Chapter Six: Counting Statistics and Uncertainty

The fundamental idea that we can describe stochastic processes precisely (given enough samples of the process) despite the intrinsic randomness is an extremely powerful concept when working with experimental data. What this means is that although we cannot know the exact outcome in any given experiment, we can know the range of expected values well enough to be able to make meaningful predictions about the outcome of future experiments. More importantly, this means that we can say with a great deal of confidence that our predictions were incorrect if our experiment yields results outside of those expected by our predictions. However, we cannot quantify what it means for a measurement to be outside of our predictions without the ability to quantify how much variation we expect in our experimental measurements as well.

Experimental uncertainties encapsulate this concept, with the underlying assumption being that fundamental properties of data will follow a Poisson or Gaussian distribution for which the measured mean and standard deviation are directly related to the intrinsic variance seen in data that follows distributions of this kind.

6.1 From Theory to Measurement

In the preceding chapter, we developed the statistical framework for understanding stochastic processes. Starting from the coin flip, we built up the binomial distribution, and showed that in the appropriate limits this simplifies to the Poisson and then the Gaussian distribution. We also explored how experimental observations sample from an underlying probability distribution, and how the uncertainty on the experimental mean decreases as we perform more measurements. The goal of this chapter is to connect these tools directly to the stochastic process we care most about in this course: radioactive decay and the measurement of radiation.

6.1.1 Radioactive Decay as a Stochastic Process

Recall from our earlier discussion of radioactive decay that the total number of radioactive atoms in a sample at time t is given by the exponential decay law, $N(t) = N_0 e^{-\lambda t}$, and that the rate of decays, the activity, is $A = \lambda N(t)$, where λ is the decay constant. The key quantity for our statistical analysis is the **probability of a single atom decaying** in some time interval Δt. From the decay law, this probability is:

$$p = \lambda \Delta t$$

For any reasonable measurement interval, this probability is extremely small. Consider a source of Cs-137, which has a half-life of about 30 years. The decay constant is $\lambda = \ln(2)/t_{1/2} \approx 7.3 \times 10^{-10}\ s^{-1}$. If we measure for $\Delta t = 2$ seconds, the probability of any single Cs-137 atom decaying during our measurement is $p = 7.3 \times 10^{-10} \times 2 \approx 1.5 \times 10^{-9}$. This is an incredibly small number, meaning any individual atom is overwhelmingly likely to survive through our two-second measurement window.

At the same time, even a modest radioactive source contains a very large number of atoms. A source with an activity of 1 kBq (1000 decays per second) of Cs-137 contains $N = A/\lambda \approx 1.4 \times 10^{12}$ atoms – more than a trillion atoms, each independently capable of decaying during our measurement.

This situation – a very large number of independent trials (atoms), each with a very small probability of a particular outcome (decaying) – is exactly the binomial distribution problem from Chapter 5. Each atom is like a coin flip, where "heads" means the atom decays and "tails" means it does not. The number of decays we observe in a time interval is the number of "heads" out of N flips, each with probability p. From Chapter 5, we know that the binomial distribution is the correct description for this process, and that the expected number of decays is simply:

$$\bar{x} = Np = N\lambda\Delta t = A\Delta t$$

This should be reassuring – the expected number of decays observed in a time interval is just the activity multiplied by the time, which is exactly what we would expect.

6.1.2 Why Counting Statistics are Poisson and Gaussian

Having established that radioactive decay is a binomial process, we can now identify which of our simplified distributions applies. From Chapter 5, the Poisson distribution applies when $p \ll 1$, and the Gaussian distribution applies when additionally N is large enough that $Np \gg 1$. For radioactive decay:

$$p = \lambda\Delta t \ll 1 \rightarrow \text{Poisson statistics apply}$$

$Np = A\Delta t$, which for any measurable activity and reasonable counting time will be much larger than 1 → Gaussian statistics apply

