Continuous Random Variables and the Uniform Distribution

1. Continuous Random Variables
    a. Variables that can take on an uncountably infinite number of different values
    b. For example, if X can be any value between 0 and 1, X is a continuous random variable because there are an infinite number of values between 0 and 1, and we cannot enumerate them
    c. The probability X is exactly 1 value is 0 because there are infinite possibilities
    d. We talk instead about the probability of X taking on some value within an interval
    e. We characterize the probability distribution using a curve where the area under the curve within some interval gives the probability of X falling within that interval
2. Probability Density Function (PDF)
    a. This function cannot compute the probability of the variable attaining 1 particular value (that probability is 0, as stated before)
    b. This function is a tool used for finding the probability of the variable falling in some interval
    c. The PDF is similar to the Probability Mass Function (PMF) for discrete random variables, but they are not the same
    d. They are both denoted as f(x), but whereas f(X=x) gives the probability of X attaining some value in a discrete distribution, f(X=x) is meaningless in a continuous distribution
3. Uniform Distribution
    a. This is a probability distribution that characterizes a continuous random variable for which all possible values are equally likely
    b. Consider a variable that can take on all values in the interval [a,b]
    c. f(x) looks like a rectangle with width (b - a) and height $\frac{1}{b-a}$
    d. $f(x) = \frac{1}{b-a} \quad for\ \{a \leq x \leq b\}$
4. Cumulative Distribution Function (CDF)

a. Again, Pr(X = x) = 0 for a continuous distribution
b. The CDF is a function that outputs the probability that the variable will take on a value within the specified interval
c. This function actually gives the area under the PDF curve for that interval
d. For the uniform distribution:
    i. $CDF = F(x) = Pr(X \leq x) = \frac{x-a}{b-a}$
    ii. b is the maximum value that X could possibly take
    iii. a is the minimum value that X could possibly take
    iv. x is the upper limit of the interval of interest
    v. F(b) = 1 such that the probability of X taking on any of its possible values is 1
    vi. If d is some number between a and b, $Pr(X \leq d) = F(D)$
    vii. If c is some number between a and b,
    $Pr(X > c) = 1 - Pr(X \leq c) = 1 - F(c)$
    $Pr(c \leq X \leq d) = F(d) - F(c)$

Expectations of the Uniform Distribution

1) PDF: $f(x) = \frac{1}{b-a}$
2) Mean: $\mu = \frac{a+b}{2}$
3) Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
4) Standard Deviation: $\sigma = \sqrt{\frac{(b-a)^2}{12}}$

The Normal Distribution

1) Parameters and Characteristics of the Normal Distribution
   a) Center of the distribution is determined by the mean, μ
   b) The spread (width of the distribution) is determined by the standard deviation, σ
   c) We denote that a continuous random variable is described by the normal distribution by writing $X \sim N(\mu, \sigma)$
   d) This distribution is shaped like a bell, symmetrical about the mean, and unimodal (the mode of the distribution is only 1 number)
   e) Often the mean is also the mode and median
2) Empirical Rule
   a) Recall that the area under the curve within some interval gives the probability of finding a sample within that interval
   b) The distribution is symmetrical about the mean, so the mean splits the area under the curve half and half
   c) 68% of the area is contained within 1 standard deviation of the mean
   d) 95% of the area is contained within 2 standard deviations of the mean
   e) 99.7% of the area is contained within 3 standard deviations of the mean
   f) This rule can be used to calculate probabilities of intervals that fall on 1, 2, or 3 standard deviations from the mean
3) The Standard Normal Distribution and Z-Transformations
   a) The Standard Normal Distribution is a normal distribution with $\mu = 0$ and $\sigma = 1$
   b) The Cumulative Distribution Function (CDF) of any old normal distribution is difficult to calculate
   c) The CDF of the Standard Normal Distribution is usually tabulated in textbooks as a look-up table
   d) Z-transformations allow us to take any normal distribution and turn it into a Standard Normal Distribution
      $$z = \frac{x - \mu}{\sigma}$$

e) To go back to your specific normal distribution, you can do a reverse Z-transformation
$x = z\sigma + \mu$
f) If you want to calculate the probability of some interval on a normal distribution, you can map the boundaries of that interval onto the Standard Normal Distribution using the Z-transformation and find the probability of that region in a look-up table

4) Approximating the Binomial and Poisson Distributions
   a) The normal distribution can be used to approximate the Binomial and Poisson distributions when the calculations become unwieldy
   b) The normal distribution has two parameters, μ and σ, so we need to find these parameters within the Binomial and Poisson distributions in order to get a good approximation
   c) Binomial Distribution
      i) Mean: $\mu = np$
      ii) Standard Deviation: $\sigma = \sqrt{np(1-p)}$
      iii) We can then plug these into the Z-Transformation: $z = \frac{x-np}{\sqrt{np(1-p)}}$
      iv) This approximation is good when n is very large, np is greater than 10, and n(1-p) is greater than 10
   d) Poisson Distribution
      i) Mean: $\mu = \lambda$
      ii) Standard Deviation: $\sigma = \sqrt{\lambda}$
      iii) $z = \frac{x-\lambda}{\sqrt{\lambda}}$
      iv) This approximation is good for very large λ