Chapter Five: Stochastic Processes and Modeling

5.1 Stochasticity

In the discussion thus far, the stochastic nature of radioactive decays has been touched on. A phenomenon is referred to as ***stochastic*** if it can be described by a probability distribution that depends on a random variable. To better understand this, consider the example of a set of radiation data collected over several minutes. In this case, a detector sensitive to gamma radiation is used to measure interactions – or counts – of gamma radiation over time. Figure 18 shows counts collected in two second intervals. In this graph, we see that the number of counts collected in each interval varies, so there is not a fixed number of interactions occurring in the detector within each two second interval. This is a result of both the stochastic nature of the radioactive decays that produce the gamma radiation being measured, and of the probability for gamma radiation to interact in the detector at all. The later component is discussed in detail in Chapter 3, where how radiation interacts in matter is discussed. Each of these properties of radiation from radioactive decay will lead to fluctuations in the number of measured interactions within each time interval.

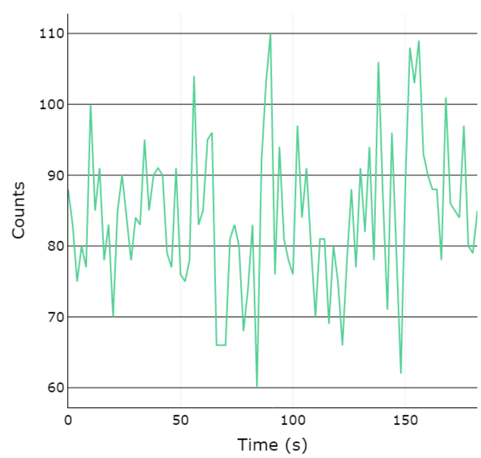

Figure 18: Graph of radiation interactions (counts) measured in two second intervals over time (s).

Despite this intrinsic randomness, there are ways that we can characterize these measurements that allow us to make meaningful predictions about what subsequent measurements of the same physical process under the same conditions (same detector, etc.) would yield. This is a powerful claim which lies at the heart of why statistical methods are so critical to our ability to make predictions about the world.

Two fundamental quantities we can determine that help to characterize this data are the experimental ***mean*** and ***variance***. The experimental mean is defined as

$$\bar{x}_e = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{Eq. 8}$$

The variance describes how much individual measurements vary from the measured mean based on the *residual*: $d = (x_i - \bar{x}_e)$. The sample variance involves essentially attempting to describe the average amount of variation from individual samples from the mean using the residual. However, we cannot simply add up all of the residuals, because by definition those will sum to zero[*]. We can use the squares of the residuals, which will always be positive. The sample variance is therefore defined as

$$\sigma^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \bar{x}_e)^2 \; ^{\dagger} \qquad \text{Eq. 9}$$

Here we use the ***standard deviation***, σ, as the measure of sample variance.

It is often the case that the mean and variance are necessary but not sufficient when attempting to characterize observed properties of the world. However, as we will see, is most of the cases we will encounter, we can use the two properties to make powerful and largely valid assumptions about the data we are investigating and what such data can tell us.

5.2 The Coin Flip

Any stochastic, or random, process with some expected set of results will involve a similar type of variation in measured results to that seen in the above example. To explore how the stochastic nature of a given process can be characterized, we will use a simple yet powerful illustrative example – the coin toss. We know that there are only two possible outcomes for a coin toss – heads or tails, so the probability of either, assuming a fair coin, is trivially 50%. It should also be clear that this does not mean that we can predict heads or tails for any single coin flip. We can take that statement a bit further to emphasize the stochastic nature of this example by pointing out that even after many coin flips, we have no more knowledge about what the outcome of the next coin flip will be than we did for the first coin flip. Meaning, even if we flip the coin ten times and the first nine times the coin landed on heads, this does not mean that the next flip is more likely to be tails just because we know that on average 50% of flips should be tails. The outcome of nine of the ten coin-flips being heads does not even necessarily tell us that the coin is not fair, even though we might intuitively expect that this outcome is very unlikely for a fair coin.

[*] It is left as an exercise for the reader to prove this using the definitions of the residual and the experimental mean.
[†] The normalization here is (N-1) rather than N for a subtle reason. The constraint that the sum of all individual measurements of x satisfy the mean equation means that there is effectively one less independent measurement of the sample than there would be if we were comparing to some unknown "true" mean, and thus our effective total number of measurements is reduced by 1.

From this example, we now see that it can be useful to quantify the probabilities involved. So the first important question to ask is how likely is a given outcome for a number of heads, x, given that we flip the coin N times, so what is the probability of x: $P(x)$. We will start by considering the case for $N=5$, so we are flipping the coin 5 times. In these cases, we are distinguishing between an outcome where we get two heads and then three tails versus getting three tails and then two heads, for example, meaning order matters. We can explore this by building up all of the possible outcomes considering the frequency for each, $F(x)$. We will denote an outcome of heads as H and an outcome of tails as T.

