In this note, we will derive the Poisson distribution. Poisson distributions are extremely common in nature, and arise wherever you have random discrete events happen that are uncorrelated with each other. The number of events that happen in a given length of time or a given area of sky will be Poisson-distributed. The distribution is characterized by a single number, the expected number of events.

1. Binomial Distribution

We'll derive the Poisson distribution from the binomial distribution. The binomial distribution tells us if we flip a coin, how many heads are we likely to get? In particular, the coin has probability p of coming heads, we're going to flip the coin n times, and we want to know the probability of getting k heads.

If we want all heads, the probability is easy - we need to get n heads in a row, each with probability p, so the probability is p^n. Similarly, no heads is easy - by the same logic, we have $(1-p)^n$. Now, how about one head? Well, the probability that the first flip is heads and the rest are all tails is $p(1-p)^{n-1}$. The probability that the second (or third, or fourth) flip is heads is the same. We have n different places we could have our single head, so the probability of getting one head in *any* location is $np(1-p)^{n-1}$.

How about two heads? Well, the probability of the first two being heads and the rest tails is $p^2(1-p)^{n-2}$. With one heads, we had n ways of spreading out our single head. How many ways are there to spread out two heads? Well, there are n places to put the first head, and $n-1$ places to put the second head (since we used up one of our locations with the first head). However, if we put the first heads in slot one, and the second heads in slot 2, that's identical to putting the first heads in slot 2 and the second head in slot 1 since the final result in both cases is the first two flips are heads. So the number of ways we have to put 2 heads is $n(n-1)/2$. This gives us the final probability of two heads of: $n(n-1)/2p^2(1-p)^{n-2}$.

When we go to k heads, the probability of any individual realization is $p^k(1-p)^{n-k}$. The number of ways we have to spread out the k flips among n heads is $\binom{n}{k}$, or $\frac{n!}{k!(n-k)!}$ where ! is the factorial operator. This gives us the final result, the probability of flipping k heads out of n flips with an individual flip probability of p is:

$$B(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

2. Extending to Poisson

The way we're going to derive the Poisson distribution is to say we have say a stretch of time over which we distribute events randomly, then ask ourselves the probability of getting k events in a small stretch of that time. As we make the background length of time go to infinity, we'll get the Poisson distribution. The thing we want to hold fixed is the expected number of events (times we flip "heads" in the binomial). If the individual probability is p and we have n total flips, the expected number of heads is np which we define to be r. We not want to get rid of n in the binomial distribution, and take the limit as p goes to zero while we hold r constant. We now rewrite the binomial distribution as:

$$\binom{r/p}{k} p^k (1-p)^{r/p-k}$$

First, look at the choose term. That will be $\frac{(r/p)!}{(r/p-k)!}\frac{1}{k!}$. The first term looks like $r/p(r/p-1)(r/p-2)...(r/p-k+1)$. If r/p is much larger than k, then each term in the multiplication goes to r/p, we have k of them, so the choose term turns into $(r/p)^k/k!$.

Fortunately, the p^k term is already good to go. That leaves us with the $(1-p)^{r/p-k}$. We can rewrite this as $(1-p)^{r/p}(1-p)^{-k}$. As p goes to zero, $(1-p)^{-k}$ goes to one, since k is the (finite!) number of events we actually got. That leaves $(1-p)^{r/p} = \left[(1-p)^{1/p}\right]^r$. Recalling that the original definition of e was $(1+1/n)^n$ as n goes to infinity, we can see that the term in brackets is just e^{-1}. The total term is then e^{-r}. Combining all the terms, we get an overall probability of $r^k p^{-k}/k! p^k e^{-r}$. Happily the p terms cancel, and we are left with the probability of getting k heads when we expected r:

$$P(k|r) = r^k e^{-r}/k!$$

3. Poisson Properties

Let's check some basic properties of the Poisson distribution. First, if we haven't screwed up, the probality should sum to one. In particular, we want to check the sum over k:

$$\sigma_k r^k e^{-r}/k!$$

We can pull e^{-r} out, leaving us with

$$e^{-r}\sigma_k r^k/k!$$

The sum is just the Taylor series expansion for e^r, so we're left with $e^{-r}e^r = 1$. So, the probability is indeed one.

Next we should check the expectation of k. Recall that we derived the distribution expecting the mean to be r, but we should check we didn't screw anything up along the way. The mean is going to be the sum over k of $kP(k|r)$, or:

$$\sigma_k r^k k e^{-r}/k!$$

The k in the numerator cancels the last k in the factorial, so we have $e^{-r}\sigma_k r^k/(k-1)!$. Let $k' = k-1$, and we have $e^{-r}\sigma_{k'} r^{k'+1}/k! = e^{-r} r \sigma_{k'} r^{k'}/k'!$. Again the sum turns into e^r, cancelling the e^{-r} out front, and we're left with $<k>= r$, as expected.

Now, how about the variance? We do *not* (yet) know what this should be, but it's critical since all of our error estimates when we actually go out and take Poisson data will depend on this. Remember that $Var k =< k^2 > - < k >^2$. We already have $< k >$, but we're going to have to find $< k^2 >$. That means finding $\sum_k k^2 r^k e^{-r}/k!$. We can start off with how we calculated the expectation, but this leaves us with $e^{-r} r \sum_{k'}(k'+1) r^{k'}/k'!$. We can drop the prime from the k and split up the sum into two terms, leaving:

$$e^{-r}r\left(\sum_k kr^k/k! + \sum_k r^k/k!\right)$$

Happily, we've already done the first sum back when we calculated the expectation, and it is just re^r. The second sum is again the Taylor series for e^r, so we have:

$$< k^2 >= re^{-r}(re^r + e^r) = r^2 + r$$

The variance is then:

$$< k^2 > - < k >^2 = r^2 + r - r^2 = r$$

. This is a foundational result in data analysis. If we expect r photons, the variance in the number of photons we actually get is r, which means the standard deviation is $\sqrt{r}$. The *fractional* uncertainty is $\sqrt{r}/r = r^{-1/2}$. If I want to measure the mean of some process, the uncertainty only goes down like the square root of the number of events I have. If I want to measure the brightness of a star to a part in a thousand, I don't need to observe a thousand photons from the star, I need to observe a million photons.

The Poisson distribution also converges to a Gaussian quite quickly. You could of course have guessed this from the central limit theorem (since a long Poisson observation can be thought of as the sum of several short Poisson observations), but I leave the actual derivation as a (homework) exercise for the reader.