Photon Counting and the Statistics of Light

Ayushi Singh, Anita Bahmanyar, Carly Berard

ayushi.singh@mail.utoronto.ca

Astronomy and Physics, University of Toronto, ON

30 September, 2013

Abstract

This experiment examined the distribution of photon collected using PMT for different number of samples and time intervals. Python programming language was used to calculate mean and standard deviation, and to plot the graphs and histograms. Various time vs. count plots show that increasing number of samples give us more precise counts per sample, where as, increasing the time intervals widens the range of count per sample. The shape of histograms (frequency vs. count graphs) and mean vs. variance graphs strongly support Poisson distribution, even at higher time intervals. It is also induced that with large number of samples N, precision of a measurement of mean increased by the factor of $\sqrt{N}$.

1 Introduction

The purpose of this report is to discuss the physical limitation of detecting light. The two main parameters: Number of samples and time interval was explored. Photons collected through PMT with these parameter produce a histogram that fits the poisson distribution. In this report, this idea was tested and compared to other distributions. The report also investigate the fundamental statistics and plots created using Python.

2 Equipment and Method

Photomultipler Tube (PMT) was used to collected the data. It is a black box that receive photon and uses photoelectric effect to multiply the number of photons. PMT outputs a pulse [**?**]. Using the python script (PMT Python Module [**?**]), provided in advance, number of counts per sample was recored. This depends on the two parameters: number of samples and time interval, which is specified before running the PMT.

3   Data Summary

The data collected from PMT using PMT Python Module was represented in python array of counts per sample and saved in a file. These files were then later used to manipulate the data. Multiple data was collected with different time interval and number of samples using the method explained in Section 2. For each combination number of samples and time interval, 6 or 10 data set was collected.

Table 1: Summary of all the data collected for the experiment

Data Number	Time Interval (s)	Number of Samples	Number of data set	Comments
1	0.001	100	6	
2	0.01	400	6	
3	0.01	400	6	comparing mean and variance
4	0.0125	400	6	comparing mean and variance
5	0.025	400	6	comparing mean and variance
6	0.0375	400	6	comparing mean and variance
7	0.05	400	6	comparing mean and variance
8	0.08	400	6	comparing mean and variance
9	0.1	400	6	comparing mean and variance
10	0.001	1000	6	for theoretical distribution
11	0.1	1000	6	for theoretical distribution
12	0.001	2	10	exploring MOM and SDOM
13	0.001	4	10	exploring MOM and SDOM
14	0.001	8	10	exploring MOM and SDOM
15	0.001	16	10	exploring MOM and SDOM
16	0.001	32	10	exploring MOM and SDOM
17	0.001	64	10	exploring MOM and SDOM
18	0.001	128	10	exploring MOM and SDOM
19	0.001	256	10	exploring MOM and SDOM
20	0.001	512	10	exploring MOM and SDOM
21	0.001	1024	10	exploring MOM and SDOM
22	0.001	2048	10	exploring MOM and SDOM
23	0.001	100	6	dark count
24	0.001	1000	6	dark count
25	0.01	400	6	dark count

4   Data Reduction

Most of the data collected went though two main methods of calculating mean and standard deviation. These are used in later sections for analyzing and understanding the behaviour of data.

$$\mu = \frac{\sum x_i}{N}, \tag{1}$$

where μ is the mean and N is number of samples. This is the average value of the data.

$$\sigma = \sqrt{\frac{\sum x_i - \mu}{N - 1}}, \tag{2}$$

where σ is the standard deviation. This value represents the uncertainty in the measurement for that data set.

