Name: ______________________________ PUID ______________________________

STAT 350 Worksheet #7

Throughout our exploration of probability, we have primarily focused on discrete random variables whose possible outcomes can be listed and whose probabilities are assigned to those distinct values. However, many real-world phenomena such as measurement data, time intervals, and physical quantities are more naturally modeled by **continuous** random variables. Unlike their discrete counterparts, continuous random variables can take on any value within an interval, and probabilities are determined via **integration** of a **probability density function (PDF)** rather than by summation of a probability mass function (PMF). In this worksheet, we will introduce the core concepts of continuous random variables in their *general form* without restricting ourselves to well-known named distributions. We will examine how to define PDFs such that they are valid and explore the utility of cumulative distribution functions (CDFs), and how to compute probabilities and expectations.

A **continuous random variable** is a random variable that can take on an infinite number of possible values within a given range. Unlike discrete random variables, which have a countable set of possible outcomes, continuous random variables are associated with probability distributions where individual points have zero probability. Instead, probabilities are determined by the **probability density function (PDF)**, and the probability of an event occurring within a given interval is found by integrating the PDF over that interval. The total area under the PDF curve must equal one, ensuring that the variable adheres to probability rules.

A probability distribution for a **continuous random variable** $\boldsymbol{X}$ is given by a smooth curve called a **density curve**, or **probability density function** (**pdf**) and is defined as:

$$\boldsymbol{f_X(x) = \lim_{\Delta \to 0^+} \frac{P(x < X \leq x + \Delta)}{\Delta}}.$$

The **support** of a **continuous random variable** (r.v.) $\boldsymbol{X}$, is the set of all possible values for which the **probability density function** is **strictly positive**.

$$\mathbf{Supp}(\boldsymbol{X}) = \{\boldsymbol{x} \in \mathbb{R} \mid \boldsymbol{f_X(x) > 0}\}.$$

This definition formalizes the idea that the **PDF describes how probability is distributed around a particular point**, rather than assigning probability to any single outcome. It measures how **densely** probability accumulates near individual values of $\boldsymbol{x}$, much like a derivative in calculus captures the rate of change of a function. However, because a continuous random variable does not assign positive probability to individual points i.e., $\boldsymbol{P(X = x) = 0}$, we compute actual probabilities by integrating over an interval:

$$\boldsymbol{P(a < X < b) = \int_a^b f_X(x)dx}.$$

A probability density function (PDF) must satisfy the following conditions to be considered a valid probability density function:

1. **Non-Negativity**: The function must always be **non-negative** for all possible values of $\boldsymbol{x}$, since probabilities cannot be negative.
$$\boldsymbol{f_X(x) \geq 0, \forall x \in \mathbb{R}}$$
2. **Total Probability Equals One**: The total area under the probability density curve must be **exactly 1**, ensuring that the function represents a proper probability distribution.
$$\boldsymbol{\int_{-\infty}^{+\infty} f_X(x)dx = 1}$$

Both conditions must be checked before declaring a function to be a valid probability density function. A reasonable approach to checking the first condition (non-negativity) is to analyze the function **algebraically** by determining its roots and ensuring that $\boldsymbol{f_X(x) > 0}$ within its support, since it is defined to be zero elsewhere. This involves solving $f_X(x) = 0$ and verifying that $\boldsymbol{f_X(x) > 0}$ within the support by testing values or using derivative analysis. Alternatively, the recommended approach in this course to use **graphical verification** provide a rough graph of the function and confirm that it does not dip below zero at any point within the support.

Review Calculus: Review basic integration rules and specific functions with a specific focus on the following. Rules for **linearity**, the **power rule**, and the distinction between **definite** and **indefinite integrals**. Be sure to revisit the **Fundamental Theorem of Calculus**, as it plays a key role in probability calculations. Additionally, recall methods for changing variables, such as **u-substitution**, and techniques for integrating products of functions, such as **integration by parts**. Another important concept is the ability to **split an integral over multiple subintervals**, which allows us to compute probabilities over piecewise-defined functions. You should also be comfortable integrating common functions, including **constants, polynomials, exponentials, and piecewise functions**. We will not be using partial fraction decomposition or trigonometric substitution, so focus on the essential operations that will be necessary for working with probability distributions.

