Poisson Statistics

MIT Department of Physics
(Dated: August 25, 2013)

In this experiment, you will explore the statistics of random, independent events in physical measurements. The random events used in this study will be pulses from a scintillation detector exposed to gamma rays from a radioactive source. The measurement will be compared to Poisson or Gaussian distributions which often govern the spread of physical data. A computer simulation using Monte-Carlo techniques will provide comparison for different numbers of events and give a demonstration of "normal" statistical fluctuations. This comparison technique is widely used in modern physics.

PREPARATORY QUESTIONS

Please visit the Poisson Statistics chapter on the 8.13x website at `mitx.mit.edu` to review the background material for this experiment. Answer all questions found in the chapter. Work out the solutions in your laboratory notebook; submit your answers on the web site.

I. THEORY OF POISSON STATISTICS

A sequence of independent random events is one in which the occurrence of any event has no effect on the occurrence of any other. One example is simple radioactive decay such as the emission of 663 keV photons by a sample of ^{137}Cs. In contrast, the fissions of nuclei in a critical mass of ^{235}U are correlated events in a "chain reaction" in which the outcome of each event, the number of neutrons released, affects the outcome of subsequent events.

A continuous random process is said to be "steady state with mean rate R" if

$$\lim_{T \to \infty} \left(\frac{x}{T} \right) = R \tag{1}$$

where x is the number of events accumulated in time T.

How can one judge whether a certain process does, indeed, have a rate that is steady on time scales of the experiment itself? The only way is to make repeated measurements of the number of counts x_i in time intervals t_i and determine whether there is a trend in the successive values of x_i/t_i. Since these ratios are certain to fluctuate, the question arises as to whether the observed fluctuations are within reasonable bounds for a fixed rate. Clearly, one needs to know the probability distribution of the numbers of counts x in a fixed interval of time T if the process does indeed have a steady rate R. That distribution is known as the Poisson distribution and is defined by the equation

$$P(x; \mu) = \frac{\mu^x e^{-\mu}}{x!}, \tag{2}$$

which is the probability of recording x counts (always an integer) when $\mu = RT$ (generally not an integer) is the expected number. It is easy to show that the standard deviation of the Poisson distribution is simply $\sqrt{\mu}$, that is, the square root of the mean. Derivations of the Poisson distribution and its standard deviation are given in References [1, 2]. You should study and understand binomial, Poisson, and Gaussian distributions and their limits of application. Keep in mind that although Eqn. (2) is a function of the variable μ (or R and T), it is a probability distribution only for the random variable x.

II. EXPERIMENT

In the first part of this experiment you will set up a scintillation counter, expose it to gamma rays from a radioactive source (and ubiquitous cosmic rays)[1], and record the frequency distribution of the numbers of counts in equal intervals of time. This will be repeated for four situations with widely different mean count rates, approximately 1, 5, 10, and 100 counts per second. You only know the "real" average rate at the end, but you should aim for these values within 25%. The experimental distributions and their standard deviations will be compared with the theoretical distributions and their standard deviations.

Later, you will generate Poisson distributions with a Monte Carlo simulation and compare them with the distributions produced by nature in your counting measurements.

II.1. Setup to Measure Poisson Statistics

Set up the scintillation counter as shown in Figure 1. Expose the detector to gamma rays from a ^{137}Cs, ^{60}Co, or ^{22}Na laboratory calibration source (a $1/2'' \times 5''$ plastic rod with the source embedded in the colored end). The voltage applied to the photomultiplier should be

―――――

[1] Note: even without external sources of radiation, scintillation counters of the kind used in this experiment will produce random small amplitude pulses at a steady rate due to thermal fluctuation in the photomultiplier tube.
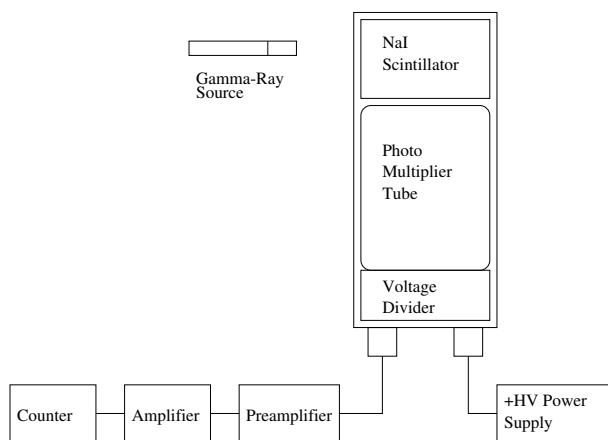
FIG. 1. The setup for measuring the number of counts from a random process (radioactive decay) in a given time interval. An oscilloscope (not shown) is used to monitor the proper functioning of the system.