5   Discussion

5.1   Variation in light count

Each data file had list of counts per sample, as mentioned in Section 3. The data from the Table 1 is used to graph time vs count plots and histograms as followed:
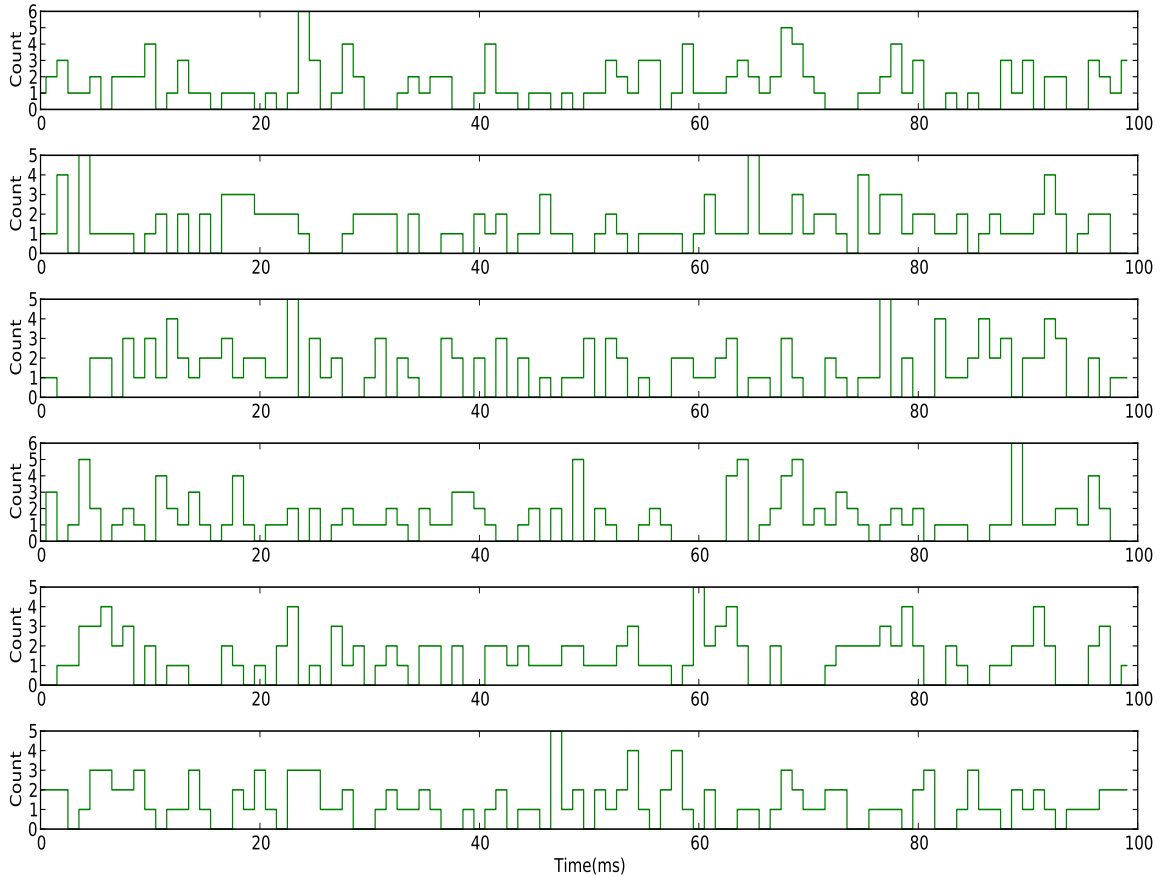
Figure 1: Time vs. count per sample graph for data number 1 from Table 1. There are six sets of data, where time interval of each sample is 0.001s for 100 samples.
