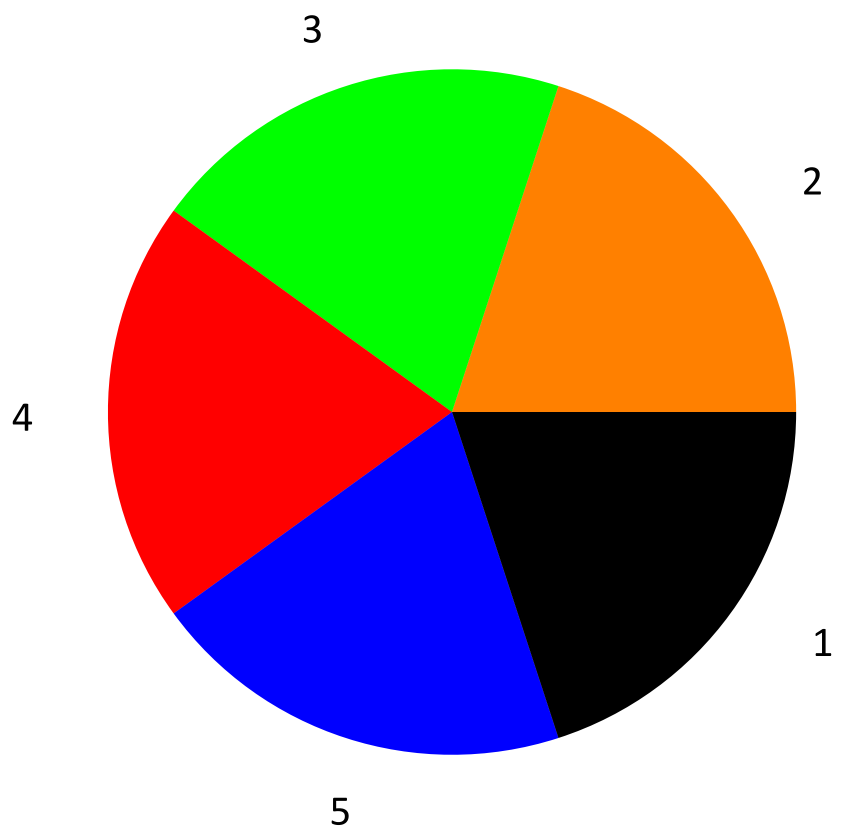
# Expected Value Lecture 1:

## Motivation

A gambler wants to play the wheel of fortune. The wheel is divided into 5 equal slices labeled 1-5. If the wheel stops spinning at 1, the player receives $20. If it stops at either 2 or 4, the player is paid $50. And if the wheel comes to rest at either 3 or 5, the player is awarded $100. The game costs $65 per round. What will be the financial consequences for the gambler that makes a habit of this game?




For each outcome of the game $s \in \{1, 2, 3, 4, 5\}$ a net pay $X(s) = win(s) - 65$ is generated. If we suppose that the gambler decides to play 20 rounds today and 20 rounds tomorrow, his net winnings could be as in the table below:

| Trial | Run 1 | Pay | Run 2 | Pay |
|---|---|---|---|---|
| 1 | 2 | -15 | 3 | 35 |
| 2 | 1 | -45 | 3 | 35 |
| 3 | 5 | 35 | 3 | 35 |
| 4 | 4 | -15 | 1 | -45 |
| 5 | 5 | 35 | 2 | -15 |
| 6 | 5 | 35 | 5 | 35 |
| 7 | 1 | -45 | 2 | -15 |
| 8 | 2 | -15 | 5 | 35 |
| 9 | 2 | -15 | 3 | 35 |
| 10 | 5 | 35 | 2 | -15 |
| 11 | 5 | 35 | 3 | 35 |
| 12 | 4 | -15 | 1 | -45 |
| 13 | 4 | -15 | 5 | 35 |
| 14 | 5 | 35 | 3 | 35 |
| 15 | 1 | -45 | 2 | -15 |
| 16 | 4 | -15 | 1 | -45 |
| 17 | 1 | -45 | 2 | -15 |
| 18 | 3 | 35 | 3 | 35 |
| 19 | 5 | 35 | 1 | -45 |
| 20 | 1 | -45 | 5 | 35 |
| **Total** | | **-50** | | **130** |

