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

## Module Leader: Dr James Abdey

## 2.3 Expectation of random variables

Certain important properties of distributions arise if we consider **probability-weighted averages** of random variables, and of functions of random variables.[1] For example, we might want to know the '*average*' value of a random variable.

It would be foolish to simply take the arithmetic average of all the values taken by the random variable, as this would mean that very unlikely values (those with small probabilities of occurrence) would receive the same weighting as very likely values (those with large probabilities of occurrence). The obvious approach is to use the probability-weighted average of the sample space values, known as the **expected value** of $X$.

If $x_1, x_2, \ldots, x_N$ are the possible values of the random variable $X$, with corresponding probabilities $p(x_1), p(x_2), \ldots, p(x_N)$, then:

$$\mathrm{E}(X) = \mu = \sum_{i=1}^{N} x_i\, p(x_i) = x_1\, p(x_1) + x_2\, p(x_2) + \cdots + x_N\, p(x_N).$$

Note that the expected value is also referred to as the **population mean**, which can be written as $\mathrm{E}(X)$ (in words 'the expectation of the random variable $X$') or $\mu$ (in words 'the (population) mean of $X$'). So, for so-called 'discrete' random variables, $\mathrm{E}(X)$ is determined by taking the product of each value of $X$ and its corresponding probability, and then summing across all values of $X$.

### Example

If the 'random variable' $X$ happens to be a **constant**, $k$, then $x_1 = k$, and $p_1 = 1$, so trivially $\mathrm{E}(X) = k \times 1 = k$. Of course, here $X$ is not 'random', but a constant and hence its expectation is $k$ as it can only ever take the value $k$!

[1] A function, $f(X)$, of a random variable $X$ is, of course, a new random variable, say $Y = f(X)$.

## Example

Let $X$ represent the value shown when a fair die is thrown once.

| $X = x$ | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| $P(X = x)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1 |
| $x\,P(X = x)$ | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 | 6/6 | $21/6 = 3.5 = \mu$ |

Hence:

$$\mu = \mathrm{E}(X) = \sum_{x=1}^{6} x\,P(X = x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5.$$

We should view 3.5 as a **long-run average** since, clearly, the score from a *single* roll of a die can never be 3.5, as it is not a member of the sample space. However, if we rolled the die a (very) large number of times, then the average of all of these outcomes would be (approximately) 3.5. For example, suppose we rolled the die 600 times and observed the frequencies of each score. Let us suppose we observed the following frequencies:

| $X = x$ | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| Frequencies | 96 | 102 | 101 | 98 | 107 | 96 | 600 |

The average observed score is:

$$\frac{1 \times 96 + 2 \times 102 + 3 \times 101 + 4 \times 98 + 5 \times 107 + 6 \times 96}{600} = 3.51 \approx 3.5.$$

So we see that in the long run the average score is approximately 3.5. Note a different 600 rolls of the die might lead to a different set of frequencies. Although we might *expect* 100 occurrences of each score of 1 to 6 (that is, taking a relative frequency interpretation of probability, as each score occurs with a probability of 1/6 we would *expect* one sixth of the time to observe each score of 1 to 6), it is unlikely we would observe *exactly* 100 occurrences of each score in practice.

## Example

Recall the toss of a fair coin, where we define the random variable $X$ such that:

$$X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails.} \end{cases}$$

Since the coin is fair, then $P(X = 0) = P(X = 1) = 0.5$, hence:

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

Here, viewed as a long-run average, $\mathrm{E}(X) = 0.5$ can be interpreted as the **proportion** of heads in the long run (and, of course, the proportion of tails too).

## Example

Let us consider the game of roulette, from the point of view of the casino ('The House').

Suppose a player puts a bet of £1 on 'red'. If the ball lands on any of the 18 red numbers, the player gets that £1 back, plus another £1 from The House. If the result is one of the 18 black numbers or the green 0, the player loses the £1 to The House.

We assume that the roulette wheel is **unbiased**, i.e. that all 37 numbers[2] have equal probabilities. What can we say about the probabilities and expected values of wins and losses?

Define the random variable $X$ = 'money received by The House'. Its possible values are $-1$ (the player wins) and 1 (the player loses). The **probability function** is:

$$P(X = x) = p(x) = \begin{cases} 18/37 & \text{for } x = -1 \\ 19/37 & \text{for } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $p(x)$ is a shortened version of $P(X = x)$. Therefore, the expected value is:

$$\mathrm{E}(X) = \left(-1 \times \frac{18}{37}\right) + \left(1 \times \frac{19}{37}\right) = +0.027.$$

*On average*, The House expects to win 2.7p for every £1 which players bet on red. This expected gain is known as the *house edge*. It is positive for all possible bets in roulette.[3]

## Expected value versus sample mean

The mean (expected value) $\mathrm{E}(X)$ of a probability distribution is analogous to the sample mean (average) $\bar{X}$ of a sample distribution (introduced in week 3). This is easiest to see when the sample space is finite.

Suppose the random variable $X$ can have $K$ different values $X_1, \ldots, X_K$, and their **frequencies** in a random sample are $f_1, \ldots, f_K$, respectively. Therefore, the sample mean of $X$ is:

$$\bar{X} = \frac{f_1 x_1 + \cdots + f_K x_K}{f_1 + \cdots + f_K} = x_1 \,\widehat{p}(x_1) + \cdots + x_K \,\widehat{p}(x_K) = \sum_{i=1}^{K} x_i \,\widehat{p}(x_i)$$

where:

$$\widehat{p}(x_i) = \frac{f_i}{\sum_{i=1}^{K} f_i}$$

are the **sample proportions** of the values $x_i$. The expected value of the random variable $X$ is:

$$\mathrm{E}(X) = x_1 \,p(x_1) + \cdots + x_K \,p(x_K) = \sum_{i=1}^{K} x_i \,p(x_i).$$

So $\bar{X}$ uses the sample proportions, $\widehat{p}(x_i)$, whereas $\mathrm{E}(X)$ uses the population probabilities, $p(x_i)$.

[2]We assume a 'European' style roulette wheel with a single zero.

[3]The house edge must be positive such that the casino expects a positive profit in the long run!