EP4130 - Data Science Analysis

Assignment - 1

ABHISHEK GUGULOTH
MS20BTECH11008

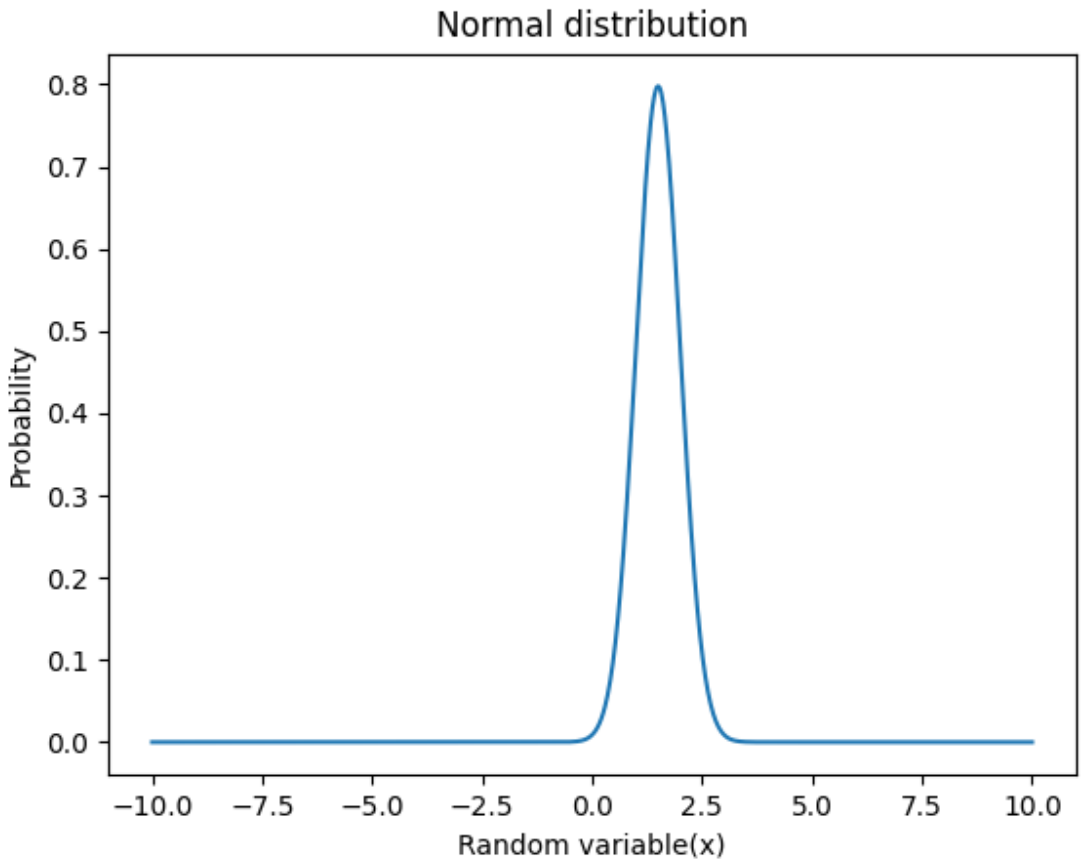
Dr. Shantanu Desai
Department of EP

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Question 1:

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats


# Part 1
x=np.linspace(-10, 10, 1000)

mean1 = 1.5
std_dev = 0.5
dist = stats.norm(mean1, std_dev)

plt.plot(x, dist.pdf(x))
y=dist.pdf(x)

plt.title('Normal distribution')
plt.xlabel('Random variable(x)')
plt.ylabel('Probability')
plt.show()


