Introduction to Probability Distribution Functions

Michael Mulhearn

April 30, 2018

1 The Binomial Distribution

The Binomial Distribution is the most general of the distributions we'll consider, but it is a bit cumbersome to use in practice. The more familiar Poisson and Gaussian distributions are limiting cases of this distribution.

Suppose you repeat a particular process n times, and each time you have the same probability ϵ of a particular outcome, which, without losing generality, we'll call "success". The probability of having exactly m successes after n trials is simply given by:

$$P = \sum_i p_i$$

where i runs over all specific outcomes with m successes and p_i is the probability of each specific outcome. However, as these specific outcomes all contain exactly m successes, they share the same probability, namely:

$$p_i = \epsilon^m (1 - \epsilon)^{n-m}$$

and so we are left to consider simply the total number of specific outcomes containing m successes.

The quantity we need is provided by the binomial theorem from mathematics, which states that:

$$(p + q)^n = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} \tag{1}$$

where the binomial coefficients are defined by

$$\binom{n}{m} = \frac{n!}{m! \, (n - m)!} \tag{2}$$

and are also often referred to in other contexts as n-choose-m. The binomial coefficient simply tells us how many times we can choose m instances of p instead of q, from n factors, and so it is precisely the combinatoric factor that we need.

The probability of obtaining m successes after n trials with probability ϵ is therefore given by:

$$P(m; \, n, \epsilon) = \binom{n}{m} \epsilon^m \, (1 - \epsilon)^{n-m} \tag{3}$$

which is called the Binomial Distribution.

2 Mean and Variance

Given a probability distribution, the most urgent questions are generally "what is the mean value we can expect from this distribution?" and "how close to the mean value are most of the outcomes?" The first answer localizes the distribution while the second answer describes its width.

To calculate the mean value in answer to the first question, we simply calculate a weighted average:

$$\langle m \rangle \equiv \bar{m} \equiv \sum_m m \, P(m) \tag{4}$$

For a continuous probability distribution, we would integrate instead:

$$\langle x \rangle \equiv \bar{x} \equiv \int x \, P(x) \, dx \tag{5}$$

We usually answer the second question in terms of the variance, σ^2, of the distribution:

$$\sigma^2 \equiv \langle (x - \bar{x})^2 \rangle$$

Other answers have problems, e.g. $\langle x - \bar{x} \rangle$ can be zero or nearly so, even for wide distributions, as long as it is symmetric. You could fix this by calculating $\langle |x - \bar{x}| \rangle$ but this is generally much harder to calculate, and less useful, than the variance. For instance, it is left as an exercise to show that:

$$\langle (x - \bar{x})^2 \rangle = \langle x^2 \rangle - \bar{x}^2 \tag{6}$$

using the fact that $\bar{x}$ is simply a number, and so $\langle \bar{x} \rangle = \bar{x}$. We often write this result equivalently as:

$$\sigma^2 \equiv \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \tag{7}$$

Which shows explicitly that we need only calculate $\langle x \rangle$ and $\langle x^2 \rangle$ in order to determine the variance of a distribution.

3 Mean and Variance of the Binomial Distribution

The mean value of Binomial Distribution is given by:

$$
\begin{aligned}
\bar{m} &= \sum_{m=0}^{n} m \, P(m) \\
&= \sum_{m=0}^{n} m \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m}
\end{aligned}
$$

which looks rather daunting! The trick is to use the Binomial Theorem (1) and define a function of two independent variables p and q given by:

$$f(p, q) = (p + q)^n = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m}$$

We then calculate:

$$\frac{\partial f}{\partial p} = n(p + q)^{n-1} = \sum_{m=0}^{n} m \binom{n}{m} p^{m-1} q^{n-m}$$

and multiplying by p we have:

$$np(p+q)^{n+1} = \sum_{m=0}^{n} m \binom{n}{m} p^m q^{n-m}$$

which is true for any p and q. We now substitute the particular values $p = \epsilon$ and $q = 1 - \epsilon$ and find that:

$$n\epsilon = \sum_{m=0}^{n} m \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m} \equiv \sum_{m=0}^{n} mP(m) = \bar{m}$$

So the mean value is given by:

$$\bar{m} = n\epsilon \tag{8}$$

or the total number of trials times the probability of success for each trial, a wholly plausible answer.

