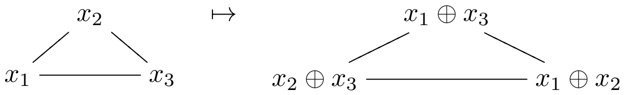
Problem 1

Let X_1, X_2, X_3 be **independent** random variables taking values in the set $\{0, 1\}$ with the same probability distribution $\Pr(X_i = 0) = p$, where $0 < p < 1$. The symbol $\oplus$ means addition modulo 2 (XOR). Please compare the following values

$$H(X_1, X_2, X_3) \qquad H(X_2 \oplus X_3, X_1 \oplus X_3, X_1 \oplus X_2)$$

$$I(X_1; X_2|X_3) \qquad I(X_2 \oplus X_3; X_1 \oplus X_3, |X_1 \oplus X_2).$$

Remark. To help intuitions, the operation considered above can be illustrated on a triangle



that is, the value in each node is replaced by the $\oplus$ of its neighbours.

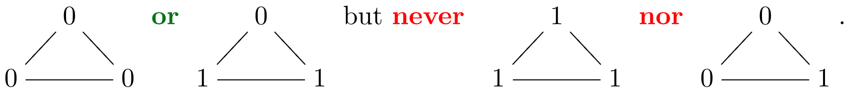
Note. In case of difficulties, please solve the problem for $p = \frac{1}{2}$. For the general case, use the Venn diagram and explore symmetry of the problem while avoiding long calculations.

Solution

We consider general case. As X_1, X_2, X_3 are independent, the entropy $H(X_1, X_2, X_3)$ achieves the maximal value (equal to $3 \cdot H(p)$). To proceed further, let us abbreviate

$$Y_1 = X_2 \oplus X_3, \; Y_2 = X_1 \oplus X_3, \; Y_3 = X_1 \oplus X_2.$$

Note that the product variable (Y_1, Y_2, Y_3) is a function of (X_1, X_2, X_3), but **not** *vice versa*, because the latter variable assumes **less** values. Indeed, while looking at the triangle, we immediately see that we can obtain only (up to permutation)



Hence

$$H(Y_1, Y_2, Y_3) \quad < \quad H(X_1, X_2, X_3).$$

Remark. An alternative argument stems from the fact that the variables Y_1, Y_2, Y_3 are **not** independent (as, e.g., $(X_2 \oplus X_3) \oplus (X_1 \oplus X_3) = X_1 \oplus X_2$), and moreover $H(Y_i) \leq H(X_i)$. The last inequality follows from examining the respective Bernoulli distributions[1]. Hence we have

$$H(Y_1, Y_2, Y_3) < H(Y_1) + H(Y_2) + H(Y_3) \leq H(X_1) + H(X_2) + H(X_3) = H(X_1, X_2, X_3).$$

Now observe that $I(X_1; X_2|X_3) = 0$, because of the independence, while

$$I(Y_1; Y_2|Y_3) = H(Y_1|Y_3) - \underbrace{H(Y_1|Y_2, Y_3)}_{0} > 0, \tag{1}$$

as Y_1 is not a function of Y_3. We further explore the equality (1) to show that $I(Y_1; Y_2|Y_3) < H(Y_1, Y_2, Y_3)$. Indeed, we have

$$I(Y_1; Y_2|Y_3) = H(Y_1|Y_3) = H(Y_1, Y_3) - \underbrace{H(Y_3)}_{>0} < H(Y_1, Y_3) \leq H(Y_1, Y_2, Y_3)$$

[1] Letting $q = \Pr(X_2 \oplus X_3 = 0) = p^2 + (1-p)^2$, we have $p = q$ if $p = \frac{1}{2}$, but otherwise the new distribution is “more polarised”, i.e., either $q < p < \frac{1}{2}$ or $\frac{1}{2} < p < q$. Hence $H(q) \leq H(p)$ is any case, with equality only for $p = \frac{1}{2}$.

(the last inequality is in fact equality, as $Y_2 = Y_1 \oplus Y_3$). Thus we obtain the **strict** ordering

$$I(X_1; X_2|X_3) < I(X_2 \oplus X_3; X_1 \oplus X_3, |X_1 \oplus X_2) < H(X_2 \oplus X_3, X_1 \oplus X_3, X_1 \oplus X_2) < H(X_1, X_2, X_3).$$

For $p = \frac{1}{2}$, these numbers are

$$0 < 1 < 2 < 3.$$

Problem 2

We consider two channels whose input and output alphabet is $\{0,1\}^n$. Channel Γ_1 inputs a word w and with probability $\frac{1}{2}$ outputs it correctly, or outputs its mirror [2] image w^R.

Whereas channel Γ_2 inputs a word w and with probability $\frac{1}{2}$ outputs it correctly, or swaps its **first** bit. For example, with $n = 7$,

$$\begin{tikzcd}
\Gamma_1 : & 1101001 \arrow[r] \arrow[dr] & 1101001 & \Gamma_2 : & 1101001 \arrow[r] \arrow[dr] & 1101001 \\
& & 1001011 & & & 0101001
\end{tikzcd}$$

Compare the capacities of the two channels.

Note. The argument may depend on the parity of n.

Solution

The capacity of Γ_2 is straightforward to compute. Each row in the matrix channel contains two non-zero values of $\frac{1}{2}$, and similarly each column. Therefore $H(B|A) = 1$ independently of the distribution of A, and B is uniform if so is A. Hence $C(\Gamma_2) = n - 1$.

We will show that, in the channel Γ_1, we can achieve strictly more. Note that the shape of a row now depends on whether the input word w is a palindrome (i.e., $w^R = w$) or not. In the former case, it has only one non-zero value $P(w \to w) = 1$, and the entropy of the row $H(B|A = w) = 0$, whereas in the latter case it admits twice $\frac{1}{2}$, and $H(B|A = w) = 1$. As every column sums up to 1, we again see that B is uniform if so is A, hence $H(B) = n$. On the other hand, $H(B|A) < 1$, because we always have $H(B|A = w) \leq 1$ and $H(B|A = w) = 0$ holds with probability $\Pr(A \text{ is a palindrom }) > 0$. Hence, without even checking if $I(A; B)$ is maximal in this case, we see

$$C(\Gamma_1) \geq I(A; B) > n - 1 = C(\Gamma_2).$$

[2] For $w = w_1 w_2 \dots w_n$, $w^R = w_n w_{n-1} \dots w_1$.