Abstract

The goal of a physics experiment is the extraction of physically meaningful parameters from data and/or testing of a theoretical hypothesis. Classical statistical analysis is the mathematical study of these goals. In this experiment, you will gain some insight into classical statistical analysis and additional experience in working with experimental data.

1 Introduction

In physics we seek a rational reductionist explanation for the world surrounding us: we want a mathematical framework that consistently and repeatably predicts (within the experimental and theoretical uncertainties involved) events that can be measured\textsuperscript{1}. The standard philosophy of physics, based on the work of Karl Popper, requires a falsifiable theory: a theory in physics only has value insofar as predictions can be compared to data and potentially shown to be incorrect. While we may never know how the world actually works, we can certainly discover ways in which the world does not work, by ruling out potential models of physics mechanisms. Once we have settled on a theory that is consistent with physical observations, we may wish to extract a physically meaningful parameter associated with that theory. For example, once we convince ourselves that hydrodynamics is a good description for fluid flow, we may ask what the viscosities of various fluids are.

Classical statistics is the mathematical study of hypothesis testing and parameter estimation based on sets of numbers; since these are precisely the goals we wish to achieve, an at least cursory knowledge of statistics is crucial for theoretical and experimental physicists alike. There are a number of books on classical statistical analysis, and many that focus specifically on questions we wish to answer in the physical sciences\textsuperscript{[1–10]}. Of the previous, Cowan’s book \textsuperscript{[5]} is an especially lucid treatment.

There are two general interpretations of statistics: frequentist and Bayesian. The two different approaches to classical statistics each has its own advantages and disadvantages. In Bayesian statistics, one simultaneously determines physical parameters and performs hypothesis testing, which is updated as data increases; however, the process is more complicated than the frequentist approach, especially in its requirement for a prior estimation for the likelihood of a hypothesis. The frequentist approach limits itself to independently determining physical parameters and testing goodness-of-fit, but is much easier to implement. It is not uncommon in a physical analysis to mix both interpretations. As an introduction to classical statistics, we will focus solely on the frequentist interpretation.

2 Preparatory Questions

1. Describe how a Geiger counter works, starting with the entrance of a charged particle or photon and ending with the electrical signal. Are subsequent signals independent?

2. Starting with the formula for the Binomial Distribution for an event to occur $x$ times in $z$ trials, where the probability of the event to occur in one trial is $p$,\n
\begin{equation}
P(x; p, z) \equiv \frac{z!}{x!(z-x)!} p^x (1-p)^{z-x}, \quad (1)\end{equation}

\textsuperscript{1}Physics is not like some other sciences which content themselves with simply cataloging phenomena—we want to describe physical phenomena in as simple a way as possible.
derive the Poisson distribution, the probability of an event to occur \( x \) times when the expected mean number of events is \( \mu \):

\[
P(x; \mu) = \frac{\mu^x e^{-\mu}}{x!}
\]  

(2)

by assuming that \( p \) is very small, \( z \) is very large, and \( p z = \mu \) is fixed. You will find Stirling’s approximation useful: for \( n \) large \( \log(n!) \approx n \log(n) - n \). In this way the Poisson distribution naturally arises in any natural phenomenon which involves low probabilities of events with very large numbers of trials, such as the spontaneous decay of radioactive nuclei. Note that unlike the Binomial Distribution, the Poisson Distribution relies on only one parameter, \( \mu \).

Using Eq. (2) prove the following:

(a) \( \langle x \rangle = \mu \)

(b) \( \langle x^2 \rangle = \mu (\mu + 1) \)

(c) \( \langle (x - \mu)^2 \rangle = \mu \)

where \( \langle x \rangle \) represents the mean value of \( x \).

3. The gamma function is defined as \( \Gamma(t) \equiv \int_0^\infty x^{t-1}e^{-x}dx \). Show that \( \Gamma(1) = 1 \) and \( \Gamma(n) = (n-1)\Gamma(n-1) \), and hence \( \Gamma(n) = (n-1)! \) for \( n \) integer. Thus the gamma function is an analytic continuation of the factorial to the full complex plane.

4. Plot the frequency distribution of counts when the average counts per interval is 1.5. Make use of the gamma function to create a smooth plot as a function of counts.

5. Suppose the mean counting rate of a certain detector of random events is 2.1 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?

