Derivation of the Poisson Distribution

Sharpe, Cal Poly, Oct 2014

Although the Poisson Distribution is derived in various books [1][2], these proofs require a fair amount of background reading. This note is designed to take one from the beginning and thread a path through the math needed for the derivation. One only needs an elementary idea of probability and we also need to use Stirling's formula.

Background

The Poisson distribution arises when one has

- Independent events
- A large number of trials, N, in an interval
- The probability, p, of any individual event occurring in any particular trial being vanishingly small.
- The average number of events occurring in the interval, given by Np, is finite and constant

These conditions would be satisfied by, for example, many nuclear counting experiments where the nuclear decays are independent of each other and we count the number of decays (events) over a certain amount of time (the interval).

The binomial distribution

We start with the binomial distribution which governs the distribution of events that occur when we have some sort of experiment that yields either "success" or "not success". An example is throwing a die and expecting a six. The probability of getting a six in any one throw is $\frac{1}{6}$. If you throw twice the probability of getting two sixes is $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$. The probability of getting three sixes is is $\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$ and so on.

But what is the probability of throwing the die three times and only getting one six? Or two sixes? We can find this if we enumerate the ways in which the die could fall each time we throw. This is shown in table 1 where S stands for a six being thrown and NS stands for six not being thrown.

S	S	S
S	S	NS
S	NS	S
S	NS	NS
NS	S	S
NS	S	NS
NS	NS	S
NS	NS	NS

Table 1: Possible ways the die could land in three successive throws. S stands for a six being thrown, NS stands for a six not being thrown.

Notice that there are three different ways of getting only one six (the fourth, sixth and seventh lines in table 1). Each of these ways have probabilities of $\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}$, $\frac{5}{6}\times\frac{1}{6}\times\frac{5}{6}$ and $\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}$, respectively. To get the final probability we add these probabilities together, giving 0.35.

So you can see that there are two parts to the problem – the first is finding the probability of any individual event occurring while the second is enumerating the number of ways that a certain number of events could occur. Let's look at this second enumeration problem more carefully. What we are trying to do is figure how many possible ways we could get a certain number of successes (say, n) out of some number of trials (say, N). And remember, we do not care about the order these events occur in. To do that lets step back a bit more and look at permutations and combinations, topics that can be found in statistics books, for example, reference [4]

An Aside. Permutations and combinations

Imagine you are a member of a club with five members. A, B, C, D and E. This club wants to elect a president, vice president and treasurer. How many possible ways could this happen? Well, there are 5 possible choices for the president. Once the president is picked there are 4 choices for the VP and once that is done there are three choices for the treasurer. The total number of possible ways is 5x4x3. Note that in this scheme of things the order of the choices is important so that selection ABC for president, VP and treasurer, respectively, is different than BCA or CAB. This is called a permutation and you can see that for N members in the club and n officials to be picked there are N(N-1)(N-2)…(N-n+1) ways. Another way of writing this is

$$\frac{N!}{(N-n)!}$$

which you can check with the example above.

Now, that was for the case that the various offices of the club are distinguishable. What if the club wants to elect an executive committee? The members of the committee could be picked as ABC or CAB or BCA and those are identical to each other so the formula for the permutation, above, will over-count the number of ways. We can correct this over count by noting that if the size of the executive committee is n then for each selection of members to be in it, there are $n!$ different ways of arranging those n members. This leads to

$$\frac{N!}{(N-n)!\,n!}$$

for the number of possible ways to arranging n things out of N things with no distinction between them . This is called a combination and is often written as $\binom{N}{n}$ and read as "N choose n".

Returning now to our die throwing and denoting the probability of throwing a six by the symbol p (so that the probability of not throwing a six is 1-p then we can write the probability of seeing n sixes in N throws of the die as

$$\binom{N}{n} p^n (1-p)^{N-n}$$

This is the binomial distribution, quoted as formula 10.6 in Taylor's book [3].

Back to the Poisson distribution

Now the number of trials, N, over some interval is going to get very big while p, the probability of seeing success (an event) in any one of the trials, is going to get very small. However, the product Np will be finite and is clearly the mean number, $\bar{n}$,of events in the interval.

$$\bar{n} = Np$$

Then writing the binomial distribution

$$P_N(n) = \binom{N}{n} p^n (1-p)^{N-n}$$

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

and taking the log of both sides gives

$$\log\big(P_N(n)\big) = \log(\,N!) - \log\big((N-n)!\big) - \log(n!) + nlog(p) + (N-n)\log\,(1-p)$$

Now we are going to use Stirling's approximation

$$\log(x!) \approx xlog(x) - x$$

where x is a large number and recall that if $z \ll 1$ then $\log(1+z) \approx z$.

We get

$$\log\big(P_N(n)\big) \approx N\log(\,N) - N - [(N-n)\log\big((N-n)\big) - (N-n)] - \log(n!) + nlog(p) \\ + (N-n)\log\,(1-p)$$

Then, noting that we can write $\log(N-n) = \log\left(N\left(1-\frac{n}{N}\right)\right) = \log(N) + \log\left(1-\frac{n}{N}\right) \approx \log(N) - \frac{n}{N}$ because n<<N and recalling that p<<1 we get after some simplification

$$\log\big(P_N(n)\big) \approx nlog(Np) - \log(n!) - Np$$

$$P_N(n) \approx \exp\,[nlog(Np) - \log(n!) - Np]$$

From above we see that Np is the average number of events so finally we have, after dropping the approximately equals sign and converting to the notations of Taylor's book [3] where $\bar{n}$ is replaced with μ and n with ν we get

$$P_\mu(\nu) = \frac{\mu^\nu e^{-\mu}}{\nu!}$$

Which is the probability of seeing ν events in an interval where the average number of events is μ. This is equation 11.2 in reference [3].

Checks and properties

You can check (by summing from 0 to infinity) that the distribution is normalized and, after multiplying by ν and summing that the mean value is indeed μ and, finally, you should check that the standard deviation is $\sqrt{\mu}$.

References

[1] Mandel and Wolf "Optical Coherence and Quantum Optics" Cambridge U. Press (1995)

[2] Gershenfeld "The Physics of Information Technology" Cambridge U. Press (2001)

[3] Taylor "An Introduction to Error Analysis" 2nd Ed. University Science Books (1997)

[4] Spence and Vanden Eynden "Finite Mathematics" Scott Forseman & Co. (1990)