# Probability and Statistics – Assignment 3

**Release Date**: $10^{th}$ October 2020

**Submission Deadline:** $20^{th}$ October 2020, 11:55 PM

**Instructions:**

- There is a total of 2 sections and each section consists of 5 questions.
- All questions are compulsory.
- Upload your submissions to Gradescope. Do not forget to assign page number(s) to each question in Gradescope when you submit.
- Use of calculator is permitted.
- Each question in Section 1 carries 4 marks and each question in Section 2 carries 6 marks.
- **Please underline, circle, or leave space after the answer** to any question subpart. Please write your answers linearly, from the top of the page to the bottom.

## Section 1

### Question 1

You are given the probability distribution function (PDF) of a continuous random variable $X$ is $f_X(x)$. Let $Y$ be a continuous random variable such that $Y = aX + b$, where $a$ and $b$ are non-zero real constants.

1. Find the PDF of $Y$ in terms of $f_X$, $a$, and $b$. [2 points]
2. Let $X$ be an exponential random variable with parameter $\lambda$. When will $Y$ also be an exponential random variable? [1 point]
3. Let $X$ be a normal random variable with mean $\mu$ and variance $\sigma^2$ . When will $Y$ also be a normal random variable? [1 point]

**Note**: Subparts 2 and 3 must be answered using only the formula derived in subpart 1. You can use Method of Transforms in subpart 1 however it is not necessary to use it.

## Question 2

Let $X$ be a normal distribution with mean $\mu$ and variance $\sigma^2$ i.e. $X \sim N(\mu, \sigma^2)$. If $Y = e^X$, then find the

1. CDF of $Y$. You can leave the answer of CDF in terms of $\phi$.
2. PDF of $Y$. Solve it in two ways: directly using Method of Transformations, and without using Method of Transformations.

## Question 3

Find an approximation to the probability that the number of 2s obtained in $12{,}000$ rolls of a die are between $1900$ and $2150$ (non-inclusive). Give up to three decimal places correct. [Hint: Use Normal Approximation of Binomial Distribution – read here https://online.stat.psu.edu/stat414/lesson/28/28.1]

## Question 4

1. You visit your friends. They have 2 children. They tell you one of them is male. What is the probability that the other is female? [0.5 points]
2. Now the parent says that his child, Atsi, is a boy. Tell me what the probability is that the other child is Female. [0.5 points]
3. Are the answers to the 2 subparts different? If so, why? Just naming the boy provides no additional information, right? Math should not change with arbitrary nomenclature, and yet it seems to, can you resolve the paradox? [2 points]
4. Can you use the answer in sub-part (a) to compute the answer to sub-part (b) using probabilistic update from any "information you gained"? Write out the computation explicitly. [1 point]

**Note for Sub-part C:** You answer should be why physically, the problem before and after telling the name is different, and not be a facet of the mathematical representation of the problem. **Hint:** Try to show if any information is being gained or lost from naming the boy.

## Question 5

Let a random variable $Z$ be the sum of random variables $X$ and $Y$, $Z = X + Y$.

Determine the PDF $f_Z(z)$ for the following cases:

1. $X$ and $Y$ are independent exponential RVs, with common parameter $\lambda$.
2. $X$ and $Y$ are independent uniform random variables in the common interval $(0,1)$.

# Section 2

## Question 6

You arrive to the class with a delay of time $T_1$. The professor arrives to the class with a delay of time $T_2$. Both $T_1$ and $T_2$ are exponential variables with parameter lambda. Both $T_1$ and $T_2$ are independent. Let the difference in time $T = T_1 - T_2$.

1. Find the CDF of $T$ when $T \geq 0$.
   [4 points]
2. Find the CDF of $T$, when $T < 0$, based on your answer to the previous part. (Do not recompute the answer from scratch, for example, by integration)
   [1 point]
3. Using the CDF, you derived in subparts 1 and 2, write down the PDF of $T$.
   [1 point]
4. What is this PDF popularly known as? [0 points, for curiosity sake]

## Question 7

Atsi arrives at a railway station at a random time. It is a hypothetical world and hence trains arrive exactly on time and the time between two successive trains is exactly $10$ minutes. Atsi will wait and take the next train that arrives on the station after his arrival. Given that the time that Atsi arrives is uniformly random and the trains arrive $24$ hours a day,

1. What is the mean time in minutes that Atsi must wait for the next train to arrive? What is the distribution of the waiting time of Atsi?

2. What is the probability that Atsi will have to wait at least 3 more minutes if he has already been waiting for 6 minutes?
3. In the real world, after the arrival of a train, the time until the next train arrives is an exponential random variable with a mean of 10 minutes. Atsi arrives at the station not knowing how long ago the previous train had come. What is the average time he must wait for the next train to arrive? What is the distribution of Atsi's waiting time?

## Question 8

Prove that -

1. The waiting time for a Poisson distribution is Exponentially distributed. i.e. For a Poisson process with parameter $\lambda$, the time for a new arrival is exponentially distributed with same parameter $\lambda$.
2. The waiting time for a Bernoulli process is Geometric.
3. An Exponential distribution can be obtained as a limit of a Geometric Distribution.

## Question 9

1. Compute the following for the Rayleigh Distribution:

$$R(x;\sigma) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

   a. Mean of the distribution [1.5 point]
   b. Variance of the distribution [1.5 points]
   c. Mode of the distribution [0.5 points]

2. For the Cauchy distribution:

$$C(x;\gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]}$$

   a. Derive the mean of the distribution [1.5 point]
   b. Write (without derivation) the characteristic function for the distribution. What are the values of its derivatives at 0? [0.5 points]

c. Explain why higher moments do not exist for the distribution. Explain what it means for variance not to be defined. [0.5 points]

## Question 10

Let $C$ be a unit circle centered at origin. Point $P1$ is chosen randomly from the circumference of $C$. Point $P2$ is chosen randomly anywhere within the circle. Imagine a rectangle with the line segment $P1 - P2$ as its diagonal and sides parallel to $x$ and $y$ axes. What is the probability that no point of the rectangle lies outside the circle $C$?

Assumption: Points $P1$ and $P2$ are chosen independently and uniformly over their respective domains.