There is an additional stochastic component to consider when making radiation measurements. The radiation emitted by a decaying atom must also interact in our detector in order to be measured, and this interaction is itself a random process with some probability. Not every gamma ray emitted by a source will interact in the detector – it must travel in the right direction, not be absorbed before reaching the detector, and then actually interact within the detector material. The probability of detection is often referred to as the detection efficiency, ε, and is typically much less than one. The number of counts we actually observe is therefore the product of the number of decays and this detection efficiency. Importantly, this additional random process does not change the statistical nature of our measurements. The detection process is itself a binomial trial – each emitted particle either interacts in the detector or it does not – and the combination of two binomial processes is still a binomial process, just with a

smaller effective probability. The end result is that our measured counts still follow Poisson (and by extension, Gaussian) statistics[24].

6.1.3 Applying the Model to Counting Data

Now consider the example from Chapter 5 of counts collected in two-second intervals using a gamma radiation detector. Each two-second interval gives us a single measurement of the number of radiation interactions, or counts, that occurred in the detector during that time. From the full set of these measurements, we can calculate the experimental mean and standard deviation using Eq. 8 and 9. The experimental mean of the number of counts per interval gives us an estimate of the rate of radiation interactions in our detector. But how do we characterize the uncertainty in this estimate? And if we change something about our measurement – say, by counting for longer – how does this affect our mean and our uncertainty? These are the practical questions that counting statistics addresses.

We have just shown that radiation counting measurements follow Gaussian statistics, for which the mean and the standard deviation are related:

$$\mu = \bar{N}_{counts}, \quad \sigma = \sqrt{\mu} = \sqrt{\bar{N}_{counts}}$$

This is a remarkable and powerful result. If we measure a mean number of counts, we automatically have an estimate of the standard deviation of the distribution those counts are drawn from. Our measured mean can therefore be used as the input for a Gaussian model of our data, and the standard deviation of that model is simply the square root of the mean. This means that we can predict the expected point-to-point variation in our measurements based on only the mean. For example, if our mean number of counts per interval is 84, we would expect a standard deviation of approximately $\sqrt{84} \approx 9.2$ counts. This tells us that the majority of our individual two-second measurements should fall within about 9 counts of 84, and we can use this prediction to check whether our data is behaving as expected.

[24] When the combined probability of decay and detection is very small, the Poisson distribution is the more accurate description of the counting statistics. However, since the mean and standard deviation have the same relationship in both the Poisson and Gaussian cases ($\sigma = \sqrt{\mu}$), the distinction is often not practically important for the purposes of uncertainty estimation.

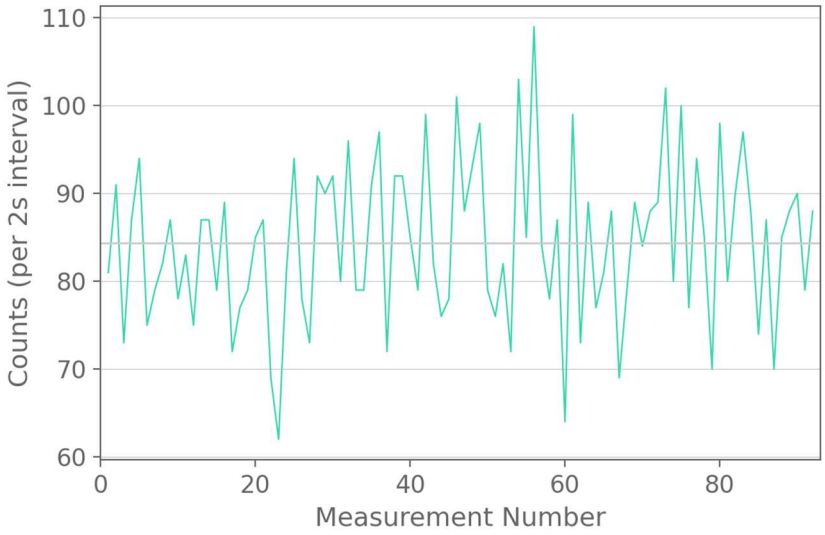

Figure 6.1: 92 measurements of counts per two-second interval from a gamma radiation detector. The data fluctuates around a mean of approximately 84 counts, with point-to-point variation consistent with counting statistics.

