PyLab 5 – Random Number Analysis

October 25, 2018 – Ayush Pandhi (1003227457) and Xi Chen (1003146222)

1 Abstract

Statistical distributions such as Poisson and Gaussian are very useful when describing the results of an experiment in which one counts events that occur at random but at an expected average rate and experiments with random measurements respectively. In Poisson statistics, $\sigma \sim \sqrt{\mu}$, and it is predicted that for large data sets (i.e. large mean and standard deviations) the Poisson distribution becomes increasingly similar to a Gaussian distribution. In the experiment, the radioactive emission counts from a fiesta plate as well as background emission are analyzed to determine how well they are represented by these forms of statistical analysis. Results show that indeed for a large data set, both Poisson and Gaussian are very similar distributions and approximate the binned data relatively well. However, for smaller data sets this does not seem to be the case and it is not clear that either distribution describes the data set well.

2 Introduction

The primary focus of this experiment was to conduct statistical analysis of the provided data sets. The mean, μ_x, and standard deviation, σ_x, of a given parameter x were computed as:

$$\mu_x = \frac{\sum_i x_i}{N} \quad and \quad \sigma_x = \sqrt{\frac{\sum_i (x_i - \mu_x)^2}{N}}$$

Where x_i is the i^{th} value of the parameter and N is the total number of data points in the set. The standard deviation represents the systematic error of the system; consistent error that is innate to the experiment as opposed to random error. The histograms produced from the data sets were analyzed with Poisson and Gaussian statistics. Firstly, Poisson statistics is particularly useful for describing the results of an experiment in which one counts events that occur at random but at an expected average rate. The probability distribution given by Poisson statistics is described by:

$$P(x, \mu) = \frac{\mu^x}{x!} e^{-\mu}$$

Which depicts the probability of having x events when the mean of that parameter is μ. It should be noted that $P(x, \mu)$ is a discrete function that exists for $x > 0$ and $x \in \mathbb{Z}$. Importantly, the standard deviation of a Poisson distribution is $\sqrt{\mu}$. Importantly, for large numbers in the independent variable (i.e.

in the limit that $x \rightarrow \infty$ because then μ is large), the Poisson distribution becomes similar to a Gaussian distribution.

Gaussian statistics is often used to describe the distribution of random measurements. The probability distribution of a Gaussian is normalized to 1 and centered at the mean value of the parameter, μ, and follows:

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

Therefore, a small σ leads to precise measurements as there is a high probability of measured values being near the mean of the parameter.

We aim to compare the radioactive emission from the fiesta plate and background emission to Poisson and Gaussian distributions for corresponding μ and σ parameters. Additionally, we analyze whether the Poisson and Gaussian distributions become more similar for a larger data set.

3 Methods and Materials

The materials utilized for this experiment were: a Geiger counter, a generator, a fiesta plate (with Uranium Oxide glaze), and the Jupyter Notebooks software.

4 Experimental Procedure

A Geiger counter was setup to measure the rate of radioactive emissions from the Uranium Oxide glaze on the fiesta plate at 3 second intervals for 60 total samples. The Geiger counter then measures the rate of radioactive emission and the data is collected into a .txt file. Additionally, the Geiger counter is run to measure the background emission of the laboratory environment at 20 second intervals for 60 total samples and this data is also stored in a .txt file.

5 Results

Before analyzing the data, the mean background emission must be removed as noise from the radioactive emission to discern its true rate of decay. This is done by computing the mean value from the background emission data file and subtracting it from each individual data point of the fiesta plate data set. Using this updated set of data, a histogram is plotted alongside Poisson and Gaussian distributions which are generated using the mean and standard deviation of the data set ($\mu = 35$ and $\sigma = 6.2$). The histogram seems to roughly follow both distributions and it is seen that both the Poisson and Gaussian distributions are very similar, which agrees with the expected outcome for large data sets (Figure 1).

