Poisson Statistics

Berat Gönültaş

March 2019

Abstract

Poisson distribution is, in general, used to model the number of events occurring in a fixed interval of time (and/or space) if these events occur with a known average rate and independently of time since the last event.[1] In this experiment, we examined the relation of and the difference between the Gaussian and Poisson distributions, and investigated if the underlying randomness of the radioactive decay process can be modeled by the Poisson distribution or not.

1 Theoretical Motivation

Radioactive decay is the process of emitting radiation by an unstable atom. Radiation can occur in the form of alpha rays, beta rays or gamma rays. It is impossible to know when a single atom will decay. However, it is possible to assign an expected decay rate for a collection of atoms. (e.g. half life) Thus, in that case, it is convenient to consider the decay process as stochastic.

Let us assign a probability p for the decay of a single atom in a certain time interval. Then, one can write the probability of decay of x many atoms in a source having total of n many atoms as

$$f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Here, n is a big number so that it is quite complicated to find the probabilities because of the factorial terms that will appear. However, these can be approximately found easily using by the Poisson approximation. It is reasonable to use the Poisson distribution to model radioactive decay since the decay events occur with a known average rate and independently of time since the last event, which are the requirements for a process to be considered as Poisson distributed.

Before discussing the Poison approximation to the binomial distribution, recall the following fact from calculus.

Proposition 1. *If $a_n \to 0$, $c_n \to \infty$, and $a_n c_n \to b$ as $n \to \infty$, then $(1+a_n)^{c_n} \to e^b$ as $n \to \infty$.*

[1] Ümit Işlak, Introductory (mathematical) statistics, `https://sites.google.com/site/umitislak/lecture-notes`

Theorem 1. For each integer n and each $p \in (0,1)$, let $f(x|n,p)$ be the binomial pmf(probability mass function) with parameters n and p. Let $f(x|\lambda)$ denote the Poisson pmf with parameter $\lambda > 0$. Also let p_n be a sequence of real numbers between 1 and 0 such that $\lim_{n\to\infty} np_n = \lambda$. Then for any $x \in \mathbb{N}$,

$$\lim_{n\to\infty} f(x|n,p_n) = f(x|\lambda)$$

Proof. We have

$$f(x|n,p_n) = \binom{n}{x} p_n^x(1-p_n)^{n-x} = \frac{n(n+1)...(n-x+1)}{x!} p_n^x(1-p_n)^{n-x}$$

Let $\lambda_n = np_n$ so that $lim_{n\to\infty} np_n = \lambda$. Then,

$$f(x|n,p_n) = \frac{\lambda_n^x}{x!}\frac{n}{n}\frac{n-1}{n}...\frac{n-x+1}{n}\left(1-\frac{\lambda_n}{n}\right)^{-x}\left(1-\frac{\lambda_n}{n}\right)^{n} \to \frac{\lambda^x e^{-\lambda}}{x!} = f(x|\lambda)$$

as $n \to \infty$, where for the last step we used Proposition 1.[2] □

So, when p_n is small and np_n is around λ, probabilities related to a binomial random variable with parameters n and p_n can be approximated with a Poisson distributed random variable with parameter $\lambda = np_n$. Thus, we can treat decay process as Poisson distributed as n is large in the case of decay process (number of atoms), p_n is small (probability of decay for one atom) and np_n can be assigned.

On the other hand, a limiting form of the Poisson distribution (and many others – see the Central Limit Theorem) is the Gaussian distribution. In deriving the Poisson distribution we took the limit of the total number of events $n \to \infty$; we now take the limit that the mean value is very large. Recall the Poisson distribution

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Now, let $x = \lambda(1+\delta)$ where $\lambda \gg 1$ and $\delta \ll 1$. Since $\langle x \rangle = \lambda$, this means that we will also be concerned with large values of x. Using Stirling's formula for $x!$, as $x \to \infty$

$$x! \to \sqrt{2\pi x}e^{-x}x^x$$

we find

$$\begin{aligned} f(x|\lambda) &= \frac{\lambda^{\lambda(1+\delta)}e^{-\lambda}}{\sqrt{2\pi}e^{-\lambda(1+\delta)}[\lambda(1+\delta)]^{\lambda(1+\delta)+1/2}} \\ &= \frac{e^{\lambda\delta}(1+\delta)^{-\lambda(1+\lambda)-1/2}}{\sqrt{2\pi\lambda}} \end{aligned}$$

The limit of a function like $(1+\delta)^{-\lambda(1+\lambda)-1/2}$ with $\lambda \gg 1$ and $\delta \ll 1$ can be found by taking the natural log, then expanding in δ to second order and using $\lambda \gg 1$

[2]The Poisson approximation proof was taken from Ümit Işlak's previously mentioned book.(see footnote1)

$$
\begin{aligned}
ln[(1+\delta)^{\lambda(1+\lambda)+1/2}] &= [\lambda(1+\delta)+1/2]\ln(1+\delta) \\
&= (\lambda+1/2+\lambda\delta)(\delta-\delta^2/2+O(\delta^3)) \\
&\simeq \lambda\delta+\lambda\delta^2/2+O(\delta^3)
\end{aligned}
$$

