Overview

- Expected Value of a Random Variable
- Linearity of Expected Value
- Variance

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 siz-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.

Expected Value of a Random Variable

Suppose we toss a six sided die. We pay €1 to play and win €5 if it comes up 6 (otherwise win nothing). Should we play the game ? If win €6 should we play ? Or €7 ?

- If we play the game N times, with N large, we expect a 6 to appear $1/6 \times N$ of the time and another number to appear $5/6 \times N$ of the time.
- So our overall winnings are expected to be

$$\frac{1}{6}N \times (5-1) + \frac{5}{6}N \times (0-1) = (\frac{4}{6} - \frac{5}{6})N = -\frac{1}{6}N$$

- and our expected winnings per play are (divide by N):

$$\underbrace{\frac{1}{6}}_{P(get\ \ a\ \ 6)} \times (5-1) + \underbrace{\frac{5}{6}}_{P(don't\ \ get\ \ a\ \ 6)} \times (0-1) = (\frac{4}{6} - \frac{5}{6}) = -\frac{1}{6}$$

Expected Value of a Random Variable

The **Expected Value** of discrete random variable X taking values in $\{x_1, x_2, \cdots, x_n\}$ 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.

Expected Value of a Random Variable

An important example:

- Suppose I is the indicator variable for event E (so $I = 1$ if 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.

Expected Value of a Random Variable

- Suppose we play the lottery and pay €1 per play. There are two possible outcomes, win or lose. Random variable X is 10^3 if win, -1 (price of ticket) if lose. $P(X = 1M) = p = 1/10^6$, $P(X = -1) = 1 - p$. If we play the lottery many times, our average return is $E[X] = 10^3/10^6 - 1 \times (1 - 1/10^6) = -0.999$. If we don't play the lottery then our average return is 0, but higher than -0.999.
- I run an investment bank. With probability 0.99 I make profit of €1000. With probability (1-0.99)=0.01 I lose €100 billion. My expected return is $0.99 \times 1000 - 0.01 \times 100 \times 10^9 = -9,999,010$.

Expected Value of a Random Variable

Suppose we keep throwing a die until a 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)
- $P(X = 1) = (1 - \frac{1}{6})\frac{1}{6}$ (we throw a non-six and then a six)
- $P(X = 2) = (1 - \frac{1}{6})^2\frac{1}{6}$ (we throw a non-six twice, and then a six)
- and so on.
- $E[X] = \sum_{i=0}^{\infty} i \times (1 - \frac{1}{6})^i \frac{1}{6}$, which has a value of 5.

Expected Value of a Random Variable

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[1]:

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

- Total no. of families:
  344,944+339,596+285,952+144,470+75,248=1,190,210
- $P(no\ children) = \frac{344944}{1190210}$, $P(1\ child) = \frac{339596}{1190210}$ etc
- Expected value $\sum_{i=1}^{n} x_i P(X = x_i) =$
  $0 \times \frac{344944}{1190210} + 1 \times \frac{339596}{1190210} + 2 \times \frac{285952}{1190210} + 3 \times \frac{144470}{1190210} + 4 \times \frac{75248}{1190210} = 1.38$

[1] http://www.cso.ie/en/census/census2011reports/census2011profile5householdsandfamilies-livingarrangementsinireland/

Expected Value of a Random Variable

What does expected value mean here ?

- Total no. of children
  $=0 \times 344,944 + 1 \times 339,596 + 2 \times 285,952 + 3 \times 144,470 + 4 \times 75,248.$
  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

Let's start by proving that

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

for any random variable X and constants a and b.
Proof: Suppose random variable X takes values x_1, $x_2, \ldots,$ x_n. Then,

$$\begin{aligned} E[aX + b] &= \sum_{i=1}^{n} (ax_i + b) P(X = x_i) \\ &= \sum_{i=1}^{n} ax_i P(X = x_i) + \sum_{i=1}^{n} bP(X = x_i) \\ &= a \sum_{i=1}^{n} x_i P(X = x_i) + b \sum_{i=1}^{n} P(X = x_i) \\ &= aE[X] + b \end{aligned}$$

Example (revisited)

