Poisson Statistics

MIT Department of Physics

In this experiment you will explore the statistics of random events both by physical measurements and by computer simulations. The random events used in this study will be pulses from a scintillation detector exposed to gamma rays from a radioactive source. The techniques of this experiment are important for two reasons. Complicated theoretical predictions often only can be calculated on computers, so comparing experimental measurements to a computer simulation is essential for comparing theory with experiment. The ability to measure experimentally properties of statistical distributions is necessary in almost any test of modern physics.

PREPARATORY PROBLEMS

1. Describe how a scintillation counter works, starting from the entrance of an energetic charged particle or photon, and ending with an electrical pulse at the output of the photomultiplier.

2. Suppose the mean counting rate of a certain detector of random events is 3 counts per second. What is the probability of obtaining zero counts in a one-second counting interval? What is the most likely interval between successive pulses?

3. Given the formula for the Poisson distribution (Equation 2), prove each of the following:

   - $< x >= \mu$
   - $< x^2 >= \mu(\mu + 1)$
   - $< (x - \mu)^2 >= \mu$

   where $< x >$ signifies the mean value of x.

4. Plot the frequency distribution of counts when the average counts per interval is 1.

5. Puzzler: Some experiments have painfully slow counting rates that try the experimenter's soul and make him or her question the performance of even the most reliable equipment. Suppose you are running an experiment that yields no counts in 23 hours and two counts in the 24th hour. Give a quantitative answer to the question, "What is the likelihood that the equipment is malfunctioning?"

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 μ" if

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

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 in a fixed interval of time if the process does indeed have a steady rate. That distribution is known as the Poisson distribution and is defined by the equation:

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

which is the probability of recording n counts (always an integer) when μ (generally not an integer) is the expected number, the mean rate times the counting time interval. It is simple 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 by Bevington & Robinson and Melissinos.

EXPERIMENTS

In the first part of this experiment you will set up a scintillation counter, expose it to gamma rays from a radioactive source, 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. The experimental distributions and their standard deviations will be compared with the theoretical distributions and their standard deviations.

Later, you will generate Poisson distributions by Monte Carlo simulations on a Junior Lab PC and will also compare them with the ones produced by nature in your counting measurements.

Measure Poisson Statistics

Set up the scintillation counter as shown in Figure 1. Expose the detector to the gamma rays from a ^{137}Cs or ^{60}Co 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 $\leq$ 1050 volts. The output of the photomultiplier is fed to the "INPUT" connector on charge-sensitive preamplifier. Use the oscilloscope to record the voltage waveform taken from 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.

The output of the preamplifier is then connected to the "INPUT" connector on the back or front of the EG&G 575A amplifier. The amplified signal should be taken from the "UNI OUT" connector on the front of the amplifier, and fed to the "POS IN(A)" connector on the EG&G 776 Counter & Timer. Adjust the settings as follows:

EG&G 113	INPUT	0 pF
EG&G 575A	FINE GAIN	2.5
	COARSE GAIN	2
	INPUT	POS
EG&G 776	DISPLAY	COUNTS(A)
	PRESET(B)	m=3
		n=1
	DISCRIMINATOR(A)	0.1 V

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!!!.

Examine the output of the amplifier on the oscilloscope (sweep speed $\sim$ 1μsec/cm, vertical amplitude $\sim$ 1 volt/cm) to confirm the proper performance of the measurement chain. Adjust the gain of the amplifier to produce signal pulses of $\sim +3$ volts. If you trigger the scope on the rising edge of the pulses and set the trigger level to $\sim +3$ volts, you should see a signal which starts on the left-hand side of the scope display at $\sim$ 3 volts, rises to a maximum of about $\sim$ 5 volts, goes negative and finally levels off at zero. If you also set the discriminator on the counter to 3 volts, there should be an approximate one-to-one correspondence between pulses counted and pulses displayed. Ask for assistance on this step if you are unfamiliar with the operation of an oscilloscope.