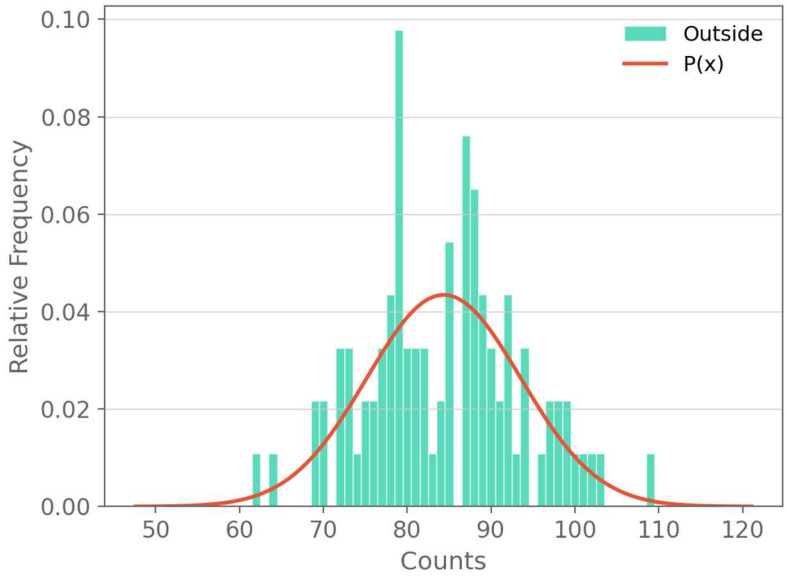

Figure 6.2: Frequency distribution of the counting data from Figure 6.1, with a Gaussian model (red curve) using $\mu = 84.4$ and $\sigma = \sqrt{\mu} = 9.2$.

6.2 Uncertainties on Individual Measurements

The relationship between the mean and standard deviation for Gaussian counting statistics extends to individual measurements as well. Each measurement of N_i counts can be thought of as a single sample from a Gaussian distribution with $\mu = N_i$ and $\sigma_i = \sqrt{N_i}$. While we

obviously do not know the true mean from a single measurement, we can use the single measurement as our best estimate of the mean, and the square root of that measurement as our estimate of the uncertainty. This allows us to assign an uncertainty to every individual counting measurement:

$$\sigma_i = \sqrt{N_i} \qquad \text{Eq. 18}$$

This is an extraordinarily useful property. In most experimental contexts, estimating the uncertainty on a measurement requires repeating the measurement many times and calculating the standard deviation from the set of results. For radiation counting, we can estimate the uncertainty from a single measurement. If we measure 88 counts in a given interval, we can state that the uncertainty on that measurement is √88 ≈ 9.4 counts. If we measure 75 counts in the next interval, the uncertainty on that measurement is √75 ≈ 8.7 counts.

Graphically, we can represent these uncertainties as **error bars** on each data point. An error bar extending one standard deviation above and below a data point indicates the range within which we expect the true value to lie with roughly 66% probability. Looking at a time series of counting data with these error bars, we would expect that roughly one-third of the data points would have the line connecting adjacent points fall outside of their error bars, and the vast majority of points should be within two standard deviations of the mean. If this is not the case, i.e. if significantly more or fewer data points fall outside the expected range, it could indicate that the measurement conditions are not stable, or that the process being observed is not purely stochastic.