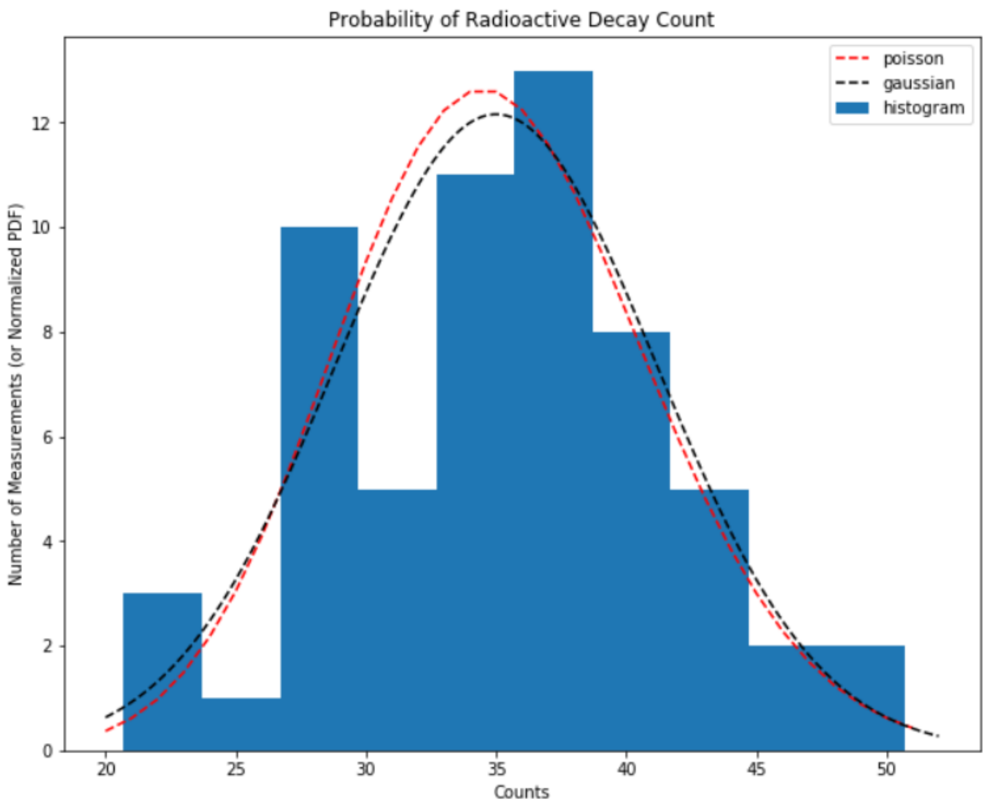

Figure 1: *Histogram of 60 fiesta plate radioactive emission count measurements (bin-width = 3.1); (Dashed Red Line) Expected Poisson distribution of* $\mu = 35$*, (Dashed Black Line) Expected Gaussian distribution of* $\mu = 35$ *and* $\sigma = 6.2$*.*

Each point in the Poisson distribution here is normalized to match the histogram using:

$$P(x_i, \mu)_{normalized} = C \frac{P(x_i, \mu)}{\sum_i P(x_i, \mu)} (x_{i+1} - x_i)$$

$$C = Total\ number\ of\ measurements(number\ of\ intervals\ in\ x/number\ of\ bins)$$

Specifically for Figure 1, $C = 60(31/10) = 186$. This normalizes the given distribution to the total number of measurements and the same technique is used to also normalize the Gaussian distribution. An alternative method to normalize the data is to normalize the histogram to 1, however to remain true to the data set and have non-decimal values for the number of measurements, the previously explained method was preferred.

In a similar manner, the background radiation emission was plotted as a histogram alongside Poisson and Gaussian distributions with parameters $\mu = 0.3$ and $\sigma = 0.5$. The histogram is not well represented by either distribution, although the Gaussian fit does seem slightly more reasonable. It is also found that, for smaller data sets such as the background emission, the Poisson and Gaussian distributions are wildly different and do match up well with each other (Figure 2).

