Lab 1: Photon Counting and the Statistics of Light

Diana Kossakowski
Lab Group: ∼foxy∼
Melanie Archipley, Christopher Agostino

September 23, 2015

Abstract

In this lab, we examined how the arrival of photons follows a Poisson probability distribution. Using a photomultiplier tube and the CoinPro interface, our data was collected from an LED, where we were able to control the brightness. By exploring the statistical features of computing the mean, the standard deviation, and the standard error of time intervals between consecutive events, we were able to plot and analyze the relationships between these calculations to find that photon counting does indeed adhere to a Poisson distribution.

1 Introduction

The main purpose of this lab is to explore the Poisson nature of the detection of light and gain insight on the importance of statistics, mainly precision and accuracy, in experimental science. Photons were detected by a very sensitive light detector known as a Photomultiplier tube (PMT). The PMT relies on the Photoelectric Effect, where incoming photons strike a metal surface with enough energy enough to eject an electron from the surface. The job of the PMT is to amplify the signal by taking that one electron caused by the incoming photon and creating a cascading effect of a plethora of electrons, which the CoinPro Interface measures the interval between pulses, or the arrival of the electrons.

Using this data, we were able to examine the statistical properties of photon counting via various python plots and calculations of the sample mean, the sample standard deviation, and histograms, to find that photon arrivals follow a Poisson distribution.

2 Equipment, Data, and Methods

For this lab, we used the CoinPro interface on the lab computer to collect the photon counts data coming from the photomultiplier tube (PMT). The PMT amplifies the signal of the photons coming in so that the CoinPro program can detect and record individual photons. To work with the data and display nice graphs, we used Python and its matplotlib library. The data collected was not questionable as it showed no anomalies or systematic errors.

3 Analysis

To begin the lab, we collected our data of 20,000 events (photon arrivals) from the photomultiplier tube. We decided to collect 20,000 events to ensure we had enough information. Because the data was collected through the CoinPro program, the time (clock ticks) was stored as 32-bit integers.

Clock Ticks

We plotted the intervals between the events as shown in Figure 1. Figure 1(a) and (b) represent the same data, though (b) the bottom graph presents the data more clearly by showing how the integers go from 2^{31} to -2^{31} with a gap in between, which is not as evidently clear in (a) the top graph. The reason there exists a gap between -2^{31} and 2^{31} is due to the CoinPro counter. The CoinPro counter uses 32 bits, meaning that when the ticks reach 2,147,483,647 (2^{31}), the next tick "rolls over" to -2,147,483,648 (-2^{31}).

When the "dtype='int32'" argument is left out, the data will load in as floats which will end up using a little bit more memory which can be problematic when reading in a large amount of data.

