Simple (Naïve) Probability

The basic ideas in probability can be illustrated by considering the rolling of a die. There are six possibilities, 1, 2, 3, 4, 5, and 6. In this case each one should be equally likely, say we call that probability P_o. Sometimes the probability is not the same for each outcome and we would use P_n for the probability for each of the n outcomes. Since they are all expected to be the same, I'll just call that probability P_o. If we roll the die once, one of those six possibilities must happen. This means that the sum of the probabilities for all the possibilities must be 1. This means that

$$\sum P_n = 1, \qquad 1.$$

where the sum is from 1 to 6. This means that

$$\sum P_n = 6P_o = 1 \text{ or } P_o = 1/6. \qquad 2.$$

This is process of demanding that the sum of the probabilities = 1 is called normalizing the probabilities. The set of P's is called the probability distribution. I will use the notation P_n = the probability for the n^{th} outcome. Of course for the die, $P_n = 1/6$ for each outcome.

What would we expect to get if we rolled the die many times and averaged the values observed for all the rolls? If we call the actual values x_n, i.e. $x_1 = 1$, $x_2=2$, $x_3=3$ etc., then we would expect that the average would be

$$\langle x \rangle = \sum x_n P_n \qquad 3.$$

where i runs from one to six. Since all the $P_n = 1/6$, that can be taken out of the sum and one has

$$\langle x \rangle = (1/6) \sum n = 21/6 = 3.5 \qquad 4.$$

The symbol $\langle x \rangle$ is called the mean value, or the expected value, or the expectation value.

The probability distribution also has a standard deviation, σ. It is defined as

$$\sigma^2 = \sum (x_n - \langle x \rangle)^2 P_n \qquad 5.$$

If d_n is how far each roll is from the expected value, i.e. $d_n = (x_n - \langle x \rangle)$, then σ squared is the average of the square of d_n. {Note that the sum of the d_n should be zero, so that is not a useful quantity.} The standard deviation is a measure of how "broad" the probabilities are distributed over the possible x's. (The probability as a function of x, P(x), is called the probability distribution.) In this case

$$\sigma^2 = (1/6)\,(2.5^2 + 1.5^2 + 0.5^2 + 0.5^2 + 1.5^2 + 2.5^2) = 2.9 \qquad 6.$$

so $\sigma = 1.7$. This is relatively large compared to the range of possible values because the probability distribution is flat, i.e. all outcomes have the same probability. (Note that the definition of $\langle x \rangle$ in eqn. 3 also means that σ^2 in eqn. 5 is a minimum. No other choice of $\langle x \rangle$ will result in a smaller σ^2.

We could do the same for rolling two dice each time where the measured outcome is the sum of the two numbers on the upward face of the dice. Then there are 11 possible outcomes, the integers from 2 thru 12. However now they do not all have the same probability. This is because there are more ways to roll a 3 with two dice than to roll a 2. There is only one way to roll a 2, both dice (e.g. A and B) have to be one. To roll a 3 one can have A=1 and B=2 or A=2 and B=1. This is twice as likely as rolling one because there are two ways of rolling a 3. If P_2 is the probability of rolling a 2, then $P_3 = 2P_2$, $P_4 = 3 P_2$, $P_5 = 4 P_2$, $P_6 = 5 P_2$, $P_7 = 6 P_2$, $P_8 = 5P_2$, $P_9 = 4P_2$, $P_{10} = 3P_2$, $P_{11} = 2P_2$, and $P_{12} = P_2$. If you normalize this distribution,

$$\sum P_n = 1 \qquad 7.$$

you will find that $P_2 = 1/36$. Then $P_3 = 2/36$, $P_4 = 3/36$ etc. If the x_n is the outcome for the n^{th} possibility, i.e. $x_2 = 2$, $x_3 = 3$ etc., then the expected value $\langle x \rangle = 7$ and $\sigma = 2.42$. You should try to calculate these to verify it. Note that the standard deviation is only slightly larger than for a single die even though the range of possible outcomes is almost twice as large. This is because the probabilities are higher for the outcomes around the expected or mean value. If I divide the possible outcomes by two for two dice so that the range of values is the same as it is for one die, the mean is 3.5 and the standard deviation is 1.2.