Then,

$$f(x|\lambda)=\frac{e^{-\lambda\delta^2/2}}{\sqrt{2\pi\lambda}}$$

Substituting back for x with $\delta=(x-\lambda)/\lambda$, yields

$$f(x|\lambda)=\frac{e^{-(x-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}}$$

This is a Gaussian distribution with mean and variance of λ.[3]

The probability of observing n counts during a time interval t is

$$f(n|\lambda)=f(n|\alpha,t)=\frac{(\alpha t)^n e^{-\alpha t}}{n!}$$

where $\alpha=\lambda/t$. Then, the probability of having one event in a time interval dt is

$$f(1|\alpha,dt)=\frac{(\alpha dt)e^{-\alpha dt}}{1!}=P(\alpha,dt,1)$$

Hence, the probability of having n events in a t interval followed by another event within a dt time is (assuming $e^{\alpha dt}\simeq 1$)

$$
\begin{aligned}
P_q(n+1,t)dt &= P(\alpha,t,n)P(\alpha,dt,1) \\
&= \frac{(\alpha t)^n e^{-\alpha t}}{n!}\frac{(\alpha dt)e^{-\alpha dt}}{1!} \\
&\simeq \frac{(\alpha t)^n e^{-\alpha t}\alpha dt}{n!}
\end{aligned}
$$

and

$$P_q(n+1,t)=\frac{(\alpha t)^n e^{-\alpha t}\alpha}{n!}$$

Again, as n becomes large, the distribution approaches the Gaussian distribution.[4]

[3] The Gaussian approximation part was taken from the link `https://roe.ac.uk/japwww/teaching/astrostats/astrostats2012_part2.pdf`

[4] Advanced Physics Experiments by Erhan Gülmez (Boğaziçi University Publications, 1999, ISBN 975-518-129-6)

2 Apparatus and Experimental Procedure

In the experiment, we used Geiger-Muller tube, chart recorder and gamma ray sources.[5]

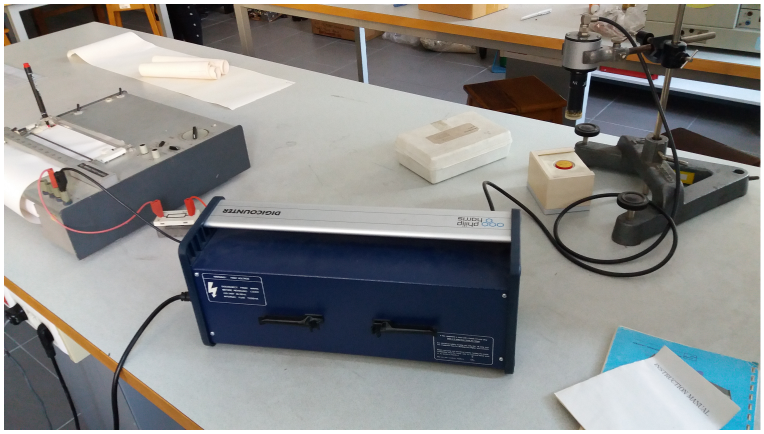

Figure 1: General view of apparatus

The GM(Geiger-Muller) tube is one type of gaseous ionization detector. Gaseous ionization detectors are radiation detection instruments used in particle physics to detect the presence of ionizing particles, and in radiation protection applications to measure ionizing radiation. They use the ionising effect of radiation upon a gas-filled sensor. If a particle has enough energy to ionize a gas atom or molecule, the resulting electrons and ions cause a current flow which can be measured. The three basic types of gaseous ionization detectors are ionization chambers, proportional counters, and Geiger-Müller tubes.[6]

The GM tube is a hollow cylinder filled with a gas at low pressure. The tube has a thin window made of mica at one end. There is a central electrode inside the GM tube. A high voltage supply is connected across the casing of the tube and the central electrode as shown at the Figure 2. When alpha, beta or gamma radiation enters the tube it produces ions in the gas. The ions created in the gas enable the tube to conduct. A current is produced in the tube for a short time. The current produces a voltage pulse. Each voltage pulse corresponds to one ionising radiation entering the GM tube. The voltage pulse is amplified and counted. The greater the flux of radiation, the more ionisation in the tube so the greater the number of counts.[7] One can use GM tube to detect particles or radiation having high energy like alpha rays, gamma rays or muons. The gas used in the tube is generally He, Ne or Ag. The GM tube can be used for the counting only. It can measure neither the position nor the type of particle.