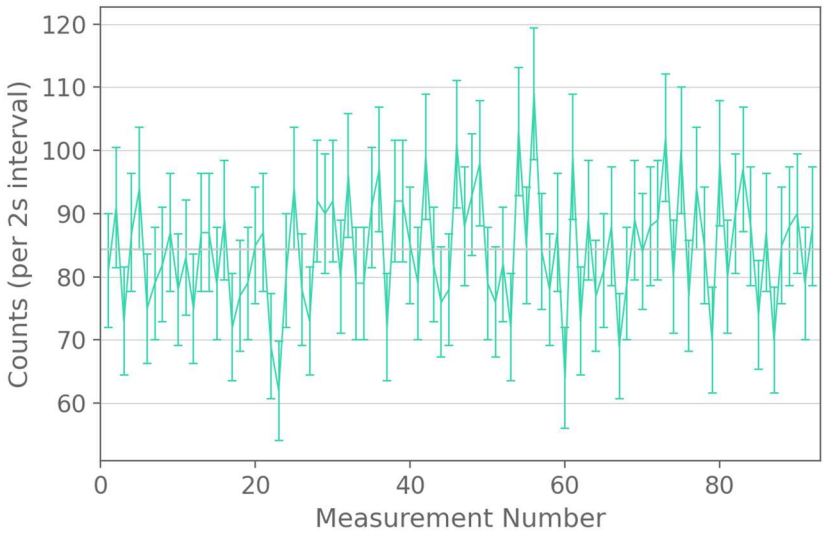

Figure 6.3: The same counting data as Figure 6.1, now shown with error bars representing $\sigma_i = \sqrt{N_i}$ for each data point. Roughly two-thirds of adjacent points should have overlapping error bars.

We can also compare this expected standard deviation from counting statistics to the experimental standard deviation calculated from the full set of data. The expected standard deviation is $\sigma = \sqrt{\mu} = \sqrt{\bar{N}_{counts}}$, while the experimental standard deviation is calculated from Eq. 9. For a well-behaved set of counting data, these two values should be close. If the experimental standard deviation is significantly larger than $\sqrt{\mu}$, it may indicate the presence of additional sources of variation beyond the inherent stochastic nature of the process, for example, a fluctuating source activity or an unstable detector. If it is significantly smaller, it may suggest that the data has been processed in some way that has artificially reduced the variation.

6.3 Combining Measurements and Propagation of Uncertainty

A natural question arises when we want to combine multiple counting measurements: what is the total uncertainty? Consider a simple example where we want to determine the total number of counts collected over several intervals. Suppose we have five measurements:

$x_1 = 88, \quad x_2 = 83, \quad x_3 = 75, \quad x_4 = 80, \quad x_5 = 77$

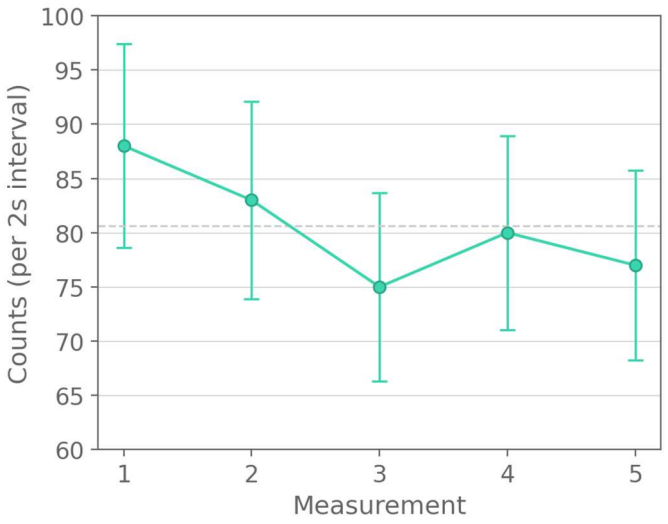

Figure 6.4: Five individual counting measurements with error bars ($\sigma_i = \sqrt{N_i}$). The dashed line shows the mean of the five measurements (80.6 counts).

Each measurement has an associated uncertainty: $\sigma_1 = \sqrt{88} = 9.38$, $\sigma_2 = \sqrt{83} = 9.11$, $\sigma_3 = \sqrt{75} = 8.66$, $\sigma_4 = \sqrt{80} = 8.94$, $\sigma_5 = \sqrt{77} = 8.77$. The total number of counts is $\Sigma N_i = 403$. What is the uncertainty on this total?

