Discrete Probability Distribution

Discrete Uniform Probability Distribution

We may have a situation where the probabilities of each event are the same. For example, if we roll a fair die, we assume that the probability of obtaining each number is $\frac{1}{6}$

If X is the random variable (r.v.) " the number showing" , the probability distribution table is

x	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The probability distribution function (p.d.f.) is

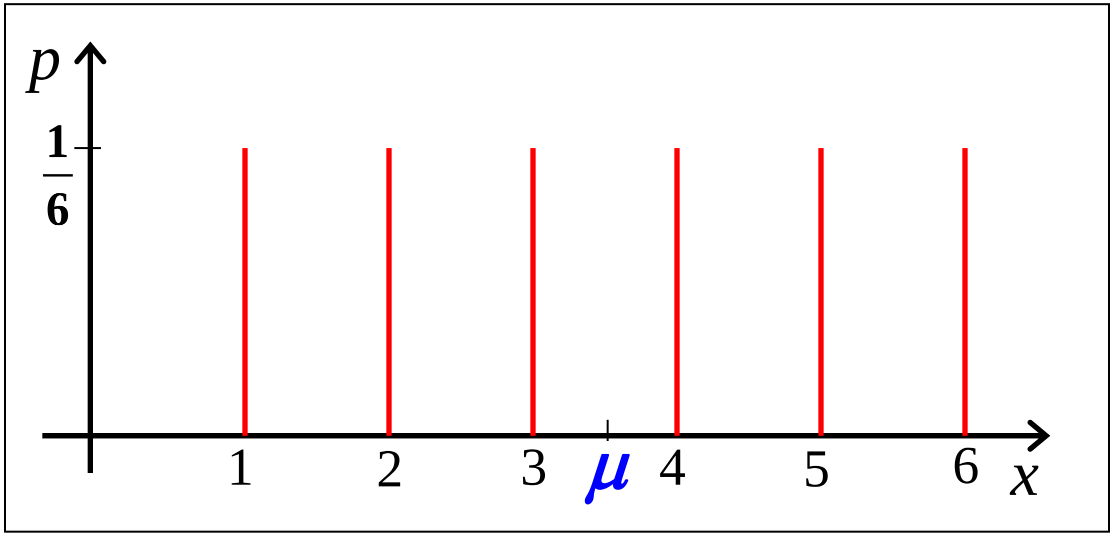
$$P(X=x)=\frac{1}{6}, \qquad x=1,\ 2,\ 3,\ 4,\ 5,\ 6$$

The distribution with equal probabilities is called "uniform"

A diagram for the distribution

$$P(X = x) = \frac{1}{6}, \qquad x = 1,\ 2,\ 3,\ 4,\ 5,\ 6$$

looks like this:



The mean value of X is given by the average of the 1^{st} and last values of x, so,

$$\mu = \frac{1+6}{2} = 3\cdot 5$$

However, we could also use the formula for the mean of <u>any</u> discrete distribution of a random variable:

$$\mu = \sum xf(x)$$

For $f(x) = P(X = x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6$

we would get

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \ldots + 6 \times \frac{1}{6}} \\ &= \mathbf{21 \times \frac{1}{6}} \\ &= \mathbf{3 \cdot 5}\end{aligned}$$

The Variance of the Uniform Distribution

We can find the variance for any discrete random variable X using

$$\mathrm{Var}(X) = \sigma^2 = \sum x^2 f(x) - \mu^2$$

e.g. The random variable X has p.d.f. given by

$$f(x) = P(X = x) = \frac{1}{6}, \qquad x = 1,\ 2,\ 3,\ 4,\ 5,\ 6$$

So, $$\mathbf{Var}(X) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \ldots + 6^2 \times \frac{1}{6} - \mu^2$$

We found earlier that $\mu = 3{\cdot}5$, so

$$\mathrm{Var}(X) = \frac{91}{6} - 3{\cdot}5^2 = 2{\cdot}92$$

The Bernoulli Process

Bernoulli process must possess the following properties:

1. The experiment consists of repeated trials.

2. Each trial results in an outcome that may be classified as a success or a failure.

3. The probability of success, denoted by p, remains constant from trial to trial.

4. The repeated trials are independent.

Binomial Distribution

The number X of successes in n Bernoulli trials is called a **binomial random variable**.

The probability distribution of this discrete random variable is called the binomial distribution and its values will be denoted by b(x; n, p) where n is number of trials and p is probability of success at each trial

A Bernoulli trial can result in a success with probability p and a failure with probability q = 1−p. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Examples

- A coin is flipped 10 times. What is the probability that exactly we will get 4 head?
- Flip a fair coin 10 times. X=4 heads

$$f(x)=\binom{10}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6=\binom{10}{4}\left(\frac{1}{2}\right)^{10}=210*.00097=0.205$$

- Die rolled for 3 times. What is the probability that at least once 6 will result?

$$f(x)=1-\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3=1-0.5787=0.4213$$

Examples

- Twelve pregnant women selected at random, take a home pregnancy test. This test give correct result with 0.8 probability. What is the probability that 10 women find a correct result?

$$f(x) = \binom{12}{10}(0.8)^{10}(0.2)^{2} = 66 * 0.107 * 0.04 = 0.283$$

- Random guessing on a multiple choice exam. 25 questions. 4 answers per question. A person get pass marks if he correctly guesses at least 15. What is the probability that a person who does not know correct answer of any question will get a pass marks?

