Mock exam (6 May 2021)

1. Let us consider three random variables $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$. Which of the following assertions are correct?
(At least one - but possibly several - answer is correct)

   (a) $I(\mathcal{X};\mathcal{Y}) \leq H(\mathcal{X})$
   ✔ Because $I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) - \underbrace{H(\mathcal{X}|\mathcal{Y})}_{\geq 0} \leq H(\mathcal{X})$.

   (b) $H(\mathcal{X},\mathcal{Y}|\mathcal{Z}) < H(\mathcal{X},\mathcal{Y})$
   ✘ We can have $H(\mathcal{X},\mathcal{Y}|\mathcal{Z}) = H(\mathcal{X},\mathcal{Y})$.

   (c) $\mathcal{Y} = g(\mathcal{X}) \Rightarrow H(\mathcal{Y}) \leq H(\mathcal{X})$
   ✔ If $g(\cdot)$ is the identity function, $H(\mathcal{Y}) = H(\mathcal{X})$ and $H(\mathcal{Y}) > H(\mathcal{X})$ is not possible.

   (d) $\mathcal{X} \not\perp \mathcal{Y} \Rightarrow I(\mathcal{X};\mathcal{Y}) < 0$
   ✘ $I(\mathcal{X};\mathcal{Y}) \geq 0$ (in all generality) and > 0 if $\mathcal{X} \not\perp \mathcal{Y}$.

   (e) $I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) < I(\mathcal{X};\mathcal{Y})$
   ✘ If $\mathcal{X} \perp \mathcal{Y}$ and $\mathcal{X} \perp \mathcal{Y}|\mathcal{Z}$, then $I(\mathcal{X};\mathcal{Y}) = 0 = I(\mathcal{X};\mathcal{Y}|\mathcal{Z})$. We can also imagine that $\mathcal{X} \perp \mathcal{Y}$ but $\mathcal{X} \not\perp \mathcal{Y}|\mathcal{Z}$ and thus $I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) > I(\mathcal{X};\mathcal{Y})$.

   (f) $H(\mathcal{X},\mathcal{Y}|\mathcal{Z}) \geq H(\mathcal{X}|\mathcal{Z})$
   ✔ because $H(\mathcal{X},\mathcal{Y}|\mathcal{Z}) = H(\mathcal{X}|\mathcal{Z}) + \underbrace{H(\mathcal{Y}|\mathcal{Z},\mathcal{X})}_{\geq 0}$.

2. Among the following codes, which one(s) **can not be** binary Huffman codes?
(At least one - but possibly several - answer is correct)

   (a) $\{0, 1\}$
   ✘ Trivially, this is a Huffman code with only two codewords.

   (b) $\{00, 01, 000\}$
   ✔ 00 is a prefix of 000.

   (c) $\{000, 001, 010, 011\}$
   ✔ The first 0 is useless, so this code can not be obtained from a Huffman tree.

   (d) $\{1, 01, 001, 0001, 0001, 00001, 00000\}$
   ✔ Two times '0001'.

   (e) $\{1, 10, 110, 1110\}$
   ✔ 1 is a prefix of other codewords.

3. Let us consider a sequence of discrete random variables $\mathcal{X}_1, \mathcal{X}_2, \ldots$ independent and identically distributed (i.i.d.) according to the distribution $P(\mathcal{X})$. The Asymptotic Equipartition Property states that

$$-\frac{1}{n}\log P(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n) \xrightarrow{P} H(\mathcal{X}). \quad (*)$$

   Which of the following assertions are correct?
   (At least one - but possibly several - answer is correct)

   (a) The self-information $-\log P(X_1, X_2, \ldots, X_n)$ of any message of length n of the source is equal to the source entropy times the length of the message.
   ✘ Not all of them.

   (b) This property $(*)$ is only valid if n is large enough.
   ✔ The convergence in probability requires n large enough.

   (c) This property $(*)$ is not valid if $P(\mathcal{X})$ is uniform.
   ✘ The property is also valid but the set of typical messages is the set of all messages.