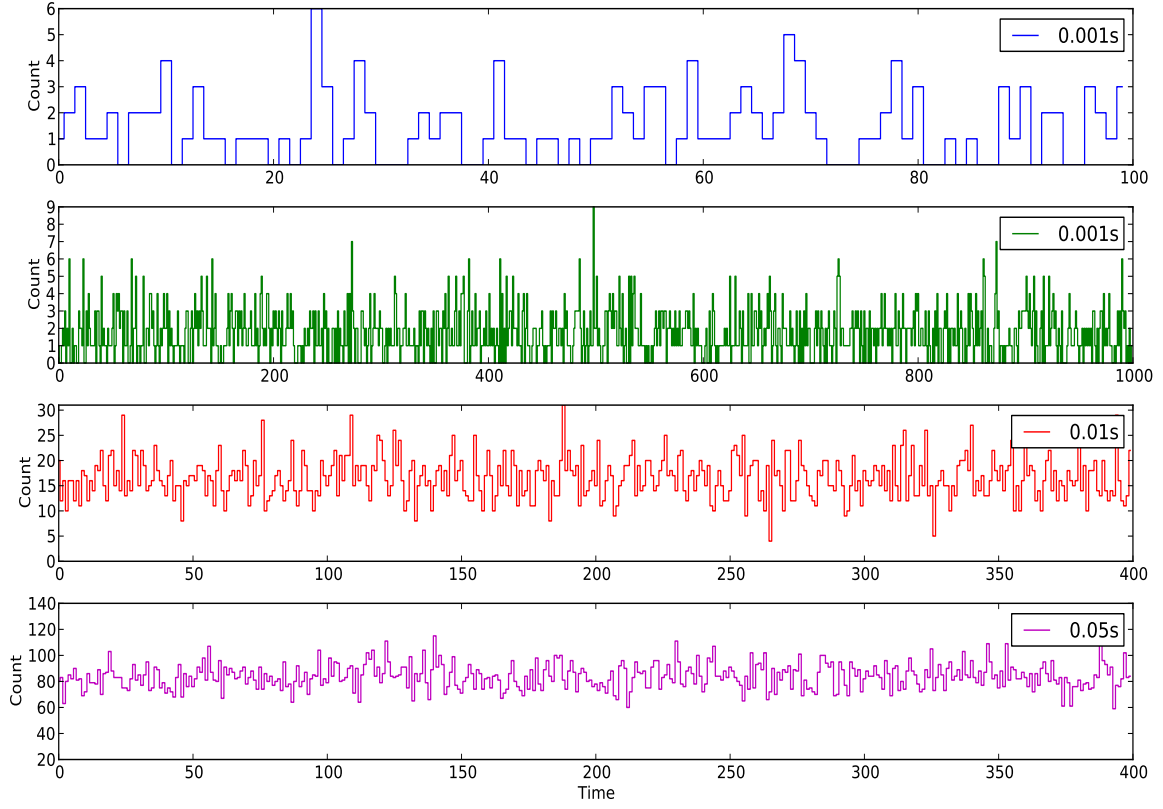
Figure 2: Time vs. count per sample graph for data number 1, 3, 7 and 10 from Table 1. Only one of each set is used. Blue and green graphs show affect of change in sample numbers, where as, red and magenta graphs show variation by changing time interval.
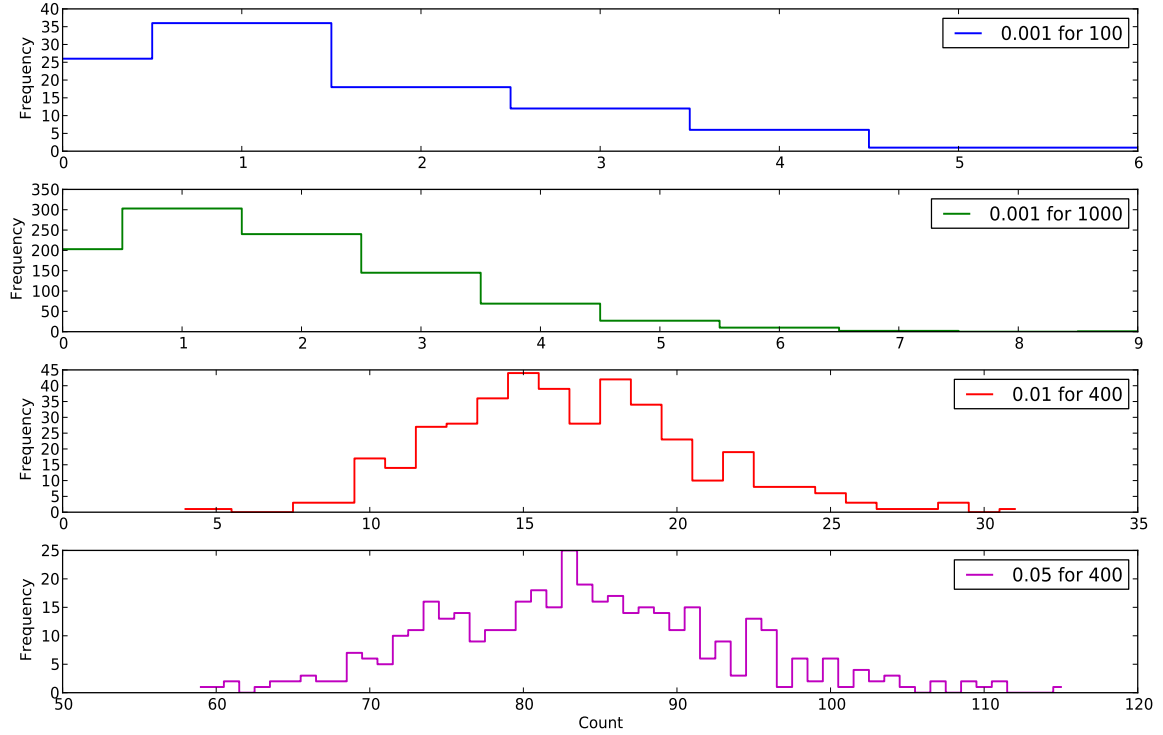