Thinking about the two extremes first, there is only one way to get all heads, $x=5$ (HHHHH), or all tails, $x=0$ (TTTTT). For the case of only one head, there are five ways this can occur: HTTTT, THTTT, TTHTT, TTTHT, or TTTTH. The case of only one tail will also have only five outcomes: THHHH, HTHHH, HHTHH, HHHTH, or HHHHT. The symmetry of these two cases should be clear. The case of two heads includes these outcomes: HHTTT, THHTT, TTHHT, TTTHH, HTHTT, HTTHT, HTTTH, THTHT, THTTH, TTHTH. There are ten possible ways to get two heads. From the symmetry of heads versus tails, we can deduce that there will also the be ten combinations of three heads and two tails. We now have all of the frequencies for the five-flip case:

$$\mathrm{F}(0) = 1, \mathrm{F}(5) = 1$$

$$\mathrm{F}(1) = 5, \mathrm{F}(4) = 5$$

$$\mathrm{F}(2) = 10, \mathrm{F}(3) = 10$$

This is a basic combinatorics problem of five trials for which you are fixing the result for zero to five of the trials. In combinatorics language this would be five choose zero, or five choose one, etc. So, we can reframe the numbers we found above as:

$$\mathrm{F}(0) = \binom{5}{0}, \mathrm{F}(5) = \binom{5}{5}$$

$$\mathrm{F}(1) = \binom{5}{1}, \mathrm{F}(4) = \binom{5}{4}$$

$$\mathrm{F}(2) = \binom{5}{2}, \mathrm{F}(3) = \binom{5}{3}$$

The formula for such combinatorics for 5 flips resulting in 2 heads, for example, is:

$$\frac{5!}{2!\,(5-2)!} = 10$$

Following this form for any combinatorics where we have N coin flips and are choosing x heads, we can generalize this to:

$$\mathrm{F}(x) = \binom{N}{x} = \frac{\mathrm{N}!}{\mathrm{x}!\,(\mathrm{N}-\mathrm{x})!} \qquad \text{Eq. 10}$$

We can represent the possible outcomes visually as a *frequency distribution*, or histogram of the possible outcomes. Each value for x is represented as a bar, or bin, in the histogram, with the height of the bar set by the frequency for that x value. Figure 19 shows the frequency distributions for the case we have just discussed, where $N=5$, and for two other cases, $N=10$ and $N=50$. For these later cases, the total number of outcomes for a given x is determined by Eq. 8.

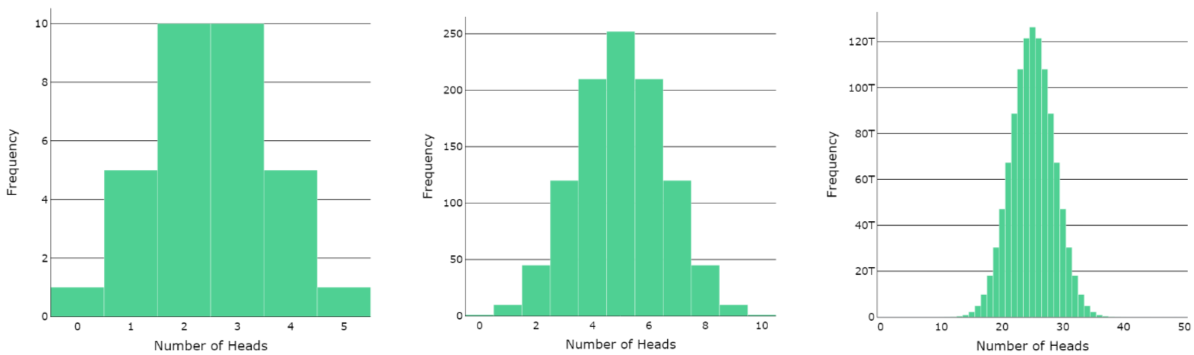

Figure 19: Graphs of the frequency distribution – the number of possible ways to get a particular number of heads – resulting from 5 (left), 10 (middle), or 50 (right) coin tosses.

We now know the general way to determine the frequency of different outcomes, but we are really interested in being able to state the likelihood of a given outcome. For this, we need to know the probability of that outcome relative to all other outcomes. To determine the probability, we need to know the total number of possible outcomes. The probability is then the relative rate for a given outcome, or in other words, the frequency for a given outcome divided by the total number of possible outcomes.

$$\mathrm{P}(x) = \frac{F(x)}{N_{outcomes}}$$

$$\mathrm{N_{outcomes}} = \sum F(x)$$

$$\Rightarrow \mathrm{P}(x) = \frac{F(x)}{\sum F(x)} \qquad \text{Eq. 11}$$

A fundamental property of probabilities is that the sum of probabilities for all possible outcomes should be one. More generally, for any probability distribution, the integral of the full distribution should be normalized to one. We can check that this form for P(x), is correct by confirming that the sum of all probabilities is one:

$$\sum \mathrm{P}(x) = \sum \frac{F(x)}{\sum F(x)} = \frac{\sum \mathrm{F(x)}}{\sum \mathrm{F(x)}} = 1$$

For N=5 we can determine what the total number of possibilities by adding up the frequencies for each number of heads:

$$1 + 5 + 10 + 10 + 5 + 1 = 32$$

So for N=5:

$$\mathrm{P(x)} = \frac{1}{32}F(x)$$

As we increase N, the total number of outcomes will increase. For coin flips we can come up with a general form for this based on the equal probability of either heads or tails, $p_H=p_T=½$, and the fact that the total probability can be determined by multiplying the probability for each outcome together. So for 5 flips, a specific outcome like HTHHT has a probability of (½)*(½)*(½)*(½)*(½), or $(½)^5$. More generally, the total number of outcomes for N flips is 2^N.

