18.05 Problem Set 4, Spring 2022

Problem 1. (25: 5,5,10,5 pts.) **Time to failure**
Recall that an exponential random variable $X \sim \exp(\lambda)$ has pdf given by $f(x) = \lambda e^{-\lambda x}$ on $x \geq 0$.

(a) Compute $P(X \geq x)$.

(b) Compute the mean and standard deviation of X. You need to set up the necessary integrals, but you can use Wolfram Alpha or another application to do the computation. (Of course, it will be good for you if you compute the integrals by hand!)

(c) Suppose that X_1 and X_2 are independent $\exp(\lambda)$ random variables. Let $T = \min(X_1, X_2)$. Find the cdf and pdf of T. (Hint: first find a formula for $P(T \geq t)$?)

Note: for independent continuous random variables X_1, X_2, you can assume the following formula:

$$P(X_1 \geq x_1,\, X_2 \geq x_2) = P(X_1 \geq x_1)P(X_2 \geq x_2).$$

(d) Suppose we are testing 3 different brands of light bulbs B_1, B_2, and B_3 whose lifetimes are exponential random variables with mean 1/2, 1/3, and 1/5 years, respectively. Assuming that all of the bulbs are independent, what is the expected time before one of the bulb fails. (Hint: part (c) was a warmup for this problem.)

Problem 2. (20: 10,10 pts.) **Elections**
To head the newly formed US Dept. of Statistics, suppose that 50% of the population supports Alessandre, 20% supports Sarah, and the rest are split between Gabriel, Sarah and So Hee. A poll asks 400 random people who they support.

(a) Use the central limit theorem to estimate the probability that at least 52.5% of those polled prefer Alessandre?

(b) Use the central limit theorem to estimate the probability that less than 31% of those polled prefer Gabriel, Sarah or So Hee?

Problem 3. (10 pts.) **A penny for your thoughts**
To save a mint, in 2012 Canada decided to do away with its pennies. The Chubby Chef in Equality, Illinois wants to be ready should the United States decide to pass a similar law. The Chubby Chef processes $n = 1000$ orders of assorted baked goods each day, and will round the price of each order to the nearest nickel (e.g., \$3.57 rounds to \$3.55 while \$3.58 rounds to \$3.60). Let p be the probability that the total rounding error over the course of a day is either greater than 100 or less than -100 cents, i.e. exceeds 100 in absolute value.

Estimate p using the central limit theorem.

Extra credit 5 points Simulate this in R with 10000 trials. (Each trial involves 1000 orders.) Print out or hand copy your code and include it. Give the result of running your code 3 times,

Problem 4. (30: 10,10,10 pts.) **Change of scale.**
In this problem we will look at scaling random variables. This is a simple, but common

thing to do. As usual with transformations, if you don't approach it systematically, it is easy to make mistakes.

(a) Suppose the random variable X has an exponential distribution with parameter 1, i.e. $X \sim \texttt{exp}(1)$. Give the range and pdf for the variables X and $Y = 3X$

Sketch the graph of the density functions for each of these variables.

(b) For the random variable X from part (a), find the range and pdf of $W = aX + b$, where a and b are constants. Assume $a > 0$.

(c) Let $V = X^3$. Find the range and pdf of V.

Problem 5. (30: 10,10,10 pts.) In this problem we will explore how the transformations in the previous problem affect the mean and median.

(a) For the variables X, Y, W in the previous problem, assume each of the variables are given in units of minutes. Find the expected value, variance and standard deviation of each variable. Be sure to include units in your answer.

What are the units on a and b in the defnition of W?

(b) For V from the previous problem, compute $E[V]$. As usual, you must set up the integral, but you can use a package like Wolfram Alpha to compute the integral.

(c) Compute the median value of both X and V.

Problem 6. (30: 5,5,10,10 pts.) **Fat tails**
This problem will explore the tails of two distributions. The tails are important when we want to think about probabilities of extreme events.

(a) As an example, in the general population IQ has mean 100 and standard deviation of 15. IQ is normally distributed. Use the R function `pnorm` to give the probability that a randomly chosen person has IQ greater than 160, i.e. more than 4 standard deviations above the mean.

(b) Now, in order to be able to use R or Wolfram Alpha without a lot of distracting algebraic manipulation, we'll modify the definition of IQ. Suppose that Modified_IQ has mean 0 and standard deviation $\sqrt{3}$.

Assuming Modified_IQ is normally distributed, find the probability that a randomly chosen person has Modified_IQ more than 4 standard deviations above the mean.

(c) Now assume that Modified_IQ follows a t-distribution with 3 degrees of freedom. Later in the class we will work extensively with t-distributions. Here, it will be enough for us to know the following about this distribution.

- Range: $(-\infty, \infty)$
- PDF: $f(x) = \dfrac{2}{3\pi}\left(1 + \dfrac{x^2}{3}\right)^{-2}$
- Mean: $\mu = 0$
- Standard deviation: $\sigma = \sqrt{3}$

(So this has the same mean and standard deviation as in part (b).)

For this problem, you can work with this pdf directly or you can look up how to use the R functions `dt` and `pt`.

Assuming it follows this t-distribution, find the probability that a randomly chosen person has Modified_IQ more than 4 standard deviations above the mean.

You can use R or another calculation package to do the calculation, but you must explicitly show the integral in terms of the probability density.

Compare this value with the probability in part (b)

(d) For this problem, compute probabilities using both the normal distribution in part (b) and the t-distribution in part (c). Do this for the following probabilities.

(i) $P(\text{Modified\_IQ} > 20)$, (ii) $P(\text{Modified\_IQ} > 40)$, (iii) $P(\text{Modified\_IQ} > 200)$.

Why do we say that the t-distribution has a 'fat tail'?

Hence the moral of this problem: Knowing the mean and standard deviation of a quantity is often not enough for predicting the frequency of extreme events (high IQ, 100-year floods, etc.); you need to know the underlying distribution itself (which often requires finding out the underlying geophysics, geochemistry, and biology). In the solutions we will show you graphs of these distributions zoomed in around 4σ above the mean. If you do that yourself, you will see that they look very different.