통계분석

Statistical Analysis

Assignment 03. Random Variables and Probability Distributions

Q1. Binomial Distribution vs Normal Distribution

Here let us compare the following binomial distribution and normal distribution for given n (the number of trials) and p (probability of success for one individual trial):

- **binomial distribution**

$$b(x;n,p) =_n C_x p^x (1-p)^{n-x}$$

- **normal distribution**

$$f\left(x;\mu=np,\sigma=\sqrt{np(1-p)}\right)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

(1) Plot binomial distribution and normal distribution and compare them when n=10,100, and 1000 and p is fixed to be 0.5. For example, when n=10 and p=0.5, you can plot the following two distributions in the same subplot,

$$b(x;n=10,p=0.5)$$

$$f\left(x;\mu=10\times 0.5,\sigma=\sqrt{10\times 0.5\times 0.5}\right)$$

Repeat the above plotting and comparison for n=100 and 1000.

(2) Describe what you observe when you do (1).

(3) What do you observe when you change p from 0.5 to 0.6 in (1)?

Q2. Root Mean Squared Deviations

When you have two sequences of numbers, $(x_1, x_2, \cdots, x_n)$ and $(y_1, y_2, \cdots, y_n)$ the root mean square deviations (RMSD) is defined as

$$(\mathrm{RMSD}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

In Python, you can calculate the RMSD for two numpy array x and y as follows:

```
x = np.array([1,3,4,5,7])
y = np.array([3,4,2,6,8])

RMSD = np.sqrt(np.sum((x-y)**2)/len(x))

print(RMSD)
```

```
1.4832396974191326
```

Now calculate the RMSD between binomial and normal distributions in Q1 by using n=10, 100, 1000 when p=0.5.

Which case has the smallest RMSD? What can you conclude from these observations?

Q3. Student's t-distribution vs Standard Normal distribution

Using Q1 and Q2, and Python, show that student's t-distribution approximately approaches the standard normal distribution when we increase the degree of freedom.

- ***student's t distribution***

$$f(t;\nu) = t_\nu = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

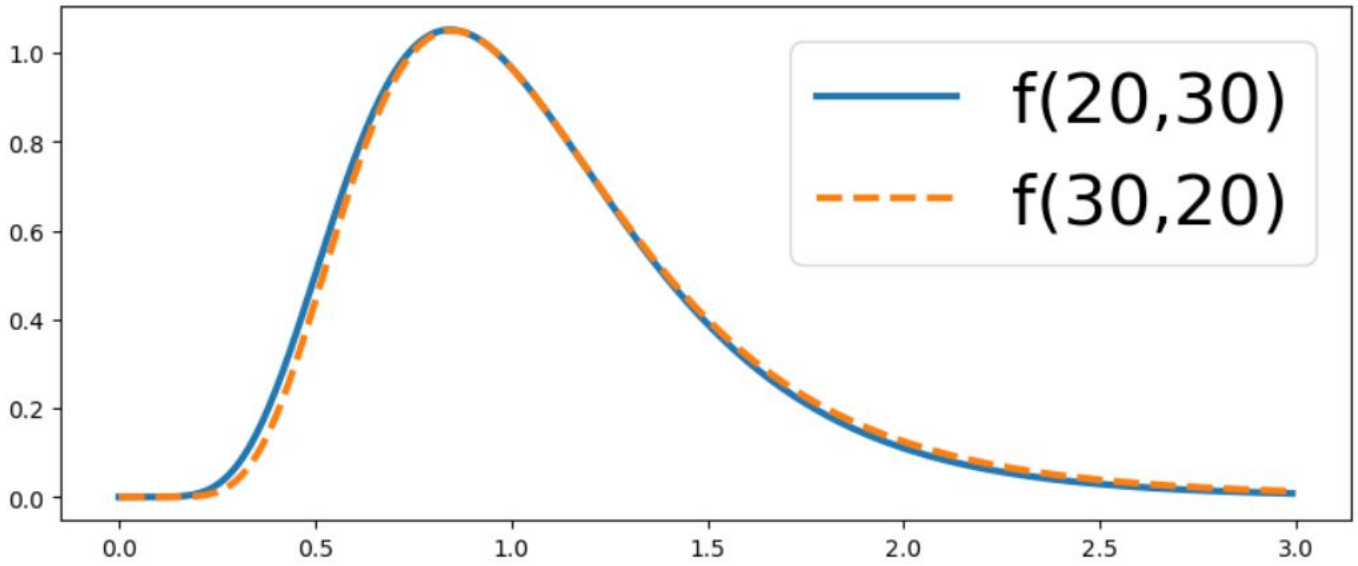
- ***standard normal distribution***

$$N(z,0,1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$$

Q4. Percentiles of F-distribution

Let us consider two F-distributions as follows:

$$f_1 = f(x; \nu_1 = 20, \nu_2 = 30) \equiv f(20, 30)$$
$$f_2 = f(x; \nu_1 = 30, \nu_2 = 20) \equiv f(30, 20)$$



(1) Calculate 95 percentile of f(20, 30) and 5 percentile of f(30,20). When you multiply 95 percentile of f(20, 30) and 5 percentile of f(30,20), what value do you have?

(2) Now consider another pair of F-distributions: f(10,40) and f(40,10). Repeat the same calculations of (1). From these two examples, what can you guess?

Submissions

- Please write down your kaggle notebook file to answer Q1 to Q4. Make a Python code to solve the assignment and submit your kaggle notebook file to https://class.ust.ac.kr.
- Due date is 2023. 05. 10.