# MAT 352 Assignment — 3

# Computer Science Department

**Question:**

For each of the distributions below:

- `Discrete Distributions`
  1. Bernoulli
  2. Binomial
  3. Poisson
  4. Geometric
- `Continuous Distributions`
  1. Bernoulli
  2. Binomial
  3. Poisson
  4. Geometric

Give:

- The key characteristics
- The PDF/PMF, Expected Value and Variance
- Sample World Problems and
- The graph of the distribution

Submitted to Dr. Adinya

April 5, 2023

## Name of Students

| S/N | Name | Matric Number |
|---|---|---|
| 1 | Adebowale Joseph Akintomiwa | 214846 |
| 2 | Adedapo Anjorin | 214864 |
| 3 | Adegbola Olatunde Williams | 207186 |
| 4 | Adeleke Sherifdeen Adeboye | 214848 |
| 5 | Adeleke Timothy Toluwani | 214849 |
| 6 | Adelowo Samuel Damilare | 214850 |
| 7 | Adeoti Warith Adetayo | 214851 |
| 8 | Adim Chimaobi Solomon | 222455 |
| 9 | Adisa Inioluwa Christiana | 214853 |
| 10 | Ahmad Animasaun | 214863 |
| 11 | Ajayi Prince Ayokunle | 215221 |
| 12 | Akinade Faith Eniola | 222459 |
| 13 | Akinrinola Akinfolarin | 205526 |
| 14 | Akinrinola, Blessing Opemipo | 214857 |
| 15 | Akinwusi Ifeoluwa | 214858 |
| 16 | Alao Tawakalit Omowunmi | 222461 |
| 17 | Alatise Oluwaseun Abraham | 214860 |
| 18 | Arowolo Ayomide Stephen | 214865 |
| 19 | Brai Daniel | 214868 |
| 20 | Chinedu Promise Okafor | 213930 |
| 21 | Daniel Emmanuel Oghenetega | 224870 |
| 22 | Denedo Oghenetega | 214873 |
| 23 | Emiade James | 214874 |

| | | |
|---|---|---|
| 24 | Farayola Joshua Olatunde | 214878 |
| 25 | Godwin Daniel | 214871 |
| 26 | Ibraheem Nuh Babatunde | 214879 |
| 27 | Ikwuegbu Michael | 214881 |
| 28 | Kareem Mustapha Babatunde | 214883 |
| 29 | Kayode Peter Temitope | 208077 |
| 30 | Kehinde Boluwatife Soyoye | 214916 |
| 31 | Kip Charles Okechukwu Emeka | 215061 |
| 32 | Matric | Number |
| 33 | Kubiat Laura | 214884 |
| 34 | Lawal Uchechukwu Adebayo | 214885 |
| 35 | Matthews Victoria Olayide | 214886 |
| 36 | Nwatu Chidinma Augustina | 214890 |
| 37 | Odulate Oluwatobi Gabriel | 214893 |
| 38 | Ogbolu Precious Chiamaka | 214894 |
| 39 | Oghie Daniel O. | 214895 |
| 40 | Ogunesan Rhoda Oluwatosin | 214897 |
| 41 | Ogunyemi Temidayo Samuel | 214898 |
| 42 | Ojewale Opeoluwa David | 214899 |
| 43 | Okafor Lisa Chisom | 214901 |
| 44 | Okoro Joshua Akachukwu | 214902 |
| 45 | Okumagba Oghenerukevwe Miracle | 222498 |
| 46 | Olagidi Joshua | 222500 |
| 47 | Olalere Khadijat Titilayo | 222502 |
| 48 | Olatunji Michael Oluwayemi | 214903 |
| 49 | Olawale Eniola Emmanuel | 214904 |

| | | |
|---|---|---|
| 50 | Olorogun Ebikabowei Caleb | 214906 |
| 51 | Oluwatade Iyanuoluwa | 214907 |
| 52 | Oluwayelu Oluwanifise | 215257 |
| 53 | Onasoga Oluwapelumi Idris | 214909 |
| 54 | Oyekanmi Eniola | 214913 |
| 55 | Sadiq Peter | 214914 |
| 56 | Salami Lateefat Abimbola | 214915 |
| 57 | Stephen Chidiebere Ivuelekwa | 214882 |
| 58 | Toluwanimi Oluwabukunmi Osuolale | 214912 |
| 59 | Ubaka Amazing-Grace Onyiyechukwu | 214918 |
| 60 | Uchechukwu Ahunanya | 214854 |
| 61 | Wisdom Oyor | 215206 |

# 1 Discrete Distributions

## 1.1 Bernoulli Distribution

The Bernoulli distribution is a probability distribution wherein the random variable can only have two possible outcomes: 1 (probability of success) with probability p or 0 (probability of failure) with probability $(1-p)$. It is a special case of the binomial distribution where a single trial is conducted (so n would be 1 for such a binomial distribution). It is represented as $X \sim Ber(p)$. This distribution's probability mass function, mean, and variance is given below.