6. Suppose a drug company is testing its newest wonder treatment for a rare disease. The company splits a large number of people (often companies use 20,000 people or more) randomly and evenly into a trial group treated with the drug and a control group that does not receive the drug. Suppose that in the control group 3 people get sick while in the trial group 2 people get sick. Evaluate the company’s claim that their drug is a stunning success which cuts the disease rate by over 30% by calculating the probability that fewer than 3 people would get sick in the drug trial group assuming the drug is completely ineffective.

7. 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 11 hours and two counts in the 12th hour. Give a quantitative answer to the question: what is the likelihood that the equipment is malfunctioning?

3 Experiment

In the first part of this experiment you will set up a Geiger counter, expose it to gamma rays from a radioactive source (and ubiquitous cosmic rays), and record the frequency distribution of the numbers of counts in equal intervals of time. This will be repeated for four situations (you will do one) with widely different mean count rates, approximately 4, 10, 30, and 100 counts per 10 seconds. You only know the “real” average rate at the end, but you should aim for these values within \( \pm 25% \). The experimental distributions and their standard deviations will be compared with the theoretical distributions and their standard deviations.

3.1 Setup to Measure Poisson Statistics

First, record 100 trials without exposing the counter to the source in order to measure your background. (Don’t forget to record your instrumental settings in your lab book!) The voltage applied to the Geiger counter should be \( \sim 600 \) volts (the voltage is found by taking the DC current listed on the Geiger counter and multiplying by 10 V/mA). Note that the voltage may drift over time; you should try to keep the voltage constant by infrequently adjusting the voltage dial if necessary. Next, expose the detector to gamma rays from the \( ^{60}\text{Co} \) source provided to you. You can control the counting rate by adjusting the distance of the source from the Geiger. When you are confident your setup is providing you with an appropriate rate, collect data for 100 more trials. Once complete share your results with others who did not use your rate such that you have the results for all of the rates.
4 Analysis

4.1 Introduction

In any one time step which is much shorter than the half life of a radioactive isotope, the probability of a nucleus decaying is very small. If the time required to conduct an experiment on a radioactive material is also very small compared to the half life of the isotope, then the probability $p$ of a nucleus decaying during any one time step in the experiment is both small and approximately constant. Finally, if the sample size of the radioactive isotope is of macroscopic size, then the total number of nuclei $N$ in the sample is astronomically large, on the order of $10^{23}$, and the total number of nuclei during the course of the experiment is approximately constant. In this case, the number of nuclei that decay in any one time step is given by the binomial distribution $B(n; p, N)$. Since $N$ is enormous and $p$ is tiny, as you showed in the preparatory questions, one may very safely approximate this binomial distribution by the Poisson distribution $P(n; \mu)$, $\mu = pN$. To justify this line of reasoning, compute the probability $p$ of any one nucleus decaying during one trial$^2$.

Now that we have our hypothesis, that the number of gamma rays emitted by the radioactive source during any one constant time period is distributed according to the Poisson distribution $P(x; \mu)$, we would like to extract the physically meaningful value of $\mu$ for the assumed Poisson distribution $P(x; \mu)$ from our measurements$^3$. One way of extracting this parameter is by the Method of Maximum Likelihood. The idea behind the MML is that one has a collection of data, $\{x_i\}$, and one has a hypothesis for how the data are distributed; $x$ is taken to be a random variable distributed according to, say $P(x; \theta)$, which depends on some vector of parameter values $\theta$. Now the entire collection of $N$ data points $\{x_i\}$ is also a random variable, which is distributed according to the joint probability distribution

$$P(\{x_i\}; \theta) = \prod_{i=1}^{N} P(x_i; \theta).$$  

(3)

The MML says that the estimators $\hat{\theta}_j$ for the parameters $\theta_j$ (of the vector of parameters $\theta$) are found at the maximum of the joint pdf given the measured data; i.e. an estimator $\hat{\theta}_j$ is a solution to

$$\partial_{\theta_j} \prod_{i=1}^{N} P(x_i; \theta) = 0.$$  

(4)

In general, Cowan$^5$ writes that the MML yields the arithmetic mean

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

(5)

as the estimator for the parameter that gives the mean of the assumed distribution. However, it is extremely instructive to derive this result oneself for the Poisson and Gaussian distributions; do this (note that the derivation is easier if you examine the maximum of the log of the joint pdf; since log is monotonic, the maximum for $\log(f(x))$ occurs at the same input value as for the maximum of $f(x)$.)

