How likely is equally likely?

Examining a 'coincidence' in a Statistics lesson

Jonny Griffiths

I am discussing with my AS statistics students the idea of the most likely value for X when X is distributed binomially. Turning to the board, I make up an example on the hoof.

'If X ~ B(31, 3/4), then what is the most likely value for X?'

My students settle dutifully to work. A few minutes later, after much button-pressing on calculators, there is a certain amount of amazement from the floor. 'The probability for X being 23 and X being 24; they're the same!' It's true – they both come out at 0.1610408092... One or two look at me accusingly, as though I have purposefully presented them with this very special case.

I'm a little amazed myself. But... how special a case is it, in fact? In statistics there is often, I have learnt down the years, a reason for this kind of coincidence. Just how likely is this apparent piece of serendipity?

The rules with which I work on this topic are these: firstly, if X ~ B(n, p) and np is an integer, that will be the most likely value. Secondly, if np is not a whole number, then the most likely value will be one of the integers on either side of np. Once students have seen the roughly bell-like shape of the Binomial Distribution (see *Waldomaths*, 2013), they readily agree that these rules make sense.

But thirdly, the probabilities for this second rule have to be checked – simply rounding np to the nearest integer won't always work. I show this to my class using the example above.

'If p is slightly larger than ¾, then np will still round to 23, but we know from what we have done that the most likely value will be 24.' There is surprise around the room at this, but no one can challenge my logic. (For a more detailed examination of these three rules, see *Making Statistics Vital*, 2013).

So equipped with these starting points, could my students make headway with deciding how special the example I'd given them was? In the end, they need plenty of scaffolding from me.

'Let's suppose X ~ B(n, p) where np is not a whole number, with

m < np < m + 1, where m is a whole number.

Let's also suppose that P(X = m) = P(X = m + 1),

as happened in our example,

and so... $^{n}C_{m}p^{m}q^{n-m} = ^{n}C_{m+1}p^{m+1}q^{n-m-1}$.'

My students are able to come up with these probabilities, but some are baffled as to where to go from there. The lightening of the atmosphere when I point out that we can merrily cancel much of this is palpable.

'Cancelling, we have $^{n}C_{m}q = ^{n}C_{m+1}p$, and so

$^{n}C_{m}(1-p) = ^{n}C_{m+1}p$ and rearranging,

$$^{n}C_{m} = (^{n}C_{m+1} + ^{n}C_{m})p\text{.'}$$

Now,' I think, 'A happy opportunity to remind everyone that

$$^{n}C_{m+1} + ^{n}C_{m} = ^{n+1}C_{m+1}\text{.'}$$

Is it only me, or do other classes find this tricky? I tend to sketch a bag containing n balls, with one loose ball outside. So if we need to pick m + 1 balls, we can either pick the loose one and n from the bag, or ignore the rogue ball and take m + 1 from the bag; the identity follows immediately. I find my students gaze at the board in silence for a while, but cross-referencing with Pascal's Triangle and how two consecutive terms add to the term below tends to lead to heads nodding in agreement.

'And so we have $p = \frac{{}^{n}C_{m}}{{}^{n+1}C_{m+1}}$. Now, we have a formula for ${}^{n}C_{m}$...'

This problem is helpfully leading us into revising statistical corners that need it. My students are grateful that they know this formula, and that it's genuinely cropping up in a homemade question.

'Using our factorial formula for ${}^{n}C_{m}$, we arrive at $p = \frac{m+1}{n+1}$.'

A satisfyingly simple conclusion to what looked initially like a complicated problem. I tell my students, perhaps dangerously, that such simplification can often tell us we are on the right track.

'It makes sense to rearrange this making m the subject, so we have

$$m = np + p - 1.'$$

We check our logic carefully, to find that np + p - 1 is a whole number if and only if the phenomenon that we ran into, where two whole numbers are equally likely, takes place.

'So we can see, folks, that in our example,

$$np + p - 1 = 31 \times \frac{3}{4} + \frac{3}{4} - 1 = 23.$$

'So once I have chosen ¾ for p, np + p - 1 will be a whole number if and only if

n = 4k - 1. In other words, if I pluck a whole number n out of the air once I have chosen 0.75 for p, there is a one in four chance that we will experience the 'two equally likely values' phenomenon.'

I look around the class, and the abiding emotion seems to be one of relief. There has been a coincidence, but not one that was as spooky as first feared.

After the lesson, I wonder about a generalisation. If I initially choose p = a/b, with a/b in its lowest terms, then what can we say? We have

$$np + p - 1 = n \times \frac{a}{b} + \frac{a}{b} - 1 = \frac{a(n+1) - b}{b}.$$

This will be a whole number if and only if b divides into n + 1, so if and only if

n = kb-1. The probability of this, if n is plucked out of the air, is 1/b.

References

1. Waldomaths, http://www.waldomaths.com/Binomial1L.jsp (accessed May 2013).

2. Making Statistics Vital, http://www.s253053503.websitehome.co.uk/msv/msv-3.html) (accessed May 2013).

Keywords: Binomial distribution, equally likely, coincidence

Jonny Griffiths, May 2013, jonny.griffiths@ntlworld.com

Maths Department,
Paston College,
Grammar School Road
North Walsham,
NR3 3NA, Norfolk.