Lab 06 – Discrete Probability Distributions

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

Nguyễn Quốc Bình
nqbinh@it.tdt.edu.vn

Required libs: math, numpy, matplotlib.

```python
import matplotlib.pyplot as plt
import math
```

1 Bernoulli distribution

Bernoulli's distribution (named after the Swiss mathematician Jacob Bernoulli) is a discrete probability distribution of random variables that only takes two values, 0 or 1, where value 1 is obtained with probability of success p and value 0 is received with failure probability q = 1 - p.

If the random variable X follows this distribution, denote X ~ Bernoulli (p). The probability density function of the Bernoulli distribution is determined by the formula:

$$p(x) = P(X = x) = \begin{cases} p & \text{nếu x=1} \\ 1 - p & \text{nếu x=0} \end{cases} \qquad (1)$$

Or in another form:

$$p(x) = P(X = x) = p^x(1 - p)^{1-x} \text{ với } x \in \{0, 1\} \qquad (2)$$

For example, call X is an event of throwing a coin of a homogeneous coin, if the coin appears a head X = 1, otherwise X = 0. The probability of success (being head) is p = 0.5. What is the probability of getting a head? Answer: 0.5^0 $(1 - 0.5)^{1-0}$ = 0.5.

Write the probability density function of Bernoulli distribution:

```python
def pmf_bernoulli(p, x):
    # your code
```

Using that pmf_bernoulli function to draw a graph which shows the relationship between the random variable X and its corresponding probability. The horizontal axis represents the x value of the random variable, the vertical axis represents the probability p (x), respectively:

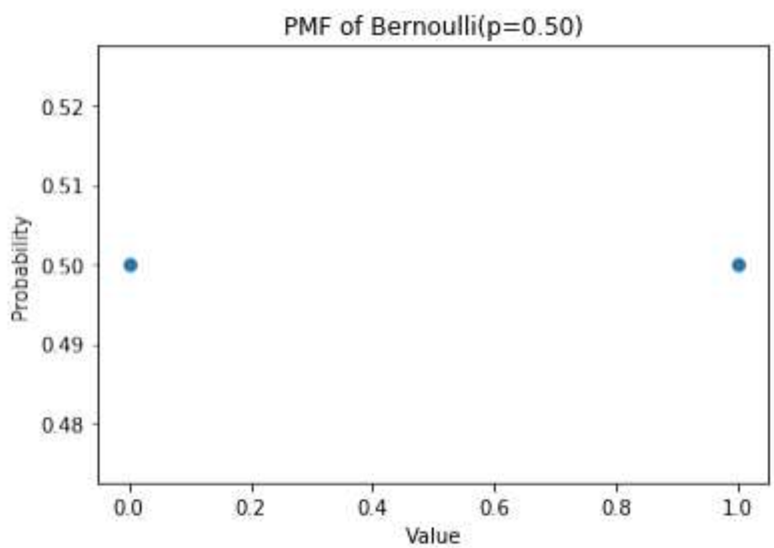
```
def plot_pmf_bernoulli(p):
    '''

    Plot the probability mass function of Bernoulli(p)
    '''
    X = [0, 1]
    P_bernoulli = [pmf_bernoulli(p, x) for x in X]
    plt.plot(X, P_bernoulli, 'o')

    plt.title('PMF of Bernoulli(p=%.2f)' % (p))
    plt.xlabel('Value')
    plt.ylabel('Probability')
    plt.show()

plot_pmf_bernoulli(0.5)
```

The result is:



2 Binomial distribution

The binomial distribution is a discrete probability distribution, which takes two parameters n and k with k as the number of successful tests in n independent tests.

The probability density function of the binomial distribution is determined by the formula:

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ với } \begin{cases} k = 0, 1, 2, ..., n \\ \binom{n}{k} = \frac{n!}{k!(n-k)!} \end{cases} \tag{3}$$

For example, toss a coin 15 times, what is the probability to get exactly 4 times of head, know that the probability of going getting head in each try is 0.5. Answer: $\binom{15}{4} 0.5^4 (1 - 0.5)^{15-4}$.