So we have that $\bar{\mu} = \bar{x}$ for our assumed Poisson distribution. There are two very important questions we need to ask before we continue. First, is $\bar{\mu} = \bar{x}$ an unbiased estimator? Unbiasedness means that the expectation of the estimator is the parameter itself; for this particular case, do we have that $E[\hat{\mu}] = \mu$? Show that this is indeed the case. (Remember that we are assuming that the $x_i$’s are distributed according to a Poisson distribution $P(x; \mu)$, although we do not know what $\mu$ is.) Second, every experimental measurement must be reported with an uncertainty estimate, so we need to estimate the uncertainty related to this parameter extraction. One often reports uncertainty based on the variance of the distribution. This scheme is motivated by the Gaussian distribution, $P(x; \mu, \sigma)$, for which $V[x] = E[(x - \mu)^2] = \sigma^2$ (show this). The Gaussian distribution is well understood (and most distributions we encounter are either Gaussian or approximately Gaussian), and one takes the convention of estimating the uncertainty of a value as given by one standard deviation of the distribution about this value, which is given by the square root of the variance$^4$.

In order to estimate the uncertainty on the mean value parameter that we extract, show that, for data distributed according to the Poisson distribution

$$E[(\bar{\mu} - \mu)^2] = E[(\bar{x} - \mu)^2] = \frac{\mu}{N} \approx \frac{\bar{x}}{N},$$  

(6)

where $N$ is the total number of data points collected.

$^2$Note that in this experiment we use Cobalt-60, which has a half-life of 5.27 years, and our trials last 10 seconds

$^3$Since $E(x) = \mu$, we know that we can interpret $\mu$ as the mean number of counts observed in any one trial in an experiment with counts distributed according to the Poisson distribution.

$^4$Note that the variance is sometimes referred to as the second central moment of the distribution.
Now we are ready to report our first result from the experiment.

But before we provide a value (with uncertainty) for this physically meaningful value, we should discuss the interpretation of the uncertainty. We are using a frequentist interpretation (as opposed to Bayesian), which means that our uncertainty implies that one standard deviation’s worth, or \( \approx 68\% \), of all (potential) future repetitions of this experiment will extract a parameter value in the range given by the uncertainty band. Equivalently, there is an \( \approx 68\% \) chance that the “true” value of the parameter (which we will never know for certain) lies within this uncertainty band (and hence there is an \( \approx 32\% \) chance that the parameter does not lie in this range!). There are other methods of estimating the uncertainty on an extracted parameter, and there are other reasonable values chosen for the probability that a future measurement of a parameter will coincide within the expressed uncertainty band. This is one of those instances where there is no right answer; one simply must clearly state what one has done and faithfully report the result.

The following analysis requires the use of repetitive arithmetic on the collected data set. You are free to use any (canned) software you like, such as VPython or Mathematica.

### 4.2 Running Mean

For each of the four runs calculate and plot the cumulative (running) average, \( r_c(j) \), of the rate as a function of the sequence number, \( j \), of the interval count and its associated uncertainty. By “cumulative average” is meant the quantity

\[
r_c(j) \equiv \bar{x}_j = \frac{1}{j} \sum_{i=1}^{j} x_i.
\]

where \( x_i \) is the number of counts detected in trial \( i \). Since our estimator is unbiased, and over the timescale of the experiment the process should have a steady mean rate, the running average should converge (within the uncertainties, which should be shrinking as \( j \) increases). In the text of your report, be sure to describe explicitly what you are plotting, including all formulas used (i.e. the formulas you use for computing the running mean and its uncertainty, which, in this case, you will take as the square root of the variance of the mean estimator, the “standard error of the mean”). Make sure to label all axes (in a large, readable font) on the plot, including units. The effectiveness of the standard error of the mean as an estimate for the uncertainty on a quantity relies on the central limit theorem and assumes a large number of measurements (\( \gtrsim 30 \)) do your error bars give a reasonable estimate for the uncertainty on the running mean for small values of \( j \)?