It is often useful to plot the probability for each outcome vs the value of the outcome. They are shown below for the one die and two dice probability distributions. I've used n for the index and the x values.

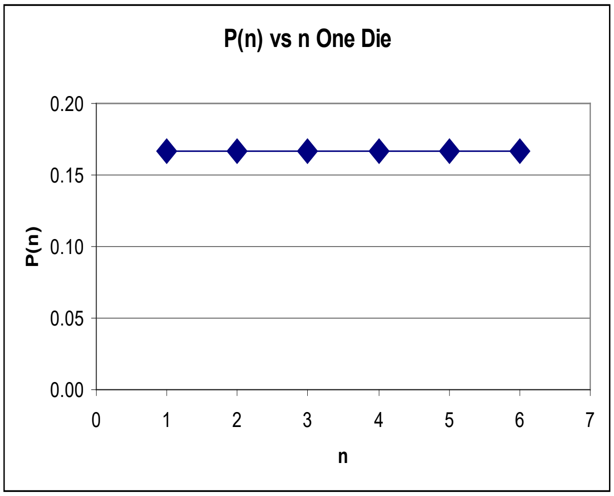

Fig. 1. P(n) for One Die

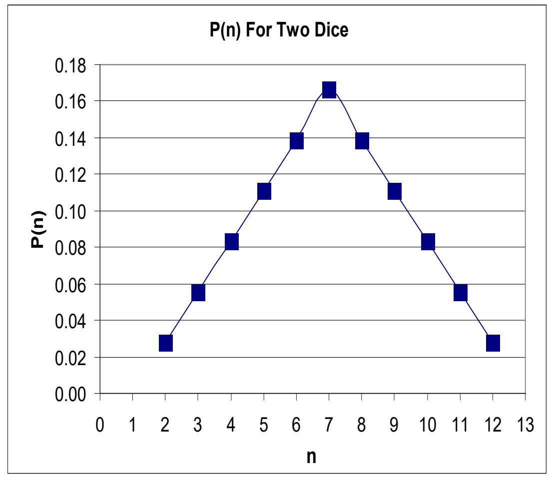

Fig. 2. P(n) for Two Dice

Often the probability distribution is not discrete, but continuous. (Often discrete distributions are approximated by continuous distributions.) For example, you might have a box that you throw a die into, but measure the position of the die from the left wall instead of measuring the number on the die. If the box is 0.5m wide, then the position can have any value between 0m and 0.5m. Then the probability would be a function of x, the distance from the wall, P(x), but now the probability of finding the die between x and x + Δx is P(x) Δx. P(x) is a "probability per unit length" and is called a probability density. {In two dimensions it would be a probability per unit area and in three dimensions a probability per unit volume.} If the die has the same probability of landing at any position between 0 and 0.5m, then P(x) is a constant P_o. Now one normalizes the distribution by say that the sum of P(x) Δx over all allowed x's is 1, but the sum will become an integral. So the normalization condition is

$$\int_{0m}^{0.5m} P(x)dx = \int_{0m}^{0.5m} P_o dx = 1 \qquad 8.$$

If you do the integral you will find that $P_o = 2/m = P(x)$. {Note that it has units of inverse meters.} The mean value or expected value of x is

$$\int_{0}^{0.5m} xP(x)dx = \int_{0}^{0.5m} x\left(\frac{2}{m}\right)dx = 0.25m \qquad 9.$$

The standard deviation squared is

$$\int_{0}^{0.5m} \left(x-\langle x\rangle\right)^2 P(x)dx = \int_{0}^{0.5m} \left(x^2-(0.25m)^2\right)\left(\frac{2}{m}\right)dx = \sigma^2 \qquad 10$$

Where again I've used <x> to indicate the mean value of x, also written as $\bar{x}$.

You should do the second integral and then find the standard deviation. Going from the first expression to the second one involves some algebra, but it isn't too bad. Remember that this integral is the square of the standard deviation. In quantum mechanics, the standard deviation, σ, is often written as Δx, and called the uncertainty in x.

