Lab 3: Distribution Functions (Part 1+2)

Author: Zain Kamal (zain.eris.kamal@rutgers.edu)

1. Introduction

In this lab, we analyze distributions of radiation counts from Cesium-137 over various time intervals, sample sizes, and radiation intensities. These data are compared with theoretical predictions of the Gaussian and Poisson distribution functions in order to better understand statistical modelling and error analysis. The functions are useful in describing the probability of obtaining different values of a physical quantity that is subject to random errors or fluctuations, such as this scenario.

2. Methods

As depicted in Figure 1 and Figure 2, we use a Geiger counter attached to the Logger Pro Interface via BNC cable, which itself is attached to a computer with Logger Pro. The Geiger counter sits on top of a metal box with slits along
the side, which allows us to slide in a tray-like aluminum plate with an adhered sample of Cesium-137 at various controlled distances from the Geiger counter.

Figure 1: Picture of experimental setup.

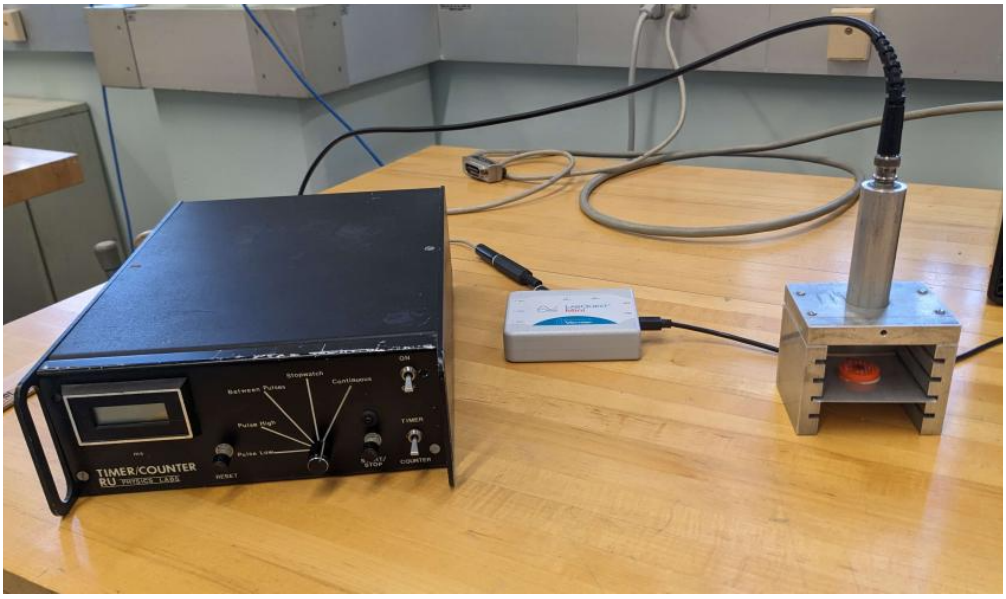
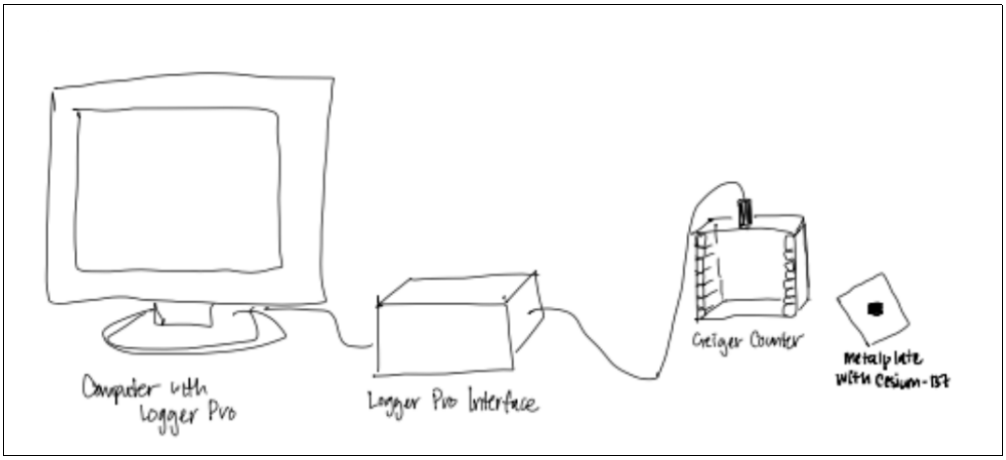


Figure 2: Drawing of experimental setup, labelling key components.



For most trials, we place the Cesium-137 sample at second slot from bottom, at which distance the measured decay is 70 counts/second. This position was chosen to achieve a count rate less than but near 100 counts/second to ensure no
counts are lost.

All data analysis, calculations, and plotting can be recreated by running the following Python notebook in browser: Google Colab (also attached as .ipynb file).

---

3. Part 1/2: Procedure & Analysis (section-by-section)

We denote frequently used symbols in Table 1 for reference.

Table 1: Frequently used symbols/values.

Symbol	Meaning
t	Total run time of experiment
Δt	Measurement interval, i.e. $(\Delta t)^{-1}$ is measurements/second
n	Number of data points $(t/\Delta t)$
N	Radiation counts per interval
$\overline{N}$	Mean of distribution of N
σ_N	Standard deviation of distribution of N
R	Decay rate in counts/second

3.1. Section C (Runs 1-2)

We want to get our first look at the distribution of measured radiation counts per interval in one run. To produce data for this run, the parameters we specified in Logger Pro (as well as summary statistics of the resulting data) are specified in Table 2.

Table 2: Logger Pro parameters and summary statistics of resulting data for Run 1.

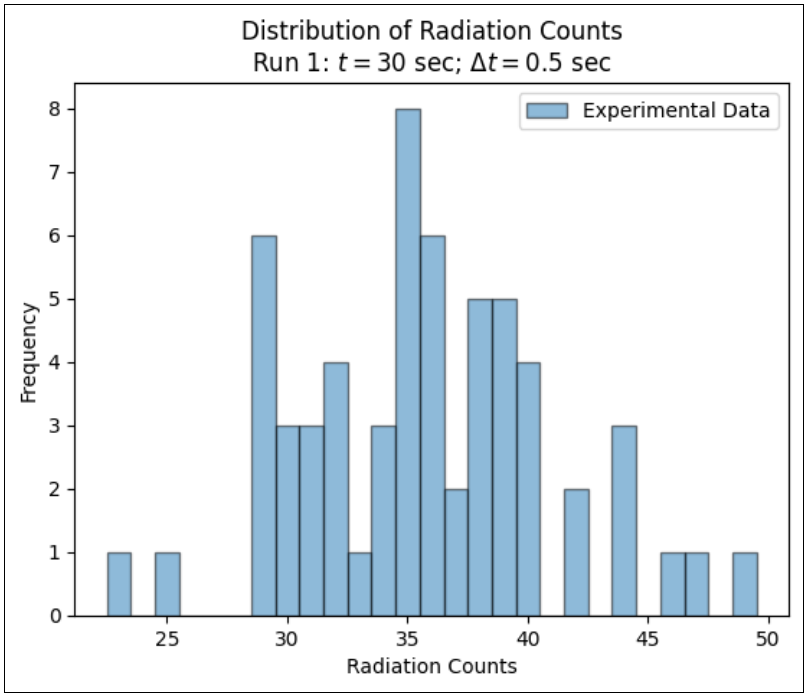
Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
30	0.5	60	36	5

