Expected Value and Variance

Have you ever wondered whether it would be "worth it" to buy a lottery ticket every week, or pondered questions such as "If I were offered a choice between a million dollars, or a 1 in 100 chance of a billion dollars, which would I choose?"

One method of deciding on the answers to these questions is to calculate the **expected** earnings of the enterprise, and aim for the option with the higher expected value.

This is a useful decision making tool for problems that involve repeating many trials of an experiment — such as investing in stocks, choosing where to locate a business, or where to fish.

(For once-off decisions with high stakes, such as the choice between a sure 1 million dollars or a 1 in 100 chance of a billion dollars, it is unclear whether this is a useful tool.)

Example

John works as a tour guide in Dublin. If he has 200 people or more take his tours on a given week, he earns €1,000. If the number of tourists who take his tours is between 100 and 199, he earns €700. If the number is less than 100, he earns €500. Thus John has a **variable** weekly income.

From experience, John knows that he earns €1,000 fifty percent of the time, €700 thirty percent of the time and €500 twenty percent of the time. John's weekly income is a **random variable** with a probability distribution

Income	**Probability**
€1,000	0.5
€700	0.3
€500	0.2

Example

What is John's average income, over the course of a 50-week year?

Over 50 weeks, we expect that John will earn

- €1000 on about 25 of the weeks (50%);
- €700 on about 15 weeks (30%); and
- €500 on about 10 weeks (20%).

This suggests that his **average** weekly income will be

$$\frac{25(€1000) + 15(€700) + 10(€500)}{50} = €810.$$

Dividing through by 50, this calculation changes to

$$.5(€1000) + .3(€700) + .2(€500) = €810.$$

Expected Value of a Random Variable

The answer in the last example stays the same no matter how many weeks we average over. This suggests the following: If X is a random variable with possible values $x_1, x_2, \ldots, x_n$ and corresponding probabilities $p_1, p_2, \ldots, p_n$, the **expected value** of X, denoted by $\mathbf{E}(X)$, is

$$\boxed{\mathbf{E}(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n.}$$

Outcomes **X**	**Probability** $\mathbf{P}(X)$	**Out.** $\times$ **Prob.** $X\mathbf{P}(X)$
x_1	p_1	x_1p_1
x_2	p_2	x_2p_2
$\vdots$	$\vdots$	$\vdots$
x_n	p_n	x_np_n
		Sum $= \mathbf{E}(X)$

Expected Value of a Random Variable

We can interpret the expected value as the long term average of the outcomes of the experiment over a large number of trials. From the table, we see that the calculation of the expected value is the same as that for the average of a set of data, with relative frequencies replaced by probabilities.

Warning: The expected value really ought to be called the expected mean. It is NOT the **value** you most expect to see but rather the average (or mean) of the values you see over the course of many trials.

Coin tossing example

Flip a coin 4 times and observe the sequence of heads and tails. Let X be the number of heads in the observed sequence. Last time we found the following probability distribution for X:

X	**P**(X)
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16

Find the expected number of heads for a trial of this experiment, that is find $\mathbf{E}(X)$.

$$\mathbf{E}(X) = \frac{1}{16} \cdot 0 + \frac{4}{16} \cdot 1 + \frac{6}{16} \cdot 2 + \frac{4}{16} \cdot 3 + \frac{1}{16} \cdot 4 = \frac{0+4+12+12+4}{16} = \frac{32}{16} = 2.$$

NFL example

The following probability distribution from "American Football" *Statistics in Sports, 1998*, by Hal Stern, has an approximation of the probabilities for yards gained on a running play in the NFL. Actual play by play data was used to estimate the probabilities. (-4 represents 4 yards lost on a running play).

x, yards	prob	x, yards	prob
-4	.020	6	.090
-2	.060	8	.060
-1	.070	10	.050
0	.150	15	.085
1	.130	30	.010
2	.110	50	.004
3	.090	99	.001
4	.070		

NFL example

Based on this data, what is the expected number of yards gained on a running play in the NFL?