If you want to go through the process for a different probability density, try one where the particle in question must be between x= 0 and x = 10cm.

$$P(x) = Cx(10cm - x). \quad 11$$

C is a constant you will need to find by normalizing P(x), so

$$\int_0^{10cm} P(x)dx = \int_0^{10cm} C\,x\left(10cm - x\right)dx = 1 \quad 12$$

Find C, <x> and Δx = the standard deviation of x.

Note that in eqn. 5 above, we **KNOW** the probability distribution, so the standard deviation is defined by equation 5. When we try to **ESTIMATE** the mean value and standard deviation of the distribution based on a set of measured values it looks a little different. The problem is that we do not know the distribution; we are trying to estimate it from our measurements. For instance if you rolled two dice and recorded the outcomes for 25 rolls and plotted the frequency of each possible outcome versus the value of that outcome it might look something like the graph below.

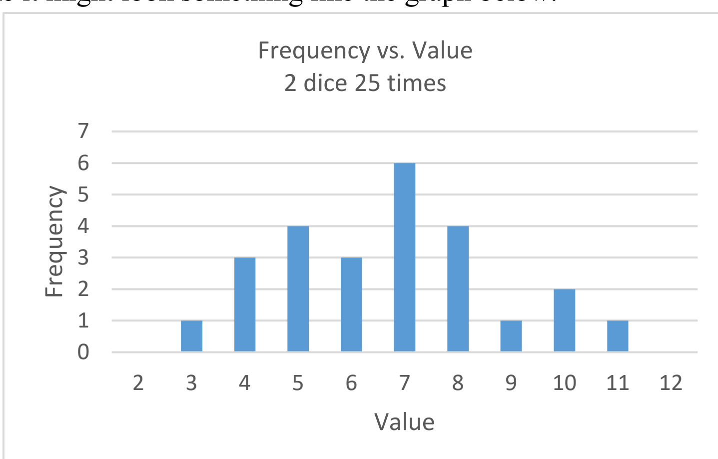

Fig. 3. Frequency plot for rolling two dice 25 times.

(I rolled two dice 25 times to get this, and I did not get a 2 or a 12 in any of the 25 rolls.) This looks vaguely like fig. 2, but you might need to roll them a lot of times to make it really resemble fig. 2. If we use these measurements to calculate an estimated mean and standard deviation, they are just that, **ESTIMATES** based on our sample. If you do the math the average is 6.7 and the standard deviation is 2.06. The mean value isn't too far from the actual mean, 7 and the standard deviation isn't too far from the actual standard deviation of 2.4, but it is off by about 15%. Here you estimate σ by

$$\sigma^2 = \frac{1}{N-1}\sum (x_n - \langle x \rangle)^2 \quad 13$$

where <x> is based on the values measured. This is slightly different than eqn. 5 where you know what the probability distribution is. Seldom do we actually know what the distribution is. Usually we try to estimate it based on a set of measured (i.e. sampled) values.

The Binomial and Gaussian Probability Distributions

One of the more common probability distributions is the binomial distribution. An example of this distribution is flipping a coin. Here there are only two results, heads and tails. A more interesting example is flipping N coins at a time and counting the number of heads, which is the same a flipping one coin N times and totaling the number of heads. The number of heads can be any number from x = 0 to x = N. I could perform this many times, say n times. I would quickly see that it is much more probable to find an x_i close to N/2 than to find one close to 0 or N. If p is the probability of getting a heads and = ½, the probability distribution for this is given by

$$P(x) = \frac{N!}{x!(N-x)!}(p^x)(1-p)^{N-x} \qquad 14$$

Where N is the number of coins tosses each time and x = number of heads. You can see this works for N = 1, where x =0 or x = 1. P(1) = 1 × 0.5 = ½ = P(0). (Remember, 0! = 1.) If N = 2, there are four possible outcomes, but two of them give 1 head, so you can have 0, 1, or 2 heads. The probability for 0 heads, P(0) = 1/4 = P(2) and P(1) = 1/2.

