Question 1.

ISBN code is 978-80-74?2-498-7, using the ISBN-13 check digit calculation we are able to determine missing number.

$$9*1+7*3+8*1+8*3+0*1+7*3+4*1+x*3+2*1+4*3+9*1+8*3+7*1 \equiv 0 \ (mod\ 10)$$

$$141 + 3x \equiv 0 \ (mod\ 10)$$

$$x = 3$$

Using the complete ISBN 978-80-7432-498-7, we are able to reveal title of the book which is *Antifragilita: Jak těžit z nahodilosti, neurčitosti a chaosu.*

Question 2.

All three codes have in common that they assign one codeword to each of ℓ-bit messages. There are 2^ℓ such messages, so for all cases, $M = 2^\ell$.

(a) $(m\ell, 2^\ell, m)$.

Obviously, $n = m\ell$ (as the original message is repeated m times) and $M = 2^\ell$.

As each message is copied m times, given two messages differing in x bits, corresponding codewords differ in mx bits. The minimum difference between messages is 1, so $d = m$.

(b) $(2\ell, 2^\ell, \min(\ell, 3))$.

Obviously, $n = 2\ell$ (as the original message of length ℓ is appended by a check bit for every neighboring pair, including the first-last pair, of which there are ℓ) and $M = 2^\ell$.

For two messages differing in x bits, the corresponding codewords differ in x bits + count of neighboring pairs of bits in the messages where one bit in the pair differs between messages and the second one does not (let's call this count y). This puts two upper bounds on minimum distance of codewords: $d \leq x + y = \ell + 0 = \ell$ (two codewords corresponding to two messages differing in all bits), and $d \leq x + y = 1 + 2 = 3$ for messages differing in one bit (yielding codewords that differ in one "message" bit and two "xor" bits).

For $\ell > 2$, d can't be any less than the upper bound, as codewords differing only in 1 or 2 bits are impossible, as seen from the definition of the "xor" part.

(c) $(\ell + (\ell(\ell-1))/2, 2^\ell, \ell)$.

Obviously, $n = \ell + (\ell(\ell-1))/2$ (an ℓ-bit message is appended by a checkbit for every pair of bits) and $M = 2^\ell$.

We now show that all pairs of codewords differ by at least ℓ bits. Let's take any two messages, which differ in x bits (meaning $\ell - x$ bits are the same). The corresponding checksums will differ in bits that correspond to pairs of message bits where one differs between messages and the second does not. That means $x(\ell - x)$ bits. The two codewords then differ by $x + x(\ell - x)$ bits, which is minimized for $x = \ell$ (this is the case when all the message bits differ, checksums are the same, which also gives a concrete example for upper bound of d), so $d = \ell$.

Question 3.

(a) $C_1 = \{001, 110\}$

(i) We need to find greater probabilities of the correct decoding for the corresponding codeword (we know that $p < \frac{1}{2}$):

i. $P(000 \mid 001) = (1-p)^2 \cdot p$
ii. $P(001 \mid 001) = (1-p)^3$
iii. $P(010 \mid 001) = (1-p) \cdot p^2$
iv. $P(011 \mid 001) = (1-p)^2 \cdot p$
v. $P(100 \mid 001) = (1-p) \cdot p^2$
vi. $P(101 \mid 001) = (1-p)^2 \cdot p$
vii. $P(110 \mid 001) = p^3$
viii. $P(111 \mid 001) = (1-p) \cdot p^2$
i. $P(000 \mid 110) = (1-p) \cdot p^2$
ii. $P(001 \mid 110) = p^3$
iii. $P(010 \mid 110) = (1-p)^2 \cdot p$
iv. $P(011 \mid 110) = (1-p) \cdot p^2$
v. $P(100 \mid 110) = (1-p)^2 \cdot p$
vi. $P(101 \mid 110) = (1-p) \cdot p^2$
vii. $P(110 \mid 110) = (1-p)^3$
viii. $P(111 \mid 110) = (1-p)^2 \cdot p$

The words $000, 001, 011, 101$ are decoded as codeword 001, while the words $010, 100, 110, 111$ are decoded as codeword 110.