Figure 3: Histogram of data plotted in Figure 2

The Figure 1 shows that even with a same time interval and number of samples, data obtained in these 6 set is quite dissimilar, that is, no two plots look alike. This exhibits that PMT gives out photons, randomly and within a range of values. In this case, the range of count is from 0-6. However, there is a peak value that occurs more often.

Table 2: Calculated values for light count

Plot colour	Mean	Standard Deviation	Count rate (Hz)
Blue	1.43	1.30	1430
Green	1.71	1.40	1710
Red	16.51	4.14	1651
Magenta	83.78	9.63	1676

Figure 2 and 3 shows the change that occur by chaining two parameters. Blue and green plot displays that by increasing the number of samples from 100 to 1000, the range of count is increased to 0-9, even though the time interval is unchanged. There are samples with count as high as 9 in green plot, where as in blue plot the maximum is only 6. This increases the precision of the result that can be seen in Table 2. Red and magenta plots, on the other hand, deal with fixed number of samples and change in time interval from 0.01s to 0.05s. In Figure 2, it is explicit that more photons are detected per sample by increasing the time interval, as can be seen from the mean value in Table 2. However, this widens the range of count per sample, even though, numbers of sample for both is 400. As exhibited in Figure 3, 15 counts occurred 45 time at 0.01s. Where as, at 0.05s, only 82 counts occurred only 25 time. In the Table 2, the count rate value increases with the increase in both the parameters.

5.2 Effect of Dark Count

Dark counts are counts taken using PMT with lights off. This generates a similar plots but as light counts (PMT with light on), but they are different in the values they provide us. This shows that PMT is not perfect black box and detects photons even when the lights are off. This is considered noise.

Table 3: Calculated values for dark count

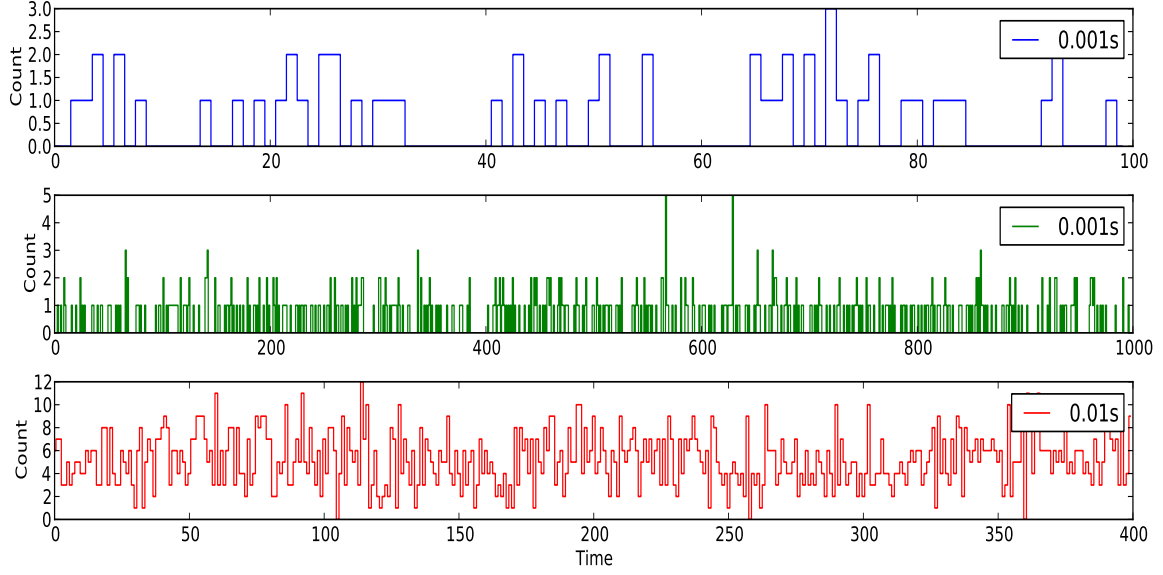| Plot colour	Mean	Standard Deviation	Count rate (Hz)
Blue	0.56	0.76	560
Green	0.49	0.69	490
Red	5.21	2.31	521

