Steve Smith Tuition:
Maths Notes

$e^{i\pi} + 1 = 0$ $a^2 + b^2 = c^2$ $z_{n+1} = z_n^2 + c$ $V - E + F = 2$

Discrete Random Variables

Contents

Prerequisites

None.

Notes

None.

Document History

Date	Version	Comments
22th November 2012	1.0	Initial creation of the document.
26th November 2013	1.1	A few minor tune-ups.

1 Intro

What are Discrete Random Variables? Well, Discrete Random Variables (DRVs) are:

- Variables : that is, they represent numbers that can take different values.
- Discrete : in Maths, the word "Discrete" means something different to its meaning in everyday life. In Maths, Discrete means that only certain values of the variable are allowed. The opposite of Discrete is "Continuous" which means that any value is allowed. Examples of discrete variables are (1) the value you get when you roll a dice and (2) the card you get when you pick one from a pack. Examples of continuous variables are (1) the height of someone in your class at school, and (2) the length of a piece of string.
- Random : this means that you can never predict what the outcome of the variable will be.

2 The Distribution of Probabilities for a Discrete Random Variable

Because DRVs are random, so that you never know what value of the variable you will get, then the best we can do is to work out probablilities of results. Usually, probabilities of DRVs are shown in a table, like this

Result	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

This table shows the probabilities of getting each possible result when you roll a fair die. In these tables, the top row shows the values that can result when you carry out your "experiment" (in this case, rolling the die). The bottom row shows the probability of getting each of those values. When you roll a fair die, each of the 6 possible results can come up, and they are all equally likely to occur. So the probability of each is $\frac{1}{6}$.

Notice that if you add up all the probabilities you get 1. This makes sense: when you roll a die, you must get one of the values, and you can't get two different values with one roll (so there is no overlap of results). This is a very important fact about DRVs: the sum of the probabilities has to be 1.

Important Point!!

The sum of the probabilities for a DRV is 1

Normally, the above table is written like this:

$\boldsymbol{x}$	1	2	3	4	5	6
$\boldsymbol{p(X = x)}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

And this is because there is a bit of notation that you need to understand. The DRV itself is usually denoted by an upper-case letter, X in this case. So X represents the action of rolling this die. Lowercase letters (x in this case, if the DRV is X) denote the values that X can take. In this example, then, our DRV is X and the values that it can take (x) are 1, 2, 3, 4, 5, or 6. The expression $p(X = x)$ means “the probability that the DRV X can take the value x”. So in our case, for example, $p(X = 2) = \frac{1}{6}$, and $p(X > 4) = \frac{2}{6}$. Can you see that? That’s because $p(X > 4)$ would mean the probability of getting a value *greater than* 4. There are only two possibilities of this: 5 or 6, each having a probability of $\frac{1}{6}$ of coming up. Hence the probability $p(X > 4)$ will be $\frac{1}{6} + \frac{1}{6} = \frac{2}{6}$.

These tables are called *probability distribution* tables, because they show how the total probability (which is 1, of course) is distributed between the various possible results.

2.1 Examples

2.1.1 Example 1

The random variable X is given by the sum of the scores when two ordinary dice are thrown.

(i) Find the probability distribution of X.

(ii) Illustrate the distribution and describe the shape of the distribution.

(iii) Find the values of:

(a) $\mathrm{p}(X > 8)$
(b) $\mathrm{p}(X\ \textit{is even})$
(c) $\mathrm{p}(|X - 7| < 3)$

2.1.2 Answer 1

(i) The easiest way to visualise the outcomes from this random variable is to draw a table of all the possible results of throwing these two dice. Here's my attempt:

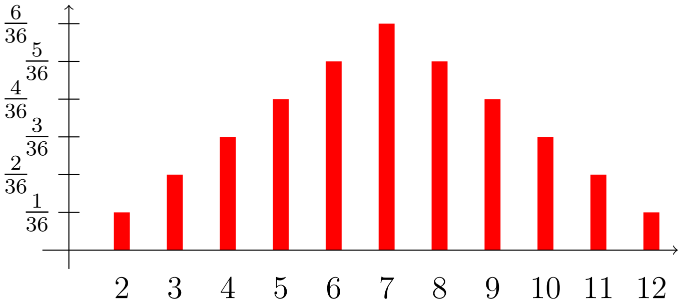
		Second die roll					
		1	**2**	**3**	**4**	**5**	**6**
First die roll	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Each entry in the table is just the values on the two dice added together. Now in order to produce a probability distribution, we have to figure out all the possible values of our random variable, and then the find the probablilities of each happening when you roll the two dice.

Looking at our table, the set of possible results we could get are 2, 3, 4, ..., 11, 12.

$\boldsymbol{x}$	2	3	4	5	6	7	8	9	10	11	12
$\boldsymbol{p(X = x)}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(ii) Here's a picture of the distribution:



The distribution is symmetric. It reaches a peak in the middle ($p(X = 7) = \frac{6}{36}$), and it goes down in the same way on either side of $x = 7$.

(iii) (a) $p(X > 8)$ will be the probability of getting the sum on the dice to be greater than 8: i.e. 9, 10, 11 or 12. If you add up the probabilities of getting those values, you get $\frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36} = \frac{5}{18}$.

(iii) (b) $p(X \textit{ is even})$ will be the probability of getting the sum on the dice to be 2, 4, 6, 8 10 or 12. If you add up the probabilities of getting those values, you get $\frac{1}{36} + \frac{3}{36} + \frac{5}{36} + \frac{5}{36} + \frac{3}{36} + \frac{1}{36} = \frac{18}{36} = \frac{1}{2}$.

(iii) (c) To get the probability of $p(|X - 7| < 3)$ we need to first understand what $|X - 7|$ means. The modulus signs (the vertical lines) around something say something very simple: "make me positive". So, for example, $|3|$ is just 3, but $|-3|$ is 3 as well. I think we need another little table:

$\boldsymbol{x}$	2	3	4	5	6	7	8	9	10	11	12
$\boldsymbol{x-7}$	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\boldsymbol{\lvert x-7\rvert}$	5	4	3	2	1	0	1	2	3	4	5

This table lists all the values x can take (on the top row), then for those values of x what the corresponding value of $x - 7$ would be (the middle row), and finally the values of $|x - 7|$ for each x (in the third row).

So the only values $|x - 7|$ can take are 0, 1, 2, 3, 4 and 5. And we want this value to be less than 3. So the probability of getting a result that corresponds to a value of $|x - 7| < 3$ would be $p(X = 5) + p(X = 6) + p(X = 7) + p(X = 8) + p(X = 9) = \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} = \frac{24}{36} = \frac{2}{3}$.

2.1.3 Example 2

The random variable Y is given by the absolute difference between the scores when two ordinary dice are thrown.

(i) Find the probability distribution of Y.

(ii) Illustrate the distribution and describe the shape of the distribution.