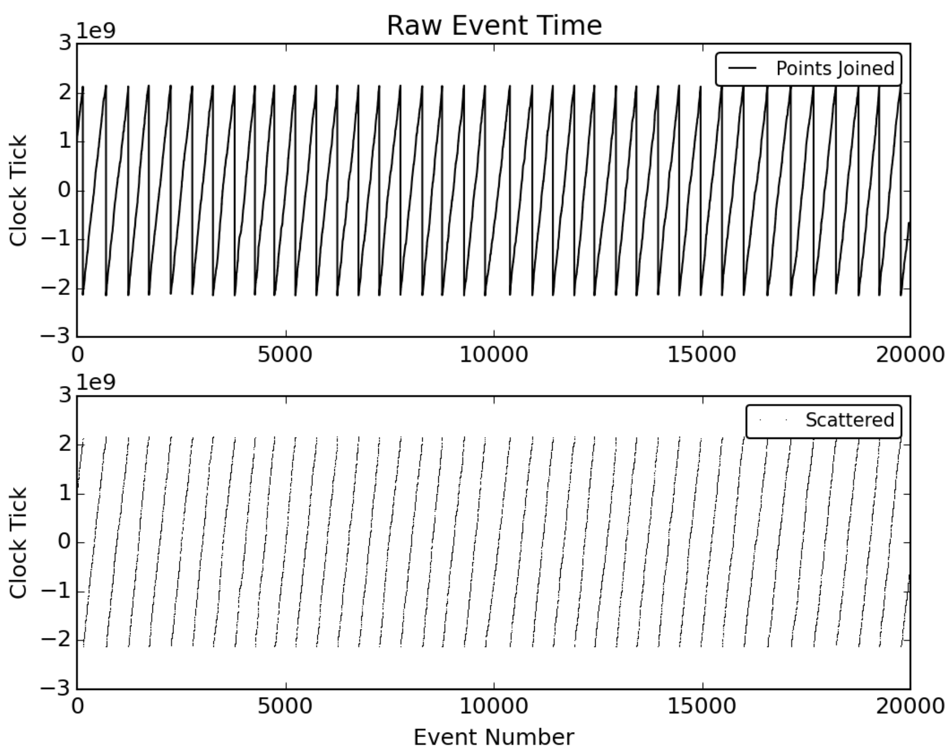

Figure 1: Time Series. Raw data is shown as event number vs. time (measured in clock ticks). (a) The top graph uses matplotlib's default plot style of joining points, and (b) the bottom graph is plotted as single points, clearly showing the gap between -2^{31} and 2^{31}.

Interval Between Events

Plotting the raw data (Figure 1) does not do the data much justice as if we were to plot the data more suggestively by calculating the time interval between two consecutive events. To calculate each time interval, dt_i, we perform $dt_i = t_{i+1} - ti$ where t_i is the i-th event. We then plotted event number and time interval, as shown in Figure 2.

Figure 2(b) suggests a better interpretation of the data as we can see that most of the time intervals are mostly scattered at less than $.2^8$ clock ticks. It is interesting to note how the time interval between events is not constant as one would think, but rather quite distributed.

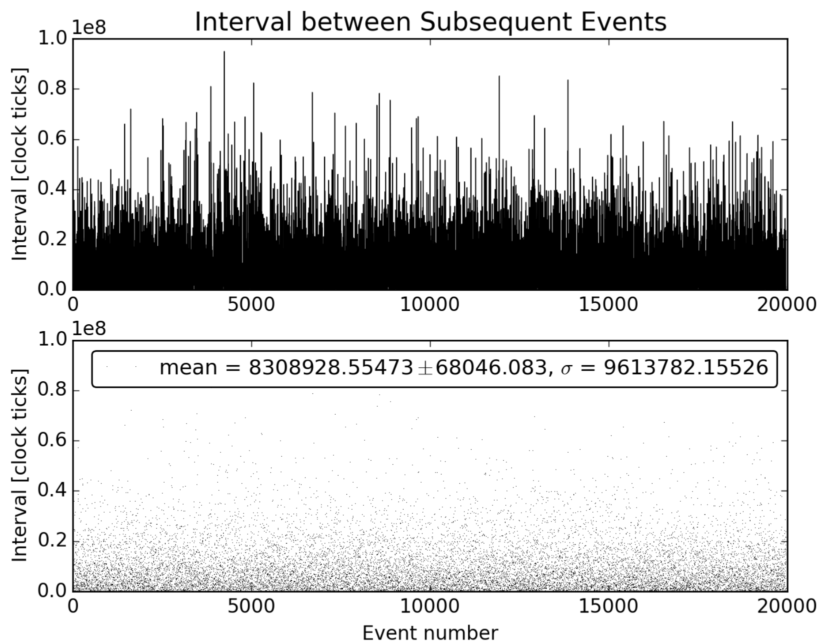

Figure 2: Time Interval between Subsequent Events Captured. (a) The top plot shows the time interval between points. (b) The bottom plot shows the time interval between points scattered giving a better description of the data and how it is spread out.

Statistical Analysis

Next, we looked at the statistical properties of the data, specifically focusing on the sample mean $\bar{x}$, the standard deviation μ, and the standard deviation of the mean, defined by $\frac{s}{\sqrt{N}}$, where s represents the sample standard deviation and N represents the number of data points.