One might be tempted to simply add the individual uncertainties: $\Sigma\sigma_i = 9.38 + 9.11 + 8.66 + 8.94 + 8.77 = 44.86$. However, this dramatically overestimates the true uncertainty. The reason is rooted in the stochastic nature of the measurements. If one measurement happens to fluctuate high, the next is just as likely to fluctuate low, so the fluctuations tend to partially cancel when measurements are added together. Recall that only about one-third of individual measurements are expected to be more than one standard deviation from the mean, so it

would be incorrect to assume all measurements fluctuate to their maximum extent in the same direction.

The correct way to combine uncertainties is to add them **in quadrature**, meaning we add the squares of the individual uncertainties and take the square root of the sum:

$$\sigma_{total} = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_N^2} \qquad \text{Eq. 19}$$

For our example, this gives $\sigma_{total} = \sqrt{(9.38^2 + 9.11^2 + 8.66^2 + 8.94^2 + 8.77^2)} = \sqrt{(88 + 83 + 75 + 80 + 77)} = \sqrt{403} \approx 20.07$. Notice that the uncertainty on the total is simply the square root of the total number of counts – exactly what we would expect from counting statistics. This makes sense: the total of 403 counts is itself a single counting measurement of the total number of radiation interactions over the combined time interval, and the uncertainty on any counting measurement is the square root of the number of counts. Adding uncertainties in quadrature is therefore not just a mathematical convention; it is a natural consequence of the statistical properties of counting measurements.

Comparing the two approaches to combining uncertainties illustrates an important point: the linear sum of uncertainties (44.86) is more than twice the correct quadrature sum (20.07). Naively adding uncertainties overestimates the uncertainty because it assumes the worst case – that all measurements fluctuate in the same direction by their full uncertainty – which is statistically unlikely. Adding in quadrature accounts for the fact that random fluctuations partially cancel.

More generally, when combining any independent measurements, their uncertainties should be added in quadrature. This is a fundamental principle of **error propagation** that extends well beyond counting statistics, though the derivation is particularly transparent for counting measurements where the relationship between individual and total uncertainties can be verified directly.

6.4 Uncertainty on the Mean

In Chapter 5, we explored how the distribution of experimental means narrows as the number of samples increases, with the standard deviation of the mean decreasing as $\sigma_\mu = \sigma/\sqrt{N_{samples}}$ (Eq. 17). We can now derive this result directly using the machinery of error propagation we have just developed.

The experimental mean is defined as the sum of all measurements divided by the number of measurements:

$$\bar{x} = (1/N)(x_1 + x_2 + \ldots + x_N)$$

The uncertainty on the sum, from Eq. 19, is $\sigma_{total} = \sqrt{(\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_N^2)}$. For counting measurements, each $\sigma_i^2 = N_i$, so $\sigma_{total}^2 = \Sigma N_i = N_{total}$. Since the mean is the total divided by the number of measurements, N, the uncertainty on the mean is:

$$\sigma_{\bar{x}} = \sigma_{total} / N = \sqrt{N_{total}} / N$$

We can also express this as:

$$\sigma_{\bar{x}} = \frac{\sqrt{\bar{x}}}{\sqrt{N}} = \frac{\sigma}{\sqrt{N}} \qquad \text{Eq. 20}$$

This is exactly the result from Eq. 17, now derived from error propagation rather than from examining the distribution of experimental means. The two approaches give the same answer, which should not be surprising – they are both consequences of the **central limit theorem**, which states that the distribution of sample means will approach a normal distribution as the number of samples increases, with a standard deviation that decreases as 1/√N.

Returning to our counting example, if we have N = 92 measurements with a mean of $\bar{x}$ = 84.2 counts and a standard deviation of σ = √84.2 ≈ 9.18, the uncertainty on the mean is σ_μ = 9.18/√92 ≈ 0.96 counts. Even though any single two-second measurement has an uncertainty of about 9 counts, the mean of 92 measurements has an uncertainty of less than 1 count. This is the power of repeated measurement: while we cannot reduce the intrinsic randomness of each individual observation, we can dramatically reduce the uncertainty on our estimate of the underlying rate by performing many observations.