   (d) The set of typical messages of length n are all messages $(X_1, X_2, \ldots, X_n)$ such that $-\frac{1}{n}\log P(X_1, X_2, \ldots, X_n)$ is close to $H(\mathcal{X})$ (up to a ϵ arbitrarily small).
   ✔ By definition of the typical set.

4. Let us consider a source S using a Q-ary alphabet, and a q-ary code. Which of the following assertions are correct?
(At least one - but possibly several - answer is correct)

   (a) If the codeword lengths of this code satisfy the Kraft inequality, then this code is uniquely decodable.
   ✘ We only know that there exists a code which is uniquely decodable with this set of codeword lengths.

   (b) If the code is instantaneous and such that $\sum_{k=1}^{s} r_k q^{-k} < 1$, we can add a codeword of any length to the code without losing the properties of the code (i.e., instantaneous code, ...).
   ✘ You can only add a codeword whose length is sufficiently large.

   (c) If the code is prefix free, then it is possible to fully characterize this code, i.e., determine whether this code is (or is not) regular, instantaneous, complete, and/or (absolutely) optimal without computing anything else.
   ✘ No, it is not possible to know if this code is complete and/or optimal.

   (d) If the code is complete, then the Kraft inequality is necessarily verified for the codeword lengths.
   ✔ Of course, and we have the equality.

   (e) If q is smaller than Q then the compression rate is smaller than 1.
   ✘ The compression rate does not only depend on the ratio between Q and q.

5. Let us consider a stationary and memoryless source $\mathcal{S}$ with an alphabet of five symbols $\{a, b, c, d, e\}$ and the following probability distribution $P(\mathcal{S}) = [0.5,\ 0.2,\ 0.1,\ 0.1,\ 0.1]$. Which of the following code(s) is (are) the most interesting (with regards of decodability and data compression):
(At least one - but possibly several - answer is correct)

   (a) $\{0,\ 100,\ 101,\ 110,\ 111\}$
   ✔ Huffman with $\bar{n}_h = 2$ and so compression rate $= \dfrac{1 \log_2 5}{2 \log_2 2} = 1.16$

   (b) $\{0, 1, 2, 3, 4\}$
   ✘ Compression rate $= 1$

   (c) $\{05,\ 02,\ 01,\ 0101\}$
   ✘ Obvious because of the useless starting 0.

   (d) $\{0,\ 10,\ 110,\ 1110,\ 1111\}$
   ✔ Huffman and compression rate $= 1.16$.

   (e) $\{0, 11, 100, 101, 110\}$
   ✘ '11' is a prefix of '110'.

6. If two random variables $\mathcal{X}$ and $\mathcal{Y}$ have the following probability distributions $P(\mathcal{X}) = [0.2,\ 0.8]$ and $P(\mathcal{Y}) = [0,\ 0.2,\ 0.8]$, then...
(Only one answer is correct)

   (a) $H(\mathcal{X}) > H(\mathcal{Y})$
   ✘

   (b) $H(\mathcal{X}) \geq H(\mathcal{Y})$
   ✘

   (c) $H(\mathcal{X}) = H(\mathcal{Y})$
   ✔ This is obvious since $0 \log_2(0) = 0$.

   (d) $H(\mathcal{X}) \leq H(\mathcal{Y})$
   ✘

   (e) $H(\mathcal{X}) < H(\mathcal{Y})$
   ✘

7. Let us consider the game "sixty-three" that consists in guessing an integer $x \in [0, \ldots, 63]$ by asking yes/no questions (e.g., "Is x greater than 10?","Is x equal to 5?"). Which of the following assertions are correct? *(At least one - but possibly several - answer is correct)*

   (a) You reduce more your uncertainty (about the true value of x) by guessing x in only one question than by guessing x in several questions.
   ✘ Once x is guessed, you have the same amount of the information (6 bits).

   (b) You reduce more your uncertainty (about the true value of x) in total for the game by dividing by systematically dividing the candidate set size (i.e., the set of all possible values for x) by two, than by applying another strategy to guess the value of x.
   ✘ No, not in total.