We focused on calculating the sample mean for different chunks of data. The sample mean is defined as:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

where N is the number of samples taken. The true mean of a data set is defined as:

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2}$$

and the true standard deviation is defined as

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2} \tag{3}$$

where μ is the true mean. The sample standard deviation is given by

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{4}$$

where $\bar{x}$ is the sample mean given by equation 1. We already notice that N needs to be quite large in order to obtain an accurate measurement of the mean. With smaller chunks (small N) of data, the mean will fluctuate more and will be more spread out, therefore making the scatter much higher; whereas with bigger chunks (big N) of data, the mean will be more stabilized and less spread out, therefore making the scatter very minimal. The standard deviation of the mean (SDOM) can tell us how precise our data is; a smaller SDOM means that most of the data is localized to the mean and a larger SDOM means that the data is generally spread out from the mean indicating a high error.

With this in mind, we partitioned the data into different sized chunks with sizes of 1000, 100, and 10 events, and took the means of each interval in order to see how standard deviation is affected (Figure 3). As we had expected, the standard deviation was higher for smaller chunks of data.

Next, we took a look at how the mean interval changes as we take progressively larger fractions of data, shown in Figure 4. Figure 4(a) and Figure 4(b) differ by the number of events the fractions are increased by, 1000 and 100 respectfully. Both plots illustrate how as we include more and more data, the sample mean value converges to a limit, which is the true mean. Recalling equations 1 & 2, the mean we are calculating is not the actual mean μ, but rather an approximation. As we increase N (number of data points considered), $\bar{x}$ converges toward μ, but at smaller values of N, $\bar{x}$ will fluctuate greatly from the real mean, which Figure 4 exhibits so nicely. Additionally, Figure 4 brings to light how the SDOM significantly decreases for the increasing fractions of 100, indicating that those points in (b) are closer to the mean.

Next, it would be interesting and fun to see how the SDOM correlates with different sized chunks of data. As Figure 5 displays, we see a general, downward trend for SDOM as we increase the size of the chunk. Because SDOM is the error of the mean, we can clearly see that adding more points leads to a decreasing value for the SDOM. We can see that the precision of the mean is quantified by the standard deviation of the mean, given by $\frac{s}{\sqrt{N}}$, where s is the standard deviation for the data set and N is the number of event intervals, where Figure 6 displays this relation quite well. The plot also shows how 100 times more samples are needed to get 10 times more precise data.

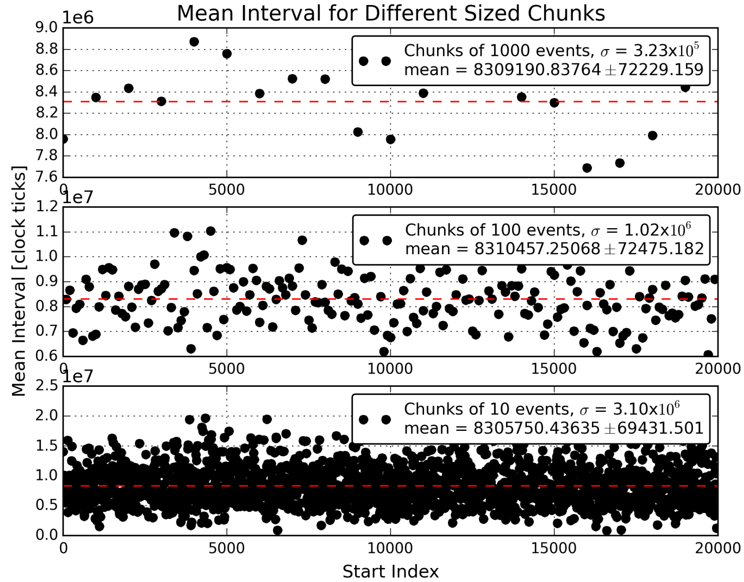

Figure 3: (a), (b), and (c) refer to chunks of 1000, 100, and 10, respectively. Mean refers to the mean of the means, and sigma refers to the standard deviation of the means. One can see how the standard deviation is increasing as chunks of data are decreasing.