Incidently, even without a 'check source' nearby, you should see signals due to "cosmic-rays" at the rate of

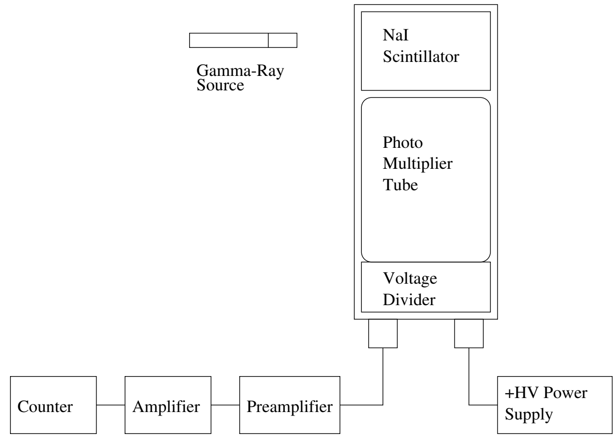

FIG. 1: The setup for measuring the number of counts from a random process (radioactive decay) in a given time interval. An oscilloscope is used to monitor the proper functioning of the system.

$\approx 1 cm^{-2} min^{-1}$

Prepare tables in your lab notebook for recording the count data in neat and compact form.

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, varying the gain of the amplifier, or changing the threshold level of the discriminator. Arrange things to yield three different mean count rates of approximately 1 sec^{-1}, 10 sec^{-1}, and 100 sec^{-1}.

At each of these approximate rates, record the counts for 100 repeated one-second intervals directly into your lab notebook.

Analysis

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

a) For each of the three runs calculate and plot the cumulative average, $r_c(j)$, of the rate as a function of the sequence number, j, of the count. By "cumulative average" is meant the quantity

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

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. Arrange the ordinate scale on the plot so that the largest positive and negative fluctuations fill the available vertical space. Include error bars to demonstrate convergence.

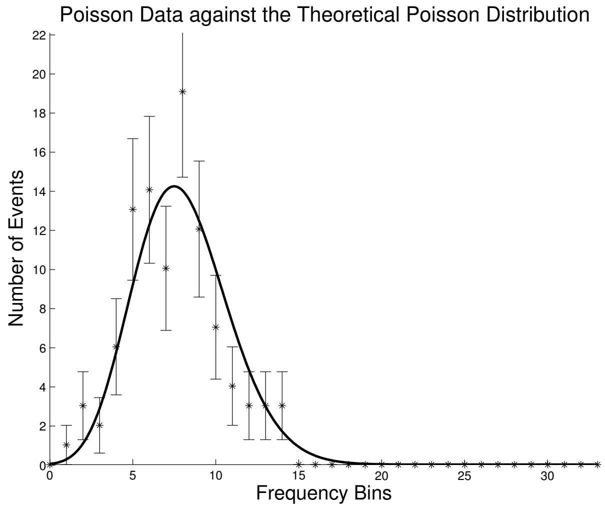

FIG. 2: A frequency distribution of observed numbers of counts. The renormalized Poisson distribution for the observed mean value is also plotted.

b) Calculate the mean and standard deviation of each of the three 100-trial distributions. Make a plot of the "number of counts" on the horizontal axis and data points for the "frequency" of occurrence for each bin on the vertical axis. Be sure to include error bars.

c) Using the mean rate just determined, calculate and plot data with errors (Fig. 2) and the Poisson frequency distribution (renormalized by multiplying by the total number of readings) on the same axes. For the observed distribution with the lowest mean rate, 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.)

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

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

Compare the Poisson and Gaussian distributions for $\mu = 10$.

Poisson Statistics Demonstration by Computer

There are two options for generating synthetic poisson data sets, one on the Windows PC's using LabVIEW and one under Athena using Matlab.

Windows Activate the program called 'Poisson Statistics' on a Junior Lab PC (downloadable from the Junior Lab Server in the software folder).

Athena Add the Junior Lab Locker by typing 'setup 8.13' at the Athena prompt. Within Matlab on Athena, type 'addpath /mit/8.13/matlab'. There are two Matlab scripts entitled 'poisson.m' and 'poissonsim.m'. 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 blue 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.

The following instructions apply to either method.

Generate 1000-trial distributions for a mean, $\mu = 1$, 10, 100 and 1000.