1. For the following functions determine if it is a valid probability density function.

a) Is the following function $\boldsymbol{f_X(x)}$ a valid pdf?

$$f_X(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \le x < 2 \\ 0, & \text{otherwise} \end{cases}$$

b) Is the following function $\boldsymbol{f_Y(y)}$ a valid pdf?

$$f_Y(y) = \begin{cases} \frac{5}{y^2}, & y \ge 5 \\ 0, & \text{otherwise} \end{cases}$$

c) Is the following function $\boldsymbol{f_Z(z)}$ a valid pdf?

$$f_Z(z) = \begin{cases} \frac{1}{88}(3z^2 - 9) & 1 \le z \le 5 \\ 0 & \text{otherwise} \end{cases}$$

d) Is the following function $\boldsymbol{f_V(v)}$ a valid pdf? For this problem first try to first rewrite the function in a piecewise form.

$$f_V(v) = \frac{\lambda}{2} e^{-\lambda|x|}, \forall x \in \mathbb{R} \text{ and } \lambda > 0$$

2. For the following functions **determine the constant** that would make them a valid probability density function.

a) Determine the constant $\boldsymbol{c}$ that would make $\boldsymbol{f_X(x)}$ a valid probability density function.

$$f_X(x) = \begin{cases} \frac{1}{5} & -1 \le x \le 0 \\ \frac{1}{5} + cx & 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

b) Determine the constant $\boldsymbol{k}$ that would make $\boldsymbol{f_Y(y)}$ a valid probability density function.

$$f_Y(y) = \begin{cases} k\,(y-6) & 6 \le y \le 8 \\ 2y & 8 < y \le 12 \\ \mathrm{k}\,(14-y) & 12 < y \le 14 \\ 0 & \textit{otherwise} \end{cases}$$

The **expected value of a continuous random variable** is the **continuously weighted average** of all values within the support of the random variable.

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

The rules we learned about in discrete cases, i.e., **linearity of expectation, additivity,** and **LOTUS** also apply in the continuous case. We also have an analogous version of variance in the continuous case:

$$\sigma_X^2 = \mathrm{Var}(X) = E[(X-\mu_X)^2] = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f_X(x) dx$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right)^2 = E[X^2] - (E[X])^2$$

Similarly the rules for variance hold in the continuous case, i.e., shifting by a constant does not change variance and if a random variable is scaled by a constant, its variance is scaled by the constant squared: $\mathbf{Var(aX+b)} = \mathbf{a^2 Var}(X)$, and the additivity of variance holds when we have independent random variables: $\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$.

3. Determine the expected value and standard deviation for the random variable X with pdf given below.

$$f_X(x) = \begin{cases} \mathbf{1/20} & \mathbf{-10 < x \le -5} \\ \mathbf{1/4} & \mathbf{-1 < x \le 1} \\ \mathbf{1/20} & \mathbf{5 < x \le 10} \\ \mathbf{0} & \textbf{otherwise} \end{cases}$$

The **Cumulative Distribution Function (CDF)** provides the probability that a random variable X is less than or equal to a given value x, making it a function of x. As the name suggests, the **CDF** accumulates the probability from the start of the support up to the value x rather than simply being an antiderivative of the probability density function (PDF). This distinction is especially important in **piecewise PDFs**, where we must ensure that the CDF correctly accumulates probability across different regions, maintaining continuity and ensuring that it reaches 1 at the upper bound of the distribution. We use $F_X(x)$ to denote the CDF for X and it must satisfy the following properties:

- The CDF is a non-decreasing function: For any $a, b \in \mathbb{R}$, if $a < b$, then $F_X(a) \le F_X(b)$.
- Limiting behavior: $\lim_{x \to -\infty} F_X(x) = \mathbf{0}$ and $\lim_{x \to +\infty} F_X(x) = \mathbf{1}.$
- Right Continuous Convention: $\lim_{x \to c^+} F_X(x) = F_X(c).$

4. Answer the following questions for the given probability density function $f_X(x)$ given below.

$$f_X(x) = \begin{cases} 1 - e^{-\frac{1}{4}} & 0 \le x \le 1 \\ \frac{1}{16} e^{-\frac{x}{16}} & x \ge 4 \\ 0 & \text{otherwise} \end{cases}$$

a) How many regions would we need to be concerned with for the cumulative distribution function?

b) Find the cumulative distribution function for the random variable X defined by the PDF $f_X(x)$.

c) Use the CDF to determine the probability that a trial from this random variable would be less than ½.

d) Use the CDF to evaluate the following probability $P(X > 5)$.

e) Use the CDF to evaluate the following probability $P(X < 5 | X > 1)$.