Activity 2
Counting Statistics

Dr. James E. Parks, Dr. Marianne Breinig, and Dr. Christine Cheney
University of Tennessee - Knoxville

September 8, 2023

1 Goals of this Activity

The objectives of this experiment are:

1. to study the Poisson probability distribution as it is applied to counting statistics;

2. to study the Gaussian probability distribution as a special case of the Poisson distribution applied to measurements with large mean value;

3. to learn some statistical analysis techniques using Excel and Python to analyze, graph, and compare results.

2 Safety

You will be participating in academic laboratory exercises which utilize ionizing radiation sources. The radioactive sources used in this laboratory contain low levels of radioactivity and are considered by the State of Tennessee to be exempt from radioactive materials licensing. You will be exposed to low levels of ionizing radiation when utilizing these items in your academic laboratory exercises.

It has been estimated by UT's Radiation Safety unit that the maximum radiation exposure you could receive for participating in these exercises over the course of the semester is 2 milliRoentgen. A milliRoentgen is a unit of measurement for radiation exposure. To compare to natural background radiation that we are exposed to each day from rocks in the soil and cosmic radiation, a person is exposed to approximately 1 milliRoentgen per day. The amount of radiation exposure you will be exposed to from participating in these lab exercises equates to two days of natural background radiation exposure.

3 Poisson Distribution

In many types of measurements, our data arises by counting something. You may count the number of dogs at the park or the number of gamma rays detected in a given amount of time t. There is uncertainty in how many counts we actually measure. The number of gamma rays we count is uncertain due to quantum mechanics. The number of dogs we count is uncertain because some dogs may leave or come to the park while we are counting.

In this laboratory you will study some aspects of the random nature of radioactive decay. You will verify that if the count rate is low, the probability of recording a particular number of counts is given by the Poisson distribution, and that if the count rate is high, the probability of recording a particular number of counts is given by the Gaussian distribution.

There are many types of events that occur at a well defined average rate, but for which we cannot predict the outcome of any particular measurement. For example, a light bulb has an average lifetime. But we cannot predict how long a particular light bulb will last. This is a characteristic of a random process. A random variable is a function that associates a unique numerical value with every outcome of an experiment. The value of the random variable will vary from trial to trial as the experiment is repeated. For the light bulb the experiment is measuring its lifetime, and the random variable is the number of days it produces light before it burns out.

The average number of rainy days in May in Nashville is 11. We cannot predict how many rainy days Nashville will have next year in May and on which days it will rain. If we record the number of rainy days next May, the new number will most likely differ from that of the preceding May. The number we will record cannot be predicted using the knowledge of a preceding observation nor can it be used to predict the following observation.

Radioactive decay is a random process. We cannot predict exactly when a certain unstable nucleus will decay, we can only predict the probability that the nucleus will decay in a certain time interval. Similarly, we cannot predict exactly how many decays will take place in a particular radioactive sample in a particular time interval, but only the average number of decays. If we actually count the number of decays in several time intervals, we get a different number of counts each time, but they have a more or less definite mean.

Assume we count the number of times n a given random event occurs in a certain time interval. Statistical theory predicts that if we repeat our experiment m times, we will find the numbers $n_1, n_2, n_3, ..., n_m$. We can calculate the average number of counts as

$$\bar{n} = \sum_{i=1}^{m} \frac{n_i}{m} \tag{1}$$

The actual number of counts will be distributed around this average value. Numbers close to the average will be recorded frequently, numbers very different from the average will be recorded infrequently. If we want to know how the actual number of events counted in a time interval Δt is distributed around the average number of events counted in Δt, we can make histogram of the number of times a certain number n appears.

For a large number of rare events we find that the probability of recording a particular number n in a given time interval is given by the Poisson distribution:

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{2}$$

where $P(n)$ is the normalized probability, i.e.

$$\sum_{n=0}^{N} P(n) = 1 \tag{3}$$

It can be shown that the mean of the number of occurrences, n, is λ. For a great video on the derivation of $n = \lambda$, see https://www.youtube.com/watch?v=e4FZ110Cceo. By replacing λ with $\bar{n}$, we get

$$P(n) = \frac{\bar{n}^n e^{-\bar{n}}}{n!} \tag{4}$$

The variance of the Poisson distribution can be shown to be λ, giving us $\sigma = \sqrt{\lambda} = \bar{n}$.

The Poisson distribution describes a wide range of phenomena in the sciences. It describes the probabilities of random occurrences and it can be applied to "intervals" on the space or time axes. For small n values, the distribution is skewed to the right, but as n increases, the distribution becomes symmetric and approaches the Gaussian (or normal) distribution. Remember these events are independent of each other, meaning one event does not influence the probability of another event.

4 Procedure for Poisson Distribution

Open the software STXx64 on the computer and turn on the counter. Set the voltage to 800 V with 1000 runs at 1 s time intervals. Place the Sr 90 source in the bottom slot of the Geiger counter stand (farthest away from the Geiger counter). Collect data for 1000 s. Save the data as a .TSV file.