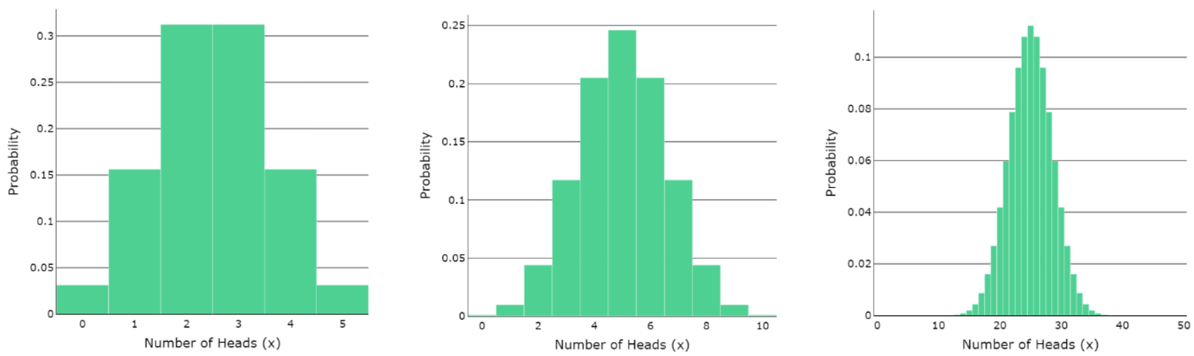

Figure 20: Probability distribution for the number of heads in coin flips for 5 flips (left), 10 flips (middle), and 50 flips (right).

Looking at the probability distributions for the three examples we have considered thus far: *N=5, 10,* and *50*, in Figure 20, there are a few useful observations we can make about these distributions. First, as we might expect from noticing the symmetry for outcomes in the case of five coin flips, and because heads and tails are equally likely, these probability distributions are symmetric. We also see that each of these distributions is centered around x = N/2, meaning that the most likely outcome in each case is for half of the coin flips to be heads.

Given that the probability of heads for a single coin flip is ½, it makes sense that the most likely outcome is for half of the flips to be heads. Intuitively, we might expect that the more coin flips, the more likely that we actually end up with half of the flips being heads. However, if we look at the actual values for P(x) in Figure 20, we see that as we as we increase the number of coin flips attempted, the probability for any single outcome for the number of heads, including the most likely outcome, will be lower. For example, the highest probability outcomes for the 5-flip case (x=2 or 3) is just over 30%, and down to ~24% for the 10-flip case. When we go to the 50-flip case, the maximum probability is drops to only ~11%.

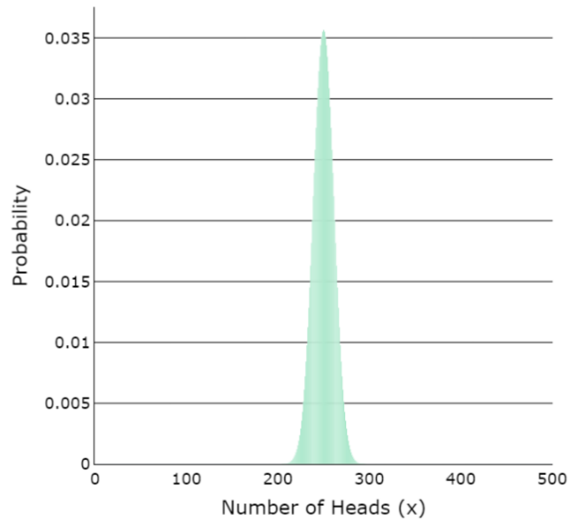

Figure 21: Probability distribution for the number of heads resulting from flipping a coin 500 times.

We also see a narrowing of the full distribution, so the probability for the most extreme results - all heads or all tails – drops more rapidly than for the most likely outcome, going from ~3% for the 5-flip case to ~.1% for the 10-flip case. If we increase N to 500 we can see this even more clearly, as shown in Figure 21. What this illustrates is that while the probability for a specific result, like exactly 250 heads resulting from flipping a coin 500 times, is only ~3.5%, the chance of getting 40% (or 60%) heads – so less than 200 heads (or more than 300 heads) – is essentially zero. We can compare this to the example of 10 coin flips, where 40% or 60% heads (4 or 6 heads) will occur ~20% of the time. So we see that the intuition that more coin tosses would get us closer to the expected rate of 50% heads is correct, but this does not mean that we are more likely to get exactly 50% heads, only that we are increasingly likely to get very close to 50% heads.

To explore this further, it is helpful to formalize the functional form for the observed distribution. From Eq. 8, we have the functional form for F(x), and we know that for coin tosses we can express P(x) as:

$$\mathrm{P}(x) = \frac{1}{2^N} \frac{\mathrm{N}!}{\mathrm{x}!\,(\mathrm{N}-\mathrm{x})!}$$

We can formalize this for arbitrary single result probabilities if we consider again how we arrive at the $(½)^N$ term. We are multiplying specific probabilities for heads or not-heads. The probability for heads is p = ½, and not heads is (1-p) = ½. The probability for a specific outcome involving x heads is the probability of getting heads x times multiplied by the probability of not getting heads N-x times:

$$\mathrm{p}^{\mathrm{x}}(1-\mathrm{p})^{\mathrm{N}-\mathrm{x}}$$

To determine the total probability for getting x heads in N flips, we then multiply by the frequency with which we would get this outcome:

$$P(x) = \mathrm{p}^{\mathrm{x}}(1-\mathrm{p})^{\mathrm{N-x}}\frac{\mathrm{N!}}{\mathrm{x!\,(N-x)!}} \qquad \text{Eq. 12}$$

This is what is known as a ***binomial distribution*** and will describe the probability distribution for any phenomenon which can be described in terms of a particular outcome either occurring (with a probability p) or not occurring.