$$f(x) = \binom{25}{15}(0.25)^{15}(0.75)^{10} + \ldots + \binom{25}{25}(0.25)^{25}(0.75)^{0}$$

- The binomial distribution derives its name from the fact that the $n + 1$ terms in the binomial expansion of $(q+p)^n$ correspond to the various values of $b(x; n, p)$ for $x = 0, 1, 2, \ldots, n$. That is

$$\begin{aligned}(q+p)^n &= \binom{n}{0}q^n + \binom{n}{1}pq^{n-1} + \binom{n}{2}p^2q^{n-2} + \cdots + \binom{n}{n}p^n \\ &= b(0; n, p) + b(1; n, p) + b(2; n, p) + \cdots + b(n; n, p).\end{aligned}$$

Since $p + q = 1$, we see that

$$\sum_{x=0}^{n} b(x; n, p) = 1,$$

Finding Mean and Variance

- Let the outcome of j^{th} trial be represented by indicator variable I_j which assumes the value 0 and 1 with probabilities q and p
- In binomial experiment number of success can be written as the sum of the n independent indicator variable
- $X = I_1 + I_2 + \ldots + I_n$
- Mean of $I_j = \boldsymbol{\Sigma [x \bullet P(x)] = 0.q + 1.p = p}$

- $\mu = E(X) = E(I_1)+E(I_2) +\ldots +E(I_n)=p+p+\ldots+p=np$
- $\sigma^2 I_j = E[(I_j - p)^2]=E(I_j^2) - p^2 = (0^2)q+(1^2)p- p^2 =p(1-p)=pq$
- For n independent variable variance would be $pq+pq+\ldots+pq=npq$

Python for Binomial Distribution

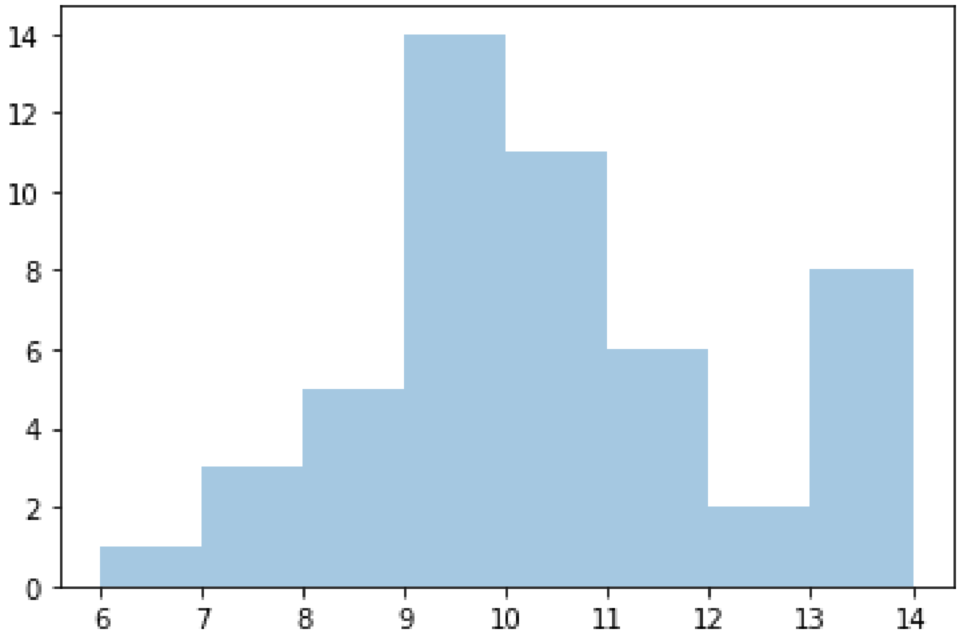
- You are tossing 20 times, probability of head is 0.5. Now you are repeating this experiment for 50 times. How many head appears each time?

```
from numpy import random, mean
x = random.binomial(n=20, p=0.5, size=50)
mean_head = mean(x)
print(x)
print (mean_head)
```

```
[10 11 10 9 8 10 9 9 7 12 12 8 10 12 9 11 12 8 11 14 10 11 8 9
8 11 14 10 11 9 6 10 11 9 11 10 13 15 10 13 9 10 7 10 6 13 6 7
16 9]
10.08
```

Visualization

```
from numpy import random
import matplotlib.pyplot as plt
import seaborn as sns
sns.distplot(random.binomial(n=20, p=0.5, size=50), kde=False)
plt.show()
```