Notice the significant variability in the gambler's fortune from one day to the next. When the number of rounds is small, much is in the hands of lady Luck! Let us now see what happens when we increase the number of rounds. But this time, to make the table shorter, we will count the number of times each outcome $s$ occurs and forget about the specific trials that produced it.

**<u>100 games per run</u>**

| S | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 |
|---|---|---|---|---|---|
| 1 | 17 | 26 | 20 | 24 | 22 |
| 2 | 26 | 18 | 21 | 17 | 20 |
| 3 | 21 | 18 | 16 | 21 | 27 |
| 4 | 22 | 20 | 18 | 24 | 14 |
| 5 | 14 | 18 | 25 | 14 | 17 |
| Total Pay | **-260** | **-480** | **-50** | **-470** | **40** |
| Ave Game | **-2.6** | **-4.8** | **-0.5** | **-4.7** | **0.4** |

Where the Total Pay and Average [Pay per] Game were computed with the formulas

$$n(1) \cdot X(1) + n(2) \cdot X(2) + n(3) \cdot X(3) + n(4) \cdot X(4) + n(5) \cdot X(5)$$

$$\frac{n(1) \cdot X(1) + n(2) \cdot X(2) + n(3) \cdot X(3) + n(4) \cdot X(4) + n(5) \cdot X(5)}{100}$$

in which $n(k)$ stands for the number of times outcome $k \in \{1,2,3,4,5\}$ was witnessed during the run and where $X(k)$ is the corresponding payoff.

As you can see, there is still a lot of variability between the runs. So let's increase the number of games per run!

**10 000 games per run**

| S | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 |
|---|---|---|---|---|---|
| 1 | 2013 | 2013 | 2040 | 2053 | 2045 |
| 2 | 1966 | 2017 | 2026 | 2066 | 2037 |
| 3 | 2102 | 2006 | 2065 | 1982 | 1959 |
| 4 | 1900 | 1976 | 1934 | 1966 | 2011 |
| 5 | 2019 | 1988 | 1935 | 1933 | 1948 |
| **Total Pay** | **-4340** | **-10 690** | **-11 200** | **-15 840** | **-16 000** |
| **Ave Game** | **-0.434** | **-1.069** | **-1.12** | **-1.584** | **-1.6** |

The calculations of the total pay and average winnings per game are exactly the same as for the previous table, except that this time, $\sum_{k=1}^{5} n(k) = 10\ 000$ and we naturally divide the total pay by the number of games, $n = 10\ 000$, to get the average payoff per game.

And finally, let us simulate one million games per run:

**1000 000 games per run**

| S | Run 1 | Run 2 | Run 3 | Run 4 | Run 5 |
|---|---|---|---|---|---|
| 1 | 199 815 | 199 680 | 199 977 | 199 872 | 200 333 |
| 2 | 200 321 | 199 613 | 200 222 | 200 217 | 200 527 |
| 3 | 200 534 | 200 581 | 200 059 | 200 371 | 199 525 |
| 4 | 199 373 | 199 984 | 199 899 | 200 226 | 199 761 |
| 5 | 199 957 | 200 142 | 199 843 | 199 314 | 199 854 |
| **Total Pay** | **-969 900** | **-954 250** | **-1 004 210** | **-1 011 910** | **-1 041 040** |
| **Ave Game** | **-0.9699** | **-0.95425** | **-1.00421** | **-1.01191** | **-1.04104** |