We should also note the distinction between the expected and experimental standard deviations. The expected standard deviation from counting statistics, σ = √μ, is based on our model for how counting measurements should behave. The experimental standard deviation from Eq. 9 is calculated directly from the data. For a small number of measurements, these may differ noticeably, since we are estimating the variance from limited sampling. As the number of measurements increases, the experimental standard deviation will converge toward the expected value. This convergence is itself a consequence of the central limit theorem – the experimental standard deviation, like the experimental mean, is a parameter being estimated from samples, and its uncertainty decreases with more data.

6.5 Count Rates

Thus far, we have been discussing uncertainties in terms of the total number of counts. In practice, it is often more useful to express measurements in terms of a count **rate** – counts per second (cps), or more generally counts per unit time. Since the count rate is simply the number of counts divided by the counting time, the conversion is straightforward. However, there is an important subtlety in how the uncertainty is propagated.

If we measure N_{counts} counts in a time interval Δt, the count rate is:

$$R = N_{counts} / \Delta t$$

The uncertainty on the count rate is determined by the uncertainty on the number of counts, divided by the same time interval:

$$\sigma_R = \frac{\sqrt{N_{counts}}}{\Delta t} \qquad \text{Eq. 21}$$

It is critical to note that $\sigma_R \neq \sqrt{R}$. This is a common error. The uncertainty on a counting measurement depends on the total number of counts collected, not on the rate. Dividing by the counting time to convert to a rate does not change the underlying statistics. To see why, consider a measurement of 102 counts in 10 seconds. The uncertainty on the number of counts is $\sqrt{102} \approx 10.1$ counts. The count rate is 10.2 cps, and the uncertainty on the count rate is 10.1/10 = 1.01 cps. If we mistakenly used $\sqrt{R}$, we would get $\sqrt{10.2} \approx 3.2$ cps, which is far too large.

A useful way to check this is through the **percent error**, which should not change when we scale a measurement. The percent error on the number of counts is:

$$\sigma/\mu = \sqrt{102}/102 = 10.1/102 \approx 10.1\%$$

The percent error on the count rate should be the same:

$$\sigma_R/R = 1.01/10.2 \approx 9.9\%$$

These agree[25], confirming that the uncertainty was propagated correctly. This invariance of the percent error under scaling is a general principle: dividing or multiplying a measurement by a constant (like the counting time) does not change the relative precision of the measurement.

This result has a practical implication: the way to improve the precision of a count rate measurement is to collect more total counts, which typically means counting for a longer time. Doubling the counting time approximately doubles the number of counts, which increases the absolute uncertainty by a factor of $\sqrt{2}$ but decreases the percent error by a factor of $\sqrt{2}$ (since the percent error is $\sqrt{N}/N = 1/\sqrt{N}$). To cut the percent error in half, we need to count four times as long.

6.6 Comparing Measurements

One of the most common and important applications of counting statistics is determining whether two measurements are **statistically different**. For example, suppose we want to determine whether the radiation level inside a building is different from the level outside. We set up a detector inside and collect data, then move it outside and collect more data. How do we determine whether any observed difference is real, or simply a result of statistical fluctuations?

Consider an initial measurement where we collect data in two-second intervals for a short period at each location. From the inside measurement, we find a mean count rate of $\bar{x}_I = 44$ cps, and from outside we find $\bar{x}_O = 42$ cps. The distributions of counts for these two measurements

[25] The slight difference is due to rounding. If we carry more decimal places, they agree exactly.

overlap significantly. Just looking at the histograms of counts, we cannot tell whether these measurements represent the same radiation level or two different levels.