$$f_X(x) = p^x(1-p)^{1-x} \tag{1}$$

$$E[X] = p \tag{2}$$

$$Var[X] = p(1-p) \tag{3}$$

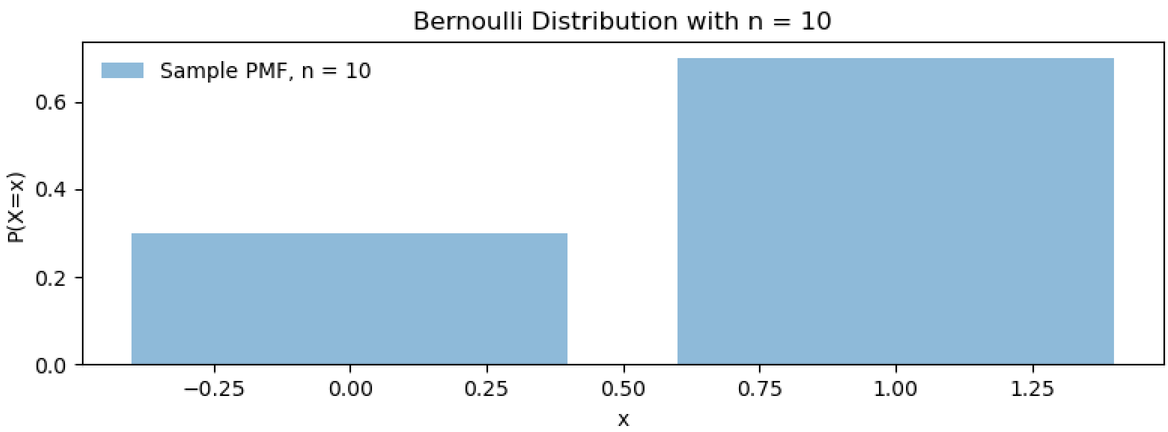
$$Std[X] = \sqrt{p(1-p)} \tag{4}$$

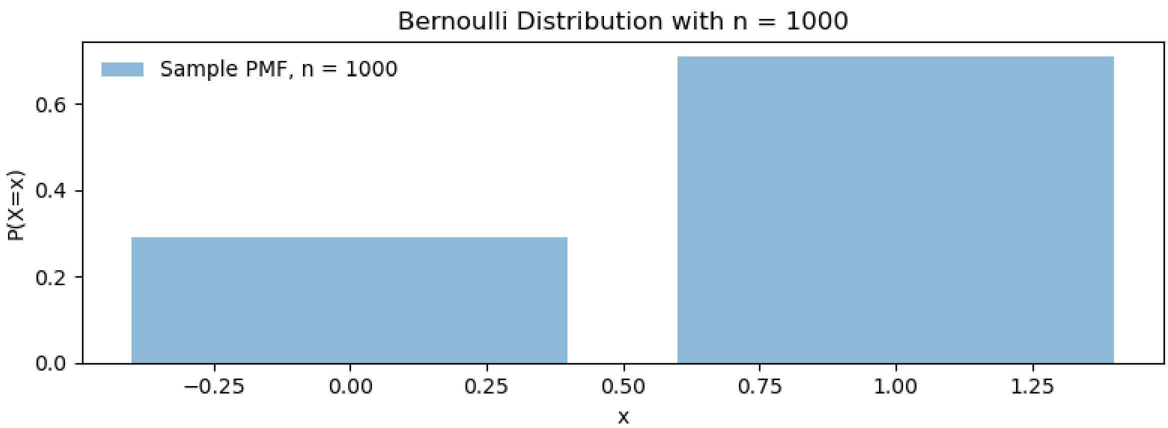
**Sample Word Problems**

1. The prevalence of a certain disease in the general population is 10%. If we randomly select a person from this population, find the probability that the person is not diseased.

2. The standard deviation of a Bernoulli random variable $X$ is $\frac{2}{5}$. Find the Expected value and the variance of $X$.

**Graph of Bernoulli Distribution**



## 1.2 Binomial Distribution

The binomial distribution is a probability distribution that models the chances of success in a series of events where the only possible outcomes are success and failure. A binomial distribution gives the probability of gaining $x$ successes out of $n$ trials if the probability of success is $p$. It is represented as $X \sim Bin(n,p)$. The probability mass function, mean, variance and standard deviation of this distribution are given below.