(iii) Find the values of:

(a) $p(Y < 3)$

(b) $p(Y \textit{ is odd})$

2.1.4 Answer 2

(i) The easiest way to visualise this random variable is to draw a table of all the possible results of throwing these dice. The phrase "the absolute difference" means subtract one value from the other, then make it positive. Here's my table:

Second die roll

First die roll

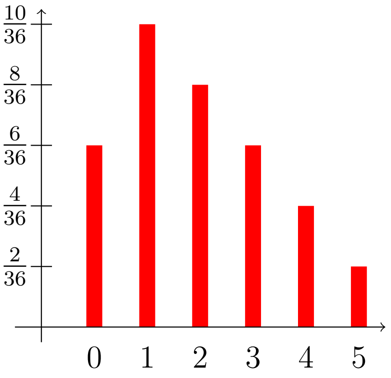
	1	**2**	**3**	**4**	**5**	**6**
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Each entry in the table is just the value on one of the two dice subtracted from the other, then made positive. Now in order to produce a probability distribution, we have to figure out all the possible values of our random variable, and then the find the probablilities of each happening when you roll the two dice.

Looking at our table, the set of possible results we could get are 0, 1, 2, 3, 4, and 5.

$\boldsymbol{y}$	0	1	2	3	4	5
$\boldsymbol{p(Y = y)}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

(ii) Here's a picture of the distribution:



The distribution is said to have "positive skew". That means that the peak is to the left of the middle.

(iii) (a) $p(Y < 3)$ will be the probability of getting the difference between the dice to be less than 3: i.e. 0, 1, or 2. If you add up the probabilities of getting those values, you get $\frac{6}{36} + \frac{10}{36} + \frac{8}{36} = \frac{24}{36} = \frac{2}{3}$.

(iii) (b) $p(Y\ is\ odd)$ will be the probability of getting the difference between the dice to be 1, 3, or 5. If you add up the probabilities of getting those values, you get $\frac{10}{36} + \frac{6}{36} + \frac{2}{36} = \frac{18}{36} = \frac{1}{2}$.

2.1.5 Example 3

The probability distribution for a discrete random variable X is given by:

$$\begin{aligned} p(X=x) &= \frac{kx}{8} && for\ x = 2,4,6,8 \\ p(X=x) &= 0 && otherwise \end{aligned}$$

(i) Find the value of k and tabulate the probability distribution.

(ii) If two successive values of X are generated independently, find the probability that:

(a) the two values are equal;

(b) the first value is greater than the second value.

2.1.6 Answer 3

(i) In order to find the value of k I think the easiest thing to do would be to tabulate the probability distribution, even though we don't know what k is yet. To do that, we have to put our probabilities in terms of k, but that's allowed! Here's my table:

$\boldsymbol{x}$	2	4	6	8
$\boldsymbol{p(X=x)}$	$\frac{2k}{8}$	$\frac{4k}{8}$	$\frac{6k}{8}$	$\frac{8k}{8}$

I calculated these probabilities from the probability distribution: the entry for $X = 4$, for example is $\frac{4k}{8}$ because in the formula, when $x = 4$, then $p(X = x) = \frac{k\times 4}{8}$.

Now, because this is a probability distribution, then the sum of the probabilities must be 1. So:

$$\begin{aligned} 1 &= \frac{2k}{8} + \frac{4k}{8} + \frac{6k}{8} + \frac{8k}{8} \\ &= \frac{2k+4k+6k+8k}{8} \\ &= \frac{20k}{8} \end{aligned}$$

In which case,

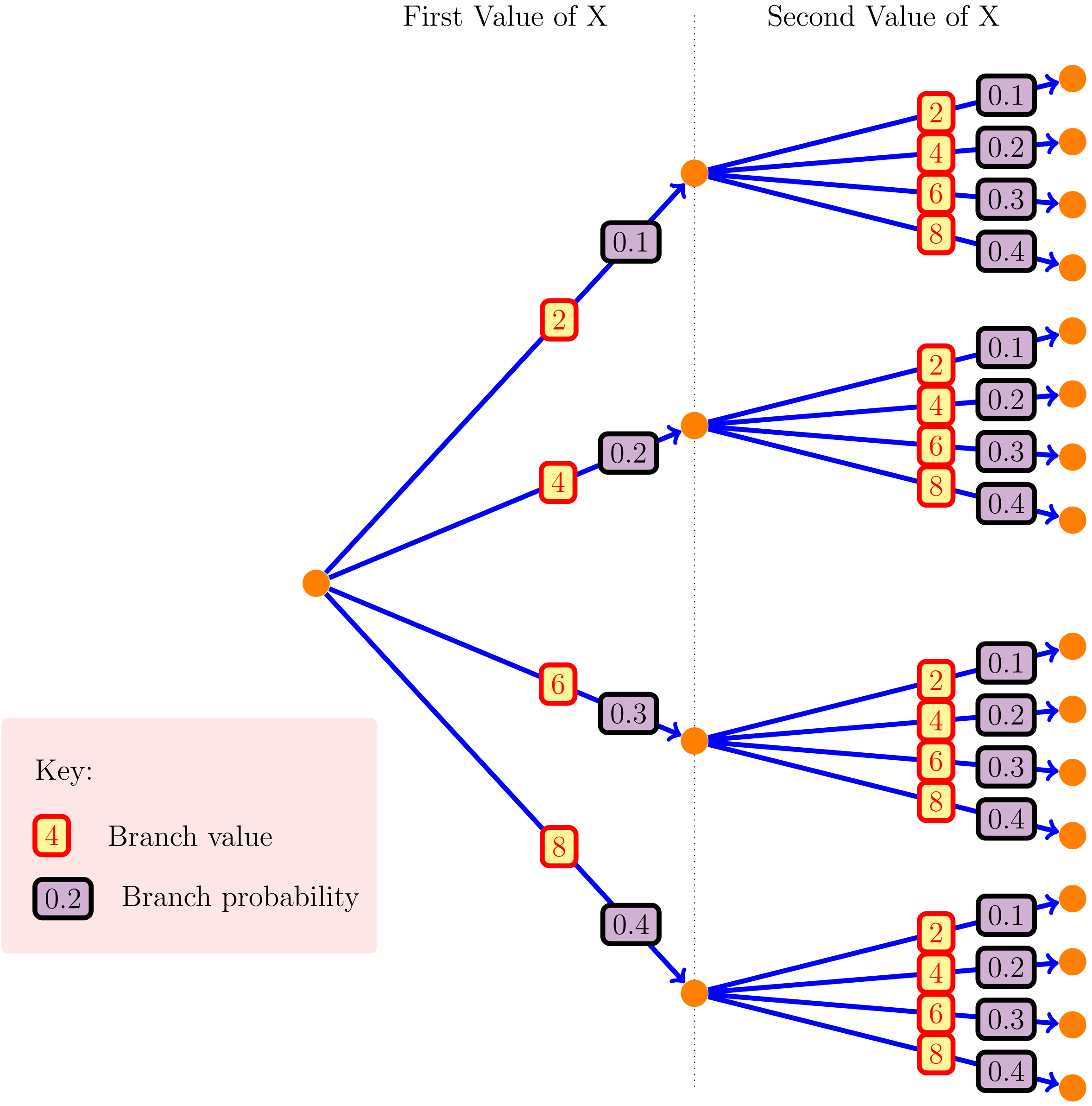
$$\begin{aligned} 20k &= 8 \\ \Longrightarrow \qquad k &= \frac{8}{20} = \frac{2}{5} = 0.4 \end{aligned}$$

