Table of Data:

Abstract:

Making use of the photoelectric effect through a photomultiplier tube (PMT) allows the investigation of photon statistics. In this experiment, two statistical models were compared to data taken from a PMT device with an attached counter. In addition, general methods of error reduction and their implementation in observation were explored. As might be expected, an increased number of samples per observational trial generally reduced the error in the measurement of the rate of photons striking the detector. It was also found that the mean value of the photon counts and the variance of the value have a linear relationship, a hallmark of Poisson statistics. As the length of the time samples increased, the Poisson distribution for the probability of counts ceases to be the superior model, as it begins to merge with the smooth Gaussian distribution.

Introduction:

This experiment was an exploration of the statistical behavior of photons. The setup was relatively simple – a PMT attached to a device the registers a count when struck by a pulse. The PMT makes use of the photoelectric effect, bouncing a single photon from the LED on dynodes. This process results in the photoelectric emission of multiple electrons from a single photon. When these electrons reach the detector, they register as a pulse, and a photon is recorded by the counter.

Using this photon counting device and some python programs, the data was shifted into probabilities – information about the number of times a set count number would appear during a sample of arbitrary length. These probabilities were compared to two probability distributions to find one that would accurately model the photon behavior. This modeling was then tested at extremes - for very high and low numbers of samples – and over multiple trials.

Observations & Data:

The primary acquisition device used in this experiment was the PMT and its attached USB-4301 counting device. The setup had only rudimentary controls – an overall power switch, a switch to turn the lamp on and off and an unlabeled brightness dial. An associated python script called PMT.py was provided to gather data from the photon counter. This program contained a function called photoncount that took a number of time samples (nsamp), and the length of each sample (tsamp). After processing information for a length of time equal to nsamp multiplied by tsamp, the function returned an array of numbers representing the number of photons counted during the time specified for each sample.

Observations were undertaken over the course of two weeks when the entire lab team could be present. Only one group member recorded data (see Table ~\ref{datatable}), but each individual analyzed the resulting output in its graphical form. Together, it was decided whether the data sets collected were in keeping with predictions or if there were explainable discrepancies.

Since data with relatively large gaps in between each session, efforts were made to ensure that conditions were similar. However, due to the fact that the brightness dial on the PMT was unmarked, it was difficult to record completely consistent data. Ideally the situation would have been such that only the parameters fed to PMT.py would have changed between each step.

Often, the PMT itself would experience issues. Even when the light was off, some count activity was expected, but for long stretches, the PMT would simply output an array of zeroes. When running a loop that gathered many data sets at once, the start of this behavior was not always obvious and had the potential to ruin the set. Another error with the machine or its partner software caused it to report counts in excess of several thousand, regardless of the length of the sample time. This was even more challenging to identify, as only a few counts were misreported, though the fix in this case was simply to remove outlying points.

As mentioned above, the counter reports photons even when the light is off in the PMT. There are a few reasons for this: first that the PMT is not perfectly insulated, and so some stray photons might have made it in from the outside. In addition, the PMT is at room temperature. For everyday materials, 273K is not hot enough to give an electron the energetic kick it needs to escape its bonds (E = frac{3}{2}kT where k is the Boltzmann constant and T is temperature). However, by virtue of the number of electrons and thermodynamic probabilities, some will escape and cause a false report of a count.

Data Reduction & Methods

Data was analyzed using the Python language, largely with original functions written for the task. Plotting the data at every step was done with the pyplot module, part of matplotlib.

The first part of the analysis done for this lab was mostly concerned with plotting data and developing intuition for expected outcomes. Plotting raw data was clearly uninformative (see Figure~\ref{fig:section4}), so a function was written to convert the array of counts into a histogram. It was simply a matter of finding how many times a given number of counts was recorded and plotting that information in a histogram. This proved to be much more relevant than the original plot of the number of counts versus time.

After using the histogram function to plot count frequency for varying parameters, the next step was to make some statistical predictions. This required writing two new functions – one for a Poission distribution, and another for a Gaussian distribution. In addition, it was necessary to calculate the mean and standard deviation of the data sets. The actual procedure and formulae are outlined below (see Section~\ref{sec:calculations}, but the functions needed were not so complex as to be challenging to model. However, due to the limitations of the written factorial function needed for the Poisson distribution, it was necessary to use the scipy.stats Poisson function on data sets with a high number of counts per sample.

Both of these distributions provide output as probability, so a final modification to the raw data was necessary. Rather than plot just the frequency of counts, as in Figure~\ref{fig:histograms}, each such frequency was divided by the total number of counts recorded to provide a probability.

The final matter of concern with regard to the methods used were possible sources of error (as outlined in Section~/ref{sec:observations}). Of particular concern were dark counts, the photon counts recorded when the LED in the PMT was turned off. After some measurements taken when the light in the room containing the PMT was both on an off, it was found that the counter was averaging 8.1 counts per millisecond $\pm something$.

Calculations & Modeling

The instructions for modeling the behavior of the photon counts were provided in the lab handout, but required some tweaking to achieve the desired results, particularly in the python implementation. One of the most significant challenges was comprehending the behavior of the probability functions and understanding each of the variables.

The first step towards plotting either of the probability distributions was calculating the mean and standard deviation of each data set (see Equations~\ref{eqn:mean},~\ref{eqn:standard deviation}).

With this information, and the histogram modified to display probabilities, it was possible to overplot the Poisson and Gaussian distributions (see Equations~\ref{eqn:poisson},~\ref{eqn:gaussian}).