Figure 4: Time vs. count per sample graph for dark count for three different set of data for Figure 2 and Figure 3. This is from data number 23, 24 and 25 from Table 1
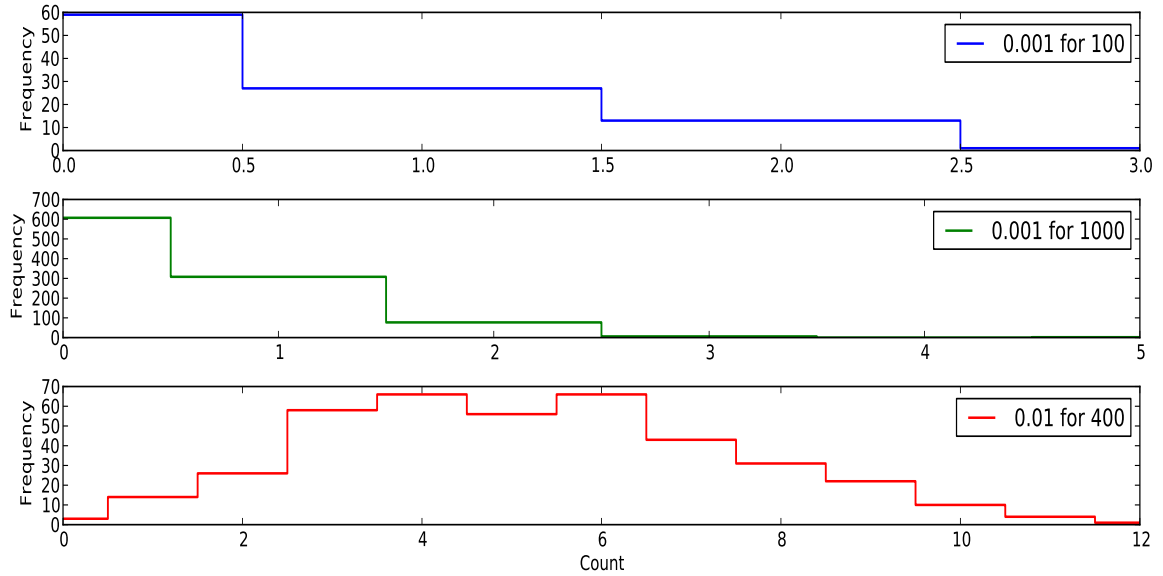

Figure 5: Histogram of data plotted in Figure 4.

From Figure 4 and 5, it is clear that there is some noise in the PMT. Blue and green plots have more counts at zero where time is 0.001s. However, in red plot, there is a peak at 4 and 6 count showing that by increasing the time to 0.01s, more noise was introduced. There are few factors that could have contributed to this noise. PMT is not perfect black box and some of the photons from outside were detected. Another contributor to the noise could possibly be the internal heat of the PMT. Upon heating up, the PMT could have released photons that were detected. It explains the larger mean

6

and standard deviation for red and green compared to blue in Table 3. PMT is working for longer period of time.

5.3 Count rates of different sample

Data number 1and 2 from Table 1 is 6 sets of data for 100 and 400 number of samples, respectively . It was used to calculate mean and standard deviation for every data set using Equation 1 and 2. Count rate was calculated by $\frac{mean}{time\ interval}$. Then the mean and standard deviation of the 6 count rates for each set was calculated using same equation. The mean and standard deviation of count rate is:

- For 100 number of samples: 1366.67 and 43.20

- For 400 number of samples: 1664.67 and 18.09

From the result it clear that by increasing the number of samples, higher count rate can be obtained with smaller standard deviation.