Areas of Application

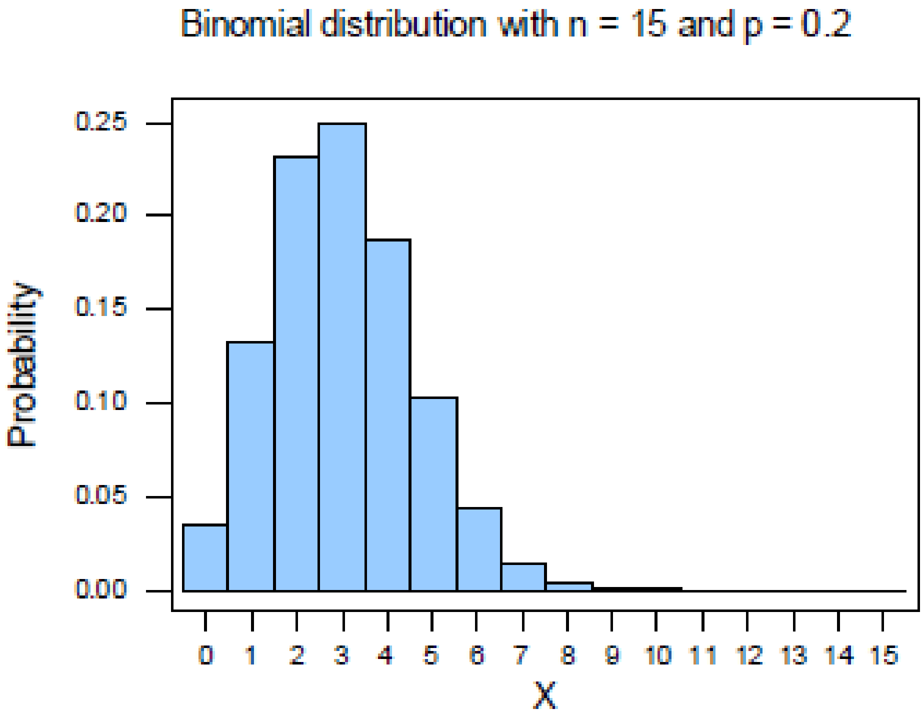
- Quality control measure in industrial process

- In epidemiology: Yes/no outcomes (dead/alive, treated/untreated, sick/well, etc.)

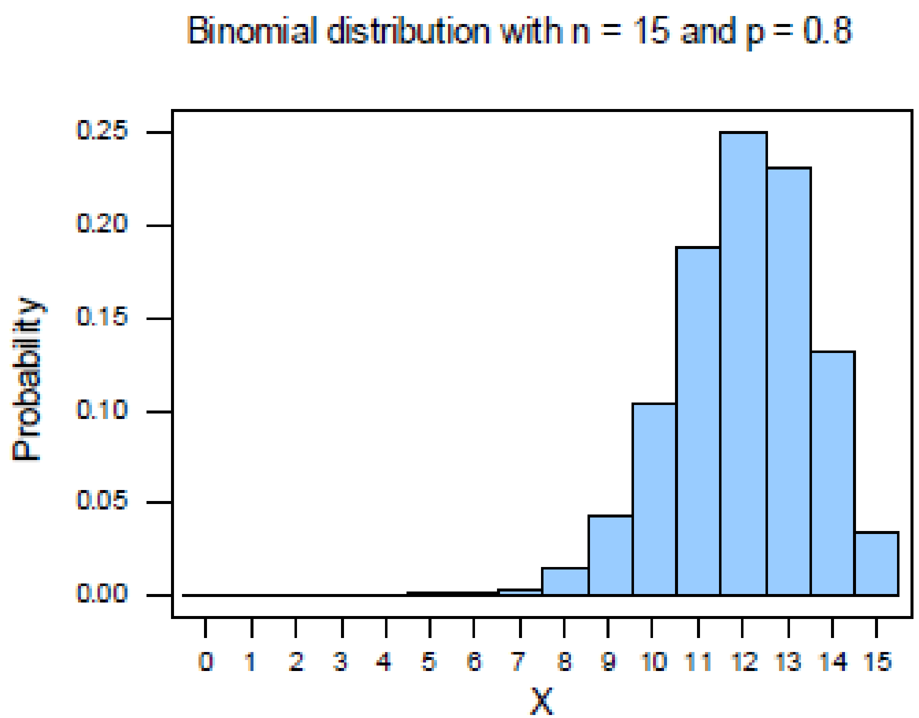
- In military application: like if a missile can successfully hit a location with probability p then what is the probability that missile hits the target k times out of n times.

Effect of n and p on shape

- For **small n** and **small p**, the binomial distribution is what we call **skewed right / positively skewed**. That is, the bulk of the probability falls in the smaller numbers 0, 1, 2,..., and the distribution tails off to the right. For example, here's a picture of the binomial distribution when n = 15 and p = 0.2:

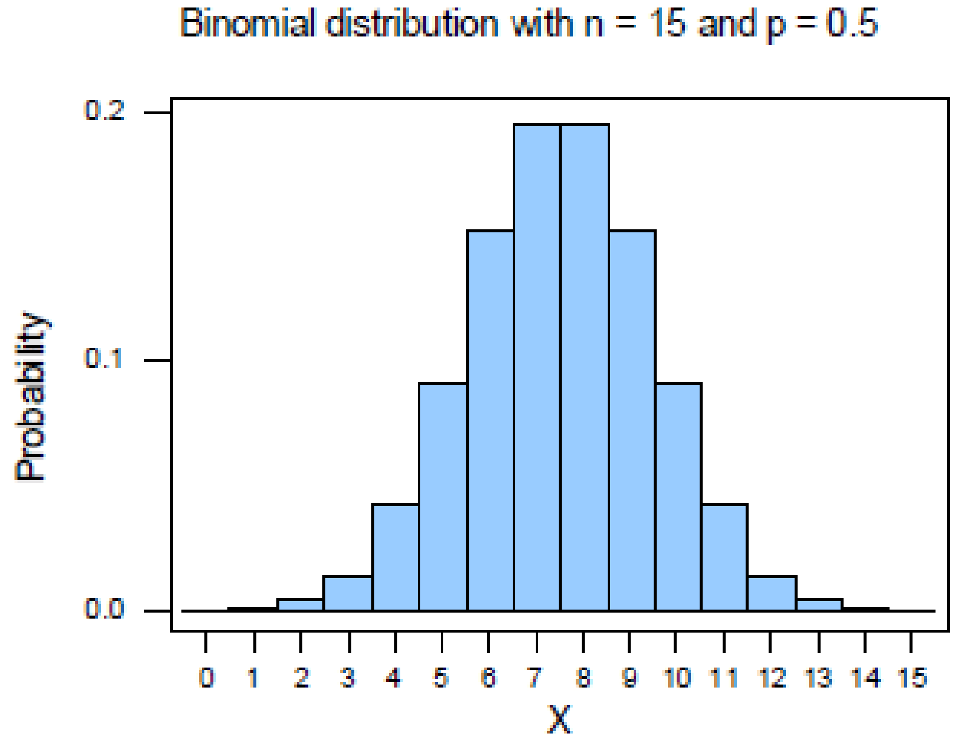


- For **small n** and **large p**, the binomial distribution is what we call **skewed left / negatively skewed**. That is, the bulk of the probability falls in the larger numbers n, $n-1$, $n-2$,... and the distribution tails off to the left. For example, here's a picture of the binomial distribution when $n = 15$ and $p = 0.8$:

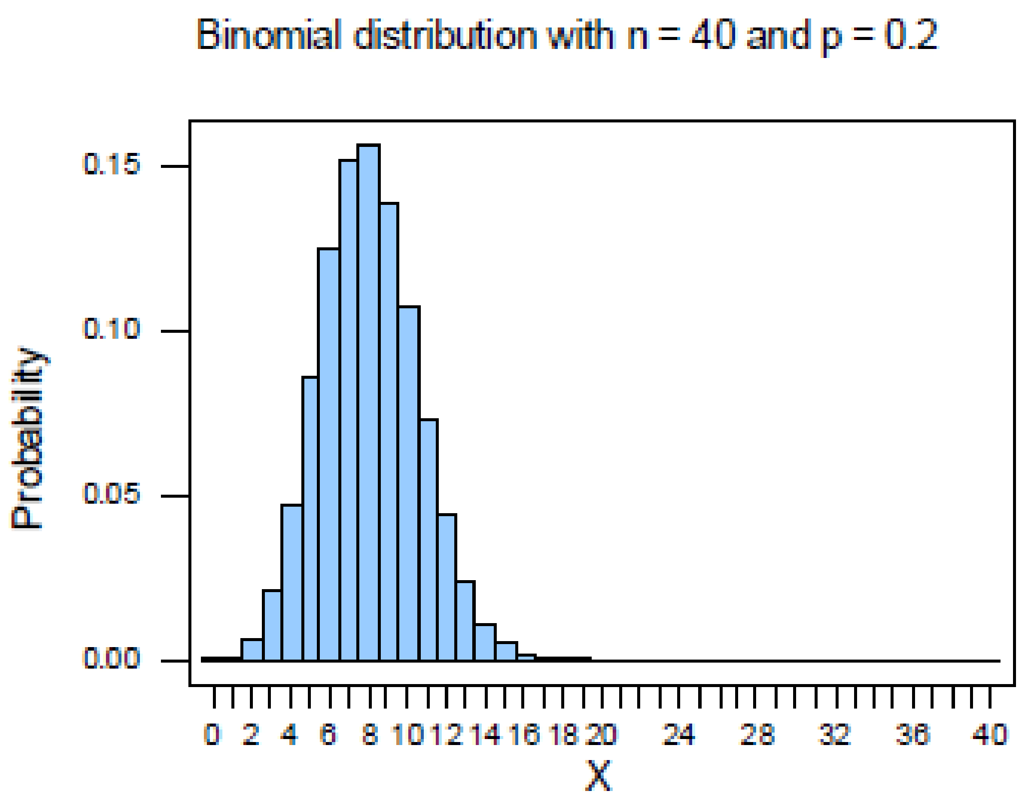


- For ***p* = 0.5** and **large and small *n***, the binomial distribution is what we call **symmetric**. That is, the distribution is without skewness. For example,

  here's a picture of the binomial distribution when $n = 15$ and $p = 0.5$:



- For **large *n and* small *p***, the binomial distribution **approaches symmetry**. For large *n*, the distribution is nearly symmetric. For example, here's a picture of the binomial distribution when $n = 40$ and $p = 0.2$:



Multinomial Distribution

- The binomial experiment becomes a **multinomial experiment** if we let each trial have more than two possible outcomes.
- In general, if a given trial can result in any one of k possible outcomes E_1, E_2, . . . , E_k with probabilities p_1, p_2, . . . , p_k, then the **multinomial distribution** will give the probability that E_1 occurs x_1 times, E_2 occurs x_2 times, . . ., and E_k occurs x_k times in n independent trials, where

$$x_1 + x_2 + \cdots + x_k = n.$$

- We shall denote this joint probability distribution by

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n).$$

- Clearly, $p_1 + p_2 + \cdots + p_k = 1$, since the result of each trial must be one of the k possible outcomes.