For the variance, we use a variation of the same trick, this time using the second partial derivative:

$$p^2 \cdot \frac{\partial^2 f}{\partial p^2} = n(n-1)p^2(p+q)^{n-2} = \sum_{m=0}^{n} m(m-1) \binom{n}{m} p^m q^{n-m}$$

and again putting $p = \epsilon$ and $q = 1 - \epsilon$ to find that:

$$
\begin{aligned}
n(n-1)\epsilon^2 &= \sum_{m=0}^{n} (m^2 - m) \binom{n}{m} p^m q^{n-m} \\
&= \sum_{m=0}^{n} (m^2 - m) P(m) \\
&= \langle m^2 - m \rangle = \langle m^2 \rangle - \langle m \rangle
\end{aligned}
$$

and as $\langle m \rangle = n\epsilon$ we have:

$$\langle m^2 \rangle = n(n-1)\epsilon^2 + n\epsilon$$

And so:

$$\sigma^2 = \langle m^2 \rangle - \langle m \rangle^2 = n(n-1)\epsilon^2 + n\epsilon - n^2\epsilon^2$$

or simply:

$$\sigma^2 = n\,\epsilon\,(1 - \epsilon) \tag{9}$$

Note that if $\epsilon = 0$ or $\epsilon = 1$, there is only one outcome (all failures or all success) and so the variation is zero.

4 The Poisson Distribution

Suppose we have some time interval over which we expect to observe a mean number of events λ. The events must be independent of one another: an event occurring at a particular time cannot effect the time at which the next event occurs. We divide the time interval over which the λ events are expected to occur into into n sub-intervals, each with an equal probability to contain an event. These intervals will be all the same size if the events are uniformly distributed in time, but if the events are not uniformly distributed, the sub-intervals are simply chosen to ensure the probability

is the same in each interval. Once cast this way, we can interpret this as a binomial distribution, with probability to contain an event, by construction, given by $\epsilon = \lambda/n$:

$$
\begin{aligned}
P(m) &= \binom{n}{m} \epsilon^m (1-\epsilon)^{n-m} \\
&= \frac{n!}{m!\,(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1-\frac{\lambda}{n}\right)^{n-m} \\
&= \left(\frac{\lambda^m}{m!}\right) \left(1-\frac{\lambda}{n}\right)^n \left[\frac{n!}{(n-m)!} \cdot \frac{1}{n^m}\right]_1 \left[\left(1-\frac{\lambda}{n}\right)^{-m}\right]_2
\end{aligned}
$$

It is left as an exercise to show that both $[\ldots]_1 \to 1$ and $[\ldots]_2 \to 1$ as $n \to \infty$. Recalling that

$$
\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}
$$

we obtain the Poisson distribution, the probability for observing m events for a mean of λ:

$$
P(m\,;\,\lambda) = \frac{\lambda^m}{m!} e^{-\lambda} \tag{10}
$$

Notice that there is no longer a parameter n, since we took $n \to \infty$, and so m now ranges from 0 to ∞.

5   Mean and Variance of The Poisson Distribution

The mean of the Poisson distribution is given by:

$$
\begin{aligned}
\bar{m} &= \sum_{m\geq 0} m\, P(m) \\
&= \sum_{m\geq 0} m\, \frac{\lambda^m}{m!} e^{-\lambda}
\end{aligned}
$$

Since the first term $(m = 0)$ is zero, we have:

$$
\begin{aligned}
\bar{m} &= e^{-\lambda} \sum_{m\geq 1} \frac{\lambda^m}{(m-1)!} \\
&= \lambda e^{-\lambda} \sum_{m\geq 1} \frac{\lambda^{m-1}}{(m-1)!} \\
&= \lambda e^{-\lambda} \sum_{n\geq 0} \frac{\lambda^n}{n!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
\bar{m} &= \lambda \tag{11}
\end{aligned}
$$

which should come as no surprise, as the assumption in the derivation was the that mean number of events was λ.

