Exercise 3.1

a).

$H = \sum_{x \in X} P_x \log \frac{1}{P_x} \approx 4.219$ bit

(b) By using Huffman code in my python algorithm, I got these results.

a → 1111	h → 0111	o → 1011	u → 01101
b → 110000	i → 1010	p → 110001	v → 0110011
c → 00100	j → 011001000	q → 011001001	w → 00111
d → 11101	k → 011000	r → 1101	x → 011001010
e → 010	l → 11001	s → 1000	y → 111001
f → 00101	m → 00110	t → 000	z → 011001011
g → 111000	n → 1001		

(c) expected length $= \sum_{x \in X} P_x \cdot \mathrm{len}(P_x) \approx 4.221$ bit

Compared to (a), is a little more than the entropy

Exercise 3.2

For (a). It is valid

For (b) It is valid

For (c) It is not valid because the number of longest symbol should be ≥ 2

In this case, it is unreasonable.

For (d) It is not valid because this code length is waisted

{0, 1} is more efficient than previous one.

For (e). It is not valid because {1} symbol is unreasonable.

In order to get reasonable, it should be {011}.

~~Exercise 3.3~~

Exercise 3.3

(a) $a \to 8.4\%$ 00 — $b \to 1.5\%$ 0110 — $c \to \cancel{1.0\%}\ 2.2\%$ ~~100~~ 0001 — $d \to 4.2\%$ 011 — $e \to 11.0\%$ 0 — $f \to 2.2\%$ 0010 — $g \to 2.0\%$ 0011 —

$h \to 6.0\%$ 010 — $i \to 7.4\%$ 10 — $j \to 0.1\%$ 1001 — $k \to 1.3\%$ 0111 — $l \to 4.0\%$ 100 — $m \to \cancel{00}\ 2.4\%$ 0000 — $n \to 6.7\%$ 000 —

$o \to 7.4\%$ 11 — $p \to 1.9\%$ 0101 — $q \to 0.1\%$ 1010 — $r \to 7.5\%$ 01 — $s \to 6.2\%$ 001 — $t \to 9.2\%$ 1 —

$u \to 2.7\%$ 110 — $v \to 0.9\%$ 1000 — $w \to 2.5\%$ 111 — $x \to 0.1\%$ 1011 — $y \to 2.0\%$ 0100 — $z \to 0.1\%$ 1100 —

(b) Expected length $= \sum_{x \in X} P_x \cdot \text{length}(P_x) = 3.457$ bit

(c) Morse code. It is a little similar if we change $0 \to \cdot$ $1 \to -$ $- \to$

Exercise 3.4

(a). From the statement, we know $\left(\sum_{j=1}^{M} 2^{-l_j}\right)^n$, for each component, they are independent

$$\left(\sum_{j=1}^{M} 2^{-l_j}\right)^n = \left(\sum_{j_1=1}^{M} 2^{-l_1}\right)\left(\sum_{j_2=1}^{M} 2^{-l_2}\right)\cdots\left(\sum_{j_n=1}^{M} 2^{-l_{j_n}}\right)$$

$$= \left(\sum_{j_1=1}^{M}\sum_{j_2=1}^{M}\cdots\sum_{j_n=1}^{M} 2^{-l_j} \cdot 2^{-l_2}\cdots 2^{-l_{j_n}}\right.$$

$$= \sum_{j_1=1}^{M}\sum_{j_2=1}^{M}\cdots\sum_{j_n=1}^{M} 2^{-(l_j+l_{j_2}+\cdots l_{j_n})}$$

(b). We rewrite $l = l_{j_1} + l_{j_2} + \cdots l_{j_n}$

$$\Rightarrow \left(\sum_{j=1}^{M} 2^{-l_j}\right)^n = \sum_{j_1=1}^{M}\sum_{j_2=1}^{M}\cdots\sum_{j_n=1}^{M} \cdot 2^{-l}$$

$$= \sum_{j=n}^{M \cdot ln} I^{-l}$$

$$= \sum_{l=n}^{n\, l_{max}} A_l \cdot 2^{-l}$$

(c) From (b), and using $\left(\sum_{j=1}^{M} 2^{-l_j}\right)^n = \sum_{n}^{n l_{max}} A_l \cdot 2^{-l}$

We know $\left(\sum_{j=1}^{M} 2^{-l_j}\right)^n \le 1 \Rightarrow A_l \cdot 2^{-l} \le 1$

$\Downarrow$

$A_l \le 2^l$

$\Downarrow$

$A_i \le 2^i$

Hence $\left(\sum_{j=1}^{M} 2^{-l_j}\right)^n \le n \cdot l_{max}$

Exercise 3.5.

(a). For any d, x be a random variable on $\{0, 1, 2, \cdots d-1\}$

First $\lceil \log_2 \frac{1}{P_X(x)} \rceil$ expression is uniquely decodable

because $\lceil \log_2 \frac{1}{P_X(x)} \rceil \ge \log_2 \frac{1}{P_X(x)}$

(b). $0 \to 0.1 \ (\frac{1}{2}) \Rightarrow 1$

$1 \to 0.\cancel{100}0010101\cdots \ (\frac{1}{6}) \Rightarrow 001$

$2 \to \cancel{0.100100\cdots}\ 0010101\cdots \ (\frac{1}{6}) \Rightarrow 001$

$3 \to \cancel{0.100100\cdots}\ 0.0010101 \ (\frac{1}{6}) \Rightarrow 001$

$H = \frac{1}{2} \times 1 + \frac{1}{6} \times 3 \times 3 = 2$

(c) If we use Huffman code, we express

$0 \to 0$

$1 \to 10$

$2 \to 110$

$3 \to 111$

$H = 1 \times 0.5 + 2 \times \frac{1}{6} + 2 \times 3 \times \frac{1}{6} = 1.8333\cdots$

Therefore, the expected length of Huffman code is shorter