Now, if we write out our probability distribution table again, this time incorporating the value of k that we have just discovered, then it turns out to be:

$\boldsymbol{x}$	2	4	6	8
$\boldsymbol{p(X=x)}$	0.1	0.2	0.3	0.4

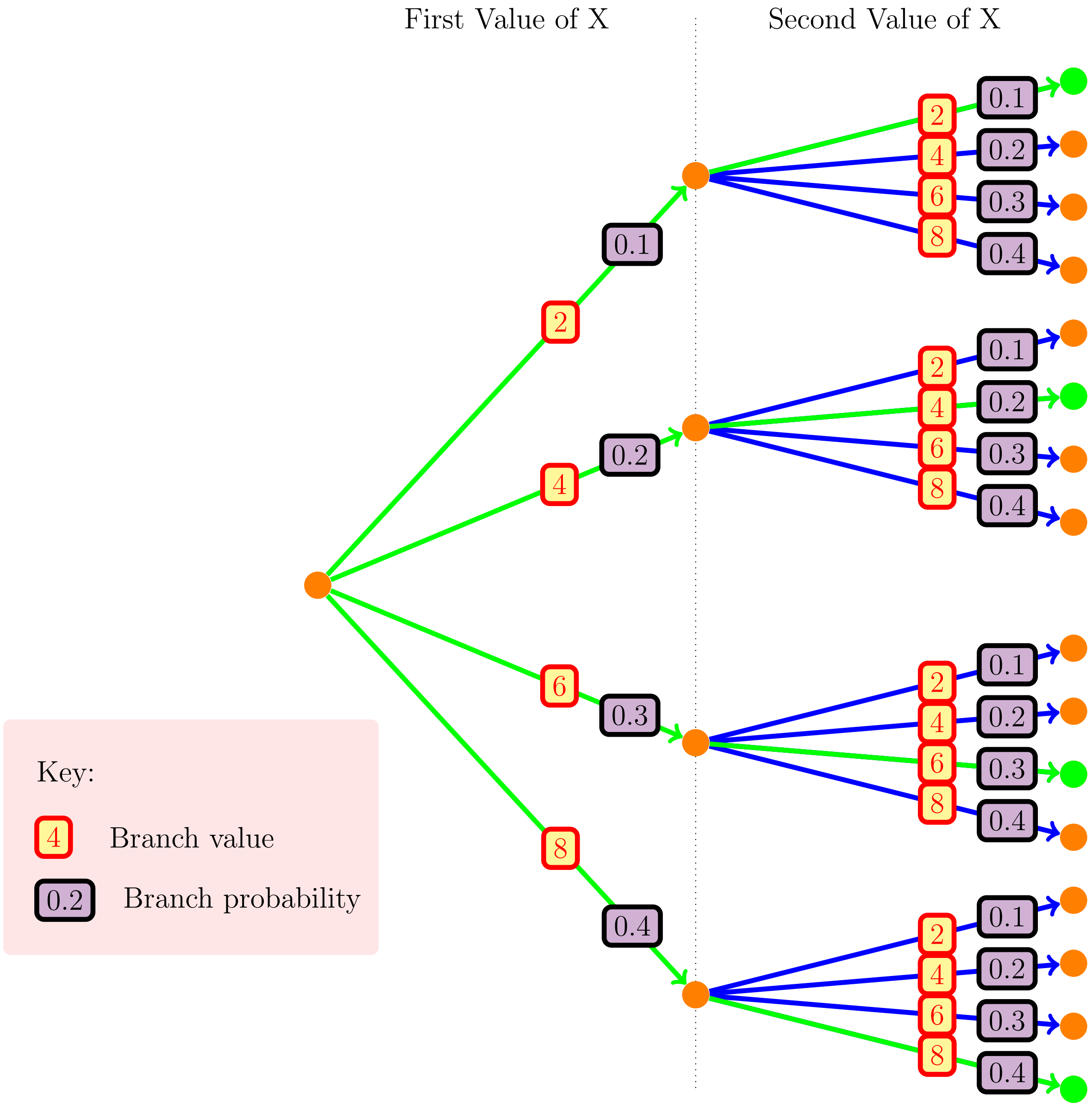
because, for example, when $k = 0.4$, then $p(X = 6) = \frac{6k}{8} = \frac{6\times 0.4}{8} = \frac{2.4}{8} = 0.3$.

(ii) When you have successive events occurring and you want to find the probabilities of combinations of events occurring, then the best thing to do is to draw a tree diagram:



Now if we want the two values of X to be equal, then the tree diagram we need to draw emphasises those outcomes where the first value of X is equal to the second value of X. In the following diagram, those branches and outcomes are marked in green:

Now we need to find the probability of any of those green outcomes occurring. Remember how to do this? We simply multiply the probabilities of the branches to get the probability of the final outcome, so:



p(the first value of X is 2 AND the second value of X is 2) is: $0.1 \times 0.1 = 0.01$

p(the first value of X is 4 AND the second value of X is 4) is: $0.2 \times 0.2 = 0.04$