The experiment consists of three parts. In the first part, we used Cs-137 as source to find the working voltage of the GM tube. We set the GM tube to its minimum voltage (300V) and the counter to the 100 seconds, radioactivity and single mode. We took data at 20V intervals. When the counts start to be almost constant as the voltage increases, we

[5]More photos of the apparatus are given at appendix

[6]https://en.wikipedia.org/wiki/Gaseous_ionization_detector

[7]https://www.bbc.com/bitesize/guides/zt9s2nb/revision/5

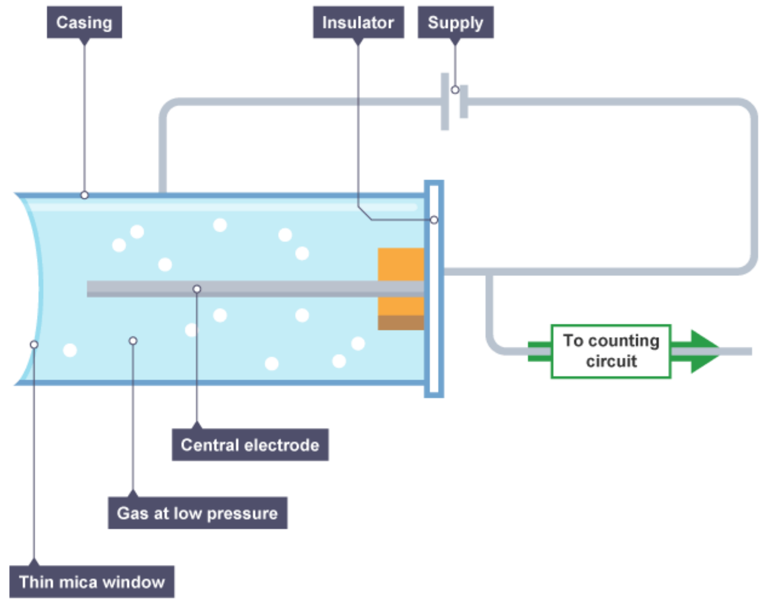

Figure 2: The Structure of the Geiger-Muller tube(an image from the link at the footnote number 5)

picked a value from the beginning of the "plateau", i.e. the flat region of the voltage vs count graph. 380V was the value we chose.

In the second part, setting the voltage to 380V, continuous mode, we took 100 data from the counts of Cs-137 with the counter set to 10 seconds and 1 seconds. So we get two sets of data consisting 100 counts each. We repeated the same procedure for Ba-133.

In the third part, we obtained the distribution of the successive counts. We arranged the position of source so that the chart is clear enough to extract data. After calibration, we ran the chart recorder for around one and a half minutes. We turn the chart recorder off and took the portion of the paper. We were interested in n=0 and n=1 cases. n=0 case is the case where there is no recordings between two recordings. So, for n=0 case, we measured the distances between the adjacent pulses. Since, the exact unit is not important in this experiment, we measured time with millimeters. n=1 case corresponds to the case where there is one recording between two recordings. So, for n=1 case, we simply added two successive pairs of time intervals of the case n=0.

3 Data and Analysis

The corresponding counts to different high voltage values are given below.

HV(Volts)	Count
300	0
320	0
340	0
360	262
380	362
400	377
420	376
440	384

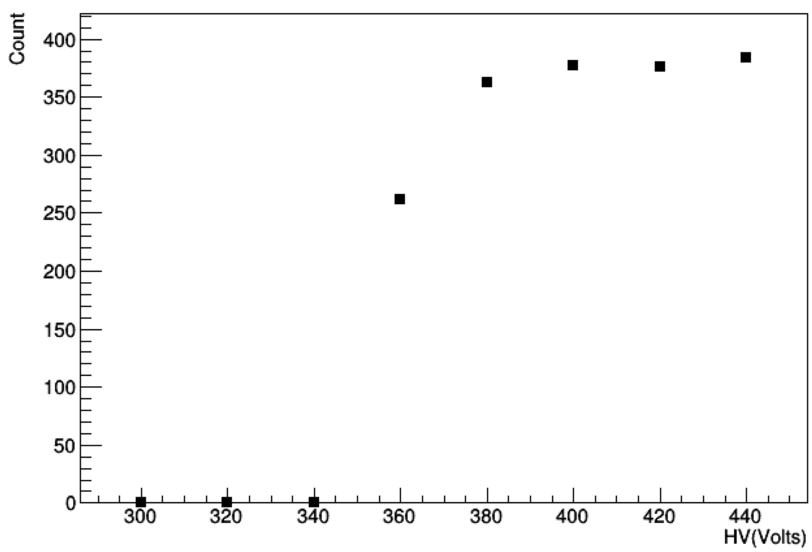

Figure 3: High Voltage versus Count Graph

So, we chose 380 V as the working voltage of the GM tube.

The counts from the part in which Cs-137 used as a source are given below.[8]

10 sec	1 sec
102	11
90	7
112	15
98	7
92	12
93	9
98	12
124	7
102	11
109	4
..	..
..	..

The counts from the part in which Ba-133 is used as source are given below.[9]

10 sec	1 sec
62	5
50	4
57	3
51	4
49	9
51	7
59	4
54	5
65	4
54	7
58	5
..	..
..	..

