Lab 05 – Random Variable

Trần Lương Quốc Đại
tlqdai@it.tdt.edu.vn

Nguyễn Quốc Bình
nqbinh@it.tdt.edu.vn

1 Definitions

"Variable" is something that can be changed, "random" is unknown. Thus, "random variable" is something that can be changed and cannot be determined what it is.

In the dice experiment, we define the random variable X = {1, 2, 3, 4, 5, 6} corresponding to the outcome of the face 1, 2, 3, 4, 5, or 6; It is clear that the result after each experiment is not determined (random).

For example, launch 8,000 times

```
import random

x = [random.randint(1,6) for i in range(8000)]
X = set(x)
print(x)
print(X)
```

2 Random variable classification

Random events can be classified into two categories: discrete random variables and continuous random variables.

• Discrete random variables: random variables are called "discrete" when their values are finite or infinitely countable x_1, x_2, ..., x_n. In other words, discrete random variables when we can list all the values of that random variable.

• Continuous random variable: random variable is called "continuous" in a range of values if its possible values fill that value range.

3 Probability distribution

A probability distribution function of a random variable is the rule that shows how to assign each probability to each value of a random variable.

Determine the probability distribution function of the dice experiment based on experimental results.

```
P = [x.count(i)/len(x) for i in X]
print(P)
```

4 Cumulative distribution function – CDF

In probability theory, the cumulative distribution function fully describes the probability distribution of a random variable X with real value. For each real number x, the cumulative distribution function is defined as follows:

$$F(x) = P(X \leq x) = \sum_{x_i < x} p_i, x \in \mathbb{R} \tag{1}$$

Calling faces 1, 2, 3 is small; faces 4, 5, 6 are big. Calculate the probability to achieve a small result:

```
FX = sum(P[:3])
print(FX)
```

5 Characteristic parameters of random variables

5.1 Expectations

Suppose the random variable X receives one of the possible values $x_1, x_2, x_3, \ldots, x_n$ with the corresponding probability $p_1, p_2, p_3, \ldots, p_n$. Then the expectation of the random variable X is the sum of products between possible values of random variables with corresponding probabilities, symbol μ or E (X) and determined by the formula:

$$\mu = E(X) = \sum_{i=1}^{n} x_i p_x \tag{2}$$

Determine the expectation of random variable X in dice experiment:

```
EX = 0
for x in X:
    EX = EX + (x * P[x-1]);
print(EX)
```

5.2 Variance

In probability and statistical theory, the variance of a random variable is the measure of the statistical dispersion of that random variable, which indicates that the values of the random variable are often how for from expected values.

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p_i \tag{3}$$

Determine the variance of random variable X in dice experiment:

```
VarX = 0
for x in X:
    VarX = VarX + (x-EX)*(x-EX)*P[x-1]
print(VarX)
```

5.3 Standard Deviation

Variance denotes the dispersion of random variables, but not the same unit with those random variables. According to (3) it is easy to see the measurement unit of the variance by square of the unit of the random variable. Therefore, we give a new parameter that has the same meaning as the variance, but with the same unit as the random variable. That quantity is called the standard deviation and is denoted by σ.

The standard deviation can be calculated by taking the square root of the variance:

$$\sigma = \sqrt{Var(X)} \tag{4}$$

Determine the standard deviation of the random variable X in the dice experiment:

```
import math
SD = math.sqrt(VarX)
print(SD)
```

5.4 Standard Score - SC

The standard deviation allows us to know the average level of dispersion of the entire data set, but does not tell us the dispersion level of a certain point. Therefore, we add one more parameter to evaluate this point as the Standard Score (SC).

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

6 Exercises

Exercise 1: Simulate a Discrete Random Variable

Simulate a discrete random variable representing the roll of a six-sided die. Compute and plot the probability mass function (pmf).

1. Simulate rolling a die 10,000 times.
2. Count the occurrences of each outcome (1-6).
3. Calculate the pmf.

Exercise 2: Continuous Random Variable and PDF

Define a continuous random variable that follows a normal distribution with mean 0 and standard deviation 1. Calculate and plot the probability density function (pdf).

1. Use scipy.stats.norm to define the distribution.
2. Generate x values from -4 to 4.
3. Calculate the pdf for those x values.

Exercise 3: Cumulative Distribution Function (CDF)

Calculate the cumulative distribution function (cdf) for a uniform distribution on the interval [0, 1].

1. Use numpy to create an array of x values from 0 to 1.
2. Compute the cdf using the formula $F(x) = \frac{x-a}{b-a}$

Exercise 4: Expectation of a Discrete Random Variable

Compute the expectation (mean) of a discrete random variable defined by its pmf.

1. Define a discrete random variable with the following pmf:
    - P(X=1) = 0.2
    - P(X=2) = 0.5
    - P(X=3) = 0.3
2. Calculate the expectation E(X).
3. Print the result.

Exercise 5: Variance of a Random Variable

Calculate the variance of the random variable defined in Exercise 4.

1. Use the same pmf from Exercise 4.
2. Calculate the variance using the formula Var(X) = E(X^2) - (E(X))^2
3. Print the variance.

Exercise 6:

The number of emergency cases in a hospital in a day is a random variable. The hospital collected data for 10 years and saved in variable x as follows:

```
x = np.random.choice([0, 1, 2, 3, 4, 5], 3650, p=[0.1, 0.2, 0.3,
    0.2, 0.15, 0.05])

# your code
```

Know that each value in list x is the number of emergency cases in a day. For example: x = [0, 5, 2] means that the data consists of 3 days with the corresponding number of emergency cases per day of 0, 5, 2. Write the program to execute:

(a) Determine the values of the random variable X and store it in variable X.

(b) Calculate the probability distribution function of the random variable X and store it in variable P (list type).

(c) Calculate the characteristic parameters of random variable X including: expectation, variance, standard deviation.

(d) Calculate the probability of having 3 or more emergency cases.

Exercise 7:

Write a program for flipping 2 coins experiment in 10,000 times, calling X the number of heads that appear on each flip.

(a) Save the results of flipping 2 coins into the variable x (list type). For example, x = [0, 2, 1] represents 3 times of launching, the first time no coin has a head, the second time both of them have a head, the third time has only one head.

(b) Find the values of random variable X and save to variable X.

(c) Calculate the probability distribution function of the random variable X and store it in variable P (list type).

(d) Calculate the characteristic parameters of random variable X including: expectation, variance, standard deviation.

(e) Calculate the probability to have at least one head.