Note that the p in the distribution above does not have to be ½. It can be any number between 0 and 1. This would also be the probability distribution for rolling N dice and letting x = the number of ones. Here the probability of rolling a one is 1/6 for each die, so p would equal 1/6 in the expression above. The binomial distribution is applicable to situations where there are only two outcomes for the "basic" trial. Flipping a coin only has two possible results. Rolling a die has six possible outcomes in principal, but if you only look to see if it is a one or not a one, your trial only has two possible outcomes. Radioactive decay follows this rule if you ask about the probability of decay in a time interval Δt; it either decays or it does not decay.

You might also recognize this P(x) from the expansion of $(a + b)^N$

$$(a+b)^N = \sum_{x=0}^{x=N} \frac{N!}{x!(N-x)!}(a^x)(b^{N-x}) \qquad 15$$

In the first expression p = a and (1-p) = b. In this case a + b = 1 and $(1)^N = 1$, so the sum of P(x) over all x between 0 and N must be 1. That means that the distribution is normalized. I could also label the P(x) so that it indicates the number N and the probability p and label it P(N,x,p).

You can calculate the mean value of x and the standard deviation. They are given by

$$\langle x \rangle = Np \quad and \quad \sigma = \sqrt{Np(1-p)} \qquad 16$$

Again I've used <x> to indicate the mean value of x.

The plot below shows the probability for N = 20 and p = 0.3. This distribution in not symmetric because p ≠ 0.5. The dots are the actual points and the line is drawn to show the shape. The one on the right also shows a normal or Gaussian distribution with the same mean and standard deviation. The normal distribution is the one with the x's instead of dots.

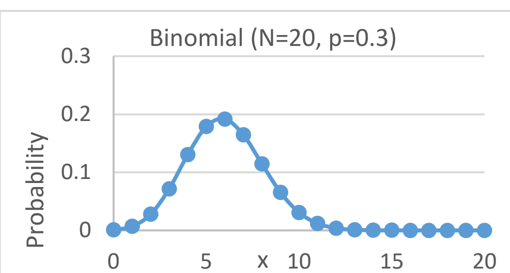

Fig. 4 Binomial Distribution

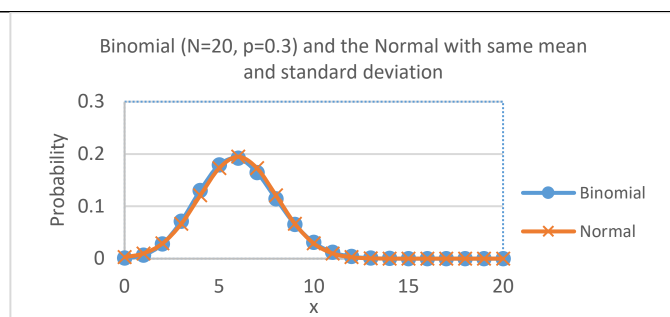

Fig. 5 Binomial and Normal Distributions

This comparison is interesting because it suggests that the binomial distribution can often be approximated by a normal or Gaussian distribution. If N is large, say N > 20, and Np is not too small, the normal distribution is a reasonably good approximation to the binomial distribution.

The normal distribution has the functional form shown below.

$$P(x) = G_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\, exp\left(-\frac{[x-m]^2}{2\sigma^2}\right) \qquad 17$$

where m is the mean of the distribution and σ is the standard deviation. (Taylor uses X as the mean. The most common notations for mean value are m and $\bar{x}$, but in Word $\bar{x}$ is an equation, so it is a little more awkward to use.) $G_{m,\sigma}(x)$ is Taylor's notation for the normal distribution or Gaussian distribution. This distribution is symmetric about x = m. It tends to be "broad" for large σ and sharp for small σ. Note that the area under the curve from x_1 to x_2 is the probability of measuring a value for x between x_1 and x_2. Of course this P(x) is really a probability density rather than a "distribution", but I tend to be a little sloppy and use the terms interchangeably, even though they are not quite the same. The probability of finding or measuring x between x_1 and x_2 is

$$P(x_1 < x < x_x) = \int_{x_1}^{x_2} G_{m,\sigma}(x)\, dx \qquad 18$$

These integrals are not trivial and one usually has to do them numerically or look them up. The two plots below show the normal probability density plotted for a mean value of 10, with a standard deviation of 2 for the left plot and σ = 5 for the right plot. The one on the right is much broader. Notice that they are both symmetrical about the mean = 10 and that the probability of actually measuring a value close to the mean is larger when the probability density is narrower. However, the probability of measuring a value within ± σ of the mean value is the same for both of these densities. Actually this is also true for measuring a value within ± cσ of the mean value where c is a constant.

