Information Theory – Exercises (All Answers)

Exercises for Chapter 6 of Steinhart, E. (2017) *More Precisely: The Math You Need to Do Philosophy*. Broadview Press. (Version 2)

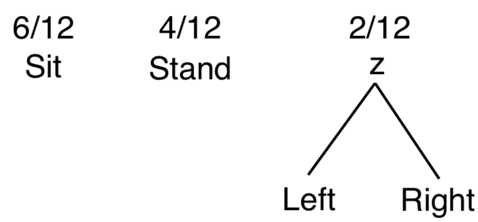
6.1 Compute the following:

$\log_3 9 = 2$ $\log_3 81 = 4$ $\log_{10} 100 = 2$ $\log_{10} 0.01 = -2$

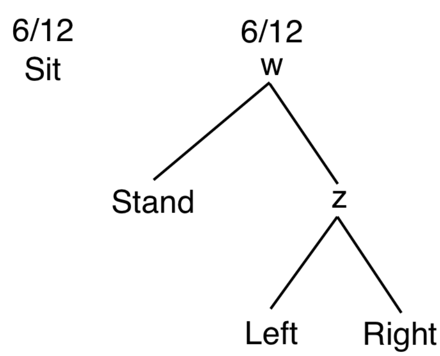
6.2A Construct the Huffman tree for these symbols and associated probabilities.

Command	Sit	Stand	Left	Right
Probability	6/12	4/12	1/12	1/12

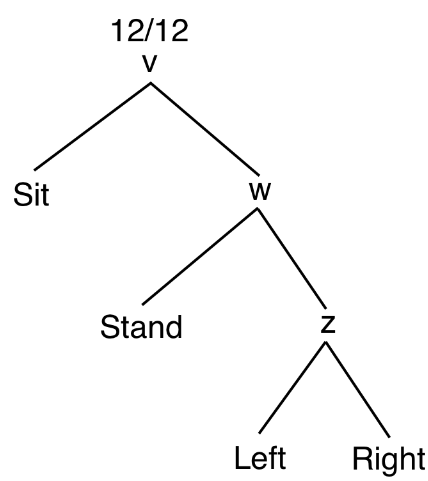
First merge Left and Right:



Second merge Stand and z:

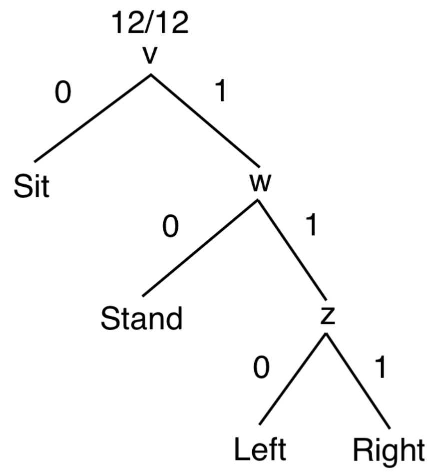


Third merge Sit and w:



6.2B Write down the binary codes for the symbols for the tree in 7.2A.

First decorate the branches of the tree with 0s and 1s:



Now derive the code by starting with v and following the path to the command:

Command	Sit	Stand	Left	Right
Code	0	10	110	111

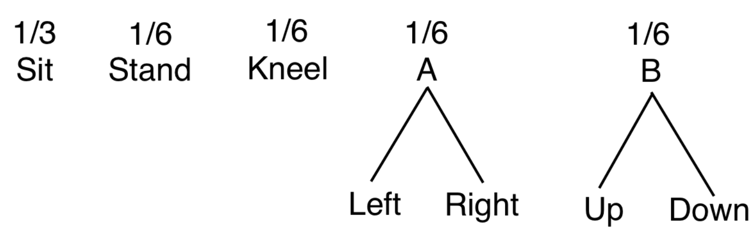
6.2C Calculate the average information of the code in 7.2B. The average is calculated by summing, for each command, its number of bits times its probability. Thus:

$$\text{Average} = \left(1\frac{6}{12} + 2\frac{4}{12} + 3\frac{1}{12} + 3\frac{1}{12}\right) = \left(\frac{6}{12} + \frac{8}{12} + \frac{3}{12} + \frac{3}{12}\right) = \frac{20}{12} = 1\frac{3}{4} \text{ bits.}$$