We use these values to estimate R_0, the true value of the decay rate in units of counts per second (cps), with

$$R = \frac{\overline{N}}{\Delta t} \pm \frac{\sigma_N}{\Delta t \sqrt{n}}$$
$$= 71 \pm 1 \text{ cps.}$$

In order to visualize the shape of the distribution of N, we plot a histogram with bin width 1 in Figure 3.

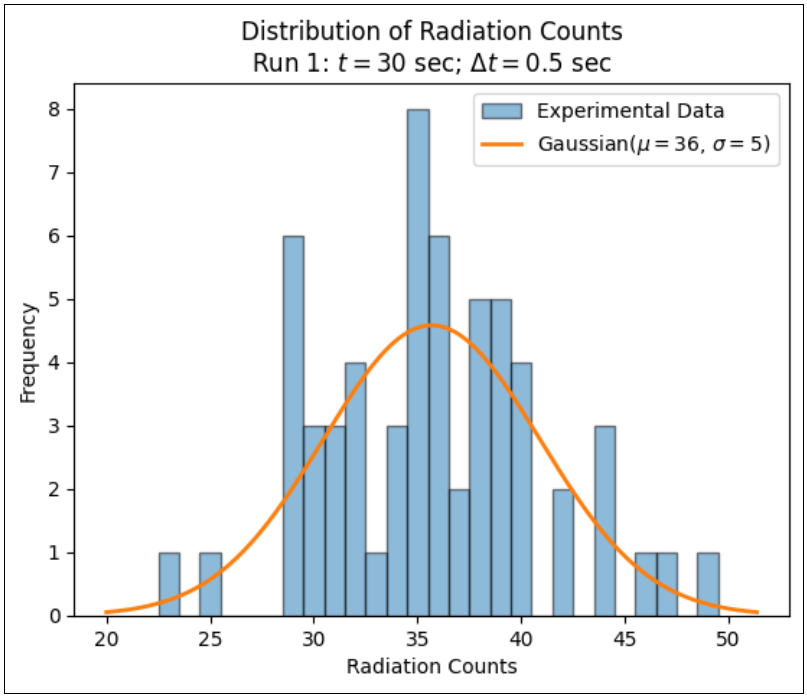
Figure 3: Distribution of radiation counts for Run 1.



To evaluate if this distribution resembles a Gaussian (which is the shape we'd expect for radiation counts), we overlay

a normalized (scaled by n) Gaussian curve centered on $\overline{N}$ in Figure 4.

Figure 4: Distribution of radiation counts and overlaid normalized Gaussian based on summary statistics for Run 1.



The experimental and theoretical distributions don't line up very well, so we can't be confident data from this run resembles a Gaussian. The likely source of this lack of agreement is our relatively small sample number ($n = 60$). In the next run, we tweak this factor.

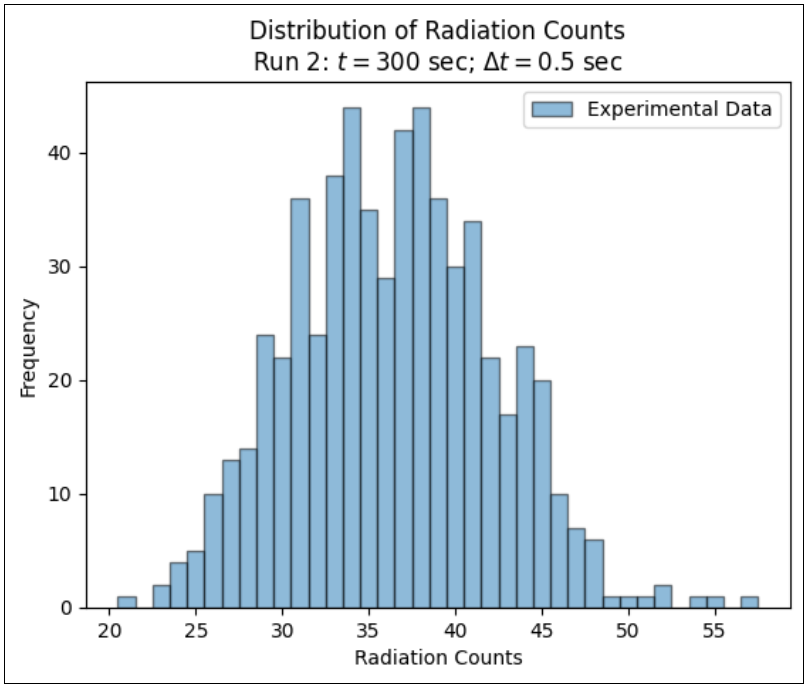
We use similar Logger Pro parameters (namely $\Delta t = 0.5$ s), but we increase run time from $t = 30$ s to $t = 300$ so that our total number of collected data points increases from $n = 60$ to $n = 600$. All parameters as well as summary
statistics for the resulting dataset are listed in Table 3.

Table 3: Logger Pro parameters and summary statistics of resulting data for Run 2.

Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
300	0.5	600	36	6

We visualize the shape of distribution of N by plotting a histogram with bin width 1 in Figure 5.

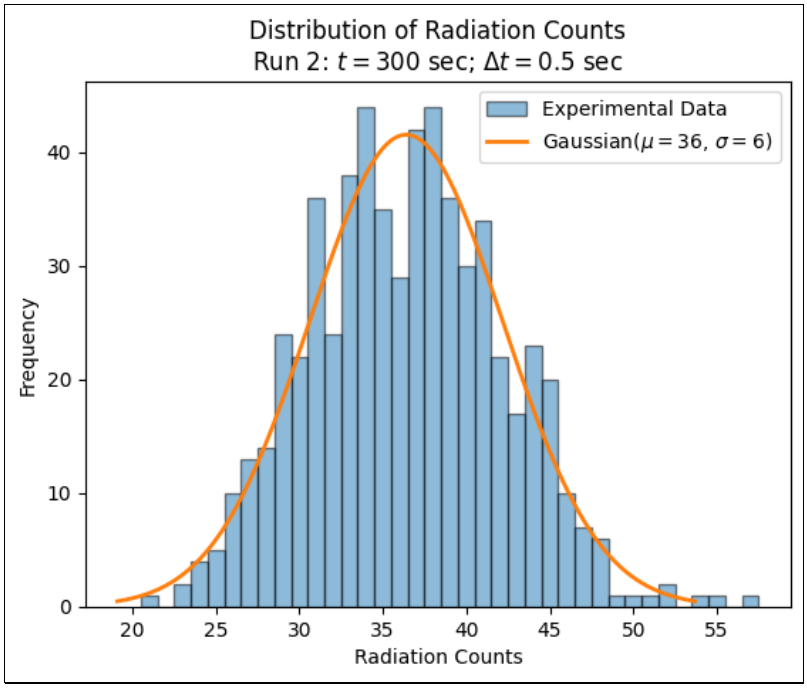
Figure 5: Distribution of radiation counts for Run 2.