The data taken at part 3 is given below. Here, x shows the position of the peaks of the pulses.[10]

[8] For whole data, see appendix.
[9] For whole data, see appendix.
[10] For whole data, see appendix.

x(cm)	n=0	n=1
2.6	1.1	1.4
3.7	0.3	3.1
4	2.8	3.5
6.8	0.7	1.7
7.5	1	2.9
8.5	1.9	2.1
10.4	0.2	0.3
10.6	0.1	1.9
..	..	..
..	..	..

The graph of data related to Cs-137 for 10 seconds is given below(Figure 4). The green curve represents the Gaussian fit while the red curve represents the Poisson fit.[11] The results for the Poisson and Gaussian fit in the Figure 4 are also given below. In the results of the Poisson fit, p0 represents the normalization constant and p1 represents λ(Mean). In the results of the Gaussian fit, p0 represent the normalization constant, p1 represents mean and p2 represents standard deviation.[12]

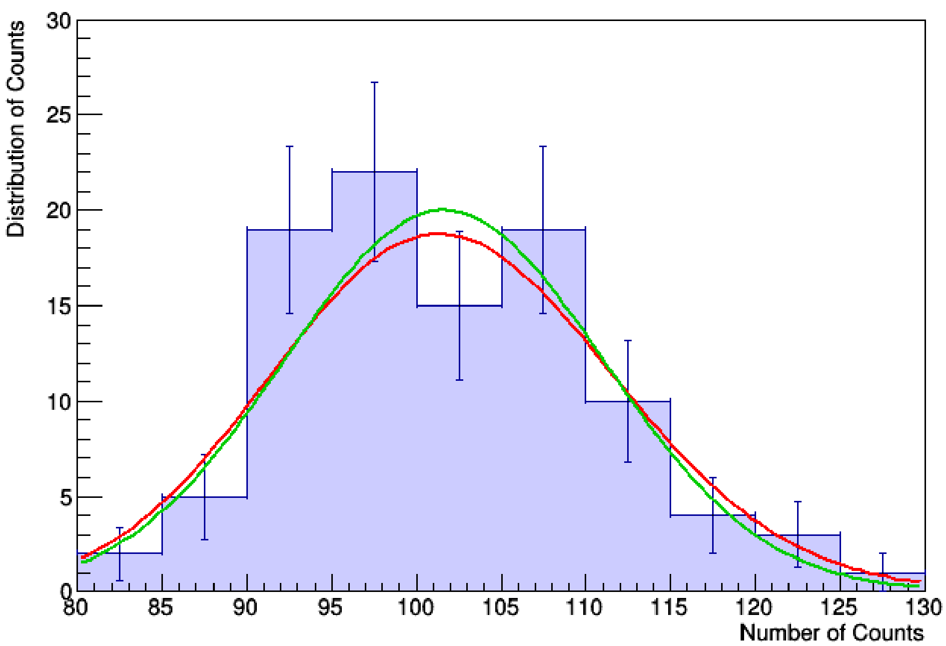

Figure 4: Number of counts of decay for Cesium-137 for 10 seconds

[11]Same convention for the colors of the curves will be assumed for the rest of the report unless stated otherwise .

[12]Same convention for p0, p1 and p2 will be assumed for the rest of the report unless stated otherwise.

Entries	100
Mean	101.1
Std Dev	9.302
Underflow	0
Overflow	0
χ^2 / ndf	6.538 / 8
Prob	0.5872
p0	474.6 ± 49.1
p1	101.8 ± 1.0

Figure 5: The results for the Poisson fit in the Figure 4

Entries	100
Mean	101.1
Std Dev	9.302
Underflow	0
Overflow	0
χ^2 / ndf	6.45 / 7
Prob	0.4883
p0	20 ± 2.7
p1	101.6 ± 1.0
p2	9.438 ± 0.857

Figure 6: The results for the Gaussian fit in the Figure 4

The graph of data related to Cs-137 for 1 seconds is given below(Figure 7).

