

Contents

Expected Value of a Random Variable	1
Linearity of Expected Value	3
Expected Value of Independent Random Variables	3
Expected Value of a Random Variable	3
Gamblers Ruin	4
Variance	4
Variance of Independent Random Variables	5
Expected Value of Binomial Random Variable	5
Anscombe's Quartet	5

Expected Value of a Random Variable

- Sometimes we want to make a decision under uncertainty, e.g. in a game of chance I throw a six-sided die and win €5 if it comes up 6 and otherwise lose €1, should I play? What if win €6? Or €7?
- Sometimes we have a number of measurements that we want to summarise by a single value, e.g. measurements of the time it takes you to travel to Trinity each day.
 - We we play the game N times, with N large, we expect a 6 to appear $\frac{1}{6} \times N$ of the time and another number of appear $\frac{5}{6} \times N$ of the time
 - So our overall winnings are expected to be $\frac{1}{6}N \times (5-1) + \frac{5}{6}N \times (0-1) = (\frac{4}{6} - \frac{5}{6})N = -\frac{1}{6}N$

The **Expected Value** of discrete random variable X taking values in $\{x_1, x_2, \dots, x_n\}$ is defined to be:

$$E[X] = \sum_{i=1}^n x_i P(X = x_i)$$

Also referred to as the **mean** or **average**

- Values x_i for with $P(X = x_i) = 0$ don't contribute
- Values of x_i with higher probability $P(X = x_i)$ contribute more

- Viewing the probability of an event as the frequency with which it occurs when an experiment is repeated many times, the expected value tells us about the overall outcome we can expect

An important example:

- Suppose I is the indicator variable for event E (so $I = 1$ if the event E occurs, $I = 0$ otherwise)
- Then $E[I] = 1 \times P(E) + 0 \times (1 - P(E)) = P(E)$
- The expected value of I is the probability that event E occurs

E.g. Suppose we play a game and RV I equals 1 when we win and I equals 0 otherwise, then $E[I]$ is the probability of winning

Suppose we keep throwing a die until six comes up. On average how many times do we need to throw the die before a six appears?

- Let random variable X be the number of die throws
- $P(X = 0) = \frac{1}{6}$ (we throw a 6 first time)
- $P(X = 1) = (X - \frac{1}{6})\frac{1}{6}$ (we throw a non-six and then a six)
- $P(X = 2) = (X - \frac{1}{6})^2\frac{1}{6}$ (we throw a non-six twice and then a six)
- and so on...
- $E[X] = \sum_{i=1}^{\infty} i \times (1 - \frac{1}{6})^i \frac{1}{6}$ which has a value of 5

Another use of the expected value: sometimes we have a number of measurements that we want to summarise by a single value.

Example: in 2011 Irish census:

No. of children	No. of families
0	344,944
1	339,596
2	285,952
3	144,470
>3	75,248

- Total no. of families: $344,944 + 339,596 + 285,952 + 144,470 + 75,248 = 1,190,210$
- $P(\text{no children}) = \frac{344944}{1190210}$, $P(1 \text{ child}) = \frac{339596}{1190210}$, etc.
- Expected value $\sum_{i=1}^n x_i P(X = x_i) = 0 \times \frac{344944}{1190210} + 1 \times \frac{339596}{1190210} + 2 \dots = 1.38$

What does expected value mean?

- Total no. of children = $0 \times 344944 + 1 \times 339596 + 2 \times 285952 + 3 \times 144470 + 4 \times 75248$
- Expected value is the Total no. of children/Total no. of families
- So if all families had the same number of children, the expected value is the value that would maintain the right total number of children

What about experiment repetition (frequency interpretation of probability)?

- Pick a family at random from the population, number of children is the “reward”. Repeat many times...