We can now formalize some fundamental properties of this distribution, namely the mean and variance, or standard deviation. To determine the mean of a probability distribution, P(x), we want to determine the average value for x. The mean value is defined as

$$\bar{\mathrm{x}} = \frac{\sum \mathrm{x\,F(x)}}{\sum \mathrm{F(x)}} = \sum \mathrm{xP(x)} \qquad \text{Eq. 13}$$

Based on the general experimental definition for a mean as the sum of results divided by the number of experiments, this analytic definition for the mean should make sense. Recall that F(x) is just the number of times a given x value occurs, so $\sum \mathrm{x\,F(x)}$ is simply the sum of all x values, and the sum over all F(x) is the total number of possible outcomes.

For the binomial distribution this simplifies to

$$\bar{\mathrm{x}} = \mathrm{Np},$$

which agrees with what we would already expect in the coin-flip example, where we know the average number of heads should be half of the total number of coin flips, and p = ½. Note that this is the "true" or theoretical mean for the number of heads. If we were to perform an experiment in which we repeatedly flipped a coin N times, each time getting a specific result for the number of heads, x, we could determine the experimental mean as defined in Eq. 8, where the N in this case is the number of times we performed the experiment, not the number of times we flipped the coin. This experimental mean will approach the true mean as $\mathrm{N_e} \to \infty$. We can think of each experiment as sampling from the probability distribution, each experiment will only yield a single result somewhere on the probability distribution curve. The more experiments are performed, the more of the distribution is sampled, and the distribution of experimental results will get closer to the true distribution.

We can determine the "true" variance in the same way, recalling that the variance is defined in terms of the residuals, $\mathrm{d} = (x_i - \bar{\mathrm{x}})$, for individual results relative to the mean. Given the probability distribution, the total variance is then

$$\sigma^2 = \sum (\mathrm{x} - \bar{\mathrm{x}})^2 \mathrm{P(x)} \qquad \text{Eq. 14}$$

For the binomial distribution this simplifies to

$$\sigma^2 = \mathrm{Np}(1-\mathrm{p}) = \bar{\mathrm{x}}(1-\mathrm{p}).$$

5.3 Poisson and Gaussian Statistics

5.3.1 Poisson Distributions

The fundamental example of binomial distribution applies for any situation where there is a specific outcome that can either occur or not occur, where there is a well defined probability that the outcome will occur, p.* We can now consider some interesting limits that will simplify this distribution further, specifically the cases of $p \ll 1$ and $N \to \infty$. First we can look at how our binomial distribution changes as we decrease p.

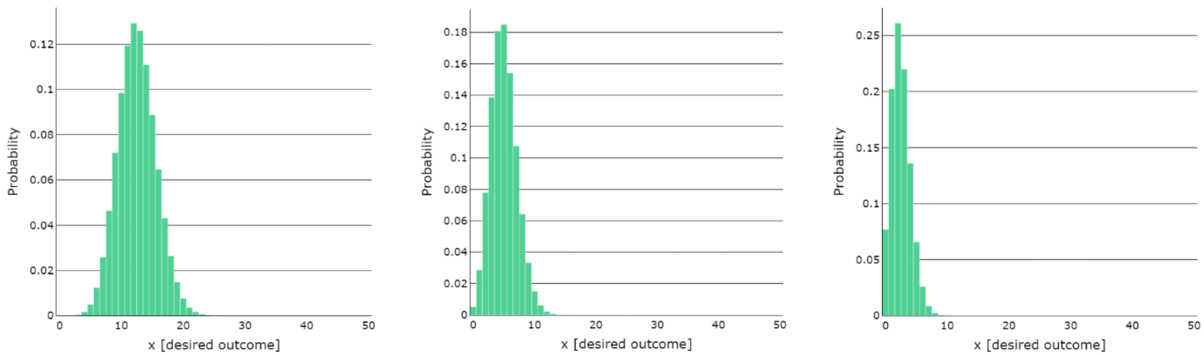

Figure 22: Probability distributions for a given outcome x from N=50 trials, where the probability of a that outcome occurring in a single test, p, is 1/4 (left), 1/10 (middle), or 1/20 (right).

As p decreases, the symmetry of the distribution is lost. Note that the mean value for x is still Np, and that there cannot be a negative number of desired outcomes. It should, therefore, make sense that as p decrease, which necessitates that Np decrease, the probability distribution will get squished up against zero, becoming non-symmetric. If we consider the original form for the binomial distribution, and simply using the fact that $p \ll 1$, we get a simplified functional form for our probability distribution of

$$P(x) = p^x(1-p)^{N-x}\frac{N!}{x!\,(N-x)!}$$

$$\mathrm{p} \ll 1 \;\rightarrow \mathrm{P}(x) = \frac{(\mathrm{p}N)^x e^{-pN}}{x!}$$

$$\Rightarrow \mathrm{P}(x) = \frac{\bar{\mathrm{x}}^{\mathrm{x}}\mathrm{e}^{-\bar{\mathrm{x}}}}{x!} \qquad \text{Eq. 15}$$