To evaluate if this distribution resembles a Gaussian (which is the shape we'd expect for radiation counts), we overlay

a normalized (scaled by n) Gaussian curve centered on $\overline{N}$ in Figure 6.

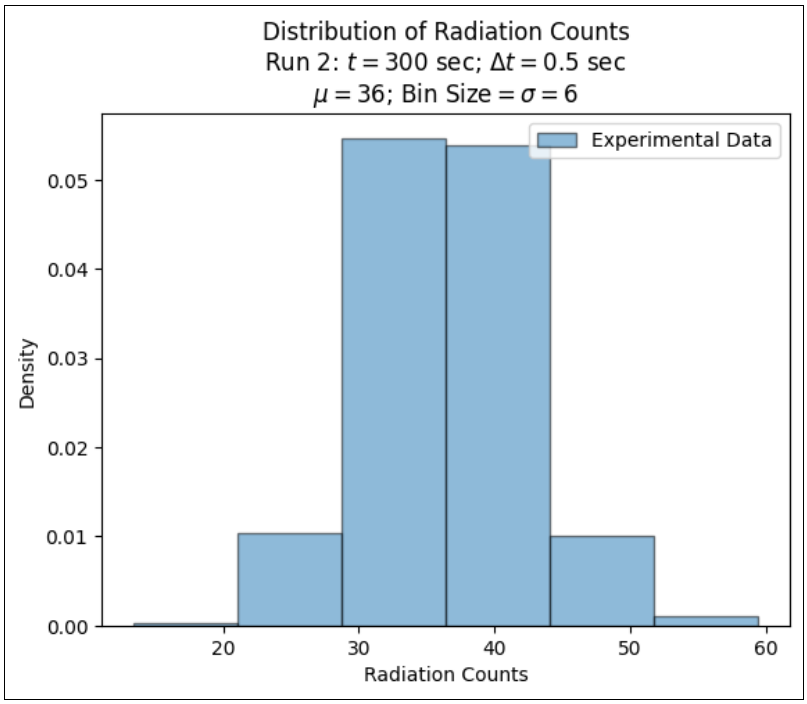
Figure 6: Distribution of radiation counts and overlaid normalized Gaussian based on summary statistics for Run 2.



As expected, we see that increasing total measurement time, and thus total number of samples, has significantly increased resemblance between distribution of N and Gaussian centered on $\overline{N}$.

Now that we're confident the distribution of N is Gaussian, we have additional context to interpret the standard deviation, specifically the probabilitiy of mesauring certain values within certain ranges. Analysis reveals that 414/600 or 69.0% of values fall within $\pm 1\sigma$ of $\overline{N}$, which is very close to the theoretical prediction of 68% of a Gaussian function. This is visually apparent in the coarse-binned and area-normalized histogram in Figure 7 by multipying base and height of central two rectangles.

Figure 7: Distribution of radiation counts for Run 2, with wide bins and normalized area/density to demonstrate Gaussian standard deviation percentages.



3.2. Section D (Run 3)

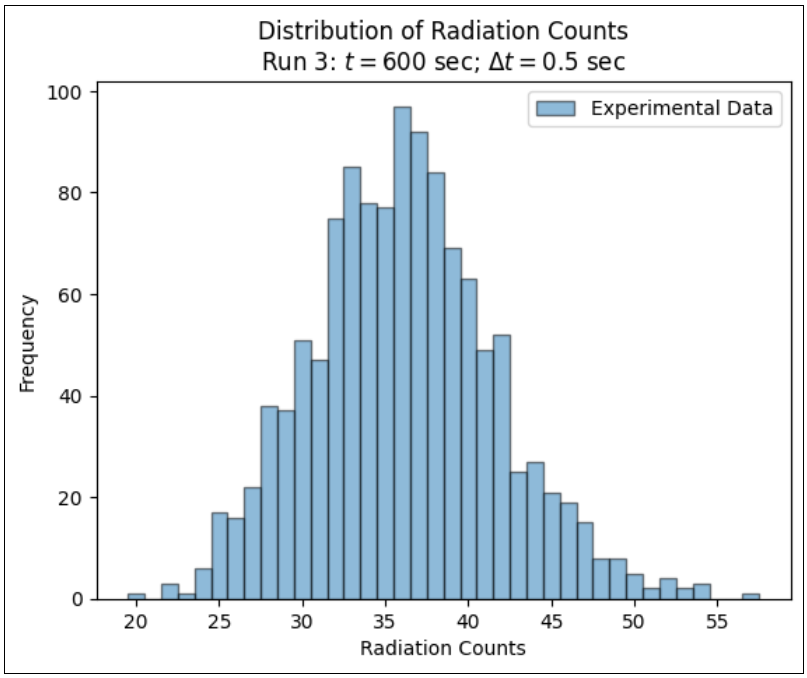
Following from the previous Section 3.1., we use similar Logger Pro parameters (namely $\Delta t = 0.5$ s), but further increase run time and thus number of samples by a factor two. All parameters as well as summary statistics for the resulting dataset are listed in Table 4.

Table 4: Logger Pro parameters and summary statistics of resulting data for Run 3.

Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
600	0.5	1200	36	6

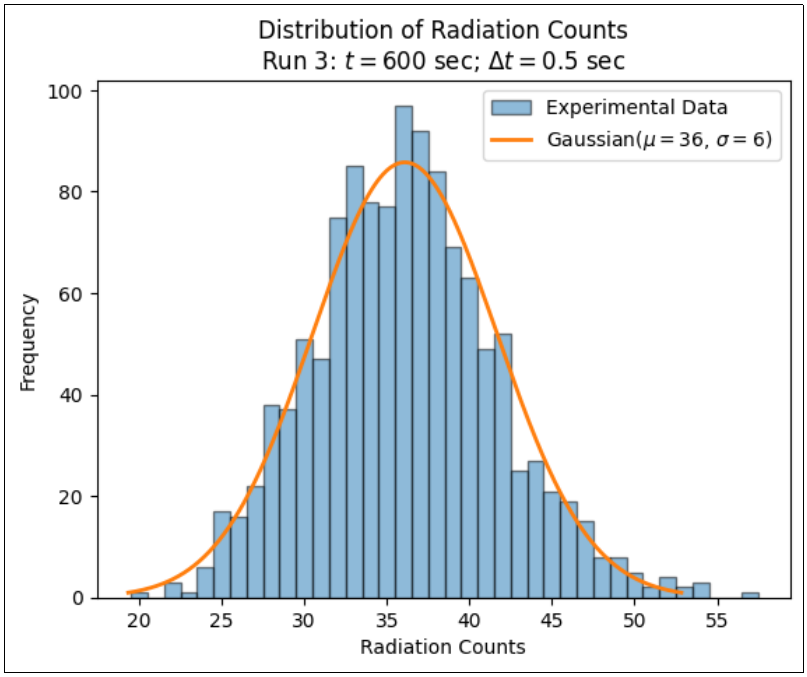
We visualize the shape of distribution of N by plotting a histogram with bin width 1 in Figure 8.

Figure 8: Distribution of radiation counts for Run 3.



We overlay a normalized (scaled by n) Gaussian curve centered on $\overline{N}$ in Figure 9.

Figure 9: Distribution of radiation counts and overlaid normalized Gaussian based on summary statistics for Run 3.



Once again, as expected, we see that increasing total measurement time and thus total number of samples has significantly increased resemblance between distribution of N and Gaussian centered on $\overline{N}$.

3.3. Section E (Run 3 cont.)

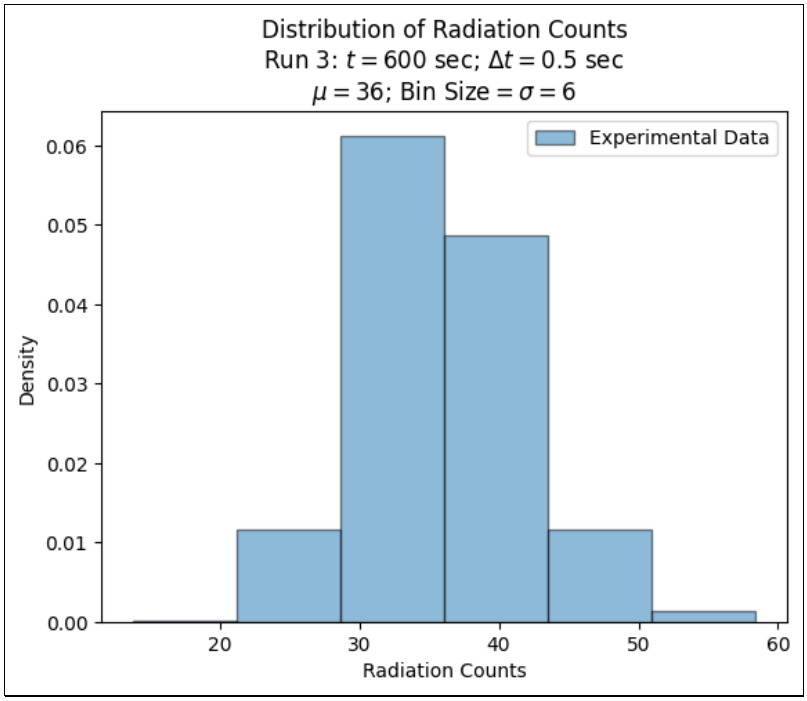
To further study the distribution of N in very large samples, we keep using data from Run 3 and calculate what percent of data falls within each standard deviation from the mean, and compare to calculated value for a theoretical Gaussian distribution in Table 5.

Table 5: Comparison of % data in each standard deviation for Run 3.

Interval	**% Data Contained in Interval, Experimental Value**	**% Data contained in Interval, Theoretical/Gaussian Value**
$\pm 1\sigma$	68.00 %	68.00 %
$\pm 2\sigma$	96.33 %	95.40 %
$\pm 3\sigma$	99.50 %	99.70 %

This is visually apparent in the coarse-binned and area-normalized histogram in Figure 10 by multipying base and height of central rectangles.

Figure 10: Distribution of radiation counts for Run 3, with wide bins and normalized area/density to demonstrate Gaussian sigma percentages.



In the previous Section 3.2. we concluded the distribution of N is Gaussian by visually comparing it with an overlaid Gaussian — in this section, we've provided quantitative evidence in the form of comparing the percent of data that falls in each interval between our distribution and an ideal Gaussian function.

— —

3.4. Section F (Run 4)

So far, the distribution of measurements of N can be accurately described by Gaussian (i.e. parametrized by mean and standard deviation) beceause $\overline{N}$ is relatively large. But what happens when $\overline{N}$ is very small (i.e. close to one)? Gaussian is a poor theoretical approximation, since we might predict some negative counts which is unphysical. Thus we turn to Poisson distributions.

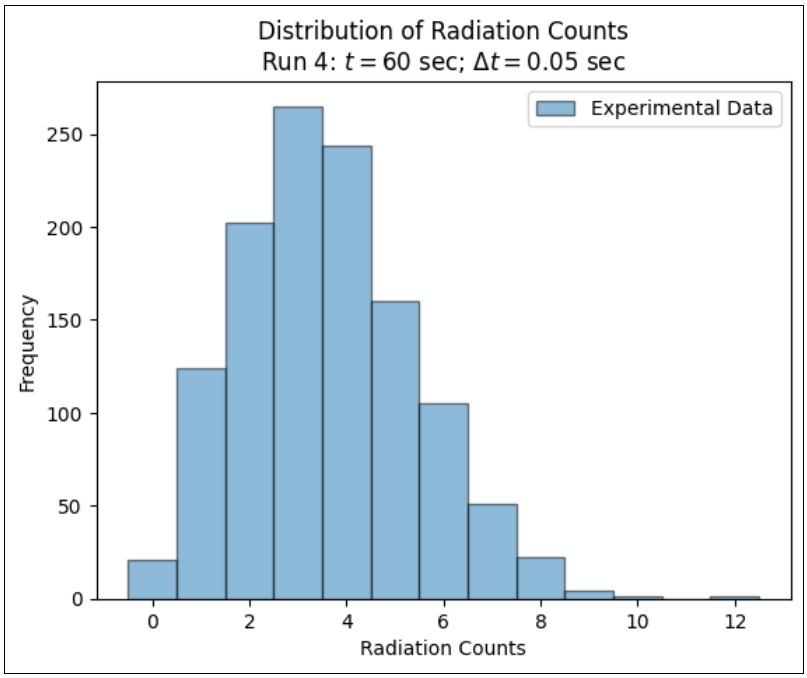
To reach the limit/conditions where Guassian fails and Poisson prevails, we decrease measurement interval from $\Delta t = 0.5$ sec to $\Delta t = 0.05$ sec. To match previous sample size $n = 1200$, we have a run time of $t = 60$ sec. Parameters and resulting summary statistic for this run are listed in Table 6.