- Suppose we toss a six sided die. We pay €1 to play and win €5 if it comes up 6 (otherwise win nothing). Recall $E[X] = -\frac{1}{6}$, where X is our winnings per play.
- Now change to using pounds, with £1=€1.15. That is, pay €1.15 to play and win €5.75 if comes up 6.
- Our expected winnings per play are now :

$$\underbrace{\frac{1}{6}}_{P(get\ \ a\ \ 6)} \times (5.75 - 1.15) + \underbrace{\frac{5}{6}}_{P(don't\ \ get\ \ a\ \ 6)} \times (0 - 1.15) = -\frac{1.15}{6}$$

- Or, using linearity, immediately have $E[1.15X] = 1.15E[X] = -\frac{1.15}{6}$.

Linearity of Expected Value

Now we extend our analysis to show that

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

for any two random variables X and Y and constants a and b.
Proof:

$$\begin{aligned}
E[aX + bY] &= \sum_x \sum_y (ax + by) P(X = x \text{ and } Y = y) \\
&= a \sum_x \sum_y x P(X = x \text{ and } Y = y) + b \sum_y \sum_x y P(X = x \text{ and } Y = y) \\
&\overset{(a)}{=} a \sum_x x P(X = x) + b \sum_y y P(Y = y) \\
&= aE[X] + bE[Y]
\end{aligned}$$

(a) Recall marginalising, $\sum_y P(X = x \text{ and } Y = y) = P(X = x)$

Linearity of Expected Value

More generally,

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

for random variables X_1, $X_2, \ldots,$ X_n. This is a v important property, and handy for exam questions.
Example:

- Suppose we toss 100 dice. We pay €1 to play each die and for each die that comes up 6 we win €5.
- X_i is the outcome for i'th die, $X_i = -1$ if lose or $X_i = 5 - 1$ if win. $E[X_i] = -\frac{1}{6}$.
- Expected overall winnings are $E[\sum_{i=1}^{100} X_i] = \sum_{i=1}^{100} E[X_i] = -\frac{100}{6}$

Linearity of Expected Value

A server has 32GB of memory. Suppose the memory usage of a job is 0.5GB with probability 0.5 and 1GB with probability 0.5, and that the memory usage of different jobs is independent.

- Suppose exactly 32 jobs are running. What is the expected memory usage ?
- Let X_i be memory usage of i'th job.
  $E[X_i] = 0.5 \times 0.5 + 1 \times 0.5 = 0.75$.
- So overall memory usage of all jobs is
  $E[\sum_{i=1}^{32} X_i] = \sum_{i=1}^{32} E[X_i] = 32 \times 0.75$.

What about when the number N of jobs is random ? Need to calculate $E[\sum_{i=1}^{N} X_i]$. Will come back to this later.

Expected Value of Independent Random Variables

- Take two **independent** random variables X and Y
- $E[XY] = E[X]E[Y]$
- Proof:

$$
\begin{aligned}
E[XY] &= \sum_x \sum_y xyP(X = x \text{ and } Y = y) \\
&= \sum_x \sum_y xyP(X = x)P(Y = y) \\
&= \sum_x xP(X = x) \sum_y yP(Y = y) \\
&= E[X]E[Y]
\end{aligned}
$$

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.

When Expected Value Isn't Enough

Game description[2]:

- We have a fair coin
- We keep flipping (perhaps infinitely many times) until we reach the first tails
- Random variable N = number of flips before first tails (so N is the number of consecutive heads)
- You win 2^N euros at the end

How much would you pay to play ?

- Random variable X = your winnings
- $E[X] = \left(\frac{1}{2}\right)^1 \times 2^0 + \left(\frac{1}{2}\right)^2 \times 2^1 + \left(\frac{1}{2}\right)^2 \times 2^2 + \cdots$
- $E[X] = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i = \sum_{i=0}^{\infty} \frac{1}{2} = \infty$
- Pay €100K each time and play 10 times. Any takers ?