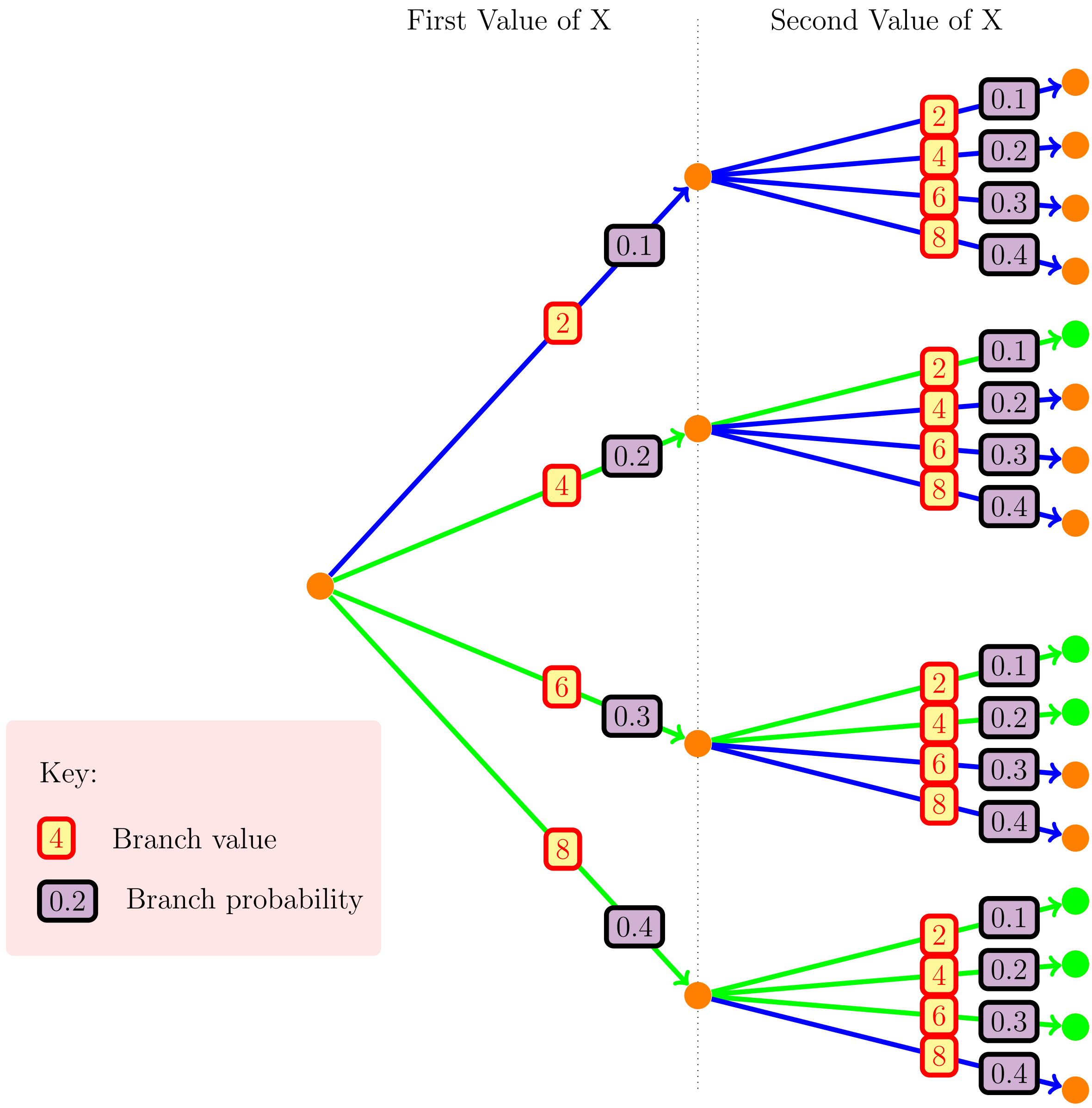
p(the first value of X is 6 AND the second value of X is 6) is: $0.3 \times 0.3 = 0.09$

p(the first value of X is 8 AND the second value of X is 8) is: $0.4 \times 0.4 = 0.16$

But we want any of these situations: so we just add up the probabilities of the events we want:

p(first X = second X) $= 0.01 + 0.04 + 0.09 + 0.16 = 0.3$

(ii) Now if we want the first value of X to be greater than the second value of X, then the tree diagram we need to draw emphasises those outcomes where the first value of X is greater than the second value of X. In the following diagram, those branches and outcomes are marked in green:



Again we need to find the probability of any of those green outcomes occurring. Remember how to do this? We simply multiply the probabilities of the branches to get the probability of the final outcome, so:

p(the first value of X is 4 AND the second value of X is 2) is: $0.2 \times 0.1 = 0.02$

p(the first value of X is 6 AND the second value of X is 2) is: $0.3 \times 0.1 = 0.03$

p(the first value of X is 6 AND the second value of X is 4) is: $0.3 \times 0.2 = 0.06$

p(the first value of X is 8 AND the second value of X is 2) is: $0.4 \times 0.1 = 0.04$

p(the first value of X is 8 AND the second value of X is 4) is: $0.4 \times 0.2 = 0.08$

p(the first value of X is 8 AND the second value of X is 6) is: $0.4 \times 0.3 = 0.12$

But we want any of these situations: so we just add up the probabilities of the events we want:

p(first $X >$ second X) $= 0.02 + 0.03 + 0.06 + 0.04 + 0.08 + 0.12 = 0.35$

2.1.7 Example 4

A curiously shaped six-faced die produces scores, X, for which the probability distribution is given by:

$$\begin{aligned} p(X = x) &= \frac{k}{x} && for\ x = 1, 2, 3, 4, 5, 6 \\ p(X = x) &= 0 && otherwise \end{aligned}$$

(i) Find the value of k and illustrate the probability distribution.

(ii) Show that, when this die is thrown twice, the probability of obtaining two equal scores is very nearly $\frac{1}{4}$.

2.1.8 Answer 4

(i) In order to find the value of k I think the easiest thing to do would be to tabulate the probability distribution, even though we don't know what k is yet. To do that, we have to put our probabilities in terms of k, but that's allowed! Here's my table:

$\boldsymbol{x}$	1	2	3	4	5	6
$\boldsymbol{p(X = x)}$	$\frac{k}{1}$	$\frac{k}{2}$	$\frac{k}{3}$	$\frac{k}{4}$	$\frac{k}{5}$	$\frac{k}{6}$

I calculated these probabilities from the probability distribution: the entry for $X = 4$, for example is $\frac{k}{4}$ because in the formula, when $x = 4$, then $p(X = x) = \frac{k}{4}$.

Now, because this is a probability distribution, then the sum of the probabilities must be 1. So:

$$\begin{aligned} 1 &= \frac{k}{1} + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} + \frac{k}{5} + \frac{k}{6} \\ &= k\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right) \\ &= k\left(\frac{60}{60} + \frac{30}{60} + \frac{20}{60} + \frac{15}{60} + \frac{12}{60} + \frac{10}{60}\right) \\ &= k\left(\frac{147}{60}\right) \end{aligned}$$

In which case,

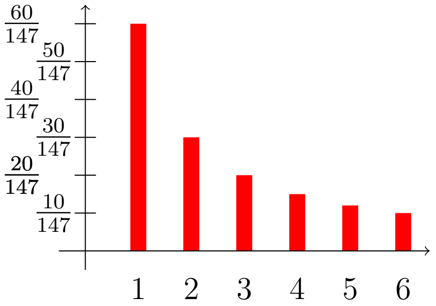
$$\begin{aligned} 60 &= 147k \\ \Longrightarrow \qquad k &= \frac{60}{147} = \frac{20}{49} \end{aligned}$$

Now, if we write out our probability distribution table again, this time incorporating the value of k that we have just discovered, then it turns out to be:

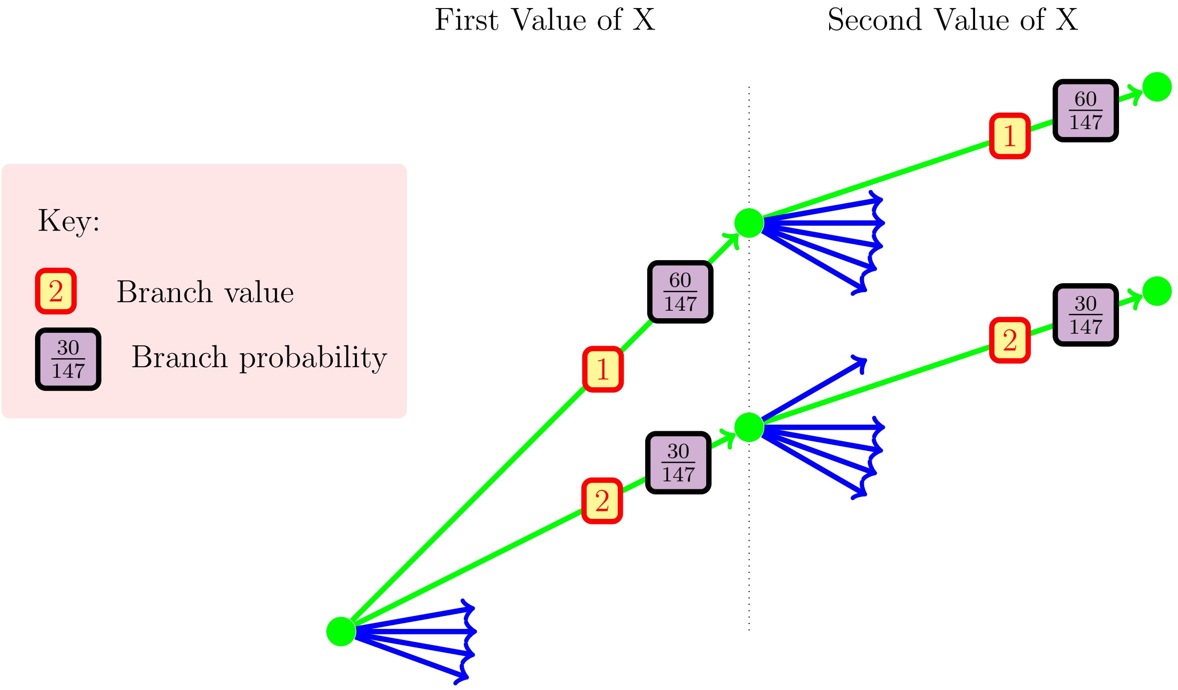
$\boldsymbol{x}$	1	2	3	4	5	6
$\boldsymbol{p(X=x)}$	$\frac{60}{147}$	$\frac{30}{147}$	$\frac{20}{147}$	$\frac{15}{147}$	$\frac{12}{147}$	$\frac{10}{147}$

because, for example, when $k = \frac{60}{147}$, then $p(X = 6) = \frac{60}{147 \times 6} = \frac{10}{147}$.

Here's a picture of the distribution:



(ii) For this part of the question, I would normally want to draw a tree diagram of all the possibilites when the die is thrown twice. But because there are so many ($6 \times 6 = 36$), instead I'm only going to show part of it:



The only bits I've drawn are the the branches and outcomes of the first two cases of the values of X being equal. Now to work out the probabilities of the outcomes, we have to multiply the probabilities of the relevant branches, so:

p(the first value of $X = 1$ AND the second value of $X = 1$) $= \frac{60}{147} \times \frac{60}{147} = \frac{3600}{21609}$;

p(the first value of $X = 2$ AND the second value of $X = 2$) $= \frac{30}{147} \times \frac{30}{147} = \frac{900}{21609}$ etc.

And we have to do this for all six outcomes where the the two values of X are the same. I've summarised the results in the following table:

	Second die roll					
First die roll	**1**	**2**	**3**	**4**	**5**	**6**
1	$\frac{3600}{21609}$					
2		$\frac{900}{21609}$				
3			$\frac{400}{21609}$			
4				$\frac{225}{21609}$		
5					$\frac{144}{21609}$	
6						$\frac{100}{21609}$

where I have only put in the values that I am interested in (where the two values of X are the same).

So, to find the total probability of the first die roll being equal to the second die roll, then we just add up all these probabilities:

$\frac{3600}{21609} + \frac{900}{21609} + \frac{400}{21609} + \frac{225}{21609} + \frac{144}{21609} + \frac{100}{21609} = \frac{5369}{21609} = 0.248$ (to 3 sf). Which is about a quarter.

2.1.9 Example 5

Three fair coins are tossed.

(i) By considering the set of possible outcomes, tabulate the probability distribution of X, the number of heads occurring.

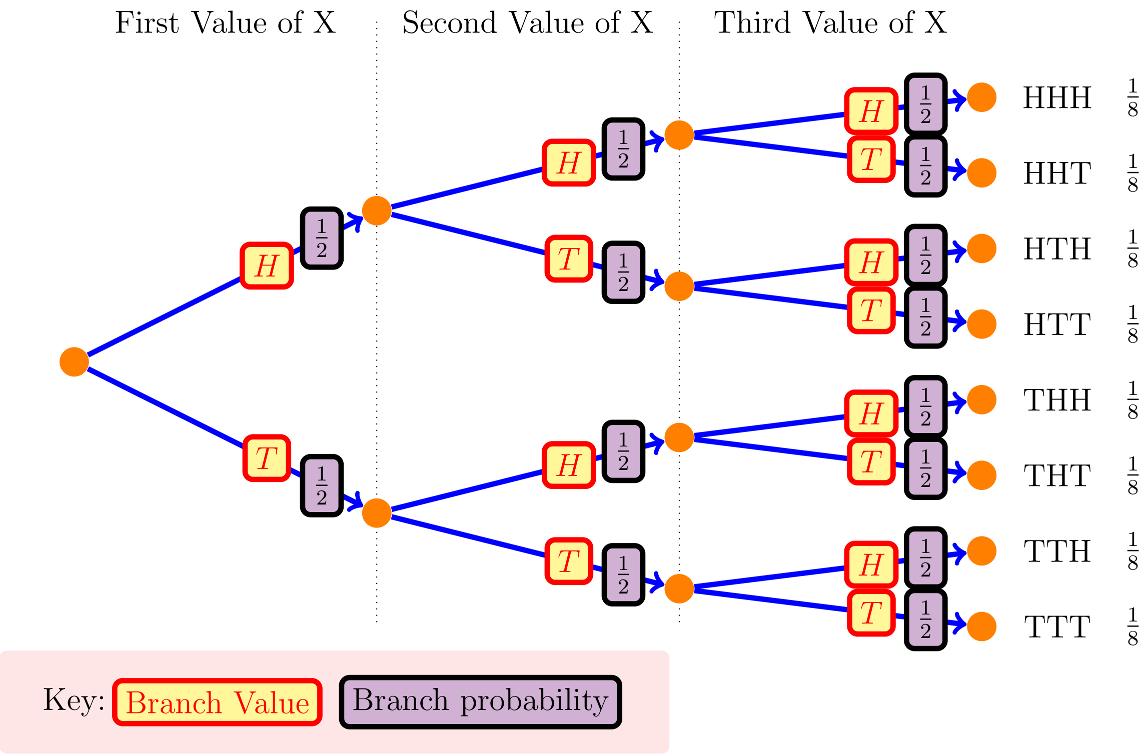
(ii) Illustrate the distribution, and describe its shape.

(iii) Find the probability that there are more heads than tails.