When we compare the red values in all three tables, we might suspect that the red values are tending to -1 as $n$ - the number of games per run - grows to infinity. Indeed,

$$\lim_{n\to\infty} \frac{n(1)\cdot X(1) + n(2)\cdot X(2) + n(3)\cdot X(3) + n(4)\cdot X(4) + n(5)\cdot X(5)}{n} =$$

$$\lim_{n\to\infty} \frac{n(1)}{n}\cdot X(1) + \lim_{n\to\infty} \frac{n(2)}{n}\cdot X(2) + \lim_{n\to\infty} \frac{n(3)}{n}\cdot X(3) + \lim_{n\to\infty} \frac{n(4)}{n}\cdot X(4) + \lim_{n\to\infty} \frac{n(5)}{n}\cdot X(5) =$$

$$0.2\cdot(-45) + 0.2\cdot(-15) + 0.2\cdot(35) + 0.2\cdot(-15) + 0.2\cdot(35) = -1$$

where we see that $\lim_{n\to\infty} \frac{n(k)}{n} = \frac{1}{5} = 0.2$, because each outcome $k$ is equally likely.

The value -1 implies that when the gambler plays more and more games, the whims of Fortuna lose their sway and the outcome becomes predictable; the gambler's payoff per game ends up being essentially the same as if the wheel of fortune were replaced by a pauper's request to give him a dollar.

The last table confirms this. It also gives us confidence that $\lim_{n\to\infty} \frac{n(k)}{n} = 0.2$. For if we look at Run 3, for instance, we obtain the estimates

$$\lim_{n\to\infty} \frac{n(1)}{n} \approx \frac{199\ 977}{1\ 000\ 000} = 0.199977$$

$$\lim_{n\to\infty} \frac{n(2)}{n} \approx \frac{200\ 222}{1\ 000\ 000} = 0.200222$$

$$\lim_{n\to\infty} \frac{n(3)}{n} \approx \frac{200\ 059}{1\ 000\ 000} = 0.200059$$

$$\lim_{n\to\infty} \frac{n(4)}{n} \approx \frac{199\ 899}{1\ 000\ 000} = 0.199899$$

$$\lim_{n\to\infty} \frac{n(5)}{n} \approx \frac{199\ 843}{1\ 000\ 000} = 0.199843$$

## The Definition of Expected Value

Expected value is a metric that measures the average consequence of a random action, when this action is preformed repeatedly. Specifically, when we are given a sample space $S$ and a random variable $X: S \to S^* \subseteq \mathbb{R}^n$, the expected value is the weighted average

$$E[X] = \sum_{s \in S} X(s)p(s)$$

We have already computed the expected value for the wheel of fortune, but let us look at a more detailed table and compute it again. We will carry out the calculation in several ways. This will reveal an essential property of expected value that will simplify many future calculations.

In the first two calculations, we are going to sort the table by outcomes s before averaging the payoffs.

## Experimental and ideal tables sorted by s

| s | Outcomes | p(s) | X(s) | X(s)p(s) |
|---|---|---|---|---|
| 1 | 199 977 | 0.199977 | -45 | -8.998965 |
| 2 | 200 222 | 0.200222 | -15 | -3.00333 |
| 3 | 200 059 | 0.200059 | 35 | 7.002065 |
| 4 | 199 899 | 0.199899 | -15 | -2.998485 |
| 5 | 199 843 | 0.199843 | 35 | 6.994505 |
| **Total** | **1 000 000** | **1** | **-5** | **-1.00421** |

| s | p(s) | X(s) | X(s)p(s) |
|---|---|---|---|
| 1 | 0.2 | -45 | -9 |
| 2 | 0.2 | -15 | -3 |
| 3 | 0.2 | 35 | 7 |
| 4 | 0.2 | -15 | -3 |
| 5 | 0.2 | 35 | 7 |
| **Total** | **1** | **-5** | **-1** |