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

with

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1.$$

- For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving
- commercial jet:
- Runway 1: $p1 = 2/9$,
- Runway 2: $p2 = 1/6$,
- Runway 3: $p3 = 11/18$.
- What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?
- Runway 1: 2 airplanes,
- Runway 2: 1 airplane,
- Runway 3: 3 airplanes

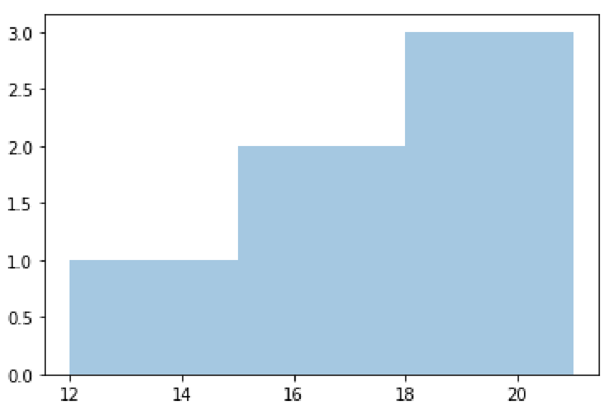
$$f\left(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6\right) = \binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$$
$$= \frac{6!}{2!\,1!\,3!}\cdot\frac{2^2}{9^2}\cdot\frac{1}{6}\cdot\frac{11^3}{18^3} = 0.1127.$$

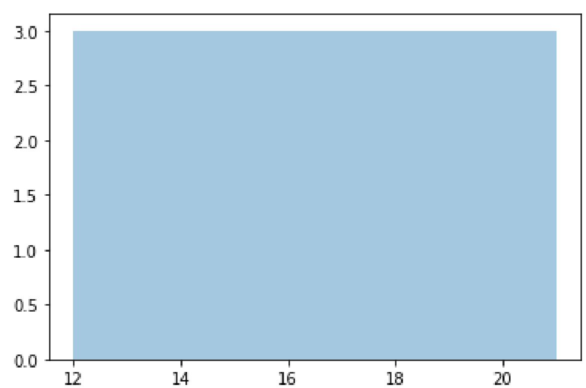
Distribution of event when probability of events are given

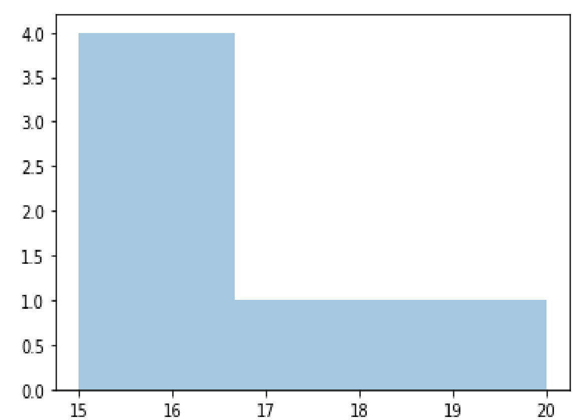
```
from numpy import random
x = random.multinomial(n=100, pvals=[1/6, 1/6, 1/6, 1/6, 1/6, 1/6])
print(x)
[ 9 13 16 19 19 24]
```

Visualization

```
from numpy import random
import matplotlib.pyplot as plt
import seaborn as sns
sns.distplot(random.multinomial(n=100, pvals=[1/6, 1/6, 1/6, 1/6, 1/6, 1/6]), kde=False)
plt.show()
```







Negative Binomial and Geometric Distributions

- Experiment is repeated until fixed number of success occurs.
- Instead of k success in n^{th} trial, we are now interested at finding number of trials needed for k success. Out of k success last success must occur at last trial.

- consider the use of a drug that is known to be effective in 60% of the cases. Find out the probability that we will have 5^{th} success at 7^{th} attempt?

- Probability of 5 success and 2 failures = $(0.6)^5*(0.4)^2$

- here are many possible arrangements of success and failure, however, last attempt must be a success. So from the first six attempts there must be 4 success and 2 failures.

- $P(X=7) = {}^6C_4* (0.6)^5*(0.4)^2=0.1866$

- What Is the Negative Binomial Random Variable?
- The number X is the number of trials required to produce k successes in a negative binomial experiment is called a **negative binomial random variable** and its probability distribution is called the **negative binomial distribution**

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \ldots.$$

- If repeated independent trials can result in a success with probability p and a failure with probability $q = 1 - p$, then the probability distribution of the random variable X, the number of the trial on which the kth success occurs, is

- In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams *A* and *B* face each other in the championship games and that team *A* has probability 0.55 of winning a game over team *B*.

- (a) What is the probability that team *A* will win the series in 6 games?

- (b) What is the probability that team *A* will win the series?