(ii) The formula for the total probability is in this case the same for the both codewords:

$$P = (1-p)^3 + 3 \cdot (1-p)^2 \cdot p$$

If $p = 0.05$, then probabilities of the correct decoding are:

$$P(001) = P(110) = 0.95^3 + 3 \cdot 0.95^2 \cdot 0.05 = 0.99275$$

(b) $C_2 = \{0000, 1110, 0101\}$

(i) The same method is applied for C_2.

i. Words which cannot be uniquely determined: 0001, 0011, 0100 and 1001 (the same probability for more codewords).
ii. Words which are decoded as codeword 0000: 0000, 0010 and 1000.
iii. Words which are decoded as word 1110: 0110, 1010, 1011, 1100, 1110 and 1111.
iv. Words which are decoded as codeword 0101: 0101, 0111 and 1101.

(ii) The formulas for the total probabilities are:

$$\begin{aligned} P(0000) &= (1-p)^4 + 2 \cdot (1-p)^3 \cdot p \\ P(1110) &= (1-p)^4 + 4 \cdot (1-p)^3 \cdot p + (1-p)^2 \cdot p^2 \\ P(0101) &= (1-p)^4 + 2 \cdot (1-p)^3 \cdot p \end{aligned}$$

If $p = 0.05$, then probabilities of the correct decoding are:

$$\begin{aligned} P(0000) &= 0.95^4 + 2 \cdot 0.95^3 \cdot 0.05 \doteq 0.9 \\ P(1110) &= 0.95^4 + 4 \cdot 0.95^3 \cdot 0.05 + 0.95^2 \cdot 0.05^2 \doteq 0.988 \\ P(0101) &= 0.95^4 + 2 \cdot 0.95^3 \cdot 0.05 \doteq 0.9 \end{aligned}$$

Question 4.

(a) yes; C_2 can be obtained from C_1 by flipping all the bits

(b) no; C_1 contains a pair of codewords with hamming distance 4 (0011, 1100), but there is no such pair in C_2

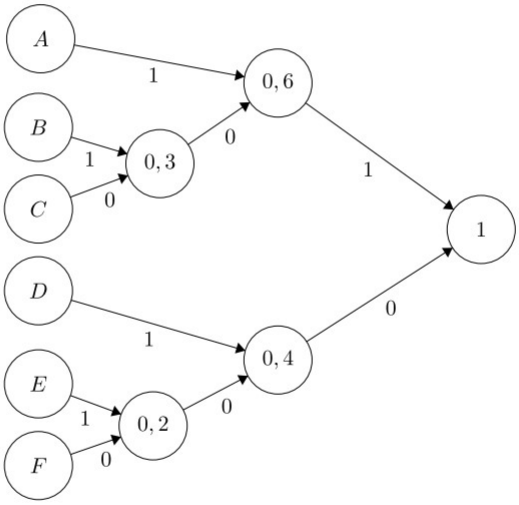
(c) no; different number of codewords (C_1 has 16, C_2 has 11)

Question 5.

(a) (i) Entropy for binary code: $S(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$
$-\frac{3}{10} * \log_2 \frac{3}{10} - 2 * \frac{3}{20} * \log_2 \frac{3}{20} - \frac{1}{5} * \log_2 \frac{1}{5} - 2 * \frac{1}{10} * \log_2 \frac{1}{10} = 2.470951$

(ii) Entropy for ternary code: $S(X) = -\sum_{x \in \mathcal{X}} p(x) \log_3 p(x)$
$-\frac{3}{10} * \log_3 \frac{3}{10} - 2 * \frac{3}{20} * \log_3 \frac{3}{20} - \frac{1}{5} * \log_3 \frac{1}{5} - 2 * \frac{1}{10} * \log_3 \frac{1}{10} = 1.559$