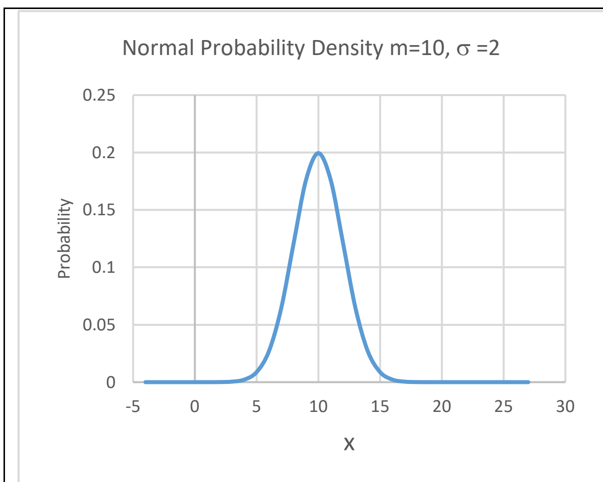

Fig. 6. Normal Distribution with m = 10 and σ = 2.

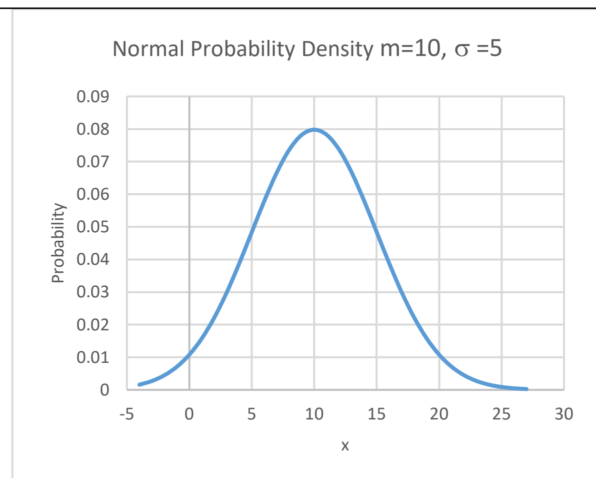

Fig. 7. Normal Distribution with m = 10 and σ = 5.

The probability of measuring a value within ± σ of the mean value is about 0.68. (I usually just say it is about 2/3.) The probability of measuring a value within ± 2σ of the mean is just over 0.95. This means that is only a 5% chance that an individual measurement would be more than 2σ from the mean value. There is a 0.5 probability of measuring a value within ± 0.67σ of the mean value. (This 50% probability is often called the probable error.) Note that these are true for a normal or Gaussian distribution. Other distributions will be different. Consider the distribution for rolling one die. It is flat and four of the six possibilities (2, 3, 4, and 5) lie within one standard deviation (1.7) of the mean (3.5).

However none of the possibilities are more than 1.5σ from the mean value; so the probability of measuring a value more than 1.5σ from the mean value is 0. For the normal distribution it is about 0.13.

One reason that the normal or Gaussian distribution or density is so useful is that it is a good approximation to many distributions or densities. For instance if a single measurement is rolling a die N times and computing the mean of those N rolls; and we repeat this measurement process many times, the distribution of the means of N rolls will look like a normal or Gaussian distribution. The larger N is, the more Gaussian it will look. (We saw this with the binomial distribution. Fig. 2 looks more Gaussian than fig. 1. Note that rolling two dice is the same as rolling one die two times. The figure below compares the distribution for rolling three dice with a Gaussian that has the same mean and standard deviation. I did not take the mean of the three dice though, just the total. To get the mean you would have to divide each result by three.)