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{5}$$

$$E[X] = np \tag{6}$$

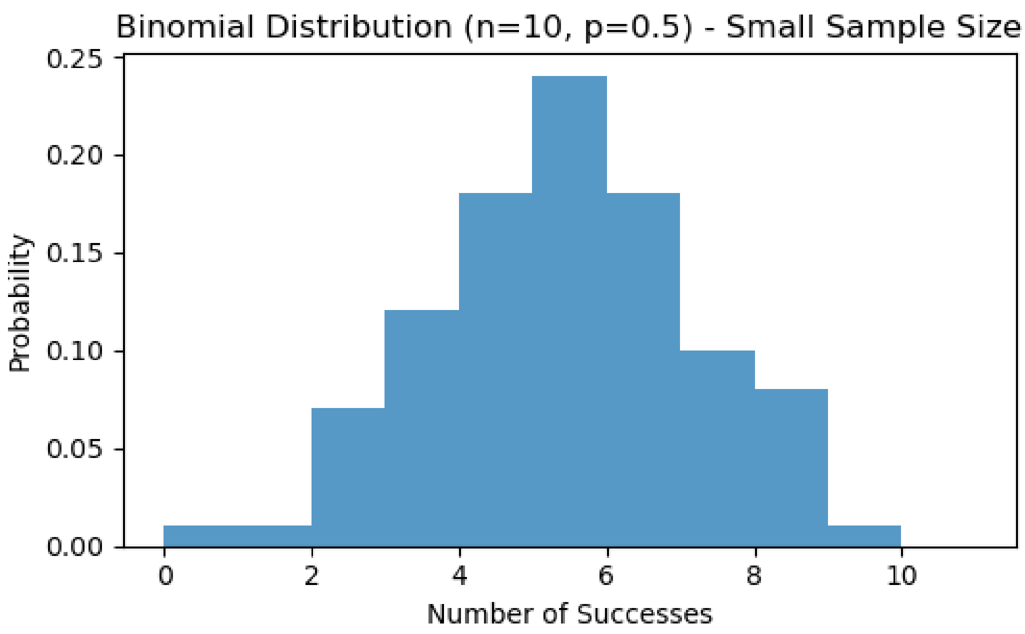
$$Var[X] = np(1-p) \tag{7}$$

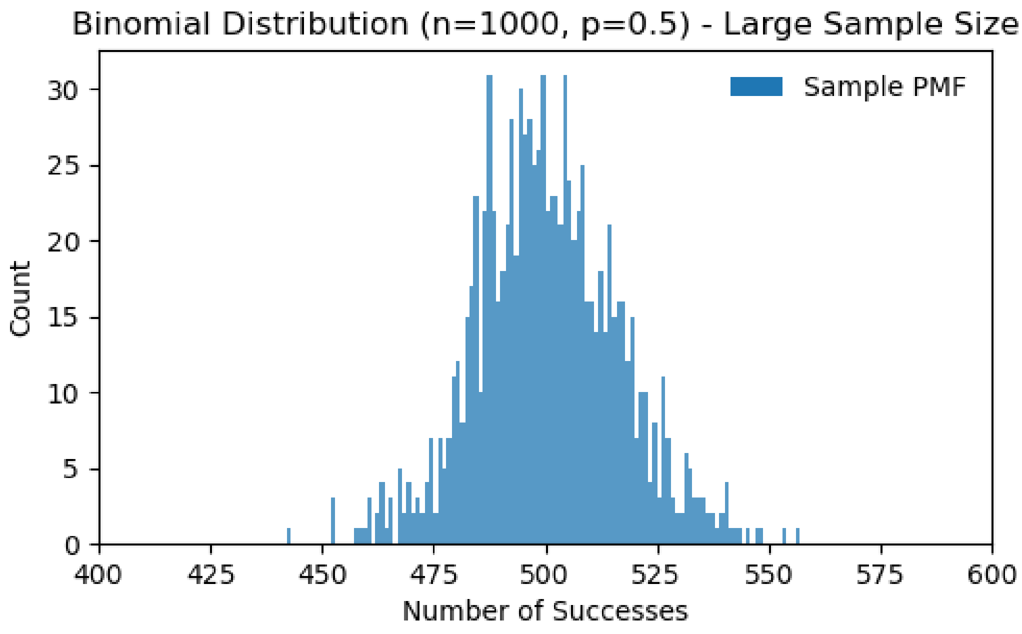
$$Std[X] = \sqrt{np(1-p)} \tag{8}$$

**Sample Word Problems**

1. There are four fused bulbs in a lot of 10 good bulbs. If three bulbs are drawn at random with replacement, find the probability of distribution of the number of fused bulbs drawn.
2. Find the probability of obtaining four or more heads in five tosses of a fair coin.
3. The probability that a motorcycle will change lanes when making a U-turn is 80%. Suppose a random sample of 16 motorcycle are observed making turns at Fordham roadss and jerome avenue intersection. Find the probability that at least one motorcycle will change lanes while making U-turn.

**Graph of Binomial Distribution**




## 1.3 Poisson Distribution

This probability distribution models the chances of a certain number of events, $x$, occurring within a time or space frame, given the average rate of occurrence of the event, $\lambda$. It is represented as $X \sim Poisson(\lambda)$.

$$f_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{for } x = 1, 2, \ldots \\ 0 & \text{elsewhere} \end{cases} \tag{9}$$