(b) (i) Huffman tree

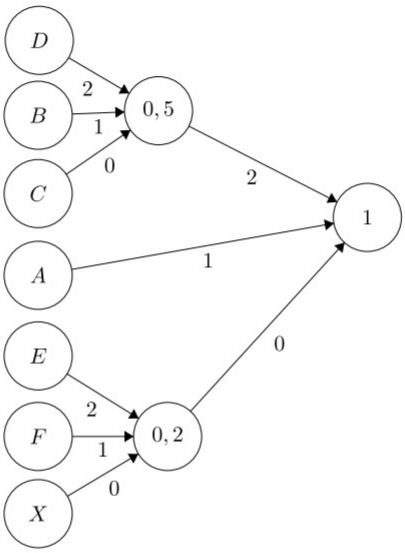


$A = 11$
$B = 101$
$C = 100$

$D = 01$
$E = 001$
$F = 000$

Average code length: $2 * 0.3 + 3 * 0.15 + 3 * 0.15 + 2 * 0.2 + 3 * 0.1 + 3 * 0.1 = 2.5$
Efficiency: $\frac{entropy}{avglength}$: $\frac{2.470951}{2.5} = 0.98838$

(ii) Huffman tree (X is added as dummy symbol)



$A = 1$
$B = 21$
$C = 20$
$D = 22$
$E = 02$
$F = 01$

Average code length: $0.3 + 2 * 0.2 + 2 * 0.15 + 2 * 0.15 + 2 * 0.1 + 2 * 0.1 = 1.7$
Efficiency: $\frac{entropy}{avglength}$: $\frac{1.559}{1.7} = 0.917$

By comparing efficiencies, we can see that binary code is more efficient than ternary code.

Question 6.

(a)

$$A_2(8,5) = 4$$

not a proof (wasn't required), only how did I get to my answer

- we can assume that $00000000 \in C$
- all other words have to have at least 5 1's.
- C can contain max 2 words with 5 1's (d = 5).
- C may contain at most 1 word with at least 6 ones
  - if $11111111 \in C$, than it cannot contain any other word (because d = 5) and M would be 2
  - if C contains word with 7 1's (e.g. 11111110), than it cannot contain any other word (because d = 5) and M would be 2
  - if C contains word with 6 1's (e.g. 11111100), than it could only contain 2 other words with 5 1's (e.g. 01010111 and 10101011) and M would be 4
  - if C would contain only 00000000 and words with 5 1's (max 2), then M would be 3

example of binary (8, 4, 5)-code:

$$C = \{00000000, 11111100, 01010111, 10101011\}$$

(b) there is no $A_q(n, d, w)$ if $d > 2w$.
Because there are no two words with weight w and distance d if $d > 2w$ Proof:

- let's take any two words v_1, v_2 of length n with weight w,
- let's say these word have x non-zero symbols on the same positions and y non-zero on different positions, where $0 \leq x \leq w \wedge 0 \leq y \leq w \wedge x + y = w$
  - if $x = 0$, then $y = w - x = w$, therefore none of non-zero symbols are on overlapping positions, therefore $h(v_1, v_2) = 2y + x = 2w$, contradiction to $d > 2w$
  - if $x = 1$, then $y = w - x = w - 1$ therefore exactly one non-zero symbol is on the same position in both words, therefore $h(v_1, v_2) \leq 2y + x \leq 2w - 1$, contradiction to $d > 2w$
    $\vdots$
  - if $x = k$, where $0 \leq k \leq w$ then $y = w - k$ therefore exactly k symbols are on overlapping positions, therefore $h(v_1, v_2) \leq 2y + x \leq 2w - 2k + k \leq 2w - k$, contradiction to $d > 2w$

  Therefore for all possible positions of non-zero symbols of any two words with weight w, is distance less or equal to $2w$