Section 8.5: Expected Value and Variance of a random variable

Have you ever wondered whether it would be "worth it" to buy a lottery ticket every week, or pondered on questions such as "If I were offered a choice between a million dollars or a 1 in 100 chance of getting 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 a higher expected value. This is most certainly a useful decision making tool when we are contemplating a strategy which involves repeating several trials of an experiment such as investing in stocks or choosing where to locate our business or where to fish. For once off decisions where the stakes are high, 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 1 John works as a tour guide in Dublin for the company Excellent Tours Ltd.. Excellent Tours has a website where tourists sign up for the tours. For any given week, if John has 200 people or more take his tours he earns €1,000. If the number of tourists who take John's tours is between 100 and 199, John earns €700 and if the number of tourists taking his tours is less than 100, John earns €500 for the week. Thus **John has a variable weekly income**. Because he has kept records over the past few years, 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. There is no discernible pattern to the variability, so John's weekly income is a random variable with a probability distribution:

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

John has a lot of fixed weekly costs, such as rent, a gas bill and an electricity bill. John's fixed costs are about to increase because a new weekly charge for water has been introduced along with a significant increase in the cost of public transport which John uses to get to work. John normally saves some of the money from good weeks to cover costs in lean weeks when his income is lower than his fixed costs. However, in order to be able to cover fixed costs (and buy food) in the long run, John's **average income must be greater than his fixed costs**.

To calculate the average income, one might consider what will happen over the next fifty weeks. For roughly half of these weeks (25 weeks), John's income will be €1,000, for roughly $(0.3 \times 50 =)$15 weeks, John's income will be €700 and for roughly $(0.2 \times 50 =)$10 weeks, John's income will be €500. Thus the average over the next fifty weeks will be roughly:

$$\frac{(25 \times €1{,}000) + (15 \times €700) + (10 \times €500)}{50}$$

$$= \frac{(50 \times 0.5 \times €1{,}000) + (50 \times 0.3 \times €700) + (50 \times 0.2 \times €500)}{50}$$

$$= \frac{\cancel{50}\big[(0.5 \times €1{,}000) + (0.3 \times €700) + (0.2 \times €500)\big]}{\cancel{50}}$$

$$= (0.5 \times €1{,}000) + (0.3 \times €700) + (0.2 \times €500) = €810.$$

We can see from the calculation above, that we would have gotten the same answer if we had use 100 weeks or any other (large) number of weeks. The number €810 is called the **expected value** of John's income and we would expect John's income to average to this amount in the long run (over the course of many weeks).

Expected Value of a Random Variable We can pull out the general principles of the above calculation to get the expected value of any random variable. 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 $E(X)$, is

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

Outcomes **X**	**Probability** **P(X)**	**Out. × Prob.** **XP(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 $= E(X)$

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.

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

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

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

Example Successful NFL running plays 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		

Find the expected number of yards gained on a running play in the NFL.

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

Example We saw last time that in a game of American roulette where you bet \$1 on red, the probability distribution for your earnings (X) is given by:

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

(a) What are your expected earnings for this bet? (What is $E(X)$?)

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

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

$$100 \cdot E(X) = -\frac{200}{18} \approx -\$11.11.$$

(c) What would the casino expect to earn if you bet \$1 on red 100 times?

\$11.11.

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.

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	**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?

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

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

$$-\frac{1}{12} \approx -\$0.08.$$

(c) How much would you expect the game host to win/lose if 100 people play this game?

$$-100 \cdot \frac{1}{12} \approx -\$8.00.$$

Variance and standard deviation of a random variable 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

To find the variance (average squared distance from the mean, $\mu =$ €810) one might again estimate that over the next 50 weeks, the (population) **variance** would be roughly

$$\frac{[25 \times (€1{,}000 - €810)^2] + [15 \times (€700 - €810)^2] + [10 \times (€500 - €810)^2]}{50}$$

$$= \frac{[50 \times 0.5 \times (€1{,}000 - €810)^2] + [50 \times 0.3 \times (€700 - €810)^2] + [50 \times (0.2) \times (€500 - €810)^2]}{50}$$

$$= \frac{\cancel{50}\Big([0.5 \times (€1{,}000 - €810)^2] + [0.3 \times (€700 - €810)^2] + [(0.2) \times (€500 - €810)^2]\Big)}{\cancel{50}}$$

$$= 0.5 \times (€1{,}000 - €810)^2 + 0.3 \times (€700 - €810)^2 + (0.2) \times (€500 - €810)^2$$

$$= 0.5 \times (190)^2 + 0.3 \times (-110)^2 + 0.2 \times (-310)^2 = 40,900$$

Recall that the **standard deviation** is the square root of the variance, so a good estimate for the standard deviation of John's income is given by €202.24.

As with the calculations for the expected value, we notice that if we had chosen any large number of weeks in our estimate, our estimates for the variance and standard deviation would have been the same as that calculated above. We can pull out the general principles to get a formula for the variance and standard deviation for and 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 $\mu = E(X)$, then

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

and

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

$\mathbf{x_i}$	$\mathbf{p_i}$	$\mathbf{x_i p_i}$	$(\mathbf{x_i} - \mu)$	$(\mathbf{x_i} - \mu)^\mathbf{2}$	$\mathbf{p_i}(\mathbf{x_i} - \mu)^\mathbf{2}$
x_1	p_1	x_1p_1	$(x_1-\mu)$	$(x_1-\mu)^2$	$p_1(x_1-\mu)^2$
x_2	p_2	x_2p_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_np_n	$(x_n-\mu)$	$(x_n-\mu)^2$	$p_n(x_n-\mu)^2$
		Sum $= \mu$			**Sum** $= \sigma^2(X)$

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.

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

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

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

$\mathbf{x}_i$	$\mathbf{p}_i$	$\mathbf{x}_i \cdot \mathbf{p}_i$	$(\mathbf{x}_i - \mu)$	$(\mathbf{x}_i - \mu)^2$	$\mathbf{p}_i \cdot (\mathbf{x}_i - \mu)^2$
-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** $= \sigma^2(X) = \frac{1572}{1728} \approx 0.9097222222$

$\sigma \approx 0.9537935952.$

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 day 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 $E(X) = 2$. Find $\sigma^2(X)$ and $\sigma(X)$.

$\mathbf{x}_i$	$\mathbf{p}_i$	$\mathbf{x}_i \cdot \mathbf{p}_i$	$(\mathbf{x}_i - \mu)$	$(\mathbf{x}_i - \mu)^2$	$\mathbf{p}_i \cdot (\mathbf{x}_i - \mu)^2$
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** $= \sigma^2(X) = 1$

$\sigma = 1$.