STAT 350 Worksheet #8

We have previously discussed the general form of probability density functions and cumulative distribution functions, just as we began with general probability mass functions before exploring special named distributions for discrete random variables. Among continuous random variables, certain widely used named distributions also capture common patterns and simplify many analyses. Two fundamental examples are the **Uniform** distribution, which represents outcomes uniformly distributed over an interval, and the **Exponential** distribution, which models waiting times between events occurring at a constant average rate.

This worksheet will focus on these two continuous distributions, examining their probability density functions, cumulative distribution functions, and properties such as expected value and variance. We will see how each distribution is applied in practice and how it relates to the general framework introduced earlier.

The **Uniform distribution** is one of the most straightforward continuous distributions. It arises when the probability density is **constant** across a specified interval $[a, b]$, meaning that any sub-interval of fixed length within $[a, b]$ has the same probability as any other sub-interval of equal length. This simplicity makes the Uniform distribution a natural starting point for modeling phenomena where each outcome within a range is equally plausible. It is also frequently employed in **simulation** to generate pseudo-random numbers and serves as a **non-informative prior in Bayesian analysis**, reflecting no initial preference among values within the specified interval.

A continuous random variable X that is Uniform on the interval $[a, b]$ has a constant probability density function **(PDF)** within that range and zero outside it. Formally,

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

such a random variable we denote as $X \sim \textbf{Uniform}(a, b)$ where a and b are the parameters of the distribution indicating the end points of the support. Its corresponding cumulative distribution function **(CDF)** increases linearly over the support and is given by

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x < b \\ 1 & x \ge b \end{cases}$$

As before, when dealing with named distributions, we have well-established formulas for their expected values and variances in terms of their parameters. The expected value is simply the midpoint of the support and this point represents both the mean and median of the distribution, i.e.,

$$E[X] = \frac{a+b}{2}.$$

The variance formula is easily obtained as a function of its parameters as well

$$\textbf{Var}(X) = E[(X - E[X])^2]$$

$$= \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{1}{b-a}\int_{-\left(\frac{b-a}{2}\right)}^{\frac{b-a}{2}} u^2 \, du$$

$$= \frac{1}{3(b-a)}\left(\frac{(b-a)^3}{8} + \frac{(b-a)^3}{8}\right) = \frac{(b-a)^2}{12}$$

where we simplified the calculations using u-substitution and found $\textbf{Var}(X) = \frac{(b-a)^2}{12}$ and therefore $SD(X) = \frac{b-a}{\sqrt{12}}$. These formulas stem from integrating the distribution's probability density function over its support and simplify many calculations, allowing us to quickly assess key characteristics of the distribution without starting from first principles every time.

1. In this question, you will explore additional properties of the uniform distribution. You will see how the length of sub-intervals determines the probability assigned within the support, reflecting the constant nature of the distribution. Moreover, you will learn that percentiles for a uniform distribution are found by solving simple linear equations, leveraging the distribution's linear cumulative distribution function. Let X be a continuous **Uniform random variable** over an unspecified support $[a, b]$.
    a) Use the **CDF** to determine a formula for computing $P(X \le c)$ where c is some constant in $[a, b]$.

    b) Use the complement rule and the **CDF** to determine a formula for computing $P(X > c)$ where c is some constant in $[a, b]$.

c) Use properties of **CDFs** to determine a formula for computing $P(c \leq X \leq d)$ where $c, d \in [a, b]$ with $d > c$.

d) Use your understanding of **conditional probabilities** and **CDFs** to determine a formula for computing $P(X > d | X > c)$ where $c, d \in [a, b]$ with $d > c$.

e) In **words**, explain how these formulas illustrate the fact that the **Uniform distribution** assigns probability in direct proportion to the length of sub-intervals within $[a, b]$.

f) Given a continuous Uniform random variable X over the interval $[a, b]$, find the p-th **percentile**, x_p in $[a, b]$ for which $P(X \leq x_p) = p$. Your answer should be a function of a, b and p.

g) Now extend your understanding of the uniform distribution by determining the **upper p-th percentile.** In other words, find the value x_p in $[a, b]$ for which $P(X > x_p) = p$. Your answer should be a function of a, b and $(1 - p)$.