For the variance, we use a similar manipulation to calculate:

$$
\begin{aligned}
\langle m^2 \rangle &= \sum_{m \geq 0} m^2 \, P(m) \\
&= \sum_{m \geq 0} m^2 \frac{\lambda^m}{m!} e^{-\lambda} \\
&= \lambda \sum_{m \geq 1} m \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} \\
&= \lambda \sum_{n \geq 0} (n+1) \frac{\lambda^n}{(n)!} e^{-\lambda} \\
&= \lambda \langle m+1 \rangle = \lambda (\lambda + 1)
\end{aligned}
$$

And so:

$$
\begin{aligned}
\sigma^2 &= \langle m^2 \rangle - \langle m \rangle^2 \\
&= \lambda (\lambda + 1) - \lambda^2 \\
\sigma^2 &= \lambda.
\end{aligned}
\tag{12}
$$

That is, the variance of a Poisson distribution is simply the mean.

6   The Gaussian Distribution

Next, we consider the Poisson Distribution in the limit $\lambda \to \infty$. In this case, we can apply the Stirling Approximation:

$$
\lim_{n \to \infty} n! = \sqrt{2\pi n} \; e^{-n} \, n^n
$$

to the Poisson distribution as follows:

$$
\begin{aligned}
P(m) &= \frac{\lambda^m}{m!} e^{-\lambda} \\
&\to \frac{\lambda^m e^{-\lambda}}{\sqrt{2\pi m} \; e^{-m} \, m^m} \\
&= \frac{e^{m-\lambda}}{\sqrt{2\pi \lambda}} \left(\frac{\lambda}{m} \right)^{m + \frac{1}{2}}
\end{aligned}
$$

Now we consider a new variable δ, defined by

$$
\delta \equiv \frac{m - \lambda}{\lambda}
$$

which measures the difference between the value m and the mean of the distribution, as a fraction of the mean. Intuitively, the function is getting very narrow, and so we expect this to be a small quantity, but let's check this. First we have:

$$
\langle \delta \rangle = \frac{\langle m \rangle - \lambda}{\lambda} = \frac{\lambda - \lambda}{\lambda} = 0
$$

but also:
$$\langle \delta^2 \rangle = \frac{\langle (m - \lambda)^2 \rangle}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$
where we have used the fact that the variance is given by $\langle (m - \lambda)^2 \rangle = \lambda$, and so as $\lambda \to \infty$ we have
$$\langle \delta^2 \rangle \to 0$$
So we can write:
$$m = \lambda(1 + \delta) \tag{13}$$
where we expect the approximation $\delta \to 0$ to hold as long as we require $\lambda \to \infty$. So now we can write the distribution in terms of the small quantity δ and the large quantity λ as:

$$
\begin{aligned}
P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \left(\frac{\lambda}{\lambda(1 + \delta)} \right)^{\lambda(1+\delta)+\frac{1}{2}} \\
&= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{X} \tag{14}
\end{aligned}
$$

where we define the quantity:
$$X = (1 + \delta)^{\lambda(1+\delta)+\frac{1}{2}}$$
which can be approximated as follows:

$$
\begin{aligned}
\ln X &= \left(\lambda(1 + \delta) + \frac{1}{2} \right) \cdot \ln(1 + \delta) \\
&= \left(\lambda(1 + \delta) + \frac{1}{2} \right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3) \right) \\
\frac{\ln X}{\lambda} &= \left(1 + \delta + \frac{1}{2\lambda} \right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3) \right) \\
&= \left(1 + \delta + \mathcal{O}\left(\delta^2\right) \right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3) \right) \\
&= \delta + \frac{\delta^2}{2} + \mathcal{O}(\delta^3),
\end{aligned}
$$

where in the second to last step we used $\mathcal{O}\left(\frac{1}{\lambda}\right) \sim \mathcal{O}\left(\delta^2\right)$. Neglecting the small quantities, we can approximate
$$X = \exp\left(\lambda\delta + \lambda\frac{\delta^2}{2} \right)$$
which, when plugged backed into Equation 14 yields:

$$
\begin{aligned}
P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{\exp\left(\lambda\delta + \lambda\frac{\delta^2}{2} \right)} \\
&= \frac{1}{\sqrt{2\pi\lambda}} \cdot \exp\left(-\lambda\frac{\delta^2}{2} \right) \tag{15}
\end{aligned}
$$