a) $b*(6; 4, 0.55) = {}^5C_3 * 0.55^4(1 - 0.55)^{6-4} = 0.1853$

b) P(team A wins the championship series) is

$b*(4; 4, 0.55) + b*(5; 4, 0.55) + b*(6; 4, 0.55) + b*(7; 4, 0.55)$

$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083.$

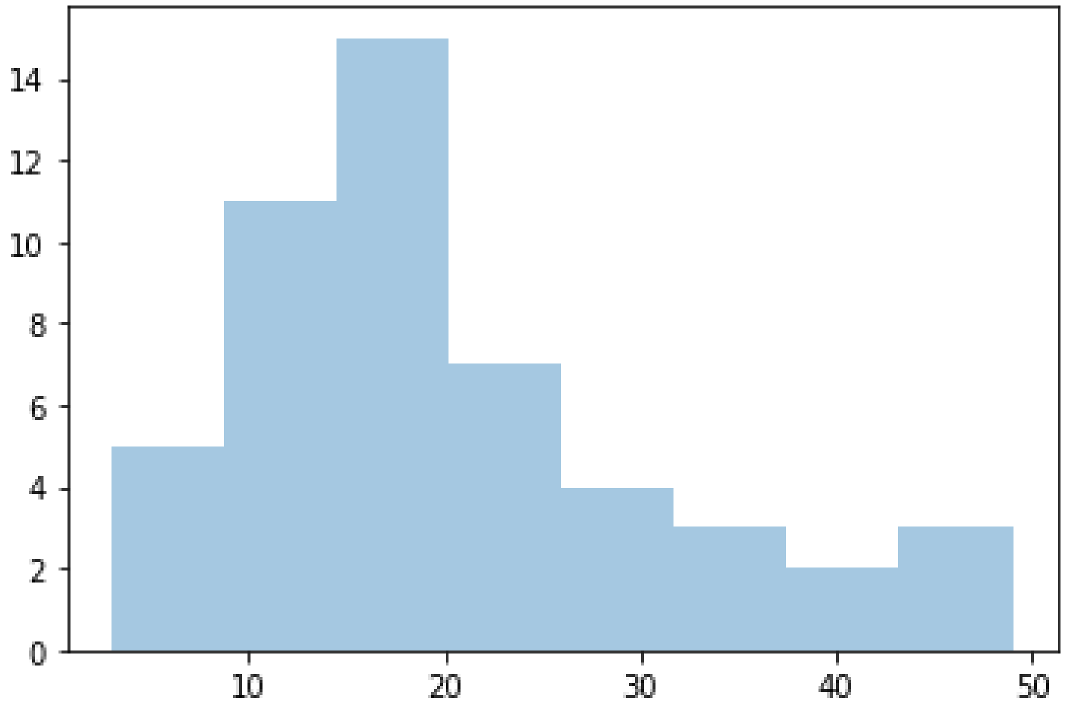
Number of failures before n^{th} success when success probability is p

```
from numpy import random, mean
x = random.negative_binomial(n=5, p=0.2, size=50)
mean_trial = mean(x)
print(x)
print (mean_trial)
```

- [12 17 11 45 18 39 5 13 10 19 33 50 27 12 27 13 14 3 27 20 20 15 21 40 13 50 9 9 49 31 22 8 14 22 8 17 33 7 27 15 5 13 18 27 24 36 6 22 10 16]
  20.44
- Average Number of experiments = 20.44+1=21.44

Visualization

- from numpy import random
- import matplotlib.pyplot as plt
- import seaborn as sns
- sns.distplot(random.negative_binomial(n=5, p=0.2, size=50), kde=False)
- plt.show()



Geometric Distribution

- If we consider the special case of the negative binomial distribution where $k = 1$, we have a probability distribution for the number of trials required before a single success.
- $b*(x; 1, p) = pq^{x-1}$, $x = 1, 2, 3, \ldots$.
- Since the successive terms constitute a geometric progression, it is customary to refer to this special case as the **geometric distribution** and denote its values by $g(x; p)$

- If repeated independent trials can result in a success with probability p and a failure with probability $q = 1 - p$, then the probability distribution of the random variable X, the number of the trial on which the first success occurs, is
- $g(x; p) = pq^{x-1}$, $x = 1, 2, 3, \ldots$.

- For a certain manufacturing process, it is known that, on an average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

- Using the geometric distribution with $x = 5$ and $p = 0.01$, we have
- $g(5; 0.01) = (0.01)(0.99)^4 = 0.0096.$

The mean and variance of a random variable following the geometric distribution are

$$\mu = \frac{1}{p} \text{ and } \sigma^2 = \frac{1-p}{p^2}.$$

Number of experiments required for first success when success prob is p

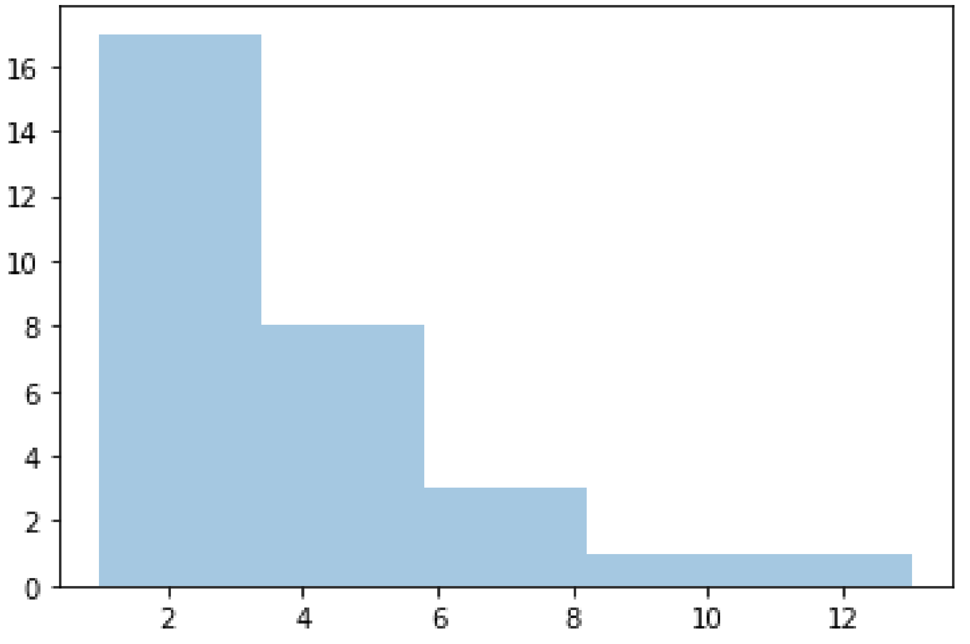
```
from numpy import random, mean
x = random.geometric(p=0.3, size=30)
mean_experiment = mean(x)
print(x)
print (mean_experiment)
[ 3 1 1 2 2 4 4 2 2 5 6 1 6 10 1 1 1 4 1 4 2 2 7 6 2 2 1 5 9 7]
3.466666666666667
```