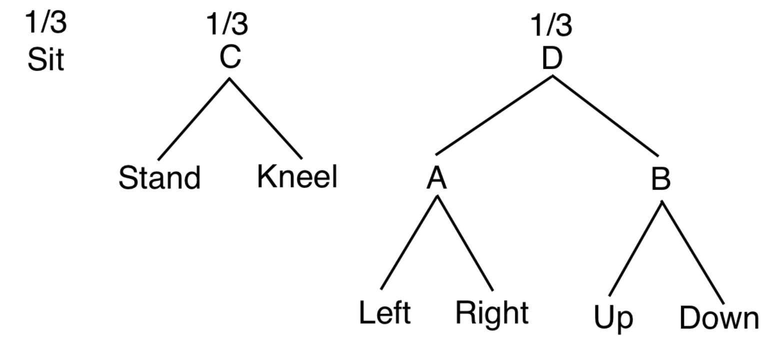
6.3A Construct a Huffman tree for these symbols and associated probabilities:

Command	Sit	Stand	Kneel	Left	Right	Up	Down
Probability	1/3	1/6	1/6	1/12	1/12	1/12	1/12

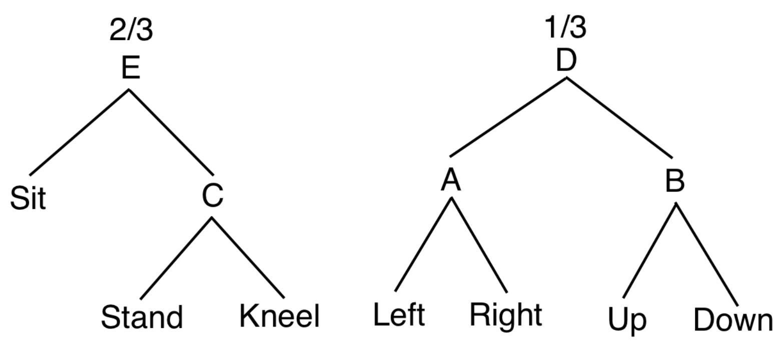
Begin by combining the symbols with 1/12 probability:



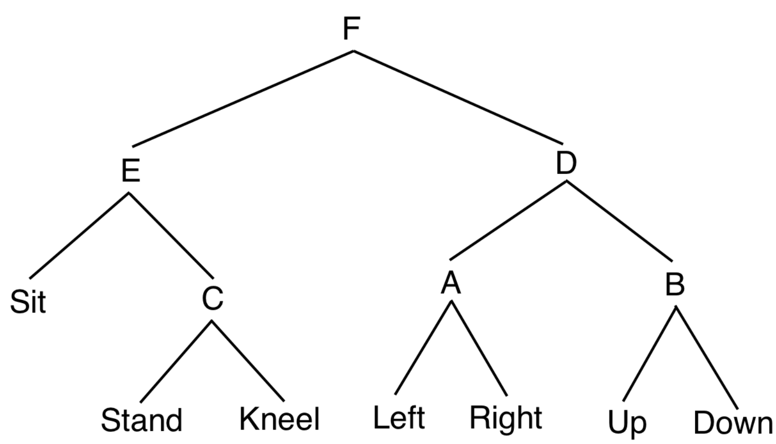
Now combine the symbols with 1/6 probability:



Combine Sit with C:

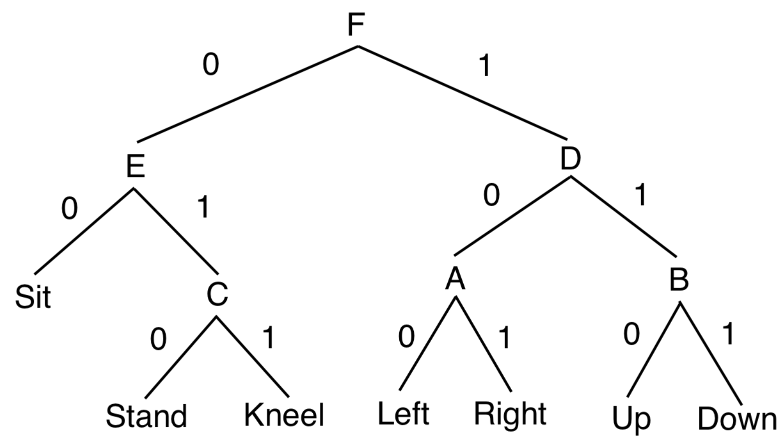


Combine E with D:



6.3B Write down the binary codes for the symbols for the tree in 7.3A.

First decorate the tree with 0s and 1s:



Now write down the binary codes by starting from F and following a path to the symbol:

Command	Sit	Stand	Kneel	Left	Right	Up	Down
Probability	00	010	011	100	101	110	111

6.3C Calculate the average information of the code in 7.3B.

$$\text{Average} = \left(1\frac{1}{3} + 3\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{12} + 4\frac{1}{12} + 4\frac{1}{12} + 4\frac{1}{12}\right) = \frac{32}{12} = 2\frac{3}{4} \text{ bits.}$$

6.4 The pie charts below illustrate the relative probabilities of the four values of a variable. That is, the area in each slice of pie represents the probability of the value. Which pie chart has the lowest entropy? Which has the highest entropy?

