Exam in information theory 31.01.2023. Problems

Problem 1

We consider random variables A and B, taking their values in the set $\{0,1\}^n$, for some $n \geq 1$, where

$$\Pr(A \neq B) \quad \leq \quad \frac{1}{n}.$$

Prove that

$$H(A \mid B) \quad \leq \quad 2,$$

and indicate, for which n (if any) the equality holds.

Hint. For words $v, w \in \{0,1\}^*$, let

$$\mathit{diff}\,(v,w) \quad = \quad \begin{cases} 0 & \text{if} \quad v = w \\ 1 & \text{if} \quad v \neq w \end{cases}$$

It may be helpful to introduce a random variable D, defined by

$$D \quad = \quad \mathit{diff}\,(A,B),$$

and consider $H(A, D \mid B)$.

Solution.

We have

$$\begin{aligned} H(A,D \mid B) &= H(A \mid B) + \overbrace{H(D \mid A,B)}^{0} \\ &= \underbrace{H(D \mid B)}_{\leq 1} + H(A \mid B, D) \end{aligned}$$

because D is a function of A and B, and takes only 2 values. Now examine possible values of $H(A \mid b,d)$. If $d = 0$ then A equals B, hence $H(A \mid b,d) = 0$. If $d = 1$ then A can take only values different from b, hence

$$H(A \mid b, 1) \leq \log(2^n - 1) < n.$$

Thus

$$H(A \mid B, D) = \sum_b H(A \mid b, 1) \cdot \Pr(B = b \wedge D = 1) < n \cdot \sum_b \Pr(B = b \wedge D = 1) < n \cdot \underbrace{\Pr(D = 1)}_{\leq \frac{1}{n}} \leq 1$$

where inequality $\Pr(D = 1) \leq \frac{1}{n}$ follows from the assumption. From the above, we obtain

$$\begin{aligned} H(A \mid B) &= \underbrace{H(D \mid B)}_{\leq 1} + \underbrace{H(A \mid B, D)}_{<1} \\ &< 2; \end{aligned}$$

in particular, the equality never holds.

Problem 2

Let $(w_n)_{n\in\mathbb{N}}$ be a sequence of different words that are random in the sense of Kolmogorov, that is $C_U(w_n) \geq n$, for some universal Turing machine U. Prove that infinitely many words in this sequence contains a subword 111.

Hint. It may be helpful to first consider the case when the length of w_n is divisible by 3.

Bonus. Propose and prove a generalization of the task of this problem.

Solution.

Any word $w \in \{0,1\}^*$ can be presented as a concatenation $w = \alpha_1\alpha_2\ldots\alpha_k\beta$, where $|\alpha_i| = 3$, for $i = 1,\ldots,k$, and $0 \leq |\beta| \leq 2$. Let $I(n)$ be the set of all words w of length n, such that in the presentation as above none of the blocks α_i is 111. We will show that the set $\bigcup_n I(n)$ contains only finitely many random words. Note that this implies that our sequence satisfies an even stronger property: almost all words w_n contain 111 as a block starting from a position $3i+1$, for some i.

Let us assume that $n = 3\cdot k + d$, where $1 \leq k$, $0 \leq d \leq 2$. Note that the number of possible words of length d (including the empty word) is 7, as is the number of 3-bit blocks different from 111. Then $m_n \stackrel{def}{=} |I(n)| \leq 7^{k+1}$, and we can list all words in $I(n)$ in the lexicographical order, say $v_1^n,\ldots,v_{m_n}^n$. Now we can construct a Turing machine T, which, given a binary representation of n and i (where $i \leq m_n$), generates the word v_i^n on the list defined above. Note that n can be represented by $\lfloor\log n\rfloor + 1$ bits, and i by at most $\lfloor(k+1)\log 7\rfloor + 1 \leq \frac{\log 7}{3}\cdot n + 3$ bits. We need to apply some encoding of pairs [1] (as explained at the tutorials), but altogether T can generate v_i^n from an input whose length is bounded by

$$2\log n + \frac{\log 7}{3}\cdot n + c.$$

Since $\frac{\log 7}{3} < 1$ this clearly implies that there is a constant $\varepsilon > 0$, such that, for sufficiently large n, for any word $v \in I(n) \subseteq \{0,1\}^n$,

$$C_U(v) \quad\leq\quad n - \varepsilon,$$

hence v is not random. This completes the proof.

We can generalize the thesis by taking any word u of length $t \geq 1$, instead of 111. The claim will follow by a similar computation, which is based on the fact that $\log(2^t - 1)$ is strictly smaller than t.

[1] To avoid pairs, we could cleverely encode just v_i^n using an appropriately chosen number of bits $< n$.