DSSC 221 – Probability and Statistics
Lab 4

(추가 패키지 설치) 아래 pip 명령어를 통해 sympy 설치

```
!pip install --upgrade pip

!pip install --upgrade

!pip install sympy
```

1. Generate vector of 10000 random numbers
    a. Plot CDF—what do you think distribution is.
    b. Find
        i. The **mean** of the numbers generated in a).
        ii. The **mean of the square** of the numbers generated in a).
        iii. The **variance** of the numbers generated in a) using the Python variance function. Show that this equals the result of (ii) minus the square of the result of (i).

2. Let's consider the **pipe leak location** problem, in which the probability density function increases linearly from the left end of the pipe to the right end of the pipe and is twice as high on the right end as the left end. Recall the that the PDF and CDF are:

$$f_X(x) = \frac{2}{3L}(1+\frac{x}{L}) \quad 0 < x < L$$
$$f_X(x) = 0 \quad otherwise$$
$$F_X(x) = \int_0^x \frac{2}{3L}(1+\frac{w}{L})dw = \frac{2}{3L}(x+\frac{x^2}{2L}) \quad 0 \le x \le L$$
$$= 1 \quad x \le L$$

   a. Solve the CDF formula for x. In other words, suppose $F_X(x) = p$, what must x be? Write formula for x as a function of p.

   b. You can now use the random numbers from 1. to generate random values of the leak location with the distribution specified above. The trick is to recognize that if we draw a value from a distribution, the CDF for that value is equally likely to take any value between 0 and 1. For each of the 10000 random values created in step 1, apply the formula from 2a. to find an associated leak location. In other words, think of each of the random numbers as a value for p and use the relationship you developed in 2a to find a value for x. Assume $L = 1$.

   c. Plot the cumulative distribution for the results in b). Does it look like what you expect?

   d. Plot a histogram for the results in b). Does it look like what you expect?

   e. Find

      i. The mean of the numbers generated in b).

      ii. The mean of the square of the numbers generated in b).

      iii. The variance of the numbers generated in b) using the Python variance function. Show that this equals the result of (ii) minus the square of the result of (i).

   f. **Challenge Question (Extra pts.) -** The PDF and CDF formulas assume that the pipe location is in inches. Suppose you want the variance in cm^2. There are two ways to make the correct conversion.

      i. Using the result in 2eiii, calculate the variance in cm^2. $1cm = 0.3937"$. (Hint: you can define a variable as: inch2cm=1/0.3937)

      ii. Convert the result in 2b to centimeters and calculate the variance.

   Compare i and ii.

3. Let's see how the expectation and variance changes when we add two random variables.

   a. **Challenge Question (Extra pts.) -** Generate two random vectors X and Y, of size 10000 following uniform distribution between 0 and 1.

      i. Calculate E(X) and E(Y)

      ii. Calculate E(X+Y)

      iii. Calculate E(2X+3Y)

      Compare i,ii,iii

      iv. Calculate Var(X) and Var(Y)

      v. Calculate Var(X+Y) and COV(X,Y)

      Compare Var(X)+Var(Y)+2COV(X,Y) and Var(X+Y)

   b. **Challenge Question (Extra pts.) -** Generate two random vectors A and B of size 10000 when A follows Uniform distribution between 0 and 1 and B follows Normal distribution with $\mu = 0$ and $\mu = 1$. (Hint: you can generate the normal random variables using `stats.norm.rvs` as we generated uniform distribution. E.g. `stats.norm.rvs(size=n)`)

      i. Calculate E(A) and E(B)

      ii. Calculate E(A+B)

      iii. Calculate E(2A+3B)

      Compare E(A)+E(B) and E(A+B)

      vi. Calculate Var(A) and Var(B)

      vii. Calculate Var(A+B) and COV(A,B)

      Compare Var(A)+Var(B)+2COV(A,B) and Var(A+B)