This is the ***Poisson*** distribution. This functional form describes cases where there are not a lot of trials (N), but where the probability of the desired outcome for a single trial is much less

* We will be considering the more complex example of dice rolls in the exercises for this chapter. In this case, the binomial distribution will not apply in the same way, though it would apply if we were only discussing the probability of a single type of outcome, such as rolling a one.

than one. The mean, often represented as μ, is unchanged from the more general binomial distribution, but the standard deviation can be simplified because $p \ll 1$:

$$\bar{x} = \mu = Np$$

$$\sigma^2 = Np(1-p) \rightarrow \sigma^2 \approx Np = \bar{x}$$

$$\Rightarrow \sigma = \sqrt{\mu}$$

5.3.1 Gaussian Distributions

We can now explore what happens as we increase the number of trials, N. We will now fix the probability, with $p = 0.05$, and increase N. Figure 23 shows the probability distributions for three examples, where the number of trials (N) ranges from 50 up to 5000. From these examples, we see that as N increases, the distribution again becomes more symmetric. This should make sense because as we increase N the mean value (Np) is increasing, moving the distribution away from the hard cutoff of x=0 that causes the asymmetric distortion for small p and small N.

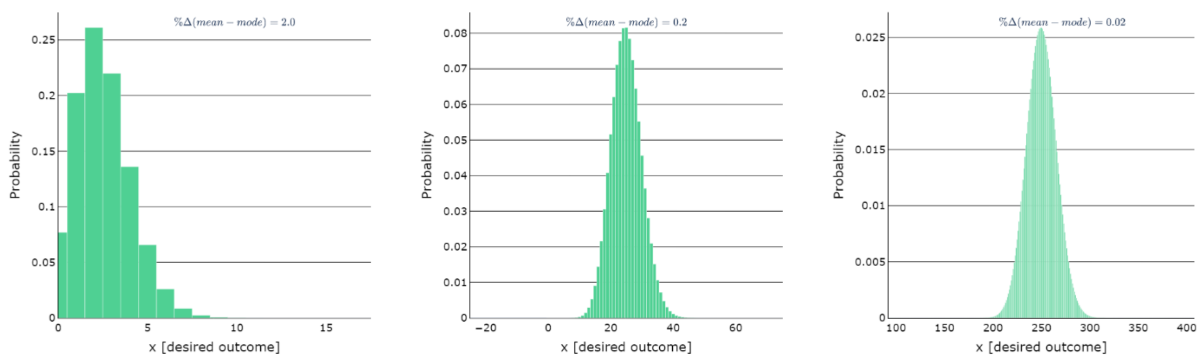

Figure 23: Probability distributions (binomial) for an expected outcome, x, with a probability of occurring, p=0.05, for N trials, where N=50 (left), N=500 (middle) and N=5000 (right).

One way to illustrate the approach towards symmetry is to look at how the mean and mode compare as we increase N. For small N, there is a clear asymmetry in the probability distribution, and the mean and mode differ by a few percent. As we increase N, this difference drops off linearly with N. As N → ∞, we approach another approximate form for this general binomial distribution for which the mean and mode are the same, the ***Gaussian distribution***:

$$P(x) = \frac{1}{\sqrt{2\pi\bar{x}}} e^{-\frac{(x-\bar{x})^2}{2\bar{x}}} = \frac{1}{\sqrt{2\pi\bar{x}}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \quad \text{Eq. 16}$$

Where the mean and standard deviation are the same as for the Poisson distribution:

$$\bar{x} = \mu = Np$$

$$\sigma = \sqrt{\mu}$$

This distribution is also often referred to as the *normal distribution*. Most stochastic processes seen in nature are assumed to follow this distribution, generally because the $p \ll 1$ and $N \rightarrow \infty$ conditions are good approximations in such processes. When considering radioactive decays, we saw from Eq. 7 that the probability for a single decay is typically small, while activities (decays per second) for any substantial quantity of radioactive material will be in kBq (thousands of decays per second). We can therefore safely assume measurements of radiation from radioactive decays to follow Gaussian statistics*.

5.4 Statistical Modeling

The value of being able to describe the probability distribution for expected outcomes using these types of distributions is that it allows us to make predictions about the range of outcomes we might expect to observe. The discussion of the probability for a particular outcome based on the binomial distribution extends to the Poisson and Gaussian distributions. If we think of the Gaussian function as giving us the infinitesimal probability continuously as a function of x – the outcome we are observing – we will find that this instantaneous probability is extremely small, and depends on p and N, which are not always known. As with the binomial distribution, it is much more meaningful to consider the probability for a range of outcomes. This can be generalized in terms of the width, σ, of the Gaussian:

$$P(x \in \mu \pm \sigma) = 66\%$$

$$P(x \in \mu \pm 2\sigma) = 95\%$$

$$P(x \in \mu \pm 3\sigma) = 99.7\%$$

What the above probabilities tell us is that we can state the probability for a measurement to fall within some range around the most likely result. Only 1/3 of all measurements will fall within one σ of the mean, so even though the mean is the most likely value it is a significant fraction of single results will be more than one standard deviation from that mean. At the same time, virtually all measurements will be within three standard deviations from the mean. This does not mean that we will never observe measurements more than three σ from the mean, however, only that there is a very small probability for this result. For example, if you imagine that your experiment involves sampling from this distribution 100,000 times, 300 samples, or tests, will not be within the three standard deviations. Thus, when discussing single measurements, we cannot say that a given measurement disagrees with the expected result if it is more than three σ from the expected mean, but we can say that the probability that this measurement is representative of the expected distribution is less than 0.003.