about[2] +1000 V. Feed the output of the photomultiplier to the `INPUT` connector on a charge-sensitive preamplifier. (Some photomultiplier assemblies are provided with a built-in preamplifier.) Use the oscilloscope to view the output of the preamplifier and draw it in your lab notebook. Note especially the rise and decay time of the signal as well as the peak amplitude and polarity.

Next, connect the output of the preamplifier to the `INPUT` connector on the back or front of the amplifier. The amplified signal should be taken from the `UNIPOLAR OUT` connector on the front of the amplifier and examined on the oscilloscope (sweep speed 1 μsec/division, vertical amplitude 1 volt/division, or whatever settings allow you to best observe the signal) to confirm the proper performance of the measurement chain. Among the pulses of various amplitude that you see on the scope, identify those that correspond to the signals of interest, *i.e.* the check source gamma rays. Adjust the gain of the amplifier such that these signal pulses are about +5 V amplitude. If you trigger the scope on the rising edge of the pulses and set the trigger level to about +3 V, then you should see a signal which starts on the left-hand side of the oscilloscope display, crosses "time zero" at about 3 V, rises to a maximum of about 5 V, goes slightly negative, and finally levels off at zero. Again, note the pulse properties on the oscilloscope and record them in your notebook. Ask for assistance on this step if you are unfamiliar with the operation of an oscil-

loscope.

Note: Throughout Junior Lab, you should pay close attention to the polarities of applied and detected voltages. Incorrectly setting the polarity on an oscilloscope trigger can be very frustrating!

Feed the amplifier output into the `POS IN(A)` connector on the scaler. ("Scaler" is another name for a "counter".) Start with the scaler's discriminator set at its lowest value, usually about 0.1 V. Set the scaler to repeatedly acquire for about 5 seconds, display the result, and then start again. Observe how increasing the discriminator threshold decreases the count rate. If you set the discriminator on the counter to about 3 V (*i.e.* the same as the trigger level on the oscilloscope), then there should be an approximate one-to-one correspondence between pulses counted and pulses displayed on the oscilloscope.

II.2.   Procedure

You can control the counting rate by adjusting the distance of the source from the scintillator, by varying the high voltage supplied to the photomultiplier, by varying the gain of the amplifier, or by changing the threshold level of the discriminator[3]. Arrange the experimental conditions to yield four different mean count rates of approximately 1 sec^{-1}, 5 sec^{-1}, 10 sec^{-1}, and 100 sec^{-1}.

Record your instrumental settings and prepare tables in your lab notebook for recording the count data in a neat and compact form[4].

At each of these four approximate rates, record the counts for 100 repeated one-second intervals directly into your lab notebook. Then, for each of these four approximate rates, also record the number of events in one long run of 100 consecutive seconds.

II.3.   Analysis

The following analysis requires the use of repetitive arithmetic on the collected data set. You could use either MATLAB or any other preferred tool.

a) For each of the four runs, calculate and plot the cumulative average, $r_c(j)$, of the rate as a function of the sequence number, j, of the count. "Cumulative

[2] As with all suggested apparatus settings in Junior Lab lab manuals, the given value will be near a value which gives good signal, but you may find that a different value works better. In general, do not use a value simply because the lab manual say so. Instead, try to understand what experimental condition would cause one to prefer one value over another, and then optimize your choice of settings for that condition.

[3] Incidentally, even without a check source nearby, you should see signals due to cosmic rays at the rate of about 1 cm^{-2} min^{-1}.

