Overview

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

In 2011 Irish census\(^1\):

<table>
<thead>
<tr>
<th>No. of children</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>&gt; 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of families</td>
<td>344,944</td>
<td>339,596</td>
<td>285,952</td>
<td>144,470</td>
<td>75,248</td>
</tr>
<tr>
<td>No. with all children &lt; 15 years</td>
<td>-</td>
<td>178,012</td>
<td>92,826</td>
<td>30,010</td>
<td>8,327</td>
</tr>
</tbody>
</table>

Pick a family at random from the population. What is the expected number of children?

- Total no. of children
  
  \[= 0 \times 344,944 + 1 \times 339,596 + 2 \times 285,952 + 3 \times 144,470 + 4 \times 75,248 = 1,645,902\]

- Total no. of families:
  
  \[344,944 + 339,596 + 285,952 + 144,470 + 75,248 = 1,190,210\]

- Ratio = \(\frac{1645902}{1190210} = 1.38\)

- NB: No family has 1.38 children

---

\(^1\)http://www.cso.ie/en/census/census2011reports/census2011profile5householdsandfamilieslivingarrangementsinireland/
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?
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

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

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

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

$$P(\text{get a 6}) + P(\text{don't get a 6})$$
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

Useful in games of chance. Another 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.
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.

(see that this is also the average number of times we have to play our dice game before we first win)
Who cares?

Suppose we have to choose an action. Once we have chosen we get a “reward” with some probability. Which action should be choose?

- Often a reasonable answer is: the one with the best expected outcome
- Delayed gratification. Suppose I have to choose between two courses. One takes 1 year but with probability 0.5 will allow me to earn €5000 extra per year for next 5 years. The other is fast, takes 1 week, but with probability 0.05 earns an extra €5000 extra per year. Which should I choose?

- Option 1: cost is 1 year, say €10K earnings missed. Reward is €5000 for next 5 years with probability 0.5 and €0 with probability 0.5. Expected reward is $-10000 + 25000 \times 0.5 + 0 \times 0.5 = 2500$.
- Option 2: cost is 1 week, say €10K/52 earnings missed. Reward is €5000 over next 5 years with probability 0.05 and €0 with probability 0.95. Expected reward is $-10000/52 + 25000 \times 0.05 + 0 \times 0.95 \approx 1000$. 
Who cares?

- 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 $E[X] = 10^3/10^6 - 1 	imes (1 - 1/10^6) = -0.999$. 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 $0.99 \times 1000 - 0.01 \times 100 \times 10^9 = -9,999,010$. 

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,

\[
E[aX + b] = \sum_{i=1}^{n} (ax_i + b)P(X = x_i)
\]

\[
= \sum_{i=1}^{n} ax_i P(X = x_i) + \sum_{i=1}^{n} bP(X = x_i)
\]

\[
= a \sum_{i=1}^{n} x_i P(X = x_i) + b \sum_{i=1}^{n} P(X = x_i)
\]

\[
= aE[X] + b
\]
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:

$$
\left(\frac{5}{6}\right) \times (5.75 - 1.15) + \left(\frac{1}{6}\right) \times (0 - 1.15) = -\frac{1.15}{6}
$$

- 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:

$$E[aX + bY] = \sum_x \sum_y (ax + by)P(X = x \text{ and } Y = y)$$

$$= a \sum_x \sum_y xP(X = x \text{ and } Y = y) + b \sum_y \sum_x yP(X = x \text{ and } Y = y)$$

$$(a) \quad = a \sum_x xP(X = x) + b \sum_y yP(Y = y)$$

$$= aE[X] + bE[Y]$$

(a) Recall $\sum_y P(X = x \text{ and } Y = y) = P(X = x)$
Linearity of Expected Value

More generally,

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

for random variables \(X_1, X_2, \ldots, X_n\). This is a very 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 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:
  
  \[ E[X] = np \]
  \[ Var(X) = np(1 - p) \]
Expected Value of a Random Variable

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

\[
E[XY] = \sum_x \sum_y xyP(X = x \text{ and } Y = y)
\]

\[
= \sum_x \sum_y xyP(X = x)P(Y = y)
\]

\[
= \sum_x xP(X = x) \sum_y yP(Y = y)
\]

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

---

\(^2\text{https://en.wikipedia.org/wiki/St._Petersburg_paradox}\)
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

\[
E[Z] = p \times 1 + (1 - p)p \times (2 - 1) + (1 - p)^2 p \times (4 - 2 - 1) + \cdots
\]
\[
= \sum_{i=0}^{\infty} (1 - p)^i p(2^i - \sum_{j=1}^{i-1} 2^j) = 1
\]

- 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 statistic that quantifies “spread”
Variance

Let $X$ be a random variable with mean $\mu$. 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 $\mu$
- $Var(X) \geq 0$
- Standard deviation is square root of variance $\sqrt{Var(X)}$. 
Variance

Discrete random variable taking values in \( D = \{x_1, x_2, \cdots, x_n\} \). Variance:

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

\[
= \sum_{i=1}^{n} (x_i^2 - 2x_i \mu + \mu^2) p(x_i)
\]

\[
= \sum_{i=1}^{n} x_i^2 p(x_i) - 2 \sum_{i=1}^{n} x_i p(x_i) \mu + \mu^2 \sum_{i=1}^{n} p(x_i)
\]

\[
= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2
\]

\[
= E[X^2] - (E[X])^2
\]
Variance

\[\text{Var}(aX + b) = a^2 \text{Var}(X).\]

Proof:

\[
\text{Var}(aX + b) = E[(aX + b)^2] - E[aX + b]^2 \\
= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
= a^2 E[X^2] + 2abE[X] + b^2 - a^2 E[X]^2 - 2abE[X] - b^2 \\
= a^2 E[X^2] - a^2 E[X]^2 \\
= a^2(E[X^2] - E[X]^2) = a^2 \text{Var}(X)
\]

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

For independent random variables $X$ and $Y$ then

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Proof.

$\text{Var}(X + Y) = E[(X + Y)^2] - E[X + Y]^2$

$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$


$= \text{Var}(X) + \text{Var}(Y)$

(recall $E[XY] = E[X]E[Y]$ when $X$ and $Y$ are independent)
Anscombe’s Quartet

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

All four datasets also have:

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

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