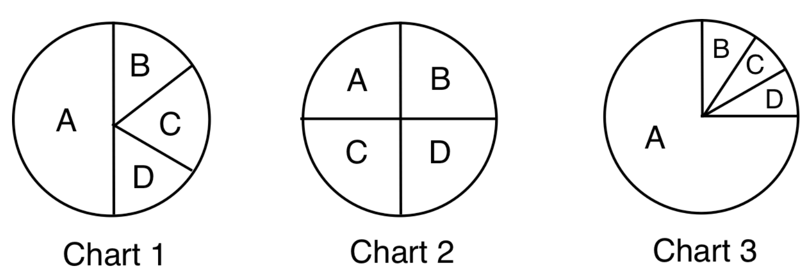


Chart 3 has the lowest entropy (the probabilities are least equally distributed, that is, the graph of the probabilities has the least flatness); Chart 2 has the highest entropy (the probabilities are most equally distributed, that is, the graph of the probabilities has the most flatness).

6.5 Calculate the Shannon entropy of a fair 8-sided die. The calculation looks like this:

$$H\left(\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8}\right) = -\left[8\left(\frac{1}{8}\log\frac{1}{8}\right)\right] = -\left[\log\frac{1}{8}\right] = -[-3] = 3 \text{ bits.}$$

6.6 Calculate the Shannon entropy of the following alphabet:

a	e	i	u	p	t	k	w
1/4	1/8	1/8	1/8	1/8	1/8	1/16	1/16

The calculation looks like this:

$$\begin{aligned} H\left(\frac{1}{4},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{16},\frac{1}{16}\right) &= -\left[\left(\frac{1}{4}\log\frac{1}{4}\right)+5\left(\frac{1}{8}\log\frac{1}{8}\right)+2\left(\frac{1}{16}\log\frac{1}{16}\right)\right] \\ &= -\left[-\left(\frac{2}{4}\right)-\left(\frac{15}{8}\right)-\left(\frac{8}{16}\right)\right] \\ &= 2\frac{7}{8}\text{bits} \end{aligned}$$

6.7 Calculate the Shannon entropy of the following alphabet:

a	e	i	u	p	t	k	w
1/4	1/16	1/16	1/8	1/4	1/16	1/16	1/8

The calculation looks like this:

$$\begin{aligned} H\left(\frac{1}{4},\frac{1}{16},\frac{1}{16},\frac{1}{8},\frac{1}{4},\frac{1}{16},\frac{1}{16},\frac{1}{8}\right) &= -\left[2\left(\frac{1}{4}\log\frac{1}{4}\right)+2\left(\frac{1}{8}\log\frac{1}{8}\right)+4\left(\frac{1}{16}\log\frac{1}{16}\right)\right] \\ &= -\left[-\left(\frac{4}{4}\right)-\left(\frac{6}{8}\right)-\left(\frac{16}{16}\right)\right] \\ &= 2\frac{3}{4}\text{ bits.} \end{aligned}$$

6.8 Calculate the entropy of the weather in Table 6.10. The calculation follows the same logic as the calculation of the entropy of the activity in Table 6.10. To calculate the entropy of the weather, it helps to use a scientific calulator or a spreadsheet like Excel. The calculation looks like this:

$$H(X \mid Y = \text{rainy}) = H\left(\frac{0}{90},\frac{0}{90},\frac{90}{90}\right) = 0 \text{ bits.}$$

6.9 Bob records his activities and the weather on the same days as Alice. Out of 256 days: 90 are sunny, 76 are cloudy, and the remaining 90 are rainy. On any day, Bob either jogs, does yoga, or just sits. He correlates his activities with the weather in Table 6X.1 (which resembles Alice's Table 6.10). The variable X ranges over his activities while the variable Y ranges over the weather. Note that Bob's activities are not correlated with the weather at all; they are random.

P(X,Y)		X			$P_M(Y)$
		jogs	yoga	sits	
Y	sunny	$\frac{30}{256}$	$\frac{30}{256}$	$\frac{30}{256}$	$\frac{90}{256}$
	cloudy	$\frac{25}{256}$	$\frac{26}{256}$	$\frac{25}{256}$	$\frac{76}{256}$
	rainy	$\frac{30}{256}$	$\frac{30}{256}$	$\frac{30}{256}$	$\frac{90}{256}$
$P_M(X)$		$\frac{85}{256}$	$\frac{86}{256}$	$\frac{85}{256}$	1

Table 6X.1 A joint probability distribution P(activity, weather) for Bob.