Now notice that Equation 13 implies that m is quite large, and so may now be treated as a continuous variable, which we will rename x (since m looks like an integer value), hence we have:
$$\delta \equiv \frac{m - \lambda}{\lambda} = \frac{x - \lambda}{\lambda}$$

and likewise we know that the variance of the original Poisson distribution is given by $\sigma^2 = \lambda$, and so we can rewrite Equation 15 in the (hopefully) more familiar form:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x - \lambda)^2}{2\sigma^2}\right) \tag{16}$$

Which is a Gaussian distribution with mean value λ and variance σ^2. The proof that these quantities are indeed the mean and the variance is left as an exercise.

7 Histograms and Distributions in Scientific Python

```
In [23]: samples = np.random.normal(size=1000)
         bins = np.arange(-4,5) # integer binning from -5 to 5
         cbins = 0.5*(bins[:-1] + bins[1:])   # bin centers, for plotting
         counts,dummy = np.histogram(samples,bins)
         errs = np.sqrt(counts)
         #plot(cbins,hist) # <- this version is for amateur hour!

         errs = np.sqrt(counts)
         plt.errorbar(cbins, counts, yerr=errs, fmt='o')


         fbins = np.arange(-4,5,0.1)
         y = mlab.normpdf( fbins , 0.0, 1.0)
         y = y*1000
         plt.plot(fbins,y,'r-')
```
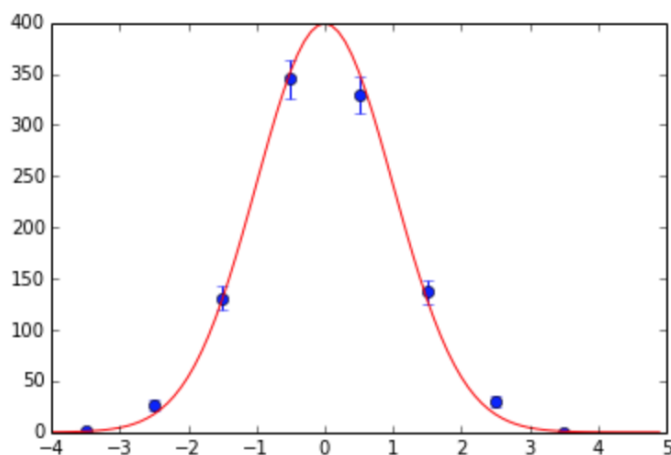
Out[23]: [<matplotlib.lines.Line2D at 0x10f5dbd10>]

Figure 1: Histogram.

8 Exercises

Problem 1: Show that the Binomial distribution, $P(m)$, in Equation 3 is properly normalized:

$$\sum_{m=0}^{n} P(m) = 1$$

as a consequence of the Binomial Theorem (Equation 1).

Problem 2: Show that Equation 6 is correct.

Problem 3: Show that the Poisson distribution, $P(m)$, in Equation 10 is properly normalized:

$$\sum_{m \geq 0} P(m) = 1.$$

Hint: recall the Taylor series expansion for e^{λ}.

Problem 4: Show that the Gaussian distribution, $P(x)$, in Equation 16 is properly normalized:

$$\int_{-\infty}^{\infty} P(x)dx = 1.$$

Problem 5: Show that the mean of the Gaussian distribution has been correctly identified in Equation 16. That is, show explicitly that:

$$\int_{-\infty}^{\infty} xP(x)dx = \lambda$$

Problem 6: Show that the variance of the Gaussian distribution has been correctly identified in Equation 16. That is, show explicitly that:

$$\int_{-\infty}^{\infty} x^2 P(x)dx = \sigma^2$$

when we take $\lambda = 0$ (which is equivalent to simply changing variables $y = x - \lambda$.)