Of course no family actually has 1.38 children...and there are choices of summary value other than the expected value, e.g. median, mode

Linearity of Expected Value

For any random variable X and constants a and b :

$$E[aX + b] = aE[X] + b$$

For any two random variables X and Y and constants a and b :

$$E[aX + bY] = aE[X] + bE[Y]$$

More generally,

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i]$$

for random variables X_1, X_2, \dots, X_n

Expected Value of Independent Random Variables

Take two **independent** random variables X and Y :

$$E[XY] = E[X]E[Y]$$

Expected Value of a Random Variable

- Expected value is the first moment of random variable X , $E[X] = \sum_{i=1}^n x_i p(x_i)$
- N 'th moment of X is $E[X^N] = \sum_{i=1}^n x_i^N p(x_i)$, will see a use for this shortly

Gamblers Ruin

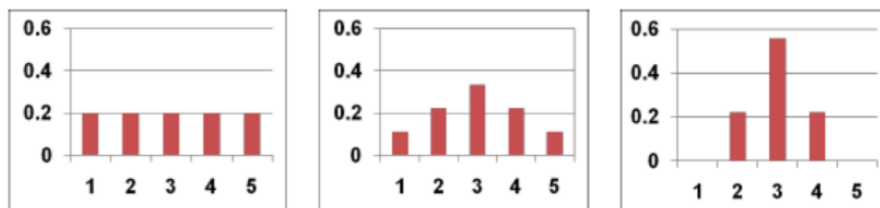
Roulette 18 red, 18 black, 1 green (37 total)

- Bet on red, $p = \frac{18}{37}$ to win €1, otherwise $1 - p$ you lose €1
- Bet €1
- If win then stop, if lose then double bet and repeat
- Random variable X is winnings on stopping

$$E[X] = p \times 1 + (1 - p)p \times (2 - 1) + (1 - p)^2 p \times (4 - 2 - 1) + \dots = \sum_{i=1}^{\infty} (1 - p)^i p (2^i - \sum_{j=1}^{i-1} 2^j)$$

- Expected winnings are > 0 so why don't we just play infinitely often?
- You have finite money! Usually also a max bet.

Variance



- All have the same expected value, $E[X] = 3$
- But “spread” is different
- Variance is a summary value (a statistic) that quantified “spread”

Let X be a random variable with mean μ . The **variance** of X is

$$Var(X) = E[(X - \mu)^2]$$

- Discrete random variable taking values in $D = \{x_1, x_2, \dots, x_n\}$

$$Var(X) = \sum_{i=1}^n (x_i - \mu)^2 P(x_i)$$

with $\mu = E[X] = \sum_{i=1}^n x_i P(x_i)$

Example: flip a coin; $X = 1$ if heads, 0 otherwise.

$$E[X] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$$

$$Var(X) = (1 - \frac{1}{2})^2 \times \frac{1}{2} + (0 - \frac{1}{2})^2 \times \frac{1}{2} = \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{1}{4}$$

- Variance is mean squared distance of X from the mean μ
- $Var(X) \geq 0$
- Standard deviation is square root of variance $\sqrt{Var(X)}$

Discrete random variable taking values in $D = \{x_1, x_2, \dots, x_n\}$. An alternative expression for variance is:

$$Var(X) = E[X^2] - (E[X])^2$$

Unlike expectation, variance is not linear. Instead we have

$$Var(aX + b) = a^2 Var(X)$$

Observe that the offset b does not affect the variance.

Variance of Independent Random Variables

For independent random variables X and Y then

$$Var(X + Y) = Var(X) + Var(Y)$$

Expected Value of Binomial Random Variable

- Bernoulli random variable, $X \sim Ber(p)$:

$$E[X] = p$$

$$Var(x) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$

- Binomial random variable, $X \sim Bin(n, p)$. Sum of n $Ber(p)$ independent random variables so:

$$E[X] = np$$

$$Var(X) = np(1 - p)$$

Anscombe's Quartet

The variable is another example of a summary value, this time out that indicated the spread in a data set. But great care is again needed.

Plot the data, don't just rely on summary values.