5.4 Relationship between mean and variance

Data number 3-9 from Table 1 was used to calculate and analyze the relationship between the mean and variance. Mean is calculated using Equation 1. And variance is standard deviation squared [?]; therefore, it is σ^2, where sigma is from Equation 2.
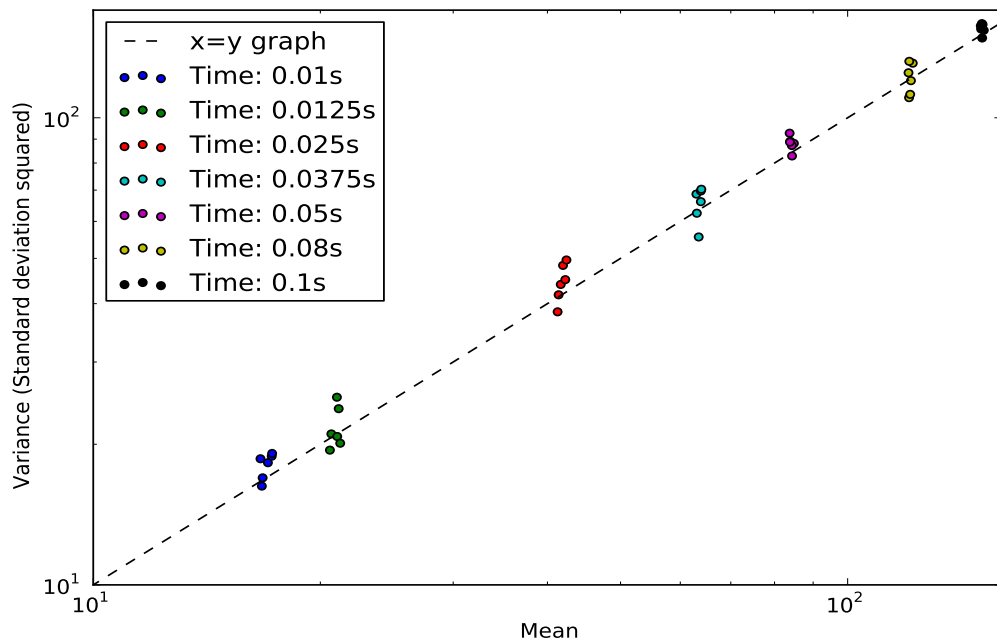

Figure 6: Mean vs. Variance (standard deviation squared) log-log graph. The sample size for all the data is 400. There are six sets of data for each of seven different time interval.

7

Data here have 400 number of sample each but various in time interval. The log-log graph (Figure 6) of mean vs. variance clearly show a linear relationship. The line $y = x$ runs through the middle of the data points, even for the large number. That is, as the time interval increases the mean and the variance increases with linear proportionality. This can be interpreted as property of poisson distribution [?], where mean is equal to variance.

5.5   Comparing results with theoretical distribution

From the Table 1, data number 10 and 11 were used to plot histogram, poisson distribution and gaussian distribution.

$$P(x; \mu) = \frac{\mu^x}{x!} \exp(-\mu), \tag{3}$$

where μ is the mean and $x = 1, 2, 3....$ This equation describes the poisson distribution. It has discrete value [?].

$$P(x; \mu; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right], \tag{4}$$

where μ is the mean and σ is the standard deviation. This is equation for gaussian distribution and it is continuos distribution [?].
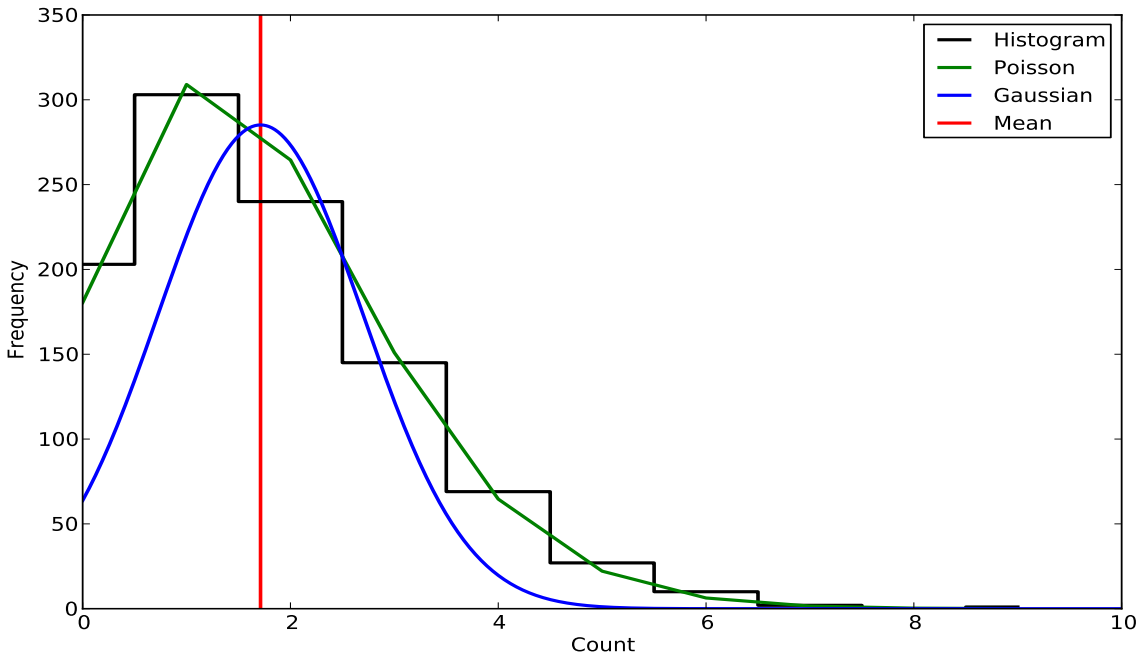