[4] Whenever experimental conditions feasibly allow, it is desirable to use your own mind and hands to record data, rather than an automated data logger, as this process more readily engages the senses and other mental pathways, bringing about a more active awareness and understanding of the experiment.

average" means the quantity

$$r_c(j) = \frac{\sum_{i=1}^{i=j} x_i}{\sum_{i=1}^{i=j} t_i}. \qquad (3)$$

where x_i is the number of counts detected in time t_i. For a process which is truly steady with mean rate μ, $r_c(j)$ should converge to μ in the asymptotic limit. Include error bars to demonstrate convergence.

b) Calculate the mean and standard deviation of each of the four 100-trial distributions, and compare the results with the mean and standard deviation of the corresponding 100-second run. Are they consistent? To within how many standard deviations?

c) Using the mean rate just determined, calculate the predicted Poisson frequency distribution (renormalized by multiplying by the total number of readings). Make a plot with "frequency bins" (counts per 1-second acquisition) on the horizontal axis and "number of events" (frequency of occurrence) for each bin on the vertical axis, as in Figure 2. Plot your data and calculated distribution on the same axes for ease of comparison. Be sure to include error bars on the data points.

d) For the observed distribution with the lowest mean rate (~ 1 sec^{-1}), take the highest deviation from that mean and test whether you might be justified in concluding that the counter was malfunctioning. Remember that there were 100 opportunities for such a deviation to occur.

e) For large values of μ, you can use the Gaussian approximation to the Poisson formula as given by the relation

$$\lim_{\mu \to \infty} p(x; \mu) = \frac{1}{\sqrt{2\pi\mu}} e^{-(x-\mu)^2/2\mu}. \qquad (4)$$

Qualitatively, compare the Poisson and Gaussian distributions for μ values equal to the averages computed from both the $\mu \approx 5$ and $\mu \approx 100$ data sets. Make the comparison quantitative by computing χ^2 values between the theoretical distributions and your data. Which distribution (Poisson or Gaussian) is preferred at each mean rate?

II.4. Simulating Poisson Statistics Using Matlab

You can run the simulations on your own copy of MATLAB (you can download the scripts from the experiment website) or the Athena version.

To access the scripts on Athena, add the Junior Lab locker by typing
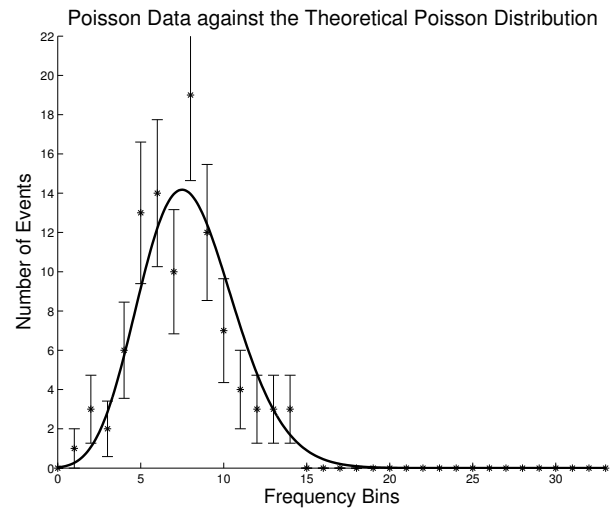
```
%> setup 8.13
```


FIG. 2. A frequency distribution of observed numbers of counts. The renormalized Poisson distribution for the observed mean value is also plotted.

Then, within MATLAB type

```
>> addpath /mit/8.13/matlab
```

The two MATLAB scripts, `poisson.m` and `poissonsim.m`, can also be downloaded off the Junior Lab website and added into MATLAB's working directory.

For information on how to use either one, type

```
>> help poissonsim
```

from within MATLAB. For example, type

```
>> poissonsim(3,20)
```

which will generate a 20 sample population with a mean of 3. The solid curve represents the theoretical Poisson distribution while the red dots represent the simulated sample population. The `poissonsim` function will output two vectors: the frequency of each rate (*i.e.* the counts in each bin shown in the graph) and the count rates for all of the trials.

1. Generate 1000-trial distributions for your measured mean rates ($\approx$ 1, 5, 10, 100).

2. Next, generate ten 100-trial distributions for each of the four mean counts you obtained in the experimental section using the scintillation counters. Record the mean values and standard deviations for each set of 10 distributions.

3. Compare the Monte Carlo-generated Poisson distributions with the experimental ones you obtained with the scintillation counter. The mean of the standard deviations should converge (within some statistical error) to the square root of the mean that is input to the Poisson generator.

4. Determine the error on μ and σ from the scatter of the ten distributions.

III. STATISTICAL EXERCISE

The following questions are a subset of the full analysis discussed in Section II.3.

1. For the 100 one-second measurements at ∼5 Hz, evaluate the average and the variance and compare the results with the mean and standard deviation of the long 100 second run. Are they consistent? To within how many standard deviations?