$$E[X] = \lambda \tag{10}$$

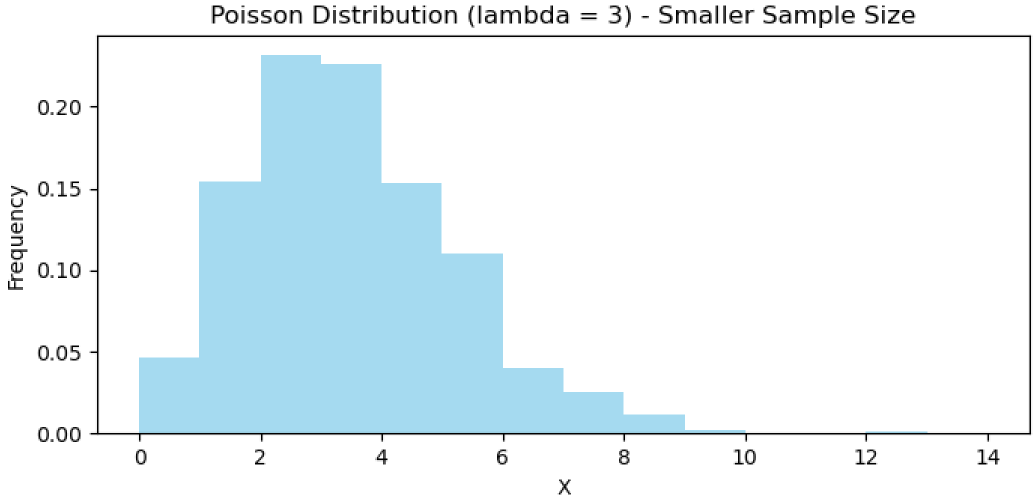
$$Var[X] = \lambda \tag{11}$$

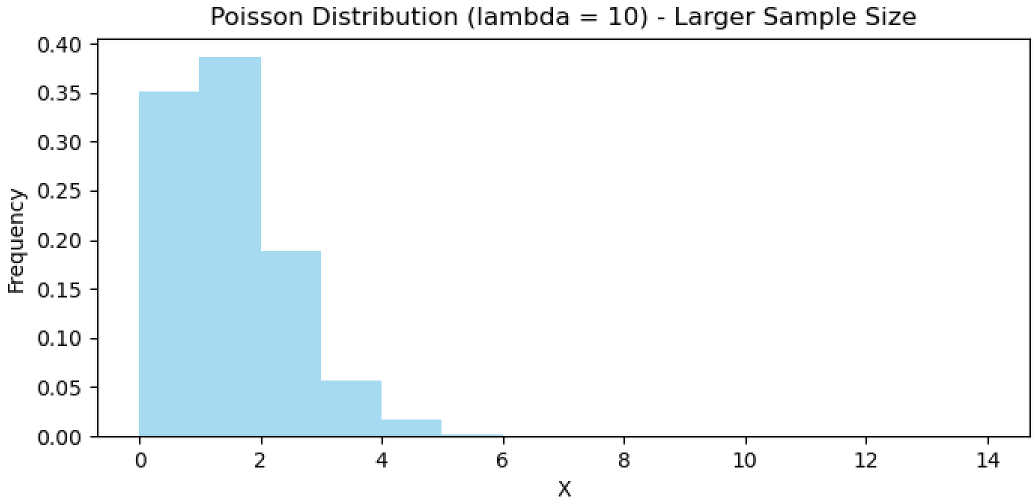
$$Std[X] = \sqrt{\lambda} \tag{12}$$

**Sample Word Problems**

1. 3 in every 1000 H-mobile phones are discovered to have fault. Find the probability that out of 5000 H-mobile phones, exactly 8 will have fault.
2. A manufacturer produces light-bulbs that are packed into boxes of 100. If quality control studies indicate that 0.5% of the light-bulbs produced are defective, what percentage of the boxes will contain:
   (a) no defective?
   (b) 2 or more defectives?
3. If 3% of electronic units manufactured by a company are defective. Find the probability that in a sample of 200 units, less than 2 bulbs are defective.

**Graph of Poisson Distribution**




## 1.4 Geometric Distribution

A geometric distribution is a probability distribution of a random variable $X$ that satisfies some conditions. The geometric distribution conditions are: A phenomenon that has a series of $n$ trials, Each trial has only two possible outcomes – either success $p$ or failure $q = 1 - p$, The probability of success is the same for each trial. This distribution gives the probability of achieving success after N number of failures. It is represented as $X \sim Geometric(p)$. This distribution's probability mass function, mean, and variance is given below.

$$f_X(x) = (1-p)^{x-1}p \tag{13}$$

$$E[X] = 1/p \tag{14}$$

$$Var[X] = (1-p)/p^2 \tag{15}$$

$$Std[X] = \sqrt{(1-p)/p^2} \tag{16}$$

**Sample Word Problems**

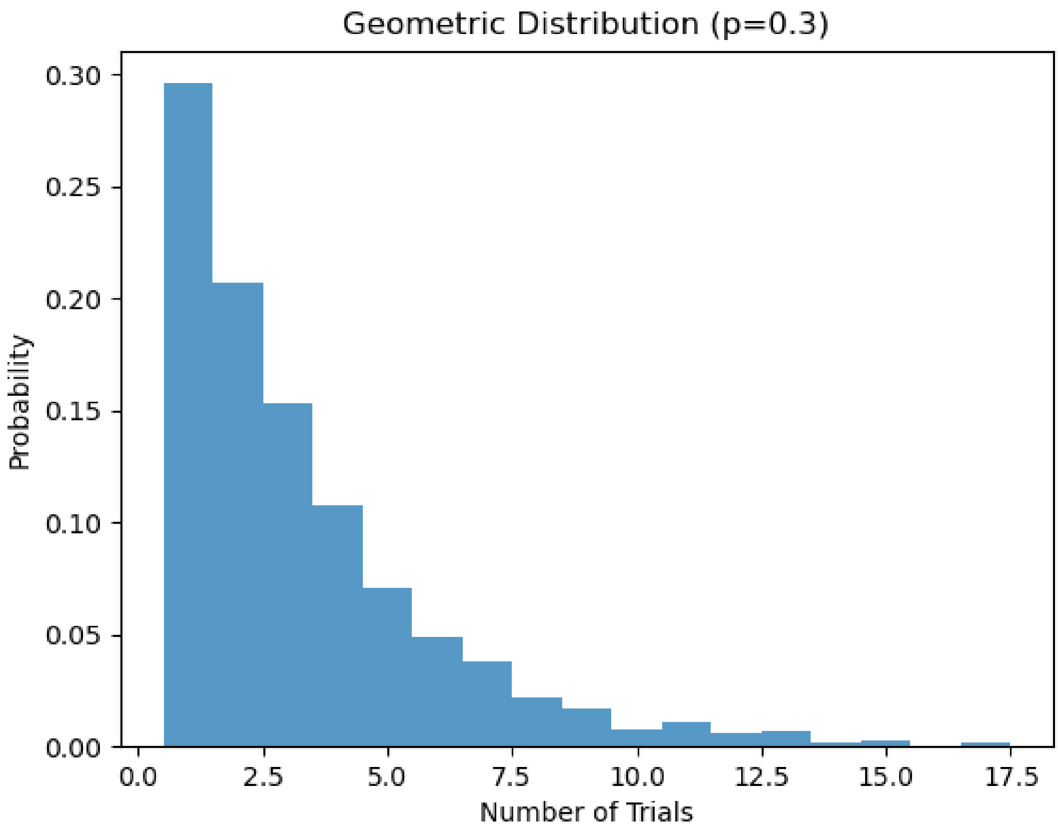
1. If a patient is waiting for a suitable blood donor and the probability that the selected donor will be a match is 0.2, then find the expected number of donors who will be tested till a match is found including the matched donor.

2. Suppose you are playing a game of darts. The probability of success is 0.4. What is the probability that you will hit the bullseye on the third try?

3. Calculate the probability density of geometric distribution if the value of $p$ is 0.42; $x = 1, 2, 3, \ldots$, also find out the mean and variance.

4. A light bulb manufacturing factory finds 3 in every 60 light bulbs defective. Calculate what will be the probability that the first defective light bulb with be found when the $6th$ one is tested?