$\mathbf{E}(X) = (-4) \cdot .020 + (-2) \cdot 0.060 + (-1) \cdot 0.070 + 0 \cdot 0.150 + 1 \cdot 0.130 + 2 \cdot 0.110 + 3 \cdot 0.090 + 4 \cdot 0.070 + 6 \cdot 0.090 + 8 \cdot 0.060 + 10 \cdot 0.050 + 15 \cdot 0.085 + 30 \cdot 0.010 + 50 \cdot 0.004 + 99 \cdot 0.001 = 4.024$

Roulette example

In roulette, when you bet \$1 on red, the probability distribution for your earnings, denoted by X, is given by:

X	**P**(X)
1	18/38
-1	20/38

(a) What are your expected earnings for this bet?

$\mathbf{E}(X) = 1 \cdot (18/38) + (-1) \cdot (20/38) = -2/38.$

(b) How much would you expect to win/lose if you bet \$1 on red 100 times? What would the casino expect to earn if you bet \$1 on red 100 times?

You would expect to win $100 \cdot \mathbf{E}(X) = -200/38 \approx -\$5.26.$

Your loss is the casino's gain so the casino's earnings are the negative of your loss: \$5.26.

A winning(?) strategy for Roulette?

Roulette seems like a fool's game. But here's a possible strategy for playing it:

1. Begin by betting a dollar on red.
2. If you win, take your winnings and go home.
3. If you lose, place two one-dollar bets in a row on red.
4. Whatever happens on those two rolls, go home (either with your winnings to date, or cutting your losses)

Question: Is this a winning strategy? Specifically, what is the probability that you will leave the Roulette wheel with more money than you began with, and is this probability more or less than 1/2?

A winning(?) strategy for Roulette?

Let X be net winnings from this strategy. Possible outcomes/values for X:

- Win on first roll, probability $18/38 \approx .474$, $X = +1$
- Lose on first, win on next two, probability $(20/38)(18/38)^2 \approx .118$, $X = +1$
- Lose on first, win exactly one of next two, probability $(20/38)2(18/38)(20/38) \approx .262$, $X = -1$
- Lose all three, probability $(20/38)^3 \approx .146$, $X = -3$.

So X takes value $+1$ with probability $\approx .592$, value -1 with probability $\approx .262$, and value -3 with probability $\approx .146$. So the strategy *is* winning — you have a get gain more often than a net loss!

On the other hand, $\mathbf{E}(X) = 1(.592) - 1(.262) - 3(.146) = -.108$. So on *average*, playing this strategy long-term, you will lose money :(

Gambling example

The rules of a carnival game are as follows:

1. The player pays \$1 to play the game.
2. The player then flips a fair coin, if the player gets a head the game attendant gives the player \$2 and the player stops playing.
3. If the player gets a tail on the coin, the player rolls a fair six-sided die. If the player gets a six, the game attendant gives the player \$1 and the game is over.
4. If the player does not get a six on the die, the game is over and the game attendant gives nothing to the player.

Gambling example

Let X denote the player's (net) earnings for this game.
Last time, we saw that X has probability distribution

X	**P**(X)
-1	5/12
0	1/12
1	1/2

(a) What are the expected earnings for the player for each play of this game?

$$\mathbf{E}(X) = (-1)\cdot\frac{5}{12} + 0\cdot\frac{1}{12} + 1\cdot\frac{1}{2} = \frac{-5+0+6}{12} = \frac{1}{12} \approx \$0.08.$$

(b) What are the expected earnings for the game host for each play of this game?

Host's earnings are negative of your earnings:
$-1/12 \approx -\$0.08$.

Variance and standard deviation

Let us return to the initial example of John's weekly income which was a random variable with probability distribution

Income	**Probability**
€1,000	0.5
€700	0.3
€500	0.2

with mean €810. Over 50 weeks, we might expect the **variance** of John's weekly earnings to be roughly

$$\frac{25(\text{€}1000\text{-}\text{€}810)^2 + 15(\text{€}700\text{-}\text{€}810)^2 + 10(\text{€}500\text{-}\text{€}810)^2}{50} = 49,900$$

or

$$.5(\text{€}1000\text{-}\text{€}810)^2 + .3(\text{€}700\text{-}\text{€}810)^2 + .2(\text{€}500\text{-}\text{€}810)^2 = 49,900$$

Variance and standard deviation

As with the calculations for the expected value, if we had chosen any large number of weeks in our estimate, the estimates would have been the same. This suggests a formula for the **variance** of a random variable.