   (c) Assuming that the answer is "yes" to your first question, you reduce more your uncertainty (about the true value of x) by asking as first question "is x smaller than 32?" than "is x equal to 32?".
   ✘ Knowing that the answer is yes, $\log_2(1/2) < \log_2(1/64)$ so no.

   (d) The best strategy to reduce the most your uncertainty (about the true value of x) on average is to systematically divide the candidate set size (i.e., the set of all possible values for x) by two.
   ✔ On average you have 1 bit of information by doing so.

8. Let us consider a stationary and memoryless source $\mathcal{S}$ which uses an alphabet of six different symbols with the following probability distribution $P(\mathcal{S}) = [0.3, 0.2, 0.2, 0.1, 0.1, 0.1]$. How many different binary Huffman codes can be built for $\mathcal{S}$?
   (Only one answer is correct)

   (a) 1
   ✘

   (b) 5
   ✘

   (c) 30
   ✘

   (d) 32
   ✘

   (e) 192
   ✘

   (f) 288
   ✔ because $C_3^2(for\ 0.10) \times C_3^1(for\ 0.20) \times 2^5$.

9. Let us consider a source $\mathcal{S}$ with an alphabet of four symbols $\{a, b, c, d\}$ with the following probability distribution [0.25, 0.25, 0.25, 0.25]. Let us consider that we have found an optimal uniquely decodable code (with $q = 2$) for this source. If $\bar{n}$ denotes the expected length per source symbol of this code, we have...
(At least one - but possibly several - answer is correct)

   (a) $\bar{n} = H(\mathcal{S})$
   ✔ Because we apply Huffman to find an optimal code, we have $\{00, 01, 10, 11\}$ and so $\bar{n} = 2 = H(\mathcal{S})$.

   (b) $\bar{n} = H(\mathcal{S}) + 1$
   ✘ because $H(\mathcal{S}) + 1 = 3 > \bar{n}$.

   (c) $\bar{n} > H(\mathcal{S})$
   ✘ because $\bar{n} = H(\mathcal{S})$.

   (d) $\bar{n} \geq H(\mathcal{S})$
   ✔ because optimal code and $\bar{n} = H(\mathcal{S})$.

   (e) $\bar{n} > H(\mathcal{S}) + 1$
   ✘ This is never the case for an optimal code anyway.

   (f) $\bar{n} \geq H(\mathcal{S}) + 1$
   ✘ This is never the case for an optimal code anyway.

   (g) $\bar{n} < H(\mathcal{S})$
   ✘ This is never the case for an optimal code anyway.

   (h) $\bar{n} < H(\mathcal{S}) + 1$
   ✔ This is always the case for an optimal code and $\bar{n} = 2 < H(\mathcal{S}) + 1$.

   (i) $\bar{n} \leq H(\mathcal{S})$
   ✔ For the equality

   (j) $\bar{n} \leq H(\mathcal{S}) + 1$
   ✔ For "<".

10. Let us consider the code $\{00, 11, 0101, 111, 1010, 100100, 0110\}$, with the following probability distribution [0.1, 0.1, 0.3, 0.1, 0.2, 0.1, 0.1]. Which of the following assertions are correct?
(At least one - but possibly several - answer is correct)

   (a) The code is regular (non-singular).
   ✔ Yes, all codewords are different.

   (b) The code is uniquely decodable
   ✘ Because, e.g., 2 ×'111' = 3 × '11'.

   (c) The code is instantaneous.
   ✘ Because not prefix free.

   (d) The code is complete
   ✘ Kraft < 1.

   (e) The code is prefix free.
   ✘ No, '11' is prefix of '111'.

   (f) The code is optimal.
   ✘ $\bar{n} = 3.7 > H + 1 = 3.64$

   (g) The code is absolutely optimal.
   ✘ $\bar{n} = 3.7 > H = 2.64$

   (h) Kraft inequality is satisfied.
   ✔ Using codeword lengths, we have $\sum_{i=1}^{Q} 2^{-n_i} = 0.828 \leq 1$.

   (i) All extensions of this code are not regular (non-singular).
   ✘ Because at least one extension (the code itself) is regular.

11. What is the name of the algorithm to build an optimal symbol code?
✔ Huffman