#### 4.3 Arithmetic Mean and Sample Variance of the Data

We have already figured out how to determine \( \bar{x} \pm \Delta \bar{x} \) for data assumed to be distributed according to the Poisson distribution. We also want to report the value for the sample variance of the data

\[
s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2.
\]

(Our first nontrivial test of whether our measured distribution is a Poisson distribution will be a comparison of the arithmetic mean and sample variance of the data.) Since all reported measurements must have an uncertainty, we need to estimate the uncertainty on this quantity. Since the variance of a Poisson distribution \( P(x; \mu) \) is \( \mu \), we may determine the uncertainty on \( s^2 \) by examining the variance of the variance: \( V[s^2] = E[(s^2 - \mu)^2] \).

Show that for data given by the Poisson distribution,

\[
E[(s^2 - \mu)^2] = \frac{2N \mu^2 + (N - 1) \mu}{N(N - 1)}.
\]

One may wish to use the equation

\[
V[s^2] = \frac{1}{N} \left( m_4 - \frac{N - 3}{N - 1} m_2^2 \right),
\]

where \( m_4 \) is the fourth central moment of the pdf for the data [11].

For each of the four runs, compute \( \bar{x} \pm \Delta \bar{x} \) and \( s^2 \pm \Delta s^2 \). Give these results in a table in your report.

Now we wish to make our first nontrivial test of the Poissonian nature of the data by seeing to what extent the sample variance agrees with the arithmetic mean (i.e. do the first two moments of the measured data behave as we would expect if the data were distributed according to the Poisson distribution). Make a plot of the four values of \( s^2/\bar{x} \) as a function of \( \bar{x} \) using the results you just found for your table. Make sure to properly propagate your uncertainties. Compare to the expected value of \( 1 \).

\[\text{Remember to keep in mind that } \mu \text{ is some unknowable “true” value for the physical distribution and } s^2 \text{ is an estimation of the variance of the distribution of the data determined from the data.}\]
4.4 Poisson Plots

We are now going to bin our data where we will keep track of the number of counts per trial on the x-axis and the number of trials with this number of counts per trial on the y-axis. In general, binned data are distributed according to the multinomial distribution (see [5]); the number of counts \( n_i \) in bin \( i \) is distributed according to a binomial distribution \( B(n_i; p_i, N) \), where \( N \) is the total number of trials in the experiment, and \( p_i \) is the probability (according to the distribution for the physical system at hand) of a single measurement giving a result in the given bin \( i \). Since in the binomial distribution \( E[n_i] = N p_i \), the entire binned distribution will follow the pdf for the underlying physical process; in our case we expect this underlying distribution to be the Poisson distribution. Since the variance of the binomial distribution is \( V[n_i] = N p_i (1 - p_i) \), we may estimate the uncertainty of the value of the measured \( n_i \) as

\[
\sqrt{n_i \left(1 \frac{n_i}{N}\right)};
\]

we arrive at this uncertainty estimate by taking the measured value in a bin as an independent measurement of the mean of the full distribution. (One may think: oh, but I already have a theoretical prediction for the distribution of the entire binned distribution, which gives an even more precise determination of the mean of the full distribution; why don’t I use that to estimate the uncertainty on each individual bin? One could do this, but the uncertainty estimates on each bin would no longer be statistical, i.e. independent of one another; they would rather be systematic uncertainties that were all mutually correlated with each other.)

Now one could do a simple-minded binning where the size of each bin is only a single number of counts per trial. This is not always a good idea, though. Generally one wants about 5 points in a bin in order to have a “good” estimate of the mean value and uncertainty on that bin (note, for instance, how the estimate above yields an uncertainty of 0 for a bin with 0 trials in it!). Also, we will be performing a \( \chi^2/\text{dof} \) analysis to quantify how well the theoretical curve describes the data. A good rule of thumb when performing \( \chi^2/\text{dof} \) analysis is that at least 80% of the bins have at least 5 trials in them [5].

For the Poisson distribution, we then face two potential challenges in terms of getting “good” statistics in each bin: the bins corresponding to the tails of the distribution, which will naturally not have very many trials in them; and the experimental setups with large mean counts per trial (there will only be a few counts per bin because of the natural width of the distribution). We will resolve the first issue by having a first and (last) bin which counts the number of trials with \( n_i \) \((n_u)\) counts per trial or fewer (greater); you will choose \( n_i \) \((n_u)\) to guarantee at least 5 trials are counted in each bin. We will resolve the second issue by expanding the size of a central bin. The \( i \)th bin will count the number of trials that had \( i, i+1, \ldots, i+k \) counts; the \( i+1 \)th bin will count the number of trials that had \( i+k+1, \ldots, i+2k \) counts; etc. Vary \( k \) until your binned data have good enough statistics.