* When considering the combined decay and measurement probabilities, the number of decays measured can be much smaller than the number of decays produced. Thus, it is often the case when considering radiation detection that the more accurate statical distribution is the Poisson distribution. Even so, the mean and standard deviation depend the number of radiation interactions in the same way in either case.

5.4.1 Modeling the coin-flip

To build our intuition for probability distributions and how to use them in the context of experimental observations, we can again consider the specific example of the coin flip. We know the theoretical probability distribution for the number of heads given N coin flips. If we perform the experiment of flipping a coin N times, we will get a single outcome. We can think of the act of performing this experiment as sampling from this theoretical probability distribution. If we imagine that we perform this experiment M times, we can look at what the experimental distribution of results looks like and explore how this distribution changes as we increase M.

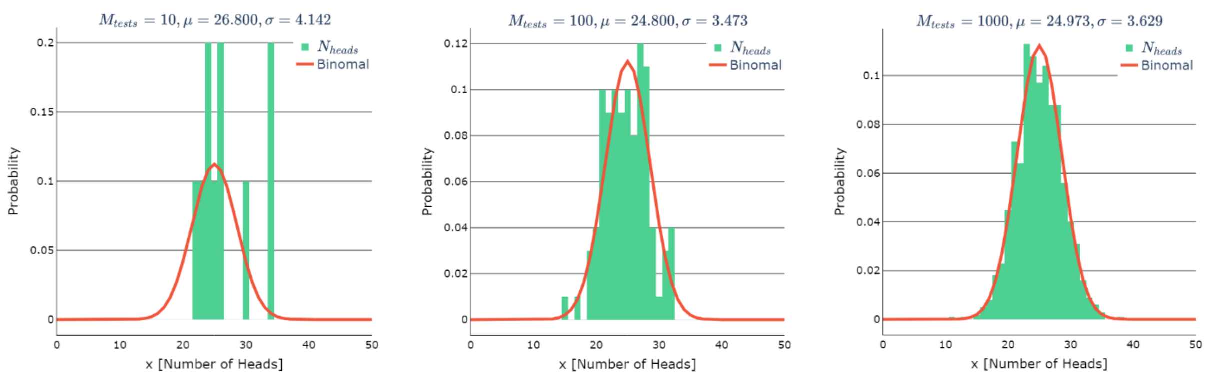

Figure 24: Binomial probability distributions (red) for the number of heads resulting from 50 coin flips. In each case, a simulated 50-coin-flip experiment was run M times where M is 10 (left), 100 (middle), and 1000 (right). The simulated results are shown in green and compared to the expected binomial distribution, along with the experimental mean and standard deviation.

We can see three examples of experimentally sampling from an underlying probability distribution in Figure 24 where we have simulated flipping a coin 50 times and observed the number of heads. In each example, we run our coin-flip experiment M times, increases M from 10 to 1000. We can calculate the experimental mean and standard deviation (Eq. 8 and 9) and compare those to what we expect: $\mu = 25, \sigma = \sqrt{25}(1 - 1/2) = 3.53$. We can see that as we perform this coin-flip experiment more times, we will get an observed distribution much closer to what we expected theoretically, with a measured mean and standard deviation that also get closer to what we expect.

The coin-flip example illustrates a fundamental property of observations made related to stochastic processes. There is often a very precise and well understood probability distribution that we expect these processes to follow. Real observations of outcomes will be sampling from that distribution, but repeated measurements (or observations) will vary in a way that only approaches the true probability distribution as the number of repeated measurements performed approaches infinity.

5.4.2 Testing against our model

It can be helpful to use an example like coin-flips, where we know the underlying probability distribution exactly, to build our understanding of how observation and expectation can vary. In practice, the true probability distribution for a stochastic process is not generally known, and we instead rely on experimentation to estimate this distribution. Thus, it is important to understand that there will always be some uncertainty on this estimation, given that we cannot generally perform enough repeated experiments to fully sample from and characterize the probability distribution. We can explore this even through further examination of the coin-flip example by asking how we would determine the fairness of a coin – that is, does it really have the same probability of landing on heads or tails. In this case, we want to ask if the coin being tested is really following the expected binomial distribution, or if it is deviating from this distribution. It should already be clear from our discussion thus far, and from Figure 24, that it would not be sufficient to only flip the coin 50 times once. And even performing the 50-coin-flip experiment 10 times would leave us with uncertainty as to the fairness of the coin – how much uncertainty is something we can quantify[*].