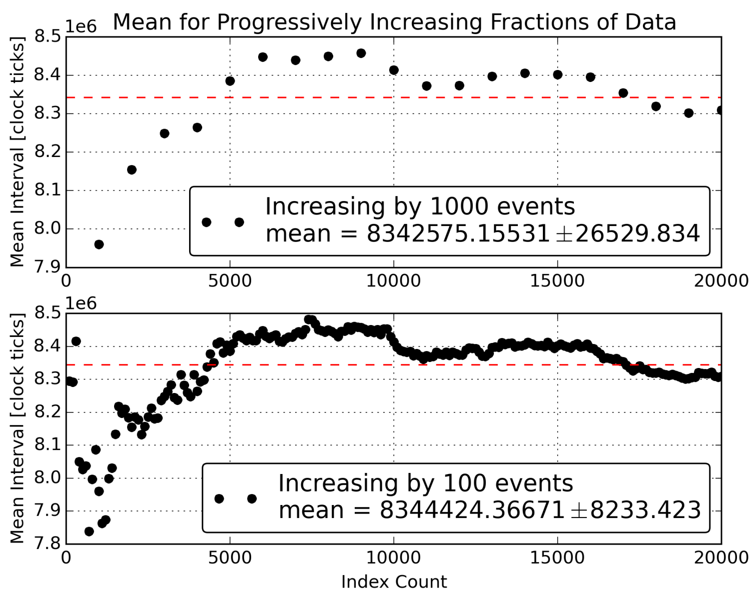

Figure 4: Computed mean time values for data sequences of different length. (a) The top and (b) the bottom plots increase the fractions of data by 1000 and by 100 events, respectively i.e. (a) takes the sample mean of the first 1000, then the first 2000, and so on.

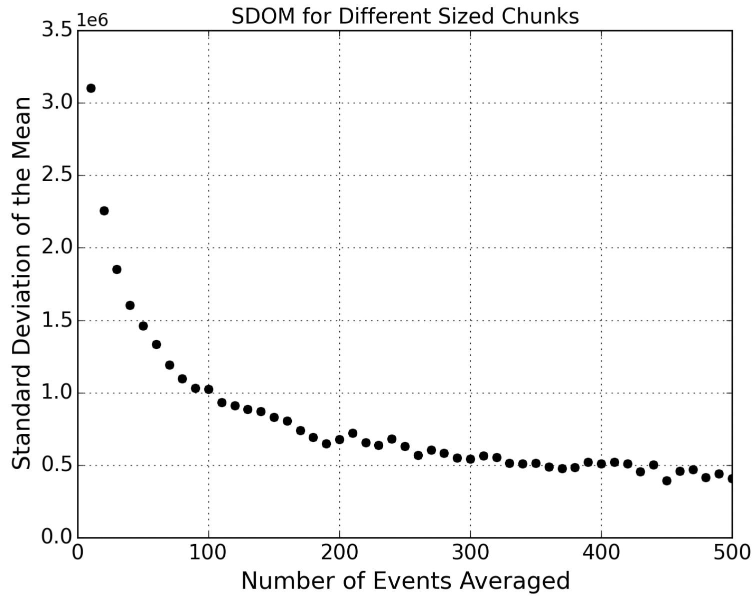

Figure 5: The standard deviation of the mean has a decreasing trend as the size of the data chunk averaged is increasing.