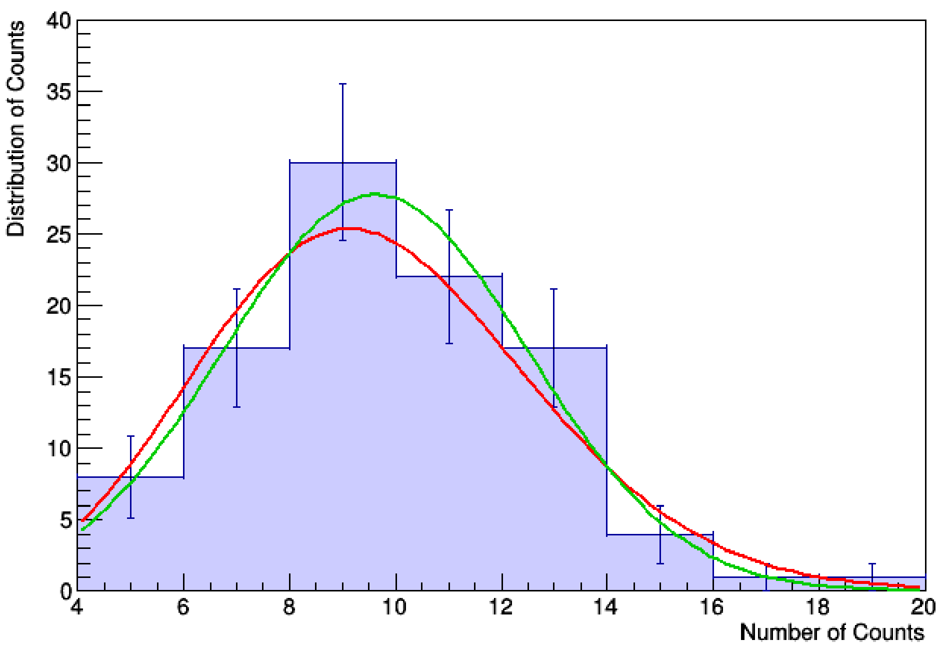

Figure 7: Number of counts of decay for Cesium-137 for 1 second

Entries	100
Mean	9.42
Std Dev	2.757
Underflow	0
Overflow	0
χ^2 / ndf	4.064 / 6
Prob	0.668
p0	196.3 $\pm$ 20.1
p1	9.625 $\pm$ 0.289

Figure 8: The results for the Poisson fit in the Figure 7

Entries	100
Mean	9.42
Std Dev	2.757
Underflow	0
Overflow	0
χ^2 / ndf	2.216 / 5
Prob	0.8186
p0	27.75 $\pm$ 3.60
p1	9.623 $\pm$ 0.326
p2	2.878 $\pm$ 0.269

Figure 9: The results for the Gaussian fit in the Figure 7

The graph of data related to Ba-133 for 10 seconds is given below(Figure 10).

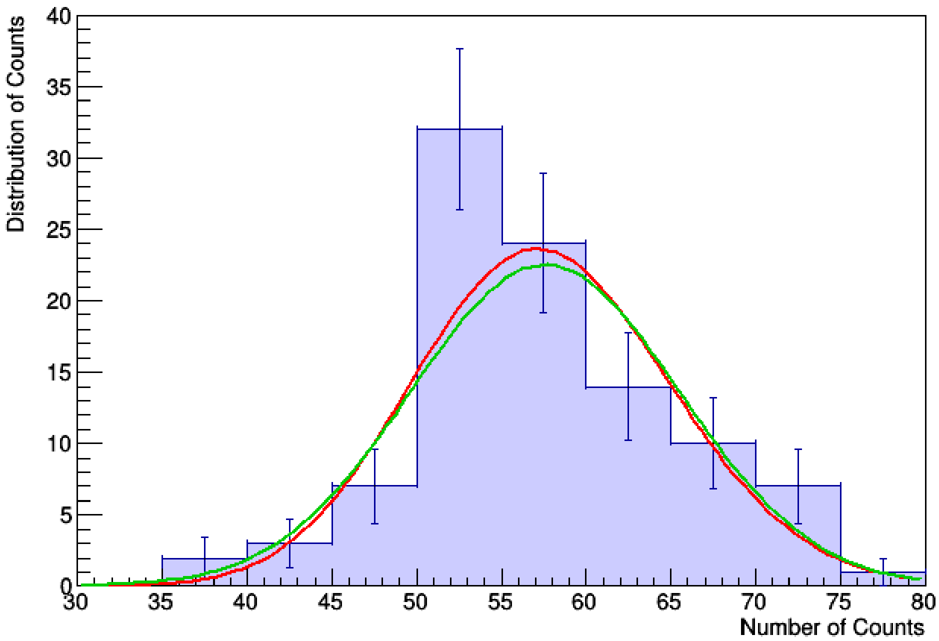

Figure 10: Number of counts of decay for Barium-133 for 10 second

Entries	100
Mean	57.09
Std Dev	7.63
Underflow	0
Overflow	0
χ^2 / ndf	10.46 / 7
Prob	0.164
p0	449 $\pm$ 47.4
p1	57.62 $\pm$ 0.84

Figure 11: The results for the Poisson fit in the Figure 10

Entries	100
Mean	57.09
Std Dev	7.63
Underflow	0
Overflow	0
χ^2 / ndf	10.93 / 6
Prob	0.09062
p0	22.5 ± 3.3
p1	57.59 ± 0.90
p2	7.927 ± 0.818

Figure 12: The results for the Gaussian fit in the Figure 10

The graph of data related to Ba-133 for 1 seconds is given below(Figure 13).