2. Plot the distributions of your measured data sets and compare to Poisson distributions of your experimental means. Are the variances in agreement with the theoretical expectation?

3. Fit the ∼5 Hz data set to both Gaussian and poisson functions. Which is preferred? Compare the χ^2 values for your fits.

IV. SUGGESTED THEORETICAL TOPICS FOR ORAL EXAM

1. The Poisson distribution, its origin from the binomial distribution and the limits of its application.

2. The Gaussian approximation to the Poisson distribution $P(x; \mu)$ for $x \gg 10$.

3. The differential distribution in the time lag between successive random pulses that occur at a fixed average rate.

---

[1] P. Bevington and D. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed. (McGraw-Hill, 2003).
[2] A. Melissinos and J. Napolitano, *Experiments in Modern Physics*, 3rd ed. (Academic Press, 2003).

Appendix A: Monte Carlo Generation of a Random Variable

Suppose we have a source of random numbers with a uniform distribution from 0 to 1. If we represent the uniform distribution by $q(y)$, so that $q(y')dy'$ is the probability that the random number y lies between y' and $y' + dy'$, then obviously $q(y') = 1$. The problem is how to convert a given random number y from this uniform distribution into a random variable x with a specified distribution $p(x')$ such that $p(x')dx'$ is the probability that the variable x will turn up with a value in the infinitesimal interval between x' and $x' + dx'$. We must find a relation between the distributions of y and x such that $p(x')dx' = q(y')dy' = dy'$.

To do this we compute, analytically if possible but otherwise numerically, the integral

$$P(x) = \int_{-\infty}^{x} p(x')dx', \qquad (A1)$$

which is the probability that the random variable will turn up with a value in the interval between $-\infty$ and x. From this definition is follows that

$$\lim_{x \to \infty} P(x) = 1, \qquad (A2)$$

since the probability that the random variable will turn up with some value is unity. Given the random number y, we set $y = P(x)$ and solve for x. To find the distribution of the resulting value of x we differentiate this expression,

using the rule for differentiating a definite integral with respect to its upper limit, and obtain $dy = dP(x) = p(x)dx$. Thus the distribution of x selected in this way is identical to the one specified.

Figure 3 shows how this works graphically. From the figure it is evident that a horizontal line at a random position y_i on the y-axis is more likely to intersect the $P(x)$ function where it is steeper than elsewhere; *i.e.*, where the differential probability is larger than elsewhere.
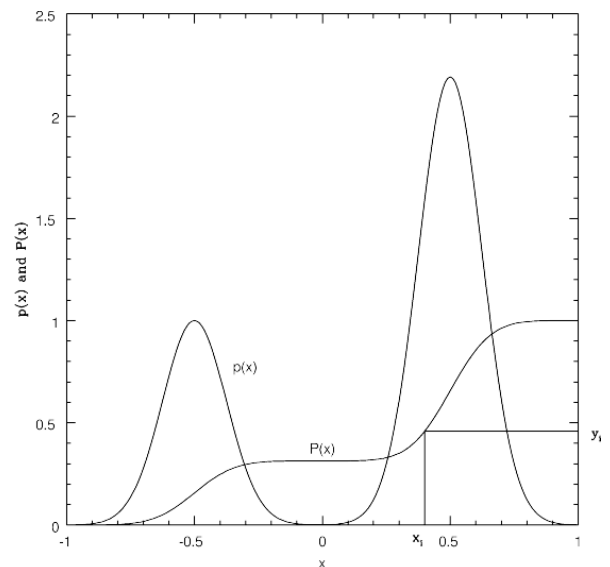

FIG. 3. Illustration of Monte Carlo selection of a random variable with a specified differential probability distribution is shown above. y_i is a random number between 0 and 1. x_i is the value of the variable for which the integral probability distribution equals y_i.

The Poisson probability, being a discontinuous func-

tion, is handled in a similar way, but with a summation rather than an integral. Given a value of $\mu < 88$ and a random number y, the Poisson simulator finds the smallest value of x for which $P(x; \mu) > y$, where P is defined by the formula

$$P(x; \mu) = \sum_{x'=0}^{x'=x} p(x'; \mu), \qquad \text{(A3)}$$

in which $p(x'; \mu)$ is the Poisson probability specified above. That value of x is the desired Poisson variate.

Inspect the MATLAB .m files where the algorithm is described in more detail.