Figure 7: Histogram of data number 10 from Table 1 with 1000 samples and time interval as 0.001. Here blues line is gaussian distribution, green is poisson distribution and red line indicated the mean value
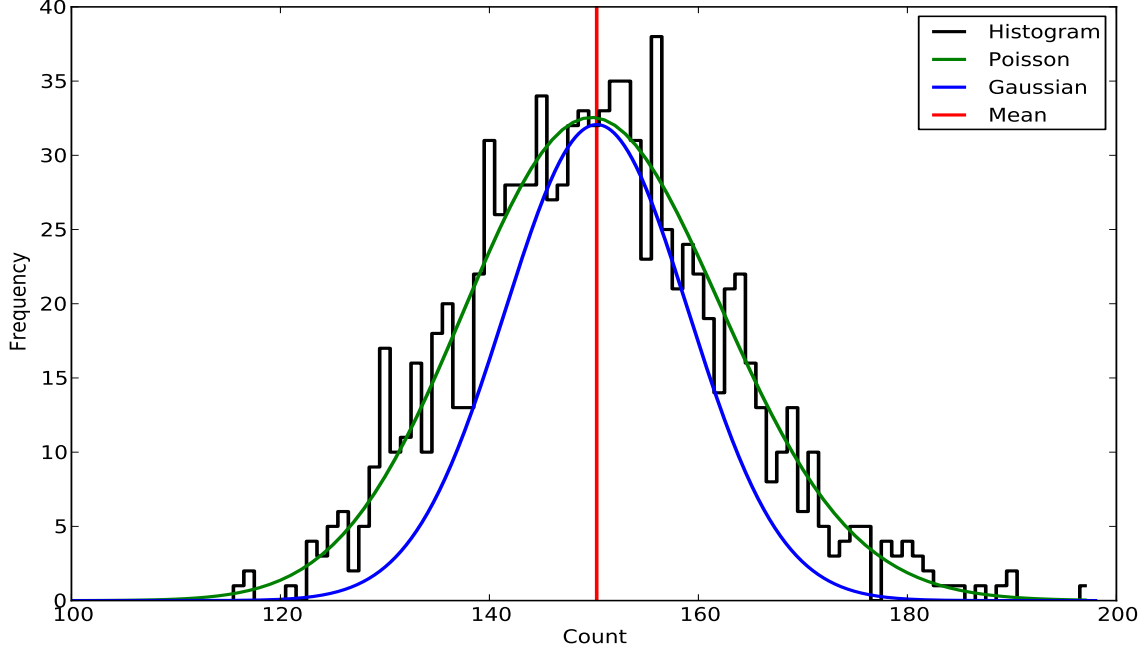
Figure 8: Histogram of data number 11 from Table 1with 1000 samples and time interval as 0.1. Here blues line is gaussian distribution, green is poisson distribution and red line indicated the mean value

The two figures (Figure 7 and 8) two data with 1000 number of samples but different time interval: 0.001s and 0.01s, respectively. There is histogram of the data in black and mean in red on each plot. Poisson and Gaussian distribution on the plots are multiplied by 1000 to obtain correct scaling for comparing with histogram. One of the main thing to notice in these two graph is that for smaller time interval gaussian distribution (by Equation 4) does not right the histogram, however it does for the larger time interval. Poisson distribution (by Equation 3) fits both the histograms. This further proves that data collected correlates with poisson distribution. Gaussian is good approximate, if the time interval of data is large, as in Figure 8.

5.6   Exploring mean of mean and standard deviation of mean

Data number 12 - 22 from Table 1 was used to calculate the mean of mean (MOM) using Equation 1 twice and standard deviation of mean (SDOM) using Equation 1 and 2. This was then plotted in count vs. number of samples plot.