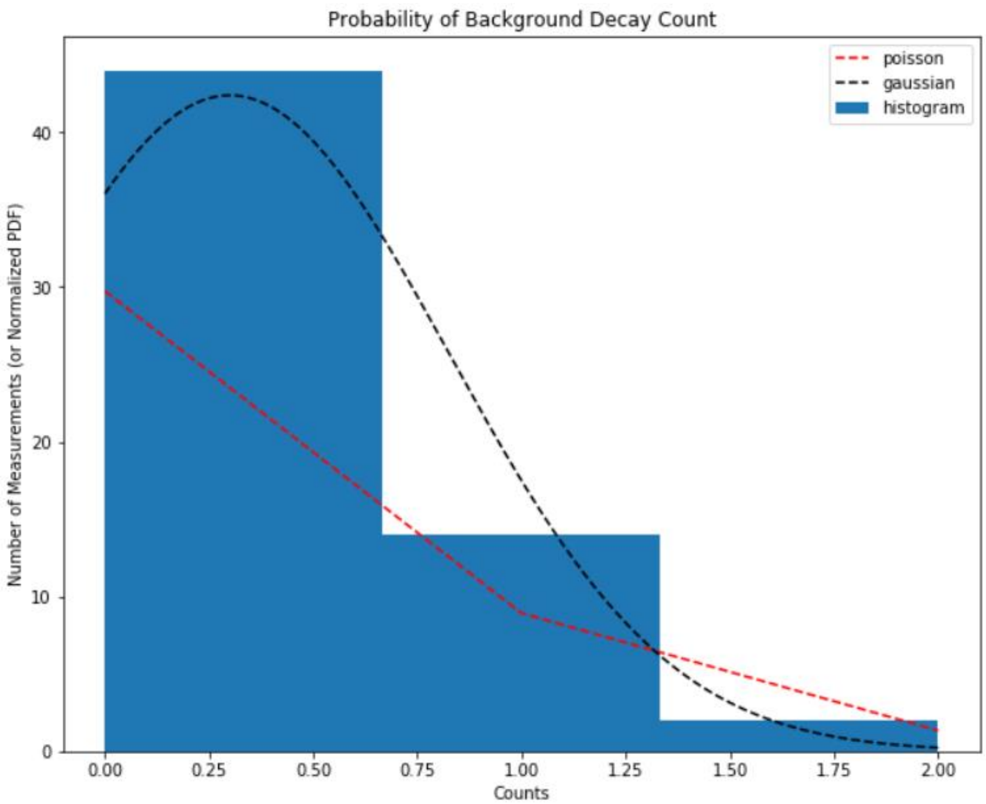

Figure 2: *Histogram of 60 background radioactive emission count measurements (bin-width = 0.67); (Dashed Red Line) Expected Poisson distribution of* $\mu = 0.30$*, (Dashed Black Line) Expected Gaussian distribution of* $\mu = 0.30$ *and* $\sigma = 0.53$*.*

6 Discussion

It is found that the radioactive emission from the fiesta plate approximately follows both a Poisson and Gaussian distribution for its mean and standard deviation parameters: $\mu = 35$ and $\sigma = 6.2$. Additionally, in this data set the independent variable, radioactive emission counts, is considered large and we see that the Poisson and Gaussian statistics line up very closely with one another. This agrees with the theoretical prediction that in the limit that the independent variable $x \rightarrow \infty$, the Poisson distribution becomes similar to a Gaussian distribution. It is also found that the background radioactive emission, which has a very small independent variable with parameters $\mu = 0.30$ and $\sigma = 0.53$, does not agree particularly well with either statistical method. The Poisson and Gaussian distributions also do not look similar for such low mean and standard deviation parameters. This also agrees with the idea that the two statistical distributions don't line up well with one another for small data sets.

The uncertainty in the experiment first arises from the error in experimental values from the Geiger counter which follows counting statistics. Thus the error in decay count is determined according to the rule:

$$\sigma_{source} = \sqrt{N}$$

Where σ_{source} is the uncertainty in the decay count and N is the number of events. Additionally, when correcting for background radiation, the error is propagated as:

$$\sigma_{signal} = \sqrt{\sigma_{source}^2 + \sigma_{background}^2} = \sqrt{N_{source} + N_{background}}$$

The strength of this experimental design is that it is performed using highly precise equipment and thus the uncertainty in the data is very low and produces clear results. The background radiation was also computed to correctly remove noise from the emission data. On the other hand, the experiment is not easily replicated as it requires access to a Geiger counter and the expertise to operate and monitor the experiment. However, for an experiment on radioactive emission, a complex setup cannot be avoided without repercussions in the accuracy of the data set and thus the experimental setup should not undergo changes.