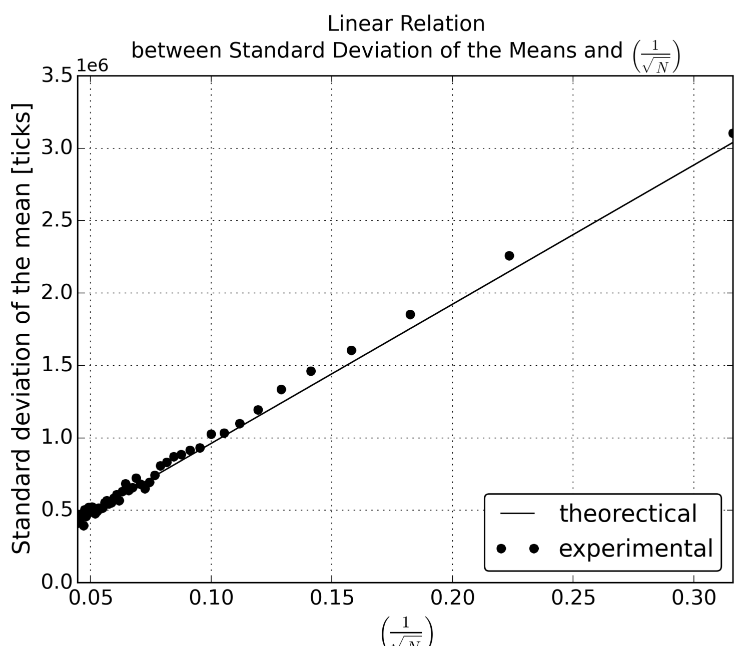

Figure 6: A linear relation is exhibited between the standard deviation of the mean and $\frac{1}{\sqrt{N}}$. The plotted line is $\frac{s}{\sqrt{N}}$, where s represents the sample standard deviation (the standard deviation of the entire data set).

Histograms

Histograms provide a visual interpretation of the distribution of data by dividing the data into bins and then plotting the frequency of events occurring in each bin. This form of categorizing the data can lead to a better explanation of what is happening with the data.

We divided the data (from Figure 2) of the clock intervals into 200 bins, as shown in Figure 7, to show the data more suggestively. Right away we notice that there is a huge spike at the very short intervals, and the histogram does not follow a normal (Gaussian) distribution, which is centered around the mean, but rather an asymmetric distribution, where there is a high frequency at shorter intervals. Figure 7(b) provides a closer look at the spike and further exposes a peak at around 186 ticks, or roughly 150 nanoseconds.

Afterpulsing

The reason for such a high spike at the shorter intervals is due to a consequence of the photomultiplier called "afterpulsing". 15% of genuine events are quickly followed by a second pulse caused by ions, not electrons. We can get rid of this effect by taking out and ignoring the events that occur with intervals shorter than 3000 ticks, as shown in Figure 8 on a (a) linear scale and (b) log scale. We might wonder if it is okay to reject those events as they may also sacrifice genuine data, but the amount of real data that may be omitted compared to the overall amount of data is only a small fraction.

For Poisson, we expect the distribution to be given by the exponential distribution

$$\rho(\Delta t) = \frac{1}{\tau} e^{-\frac{\Delta t}{\tau}} \tag{5}$$

The theoretical distribution plotted in Figure 8 is given by multiplying Equation 5 by the binwidth w and the sample size N of the data to get the theoretical frequency.

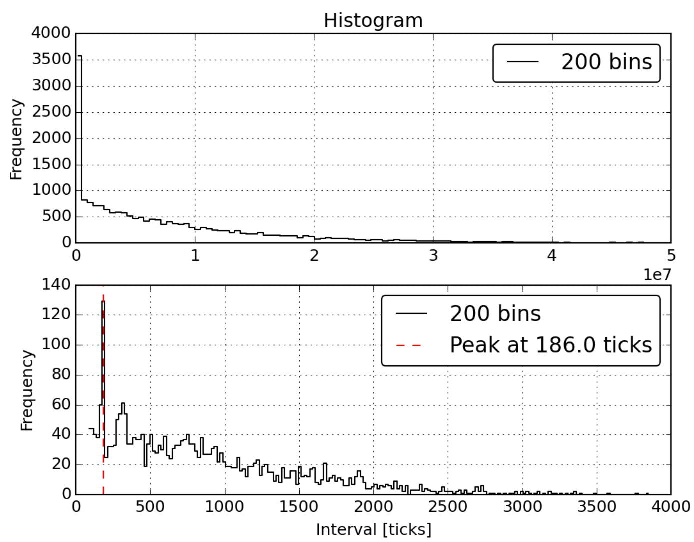

Figure 7: The purpose of both graphs is to visualize the effects of "afterpulsing". (a) The top graph includes all data points and (b) the bottom graph is zoomed in at the shorter intervals to emphasize the peak.