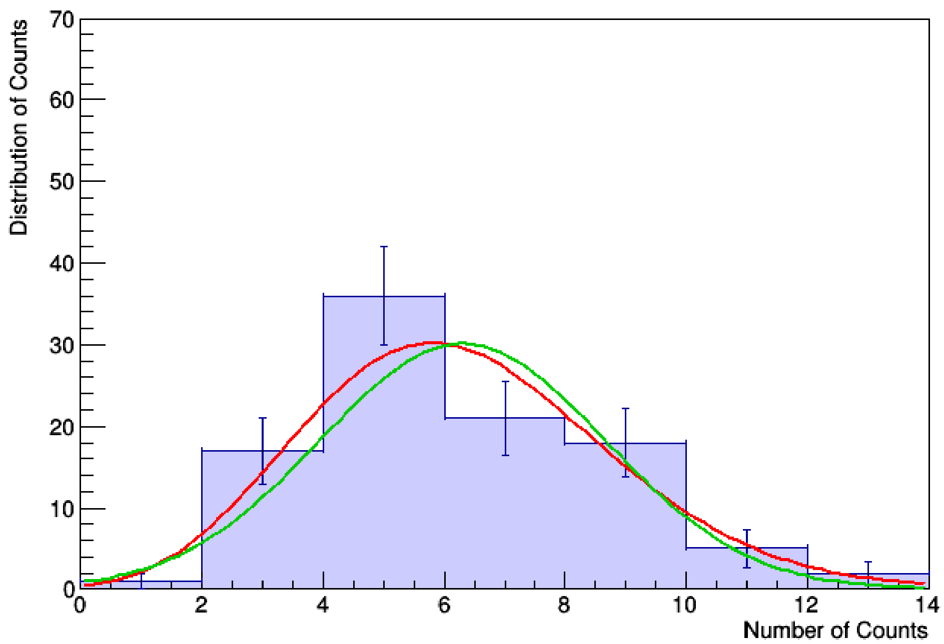

Figure 13: Number of counts of decay for Barium-133 for 1 second

Entries	100
Mean	5.73
Std Dev	2.357
Underflow	0
Overflow	0
χ^2 / ndf	5.793 / 5
Prob	0.3269
p0	189 ± 19.5
p1	6.313 ± 0.258

Figure 14: The results for the Poisson fit in the Figure 13

Entries	100
Mean	5.73
Std Dev	2.357
Underflow	0
Overflow	0
χ^2 / ndf	11.19 / 4
Prob	0.02453
p0	30.12 ± 4.08
p1	6.303 ± 0.321
p2	2.361 ± 0.203

Figure 15: The results for the Gaussian fit in the Figure 13

The graphs and the results of the third part are given below.[13](Figure 21 and 22) While fitting these curves, we used the distribution derived at the end of the theoretical motivation section, i.e.:

$$P_q(n+1,t) = \frac{(\alpha t)^n e^{-\alpha t}\alpha}{n!}$$

Respectively, we put $n = 0$ and $n = 1$ and found as follows:

$$P(0+1,t) = e^{-\alpha t}\alpha$$

$$P(1+1,t) = \alpha^2 t e^{-\alpha t}$$

[13]p0's are normalization constants and p1's are α

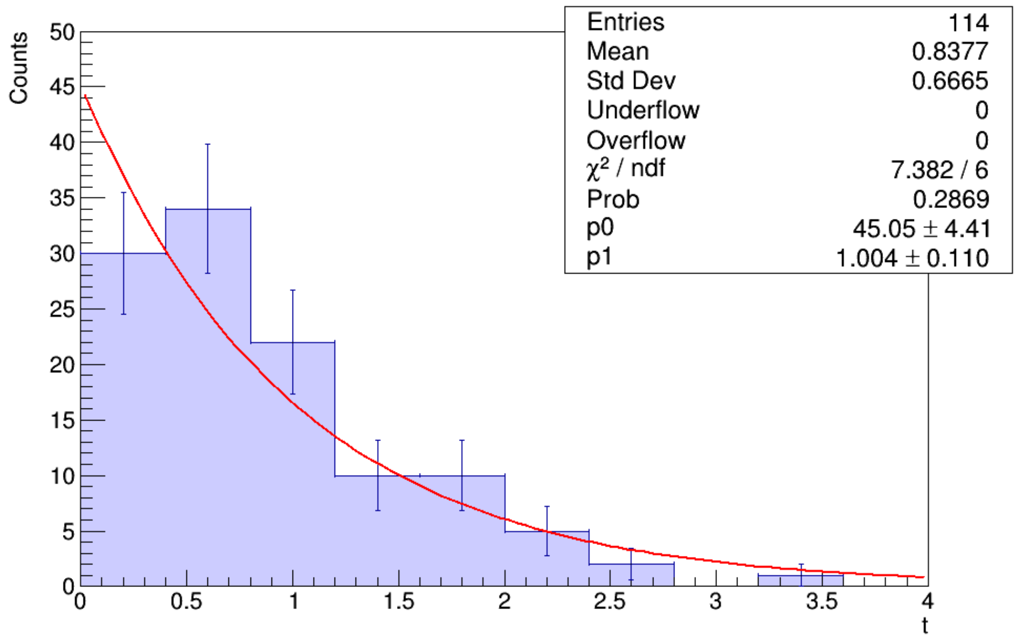

Figure 16: Distribution of 0 event(pulse) followed by another event

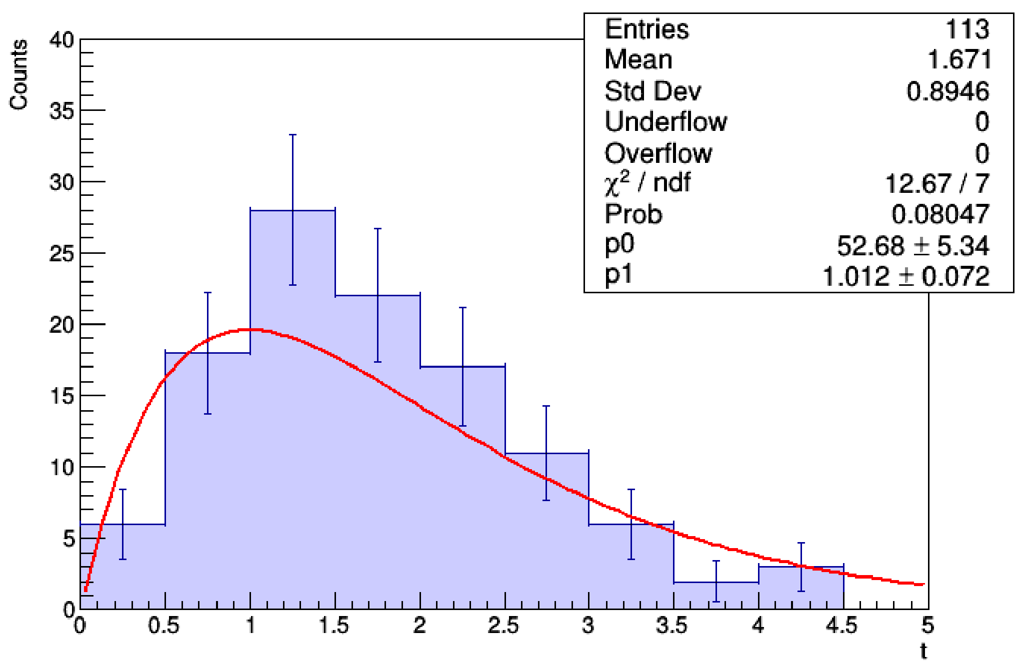

Figure 17: Distribution of 1 event(pulse) followed by another event

4 Conclusion

We were expecting that the data taken from the radioactive decay process fits the Poisson distribution and, approximately, the Gaussian distribution. So, in part 1, except the results of the decay of Cesium-137 for 1 second(Figure 7), we got the expected result which is a considerably high p-value for the Poisson fit and a near but smaller p-value for the Gaussian fit. In the exceptional case of the decay of Cesium-137 for 1 second, the p-value of Gaussian fit showed up being higher than the p-value of the Poisson fit. It is a bit suspicious since the chi2 value per ndf is also smaller than our expectation which is 1. The ndf value for that exceptional histogram is a bit smaller than the others since the bins are less. Having less ndf could be a reason for getting distorted results. We see a similar distortion at the results of Gaussian fit of the decay of Barium-133 for 1 second(Figure 13). However, since the x-axis range were smaller than 10 seconds cases, and since I was trying to arrange number of bins so that the histograms show the behavior of the data and also the bin boundaries are integers, there wasn't any possible ways to make more number of bins.

Regarding part 3, the p-values of this part are not so good, either. However, considering the being sensitive of the data of this part, it is understandable. The data of the part 3 is much more sensitive because any false counts change the data crucially since in this case we are taking the time interval of successive pair of pulses as data, not a count for a time period as in part 2. Taking into account the fact that the GM tube counts some counts even if there is no source at all because of the external (not considered in theory) radioactive sources, the p-values of the part 3 can also be seen as good enough. Looking at the α values we found, they agree with each other and the real value. (We arranged the source so that GM tube counts 1-2 counts per second at part 3.)

Of course, the error originated from us, observers, must also be taken into account. In such an experiment which involves very much data to be read and noted down carefully, a considerable amount of observer error is inevitable.

After all, the experiment successfully confirmed that the radioactive decay process is Poisson distributed as expected.

The experiment can be improved by several ways. Taking more data is a basic way to do it. We could use a system that automatically keeps record of the data given by the GM tube to avoid observer error. Using a more sensitive GM tube and a kind of more 'closed' system can improve the result of the experiment.