It’s almost always a good idea when plotting one’s data to simultaneously compare to a theoretical prediction. In this case, our theoretical prediction is a Poisson distribution \( P(n; \mu) \). Now \( \mu \) is (forever) unknown to us; however, we have extracted an estimate for its value using the MML, \( \mu \approx \bar{x} \). Plot \( NP(n; \bar{x}) \) (or the relevant version of this if the bins are of nontrivial width and also at the ends of the data range). See Fig. (1) for an example.

4.5 Fraction of Bins Not Described by Poisson

We can do a check to see whether we made a good estimate for our error bars. (This interpretation assumes that our hypothesis for the theoretical description is correct; one might alternatively say this is a test of our theoretical prediction. For the sake of this lab, let’s view this part of the exercise as getting a feel for how “good” a description we should expect, and how large statistical fluctuations might be.) If our error bars truly represent the probability that 68% of all future measurements will lie within the error bars, then there is a 68% chance that the Poisson prediction with the “true” value of the mean lies within the error bars of each bin. If we assume that our estimated \( \mu \approx \bar{x} \) is a good approximation, we should therefore expect that about 68% of the binned data agrees with the Poisson curve, within the uncertainty of the binned data. Conversely, we actually expect that about 32% of our data points with uncertainty disagree with the Poisson prediction curve! Check this prediction for your data.

We can easily count the number of bins that agree with the Poisson curve within one standard deviation (i.e. the curve goes through the uncertainty band). But we should associate an
uncertainty to every measurement, and the number of bins that agree with the Poisson curve is a measurement. In order to make an estimate for the uncertainty, we need to figure out the probability distribution for the number of bins which agree with the Poisson curve. Well, since each bin has some probability \( p_a \) of agreeing with the curve and there are \( N_{\text{bins}} \) bins, the number \( n_a \) bins that agree with the Poisson curve is given by the binomial distribution, \( B(n_a; p_a, N_{\text{bins}}) \). Hence the uncertainty is given by the variance of the binomial distribution, \( N_{\text{bins}} p_a (1 - p_a) \).

Taking the measured number \( n_a \) as an estimate for the probability that any given bin will agree within uncertainty with the Poisson prediction means we should report

\[
n_a \pm \sqrt{n_a \left(1 - \frac{n_a}{N_{\text{bins}}}ight)}
\]

for the four runs. Again, this uncertainty estimate is not reliable if \( n_a = 0 \) (or really whenever \( n \) is small). The estimate is also unreliable if \( n_a \lesssim N_b \), which we should see about \( p_a^{N_{\text{bins}}} \) times (\( \simeq 2\% \) of the time taking \( p_a \simeq 0.68 \) and \( N_{\text{bins}} = 10 \)).

Make a plot of the fraction of bins that agree with the Poisson curve within one standard deviation (with uncertainty) plotted against the mean rate, and compare to the expected value of \( \simeq 0.68 \).

### 4.6 Expected Size of Deviations

We can perform another analysis to give ourselves a sense of how much and how frequently data naturally fluctuate due to statistics by asking: what is the biggest fluctuation away from the theoretical prediction, and how often should I expect a fluctuation of this size to occur? (This view again assumes that the theoretical description is correct, and, again, one could view this check as a test of the theoretical description.)

Find the bin that has fluctuated the most away from the expectation based on the plotted Poisson distribution. Compute the probability that should you repeat the measurement of 100 trials, your new set of data would contain a bin that fluctuates further away from the theoretical description than your bin with the largest deviation in your original measured data.

### 4.7 Conclusions

What can you conclude about the hypothesis as a result of analyzing your data as laid out above? Are there any systematic deviations? If so, can you think of a good explanation for such a systematic deviation?

---

\(^6\)In either of the cases that \( n \) is small of approximately equal to \( N_b \), one should really use Bayesian statistics; in this one case do not worry here if \( n \) is not large enough or is approximately equal to \( N_b \).
References


