Other distributions and expected values

Other distributions.

There are many, many other distributions other than just the binomial or the normal. Some, like the binomial, are discrete, others, like the normal, are continuous. Here are just a few examples.

1. Poisson distribution. This distribution is discrete and is used to determine the probabilities of rare events which do not have an upper limit. For example, you might be interested in the probability that a particular individual has three ticks on them. As you might imagine, this is a fairly rare occurance (most people don't have ticks), but at the same time, there is no limit to the number of ticks a person might have.

Let's take a look at the distribution and do a simple example:

$$f(y) = \frac{e^{-\mu}\mu^y}{y!}$$

Notice that this distribution only has one paramater (μ). Of course, since we almost never know what μ is, we have to pretend again.

So let's pretend we know $\mu = 4$ for the true average number of ticks in a dog. We want to know the probability a dog has 6 ticks ($Pr\{Y = 6\}$). We just plug everything into the Poisson distribution and get:

$$Pr\{Y = 6\} = \frac{e^{-4}4^6}{6!} = 0.104$$

2. Discrete uniform distribution. This distribution is discrete (a continuous version also exists; see below). This distribution simply assigns each outcome the same probability. For example, if you roll a dice, the probability of each outcome is exactly the same ($\frac{1}{6}$).

Here it is, although since it's so simple, we won't bother with an example.

$$f(y) = \frac{1}{N}$$

Where N is the number of possible outcomes (e.g., 6, if you're rolling a dice).

3. Negative binomial. This is not meant to be a list of every possible distribution, so we'll only do one more discrete distribution, the negative binomial. This is very useful when you're trying to figure out how many trials you need for a given number of successes. We'll do an example below, but note that the negative binomial looks very similar to the binomial:

$$f(y) = \binom{r+y-1}{r-1} p^r (1-p)^y$$

Where $r =$ the number of successes, $y =$the number of failures before you get to r successes, and $p =$ the probability of success (same as in the binomial). (Note that the number of trials, n, is $n = r + y$, and the number of failures is $y = n - r$). So how do we use it?

Let's go back to our example of red headed people in the UK. We want to keep picking people for our sample until we have three red headed people. What's the probability of having three red headed people on our fifth attempt ($n = 5$)? sixth attempt ($n = 6$)? Notice that we need a minimum of three people in our sample, but there is no *guarantee* of three red headed people no matter how many people we put in our sample. We could keep going forever and not get three red headed people (although this is extremely unlikely!)

So let's figure out the probability we have three red headed people in five attempts. Note that the number of success is $r = 3$ and so the number of failures (y) must be $5-3 = 2 = y$. So we have:

$$f(y) = \binom{3+2-1=4}{3-1=2} 0.1^3 0.9^2 = .00486$$

For $n = 6$ attempys, we get:

$$f(y) = \binom{3+3-1}{2} 0.1^3 0.9^3 = 0.0729$$

or the probability of three read headed people when we get to 25 attempts:

$$f(y) = \binom{3+22-1}{2} 0.1^3 0.9^{22} = 0.0252$$

Notice again, that like the Poisson, this distribution has no upper limit.

4. The continuous uniform distribution. This is our first continuous distribution (other than the normal):

$$f(y) = \frac{1}{b-a}$$

Where a and b are the endpoints of the distribution. Notice that just as with the discrete uniform distribution, every outcome is equally likely.

Let's see how this might work. Suppose someone told you that there's a waterfall on a 2 hour hike, but didn't tell you where. Let's figure out the probability you'll see the waterfall in the first 15 minutes. We note that $a = 0$ and $b = 120$ minutes. We put the time interval we're interested in the numerator, so we have:

$$Pr\{0 < Y < 15\} = \frac{15-0}{120-0} = \frac{15}{120} = 0.125 = \frac{1}{8}$$

This is exactly what you would expect, as 15 minutes is $\frac{1}{8}$ of two hours.

5. The T distribution. We won't actually say much about the T distribution here except that it is very useful for calculating probabilities when you have normally distributed data but don't know σ. You might guess this distribution will come up a lot soon (since we almost never know σ).

If you're really interested, you can check out the Wikipedia page on the T distribution. You'll appreciate why the equation isn't given here.

6. The χ^2 distribution. This is another continuous distribution that, like the T distribution, we will use a lot. Again, if you are interested, you can go to Wikipedia for the equation and many other details.

Expected values.

Now we have to deal with a somewhat more theoretical topic. For all these distributions, we need to figure out what the mean (μ) and standard deviation (σ) are. This might seem absurdly simple with something like the normal distribution where the mean is simply μ. But what about other distributions? What is the mean for a binomial distribution? It exists, but so far the only thing we know about a binomial distribution are the parameters, n and p. How does this get us the mean?

Let's just give the answers first. For the binomial, the mean is given as follows:

$$\mu = np$$

Similarly, the standard deviation is given by:

$$\sigma = \sqrt{np(1-p)}$$

Let's see if that makes sense first. For the mean, this is easy to see. Suppose we toss a coin 10 times. What is the average number of heads (how many heads do we expect)? Using the equation for μ we get:

$$\mu = 10 \times 0.5 = 5$$

Which is exactly what we would expect. It's not obvious why the standard deviation is what is given above, but that doesn't stop us from using it (more below):

$$\sigma = \sqrt{10 \times 0.5 \times 0.5} = \sqrt{2.5} = 1.581$$

So how did we know that for the binomial $\mu = np$? Or that $\sigma = \sqrt{np(1-p)}$? As it turns out, there are mathematical definitions for the mean and standard deviation of a distribution. As usual, this is a little different for discrete and continuous distributions.

We will worry just about the mean (see the box for standard deviation if you're interested). For a discrete distribution, the equation for the mean (μ) is given as follows:

$$\mu = E[Y] = \sum_{i=0}^{n} y_i f(y_i)$$

What is $E[X]$? This is called the expected value of our random variable and is defined as given in the equation above. We won't use this equation, but if you plug in the binomial for $f(y_i)$ in the equation above, you will get:

$$\mu = \sum_{i=0}^{n} y_i \binom{n}{y_i} p^{y_i}(1-p)^{n-y_i} = np$$

The math to show this is actually a bit annoying, so we won't show it here; an easier way to see this is with an example. If you plug in $p = .5$ and $n = 3$, you'll find that you get 1.5.

What about a continuous distribution? As you might expect, this involves calculus, so you're not responsible for the following:

$$\mu = E[Y] = \int_{-\infty}^{\infty} y f(y)$$

Notice that the equation is essentially the same, except that the continuous version uses the integral symbol instead of the sum symbol. Again, the point is that using this equation we can figure out the mean (μ) for any continuous distribution (given that it has a mean).

We will only give one equation for the standard deviation:

$$\sigma = \sqrt{E[(Y-\mu)^2]}$$

Notice that this equation contains $E[(Y-\mu)^2]$, so you can probably see that depending on whether our distribution is discrete or continuous, this expression can become quite complicated. If you're really interested, you can refer, as usual, to Wikipedia.

So why do we need to know about how to calculate μ and Expected values? Because we can use this technique to find the mean not just of Y, but also of variables like $\bar{Y}$. If, for example, we know the distribution of $\bar{Y}$, we can calculate both the mean and standard deviation for $\bar{Y}$. This will be important in the next chapter when we start examining probabilities associated not with Y, but with $\bar{Y}$.