry in 45 minutes or less?

   (b) What is the probability that the paint will dry between 50 and 70 minutes?

   (c) How many minutes represents the 35^{th} percentile in paint drying?

7. A specific rock is randomly selected and weighed two different times. Let w denote the true weight (a number) of the rock, and let X_1 ad X_2 be the two measured weights. Then, $X_1 = w + E_1$ and $X_2 = w + E_2$, where E_1 and E_2 are the two measurement errors. Suppose that E_1 and E_2 are independent and distributed normally with mean 0 and variance 0.1 (*i.e.*, E_1, $E_2 \sim \mathrm{N}(0, 0.1)$).

   (a) What is the mean of X_1? What is the mean of X_2?

   (b) What is $V(X_1)$? What is $V(X_2)$?

   (c) What is $\mathrm{Corr}(X_1, X_2)$?

8. Let

$$f(x) = \begin{cases} \frac{x^3}{4} & \text{if } 0 < x < c \\ 0 & \text{otherwise.} \end{cases}$$

   (a) What value of c is needed so that $f(x)$ is a pdf?

   (b) Calculate $F(x)$.

   (c) In class we discussed that for any cdf, $F(x) \sim U(0, 1)$. For this problem, calculate the inverse of $F(x)$ using this property. i.e., if $F(x) = U$, solve for x in terms of U.

9. **APPM 5570 students only:** Let $X_1, X_2, \ldots, X_{15} \overset{iid}{\sim} f(x)$.

   (a) Write down the equation for a confidence interval for the mean if $f(x)$ is the normal distribution with known parameter σ.

   (b) Write down the equation for a confidence interval for the mean if $f(x)$ is the normal distribution with unknown σ.

   (c) If $f(x)$ is a Poisson distribution with parameter λ, derive a reasonable equation you could use to calculate an α-level confidence interval for the mean.

   (d) If $f(x)$ is a Uniform distribution with parameters a and b, derive a reasonable equation you could use to calculate an α-level confidence interval for the mean.

COMPUTATIONAL PORTION

The computational portion of your homework should be neatly done and include all graphs, code, and comments, labeled and in order based on the problem you are addressing. Do *not* put graphs in at the end, stick code in random locations, or do anything else that will make this homework difficult to read and grade. **LABELS ARE YOUR FRIEND, USE THEM.** If you turn in something that is messy or out of order, it will be returned to you with a zero. All computations should be done using R, which can be downloaded for free at https://cran.r-project.org/.

1. Aphid infestation of fruit rees is usually controlled via pesticides or via ladybug inundation. In a particular area, 2 different (and well isolated) groves, with 15 fruit trees each, are selected for an experiment. The trees in both groves are of the same age, roughly the same size and can be assumed to be independent. One grove is sprayed with pesticides, and one is infested with ladybugs. The fruit yield (in pounds) for each tree is given below:

   Treatment #1, Grove with pesticide:

   $$55.57109,\ 36.50319,\ 47.80090,\ 33.34822,\ 36.16251,$$
   $$35.28337,\ 41.50154,\ 44.18931,\ 40.81439,\ 33.88648,$$
   $$44.90427,\ 49.97089,\ 22.85414,\ 27.84301,\ 38.49843$$

   Treatment #2, Grove with ladybugs:

   $$45.44505,\ 35.52320,\ 46.97865,\ 45.76921,\ 41.66216,$$
   $$54.69599,\ 58.77678,\ 49.08538,\ 48.53812,\ 70.17137,$$
   $$51.86253,\ 39.59365,\ 42.10194,\ 47.39945,\ 39.04648$$

   You can read in this data using the "HW3TreeData.txt" data set.

   (a) Plot a relative frequency histogram of the yields in the two groves. Make sure both histograms have the same range on the x-axis.

   (b) Comment on the histogram shapes. Which densities do they resemble? In particular do they appear normal?

   (c) Find the sample mean of yields for the two groves.

   (d) Assuming both samples come from a normal population and using your answer from part (c), provide the two 95% confidence intervals for the true mean yields for trees under the two treatments.

   (e) Interpret the confidence interval you constructed for the grove treated with pesticides (*i.e.* what does a confidence interval really signify?)

   (f) Find(or approximate) the sample mean variance, *i.e.* variance of the sample mean, $\overline{X}$, for the yields in the two groves.

   (g) Using a chi-square distribution, construct the 95% confidence interval for the true variance of yield for both groves.

2. Generate 500 samples (each with $n = 300$) from an exponential distribution with a mean equal to 1/2. Retain the entire first sample and store it in "expSample", and then in "means.of.expSamples" retain only the means of the remaining 499 samples.

   (a) Report the mean and standard deviation for "expSample" and "means.of.expSamples".

   (b) Create a histogram for "expSample" and for "means.of.expSamples".

   Both of these data sets originate from the same distribution. Why are they so different? What do you notice about the the means and the standard deviations of the samples?

3. Let X be a normally distributed random variable with mean 3 and variance 4.

   (a) Let $Y = 5X + 2$, what is the distribution of Y? What are it's mean and variance? (You don't need the computer for this first one.)

   (b) Find $P(Y < 10)$ and $P(X < 10)$.

(c) What value of Y marks the 67^{th} percentile?

(d) The 25^{th} percentile of the standard normal distribution is -0.674. How can you use this information to find the value of X that marks the 25^{th} percentile?

(e) Use a normal table to calculate the answers for parts (b)-(c). Make sure your answers match. I have no way of knowing if you actually do this or not, but this is a good time to practice this skill.

4. Refer back to theoretical problem number 8. Generate a random sample of size 1000 from your distribution and make a histogram and a boxplot. What features do you notice?

5. **APPM 5570 students only:** Refer back to problem 9 in the theoretical section. Create 3 data sets, where each row of the data set is a sample of 15 i.i.d random variables, and each sample is taken 1000 times (i.e., your dataset is a 1000 by 15 matrix). The random variables are distributed as follows:

   - Dataset 1: Normally distributed with mean 5 and standard deviation equal to 2.3.
   - Dataset 2: Poisson distributed with λ equal to 5.
   - Dataset 3: Uniform distributed with $a = 1, b = 9$.

   For each row of each data set, you should calculate the confidence interval using each of your four methods, and count how many times the true mean (which is equal to 5 for all data sets) is in the interval. Keep track of this for each data set for each type of interval and report these numbers in a table.

   What do you notice about your results? Did anything surprise you?