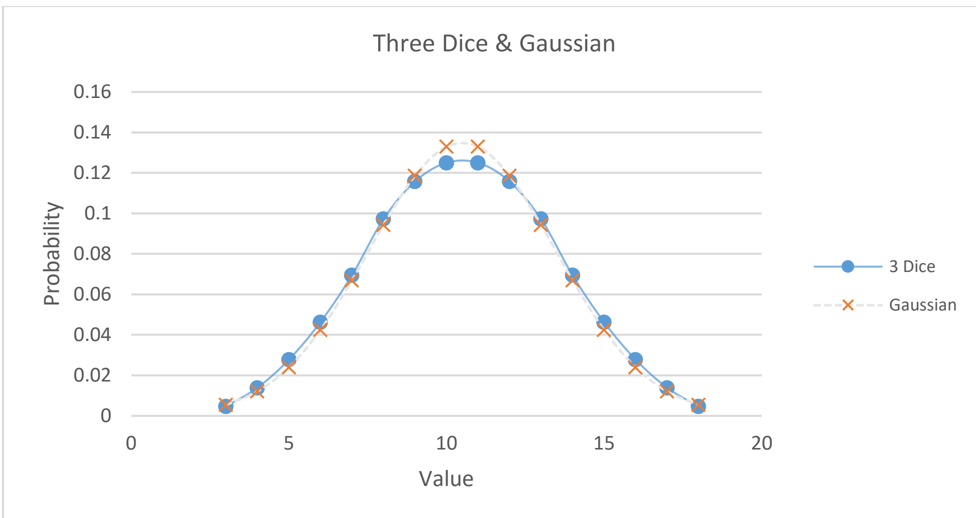

Fig. 8. Distribution for Rolling Three Dice Compared to a Gaussian

Poisson Distribution

The Poisson distribution is a special case of the binomial distribution where p is very small and N is very large so that Np is not too small. This case is usually illustrated by radioactive decay. Here you have a large number of radioactive atoms (often > 10^{15}) that each have a very small chance of decaying in the time interval of question (often minutes or hours). Say N = 10,000 and p = 0.001. The mean number of decays should be Np = 10. The standard deviation squared is $\sigma^2 = Np(1-p)$, but $1-p \approx 1$, so $\sigma^2 \approx Np$. So $\sigma \approx \sqrt{Np} \approx \sqrt{m}$. This means that the standard deviation is the square root of the mean value! A somewhat surprising result. One can make some approximations and so that in the limit of large N and small p with finite mean = Np that the binomial distribution can be approximated by

$$P_{Poisson}(x) = \frac{m^x}{x!}\, e^{-m}$$

Where x = number of events and m= the mean (Np). The Poisson distribution is not symmetric about the mean value and you are more likely to measure a value below the mean value than above the mean value. Another interesting thing is that the Poisson distribution is often approximated by the Normal or Gaussian distribution. The larger the mean value, the more it looks like a Gaussian. In radioactive decay, you often have large mean values, i.e. you try to count long enough so you get more than a hundred counts.