**Graph of Geometric Distribution**




## 2 Continuous Distributions

### 2.1 Uniform Distribution

The uniform distribution is a type of probability distribution in which all outcomes are equally likely. It is defined by two parameters, $a$ and $b$, where x = minimum value and y = maximum value. It is generally denoted by $u(a, b)$. A continuous random variable $X$ is said to have a Uniform distribution over the interval $[a, b]$, shown as $X \sim Uniform(a, b)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ or } x > b \end{cases} \tag{17}$$

$$E[X] = \frac{a+b}{2} \tag{18}$$

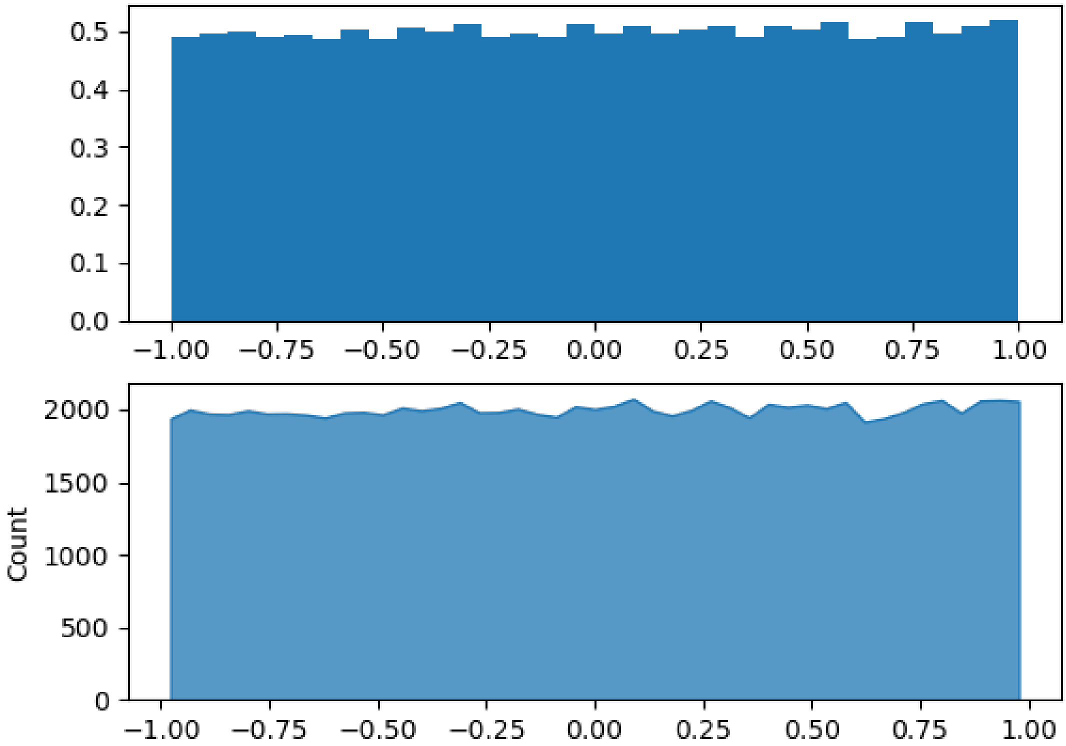
$$Var[X] = \frac{(b-a)^2}{12} \tag{19}$$

$$Std[X] = \sqrt{\frac{(b-a)^2}{12}} \tag{20}$$

**Sample Word Problems**

1. Bus is uniformly late between 2 and 10$minutes$.
   (a) How long can you expect to wait?
   (b) with what standard deviation?
   (c) If it is $> 7mins$, you'll be late for work. What is the probability of being late.
2. The average weight gained by a person over the winter months is uniformly distributed and ranges from 0 to 30$lbs$. Find the probability of a person that he will gain between 10 and 15$lbs$ in the winter months.

**Graph of Uniform Distribution**




## 2.2 Normal Distribution

The normal distribution is a probability distribution that models many natural phenomena such as the distribution of a particular feature in a population. It is represented as $X \sim N(\mu, \sigma)$. It is easily recognised by its bell-shaped curve that is centered around the mean and a spread determined by the standard deviation. Its probability distribution function, mean, variance and standard deviation are given below.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \tag{21}$$

$$E[X] = \mu \tag{22}$$

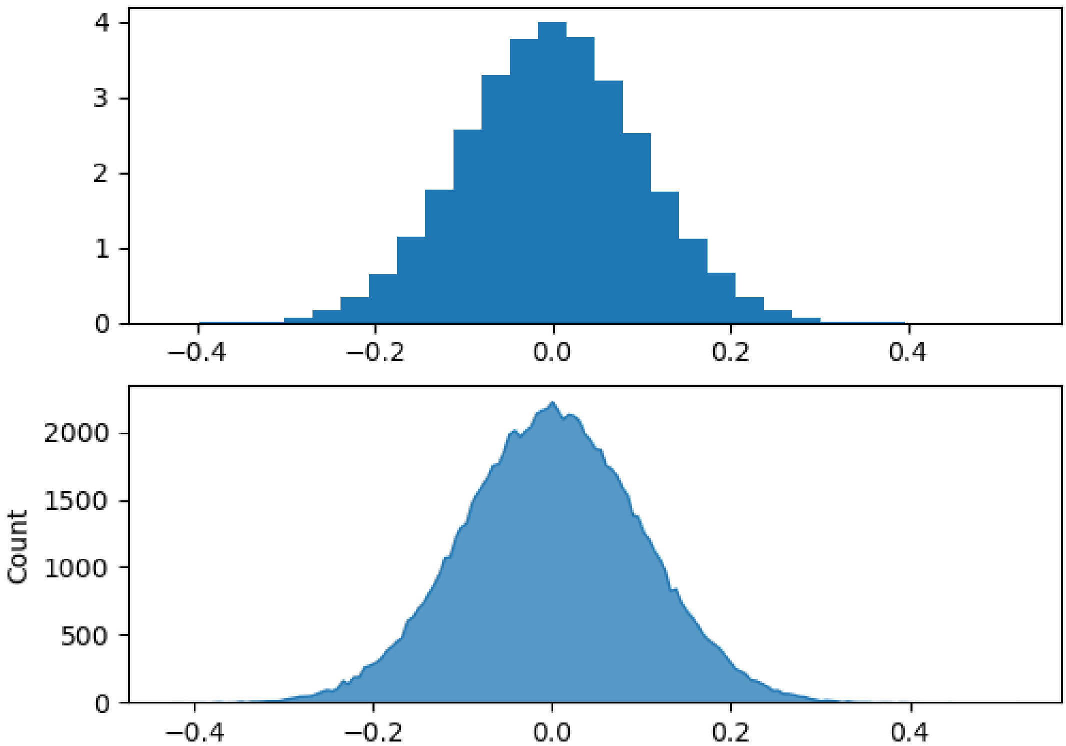
$$Var[X] = \sigma^2 \tag{23}$$

$$Std[X] = \sigma \tag{24}$$

**Sample Word Problems**

1. $X$ is a normally distributed variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find:
   (a) $P(x < 40)$
   (b) $P(x > 21)$
   (c) $P(30 < x < 35)$
2. Act score are normally distributed with mean of 24.2 and standard deviaton 42. What is the probability that a student score greater than 31?.
3. Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585 Will he be admitted to this university?