Write the probability density function of the binomial distribution:

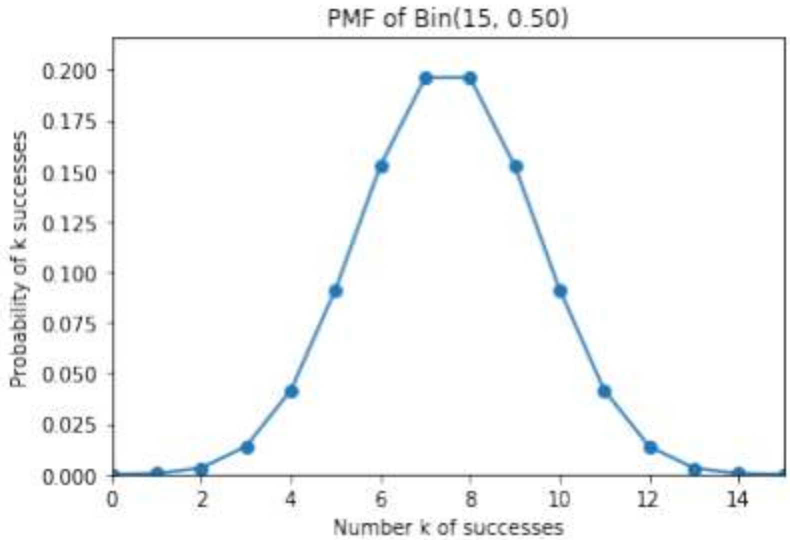
```
def pmf_binom(k, n, p):
    # your code
```

Using the pmf_binom function above, graph the relationship between k tests in the binomial distribution and the corresponding probability of the above example. We have, n = 15 tests, the probability of success at each try is p = 0.5; the horizontal axis represents k tests, the vertical axis represents the probability p (k) respectively:

```
def plot_pmf_binom(n, p):
    '''
    Plot the probability mass function of Binom(n, p)
    '''
    K = list(range(0, n + 1))
    P_binom = [pmf_binom(k, n, p) for k in K]
    plt.plot(K, P_binom, '-o')
    axes = plt.gca()
    axes.set_xlim([0, n])
    axes.set_ylim([0, 1.1 * max(P_binom)])
    plt.title('PMF of Bin(%i, %.2f)' % (n, p))
    plt.xlabel('Number k of successes')
    plt.ylabel('Probability of k successes')
    plt.show()

plot_pmf_binom(15, 0.5)
```

The result is:



3 Poisson distribution

In probability and statistical theory, Poisson distribution is a discrete probability distribution that indicates the average number of successful occurrences of an event in a given time period. This average value is denoted as lambda (λ).

Let X be a random variable whose event occurs randomly and discretely, we count its occurrences in a given time interval t, expected value or average number of times that that random variable that happens in the period that t is λ. So the value of the random variable is the number of successful occurrences of the event (symbol k). And the probability density function indicates the probability that k will succeed in the test.

The probability density function is determined by the formula:

$$p(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{4}$$

For example, a consulting company surveys that there are averages 5 calls for advice per minute. So called X is the number of incoming calls for consulting during the period t = 1 minute, then X follows Poisson distribution with λ = 5. What is the probability of having 10 incoming calls in 1 minute? Answer: $\frac{5^{10}e^{-5}}{10!} = 0.018.$

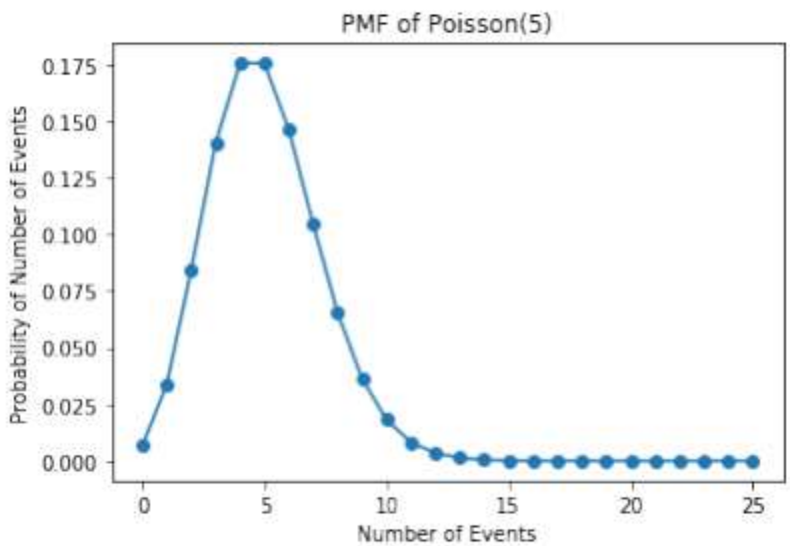
```
def pmf_poisson(k, lam):
    # your code
```