[2] https://en.wikipedia.org/wiki/St._Petersburg_paradox

Demo

Gamblers Ruin[3]

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

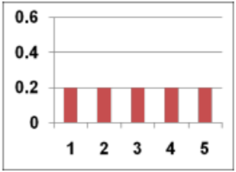
- Bet on red, $p = \frac{18}{37}$ to win €1 otherwise $1 - p$ you lose €1
- Bet €1
- If win then stop, if lose then double bet and repeat
- Random variable X is winnings on stopping

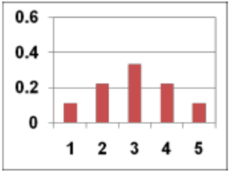
$$
\begin{aligned}
E[X] &= p \times 1 + (1-p)p \times (2-1) + (1-p)^2 p \times (4-2-1) + \cdots \\
&= \sum_{i=0}^{\infty} (1-p)^i p (2^i - \sum_{j=1}^{i-1} 2^j)
\end{aligned}
$$

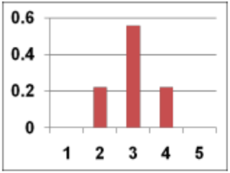
- Expected winnings are > 0 so why don't we play infinitely often ?
- You have finite money ! Usually also a max bet.

[3] https://en.wikipedia.org/wiki/Martingale_(betting_system)

Variance







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

Variance

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 $D = \{x_1, x_2, \cdots, x_n\}$.

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

  with $\mu = E[X] = \sum_{i=1}^{n} x_i p(x_i)$
- Example. Flip coin, $X = 1$ if heads, 0 otherwise.
  $E[X] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$.
  $Var(X) = (1-\frac{1}{2})^2 \times \frac{1}{2} + (0-\frac{1}{2})^2 \frac{1}{2} = \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2} = \frac{1}{4}$.
- Variance is mean squared distance of X from the mean μ
- $Var(X) \geq 0$
- Standard deviation is square root of variance $\sqrt{Var(X)}$.

Variance

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

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

Proof:

$$\begin{aligned}
Var(X) &= \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i) \\
&= \sum_{i=1}^{n} (x_i^2 - 2x_i\mu + \mu^2) p(x_i) \\
&= \sum_{i=1}^{n} x_i^2 p(x_i) - 2\sum_{i=1}^{n} x_i p(x_i)\mu + \mu^2 \sum_{i=1}^{n} p(x_i) \\
&= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}$$

Variance

Unlike expectation, variance is not linear. Instead we have $Var(aX + b) = a^2 Var(X)$. Observe that offset b does not affect the variance.

Proof:

$$
\begin{aligned}
Var(aX + b) &= E[(aX + b)^2] - E[aX + b]^2 \\
&= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
&= a^2E[X^2] + 2abE[X] + b^2 - a^2E[X]^2 - 2abE[X] - b^2 \\
&= a^2E[X^2] - a^2E[X]^2 \\
&= a^2(E[X^2] - E[X]^2) = a^2 Var(X)
\end{aligned}
$$

(recall $E[aX + b] = aE[X] + b$).

Variance of Independent Random Variables

For independent random variables X and Y then $Var(X+Y) = Var(X) + Var(Y)$.

Proof.

$$\begin{aligned}
Var(X+Y) &= E[(X+Y)^2] - E[X+Y]^2 \\
&= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2 \\
&= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2E[XY] - 2E[X]E[Y] \\
&= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 \\
&= Var(X) + Var(Y)
\end{aligned}$$

(recall $E[XY] = E[X]E[Y]$ when X and Y are independent)

Expected Value of Binomial Random Variable

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

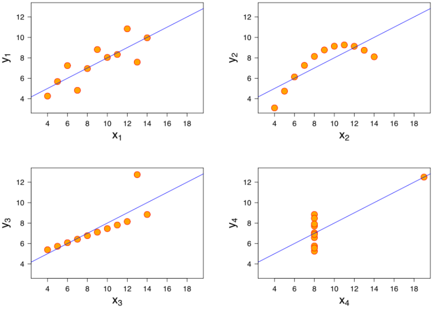
$$E[X] = p$$
$$Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$$

- 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 variance is another example of a summary value, this time one that indicates the spread in a data set. But great care is again needed.



source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet

All four datasets also have:

- $E[X] = 9$, $Var(X) = 11$
- $E[Y] \approx 7.50$, $Var(Y) \approx 4.12$
- Take home message: plot the data, don't just rely on summary values.