**Graph of Normal Distribution**



## 2.3 Exponential

The exponential distribution is a continuous probability distribution that concerns the amount of time until some specific event happens. It is a process in which events happen continuously and independently at a constant average rate.It is represented as $X \sim Exp(\lambda)$. This distribution's probability mass function, mean, and variance is given below.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda e} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \tag{25}$$

$$E[X] = \frac{1}{\lambda} \tag{26}$$

$$Var[X] = \frac{1}{\lambda^2} \tag{27}$$

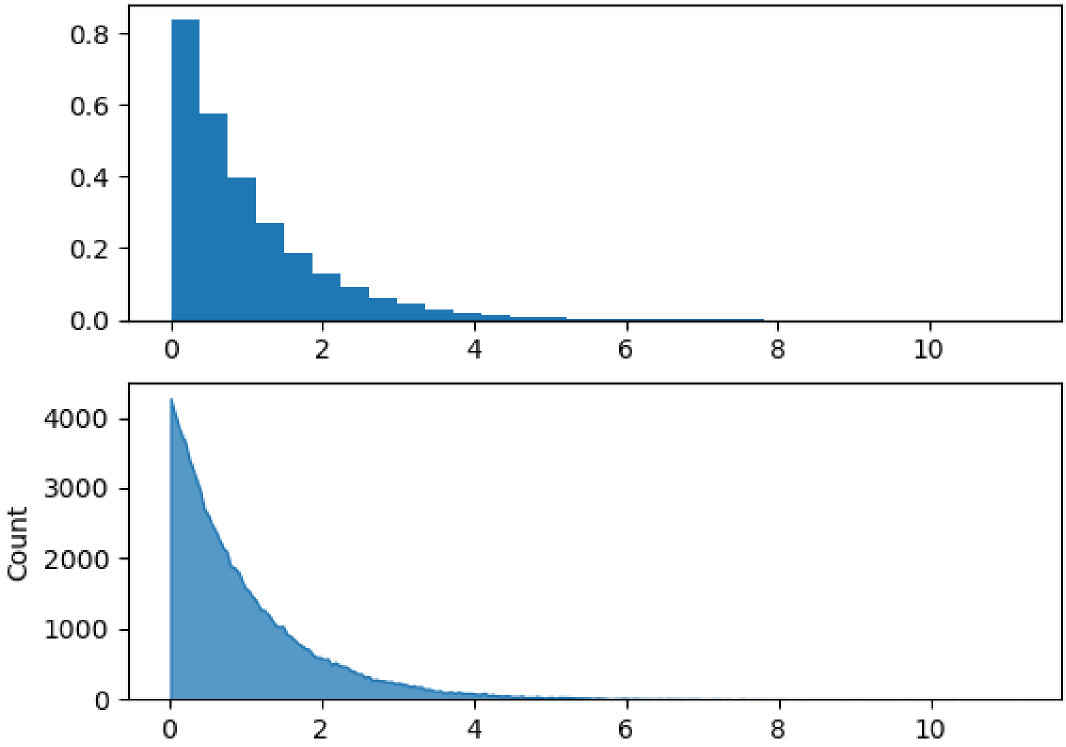
$$Std[X] = \sqrt{\frac{1}{\lambda^2}} \tag{28}$$

**Sample Word Problems**

1. On the average, a certain computer part lasts ten years. The length of time the computer part lasts is exponentially distributed. What is the probability that a computer part lasts more than 7 years?

2. Suppose that the length of a phone call, in minutes, is an exponential random variable with decay parameter 112. The decay parameter is another way to view $1/\lambda$. If another person arrives at a public telephone just before you, find the probability that you will have to wait more than five minutes. Let $X$ = the length of a phone call, in minutes. Calculate $\mu$, and $\sigma$.

**Graph of Exponential Distribution**




## 2.4 Gamma Distribution

The gamma distribution, like the normal distribution, models natural phenomena that are positively skewed such as the waiting time between events. Unlike the normal distribution, its graph is skewed to the right. Gamma distributions can be represented as $X \sim Gamma(\alpha, \beta)$.

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1} e^{\frac{-1}{\beta}x}}{\beta^{\alpha}\Gamma_{\alpha}} & ;, x > 0; (\alpha, \beta > 0) \\ 0 & \text{elsewhere} \end{cases} \tag{29}$$