# Part 2

# Mean calculation
mean2 = np.mean(y)
print(mean2)

# Median calculation
median = np.median(y)
print(median)

# Skewness calculation
skew=stats.skew(y)
print(skew)

# Kurtosis calculation
kurto=stats.kurtosis(y)
```

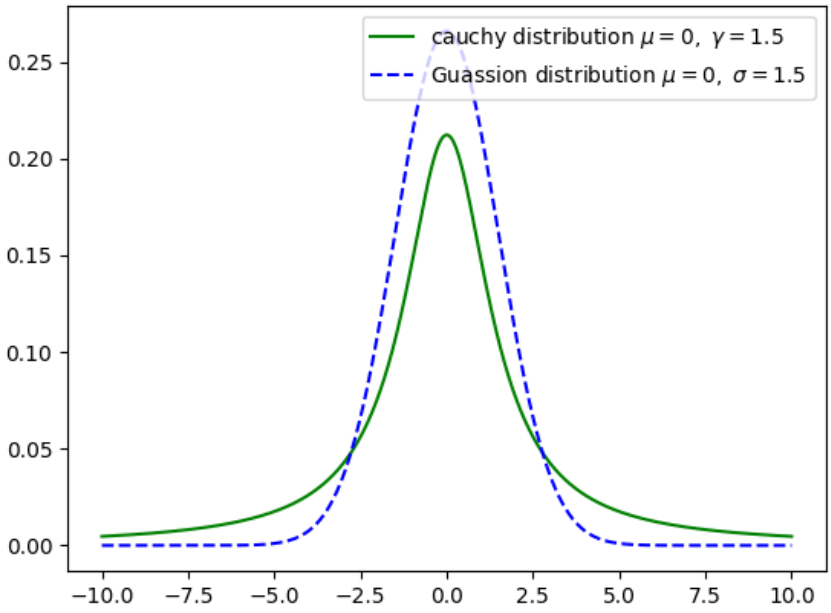
```
print(kurto)

# standard deviation
mad = stats.median_abs_deviation(y)
new_std=1.482*mad
print(new_std)

# sigma G calculation
q_25, q_75 = np.percentile(y,[25,75])
sigma = 0.7413 * (q_75 - q_25)
print(sigma)
```



Mean = 0.04995000000000003
Median = 1.4123062779902294e-22
Skewness = 3.494416489724312
Kurtosis = 11.27898248243111
standard deviation = 2.09303790398152e-22
Sigma G = 2.0655537777122252e-06

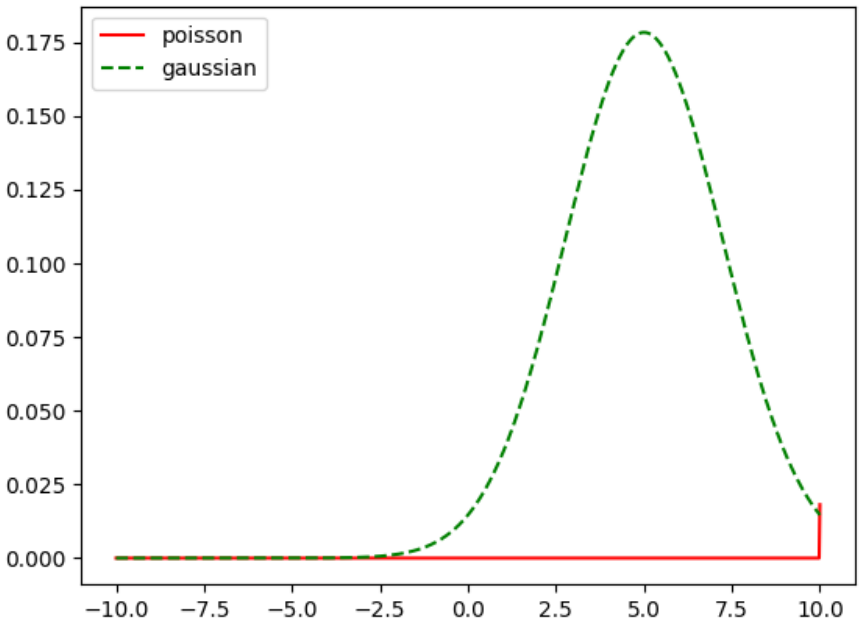
Question 2:

```
import numpy as np
from scipy import stats
from matplotlib import pyplot as plt


mua = 0
gamma = 1.5
sigma = 1.5

x = np.linspace(-10, 10, 1000)

cauchy_dist = stats.cauchy(mua, gamma)
gaus_dist = stats.norm(mua, sigma)

plt.plot(x, cauchy_dist.pdf(x), ls='-', color = 'green')
plt.plot(x, gaus_dist.pdf(x), ls='--', color = 'blue')
plt.legend(["cauchy distribution $\mu=0,\ \gamma=1.5$", "Guassion
distribution $\mu=0,\ \sigma=1.5$"], loc ="upper right")
plt.show()
```



Question 3:

```
import math
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt


std_dev = 5**0.5
mean = 5

poi_dist = stats.poisson(mean)
gaus_dist = stats.norm(mean,std_dev)

x = np.linspace(-10,10,1000)
plt.plot(x, poi_dist.pmf(x), ls='-', color='red', label='poisson')
plt.plot(x, gaus_dist.pdf(x), ls='--', color='green', label='gaussian')
plt.legend()
plt.show()
```



Question 4:

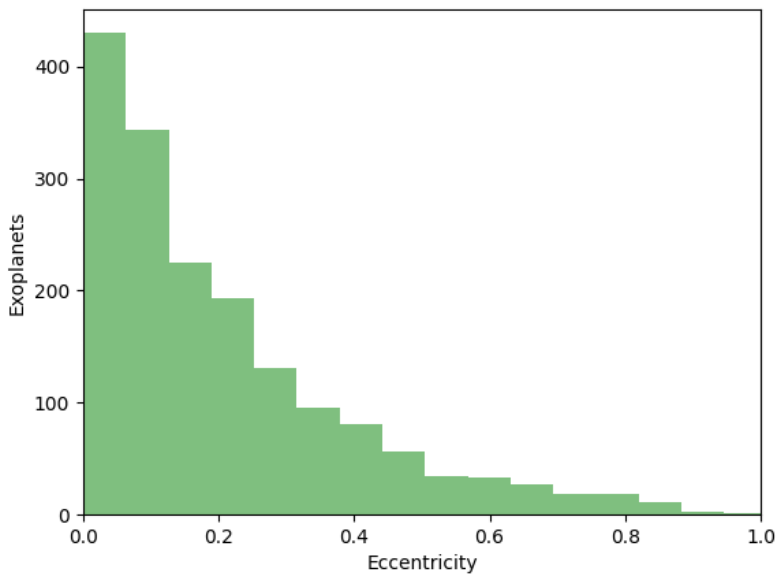
```
import numpy as np


lifetime = [0.8920, 0.881, 0.8913, 0.9837, 0.8958]
error = [0.00044, 0.009, 0.00032, 0.00048, 0.00045]

sum1 = 0
for i in range(5):
    sum1 += (lifetime[i] / ((error[i])**2))

sum2 = 0
for i in range(5):
    sum2 += (1 / ((error[i])**2))

avg_mean = sum1 / sum2
print(avg_mean)

error_mean = (1 / sum2)**0.5
print(error_mean)
```

weighted mean lifetime = 0.9089185199574897
uncertainty of the mean = 0.00020318737026848627

Question 5:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats


f = pd.read_csv('exoplanet.eu_catalog.csv')
f = f['eccentricity']

f_2 = f.dropna()
f_2.drop(f_2[f_2 == 0].index, inplace = True)

plt.hist(f_2, bins = 50, color = 'green', alpha = 0.5)
plt.xlim(xmin = 0, xmax = 1)
plt.ylabel('Exoplanets')
plt.xlabel('Eccentricity')
plt.show()
```

```
e_new=stats.boxcox(f_2)
plt.hist(e_new, bins=20, alpha=0.5)

plt.ylabel('no of exoplanets')
plt.xlabel('eccentricity')
plt.show()
```



The End