It is interesting to note that the Poisson distribution (unlike the Gaussian distribution), does not take standard deviation into account. Instead, it bases the standard deviation on the mean ($\sigma$=$\sqrt{\mu}$). This relationship is later demonstrated in the photon data in Figure~\ref{fig:section7}. The other relevant feature of the Poisson distribution is its discrete nature. Since it depends on the factorial of x, it is only meaningful for integer values. This makes sense for a photon count distribution, where the number of counts per time sample is necessarily discrete in order to be physically sensible.

The Gaussian distribution is more widely applicable than the Poisson distribution in that it is continuous. However the question remained as to whether it was a good fit for the photon data collected over the course of this experiment.

Calculations using the above equations were done entirely through python.

Discussion

The first step of the investigation was to test the PMT and get a sense for reasonable output. To this end, data was recorded at the same tsamp and nsamp values with six repetitions (see Figure~\ref{fig:section5} for three of these repetitions). The shape of histogram varies considerably, despite having been given the same input parameters. This speaks the randomness of the process, especially at a low number of samples. In the adjacent set of plots, Figure~\ref{fig:section6}, the number of samples was increased with each iteration. Here, the shape of the histogram becomes more regular, peaking towards the mean with less fluctuation.

Once the standard data behavior was clear, the next task was to develop a fit to the data. To do this, the mean and standard deviation (Equation~\ref{eqn:standard deviation}) were calculated for six data sets with increasing tsamp values. It was assumed that since the lower limit on tsamp was 0.0001s, starting at 0.001s and increasing by amounts 0.001s or larger, the quantization of sample time would not effect the data.

The mean increased with the length of the sample, as would be expected (a longer time means more counts, after all), but so did the standard deviation. However, this isn’t as surprising as it initially seemed – a longer sample times allows for more fluctuation in count values. The variance, which is simply the square of the standard deviation, was then plotted against the mean count value, on both a linear and log scale (Figure~\ref{fig:section7}). The overplotted x=y fit makes it clear that the two variables are linearly related. This is what would be expected from a phenomenon that follows Poisson statistics. A linear relation between variance and mean value implies that standard deviation is proportional to the square root of the mean, a requirement to fit Poisson statistics.

As the mean increases, however, it is clear on the linear plot that the scatter away from the linear fit is also increasing (albeit on a small scale). This implies that the relationship between standard deviation and mean dictated by Poisson statistics might not hold as well for larger means. However, this possible departure from Poisson statistics is not completely determined just by observing Figure~\ref{fig:section7}. On the linear plot, the points representing small time scales are clustered near the origin, and it is difficult to discern how far they scatter from the fit line. The scatter at higher means is hardly present on the log scale plot. Clearly more exhaustive check of photon behavior is needed to posit that Poisson statistics do not hold for large means.

The claim was further investigated by comparing two sets of count data taken over 0.3s and 0.003s times at 1000 samples each. Each set was plotted as a histogram, with Gaussian and Poisson statistical predictions plotted over top based on the mean and standard deviation of the set (see Figure~\ref{fig:section8}).

A few qualitative effects are immediately obvious. As was previously established, a higher mean value means a higher count value and standard deviation. In keeping with this, Figure~\ref{fig:section8s} has a wider and more symmetric distribution than its shorter timescale counterpart. Of course, Figure~\ref{fig:section8ms} lacks symmetry partly because its mean value is so close to zero, and obviously a count of less than zero photons points to a device malfunction, not a physical phenomenon.

More striking is the difference between the fits for each plot. In the first, it is clear that the Poisson prediction much better describes the data, fitting far closer to the measured data than the highly peaked Gaussian curve. However, in the second plot, the two predictions of probability are nearly the same curve, and both a reasonable description of the data. In the case of the longer sample time, the count probability fluctuates more extremely, but still seems to follow a standard normal distribution. It appears that a Gaussian probability distribution is a good approximation of both the data and a Poisson distribution for high enough means.

However, photon statistics at low mean values are clearly well described by the Poisson distribution. The next matter of concern was how to refine this mean value to greater precision.

Obviously, there are systematic errors associated with the PMT, but the aim was to minimize the statistical error of the measurements. To this end, data was taken with a 0.01s sample time and a number of samples that increased by a power of two with each repetition. For each number of samples, fifteen data sets were recorded. A mean value and standard deviation was calculated for each set, and from these fifteen values, a mean of the means (MOM) and a standard deviation of the means (SDOM) (Equation~\ref{eqn:sdom}) was calculated. Here, $\sigma\_\mu$ is the standard deviation in the mean and N is the number of samples.

It was expected the MOM values would approach a relatively constant number and the SDOM would decrease as nsamp increased. Fluctuations were expected for low nsamp values, but since the photon measurement has been shown to be a mostly statistical process, more measurements should increase the precision of the knowledge. However, this prediction does not entirely match the recorded data. Figure~\ref{fig:section9} shows the results of plotting the MOM and SDOM values against the number of samples used in the calculation. Contrary to expectations, the MOM increases with nsamp, though not in any reliable way. Since the data taken to form this plot was taken in succession with no interference with PMT controls, it is likely that this effect is due to the systematic errors (or dark counts, as they have been referred to previously).

The SDOM values do match expectations, fitting remarkably well to predicted values. These predictions were based on the accuracy of Poisson statistics in fitting earlier data (Figure~\ref{fig:section8}). In Poisson statistics, the standard deviation is simply the root of the mean. For each of the fifteen sets for a given nsamp value, the square root of the mean was calculated according to $\sigma = \sqrt{\mu}$. From these predicted values, the SDOM was calculated with Equation~\ref{eqn:sdom} as with the raw data. The SDOM values decrease sharply as the number of samples increases. This is quite logical, as the SDOM is inversely proportional to the number of samples. However, the increase in precision begins to slow down even as nsamp continues to increase, because the proportionality is limited to the square root of nsamp. After a certain nsamp, it is clear that the advantage of increasing the number of samples no longer results in a significant reduction in standard deviation.