Table 6: Logger Pro parameters and summary statistics of resulting data for Run 4.

Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
60	0.05	1200	4	2

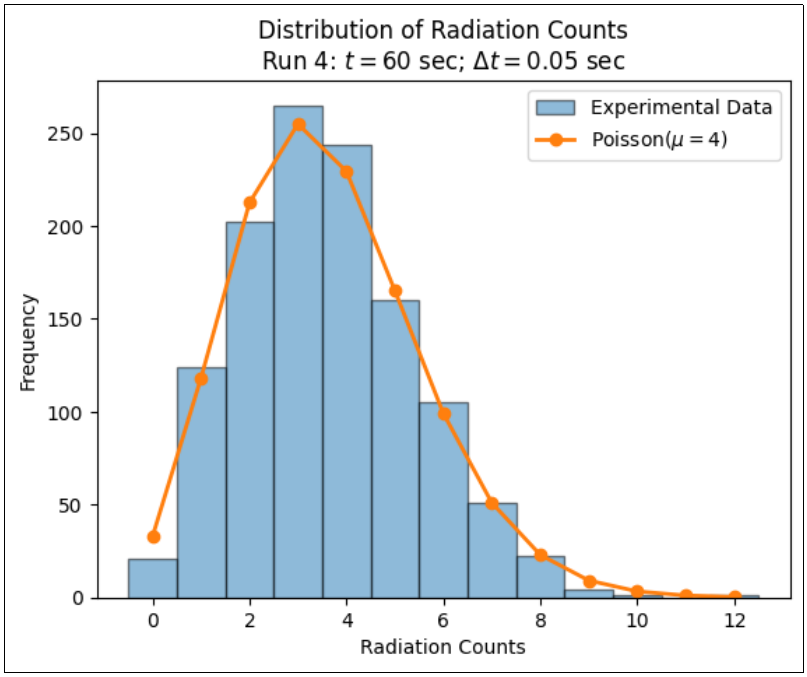
We visualize the shape of distribution of N by plotting a histogram with bin width 1 in Figure 11.

Figure 11: Distribution of radiation counts for Run 4.



To evaluate if this distribution resembles a Poisson function, we overlay a normalized (scaled by n) Poisson curve centered on $\overline{N}$ in Figure 12.

Figure 12: Distribution of radiation counts and overlaid normalized Poisson based on summary statistics for Run 4.



As expected, we see that mesauring a scenario where $\overline{N}$ is very small has increased resemblance between distribution of N and Poisson centered on $\overline{N}$.

3.5. Section G (Run 5)

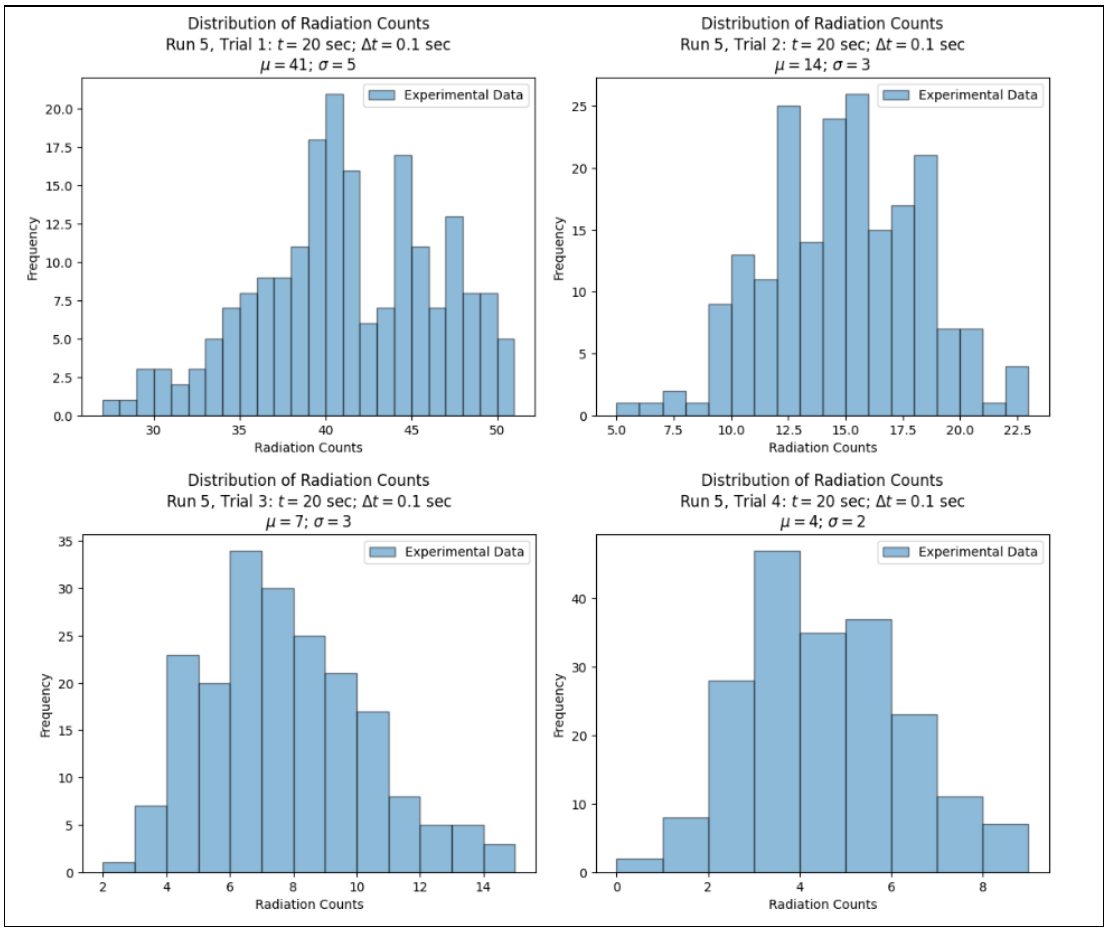
We now seek to observe how the distribution of N transitions from Gaussian to Poisson as $\overline{N}$ decreases, which is achieved by moving the radiation further away from the Geiger counter while keeping the same sampling settings. Specifically, we have have 4 trials — starting at first (top-most) slot, then moving to fourth (bottom-most) slot. All parameters as well as summary statistics for the resulting dataset are listed in Table 7.