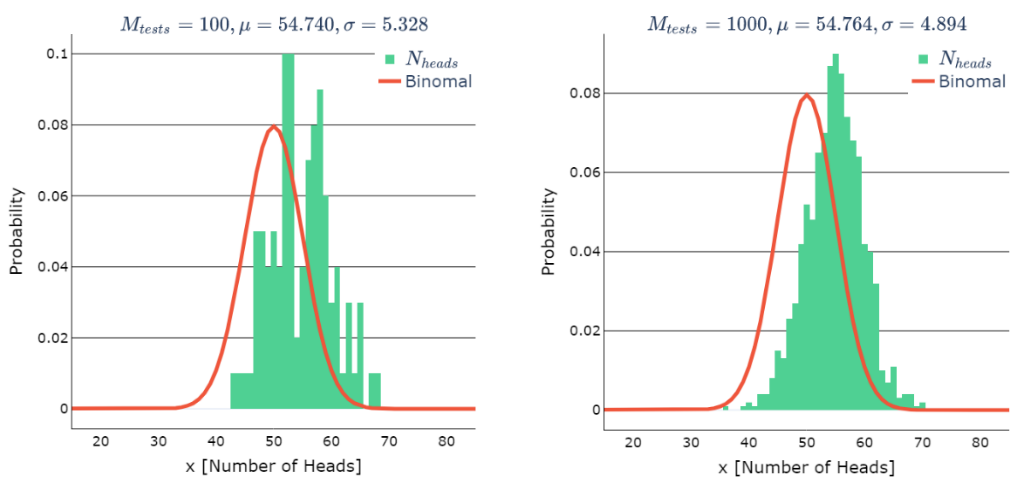

Figure 25: Distribution of the number of heads resulting from a performing 50-coin-flip test of a coin M times. The resulting experimental mean and standard deviation are also shown. In red is the expected binomial distribution for a fair coin. Examples with M=100 (left) and M=1000 (right) are shown.

We can start exploring how to proceed by performing the 50-coin-flip experiment some number of times and compare what we observe with the expectation given that the coin is fair. We can also specify how we are modeling our fair coin. We know both the type of distribution a fair coin should follow – namely the binomial – and the expected mean. If we do this, we might get something like the results shown in Figure 25. In this example, we have simulated a coin that is not fair. We see that the mean number of heads clearly differs from the expected mean of 50. The task is to understand how we can quantify the likelihood that this distribution might arise from a fair coin. The standard deviation is also shown in the figure. From our discussion of

[*] You may be asking why only flip the coin 50 times. What if we simply flipped the coin more – instead of 10 trials of flipping the coin 50 times, what if we flipped it 500 times just once. I will leave it as an exercise to compare these two options and to explore what, if any, difference there is between repeated trials with fewer flips or a single trial with the equivalent number of coin flips.

standard deviations on Gaussian distributions, we know that only ~66% of all trials will fall within the 1-sigma range. The standard deviation shown is large enough that the observed mean is not more than 1 sigma away from the expected mean. Does this mean there is still a 33% chance that this coin is fair? The expected distribution and the observed distribution of heads are clearly very different. The issue is that our knowledge of the mean of the distribution is not reflected by the standard deviation of this same distribution. The standard deviation in this case is telling us about how much variation we would expect when comparing individual coin flips, not when considering the full set of all coin-flip results.

This example seems to indicate that even with this single experiment of testing the coin with a 50-flip trial 1000, or even 100, times should be enough to state that it is unlikely to be a fair coin, because the resulting distribution shows a clear shift from what is expected for a fair coin. To be able to assign a probability that this coin is fair, however, we need to quantify the likelihood of getting the observed mean number of heads from a fair coin[*]. To state how likely the result shown in Figure 25 would be from a fair coin, we need to quantify our experimental uncertainty on the measured mean obtained in this experiment. Note that we cannot state the probability that a coin is not fair, only the probability that it is fair[†]. The challenge now is how to quantify our uncertainty on the mean.

Referring to Figure 24 we see that as we increase the number of trials for our fair coin, the experimental mean gets closer to the theoretical mean and the standard deviation also approaches the expected theoretical variance. The experimental mean is getting closer to what we expect, but it will never be exactly 50, nor will it be the same if we were to perform this same experiment again. We can extend this idea to explore how the experimental mean of our experiment will vary. If we continued running experiments, we would get a distribution of experimental means. A fundamental property of statistics is that we can expect this distribution of the means to follow a Gaussian, or normal distribution, as long as we run enough 50-coin-flip tests[‡].

The resulting normal distribution for the experimental mean of the number of heads is shown in Figure 26. Here we see that the distribution of the experimental means appears to follow a Gaussian (shown in the red curve), even for only a small number of tests of the coin (M_{test}). In general, this is not always a safe assumption, but for stochastic processes this will hold even for small numbers of samples because the distribution being sampled from is also roughly Gaussian. Similarly, as the number of samples, M_{test}, is increased, the standard deviation of the experimental mean, which represents the uncertainty, decreases. This uncertainty on the mean depends on the intrinsic variance in the number of heads but will decrease by the square root of

[*] It is worth noting here that we will also have information about the experimental standard deviation, would that be useful for testing the fairness of a coin?

[†] This is an oversimplification. We could state the probability that the coin was unfair if we tested against a specific weighting – meaning that we could state the expected unfair mean for the coin and test against that.

[‡] In fact, this observation is a result of the central limit theorem, which states that the distribution of values for any parameter of a probability distribution sampled from experimentally will approach a normal distribution as the number of times the probability distribution is sampled approaches infinity.

the number of times the number of heads is sampled, so the number of times the coin is tested (see Eq. 17, where $M_{test} = N_{sample}$). Again, this functional form only holds if the number of samples is high enough that the experimental means would follow a Gaussian if many experiments were performed.