Open the file in Excel. You will have to make sure that you search for all files when you try to open the file in Excel since it is not an Excel file. Delete the header and the High Voltage, Elapsed Time, and Date/Time columns. Save the file as an Excel file. Add a column for Average, n, Frequency Distribution, and Poisson Distribution. In the Average column, find the average number of counts that you counted during the 1000 s. In the n column, insert numbers 0, 1, 2, 3,..., 100. Select the corresponding cells in the Frequency column, go to Formulas, and hit Insert Function. Select the Frequency function. Select your counts data for the data array enrty and the n column for the bins array entry. **While holding control and shift, click OK**. This will bin your data based on the frequency of counts that occurred. For example, in the corresponding box for n = 7, the number of times there were 7 counts in the 1000 s interval will appear. You can plot the data as a scatter plot or as a histogram plot (try both!). In the Poisson Distribution, use the Poisson distribution equation, the n column, and the average number of counts you collected to get a Poisson distribution. Remember to multiply the equation by 1000 since Pn is a normalized distribution. Compare it to your data by overlaying it on the same graph as the frequency plot. Is it different? What happens if you truncate your data such that you collected data for only 100 seconds? Does your distribution change? Does the average number of counts change?

Copy your original .tsv file and name the copied file counts.tsv. Run the Poisson Distribution Python code. One graph will plot your data with the Poisson distribution overlayed on it. The other plot will truncate the data to 100 s and plot the frequency and Poisson distribution together. What is the average number of counts for your distribution? If you binned your data into 2 second intervals, how would your data change? Does your histogram change if you truncate your data to 100 runs?

Now add an aluminum attenuator between the Sr 90 source and the Geiger counter. This will reduce the number of counts/second. How do your results change? Is your distribution skewed to the right?

5 Gaussian Distribution

For a large number of frequent events, Poisson's equation is very difficult to deal with because of the large values of $n!$. We find that the probability of recording a particular number n is better characterized by the Gaussian distribution. Recall that the Gaussian distribution is given by

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \tag{5}$$

where $\bar{x}$ is the mean value about which the distribution is centered, σ is the standard deviation from the mean, and $P(x)$ is the normalized probability of measuring a value x. $P(x_i)$ is a continuous function and must be multiplied by an increment in x, Δx, in order for it to represent a finite number. Therefore, the mean value is given by

$$\bar{x} = \sum_{i=1}^{N} x_i P(x_i) \Delta x \tag{6}$$

Recall that σ, the standard deviation, is the square root of the variance, σ^2, given by

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{n-1} \tag{7}$$

Since the distribution is normalized,

$$\sum_{i=1}^{N} P(x_i) \Delta x = 1 \tag{8}$$

For integers, the smallest increment in x, $\Delta x = \Delta n$, is 1. The continuous function $P(x)$ is changed to an integer form $P(n)$ such that

$$P(n)\Delta n = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}} \Delta n \tag{9}$$

and

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(n-\bar{n})^2}{2\sigma^2}} \tag{10}$$

since $\Delta n = 1$.

The standard deviation σ is a measure of the width of the distribution. Approximately $\frac{1}{3}$ of the counts will lie outside the interval $n - \sigma$ to $n + \sigma$. The standard deviation can be approximated as

$$\sigma = \sqrt{\bar{n}} \tag{11}$$

where this approximation is improved for increasing n. This then gives us

$$P(n) = \frac{1}{\sqrt{2\pi\bar{n}}} e^{-\frac{(n-\bar{n})^2}{2\bar{n}}} \tag{12}$$

6 Procedure for Gaussian Distribution

Open the software STXx64 on the computer and turn on the counter. Set the voltage to 800 V with 2500 runs at 1 s time intervals. Place the Cs 137 source in one of the slots of the Geiger counter stand. Collect data for 2500 s. Save the data as a .TSV file.

Open the file in Excel. Delete the header and the High Voltage, Elapsed Time, and Date/Time columns. Save the file as an Excel file. Add a column for Average, n, Frequency Distribution, and Gaussian Distribution. In the Average column, find the average number of counts that you counted during the 2500 s. In a cell in the Average column, find the standard deviation, of the number of counts. In the n column, insert numbers 0, 1, 2, 3,..., 100. Select the corresponding cells in the Frequency column, go to Formulas, and hit Insert Function. Select the Frequency function. Select your counts data for the data array entry and the n column for the bins array entry. **While holding control and shift, click OK**. This will again bin your data based on the frequency of counts that occurred. You can plot the data as a scatter plot or as a histogram plot. In the Gaussian Distribution, use the Gaussian distribution equation, the n column, the average number of counts you collected, and the standard deviation of your counts. Remember to multiply it by 2500 since Pn is a normalized distribution. Compare it to your data. Is it different? What happens if you truncate your data such that you collected data for only 100 seconds? Does your distribution change? Does the average number of counts change? Approximate the full-width half-maximum, , of your distribution. Is it related to σ by $\Gamma = 2\sqrt{2\ln 2}\sigma$? This relationship can be easily shown (http://hyperphysics.phy-astr.gsu.edu/hbase/Math/gaufcn2.html).

Copy your original .tsv file and name the copied file counts.tsv or something else. If you call it something else, just edit the code for to account for the different name. Run the Gaussian Distribution Python code. What is the average number of counts for your distribution? If you binned your data into 2 second intervals, how would your data change? Does your histogram change if you truncate your data to 100 runs? You learned about standard deviation in the previous lab. What is standard error, and how can you decrease your standard error during this lab?