2. A company is testing **two independent manufacturing processes**, A and B, each with its cost of producing an item modeled by a Uniform distribution. Since **lower costs** are preferred, management wants to compare these processes under various scenarios:

- The cost of process A follows a Uniform distribution $X \sim \textbf{Uniform}(10, 30)$.
- The cost of process B follows a Uniform distribution $Y \sim \textbf{Uniform}(12, 28)$.

a) Which process is more likely to produce an item costing at most 18?

b) Determine the 75th percentile for each process.

c) If you require that at least 75% of your items fall below a certain cost threshold, which process has the lower threshold?

d) Determine the expected value and variability of each process.

e) For which probability level p (with $0 < p < 1$) do the two processes yield the **same p-th percentile** cost?

f) What are the advantages and disadvantages of each process? Is one process exclusively better?

The Exponential distribution is a cornerstone for modeling **waiting times** or **inter-arrival times** under a **constant average rate**. It is closely related to the Poisson distribution: when the total number of events in a fixed interval is governed by a Poisson random variable, the time between consecutive events follows an **Exponential distribution**. Furthermore, if we know exactly n events occur in a fixed interval $[a, b]$, then each event time is **Uniformly distributed** within $[a, b]$, and the number of events that land in any sub-interval $[a, c]$ follows a **Binomial distribution**. Furthermore, the sum of multiple independent Exponential waiting times follows a **Gamma distribution**, which is another key connection within the family of continuous distributions.

These relationships highlight how the Exponential distribution naturally bridges continuous waiting-time models (Poisson inter-arrivals) with discrete counts (Poisson, Binomial) and uniform spacing. In this worksheet, you will examine the Exponential distribution's probability density function and cumulative distribution function, along with its defining **memoryless property**, which distinguishes it from other continuous named distributions.

An **Exponential** random variable X with rate parameter $\lambda > 0$ models the waiting time (or interarrival time) until an event occurs, assuming events happen at a constant average rate and independently of each other. We denote such a variable as $X \sim \text{Exp}(\lambda)$. Its **probability density function (PDF)** is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and its corresponding cumulative distribution function (CDF) is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

As with other named distributions, the Exponential distribution has **well-established formulas** for its expected value and variance in terms of its parameter λ. By using integration by parts and integrating the PDF over its support $[0, \infty)$ one finds

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} \, dx = \frac{1}{\lambda}$$

$$\text{Var}(X) = \int_0^\infty (x - E[X])^2 \, \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

Consequently, the standard deviation is simply $\frac{1}{\lambda}$. These results simplify many calculations and allow us to **quickly assess** the core characteristics of an Exponential random variable without deriving everything from first principles each time.

Sometimes we reparametrize the Exponential distribution in terms of its **mean** $\mu = \frac{1}{\lambda}$, then we may write $X \sim \text{Exp}(\mu)$. Under this parameterization, the probability density function (PDF) becomes

$$f_X(x) = \begin{cases} \dfrac{1}{\mu}e^{-\frac{1}{\mu}x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and its corresponding cumulative distribution function (CDF) is

$$F_X(x) = \begin{cases} 1 - e^{-\frac{x}{\mu}}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

In practical settings, μ can be more intuitive: if events occur on average every 10 hours, it might feel more direct to say $\mu = 10$ then to use $\lambda = 0.1$ which denotes that on average 0.1 events happen every hour. Both forms describe the **same** Exponential distribution, but choosing the parameter that matches the **natural measure** of your scenario can simplify modeling and interpretation.

3. In this question, you will explore additional properties of the **Exponential distribution**.
   a) Use the CDF and complement rule to derive a formula for $P(X > c)$ for any $c \geq 0$. In other words, express the probability that X exceeds a particular value c in terms of λ and c. This probability is often referred to as a tail probability.

   b) A hallmark of the **Exponential distribution** is its **memorylessness**: once a certain amount of time has passed without an event, the additional waiting time until that event occurs has the **same** distribution as the original waiting time. This property is useful for modeling processes like time-to-failure of components where the past does not influence the future. Using properties of conditional probabilities and the tail probability formula show that X satisfies the memoryless property:
   $$P(X > t \mid X > s) = P(X > t - s), \text{ for any } 0 \leq s < t.$$

4. An elevator in the Purdue Math building experiences breakdowns with a **constant average rate** of 1 breakdown every 90 days. Let X denote the time (in days) until the next breakdown. We assume X follows an Exponential distribution.
   a) Determine the probability that the elevator goes at least 60 days without breaking down.

   b) Suppose the elevator has already run for 30 days without breaking down. Determine the probability of running an additional 20 days without a breakdown.

   c) Determine the top 10th percentile time; which is the point beyond which only 10% of breakdowns exceed this value.

   d) Maintenance will observe the elevator over a 180-day span, let Y denote the number of breakdowns in that period. Determine the probability that there are at least 3 breakdowns over this period.