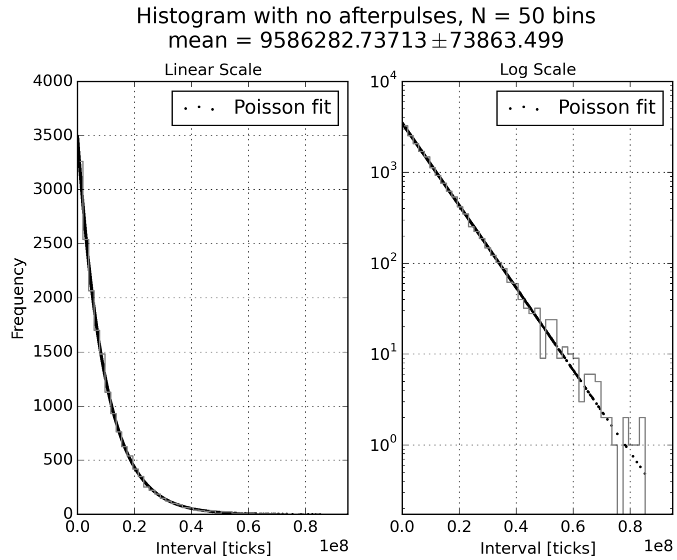

Figure 8: Both plots have a Poisson fit and do not include the afterpulse (intervals < 3000 ticks). (a) The top graph is plotted normally and (b) the bottom graph is plotted on a log scale.

Varying LED Brightness

Now, we decided to delve into how these statistics are affected when we vary the LED brightness. We took 6 additional data sets for a total of 7 data sets, including the one we have been working with. We accounted for the afterpulses for each data set and then found the sample mean and the sample standard deviation and plotted those two against each other, as in Figure 9, expecting them to fit a y=x fit since the sample mean and standard deviation are supposed to be equal for an exponential distribution.

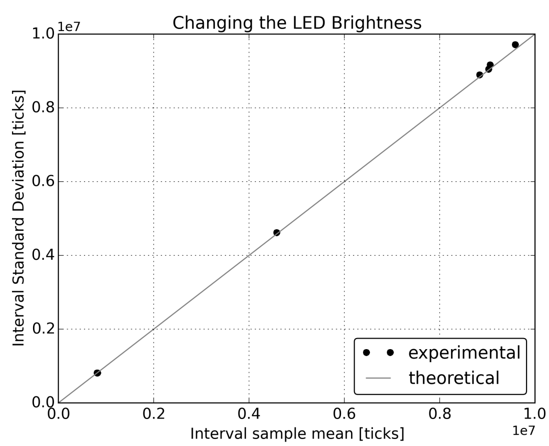

Figure 9: Varying LED Brightness exhibits the y=x relation between the sample mean and the sample standard deviation.

We can prove mathematically that the sample mean equals the standard deviation. If we look at the exponential distribution and look at the variance in the arrival time, we must calculate the expectation value of t and of t^2.

$$< t >= \int_0^\infty e^{-\frac{t}{\tau}} dt = \tau \tag{6}$$

$$< t^2 >= \int_0^\infty e^{-\frac{t}{\tau}} dt = 2\tau^2 \tag{7}$$

then we can calculate the variance

$$\sigma^2 =< t^2 > - < t >^2 = 2\tau^2 - \tau^2 = \tau^2 \tag{8}$$

which reduces to $\sigma = \tau$ where τ is the mean value of the dataset. The data, exhibited in Figure 9, clearly verifies this relation.

Binning

We will now bin the events into time samples of a constant, fixed width, where the counts from the photon detector are measured for each interval of time. In doing this, we are making sure that the data we have recorded can follow a Poisson distribution. We can confidently say that light follows the Poisson distribution once we ensure that each data point is (i) independent of each other, meaning that the occurrence of one data point does not increase or decrease the chance of another; (ii) the average (sample mean) can be calculated and; (iii) it is possible to count the number of events.

In Figure 10(a), we plotted the cumulative sum of the event intervals, which fortunately shows a nice linearly increasing trend free of any sudden bumps or changes, confirming that each point is independent of the other. However, this plot makes it seem like photon arrival is pretty constant, which is not the actual case. What was not so explicit in Figure 10(a), Figure 10(b) assures that the photon arrivals were random and that the randomness of photon arrival was consistent, in the sense that nothing strange was happening with the LED that would negatively affect the data or produce strange results.

In Figure 10(c) and Figure 11(a)(b)(c) the histograms are plotted along with the Poisson probability distribution given by the equation below,

$$p(x, \mu) = \frac{e^{-\mu}\mu^x}{x!} \tag{9}$$

where μ is the sample mean of the counts per bin.