(iv) Without further calculation, state whether your answer to part (iii) would be the same if four fair heads were tossed. Give a reason for your answer.

2.1.10 Answer 5

(i) Well, I think another tree diagram is in order:



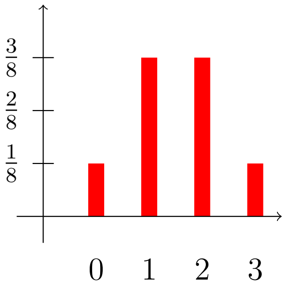
And because each branch has a probability of $\frac{1}{2}$ associated with it, then each outcome will have a probability of $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$.

So, in order to compile our distribution table, we just add up all the probabilities for a given number of heads. Looking at the possible outcomes, we could have no heads, 1 head, 2 heads or 3 heads, so:

$\boldsymbol{x}$	0	1	2	3
$\boldsymbol{p(X = x)}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

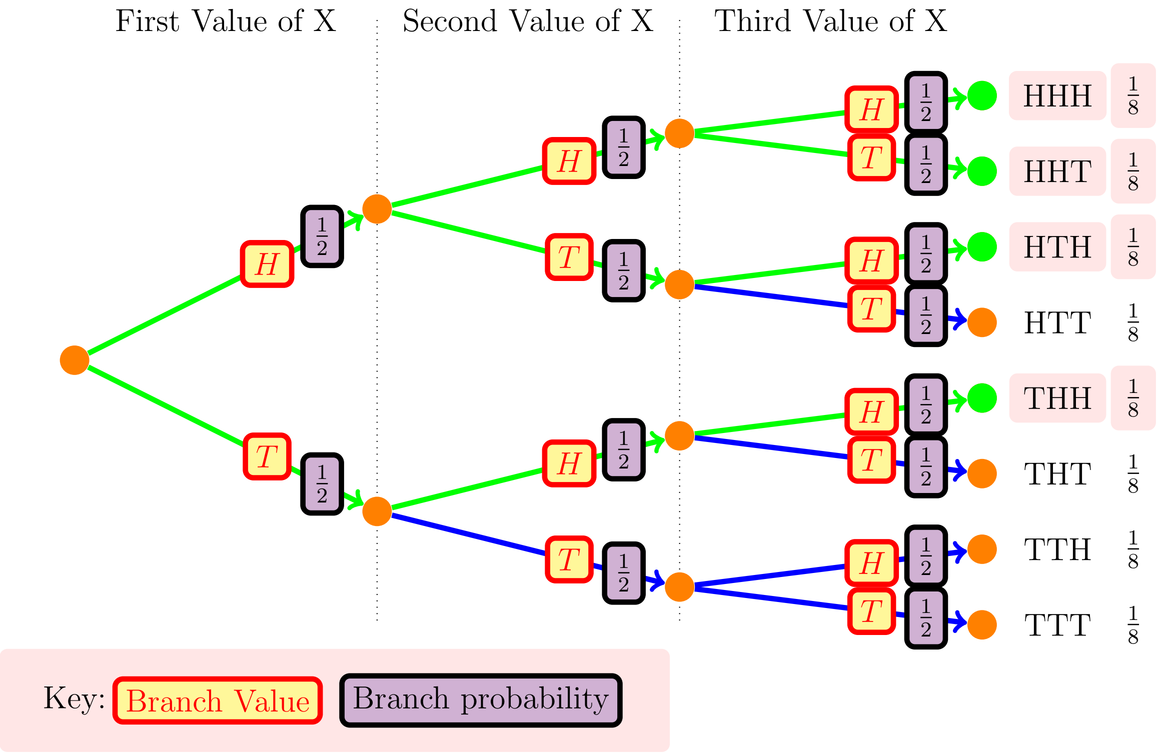
where, for example, the probability of getting 2 heads is $\frac{3}{8}$ since there are three outcomes, each having a probability of $\frac{1}{8}$, of getting 2 heads (HHT, HTH, THH).

(ii) A picture of the distribution is:



The distribution is clearly symmetric.

(iii) For the probability of getting more heads than tails, we look for the outcomes that give us more heads than tails. I've marked the outcomes with green dots, and highlighted the probabilities:



So the probability of getting more heads than tails will be $\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{4}{8}=\frac{1}{2}$.

(iii) I don't think the answer would be the same if we had four tosses. My reasoning goes like this: with an odd number of tosses, then you can never have the same number of heads and tails. So you will either have more heads than tails, or more tails than heads. Since it is equally likely to get a head or a tail on each toss, then the probability of getting either more heads than tails or more tails than heads must be the same. This can only happen if those probabilities are each $\frac{1}{2}$.

However, if you have an even number of tosses, then there must be some probability of getting the same number of heads as tails. Since, as before, the probability of getting more heads than tails, or more tails than heads will still be the same, then those probabilities must be less than $\frac{1}{2}$ if you have an even number of tosses.

3 Expectation and Variance

Now you've probably come across the idea that a set of data can have a mean (an estimate of a typical value of the data in your set) and a standard deviation (an estimate of how widely your data is spread out).

Well, you can have the same things with probability distributions. With probability distributions, the typical-value-thing is called the ***Expectation***, and the spread-measure-thing is called the ***Variance***. The variance is just the square of the standard deviation, actually.

In fact, you can think of a discrete random variable as a means of *generating* a set of data. Each time you roll a die, for example, you are generating a piece of data that you can turn into a set. The set might look like this:

$$3,\ 6,\ 2,\ 6,\ 3,\ 1,\ 5,\ 3,\ ...$$

and of course when we have finished generating our set of data, then we could work out the mean and the standard deviation of it, just like we could with any other set of data.

3.1 The Definition of Expectation

Well, to start with, $E(X)$ is the shorthand for "the expected value (i.e. the expectation) of the DRV X". Now we know this, we can be told about the formula for calculating it:

$$E(X) \ = \ \sum_{i=1}^{i=n} x_i \times p(X = x_i)$$

Now what on Earth does that mean? If you need a refresher on this kind of notation, check out Appendix A. Assuming you're OK with the $\sum$ notation, then what this formula tells us is that to calculate the expected value of a DRV (another way of putting this would be: "the mean of values generated by that DRV") what you do is to multiply all the possible values of the DRV by the probability of that value occurring, then add up (that's the $\sum$ bit) all of those products.

It sounds harder than it actually is. Check out these examples.

3.1.1 Expectation Example: Rolling A Die

Let's take the example of the die. You can only get values of 1, 2, 3, 4, 5 or 6 when you roll a die. And (assuming it is a fair die) each of the six values is equally likely to come up on any given roll. So in this case, we could work out an estimate of the "average" roll like this.

Let's say we rolled this die 6000 times. Because this is a lot of rolls, then we would expect to get roughly 1000 of each number come up. We won't get exactly 1000 of each number, but the more rolls we do, the closer we get to having $\frac{1}{6}$ of them being 1s, etc. So let's assume for the moment that we did have exactly 1000 1s, 1000 2s, etc. How could we calculate the mean?