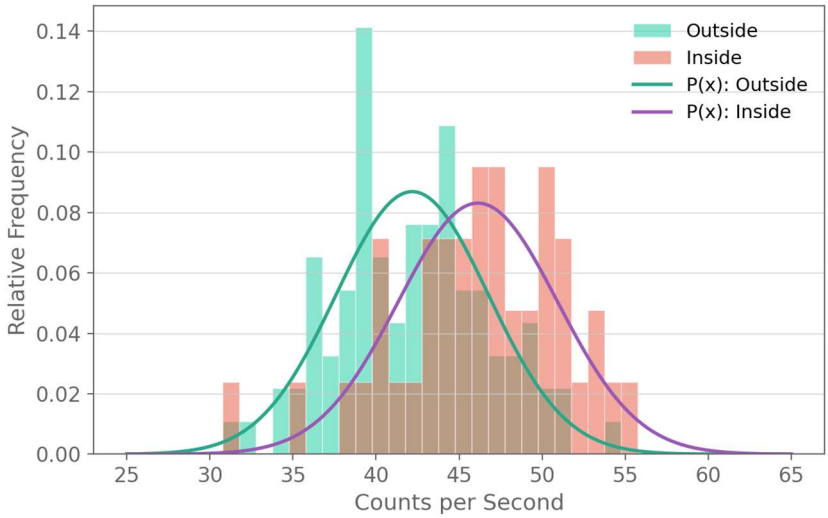

Figure 6.5: Histograms of inside and outside count rates (counts per second) with Gaussian model overlays. With limited data (N = 42 inside, N = 92 outside), the distributions overlap significantly.

To approach this more rigorously, we need to convert to total counts and determine the uncertainty on each mean. Working in counts for two-second intervals: $\bar{x}_I = 87.9$ counts with $\sigma_I = \sqrt{87.9} = 9.38$, and $\bar{x}_O = 84.2$ counts with $\sigma_O = \sqrt{84.2} = 9.18$. The means differ by 3.7 counts, but the standard deviations are about 9 counts each. Just as we saw in the coin-flip example from Chapter 5, the standard deviation describes the point-to-point variation, not the uncertainty on the mean. The standard deviation alone cannot tell us whether the means are statistically different.

To determine whether the difference is statistically significant, we need the uncertainty on each mean. If we have $N_I = 42$ inside measurements and $N_O = 92$ outside measurements, then:

$$\sigma_{\mu,I} = 9.38/\sqrt{42} = 1.45, \quad \sigma_{\mu,O} = 9.18/\sqrt{92} = 0.96$$

Now we can compare the difference in means to these uncertainties. The difference is $\mu_I - \mu_O = 87.9 - 84.2 = 3.7$ counts. How many standard deviations of the mean does this represent? Using the larger of the two uncertainties, this difference is $3.7/1.45 = 2.55\sigma_\mu$. Recalling from Chapter 5 that approximately 95% of measurements fall within 2σ and about 99% fall within 2.58σ, we can state that there is roughly a 1% probability that the inside and outside radiation levels are the same[26]. This is suggestive but not conclusive, we might want stronger evidence before claiming a definitive difference.

[26] More precisely, to determine the significance of the difference between two means, one should consider the combined uncertainty on the difference, which involves adding the uncertainties in quadrature: $\sigma\_diff = \sqrt{(\sigma^2\_\mu,I + \sigma^2\_\mu,O)}$. This gives a slightly smaller number of sigma for the difference. We use the simpler comparison here for illustration.

The natural way to improve our ability to distinguish between these two measurements is to collect more data. If we increase our measurements to $N_O = 388$ and $N_I = 537$, the uncertainties on the means shrink considerably:

$$\sigma_{\mu,O} = 8.83/\sqrt{388} = 0.45, \quad \sigma_{\mu,I} = 9.06/\sqrt{537} = 0.39$$

The measured means with the expanded datasets are $\mu_I = 82.1$ and $\mu_O = 77.9$, giving a difference of 4.2 counts. Expressed in terms of the uncertainty on the inside mean, this is $4.2/0.39 = 10.8\sigma_\mu$. This is far beyond the 3σ threshold at which we can be more than 99.7% confident the measurements are different. With enough data, even a small difference in radiation levels becomes clearly statistically significant.