Figure 11 proved to be quite interesting in exposing how the peak of the Poisson distribution moves more and more to the left as the width of the bins decreases since more bins are considered. With less bins, it is curious to note that the distribution seems Gaussian. Despite the fact that the peak moves as the bin width decreases, the unique relationship within a Poisson distribution between the mean and the variance (square of standard deviation of counts per bin (as shown below) stayed intact,

$$\sigma^2 = \mu \tag{10}$$

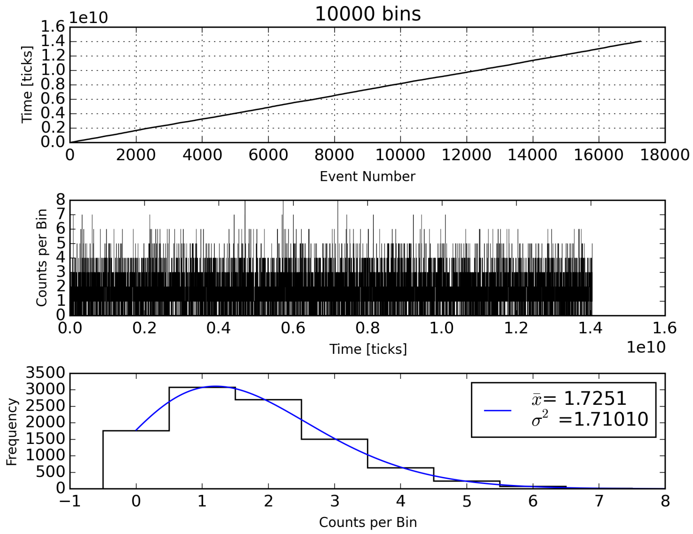

Figure 10: (a) Cumulative sum of the counts per bin. (b) Number of events in 10,000 evenly spaced time bins. (c) Histogram of events per bin. Binning the data helps show how lights follows a Poisson probability distribution. We can visually see the sample mean of counts per bin is equal to the variance of counts per bin.

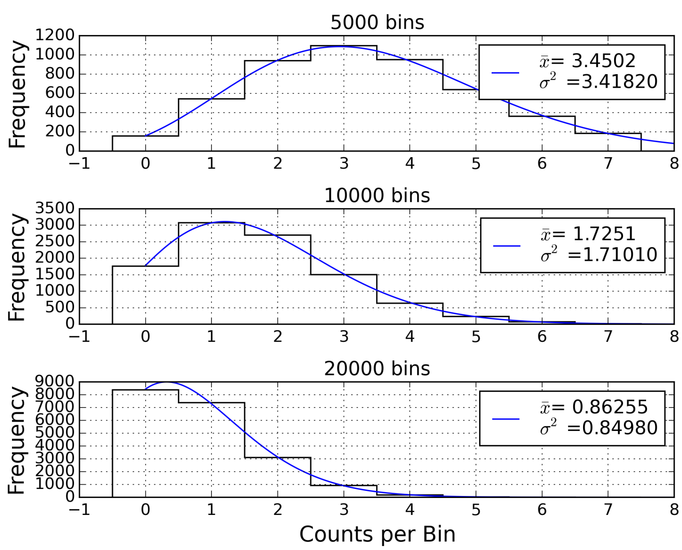

Figure 11: (a), (b), and (c) represent how different sized bins of 5000, 10,000, and 20,000, respectively, affect the Poisson probability distributions.

4 Conclusion

After completing this lab, I was able to see how light follows a Poisson distribution, entailing that the arrival of one photon does not increase or decrease the chances of another photon arrival. In addition, the Poisson distribution can predict the probability of photon detection. Exploring the statistics behind the data and doing the calculations for the mean, standard deviation, and the standard error (standard deviation of the mean), I was able to gain a better insight on how to interpret data and deduce how precise and accurate it may be. I had grasped the importance of the number of data points considered and how that number affects the sample mean and standard deviation, and in turn, how it affects the interpretation of data. I realized how histograms are an important plotting tool that can provide information more suggestively.