7 Conclusion

From the experiment we conclude that for the large data set (fiesta plate), the histogram of the data is approximated well by Poisson and Gaussian statistics for the corresponding mean and standard deviation parameters ($\mu = 35$ and $\sigma = 6.2$). In addition, the two statistical distributions are very similar for this large data set. For the background emission, which was a much smaller data set in terms of emission counts ($\mu = 0.30$ and $\sigma = 0.53$), the histogram was wildly different from both distributions and we find that the Poisson and Gaussian distributions do not line up well with one another when the data set is small.

8 References

[1] Pandhi, A. and Chen, X., University of Toronto, Toronto, ON. "PHY224 Laboratory Notes: PyLab 5 – Random Number Analysis", October 2018.

9 Appendices

The full code used for this lab can be found below as well as on the author's Github as "PyLab 5.py": https://github.com/AyushPandhi/Pandhi_Ayush_PHY224/tree/master/PyLab%205.

```
#PyLab 5: Random Number Analysis
#Author: Ayush Pandhi (1003227457)
#Date: 10/25/2018

#Importing required modules
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
from scipy.stats import poisson

#Loading the radioactive decay data file
sample = np.loadtxt('TuesdayPMPlate_20181009.txt', skiprows=2, usecols=(0,))
count = np.loadtxt('TuesdayPMPlate_20181009.txt', skiprows=2, usecols=(1,))

#Loading the background radiation data file
sample_bg = np.loadtxt('TuesdayPMBackground_20181009.txt', skiprows=2, usecols=(0,))
count_bg = np.loadtxt('TuesdayPMBackground_20181009.txt', skiprows=2, usecols=(1,))

#Getting the mean background radiation
mean_count_bg = np.mean(count_bg)

#Adjusting decay data count by subtracting mean background count
count = count - mean_count_bg

#Histogram for the Fiesta Plate data
plt.figure(figsize=(10,8))
plt.hist(count, bins=10, label='histogram')

#Adding in the Poisson Distribution
xlin = np.arange(20, 52)
mu = np.mean(count)
pd = poisson(mu)
plt.plot(xlin, 186*(pd.pmf(xlin)/np.sum(pd.pmf(xlin))/(xlin[1] - xlin[0])), 'r--', label='poisson')

#Adding in the Gaussian Distribution
xlin2 = np.linspace(20, 52, 1000)
stddev = np.std(count)
gd = norm(mu, stddev)
```

```
plt.plot(xlin2, 186*(gd.pdf(xlin2)/np.sum(gd.pdf(xlin2))/(xlin2[1] - xlin2[0])), 'k--', label='gaussian')

#Plot title, legend, labels
plt.title('Probability of Radioactive Decay Count')
plt.xlabel('Counts')
plt.ylabel('Number of Measurements (or Normalized PDF)')
plt.legend()
plt.show()

#Histogram for the background data
plt.figure(figsize=(10,8))
plt.hist(count_bg, bins=3, label='histogram')

#Adding in the Poisson Distribution
xlin = np.arange(0, 3)
mu_bg = np.mean(count_bg)
pd_bg = poisson(mu_bg)
plt.plot(xlin, 40*(pd_bg.pmf(xlin)/np.sum(pd_bg.pmf(xlin))/(xlin[1] - xlin[0])), 'r--', label='poisson')

#Adding in the Gaussian Distribution
xlin2 = np.linspace(0, 2, 1000)
stddev_bg = np.std(count_bg)
gd_bg = norm(mu_bg, stddev_bg)
plt.plot(xlin2, 40*(gd_bg.pdf(xlin2)/np.sum(gd_bg.pdf(xlin2))/(xlin2[1] - xlin2[0])), 'k--', label='gaussian')

#Plot title, legend, labels
plt.title('Probability of Background Decay Count')
plt.xlabel('Counts')
plt.ylabel('Number of Measurements (or Normalized PDF)')
plt.legend()
plt.show()
```