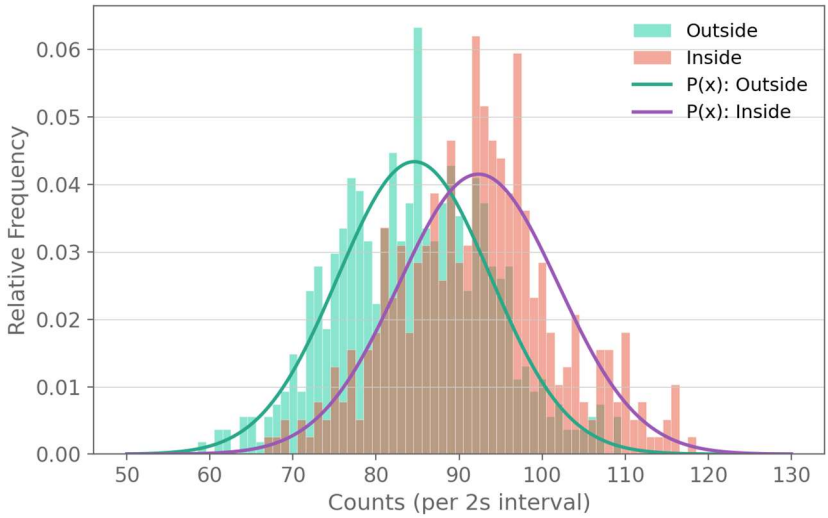

Figure 6.6: The same inside and outside measurement repeated with many more observations (N = 388 inside, N = 537 outside). The distributions are now clearly separated, and the difference in means is statistically significant at >10σ.

This example illustrates a key principle: the ability to resolve a difference between two measurements depends not just on the size of the difference, but on the uncertainty with which we know the mean of each measurement. A difference of a few counts is undetectable with 42 measurements but unmistakable with 537. This is directly analogous to the coin-fairness test from Chapter 5, where detecting a slightly weighted coin required many more trials than detecting a heavily weighted one.

Chapter Summary/Key Takeaways

- For counting measurements that follow Gaussian statistics, the standard deviation is the square root of the mean: $\sigma = \sqrt{\mu}$. This allows us to estimate the uncertainty of a counting measurement from the measurement itself.
- Each individual counting measurement of N_i counts has an uncertainty of $\sigma_i = \sqrt{N_i}$.
- When combining independent measurements, uncertainties are added in quadrature: $\sigma_{total} = \sqrt{(\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_N^2)}$. For counting measurements, this is equivalent to $\sigma_{total} = \sqrt{N_{total}}$.
- The uncertainty on the mean of N measurements is $\sigma_\mu = \sigma/\sqrt{N}$. This is a direct consequence of the central limit theorem and of adding uncertainties in quadrature.
- When converting from total counts to a count rate (dividing by counting time), the uncertainty on the rate is $\sigma_R = \sqrt{N_{counts}}/\Delta t$, not $\sqrt{R}$. The percent error is preserved under this scaling.
- To compare two measurements, the difference in means should be expressed in units of the uncertainty on the mean. A difference of more than $2\sigma_\mu$ is suggestive (~95% confidence), and more than $3\sigma_\mu$ is strong evidence (~99.7% confidence) that the measurements are genuinely different.
- The precision of any counting measurement improves with the total number of counts collected. Reducing the percent error by a factor of two requires collecting four times as many counts.

Review/Example Problems

A detector measures background radiation and records 1024 counts in a 10-second interval.

- What is the count rate and its associated uncertainty?
- If you wanted to reduce the percent error on the count rate to 1%, how long would you need to count (assuming the same count rate)?

A researcher measures radiation levels at two locations. At location A, they collect 50 measurements with a mean of 120 counts per interval and a standard deviation from counting statistics of $\sigma = \sqrt{120}$. At location B, they collect 200 measurements with a mean of 115 counts per interval.

- What is the uncertainty on the mean for each location?
- How many standard deviations of the mean separate the two measurements?
- Would you conclude the radiation levels are statistically different? What confidence level does your answer correspond to?

If the researcher could only collect a total of 250 measurements split between the two locations, how should they divide the measurements to maximize their ability to detect a difference? Does it matter?