Table 7: Logger Pro parameters and summary statistics of resulting data for Run 5 trials 1-4.

Trial	**Logger Pro Parameters**			**Summary Stats of Resulting Data**	
	t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
1	20	0.1	200	41	5
2	20	0.1	200	14	3
3	20	0.1	200	7	3
4	20	0.1	200	4	2

We visualize the shape of each resulting distribution of N by plotting a histogram with bin width 1 in Figure 13.

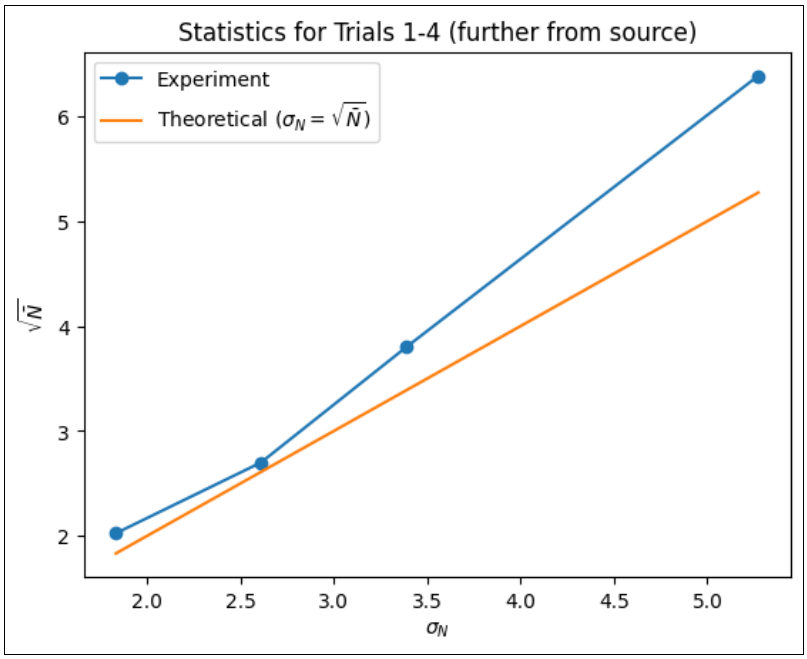
Figure 13: Distribution of radiation counts for Run 5, trials 1-4.



All of these appear to be Gaussian, with $\overline{N}$ roughly aligning with the peak bin — this makes sense in the context of $\overline{N} \gg 1$.

We can use these four datasets to prove one of our theoretical assumptions. We know that the standard deviation of the Poisson function is $\sigma = \sqrt{\mu}$ (given in the lab manual, various textbooks, as well as inherent upon inspecting/comparing the formulas themselves). For large μ, the Poisson function becomes almost symmetric and can be approximated with a Gaussian function. For the four trials/datasets from this run, we can plot each $\sqrt{\overline{N}}$ vs σ (summarized in Table 7) which is shown in Figure 14.

Figure 14: Evaluating validity of $\sigma = \sqrt{\mu}$ assumption using datasets from trials 1-4 in run 5.



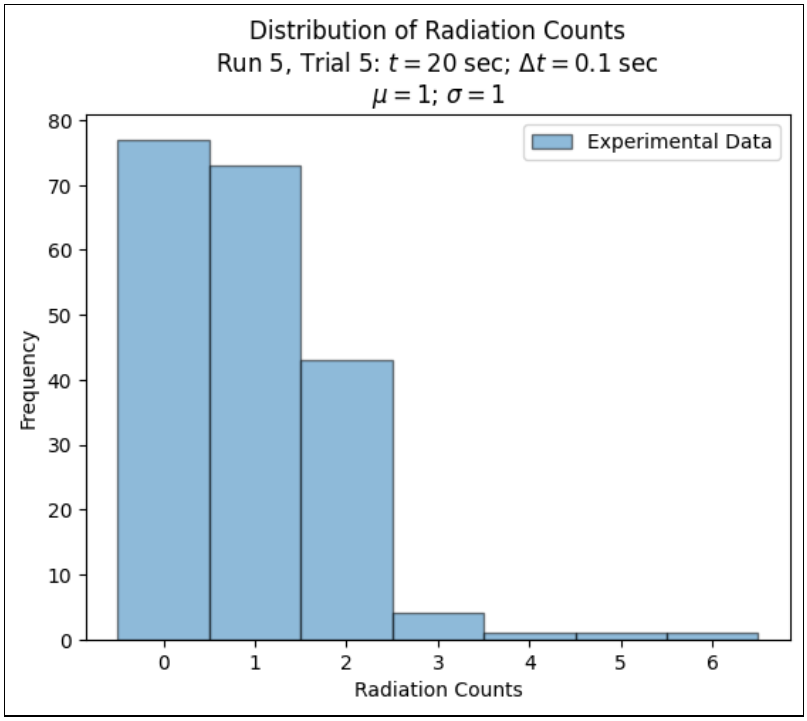
We notice that even in trial 4 where the sample is furthest away from the Geiger counter, we still have $\overline{N} > 1$ which results in a Gaussian distribution. In order to force a dataset with Poisson distribution with the same Logger Pro sampling settings, we flip the sample plate such that radiation will need to pass through a thin aluminum plate to reach the Geiger counter, which stops a significant portion of radiation. All parameters as well as summary statistics for the resulting dataset are listed in Table 8.

Table 8: Logger Pro parameters and summary statistics of resulting data for flipped sample.

Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
20	0.1	200	0.9	0.9

We visualize the shape of distribution of N by plotting a histogram with bin width 1 in Figure 15.

Figure 15: Distribution of radiation counts for Run 5 with flipped sample.



As expected, with all else equal (sampling rate/time/number), once the value of $\overline{N} < 1$, the final distribution of N is Poisson.

---

4. Part 2/2: Procedure & Analysis (section-by-section)

Note that the sample and instruments used in this section are different from Section 3., so the strength/decay rate of the radioactive source is likely different. As such, some runs are repeated and average values might be different for the same placement.

4.1. Section C (Run 1)

Repeating Section 3.1., we want to get our first look at the distribution of measured radiation counts per interval in one run. To produce data for this run, the parameters we specified in Logger Pro (as well as summary statistics of the
resulting data) are specified in Table 9.

Table 9: Logger Pro parameters and summary statistics of resulting data for Run 1.