$$E[X] = \alpha(\alpha+1)\beta^2 \tag{30}$$

$$Var[X] = \alpha\beta^2 \tag{31}$$

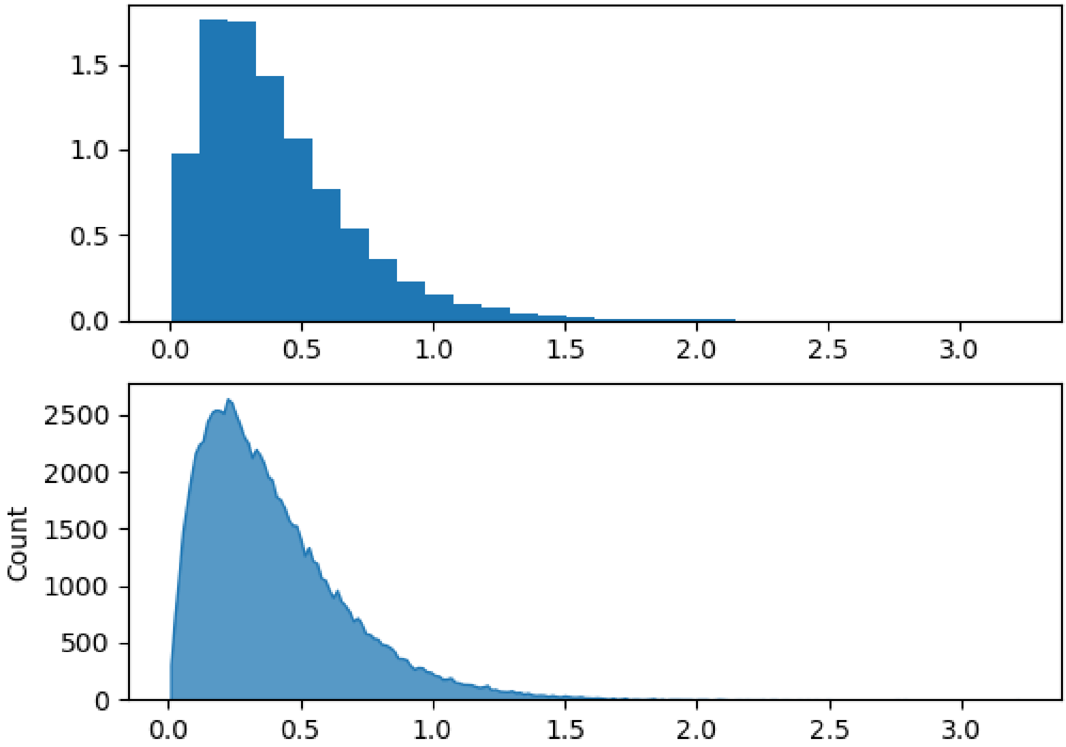
$$Std[X] = \sqrt{\alpha}\beta \tag{32}$$

**Sample Word Problems**

1. In a certain city, the daily consumption of water (in millions of liters) follows approximately a gamma distribution with $\alpha = 2$ and $\beta = 3$. If the daily capacity of that city is 9 million liters of water, what is the probability that on any given day the water supply is inadequate?

2. Suppose $X$ has gamma distribution with parameters $\alpha = 8$ and $\beta = 15$. Compute $P(60 \leq X \leq 120)$.

**Graph of Gamma Distribution**




# 3 Python Source Codes used to generate the graphs

## discrete_distribution.py

```python
import numpy as np
from scipy.stats import bernoulli
import matplotlib.pyplot as plt
import seaborn as sns

np.random.seed(0)
# sns.set_style("ticks", {'axes.grid' : False})

plt.rcParams['patch.linewidth'] = 0


def plot_bernoulli_graph():
    # Define the probability of success
    p = 0.7

    # Generate an array of values for x
    x = np.array([0, 1])

    # Define the sample sizes
    n_small = 10
    n_large = 1000

    # Generate the data for the small sample size
    data_small = bernoulli.rvs(p, size=n_small)
    hist_small, bins_small = np.histogram(data_small, bins=[-0.5, 0.5, 1.5], density=True)