Well, we'd have to add up all the values, and divide by 6000. But how do we add up the values? Have you thought of a way of doing this? What we could do is this: if we have 1000 1s, then they will add up to 1000 × 1 = 1000; if we have 1000 2s, then they will add up to 1000 × 2 = 2000, etc.

So we could calculate our mean as:

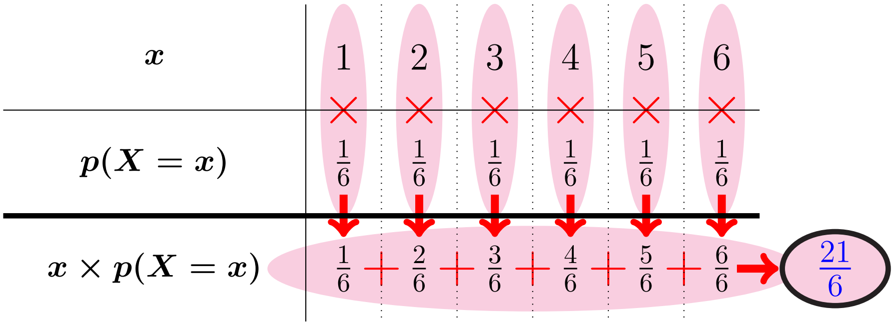
$$mean\ of\ dice\ rolls \ = \ \frac{1000 \times 1 \ + 1000 \times 2 + 1000 \times 3 + 1000 \times 4 \ + 1000 \times 5 + 1000 \times 6}{6000} \ = \ 3.5$$

Which seems entirely reasonable when you think about it: 3.5 is exactly in the middle of the set of numbers 1, 2, 3, 4, 5 and 6!

OK. Now let's check out our formula for calculating the Expectation of this DRV. The formula is:

$$E(X) \;=\; \sum_{i=1}^{i=n} x_i \times p(X = x_i)$$

And the best way to do these calculations is to draw up a probability distribution table, like we've done before, but this time with an extra row in it, representing the products that we will get:



$\boldsymbol{x}$	1	2	3	4	5	6	
$\boldsymbol{p(X = x)}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	
$\boldsymbol{x \times p(X = x)}$	$\frac{1}{6}$ +	$\frac{2}{6}$ +	$\frac{3}{6}$ +	$\frac{4}{6}$ +	$\frac{5}{6}$ +	$\frac{6}{6}$ →	$\frac{21}{6}$

So this is how the Expectation calculation works: multiply each value x by it's probabilty $p(X = x)$. Then add up all the products:

$$\begin{aligned}
E(X) &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} \\
&= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \\
&= \frac{21}{6} \\
&= 3.5
\end{aligned}$$

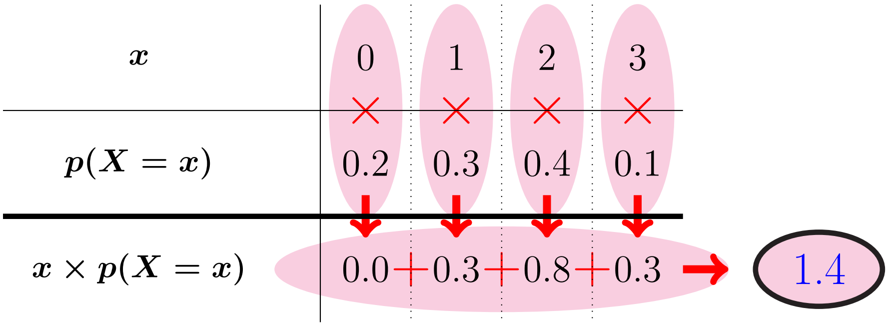
And the result is the same as we had before. This gives us a good feeling that the Expectation of the DRV is equivalent to the mean of data generated by it. Let's have a look at another example.

3.1.2 Another Example of Calculating an Expectation

Let's say we had a DRV described by the following probability distribution:

$\boldsymbol{x}$	0	1	2	3
$\boldsymbol{p(X = x)}$	0.2	0.3	0.4	0.1

Then proceeding as before, we multiply each value x by the probability of the value $p(X = x)$ and put that result in a new row of our table:



x	0	1	2	3
$p(X = x)$	0.2	0.3	0.4	0.1
$x \times p(X = x)$	0.0	0.3	0.8	0.3

Then we add up all the products, and get the answer of :

$$\begin{aligned} E(X) &= 0 \times 0.2 \ + \ 1 \times 0.3 \ + \ 2 \times 0.4 \ + \ 3 \times 0.1 \\ &= 0.0 \ + \ 0.3 \ + \ 0.8 \ + \ 0.3 \\ &= 1.4 \end{aligned}$$

So for this DRV, if we took lots of values of the DRV and found the mean of those values, we'd get 1.4. This again seems about right to me, since 1.4 is roughly in the middle of the values 0, 1, 2 and 3.

3.2 The Definition of Variance

Variance is defined in a similar way to Expectation, in that it is defined in terms of the $\sum$ notation. It's just a bit more complicated:

$$Var(X) \ = \ \left\{ \sum_{i=1}^{i=n} x_i^2 \times p(X = x_i) \right\} \ - \ \{E(X)\}^2$$

Actually the $\sum$ bit is very similar to the Expectation: the only difference is that we've got an x^2 instead of an x.

And there's a new bit at the end: once we've done the $\sum$ calculations, we have to subtract the square of the Expectation from the result! Phew!

That means we have to calculate the Expectation before we can calculate the Variance, as the Expectation is used in the calculation of Variance. Let's see what we'd get then if we used this definition on the two examples we've just calculated the Expectation for.

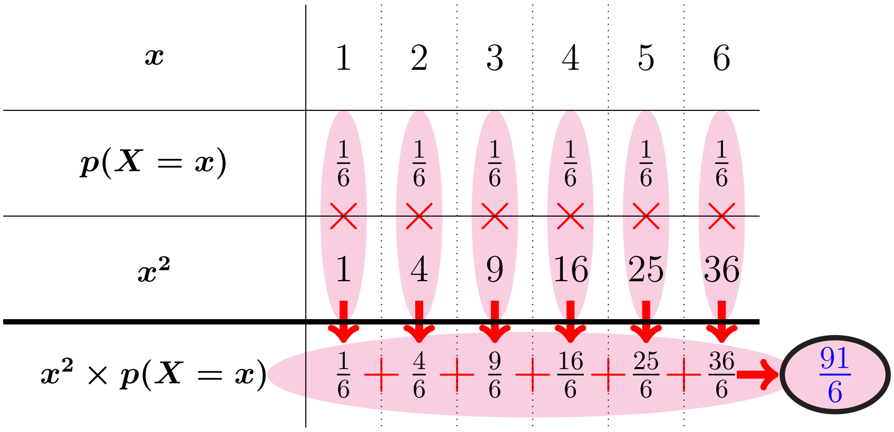
3.2.1 Variance Example: Rolling A Die

Let's check out our formula for calculating the Variance of this DRV. The formula is:

$$Var(X) \ = \ \left\{ \sum_{i=1}^{i=n} x_i^2 \times p(X = x_i) \right\} \ - \ \{E(X)\}^2$$

We've already calculated the Expectation $E(X)$ for this DRV: it's 3.5.