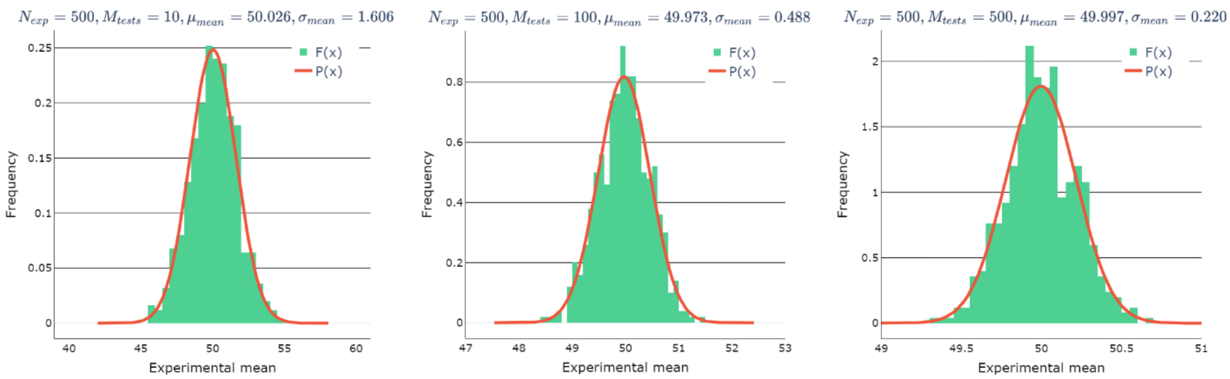

Figure 26: The distribution of experimental means resulting from performing N experiments of running M tests measuring the number of heads from a 50-coin-flip. In this case, 500 experiments are run, while the number of 50-coin-flip tests run in each experiment is varied from M=10 (left) to M=500 (right). In each case, the mean of experimental means is very close to 50, with a standard deviation that gets smaller as M is increased.

As $N_{samples} \rightarrow \infty$ $\qquad \mu_{mean} \rightarrow \mu_{true}$

$$\sigma_{\mu} \rightarrow \frac{\sigma_{dist}}{\sqrt{N_{samples}}} \qquad \text{Eq. 17}$$

The power of this result is that we can now state the uncertainty on the mean number of heads based on how many 50-coin-flip tests are performed. If we refer back to the example shown in Figure 25, for $M_{test} = 100$, the experimental mean was 54.78 and the standard deviation was 5.328. This means that the uncertainty on the mean is $\sigma_{\mu} = 5.328/\sqrt{100} = .5328$, or more than nine standard deviations (of the mean) away from the expected mean for a fair coin (50). Thus, even only testing the coin 100 times, we can determine that for this particular example the probability that the coin is fair is vanishingly small (much less even than the .3% a three-sigma variation would allow).

Chapter Summary/Key Takeaways

- ***Stochastic processes*** are those for which there is some underlying randomness to the process – the process can at best be described by a probability distribution that depends on a random variable.
  - Examples include the random process of radioactive decay, a coin landing on heads or tails, or a dice roll resulting in 1-6.

- We can describe these stochastic processes by determining the expected result using the ***mean*** observed out of a set of results and by determining the intrinsic variance in the process using the ***standard deviation*** of the set of results (Equations 8 and 9).

- The binomial distribution is generally the fundamental probability distribution any stochastic process will follow, if it can be described as either happening with a probability p or not happening with a probability q=1-p (which is generally possible).

- In many natural processes, where the probability of the desired outcome resulting from a single instance of the process– such as a single radioactive isotope actually decaying – is typically very small, the binomial distribution will simplify to the ***Poisson distribution***. If the total number of instances – like the total number of atoms of that radioactive isotope – of the process is also very large, this will further simplify to the ***Gaussian distribution***.

- The Gaussian (and Poisson) distributions have a variance defined by a standard deviation that is simply the square root of the mean.
  - $\mu = Np, \sigma = \sqrt{\mu}$

- When modeling data using a something like a Gaussian distribution, we can approximate the "true" mean and standard deviation of the model using the experimental mean and standard deviation of the data:
  - $\overline{x}_e = \frac{1}{N}\sum_{i=1}^{N} x_i$
  - $\sigma^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \overline{x}_e)^2$

- When using data from a stochastic process to estimate the mean and standard deviation for our model, we can think of this as sampling from the underlying probability distribution the data is following. There will be an intrinsic uncertainty in these estimates because of the intrinsic variance in the. The uncertainty on the mean depends on how many times the underlying probability distribution is sampled from – how many experimental tests of the process are performed.
  - $\sigma_\mu \rightarrow \frac{\sigma_{dist}}{\sqrt{N_{samples}}}$

Review/Example Problems

Consider the example discussed in this chapter of how we might determine if a coin is fair (has a 50% probability of landing on heads or tails). In that discussion, we saw that with enough 50-coin-flip tests, we could determine that the coin was unfair. In the specific example explored, the simulated unfair coin was given a probability of landing on heads of 55%, rather than 50%.

- If we require that the cutoff for the probability that the coin is fair is that the uncertainty on the mean be within three standard deviations of the expected 50% mean for a fair coin, how many 50-coin-flip tests of the 55% heads coin would we have to perform to say that it was unlikely (by our 3-sigma criteria) to be fair?
- What if the coin was only weighted by 1%, so the probability of the coin landing on heads was 51% instead of 50%, would the same number of 50-coin-flip tests be sufficient? If not, now many coin-flip tests would need to be performed now?