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

Module Leader: Dr James Abdey

2.5 Parameters

Probability distributions may differ from each other in a broader or narrower sense. In the broader sense, we have different **families** of distributions which may have quite different characteristics, for example:

- discrete distributions versus continuous distributions
- among discrete distributions: a finite versus an infinite number of possible values
- among continuous distributions: different sets of possible values (for example, all real numbers x, $x > 0$, or $x \in [0, 1]$); symmetric versus skewed distributions.

These 'distributions' are really families of distributions in this sense.

In the narrower sense, individual distributions *within* a family differ in having different values of the **parameters** of the distribution. The parameters determine the mean and variance of the distribution, values of probabilities from it etc.

In the statistical analysis of a random variable X we typically:

- select a *family* of distributions based on the basic characteristics of X
- use observed data to choose (**estimate**) values for the parameters of that distribution, and perform statistical inference on them.

Example

An opinion poll on a referendum, where each X_i is an answer to the question 'Will you vote 'Yes' or 'No' to joining/leaving[1] the European Union?' has answers recorded as $X_i = 0$ if 'No' and $X_i = 1$ if 'Yes'. In a poll of 950 people, 513 answered 'Yes'.

[1] In light of Brexit, you can choose which you prefer!

How do we choose a distribution to represent X_i?

- Here we need a family of discrete distributions with only two possible values (0 and 1). The **Bernoulli distribution** (discussed below), which has one parameter π (the probability that $X_i = 1$) is appropriate.
- Within the family of Bernoulli distributions, we use the one where the value of π is our best estimate based on the observed data. This is $\widehat{\pi} = 513/950 = 0.54$ (where $\widehat{\pi}$ denotes an **estimate of the parameter** π).

A **Bernoulli trial** is an experiment with only *two* possible outcomes. We will number these outcomes 1 and 0, and refer to them as 'success' and 'failure', respectively. Note these are notional successes and failures – the success does not necessarily have to be a 'good' outcome, nor a failure a 'bad' outcome!

Example

Examples of outcomes of Bernoulli trials are:

- agree / disagree
- pass a test / fail a test
- employed / unemployed
- owns a car / does not own a car
- business goes bankrupt / business continues trading.

The **Bernoulli distribution** is the distribution of the outcome of a single Bernoulli trial, named after Jacob Bernoulli (1654–1705). This is the distribution of a random variable X with the following **probability function**:[2]

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

Therefore:

$$P(X = 1) = \pi^1\,(1-\pi)^{1-1} = \pi$$

and:

$$P(X = 0) = \pi^0\,(1-\pi)^{1-0} = 1 - \pi$$

and no other values are possible. We could express this family of Bernoulli distributions in tabular form as follows:

$X = x$	0	1
$P(X = x)$	$1 - \pi$	π

[2]A probability function is simply a function which returns the probability of a particular value of X.

where $0 \leq \pi \leq 1$ is the probability of 'success'. Note that just as a **sample space** represents all possible values of a random variable, a **parameter space** represents all possible values of a parameter. Clearly, as a probability, we must have that $0 \leq \pi \leq 1$.

Such a random variable X has a Bernoulli distribution with (probability) parameter π. This is often written as:

$$\boldsymbol{X \sim \text{Bernoulli}(\pi).}$$

If $X \sim \text{Bernoulli}(\pi)$, then we can determine its expected value, i.e. its mean, as the usual **probability-weighted average**:

$$\mathbf{E}(\boldsymbol{X}) = \mathbf{0} \times (\mathbf{1} - \boldsymbol{\pi}) + \mathbf{1} \times \boldsymbol{\pi} = \boldsymbol{\pi}.$$

Hence we can view π as the long-run average (proportion) of successes if we were to draw a large random sample from this distribution.

Different members of this family of distributions differ in terms of the value of π.

Example

Consider the toss of a *fair* coin, where $X = 1$ denotes 'heads' and $X = 0$ denotes 'tails'. As this is a fair coin, heads and tails are equally likely and hence $\pi = 0.5$ leading to the specific Bernoulli distribution:

$X = x$	0	1
$P(X = x)$	0.5	0.5

Hence:

$$\text{E}(X) = 0 \times 0.5 + 1 \times 0.5 = 0.5$$

such that if we tossed a fair coin a large number of times, we would *expect* the proportion of heads to be 0.5 (and in practice the long-run proportion of heads would be *approximately* 0.5)[3].

[3] Get a fair coin, and test this!