Using the pmf_poisson function above to graph the relationship between the number of occurrences in the Poisson distribution and the corresponding probability in the above example. Let k = 0,1, ..., 25, the horizontal axis represent k the test, the vertical axis represents the probability p (k; λ) respectively:

```
def plot_pmf_poisson(n, lam):
    '''
    Plot the probability mass function of Poisson(n, lambda)
    '''
    K = list(range(0, n + 1))
    P_poisson = [pmf_poisson(k, lam) for k in K]
    plt.plot(K, P_poisson, '-o')
    plt.title('PMF of Poisson(%i)' %lam)
    plt.xlabel('Number of Events')
    plt.ylabel('Probability of Number of Events')
    plt.show()

plot_pmf_poisson(25, 5)
```

The result is:



4 Geometric distribution

Geometric Distribution is the distribution of the probability of the first occurrence of event X in the Bernoulli test. The probability density function of the geometric distribution is determined by the formula:

$$p(x) = P(X = x) = p(1 - p)^x \tag{5}$$

For example, in a game, candidates are given a ring and have to throw the ring on the hook from a fixed distance. As observed, only 30% of candidates can do this. So if the candidate has 5 throws, what is the probability that he will win the prize if he misses 4 times? Answer: 0.3 (1 - 0.3) $^{5-1}$ = 0.072.

Write the probability density function of Poisson distribution:

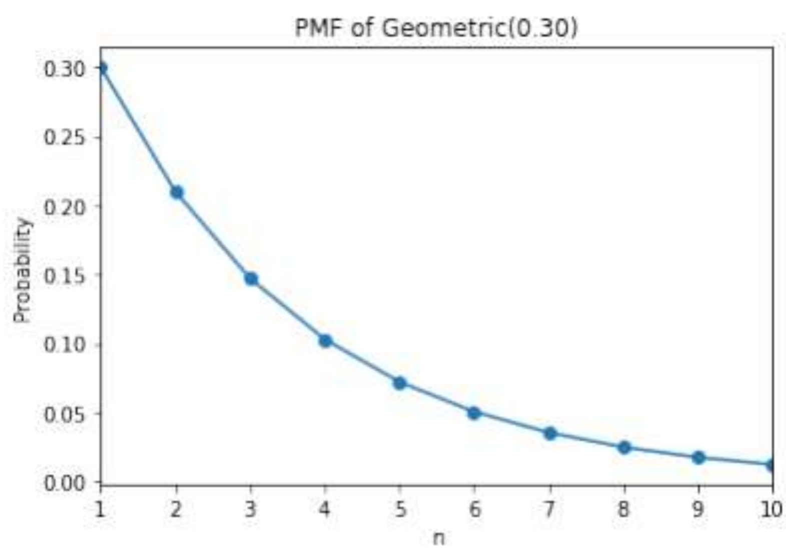
```
def pmf_geo(p, x):
    # your code
```

Using the pmf_geo function above to graph the relationship between the first hit after x tries and the corresponding probability. Given X = {1, 2, ..., 10}, the horizontal axis represents x tries, the vertical axis represents the probability p (p, x) respectively:

```
def plot_pmf_geo(p, n):
    '''
    Plot the probability mass function of Geometric
    '''
    # your code

plot_pmf_geo(0.3, 10)
```

The result is:



5 Exercises

1. One factory has 5 machines. The probability of each machine is broken in 1 session is 0.1.

(a) Use probability distribution functions to calculate the probability that 2 machines are broken in 1 session. (Answer: 0.073)

(b) Call X = {0, 1, 2, 3, 4, 5} is the event that in 1 session there are 0, 1, 2, 3, 4, and 5 broken machines respectively. Draw a graph representing the relationship between X and the corresponding probability.

2. A post office receives an average of 3 phone calls per minute.

(a) Use probability distribution functions to calculate the probability that the center receives 2 calls in 1 minute. (Answer: 0.224)

(b) Call X = {1, 2, 3, 4, 5} as an event in 1 minute there are 1, 2, 3, 4, 5 respectively calls to the post office. Draw a graph representing the relationship between X and the corresponding probability.

3. Each person is given 10 bullets and fired until 1 member hits the target. Knowing the probability of each bullet being hit is 40%.

(a) Use probability distribution functions to calculate the probability that a person hits the target in his third try. (Answer: 0.144)

(b) Call X = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} as the event where a person hits the target in his 1st, 2nd, 3rd, ..., 10^{th} try. Draw a graph representing the relationship between X and the corresponding probability.