6.9A Calculate the entropy of Bob's activity. The calculation looks like this:

$$\begin{aligned}
\text{H(activity)} &= \text{H}(\frac{85}{256},\frac{86}{256},\frac{85}{256}) \\
&= -\left[\left(\frac{85}{256}\log\frac{85}{256}\right)+\left(\frac{86}{256}\log\frac{86}{256}\right)+\left(\frac{85}{256}\log\frac{85}{256}\right)\right] \\
&= 1.5849 \text{ bits.}
\end{aligned}$$

6.9B Calculate the entropy of the weather. The entropy of the weather is the same as in exercise 6.8.

H(weather) = 1.5806 bits.

6.9C Calculate the conditional entropy H(X|Y).

H(X|Y) = 1.5849 bits.

6.9D Calculate the mutual information I(activity; weather) for Bob.

I(activity; weather) = H(activity) – H(activity | weather)
= 1.5849 – 1.5849
= 0 bits.

Thus Bob's activity carries 0 bits of information about the weather. Bob's activity does not represent the weather at all. His activity and the weather are independent.

6.9E Carefully compare the case of Bob with the case of Alice to see the difference between carrying some information (Alice) and carrying none (Bob).

6.10 Stacy records her activities and the weather on the same days as Alice. Out of 256 days: 90 are sunny, 76 are cloudy, and the remaining 90 are rainy. On any day, Stacy either jogs, does yoga, or just sits. She correlates her activities with the weather in Table 6X.2 (which resembles Alice's Table 6.10). The variable X ranges over her activities while the variable Y ranges over the weather. Note that Stacy's activities are perfectly correlated with the weather.

P(X,Y)		X			$P_M(Y)$
		jogs	yoga	sits	
Y	sunny	$\frac{90}{256}$	$\frac{0}{256}$	$\frac{0}{256}$	$\frac{90}{256}$
	cloudy	$\frac{0}{256}$	$\frac{76}{256}$	$\frac{0}{256}$	$\frac{76}{256}$
	rainy	$\frac{0}{256}$	$\frac{0}{256}$	$\frac{90}{256}$	$\frac{90}{256}$
$P_M(X)$		$\frac{90}{256}$	$\frac{76}{256}$	$\frac{90}{256}$	1

Table 6X.2 A joint probability distribution P(activity, weather) for Stacy.

6.10A Calculate the entropy of the activity. The entropy of the activity is the same as in exercise 6.8.

H(activity) = 1.5806 bits.

6.10B Calculate the entropy of the weather. The entropy of the weather is the same as in exercise 6.8.

H(weather) = 1.5806 bits.

6.9C Calculate the conditional entropy H(X|Y). Here are the intermediate calculations:

$$H(X \mid Y = \text{sunny}) = H\left(\frac{90}{90}, \frac{0}{90}, \frac{0}{90}\right) = 0 \text{ bits;}$$

$$H(X \mid Y = \text{cloudy}) = H\left(\frac{0}{90}, \frac{76}{90}, \frac{0}{90}\right) = 0 \text{ bits;}$$

$$H(X \mid Y = \text{rainy}) = H\left(\frac{0}{90}, \frac{0}{90}, \frac{90}{90}\right) = 0 \text{ bits.}$$

The result is H(X|Y) = 0 bits.

6.9D Calculate the mutual information I(activity; weather) for Stacy.

I(activity; weather) = H(activity) – H(activity | weather)
= 1.5806 – 0
= 1.5806 bits.

Thus Stacy's activity carries 1.5806 bits of information about the weather. Her activity perfectly represents the weather; it carries all the information about the weather that can be carried. That is, there is as much information about the weather in Stacy's activity as there is in the weather itself.

6.9E Carefully compare the case of Bob with the case of Alice to see the difference between carrying some information (Alice) and carrying none (Bob).

6.10 One of the striking features of mind-body dualism, theism, and telepathy is that advocates of those positions never talk about them using information theory. Why not?

6.11 It can be shown that $H(X) - H(X|Y) = H(Y) - H(Y|X)$. Use this and the Chain Rule to show that the definition of $I(X;Y)$ in terms of joint entropy is equivalent to the definition of $I(X;Y)$ in terms of conditional entropy.

The algebra looks like:

$$
\begin{aligned}
I(X; Y) &= H(X) + H(Y) - H(X, Y) \\
&= H(X) + H(Y) - (H(X) + H(Y|X)) \quad \text{Chain Rule} \\
&= H(Y) - H(Y|X) \\
&= H(X) - H(X|Y) \\
&= I(X; Y)
\end{aligned}
$$