If X is a random variable with values $x_1, x_2, \ldots, x_n$, corresponding probabilities $p_1, p_2, \ldots, p_n$, and expected value $\mu = \mathbf{E}(X)$, then

$$\boxed{\textbf{Variance} = \boldsymbol{\sigma}^2(X) = p_1(x_1 - \mu)^2 + p_2(x_2 - \mu)^2 + \cdots + p_n(x_n - \mu)^2}$$

and

$$\boxed{\textbf{Standard Deviation} = \boldsymbol{\sigma}(X) = \sqrt{\text{Variance}}}.$$

Variance and standard deviation

$$\boxed{\textbf{Variance} = \boldsymbol{\sigma}^2(X) = p_1(x_1 - \mu)^2 + p_2(x_2 - \mu)^2 + \cdots + p_n(x_n - \mu)^2}$$

$$\boxed{\textbf{Standard Deviation} = \boldsymbol{\sigma}(X) = \sqrt{\text{Variance}}}.$$

x_i	p_i	$x_i p_i$	$(x_i - \mu)$	$(x_i - \mu)^2$	$p_i(x_i - \mu)^2$
x_1	p_1	$x_1 p_1$	$(x_1 - \mu)$	$(x_1 - \mu)^2$	$p_1(x_1 - \mu)^2$
x_2	p_2	$x_2 p_2$	$(x_2 - \mu)$	$(x_2 - \mu)^2$	$p_2(x_2 - \mu)^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	p_n	$x_n p_n$	$(x_n - \mu)$	$(x_n - \mu)^2$	$p_n(x_n - \mu)^2$
		Sum $= \mu$			**Sum** $= \boldsymbol{\sigma}^2(X)$

Gambling example

The rules of a carnival game are as follows:

1. The player pays \$1 to play the game.
2. The player then flips a fair coin, if the player gets a head the game attendant gives the player \$2 and the player stops playing.
3. If the player gets a tail on the coin, the player rolls a fair six-sided die. If the player gets a six, the game attendant gives the player \$1 and the game is over.
4. If the player does not get a six on the die, the game is over and the game attendant gives nothing to the player.

Gambling example

Let X denote the player's (net) earnings for this game. Last time we saw that the probability distribution of X is given by:

X	$\mathbf{P}$(X)
-1	5/12
0	1/12
1	1/2

Use the value for $\mu = \mathbf{E}(X)$ found above to find the variance and standard deviation of X, that is find $\boldsymbol{\sigma}^2(X)$ and $\boldsymbol{\sigma}(X)$.

Gambling example

$\mathbf{x}_i$	$\mathbf{p}_i$	$\mathbf{x}_i \cdot \mathbf{p}_i$	$(\mathbf{x}_i - \mu)$	$(\mathbf{x}_i - \mu)^2$	$\mathbf{p}_i \cdot (\mathbf{x}_i - \mu)^2$
-1	5/12	$\frac{-5}{12}$	$\frac{-13}{12}$	$\frac{169}{144}$	$\frac{845}{1728}$
0	1/12	$\frac{0}{12}$	$\frac{-1}{12}$	$\frac{1}{144}$	$\frac{1}{1728}$
1	6/12	$\frac{6}{12}$	$\frac{11}{12}$	$\frac{121}{144}$	$\frac{726}{1728}$
		Sum $= \mu = \frac{1}{12}$			**Sum** $= \boldsymbol{\sigma}^2(X) = \frac{1572}{1728} \approx 0.909$

$\boldsymbol{\sigma} \approx 0.953.$

Coin tossing example

An experiment consists of flipping a coin 4 times and observing the sequence of heads and tails. The random variable X is the number of heads in the observed sequence. Last time we found the following probability distribution for X:

X	**P**(X)
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16

We saw above that the expected value for this random variable is $\mathbf{E}(X) = 2$. Find $\boldsymbol{\sigma}^2(X)$ and $\boldsymbol{\sigma}(X)$.

Coin tossing example

x_i	p_i	$x_i \cdot p_i$	$(x_i - \mu)$	$(x_i - \mu)^2$	$p_i \cdot (x_i - \mu)^2$
0	$\frac{1}{16}$	$\frac{0}{16}$	-2	4	$\frac{4}{16}$
1	$\frac{4}{16}$	$\frac{4}{16}$	-1	1	$\frac{4}{16}$
2	$\frac{6}{16}$	$\frac{12}{16}$	0	0	$\frac{0}{16}$
3	$\frac{4}{16}$	$\frac{12}{16}$	1	1	$\frac{4}{16}$
4	$\frac{1}{16}$	$\frac{4}{16}$	2	4	$\frac{4}{16}$
		Sum $= \mu = 2$			**Sum** $= \boldsymbol{\sigma}^2(X) = 1$

$\boldsymbol{\sigma} = 1$.

