UNIVERSITY OF LONDON

Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

2.6 The distribution zoo

Suppose we carry out n **Bernoulli trials** such that:

- at each trial, the probability of success is π
- different trials are statistically independent events.

Let X denote the total number of successes in these n trials, then X follows a **binomial distribution** with parameters n and π, where $n \geq 1$ is a known integer and $0 \leq \pi \leq 1$. This is often written as:

$$\boldsymbol{X \sim \mathrm{Bin}(n,\, \pi).}$$

If $X \sim \mathrm{Bin}(n,\, \pi)$, then:

$$\boldsymbol{\mathrm{E}(X) = n\,\pi.}$$

Example

A multiple choice test has 4 questions, each with 4 possible answers. James is taking the test, but has no idea at all about the correct answers. So he guesses every answer and, therefore, has the probability of $1/4$ of getting any individual question correct.

Let X denote the number of correct answers in James' test. X follows the binomial distribution with $n = 4$ and $\pi = 0.25$, i.e. we have:

$$X \sim \mathrm{Bin}(4,\, 0.25).$$

For example, what is the probability that James gets 3 of the 4 questions correct?

Here it is assumed that the guesses are independent, and each has the probability $\pi = 0.25$ of being correct.

The probability of any particular sequence of 3 correct and 1 incorrect answers, for example 1110, is $\pi^3 (1-\pi)^1$, where '1' denotes a correct answer and '0' denotes an incorrect answer.

However, we do not care about the order of the 0s and 1s, only about the number of 1s. So 1101 and 1011, for example, also count as 3 correct answers. Each of these also has the probability $\pi^3 (1-\pi)^1$.

The total number of sequences with three 1s (and, therefore, one 0) is the number of locations for the three 1s which can be selected in the sequence of 4 answers. This is $\binom{4}{3} = 4$ (see below). Therefore, the probability of obtaining three 1s is:

$$\binom{4}{3} \pi^3 (1-\pi)^1 = 4 \times (0.25)^3 \times (0.75)^1 \approx 0.0469.$$

In general, the **probability function** of $X \sim \mathrm{Bin}(n, \pi)$ is:

$$P(X = x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

where $\binom{n}{x}$ is the **binomial coefficient** – in short, the number of ways of choosing x objects out of n when sampling without replacement when the order of the objects does not matter.

$\binom{n}{x}$ can be calculated as:

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

where $k! = k \times (k-1) \times \cdots \times 3 \times 2 \times 1$, for an integer $k > 0$. Also note that $0! = 1$. For example:

$$\binom{4}{3} = \frac{4!}{3!\,(4-3)!} = \frac{4!}{3!\,1!} = \frac{4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) \times 1} = \frac{24}{6 \times 1} = 4.$$

Example

Now assume there are $n = 20$ questions, each with 4 possible answers.

More generally, consider a student who has the same probability π of the correct answer for every question, so that $X \sim \mathrm{Bin}(20, \pi)$.

The figure below shows plots of the probabilities for $\pi = 0.25$, 0.5, 0.7 and 0.9 (reflecting students of differing abilities, i.e. the better the student the more likely s/he is to get the answer correct and hence a higher π).

Note that as π increases, the probability of obtaining a large number of correct answers increases (and hence the probability of obtaining a small number of correct answers decreases) as we would expect because better-prepared students tend to score higher marks. Of course, there is an opportunity cost: it takes more time and effort to prepare, but this on average is rewarded with high marks!