```

```python
    # Generate the data for the large sample size
    data_large = bernoulli.rvs(p, size=n_large)
    hist_large, bins_large = np.histogram(data_large, bins=[-0.5, 0.5, 1.5], density=True)

    # Create a grid of subplots
    fig, axs = plt.subplots(2, 1, figsize=(8, 6))

    # Plot the small sample size data on the top subplot
    # axs[0].bar(x, pmf_small, label=f'Theoretical PMF, p = {p}')
    axs[0].bar(x, hist_small, alpha=0.5, label=f'Sample PMF, n = {n_small}')
    axs[0].set_xlabel('x')
    axs[0].set_ylabel('P(X=x)')
    axs[0].set_title('Bernoulli Distribution with n = {}'.format(n_small))
    axs[0].legend()

    # Plot the large sample size data on the bottom subplot
    # axs[1].bar(x, pmf_large, label=f'Theoretical PMF, p = {p}')
    axs[1].bar(x, hist_large, alpha=0.5, label=f'Sample PMF, n = {n_large}')
    axs[1].set_xlabel('x')
    axs[1].set_ylabel('P(X=x)')
    axs[1].set_title('Bernoulli Distribution with n = {}'.format(n_large))
    axs[1].legend()

    # Show the plot
    plt.tight_layout()
    plt.savefig(fname=plot_name("Bernoulli"))


def plot_binomial_graph():
    # Parameters for the binomial distribution
    n = 10
    p = 0.5

    n_large = 1000
    # Generate random samples for small sample size
    data_small = np.random.binomial(n, p, size=100)

    # Generate random samples for large sample size
    data_large = np.random.binomial(n_large, p, size=1000)

    # Create a figure with two subplots
    fig, axs = plt.subplots(2, 1, figsize=(6, 8))

    # Plot the binomial distribution with small sample size using Seaborn histplot
    sns.histplot(data_small, bins=np.arange(0, n + 2), stat='probability', kde=False, ax=axs[0])
    axs[0].set_title(f'Binomial Distribution (n={n}, p={p}) - Small Sample Size')
    axs[0].set_xlabel('Number of Successes')
    axs[0].set_ylabel('Probability')

    # Plot the binomial distribution with large sample size using Seaborn histplot
    sns.histplot(data_large, bins=np.arange(0, n_large + 2), kde=False, ax=axs[1])
    axs[1].set_title(f'Binomial Distribution (n={n_large}, p={p}) - Large Sample Size')
    axs[1].set_xlabel('Number of Successes')
    axs[1].set_ylabel('Count')

    # Zooming in

    axs[1].set_xlim(400, 600)

    # Show the legend
    axs[1].legend(labels=['Sample PMF'])

    # Adjust spacing between subplots
    plt.subplots_adjust(hspace=0.4)

    # Save the plot
    plt.savefig(fname=plot_name("Binomial"))


def plot_geometric_graph():
    np.random.seed(1)
```

```python

    # Generate random data for the geometric distribution
    p = 0.3  # probability of success
    sample_size = 1000  # sample size
    data = np.random.geometric(p, sample_size)

    # Create the histogram using seaborn
    sns.histplot(data, bins=np.arange(1, np.max(data) + 2) - 0.5, stat='probability',
    kde=False)
    plt.xlabel('Number of Trials')
    plt.ylabel('Probability')
    plt.title('Geometric Distribution (p=0.3)')
    # plt.savefig(fname='Geometric_distribution_visualization')
    plt.savefig(fname=plot_name("Geometric"))


def plot_poisson_graph():
    # Testing between the different lambda values
    lambda_small = 3
    lambda_large = 10

    # Generate data for smaller sample size
    data_small = np.random.poisson(lambda_small, size=1000)

    # Generate data for larger sample size
    data_large = np.random.poisson(1, size=1000)

    # Set up the figure with subplots
    fig, axs = plt.subplots(2, 1, figsize=(8, 8))

    # Plot histogram for smaller sample size
    sns.histplot(data_small, kde=False, color='skyblue', bins=range(0, 15), stat='
    probability', ax=axs[0])
    axs[0].set_title(f'Poisson Distribution (lambda = {lambda_small}) - Smaller Sample
    Size')
    axs[0].set_xlabel('X')
    axs[0].set_ylabel('Frequency')

    # Plot histogram for larger sample size
    sns.histplot(data_large, kde=False, color='skyblue', bins=range(0, 15), stat='
    probability', ax=axs[1])
    axs[1].set_title(f'Poisson Distribution (lambda = {lambda_large}) - Larger Sample
    Size')
    axs[1].set_xlabel('X')
    axs[1].set_ylabel('Frequency')

    plt.subplots_adjust(
        hspace=0.33)
    # Save the plot
    plt.savefig(fname=plot_name("Poisson"))


def plot_name(dist_name):
    SAVE_DIR = "Assignment_3_Res/Graphs"

    return f"{SAVE_DIR}/(D)-{dist_name.title()}_distribution_visualization"



if __name__ == '__main__':
    plot_bernoulli_graph()
    plot_binomial_graph()
    plot_geometric_graph()
    plot_poisson_graph()
```

### continuous_distribution.py

```python
"""
    This script was written by computer science student 300 level.
    As an assignment submission for MAT353
"""

import numpy as np
```

```python
import scipy as sp
import seaborn as sb
import matplotlib.pyplot as plt


def visualize_uniform_distribution():
    """This function visualizes a random variable of a uniform distrbution."""
    _, plot_axis = plt.subplots(2, 1)

    # uniformly select 90000 random samples between -1 and 1
    samples = np.random.uniform(-1, 1, 90000)

    # plot histogram of samples distribution
    _, bins, _ = plot_axis[0].hist(samples, bins=30, density=True)

    # draws line plot
    sb.histplot(samples, element='poly', fill=True, ax=plot_axis[1])

    # save visualisation
    plt.savefig(fname=plot_name("Uniform"))
    plt.close()


def visualize_normal_distribution():
    """This function visualizes a random variable of a normal distrbution."""

    _, plot_axis = plt.subplots(2, 1)

    # select 90000 random samples using the nomal distibution given a mean=0 and standard_deviation=0.1
    mean, standard_deviation = 0, 0.1
    samples = np.random.normal(loc=0, scale=0.1, size=90000)

    # plot histogram to show sample distribution
    _, bins, _ = plot_axis[0].hist(samples, 30, density=True)

    # draws line plot
    sb.histplot(samples, element='poly', fill=True, ax=plot_axis[1])

    # save visualisation
    plt.savefig(fname=plot_name("Normal"))
    plt.close()


def visualize_exponential_distibution():
    """This function visualizes a random variable of a exponential distrbution."""
    _, plot_axis = plt.subplots(2, 1)

    # select 90000 random samples using the exponential distribution
    samples = np.random.exponential(1, size=90000)

    # plot histogram to show sample distribution
    _, bins, _ = plot_axis[0].hist(samples, 30, density=True)

    # draws line plot
    sb.histplot(samples, element='poly', fill=True, ax=plot_axis[1])

    plt.savefig(fname=plot_name("Exponential"))
    plt.close()


def visualize_gamma_distribution():
    """This function visualizes a random variable of a gamma distrbution."""
    _, plot_axis = plt.subplots(2, 1)

    # select 90000 random samples using the nomal distibution given a mean=2 and standard_deviation=0.2
    mean, standard_deviation = 2, 0.2
    samples = np.random.gamma(mean, standard_deviation, size=90000)

    # plot histogram to show sample distribution
    _, bins, _ = plot_axis[0].hist(samples, 30, density=True)

```

```python
    # # draws line to highlight approximate distribution shape
    # scale_y = bins**(mean-1)*(np.exp(-bins/standard_deviation)/(sp.special.gamma(mean)
    *standard_deviation**mean))
    # plt.plot(bins, scale_y, linewidth=2, color='g')

    # draws line plot
    sb.histplot(samples, element='poly', fill=True, ax=plot_axis[1])

    plt.savefig(fname=plot_name("Gamma"))
    plt.close()


def plot_name(dist_name):
    SAVE_DIR = "Assignment_3_Res/Graphs"

    return f"{SAVE_DIR}/(C)-{dist_name.title()}_distribution_visualization"


if __name__ == '__main__':
    visualize_uniform_distribution()
    visualize_normal_distribution()
    visualize_exponential_distibution()
    visualize_gamma_distribution()
```