And the best way to calculate the $\sum$ bit is to draw up a probability distribution table, like we've done before, but this time with extra rows in it:



x	1	2	3	4	5	6
$p(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
x^2	1	4	9	16	25	36
$x^2 \times p(X = x)$	$\frac{1}{6}$ +	$\frac{4}{6}$ +	$\frac{9}{6}$ +	$\frac{16}{6}$ +	$\frac{25}{6}$ +	$\frac{36}{6}$ $\rightarrow \frac{91}{6}$

This time we need two extra rows: one for the x^2 values, and then the other for the products. And this is how the Variance calculation works: multiply each value x^2 by the probabilty $p(X = x)$. Then add up all the products. And don't forget we have to subtract the square of the Expectation at the end as well:

$$\begin{aligned} Var(X) &= 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + 25 \times \frac{1}{6} + 36 \times \frac{1}{6} - 3.5^2 \\ &= \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} - 12.25 \\ &= \frac{91}{6} - 12.25 \\ &= 2.917 \ (to\ 4\ sf) \end{aligned}$$

Remember that I said that the Variance is the square of the Standard Deviation? Then the Standard Deviation measure for rolling a fair die would be

$$\begin{aligned} SD(X) &= \sqrt{\frac{91}{6} - 12.25} \\ &= 1.708 \ (to\ 4\ sf) \end{aligned}$$

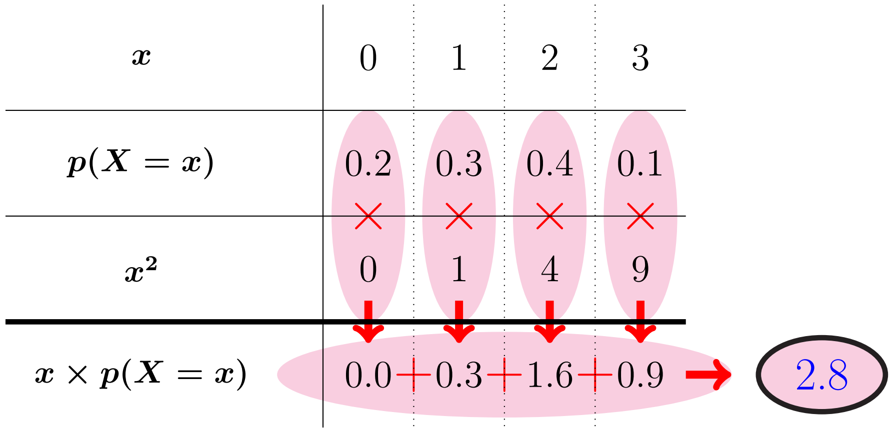
This is the number that gives you an idea of the typical distance a value is from the mean of the DRV.

3.2.2 Another Example of Calculating a Variance

Let's say we had the same DRV described by the probability distribution that we saw a bit earlier:

x	0	1	2	3
$p(X = x)$	0.2	0.3	0.4	0.1

Then proceeding as before, we multiply each value x^2 by the probability of the value $p(X = x)$ and put that result in a new row of our table:



x	0	1	2	3	
$p(X = x)$	0.2	0.3	0.4	0.1	
	×	×	×	×	
x^2	0	1	4	9	
$x \times p(X = x)$	0.0 +	0.3 +	1.6 +	0.9	→ 2.8

Remembering that the Expectation of this DRV was 1.4, then the Variance is calculated by:

$$\begin{aligned} Var(X) &= 0 \times 0.2 \ + \ 1 \times 0.3 \ + \ 4 \times 0.4 \ + \ 9 \times 0.1 \ - \ 1.4^2 \\ &= 0.0 \ + \ 0.3 \ + \ 1.6 \ + \ 0.9 \ - \ 1.4^2 \\ &= 2.8 \ - \ 1.96 \\ &= 0.84 \end{aligned}$$

So for this DRV, if we took lots of values of the DRV and found the Variance of those values, we'd get 0.84.

At this point, you should check out Appendix B, where I show you something interesting!

A Sigma ($\sum$) Notation

Often in science and mathematics, you'll come across the $\sum$ notation. A typical example is this kind of thing:

$$E(X) \ = \ \sum_{i=1}^{i=n} x_i \times p(X = x_i)$$

So what does this mean? Well, the $\sum$ notation just represents a lot of things added together. OK, but what things? And how many?

Well in this example, below the $\sum$ sign we have $i = 1$. i is a count variable (that could have any name: here I've called it i, but you can use any letter you like) that's used to show how many things we are going to add together. In this case, the count starts at 1. Above the $\sum$ we have $i = n$. That tells us that the count stops at n. What's n? In statistics, n is most often used to denote how many items there are in our set of data. So, translated into English,

$$\sum_{i=1}^{i=n}$$

simply means "add together the following set of n things...".

Right. But what is it that we are going to add together? Well, in this situation, the bit after the $\sum$ sign is

$$x_i \times p(X = x_i)$$

so what's that all about? If you ignore the i for a moment, and remember that we are dealing here with DRVs, you should be able to twig that x represents a value that the DRV can take, and $p(X = x)$ is the probability of that value turning up when the DRV is actioned[1]. But what about the i? Remember that it is a count variable: it counts through each item in the set of things you are adding up. So what

$$\sum_{i=1}^{i=n} x_i \times p(X = x_i)$$

means is "for each value X can take, multiply that value by the probability of that value turning up. Add all those products[2] together".

---

[1] E.g. when a die is rolled, or a coin tossed.

[2] A *product* is what mathematicians call two things multiplied together.

B An Alternative Formula For the Variance

Check this out. I want to figure out what

$$E\left[(X-\mu)^2\right]$$

is.

This is the Expectation of $X-\mu$. Using the definition of Expectation, this will be

$$E\left[(X-\mu)^2\right] = \sum_{i=1}^{i=n}(x_i-\mu)^2 \times p(X=x_i)$$

Now if you think about it, we could multiply out those brackets within the $\sum$, and get this:

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

since μ is just a constant, so we can take it outside the $\sum$s. But the middle term in the above equation is just $-2\mu E(X)$ from the definition of $E(X)$, and the third term is just μ^2 because the sum of the probabilities of a DRV add up to 1!! So:

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

So far, I haven't said anything about this μ. It could be *anything.* What if I made it equal to $\mu = E(X)$? What would we have then? Let's see:

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

Wow!!

That means that we have *two* ways of working out the Variance of a DRV. They are:

$$Var(X) \;=\; \left\{\sum_{i=1}^{i=n} x_i^2 \times p(X = x_i)\right\} \;-\; \{E(X)\}^2 \;=\; E(X^2) \;-\; [E(X)]^2$$

which is the typical definition, and also

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

Two Ways To Calculate Variance:

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

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