Another formula for variance

Using $(x-\mu)^2 = x^2 - 2\mu x + \mu^2$, we get another formula for variance:

$$
\begin{aligned}
\boldsymbol{\sigma}^2(X) &= p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \cdots + p_n(x_n-\mu)^2 \\
&= p_1(x_1^2 - 2\mu x_1 + \mu^2)^2 + \cdots + p_n(x_n^2 - 2\mu x_n + \mu^2)^2 \\
&= [p_1 x_1^2 + \cdots + p_n x_n^2] + \\
&\quad -2\mu[p_1 x_1 + \cdots p_n x_n] + \\
&\quad \mu^2[p_1 + \cdots p_n] \\
&= \mathbf{E}(X^2) - 2\mu\mathbf{E}(X) + \mu^2 \\
&= \mathbf{E}(X^2) - 2\mathbf{E}(X)\mathbf{E}(X) + \mathbf{E}(X)^2 \\
&= \mathbf{E}(X^2) - \mathbf{E}(X)^2.
\end{aligned}
$$

$$\boxed{\boldsymbol{\sigma}^2(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2}.$$

Redoing the coin example

Using the definition of variance:

x_i	p_i	$x_i \cdot p_i$	$(x_i - \mu)$	$(x_i - \mu)^2$	$p_i \cdot (x_i - \mu)^2$
0	$\frac{1}{16}$	$\frac{0}{16}$	-2	4	$\frac{4}{16}$
1	$\frac{4}{16}$	$\frac{4}{16}$	-1	1	$\frac{4}{16}$
2	$\frac{6}{16}$	$\frac{12}{16}$	0	0	$\frac{0}{16}$
3	$\frac{4}{16}$	$\frac{12}{16}$	1	1	$\frac{4}{16}$
4	$\frac{1}{16}$	$\frac{4}{16}$	2	4	$\frac{4}{16}$
		Sum $= \mu = 2$			**Sum** $= \boldsymbol{\sigma}^2(X) = 1$

$\boldsymbol{\sigma} = 1.$

Redoing the coin example

Using the new formula:

x_i	p_i	$p_i x_i$	x_i^2	$p_i x_i^2$
0	$\frac{1}{16}$	$\frac{0}{16}$	0	$\frac{0}{16} = 0$
1	$\frac{4}{16}$	$\frac{4}{16}$	1	$\frac{4 \cdot 1}{16} = \frac{4}{16}$
2	$\frac{6}{16}$	$\frac{12}{16}$	4	$\frac{6 \cdot 4}{16} = \frac{24}{16}$
3	$\frac{4}{16}$	$\frac{12}{16}$	9	$\frac{4 \cdot 9}{16} = \frac{36}{16}$
4	$\frac{1}{16}$	$\frac{4}{16}$	16	$\frac{1 \cdot 16}{16} = \frac{16}{16}$
		$\textbf{Sum} = \mathbf{E}(X) = \mu = 2$		$\textbf{Sum} = \mathbf{E}(X^2) = \frac{80}{16} = 5$

Hence

$$\boldsymbol{\sigma}^2(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = 5 - 2^2 = 1$$

Back to the Roulette strategy

Recall that we had a Roulette strategy where your winnings X had probability distribution

$$P(X = 1) = .592, \quad P(X = -1) = .262, \quad P(X = -3) = .146.$$

We calculated $\mathbf{E}(X) = 1(.592) - 1(.262) - 3(.146) \approx -.108$

We can easily calculate
$\mathbf{E}(X^2) = 1^2(.592) + (-1)^2(.262^2) + (-3)^2(.146^2) \approx 2.16$

From this we get $\sigma^2(X) = .216 - (-.108)^2 \approx 2.15$, so the standard deviation of your winnings is roughly $\sqrt{2.15} \approx 1.46$.

The strategy "bet once on Red" has expected winnings $-.053$, with standard deviation .998 — our complicated strategy is a little bit worse on average, and much more unstable