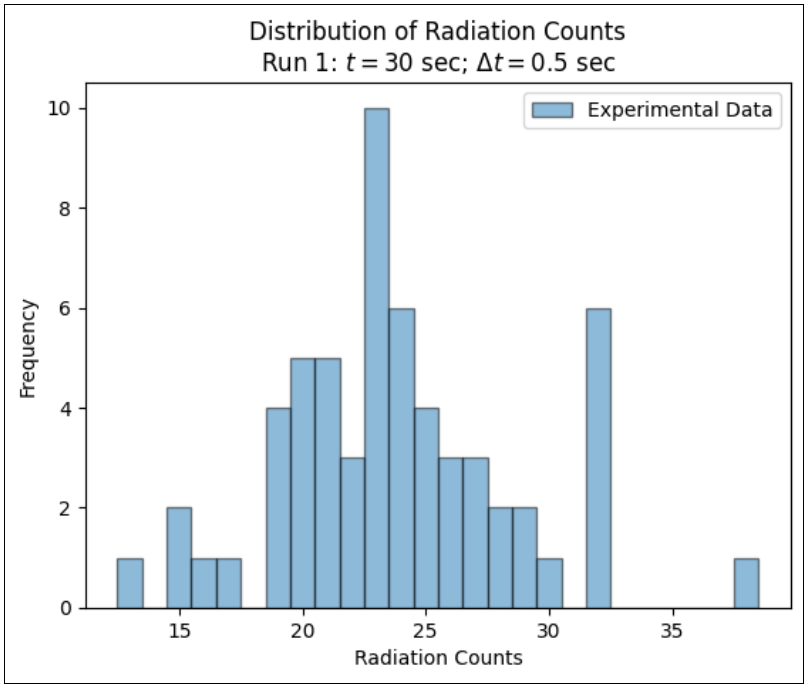
Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
30	0.5	60	24	5

We use these values to estimate R_0, the true value of the decay rate in units of counts per second (cps), with

$$R = \frac{\overline{N}}{\Delta t} \pm \frac{\sigma_N}{\Delta t \sqrt{n}}$$
$$= 48 \pm 1 \text{ cps.}$$

In order to visualize the shape of the distribution of N, we plot a histogram with bin width 1 in Figure 16.

Figure 16: Distribution of radiation counts for Run 1.



As expected, this roughly resembles a Gaussian centered on $\overline{N} = 24$, with the vast majority of data falling withing $\pm 2\sigma = \pm 10$ of $\overline{N}$. However we emphasize this is quite rough, and previous runs with the same sample time/rate/size in Section 3.1. have been less reliable.

— —

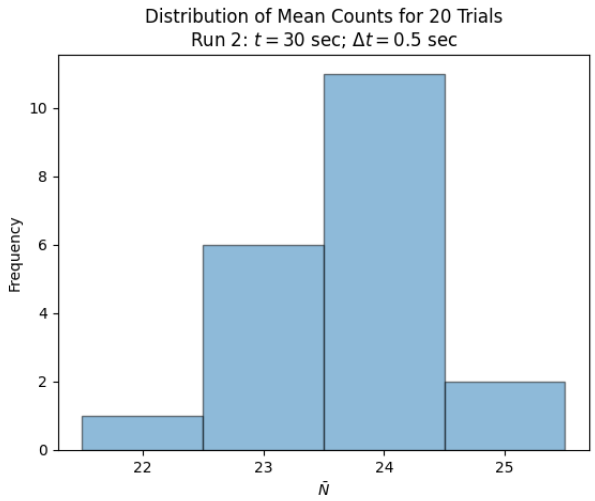
4.2. Section D (Run 2)

Now we seek to study the distribution of $\overline{N}$ in a set of runs. If a Gaussian function is an accurate model of radiation counts, then if we were to repeat a run manh times, there should be a 68% probability that the measured $\overline{N}$ falls within 68% of the true decay rate in that interval, N_0.

We test this hypothesis by repeating the previous run (settings specified in Table 9) 20 more times.

We visualize the distribution of $\overline{N}$ by plotting a histogram with bin width 1 (which also matches the "coarse" bin width $2\sigma / \sqrt{60}$) centered on mean$\left(\overline{N}\right) = 24$ in Figure 17.

Figure 17: Distribution of $\overline{N}$ for 20 trials.



The high "clustering" about the mean supports hypothesis it's Gaussian, although it's a little difficult to interpret since there are only 20 values of $\overline{N}$.

We use these values to estimate R_0, the true value of the decay rate in units of counts per second (cps), with

$$\begin{aligned} R &= \frac{\overline{N}}{\Delta t} \pm \frac{\sigma_N}{\Delta t \sqrt{n}} \\ &= 47 \pm 1 \text{ cps.} \end{aligned}$$

Rather than doing 20×30 second trials, we can do a single 600 second run (still measuring 1200 data points in the end) and compare the results. The parameters we specified in Logger Pro (as well as summary statistics of the resulting data) are specified in Table 10.

Table 10: Logger Pro parameters and summary statistics of resulting data for Run 3.

Logger Pro Parameters			**Summary Stats of Resulting Data**	
t [s]	Δt [s]	n	$\overline{N}$ [counts/interval]	σ_N [counts/interval]
600	0.5	1200	24	5

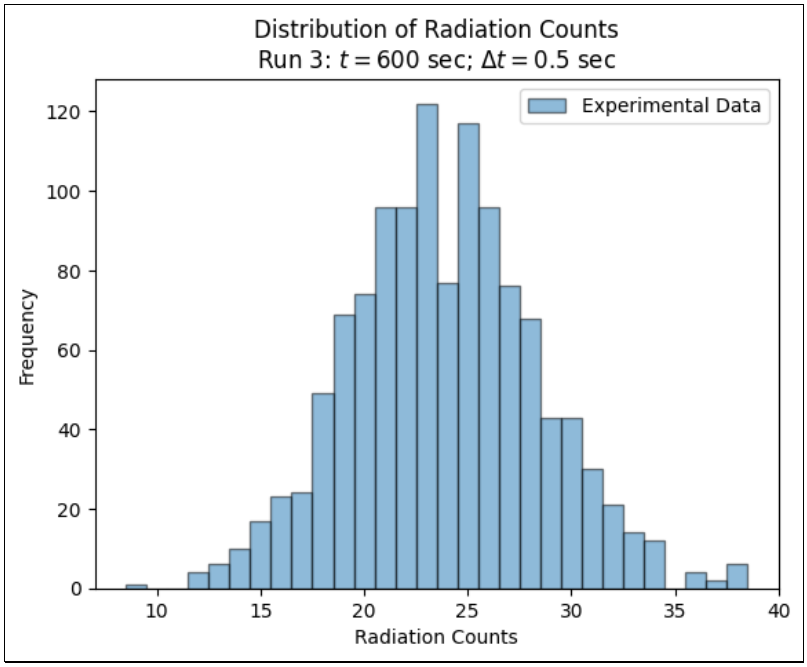
We use these values to estimate R_0, the true value of the decay rate in units of counts per second (cps), with

$$R = \frac{\overline{N}}{\Delta t} \pm \frac{\sigma_N}{\Delta t \sqrt{n}}$$

$$= 48 \pm 0 \text{ cps}.$$

We have much higher precision here. We visualize the shape of distribution of N by plotting a histogram with bin width 1 in Figure 18.