- Draw one thousand values from the geometric distribution, with the probability of an individual success equal to 0.35. What is the probability of success after a single run?
- import numpy as np
- z = np.random.geometric(p=0.35, size=1000)
- sum_item = 0
- for item in z:
-   if item == 1:
-     sum_item +=1;
- prob = sum_item/(1000.0)
- print(prob)

0.354

Visualization of geometric distribution

- from numpy import random
- import matplotlib.pyplot as plt
- import seaborn as sns
- sns.distplot(random.geometric(p=0.3, size=30), kde=False)
- plt.show()



Properties of the Poisson Process

- The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. In this sense we say that the Poisson process has no memory.

- The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region.

- The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

- The number X of outcomes occurring during a Poisson experiment is called a **Poisson random variable**, and its probability distribution is called the **Poisson distribution**.

- The mean number of outcomes is computed from $\mu = \lambda$, Since the probabilities depend on λ, the rate of occurrence of outcomes, we shall denote them by $p(x; \lambda)$.
- The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in t time interval with mean rate λ

$$\boxed{p(x, \lambda t) = \frac{e^{(-\lambda t)}(\lambda t)^{x}}{x!}}$$

- where λ is the average number of outcomes for the unit time interval, distance, area, or volume.

- Arrivals of bus at a bus-stop follow a Poisson distribution with an average of 18 bus every hour. Calculate the probability of fewer than 3 arrivals in a quarter of an hour.

$$\boxed{p(x,\lambda t) = \frac{e^{(-\lambda t)}(\lambda t)^x}{x!}}$$

- λt=4.5, P(0,4.5)= 0.01111, p(1,4.5)=0.04999 and p(2,4.5)=0.11248

- So the probability of fewer than 3 arrivals in a quarter of an hour time is 0.01111+ 0.04999 + 0.11248 =0.17358

- During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

- $p(6; 4) = e^{-4}4^6/6! = 0.1042$

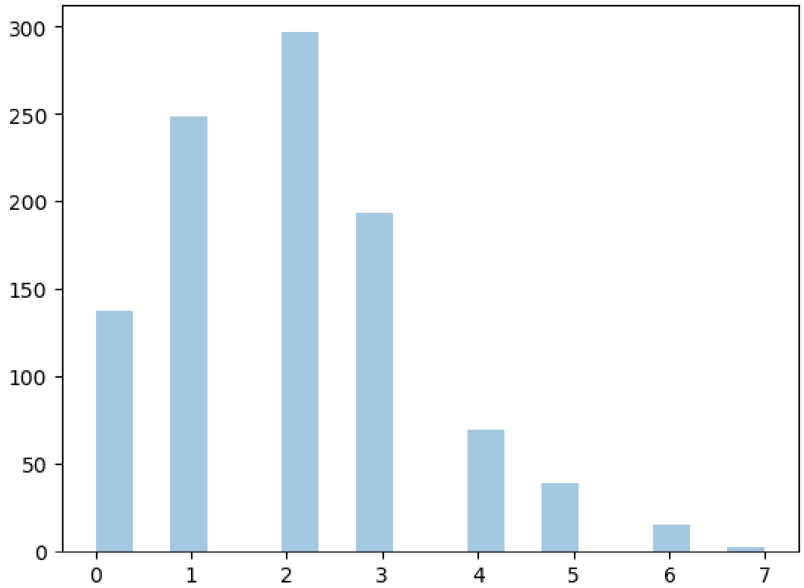
- Both the mean and the variance of the Poisson distribution $p(x; \lambda)$ are λ.