We can also rearrange the tables by grouping together equal payoffs. This yields

## Experimental and ideal tables sorted by X

| s | Outcomes | p(s) | X(s) | X(s)p(s) |
|---|---|---|---|---|
| 1 | 199 977 | 0.199977 | -45 | -8.998965 |
| 2 | 200 222 | 0.200222 | -15 | -3.00333 |
| 3 | 200 059 | 0.200059 | 35 | 7.002065 |
| 4 | 199 899 | 0.199899 | -15 | -2.998485 |
| 5 | 199 843 | 0.199843 | 35 | 6.994505 |
| **Total** | **1 000 000** | **1** | **-5** | **-1.00421** |

| s | Outcomes | p(s) | X(s) | X(s)p(s) |
|---|---|---|---|---|
| 1 | 199 977 | 0.199977 | -45 | -8.998965 |
| 2 | 200 222 | 0.200222 | -15 | -3.00333 |
| 4 | 199 899 | 0.199899 | -15 | -2.998485 |
| 3 | 200 059 | 0.200059 | 35 | 7.002065 |
| 5 | 199 843 | 0.199843 | 35 | 6.994505 |
| **Total** | **1 000 000** | **1** | **-5** | **-1.00421** |

⇩

| s | Outcomes | p(X) | X | Xp(X) |
|---|---|---|---|---|
| 1 | 199 977 | 0.199977 | -45 | -8.998965 |
| 2&4 | 400 121 | 0.400121 | -15 | -6.001815 |
| 3&5 | 399 902 | 0.399902 | 35 | 13.99657 |
| **Total** | **1 000 000** | **1** | **-5** | **-1.00421** |

Similarly, we can rearrange the ideal table by the column of payoffs X:

| s | p(s) | X(s) | X(s)p(s) |
|---|---|---|---|
| 1 | 0.2 | -45 | -9 |
| 2 | 0.2 | -15 | -3 |
| 3 | 0.2 | 35 | 7 |
| 4 | 0.2 | -15 | -3 |
| 5 | 0.2 | 35 | 7 |
| **Total** | **1** | **-5** | **-1** |

⇩

| s | p(X) | X | Xp(X) |
|---|---|---|---|
| 1 | 0.2 | -45 | -9 |
| 2&4 | 0.4 | -15 | -6 |
| 3&5 | 0.4 | 35 | 14 |
| **Total** | **1** | **-5** | **-1** |

We now have a general sense of how to calculate expected value in two seemingly different ways and can now describe why the two calculations provide the same solution.

Let $X: S \to S^* \subseteq \mathbb{R}^q$ be a random variable with range $S^* = \{x_1 \;\; x_2 \;\; \dots \;\; x_m\}$. Then

$$E[x] = \lim_{n\to\infty} \frac{\sum_{s\in S} X(s)n(s)}{n} = \sum\nolimits_{s\in S} X(s) \lim_{n\to\infty} \frac{n(s)}{n} = \sum\nolimits_{s\in S} X(s)p(s)$$

The last sum can be rearranged as

$$\sum\nolimits_{s\in S} X(s)p(s) = \sum_{k=1}^{m} x_k \sum\nolimits_{s:X(s)=x_k} p(s) = \sum_{k=1}^{m} x_k p(x_k)$$

Hence we may define

$$E[x] = \sum\nolimits_{s\in S} X(s)p(s) = \sum_{k=1}^{m} x_k p(x_k)$$

Again, this is not as scary as it looks. In the fortune wheel example, these equations are simply

$$X: \{1, 2, 3, 4, 5\} \to \{x_1 = -45, x_2 = -15, x_3 = 35\}$$

$$E[X] = \sum\nolimits_{s\in S} X(s)p(s) = \sum_{j=1}^{5} X(j)p(j) =$$

$$(-45)\cdot 0.2 + (-15)\cdot 0.2 + (35)\cdot 0.2 + (-15)\cdot 0.2 + (35)\cdot 0.2 =$$

$$(-45)\cdot 0.2 + (-15)\cdot (0.2 + 0.2) + (35)\cdot (0.2 + 0.2) = \sum_{k=1}^{3} x_k \, p(x_k)$$