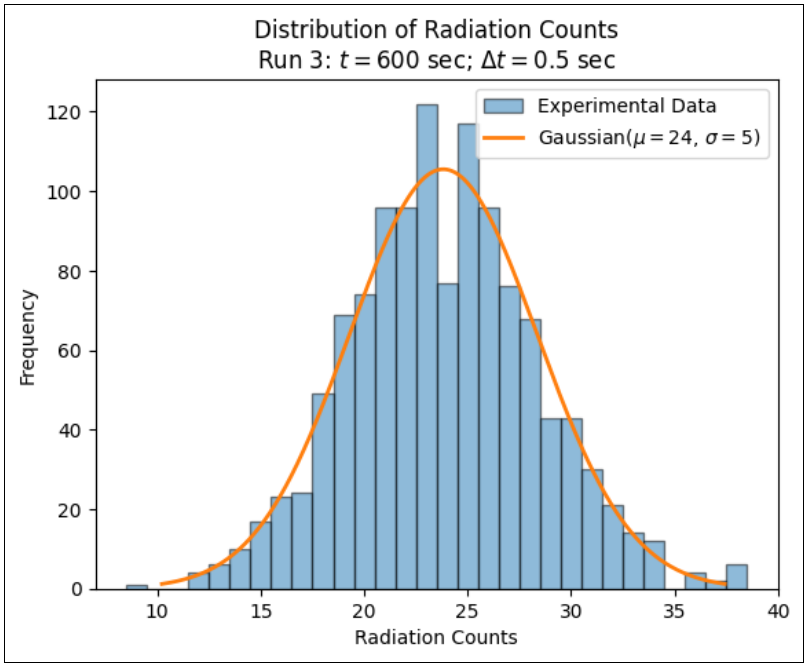
Figure 18: Distribution of radiation counts for Run 3.



To evaluate if this distribution resembles a Gaussian (which is the shape we'd expect for radiation counts), we overlay

a normalized (scaled by n) Gaussian curve centered on $\overline{N}$ in Figure 19.

Figure 19: Distribution of radiation counts and overlaid normalized Gaussian based on summary statistics for Run 3.



As expected, we see strong resemblance with a Gaussian function centered on $\overline{N}$.

— —

4.3. Section E (Run 4-13)

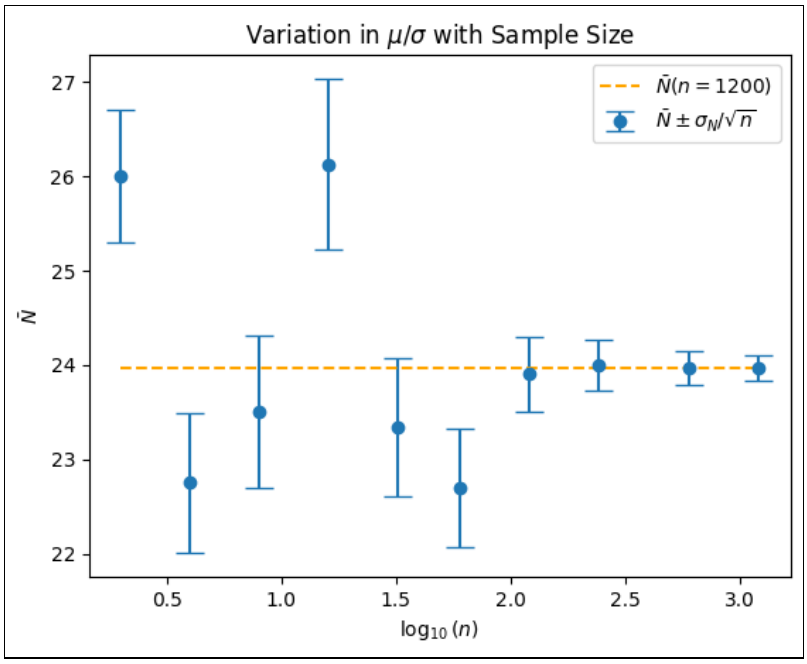
Now we aim to study the variation in mean and standard deviation of mean as sample size varies. The parameters we specified in Logger Pro (as well as summary statistics of the resulting data) are specified in Table 11.

Table 11: Logger Pro parameters and summary statistics for various sample sizes

Trial	Logger Pro Parameters			Summary Stats of Resulting Data	
	t [s]	Δt [s]	n	$\overline{N}$	σ_N
1	600	0.5	1200	24	4
2	300	0.5	600	24	4
3	120	0.5	240	24	4
4	60	0.5	120	24	4
5	30	0.5	60	23	5
6	16	0.5	32	23	4
7	8	0.5	16	26	4

We notice that σ_N stays mostly stable, up until trial 8 when the sample size is too small for the results to be meaningful. However, we expect the uncertainty of the mean $\sigma_N / \sqrt{n}$ to decrease as the sample size increases. We can visualize this by plotting $\overline{N}$ vs $\log_{10}(n)$ with error bars on $\overline{N}$ for $\pm\frac{\sigma_N}{\sqrt{n}}$ in Figure 20.

Figure 20: Variations in $\overline{N} \pm \left(\sigma_N / \sqrt{n}\right)$ with increasing sample size.



As expected, our measurements become both more accurate and precise as sample size becomes significantly larger. We see that 4/10 measurements fall outside of our best estimate of N_0 by more than one standard deviation. For a Gaussian distribution there is a 32% chance (or 3-4 out of 10) that a single measurement will fall outside of $\mu \pm \sigma$, which nearly matches our observed values.

---

5. Conclusion

In this lab, we used a Geiger counter to measure radiation counts over small intervals while varying the sampling time/interval/rate/size in order to study the Gaussian and Poisson distributions.

We observed that for $\overline{N} \gg 1$, increasing the sample size increases resemblance between distribution of the data and a Gaussian function. This is consistent with the Central Limit Theorem, which states that the sum of a large number of independent and identically distributed random variables will be approximately normally distributed, regardless of

the original distribution of the variables. The importance/applicability of this came into play when we explored of the properties of Gaussian functions, particularly the predictable nature of data distribution in relation to standard deviations. By leveraging these properties, we can make informed predictions about the range in which a certain percentage of data points will fall.

We also observed that when $\overline{N} \lesssim 1$, our distribution is better represented by a Poisson function, which is only parameterized by the mean of the data. Similar to our study of the Gaussian function, the importance here is understanding the probability of a random data point falling within a specific range.

In general, the value of finding an appropriate Gaussian/Poisson function based on some distribution of data is the ability to predict how often data should fall within a specific range, as well as identifying when the underlying process
you're studying may not actually be normal.