Next, generate ten 100-trial distributions for each of the three 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. 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. Determine the error on μ and σ from the scatter of the ten distributions.

For whichever poisson simulator you've used, inspect the code. In the LabVIEW code, for $\mu < 88$ it works according to a general scheme for Monte Carlo simulations which employs a generator of random numbers with a *uniform* distribution between 0 and 1 and yields a random variable with a specified distribution. The theory of these schemes is presented in the Appendix.

[1] P. R. Bevington & D. K. Robinson. 1992, *Data Reduction and Error Analysis for the Physical Sciences,* (2nd edition), McGraw Hill.

[2] A. Melissinos. 1966, *Experiments in Modern Physics*, Academic Press.

SUGGESTED THEORETICAL TOPICS

1. The Poisson distribution.
2. The Gaussian approximation to the Poisson distribution $P(m,n)$ for $m \gg 10$.
3. The differential distribution in the time lag between successive random pulses that occur at a fixed average rate.

MONTE CARLO GENERATION OF A RANDOM VARIABLE WITH A SPECIFIED PROBABILITY DISTRIBUTION

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

$$P(x) = \int_{-\infty}^{x} p(x')dx' \tag{5}$$

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

$$\lim_{x\to\infty} (P(x)) = 1 \tag{6}$$

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.

The Poisson probability, being a discontinuous function, is handled in a similar way, but with a summation rather than an integral. Given a value of μ <88 and a random number y, the Poisson Simulator find the smallest value of x for which $P(x;\mu) > y$, where P is defined by the formula

$$Pp(x;\mu) = \sum_{x'=0}^{x'=x} p(x';\mu) \tag{7}$$

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

If $\mu \geq 87$ the (LabVIEW) Poisson Simulator switches to the Gaussian approximation, calling a subroutine GAUSS which generates a random number with a Gaussian distribution and $\sigma = \sqrt{\mu}$. The GAUSS algorithm is

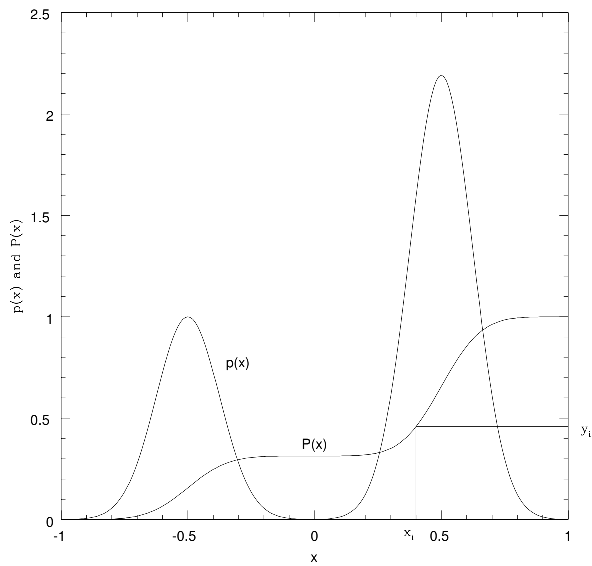

FIG. 3: Illustration of Monte Carlo selection of a random variable with a specified differential probability distribution is shown above. y_i is a random number between 0 and 1. x_i is the value of the variable for which the integral probability distribution equals y_i.

based on the fact that the distribution of the sum of 12 uniformly distributed random numbers is approximately Gaussian with a mean of 6 and a standard deviation of 1 (see if you can prove this). The rest of the simulator program plots the frequency distribution and calculates the mean and standard deviation.

In the Matlab version, the algorithm is described in more detail in the code.

EQUIPMENT LIST

Item	Description	Website
Tektronix 2225	50 MHz Analog Oscilloscope	www.agilent.com
Canberra 802-3	Scintillation Counter	www.canberra.com
Canberra 805	Scintillation Counter Pream	www.canberra.com
Canberra 3002D	High voltage power supply	www.canberra.com
Ortec 575A	NIM Amplifier	www.ortec-online.com
Ortec 776	NIM Timer/Counter	www.ortec-online.com