The Figure 9 also have a theoretical graph, which was derived by the following method. For poisson distribution variance (standard deviation squared) is same as mean value (μ) of the data as explained in Section 5.4. Therefore, the standard deviation of poisson distribution is $\sigma = \sqrt{\mu}$. And for expected error of mean, the formula is $\frac{\sigma}{\sqrt{N}}$ [?]. Combining these two we get:

$$\sigma_{exp} = \sqrt{\frac{\mu}{N}}, \tag{5}$$

where μ is the mean and N is number of samples for one data. These value was for each set of data. To get σ_{exp} for each number of samples. The mean of σ_{exp} for a corresponding number of sample was taken using Equation 1. This is then the theoretical standard deviation representing poisson distribution.
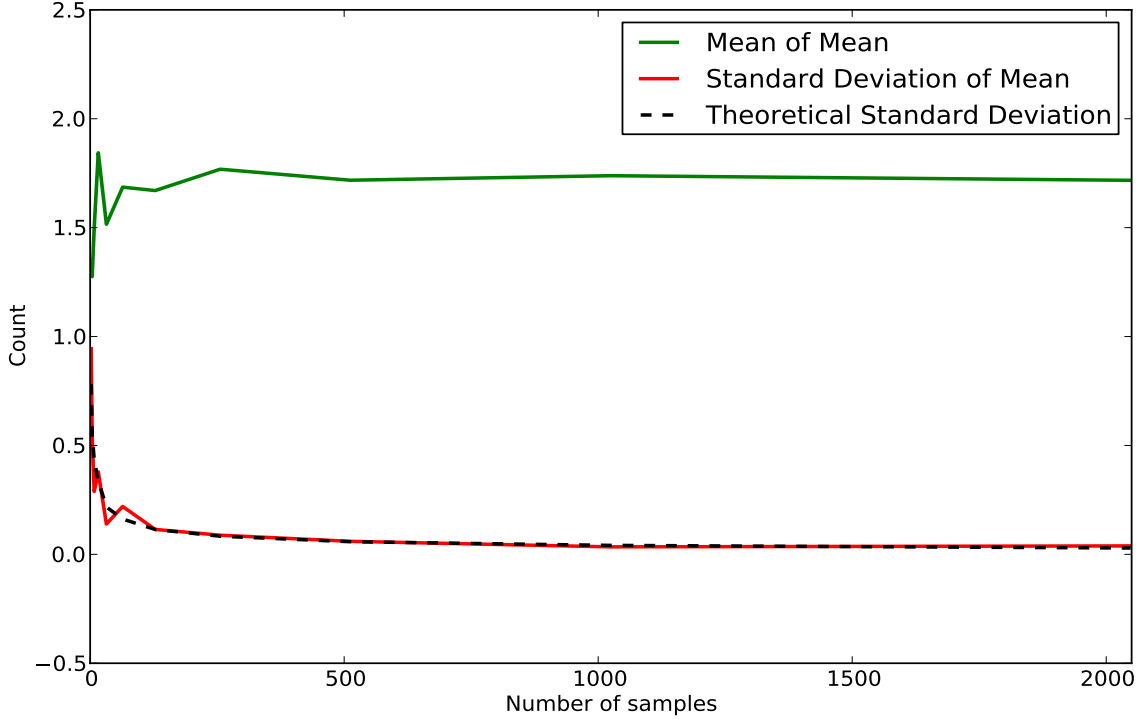

Figure 9: Plot of mean of mean and standard deviation of mean vs. number of samples. This plot also show the theoretical prediction of the mean by Equation 5.

It can be induced from Figure 9 that as the number of samples increases, then MOM converge to a certain value and SDOM get closer to zero. Therefore with high number of samples, not only more accurate MOM can be obtained but also the error can be reduced. However, having too big number of samples does not make a huge difference because the two values stay very much consistent. And, it is clear in the figure that the theoretical standard deviation fits the SDOM very well. Therefore, to improve the precision of a measurement of mean by factor of 2, number of samples should be increased by the 4. This can be induced from Equation 5.

6 Conclusion

Photons detected using PMT based on different number of samples and time intervals show high correlation with poisson distribution. Mean value is equal to the variance

for all data set. However, at larger time interval, the data set is also approximated to gaussian distribution, but not completely. It also shown that at larger number of samples the MOM approaches a certain value and SDOM approaches zero. This also fits with theoretical standard deviation calculated by using poisson statistics and error propagation.