Average outcome due to Poisson process

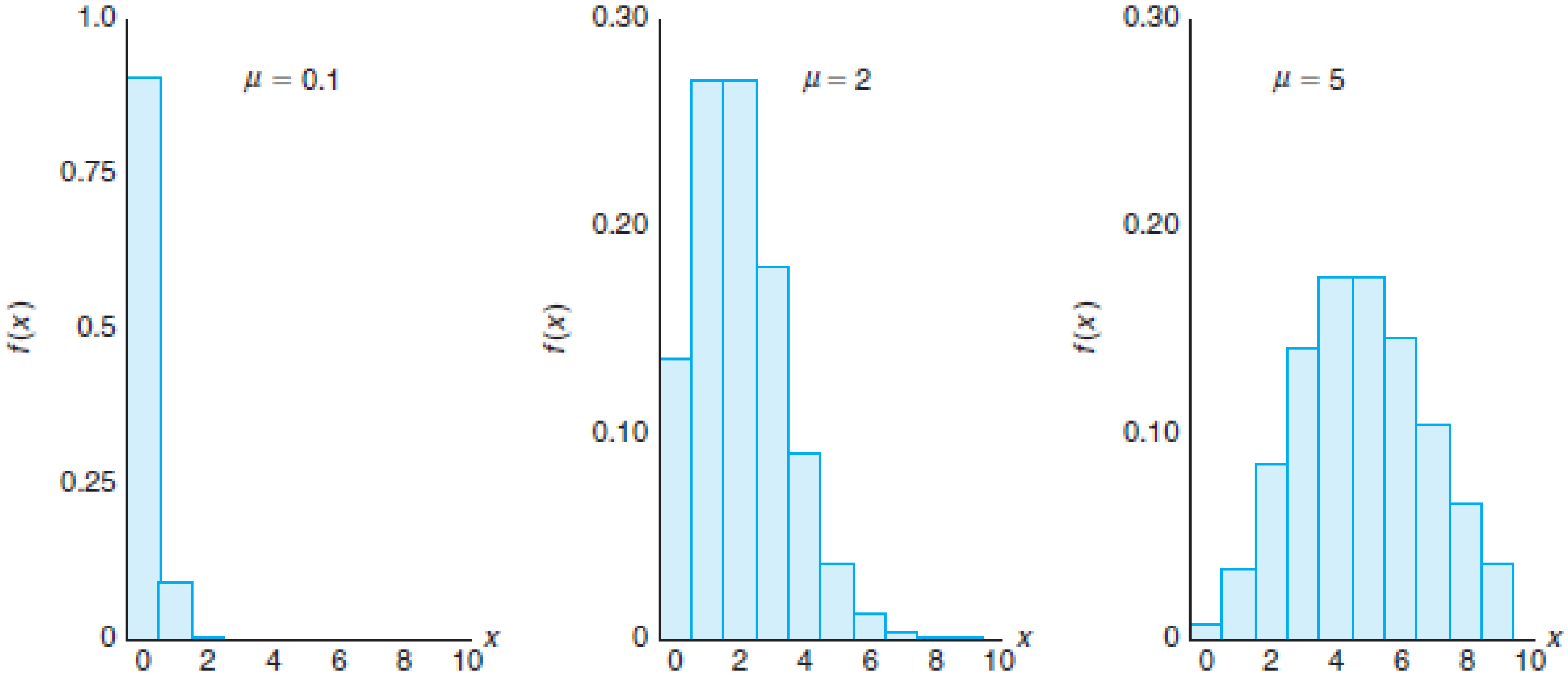
```
from numpy import random, mean
x = random.poisson(lam=2, size=100)
mean_experiment = mean(x)
print(x)
print (mean_experiment)
```

```
[3 1 1 1 2 6 1 1 1 2 1 2 3 2 2 3 1 2 3 3 2 1 0 2 3 4 2 3
5 5 1 2 1 0 4 1 3 2 3 3 0 1 3 0 5 4 1 2 2 0 3 1 0 2 2 2 2
3 2 2 3 5 1 1 2 2 0 1 3 2 4 4 1 3 2 3 1 5 1 3 2 1 5 0 1 1
1 0 2 2 3 1 3 1 1 2 1 1 1 1]
2.02
```

- from numpy import random
  import matplotlib.pyplot as plt
  import seaborn as sns
  sns.distplot(random.poisson(lam=2, size=1000), kde=False)
  plt.show()



Poisson Density Function for different Mean



Approximation of Binomial Distribution by a Poisson Distribution

- If n is large and p is close to 0, the Poisson distribution can be used, with $\mu = np$, to approximate binomial probabilities
- If p is close to 1, we can still use the Poisson distribution to approximate binomial probabilities by interchanging what we have defined to be a success and a failure, thereby changing p to a value close to 0.
- Binomial situation, n= 100, p=0.075. Calculate the probability of fewer than 10 successes.
- Using binomial distribution value is 0.7832687
- The Poisson approximation to the Binomial states that λ will be equal to np, i.e. 100 x 0.075 so λ=7.5
- Using Poisson distribution value is 0.7764076

- In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

a) What is the probability that in any given period of 400 days there will be an accident on one day?

b) What is the probability that there are at most three days with an accident?

- Let X be a binomial random variable with n = 400 and p = 0.005. Thus, np = 2. Using the Poisson approximation,

a) $P(1,2) = e^{-2}2^{1} = 0.271$ and

b) $P(X \leq 3,2) = \sum_{x=0}^{3} e^{-2}2^{x}/x! = 0.857.$

Area of Application

- Number of deaths due to some epidemic
- Number of faulty item produced in a year
- Number of days it is going to rain

Exercise

- The probability that a student pilot passes the written test for a private pilot's license is 0.7. Find the probability that a given student will pass the test
- (a) on the third try;
- (b) before the fourth try.
- Using the geometric distribution

a) $P(X = 3) = g(3; 0.7) = (0.7)(0.3)^2 = 0.0630.$

b) $P(X < 4) = \sum_{x=1}^{3} g(x, 0.7) = \sum_{x=1}^{3} (0.7)(0.3)^{x-1} = 0.973$

Exercise

- On an average, 3 traffic accidents per month occur at a certain intersection. What is the probability that in any given month at this intersection
- exactly 5 accidents will occur?
- fewer than 3 accidents will occur?
- Using the Poisson distribution with $x = 5$ and $\lambda t = 3$,

  $p(5; 3) = e^{-3} * 3^5/5! = 0.1008$
- $P(X < 3) = P(X \leq 2) = p(0,3)+p(1,3)+p(2,3)=0.4232.$