5 Appendix

The full data of the experiment and the codes used to analyze data can be found at the following github link:

`https://github.com/beratgonultas/phys442/tree/master/poisson`

More photos of the apparatus are below:

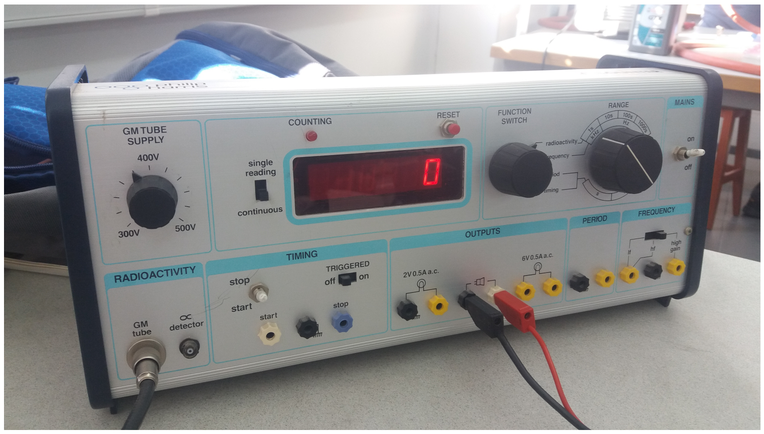

Figure 18: GM tube

Figure 19: GM tube

Figure 20: Cs-137

Figure 21: Ba-133

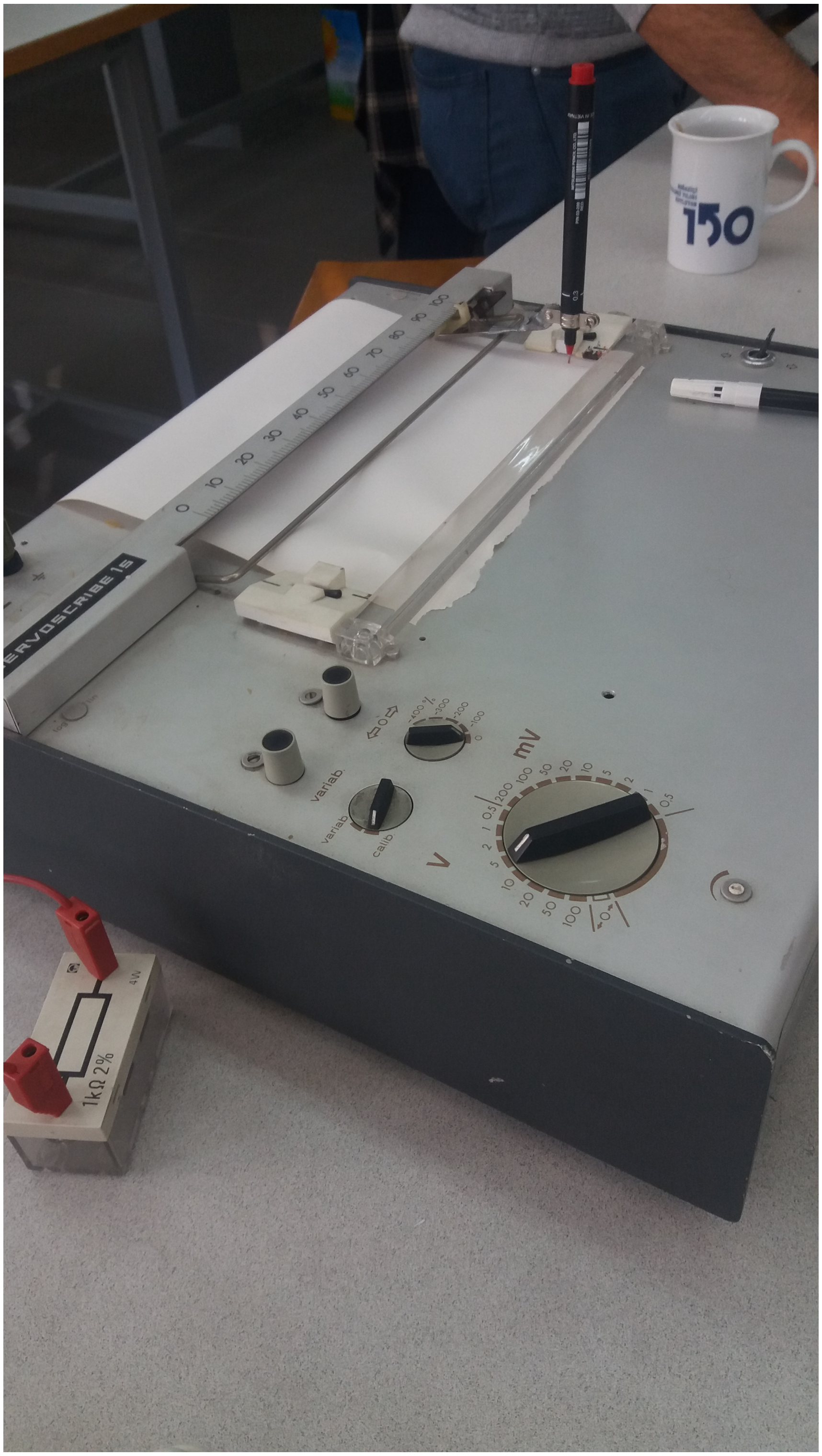

Figure 22: Chart recorder