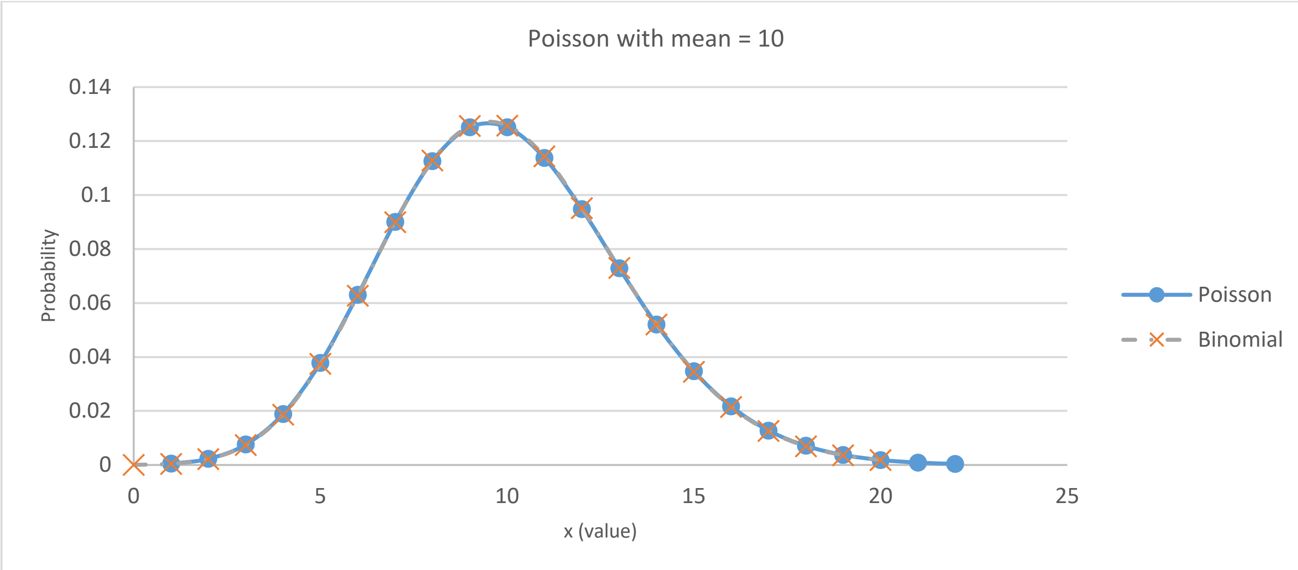

Fig. 9 Poisson distribution vs a Binomial of the same mean & standard deviation.

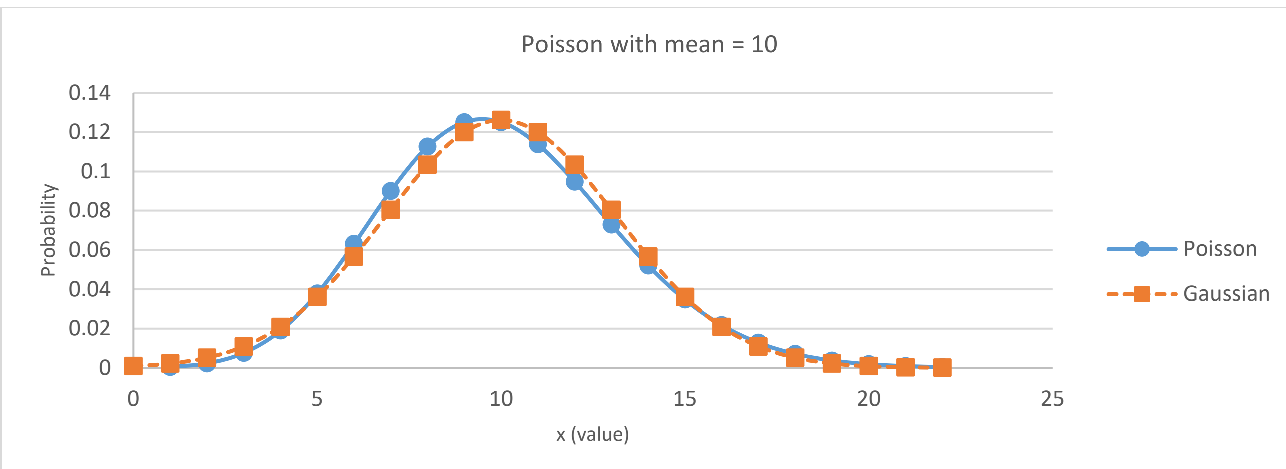

Fig. 10 Poisson distribution compared to a Gaussian of the same mean and standard deviation

It should be pointed out that the Poisson distribution is a discrete distribution whereas the Gaussian or Normal distribution is continuous. However the mean value is not necessarily an integer.

Rules of Thumb

A rough rule of thumb is that if I make two different measurements of the same quantity, e.g. e/m for the electron, and they are within one standard deviation of each other, they look like they could truly be measuring the same quantity, i.e. the two measurements are consistent with measuring the same quantity. However, if they are more than two standard deviations apart, then it looks like they might not be measuring the same quantity. It suggests that there might be systematic errors in one or both of the measurements causing this difference.